Solutions to Axler's Linear Algebra Done Right

Aayush Bajaj

2026-08-29T13:00:00+10:00

Contents

1 Vector Spaces **3**
1.1 Exercises 1A . . . 4
1.2 Exercises 1B . . . 13
1.3 Exercises 1C . . . 20

2 Finite-Dimensional Vector Spaces **35**
2.1 Exercises 2A . . . 36
2.2 Exercises 2B . . . 48
2.3 Exercises 2C . . . 56

3 Linear Maps **70**
3.1 Exercises 3A . . . 71
3.2 Exercises 3B . . . 84
3.3 Exercises 3C . . . 105
3.4 Exercises 3D . . . 121
3.5 Exercises 3E . . . 139
3.6 Exercises 3F . . . 155

4 Polynomials **179**
4.1 Exercises . . . 180

5 Eigenvalues and Eigenvectors **191**
5.1 Exercises 5A . . . 192
5.2 Exercises 5B . . . 221
5.3 Exercises 5C . . . 244
5.4 Exercises 5D . . . 257
5.5 Exercises 5E . . . 277

6 Inner Product Spaces **287**
6.1 Exercises 6A . . . 288
6.2 Exercises 6B . . . 318
6.3 Exercises 6C . . . 342

7 Operators on Inner Product Spaces **361**
7.1 Exercises 7A . . . 362
7.2 Exercises 7B . . . 388
7.3 Exercises 7C . . . 407
7.4 Exercises 7D . . . 425
7.5 Exercises 7E . . . 444
7.6 Exercises 7F . . . 459

8 Operators on Complex Vector Spaces **486**
8.1 Exercises 8A . . . 487
8.2 Exercises 8B . . . 504
8.3 Exercises 8C . . . 526
8.4 Exercises 8D . . . 539

9 Multilinear Algebra and Determinants **551**
9.1 Exercises 9A . . . 552
9.2 Exercises 9B . . . 560
9.3 Exercises 9C . . . 567
9.4 Exercises 9D . . . 586

Solutions to every exercise in the fourth edition of Sheldon Axler's *Linear Algebra Done Right* (Springer UTM, 2024) — 732 exercises across sections 1A–9D. The book is open access at linear.axler.net and is filed at Linear Algebra Done Right (Axler).

1 Vector Spaces

Contents

1.1 Exercises 1A 4
1.2 Exercises 1B 13
1.3 Exercises 1C 20

1.1 Exercises 1A

Problem 1A.1. Show that $\alpha + \beta = \beta + \alpha$ for all $\alpha, \beta \in \mathbf{C}$.

Solution. By 1.1, every complex number is by definition an ordered pair of real numbers written in the form $a + bi$, and addition on $\mathbf{C}$ is defined by

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$

where $a, b, c, d \in \mathbf{R}$.
So suppose $\alpha, \beta \in \mathbf{C}$, and write

$$\alpha = a + bi, \qquad \beta = c + di$$

with $a, b, c, d \in \mathbf{R}$. Then the definition of complex addition gives

$$\alpha + \beta = (a + bi) + (c + di) = (a + c) + (b + d)i$$

and, applying the same definition with the roles of the two numbers interchanged,

$$\beta + \alpha = (c + di) + (a + bi) = (c + a) + (d + b)i.$$

Now a, b, c, d are real numbers, and addition of real numbers is commutative. Hence $a + c = c + a$ and $b + d = d + b$. Therefore the two complex numbers displayed above have the same real part and the same imaginary part, so by the definition of equality of complex numbers as ordered pairs of reals (1.1 together with 1.8, since two lists are equal exactly when they have the same entries in the same order),

$$\alpha + \beta = (a + c) + (b + d)i = (c + a) + (d + b)i = \beta + \alpha.$$

Since α and β were arbitrary elements of $\mathbf{C}$, this proves the commutativity of addition asserted in 1.3.

Problem 1A.2. Show that $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

Solution. Suppose $\alpha, \beta, \lambda \in \mathbf{C}$, and write

$$\alpha = a + bi, \qquad \beta = c + di, \qquad \lambda = e + fi$$

with $a, b, c, d, e, f \in \mathbf{R}$.
Applying the definition of complex addition from 1.1 twice, first to α and β and then to the result and λ, we get

$$\begin{aligned}(\alpha + \beta) + \lambda &= \big((a + bi) + (c + di)\big) + (e + fi) \\ &= \big((a + c) + (b + d)i\big) + (e + fi) \\ &= \big((a + c) + e\big) + \big((b + d) + f\big)i.\end{aligned}$$

Applying the same definition in the other grouping, first to β and λ and then to α and the result, we get

$$\begin{aligned}\alpha + (\beta + \lambda) &= (a + bi) + \big((c + di) + (e + fi)\big) \\ &= (a + bi) + \big((c + e) + (d + f)i\big) \\ &= \big(a + (c + e)\big) + \big(b + (d + f)\big)i.\end{aligned}$$

The numbers a, c, e are real and addition of real numbers is associative, so $(a + c) + e = a + (c + e)$. Likewise b, d, f are real, so $(b + d) + f = b + (d + f)$. Thus the two complex numbers computed above have the same real part and the same imaginary part, and hence are equal:

$$(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda).$$

Since α, β, λ were arbitrary elements of $\mathbf{C}$, this proves the associativity of addition asserted in 1.3.

Problem 1A.3. Show that $(\alpha\beta)\lambda = \alpha(\beta\lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

Solution. Recall from 1.1 that multiplication on $\mathbf{C}$ is defined by

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i,$$

where $a, b, c, d \in \mathbf{R}$.
Suppose $\alpha, \beta, \lambda \in \mathbf{C}$, and write

$$\alpha = a + bi, \qquad \beta = c + di, \qquad \lambda = e + fi$$

with $a, b, c, d, e, f \in \mathbf{R}$.
First compute $(\alpha\beta)\lambda$. The definition of multiplication gives $\alpha\beta = (ac - bd) + (ad + bc)i$, and applying the definition again with the real number $ac - bd$ in place of a and $ad + bc$ in place of b yields

$$\begin{aligned}(\alpha\beta)\lambda &= \big((ac - bd) + (ad + bc)i\big)(e + fi) \\ &= \big((ac - bd)e - (ad + bc)f\big) + \big((ac - bd)f + (ad + bc)e\big)i \\ &= (ace - bde - adf - bcf) + (acf - bdf + ade + bce)i,\end{aligned}$$

where the last step uses only the distributive and associative laws for real numbers.
Now compute $\alpha(\beta\lambda)$. The definition of multiplication gives $\beta\lambda = (ce - df) + (cf + de)i$, so

$$\begin{aligned}\alpha(\beta\lambda) &= (a + bi)\big((ce - df) + (cf + de)i\big) \\ &= \big(a(ce - df) - b(cf + de)\big) + \big(a(cf + de) + b(ce - df)\big)i \\ &= (ace - adf - bcf - bde) + (acf + ade + bce - bdf)i.\end{aligned}$$

Comparing the two results, the real parts are

$$ace - bde - adf - bcf \qquad \text{and} \qquad ace - adf - bcf - bde,$$

which are equal because addition of real numbers is commutative and associative, so the same four real summands may be rearranged freely. The imaginary parts are

$$acf - bdf + ade + bce \qquad \text{and} \qquad acf + ade + bce - bdf,$$

which are equal for the same reason.
Hence $(\alpha\beta)\lambda$ and $\alpha(\beta\lambda)$ have the same real part and the same imaginary part, so they are equal. Since α, β, λ were arbitrary elements of $\mathbf{C}$, this proves the associativity of multiplication asserted in 1.3.

Problem 1A.4. Show that $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbf{C}$.

Solution. Suppose $\lambda, \alpha, \beta \in \mathbf{C}$, and write

$$\lambda = e + fi, \qquad \alpha = a + bi, \qquad \beta = c + di$$

with $a, b, c, d, e, f \in \mathbf{R}$.
By the definition of complex addition in 1.1 we have $\alpha + \beta = (a + c) + (b + d)i$, so the definition of complex multiplication in 1.1 gives

$$\begin{aligned}\lambda(\alpha + \beta) &= (e + fi)\big((a + c) + (b + d)i\big) \\ &= \big(e(a + c) - f(b + d)\big) + \big(e(b + d) + f(a + c)\big)i \\ &= (ea + ec - fb - fd) + (eb + ed + fa + fc)i,\end{aligned}$$

where the last equality uses the distributive law for real numbers.
On the other hand, the definition of complex multiplication gives

$$\lambda\alpha = (ea - fb) + (eb + fa)i \qquad \text{and} \qquad \lambda\beta = (ec - fd) + (ed + fc)i,$$

so by the definition of complex addition,

$$\lambda\alpha + \lambda\beta = \big((ea - fb) + (ec - fd)\big) + \big((eb + fa) + (ed + fc)\big)i.$$

The real part $(ea-fb)+(ec-fd)$ equals $ea+ec-fb-fd$, and the imaginary part $(eb+fa)+(ed+fc)$ equals $eb + ed + fa + fc$; in each case the two expressions are sums of the same four real numbers, hence equal by the commutativity and associativity of addition in $\mathbf{R}$.
Therefore $\lambda(\alpha + \beta)$ and $\lambda\alpha + \lambda\beta$ have the same real part and the same imaginary part, so

$$\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta.$$

Since λ, α, β were arbitrary elements of $\mathbf{C}$, this proves the distributive property asserted in 1.3.

Problem 1A.5. Show that for every $\alpha \in \mathbf{C}$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha + \beta = 0$.

Solution. Throughout, 0 denotes the complex number $0 + 0i$; this is consistent with the identification in 1.1 of $a + 0i$ with the real number a. The proof below works directly from the definitions in 1.1 and assumes no part of 1.3, since 1.3 is what this exercise and its companions establish. (The optional remark at the end does invoke the identity clause $\lambda + 0 = \lambda$ of 1.3, but that clause is not among the properties left to these exercises, and it is immediate from 1.1.)
Suppose $\alpha \in \mathbf{C}$ and write $\alpha = a + bi$ with $a, b \in \mathbf{R}$.
Existence. Let $\beta = (-a)+(-b)i$, which is an element of $\mathbf{C}$ because $-a$ and $-b$ are real numbers. Then the definition of complex addition in 1.1 gives

$$\alpha + \beta = (a + bi) + \big((-a) + (-b)i\big) = \big(a + (-a)\big) + \big(b + (-b)\big)i = 0 + 0i = 0.$$

Uniqueness. Suppose $\beta \in \mathbf{C}$ satisfies $\alpha + \beta = 0$, and write $\beta = c + di$ with $c, d \in \mathbf{R}$. Then

$$0 + 0i = \alpha + \beta = (a + c) + (b + d)i.$$

Two complex numbers are equal precisely when they are equal as ordered pairs of real numbers, that is, when their real parts agree and their imaginary parts agree (1.1). Hence

$$a + c = 0 \qquad \text{and} \qquad b + d = 0.$$

Since a, b, c, d are real numbers and each real number has a unique additive inverse in $\mathbf{R}$, these two equations force $c = -a$ and $d = -b$. Thus $\beta = (-a) + (-b)i$, which is exactly the element exhibited in the existence step.
Therefore for every $\alpha \in \mathbf{C}$ there is one and only one $\beta \in \mathbf{C}$ with $\alpha + \beta = 0$, namely $\beta = (-a) + (-b)i$. This proves the additive inverse property asserted in 1.3 and justifies the notation $-\alpha$ introduced in 1.5.
Remark. Uniqueness also follows abstractly from the properties already established, without coordinates: if $\alpha + \beta = 0$ and $\alpha + \beta' = 0$, then using Exercise 1A.1 (commutativity), Exercise 1A.2 (associativity), and the additive identity property from 1.3,

$$\beta = \beta + 0 = \beta + (\alpha + \beta') = (\beta + \alpha) + \beta' = (\alpha + \beta) + \beta' = 0 + \beta' = \beta'.$$

Problem 1A.6. Show that for every $\alpha \in \mathbf{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha\beta = 1$.

Solution. Suppose $\alpha \in \mathbf{C}$ with $\alpha \neq 0$, and write $\alpha = a + bi$ with $a, b \in \mathbf{R}$. Here 1 denotes the complex number $1 + 0i$ and 0 denotes $0 + 0i$.
First note that $a^2 + b^2 > 0$. Indeed, since $\alpha \neq 0$, the ordered pair (a, b) is not $(0, 0)$, so at least one of a, b is nonzero; squares of real numbers are nonnegative, and the square of a nonzero real number is positive, so $a^2 + b^2 > 0$. In particular we may divide real numbers by $a^2 + b^2$.

Existence. Let

$$\beta = \frac{a}{a^2+b^2} + \frac{-b}{a^2+b^2}\,i,$$

which lies in $\mathbf{C}$ because both coefficients are real numbers. By the definition of complex multiplication in 1.1,

$$\begin{aligned}\alpha\beta &= \left(a\cdot\frac{a}{a^2+b^2} - b\cdot\frac{-b}{a^2+b^2}\right) + \left(a\cdot\frac{-b}{a^2+b^2} + b\cdot\frac{a}{a^2+b^2}\right)i \\ &= \frac{a^2+b^2}{a^2+b^2} + \frac{-ab+ab}{a^2+b^2}\,i \\ &= 1 + 0i = 1.\end{aligned}$$

Uniqueness. Suppose $\beta \in \mathbf{C}$ satisfies $\alpha\beta = 1$, and write $\beta = c + di$ with $c, d \in \mathbf{R}$. By the definition of complex multiplication,

$$1 + 0i = \alpha\beta = (ac - bd) + (ad + bc)i,$$

and equality of complex numbers means equality of real and imaginary parts, so

$$ac - bd = 1 \qquad \text{and} \qquad ad + bc = 0.$$

Multiply the first equation by a and the second by b and add:

$$a^2c - abd + abd + b^2c = a, \qquad \text{that is,} \qquad (a^2+b^2)c = a.$$

Multiply the first equation by $-b$ and the second by a and add:

$$-abc + b^2d + a^2d + abc = -b, \qquad \text{that is,} \qquad (a^2+b^2)d = -b.$$

Since $a^2 + b^2 \neq 0$, these determine c and d uniquely:

$$c = \frac{a}{a^2+b^2}, \qquad d = \frac{-b}{a^2+b^2}.$$

Thus β must be the complex number exhibited in the existence step, and no other element of $\mathbf{C}$ can satisfy $\alpha\beta = 1$.
Therefore every nonzero $\alpha \in \mathbf{C}$ has exactly one multiplicative inverse, proving the multiplicative inverse property asserted in 1.3 and justifying the notation $1/\alpha$ introduced in 1.5.
Remark. Uniqueness again also follows abstractly: if $\alpha\beta = 1$ and $\alpha\beta' = 1$, then using the commutativity of complex multiplication proved in 1.4, the associativity proved in Exercise 1A.3, and the multiplicative identity property from 1.3,

$$\beta = \beta\cdot 1 = \beta(\alpha\beta') = (\beta\alpha)\beta' = (\alpha\beta)\beta' = 1\cdot\beta' = \beta'.$$

Problem 1A.7. Show that

$$\frac{-1+\sqrt{3}\,i}{2}$$

is a cube root of 1 (meaning that its cube equals 1).

Solution. Let

$$\omega = \frac{-1+\sqrt{3}\,i}{2}.$$

First we identify ω in the standard form $a + bi$ with $a, b \in \mathbf{R}$. By the definition of division in 1.5, $\omega = (-1+\sqrt{3}\,i)\big(1/2\big)$, where $1/2$ is the multiplicative inverse of the complex number $2 = 2 + 0i$; that inverse is the real number $\frac{1}{2}$, viewed as $\frac{1}{2} + 0i$, since $(2+0i)(\frac{1}{2}+0i) = 1 + 0i$. Hence by the definition of complex multiplication in 1.1,

$$\omega = (-1+\sqrt{3}\,i)\left(\tfrac{1}{2}+0i\right) = \left(-1\cdot\tfrac{1}{2} - \sqrt{3}\cdot 0\right) + \left(-1\cdot 0 + \sqrt{3}\cdot\tfrac{1}{2}\right)i = -\frac{1}{2} + \frac{\sqrt{3}}{2}\,i.$$

By the definition of powers given in Section 1A, $\omega^3 = \omega\omega\omega$; the associativity proved in Exercise 1A.3 shows that the grouping is irrelevant, so we may compute $\omega^3 = (\omega^2)\omega$.
Compute ω^2 using the definition of complex multiplication, with $a = -\frac{1}{2}$ and $b = \frac{\sqrt{3}}{2}$:

$$\begin{aligned}
\omega^2 &= \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}\,i\right)\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}\,i\right) \\
&= \left(\left(-\frac{1}{2}\right)\left(-\frac{1}{2}\right) - \frac{\sqrt{3}}{2}\cdot\frac{\sqrt{3}}{2}\right) + \left(\left(-\frac{1}{2}\right)\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2}\left(-\frac{1}{2}\right)\right) i \\
&= \left(\frac{1}{4} - \frac{3}{4}\right) + \left(-\frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{4}\right) i \\
&= -\frac{1}{2} - \frac{\sqrt{3}}{2}\,i,
\end{aligned}$$

where we used $\sqrt{3}\cdot\sqrt{3} = 3$.
Now multiply by ω once more:

$$\begin{aligned}
\omega^3 &= \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}\,i\right)\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}\,i\right) \\
&= \left(\left(-\frac{1}{2}\right)\left(-\frac{1}{2}\right) - \left(-\frac{\sqrt{3}}{2}\right)\frac{\sqrt{3}}{2}\right) + \left(\left(-\frac{1}{2}\right)\frac{\sqrt{3}}{2} + \left(-\frac{\sqrt{3}}{2}\right)\left(-\frac{1}{2}\right)\right) i \\
&= \left(\frac{1}{4} + \frac{3}{4}\right) + \left(-\frac{\sqrt{3}}{4} + \frac{\sqrt{3}}{4}\right) i \\
&= 1 + 0i = 1.
\end{aligned}$$

Thus $\omega^3 = 1$, so $\omega = \dfrac{-1 + \sqrt{3}\,i}{2}$ is a cube root of 1.
Remark. Note that $\omega \neq 1$, since ω has imaginary part $\frac{\sqrt{3}}{2} \neq 0$ while $1 = 1 + 0i$ has imaginary part 0. So ω is a cube root of 1 different from the obvious one. The computation above also shows that $\omega^2 = \dfrac{-1 - \sqrt{3}\,i}{2}$, and since $(\omega^2)^3 = \omega^6 = (\omega^3)^2 = 1$ by the power rules recorded in Section 1A, that number is a cube root of 1 as well. It is distinct from both 1 and ω, because its imaginary part is $-\frac{\sqrt{3}}{2}$, which differs from 0 and from $\frac{\sqrt{3}}{2}$. Hence 1, ω, ω^2 are three distinct cube roots of 1. Equivalently, one can see all of this at once from the factorization

$$z^3 - 1 = (z - 1)(z^2 + z + 1),$$

because ω satisfies $\omega^2 + \omega + 1 = \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) + \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) + 1 = 0$.

Problem 1A.8. Find two distinct square roots of i.

Solution. We look for $z \in \mathbf{C}$ with $z^2 = i$. Write $z = a + bi$ with $a, b \in \mathbf{R}$. Then, using the definition 1.1 of complex multiplication,

$$z^2 = (a + bi)(a + bi) = (a^2 - b^2) + (2ab)i.$$

Two complex numbers are equal exactly when their real parts agree and their imaginary parts agree, so $z^2 = i = 0 + 1i$ forces

$$a^2 - b^2 = 0 \quad \text{and} \quad 2ab = 1.$$

The first equation gives $b = a$ or $b = -a$. If $b = -a$, then $2ab = -2a^2 \leq 0$, which contradicts $2ab = 1$. Hence $b = a$, and then $2a^2 = 1$, so $a = \frac{1}{\sqrt{2}}$ or $a = -\frac{1}{\sqrt{2}}$.

Thus the only candidates are

$$z = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\,i = \frac{\sqrt{2}}{2}(1+i) \qquad \text{and} \qquad z = -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\,i = -\frac{\sqrt{2}}{2}(1+i).$$

Both really are square roots of i. Indeed,

$$\left(\frac{1+i}{\sqrt{2}}\right)^2 = \frac{(1+i)^2}{2} = \frac{1+2i+i^2}{2} = \frac{1+2i-1}{2} = \frac{2i}{2} = i,$$

and for the second candidate, applying the definition 1.1 of complex multiplication directly,

$$\left(-\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\,i\right)^2 = \left(\tfrac{1}{2} - \tfrac{1}{2}\right) + 2\left(-\tfrac{1}{\sqrt{2}}\right)\left(-\tfrac{1}{\sqrt{2}}\right) i = 0 + i = i.$$

Finally the two numbers are distinct: their difference is $\sqrt{2}(1+i) \neq 0$. So

$$\frac{\sqrt{2}}{2}(1+i) \qquad \text{and} \qquad -\frac{\sqrt{2}}{2}(1+i)$$

are two distinct square roots of i. The computation above also shows that these are the only square roots of i, which is consistent with the fact that a nonzero complex number has exactly two square roots.

Problem 1A.9. Find $x \in \mathbf{R}^4$ such that

$$(4, -3, 1, 7) + 2x = (5, 9, -6, 8).$$

Solution. Write $x = (x_1, x_2, x_3, x_4)$ with each $x_k \in \mathbf{R}$. By the definition 1.18 of scalar multiplication we have $2x = (2x_1, 2x_2, 2x_3, 2x_4)$, and then by the definition 1.13 of addition in $\mathbf{F}^n$,

$$(4, -3, 1, 7) + 2x = (4 + 2x_1,\ -3 + 2x_2,\ 1 + 2x_3,\ 7 + 2x_4).$$

Two lists are equal exactly when they agree in each coordinate, so the required equation is equivalent to the four real equations

$$4 + 2x_1 = 5, \qquad -3 + 2x_2 = 9, \qquad 1 + 2x_3 = -6, \qquad 7 + 2x_4 = 8,$$

whose unique solutions are

$$x_1 = \tfrac{1}{2}, \qquad x_2 = 6, \qquad x_3 = -\tfrac{7}{2}, \qquad x_4 = \tfrac{1}{2}.$$

Hence

$$x = \left(\tfrac{1}{2},\ 6,\ -\tfrac{7}{2},\ \tfrac{1}{2}\right).$$

Check: $2x = (1, 12, -7, 1)$, and

$$(4, -3, 1, 7) + (1, 12, -7, 1) = (5, 9, -6, 8),$$

as required. Since each coordinate equation had a unique solution, this x is the only element of $\mathbf{R}^4$ with the stated property.

Problem 1A.10. Explain why there does not exist $\lambda \in \mathbf{C}$ such that

$$\lambda(2 - 3i,\ 5 + 4i,\ -6 + 7i) = (12 - 5i,\ 7 + 22i,\ -32 - 9i).$$

Solution. Suppose, for the sake of contradiction, that such a $\lambda \in \mathbf{C}$ exists. By the definition 1.18 of scalar multiplication,

$$\lambda(2-3i,\ 5+4i,\ -6+7i) = \big(\lambda(2-3i),\ \lambda(5+4i),\ \lambda(-6+7i)\big),$$

and equality of lists means equality in each coordinate. So λ would have to satisfy all three of

$$\lambda(2-3i) = 12-5i, \qquad \lambda(5+4i) = 7+22i, \qquad \lambda(-6+7i) = -32-9i.$$

The first equation already pins down λ. Since $2-3i \neq 0$, it has a multiplicative inverse in $\mathbf{C}$ (see 1.3 and 1.5), and multiplying the first equation by $\dfrac{1}{2-3i}$ gives

$$\lambda = \frac{12-5i}{2-3i} = \frac{(12-5i)(2+3i)}{(2-3i)(2+3i)} = \frac{24+36i-10i-15i^2}{4+9} = \frac{39+26i}{13} = 3+2i.$$

Thus $\lambda = 3+2i$ is the only possible value.
Now test this value in the third coordinate:

$$(3+2i)(-6+7i) = -18+21i-12i+14i^2 = -18+9i-14 = -32+9i.$$

But the third coordinate on the right side of the desired equation is $-32-9i$, and $-32+9i \neq -32-9i$ because their imaginary parts 9 and -9 differ. This contradicts the third equation above.
Hence no $\lambda \in \mathbf{C}$ satisfies all three coordinate equations, so no such λ exists.
Remark: it is only the third coordinate that fails. The first two are consistent with $\lambda = 3+2i$, since

$$(3+2i)(5+4i) = 15+12i+10i+8i^2 = 15+22i-8 = 7+22i.$$

Geometrically, the point is that the list $(12-5i,\ 7+22i,\ -32-9i)$ is not a scalar multiple of $(2-3i,\ 5+4i,\ -6+7i)$: a single scalar must act simultaneously on every coordinate, and no one scalar works here.

Problem 1A.11. Show that $(x+y)+z = x+(y+z)$ for all $x, y, z \in \mathbf{F}^n$.

Solution. Recall that $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$ (see 1.6), and that in either case addition in $\mathbf{F}$ is associative: for $\mathbf{C}$ this is the associativity statement in 1.3, and $\mathbf{R}$ inherits it as a subset of $\mathbf{C}$ (or one may take it as a familiar property of the real numbers).
Let $x, y, z \in \mathbf{F}^n$. By the definition 1.11 of $\mathbf{F}^n$ we may write

$$x = (x_1, \ldots, x_n), \qquad y = (y_1, \ldots, y_n), \qquad z = (z_1, \ldots, z_n),$$

where $x_k, y_k, z_k \in \mathbf{F}$ for each $k \in \{1, \ldots, n\}$. Using the definition 1.13 of addition in $\mathbf{F}^n$ twice, then associativity of addition in $\mathbf{F}$ in each of the n coordinates, then the definition 1.13 twice more, we compute

$$\begin{aligned}
(x+y)+z &= \big((x_1, \ldots, x_n) + (y_1, \ldots, y_n)\big) + (z_1, \ldots, z_n) \\
&= (x_1+y_1, \ldots, x_n+y_n) + (z_1, \ldots, z_n) \\
&= \big((x_1+y_1)+z_1,\ \ldots,\ (x_n+y_n)+z_n\big) \\
&= \big(x_1+(y_1+z_1),\ \ldots,\ x_n+(y_n+z_n)\big) \\
&= (x_1, \ldots, x_n) + (y_1+z_1, \ldots, y_n+z_n) \\
&= (x_1, \ldots, x_n) + \big((y_1, \ldots, y_n) + (z_1, \ldots, z_n)\big) \\
&= x+(y+z).
\end{aligned}$$

The fourth equality is the only nonformal step: it holds because $(a+b)+c = a+(b+c)$ for all $a, b, c \in \mathbf{F}$, applied with $a = x_k$, $b = y_k$, $c = z_k$ for each $k = 1, \ldots, n$. All other equalities are simply the definition of addition in $\mathbf{F}^n$.
Since $x, y, z \in \mathbf{F}^n$ were arbitrary, addition in $\mathbf{F}^n$ is associative.

Problem 1A.12. Show that $(ab)x = a(bx)$ for all $x \in \mathbf{F}^n$ and all $a, b \in \mathbf{F}$.

Solution. Note first that the two sides of the claimed identity are built from different operations: on the left, ab is a product of two elements of $\mathbf{F}$ and then that scalar multiplies the vector x; on the right, bx is a scalar times a vector and then a multiplies the resulting vector. So the statement is an associativity statement linking multiplication in $\mathbf{F}$ with scalar multiplication in $\mathbf{F}^n$.
Multiplication in $\mathbf{F}$ is associative: for $\mathbf{F} = \mathbf{C}$ this is the associativity statement in 1.3, and for $\mathbf{F} = \mathbf{R}$ it is the familiar property of real multiplication (also a special case of 1.3, since $\mathbf{R} \subseteq \mathbf{C}$).
Let $a, b \in \mathbf{F}$ and let $x = (x_1, \ldots, x_n) \in \mathbf{F}^n$, where each $x_k \in \mathbf{F}$ (definition 1.11). Applying the definition 1.18 of scalar multiplication in $\mathbf{F}^n$ repeatedly,

$$\begin{aligned}(ab)x &= (ab)(x_1, \ldots, x_n)\\ &= \big((ab)x_1,\ \ldots,\ (ab)x_n\big)\\ &= \big(a(bx_1),\ \ldots,\ a(bx_n)\big)\\ &= a(bx_1, \ldots, bx_n)\\ &= a\big(b(x_1, \ldots, x_n)\big)\\ &= a(bx).\end{aligned}$$

The third equality uses associativity of multiplication in $\mathbf{F}$, applied in each coordinate with the elements a, b, x_k of $\mathbf{F}$ for $k = 1, \ldots, n$. The second and fourth equalities are the definition 1.18 (in the fourth, read from right to left, with scalar a), and the fifth is again 1.18 with scalar b.
Since $a, b \in \mathbf{F}$ and $x \in \mathbf{F}^n$ were arbitrary, we conclude that $(ab)x = a(bx)$ always holds. This is the associativity axiom for scalar multiplication that will appear again in the definition of a vector space in Section 1B.

Problem 1A.13. Show that $1x = x$ for all $x \in \mathbf{F}^n$.

Solution. Here 1 denotes the multiplicative identity of $\mathbf{F}$, and $1x$ means the scalar 1 multiplying the vector $x \in \mathbf{F}^n$ in the sense of definition 1.18.
The key fact about $\mathbf{F}$ is the multiplicative identity property from 1.3: $\lambda 1 = \lambda$, and hence by commutativity of multiplication $1\lambda = \lambda$, for every $\lambda \in \mathbf{F}$. (For $\mathbf{F} = \mathbf{C}$ this is exactly the identities statement in 1.3; for $\mathbf{F} = \mathbf{R}$ it is the familiar property of 1, and also a special case since $\mathbf{R} \subseteq \mathbf{C}$.)
Let $x \in \mathbf{F}^n$ and write $x = (x_1, \ldots, x_n)$ with each $x_k \in \mathbf{F}$, as in definition 1.11. Then by the definition 1.18 of scalar multiplication,

$$1x = 1(x_1, \ldots, x_n) = (1x_1, \ldots, 1x_n) = (x_1, \ldots, x_n) = x,$$

where the third equality holds because $1x_k = x_k$ in $\mathbf{F}$ for each $k \in \{1, \ldots, n\}$.
Since $x \in \mathbf{F}^n$ was arbitrary, $1x = x$ for all $x \in \mathbf{F}^n$. Note that this property is genuinely a separate assumption in the abstract setting: when the definition of a vector space is given in Section 1B, the equation $1v = v$ must be listed as an axiom, because it does not follow from the other axioms. To see this, take $V = \mathbf{R}$ with its usual addition but with scalar multiplication redefined by $\lambda \cdot v = 0$ for all $\lambda \in \mathbf{R}$ and all $v \in V$. Every other vector space axiom holds for this V (both sides of the associativity and distributivity axioms are 0), yet $1 \cdot 1 = 0 \neq 1$.

Problem 1A.14. Show that $\lambda(x + y) = \lambda x + \lambda y$ for all $\lambda \in \mathbf{F}$ and all $x, y \in \mathbf{F}^n$.

Solution. This is the distributive property for scalar multiplication over vector addition in $\mathbf{F}^n$; it will be proved by reducing it, coordinate by coordinate, to the distributive property in $\mathbf{F}$.
Multiplication distributes over addition in $\mathbf{F}$: for $\mathbf{F} = \mathbf{C}$ this is the distributive property in 1.3, namely $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbf{C}$; for $\mathbf{F} = \mathbf{R}$ it is the same statement restricted to real numbers.
Let $\lambda \in \mathbf{F}$ and let $x, y \in \mathbf{F}^n$, written as $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ with all $x_k, y_k \in \mathbf{F}$

(definition 1.11). Then

$$\begin{aligned}\lambda(x+y) &= \lambda\big((x_1,\dots,x_n)+(y_1,\dots,y_n)\big)\\ &= \lambda(x_1+y_1,\ \dots,\ x_n+y_n)\\ &= \big(\lambda(x_1+y_1),\ \dots,\ \lambda(x_n+y_n)\big)\\ &= (\lambda x_1+\lambda y_1,\ \dots,\ \lambda x_n+\lambda y_n)\\ &= (\lambda x_1,\dots,\lambda x_n)+(\lambda y_1,\dots,\lambda y_n)\\ &= \lambda(x_1,\dots,x_n)+\lambda(y_1,\dots,y_n)\\ &= \lambda x+\lambda y.\end{aligned}$$

Here the second equality is the definition 1.13 of addition in $\mathbf{F}^n$, the third is the definition 1.18 of scalar multiplication, the fourth applies the distributive property in $\mathbf{F}$ separately in each of the n coordinates, the fifth is again 1.13 (read from right to left), and the sixth is again 1.18.
Since $\lambda \in \mathbf{F}$ and $x, y \in \mathbf{F}^n$ were arbitrary, the identity $\lambda(x+y) = \lambda x + \lambda y$ holds throughout $\mathbf{F}^n$.

Problem 1A.15. Show that $(a+b)x = ax + bx$ for all $a, b \in \mathbf{F}$ and all $x \in \mathbf{F}^n$.

Solution. Fix $a, b \in \mathbf{F}$ and $x \in \mathbf{F}^n$, and write $x = (x_1, \dots, x_n)$ with each $x_j \in \mathbf{F}$.
The definition of scalar multiplication in $\mathbf{F}^n$ (1.18) says that a scalar multiplies a list coordinate by coordinate. Applying it with the scalar $a+b$ gives

$$(a+b)x = \big((a+b)x_1, \dots, (a+b)x_n\big).$$

Now compute the right side of the desired identity. Two more applications of 1.18 give $ax = (ax_1, \dots, ax_n)$ and $bx = (bx_1, \dots, bx_n)$, and then the definition of addition in $\mathbf{F}^n$ (1.13), which adds corresponding coordinates, gives

$$ax + bx = (ax_1, \dots, ax_n) + (bx_1, \dots, bx_n) = (ax_1 + bx_1, \dots, ax_n + bx_n).$$

Thus both $(a+b)x$ and $ax+bx$ are lists of length n, and it suffices to show that their j-th coordinates agree for each $j \in \{1, \dots, n\}$. The j-th coordinate of $(a+b)x$ is $(a+b)x_j$ and the j-th coordinate of $ax + bx$ is $ax_j + bx_j$. These are equal because a, b, and x_j all lie in $\mathbf{F}$, and multiplication in $\mathbf{F}$ distributes over addition in $\mathbf{F}$. Recall that by 1.6 the symbol $\mathbf{F}$ stands for either $\mathbf{R}$ or $\mathbf{C}$, so there are two cases and both are already available to us. For $\mathbf{F} = \mathbf{R}$ the equality $(a+b)x_j = ax_j + bx_j$ is the familiar distributive property of real arithmetic. For $\mathbf{F} = \mathbf{C}$, the distributive property recorded in 1.3 is stated in the form $\lambda(\alpha+\beta) = \lambda\alpha + \lambda\beta$, with the sum inside the second factor, which is not literally the form we need; but 1.3 also records that complex multiplication is commutative, and combining the two gives

$$(a+b)x_j = x_j(a+b) = x_j a + x_j b = ax_j + bx_j,$$

which is exactly the required identity. (Nothing here uses any special feature of $\mathbf{R}$ or $\mathbf{C}$ beyond commutativity and distributivity of multiplication, so the same argument works verbatim over any field, as the digression on fields at the end of Section 1A points out.)
Since the coordinates agree in every slot, the two lists are equal:

$$(a+b)x = ax + bx.$$

As $a, b \in \mathbf{F}$ and $x \in \mathbf{F}^n$ were arbitrary, the identity holds in general.
This is one of the two distributive properties required in the definition of a vector space (1.20). Together with Exercises 11 through 14 of this section (associativity of addition, associativity of scalar multiplication, the multiplicative identity, and the other distributive property) and the commutativity of addition (1.14) along with the additive identity $0 \in \mathbf{F}^n$ (defined in 1.15, with $x + 0 = x$ noted in 1.16) and additive inverses (1.17), it confirms that $\mathbf{F}^n$, equipped with the operations of 1.13 and 1.18, is a vector space over $\mathbf{F}$.

1.2 Exercises 1B

Problem 1B.1. Prove that $-(-v) = v$ for every $v \in V$.

Solution. Let $v \in V$. By 1.28, the notation $-v$ denotes the additive inverse of v, and the notation $-(-v)$ denotes the additive inverse of $-v$; these notations are unambiguous because every element of a vector space has a unique additive inverse (1.27).
By the definition of $-v$ we have

$$v + (-v) = 0.$$

Using commutativity of addition, this says

$$(-v) + v = 0,$$

which is exactly the statement that v is an additive inverse of the vector $-v$. But $-(-v)$ is also an additive inverse of $-v$. Since additive inverses are unique (1.27), these two vectors are equal:

$$-(-v) = v,$$

as desired.
(Alternatively, one may argue directly:

$$-(-v) = -(-v) + 0 = -(-v) + \big((-v) + v\big) = \big(-(-v) + (-v)\big) + v = 0 + v = v,$$

where the first equality is the additive identity condition, the second uses $(-v) + v = v + (-v) = 0$ (commutativity and the definition of $-v$), the third is associativity, the fourth uses $-(-v) + (-v) = (-v) + \big(-(-v)\big) = 0$ (commutativity and the definition of $-(-v)$), and the last uses commutativity together with the additive identity condition. This is just the proof of 1.27 written out for this special case.)

Problem 1B.2. Suppose $a \in \mathbb{F}$, $v \in V$, and $av = 0$. Prove that $a = 0$ or $v = 0$.

Solution. Recall that $\mathbb{F}$ denotes $\mathbb{R}$ or $\mathbb{C}$ (1.1), so every nonzero element of $\mathbb{F}$ has a multiplicative inverse in $\mathbb{F}$.
Suppose $a \in \mathbb{F}$, $v \in V$, and $av = 0$. If $a = 0$, then the conclusion "$a = 0$ or $v = 0$" holds and there is nothing more to prove. So assume $a \neq 0$. Then $\frac{1}{a} \in \mathbb{F}$, and we compute

$$v = 1v = \left(\frac{1}{a}\, a\right) v = \frac{1}{a}(av) = \frac{1}{a}\, 0 = 0,$$

where the first equality holds by the multiplicative identity property in the definition of a vector space (1.20), the second holds because $\frac{1}{a}a = 1$ in $\mathbb{F}$, the third holds by the associativity property $(bc)v = b(cv)$ in 1.20, the fourth holds because $av = 0$ by hypothesis, and the last equality holds by 1.31 (a number times the vector 0 equals 0).
Thus if $a \neq 0$ then $v = 0$. Hence in every case $a = 0$ or $v = 0$, as desired.

Problem 1B.3. Suppose $v, w \in V$. Explain why there exists a unique $x \in V$ such that

$$v + 3x = w.$$

Solution. Here 3 denotes the scalar $3 \in \mathbb{F}$; since $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, we have $3 \neq 0$ and hence $\frac{1}{3} \in \mathbb{F}$.
Existence. Put

$$x = \tfrac{1}{3}(w - v) = \tfrac{1}{3}\big(w + (-v)\big) \in V.$$

Then, using associativity of scalar multiplication and the multiplicative identity property from 1.20,

$$3x = 3\left(\tfrac{1}{3}(w-v)\right) = \left(3\cdot\tfrac{1}{3}\right)(w-v) = 1(w-v) = w+(-v).$$

Therefore, using commutativity and associativity of addition,

$$v+3x = v+\big(w+(-v)\big) = w+\big(v+(-v)\big) = w+0 = w,$$

where the last equality uses the additive identity property. So this x satisfies the equation.
Uniqueness. Suppose $x, x' \in V$ both satisfy $v+3x = w$ and $v+3x' = w$. Then $v+3x = v+3x'$. Add $-v$ to both sides on the left, using commutativity, the additive identity condition, and associativity:

$$3x = 0+3x = \big((-v)+v\big)+3x = (-v)+(v+3x) = (-v)+(v+3x') = \big((-v)+v\big)+3x' = 3x'.$$

Multiplying both sides by the scalar $\frac{1}{3}$ and using associativity of scalar multiplication together with the multiplicative identity property gives

$$x = 1x = \left(\tfrac{1}{3}\cdot 3\right)x = \tfrac{1}{3}(3x) = \tfrac{1}{3}(3x') = \left(\tfrac{1}{3}\cdot 3\right)x' = 1x' = x'.$$

Hence there is exactly one $x \in V$ with $v+3x = w$, namely $x = \frac{1}{3}(w-v)$.
Note that the hypothesis that $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$ was used only through the fact that the scalar 3 is invertible in $\mathbb{F}$; over a field of characteristic 3 the statement would fail.

Problem 1B.4. The empty set is not a vector space. The empty set fails to satisfy only one of the requirements listed in the definition of a vector space (1.20). Which one?

Solution. The requirement that fails is the **additive identity** condition: there is no element $0 \in \emptyset$, because $\emptyset$ has no elements at all.
To see that this is the only failure, go through the definition 1.20 with $V = \emptyset$.
First, the empty set does carry an addition and a scalar multiplication in the sense of 1.19. An addition on V is a function from $V \times V$ to V; here $\emptyset \times \emptyset = \emptyset$, and the empty function $\emptyset \to \emptyset$ is a (unique, legitimate) such function. Likewise a scalar multiplication is a function from $\mathbb{F}\times V$ to V; here $\mathbb{F}\times\emptyset = \emptyset$, and again the empty function serves. So there is nothing wrong with the operations.
Now examine each listed property.

- Commutativity: "$u+v = v+u$ for all $u, v \in V$" is a statement universally quantified over elements of V. Since $V = \emptyset$ has no elements, the statement holds vacuously.
- Associativity: "$(u+v)+w = u+(v+w)$ for all $u, v, w \in V$" is likewise vacuous, and "$(ab)v = a(bv)$ for all $v \in V$ and all $a, b \in \mathbb{F}$" is vacuous in v.
- Additive inverse: "for every $v \in V$, there exists $w \in V$ such that $v+w = 0$" begins with a universal quantifier over V, so it too is vacuously satisfied; there is no v for which a w must be produced.
- Multiplicative identity: "$1v = v$ for all $v \in V$" is vacuous.
- Distributive properties: both "$a(u+v) = au+av$" and "$(a+b)v = av+bv$" are universally quantified over elements of V, hence vacuous.

The additive identity condition is the only one of the six that asserts the **existence** of an element of V rather than a property of all elements of V. An existence statement cannot hold in the empty set. Hence $\emptyset$ satisfies every requirement in 1.20 except the additive identity condition, and that single failure is what disqualifies it from being a vector space.

Problem 1B.5. Show that in the definition of a vector space (1.20), the additive inverse condition can be replaced with the condition that

$$0v = 0 \quad \text{for all } v \in V.$$

Here the 0 on the left side is the number 0, and the 0 on the right side is the additive identity of V.

[The phrase a "condition can be replaced" in a definition means that the collection of objects satisfying the definition is unchanged if the original condition is replaced with the new condition.]

Solution. Let V be a set equipped with an addition and a scalar multiplication (1.19), and consider the five conditions of 1.20 other than the additive inverse condition, namely commutativity, associativity, additive identity, multiplicative identity, and the distributive properties. Call these the **common conditions**. We must show:
V satisfies the common conditions together with the additive inverse condition if and only if V satisfies the common conditions together with the condition $0v = 0$ for all $v \in V$.
First note that under the common conditions alone the phrase "the additive identity of V" is unambiguous: the proof of 1.26 shows that if 0 and $0'$ are both additive identities then $0' = 0' + 0 = 0 + 0' = 0$, and this argument uses only the additive identity condition and commutativity. So the new condition $0v = 0$ has a well-defined meaning in either setting.
Suppose first that V satisfies the common conditions and the additive inverse condition, i.e. V is a vector space. Then $0v = 0$ for every $v \in V$ by 1.30. Explicitly, the distributive property gives

$$0v = (0 + 0)v = 0v + 0v,$$

and adding an additive inverse w of $0v$ to both sides gives

$$0 = 0v + w = (0v + 0v) + w = 0v + (0v + w) = 0v + 0 = 0v,$$

using associativity and the additive identity condition. Hence the new condition holds.
Conversely, suppose V satisfies the common conditions and satisfies $0v = 0$ for all $v \in V$. Let $v \in V$ and set $w = (-1)v$, which lies in V because scalar multiplication maps $\mathbb{F} \times V$ into V. Then

$$v + w = 1v + (-1)v = \big(1 + (-1)\big)v = 0v = 0,$$

where the first equality uses the multiplicative identity condition, the second uses the distributive property $(a + b)v = av + bv$, the third uses $1 + (-1) = 0$ in $\mathbb{F}$, and the fourth is the assumed new condition. Thus v has an additive inverse in V, so the additive inverse condition holds.
Since each version of the definition implies the other, the collection of objects satisfying the definition is unchanged when the additive inverse condition is replaced by the condition $0v = 0$ for all $v \in V$. In other words, the replacement is legitimate.

Problem 1B.6. Let ∞ and $-\infty$ denote two distinct objects, neither of which is in $\mathbb{R}$. Define an addition and scalar multiplication on $\mathbb{R} \cup \{\infty, -\infty\}$ as you could guess from the notation. Specifically, the sum and product of two real numbers is as usual, and for $t \in \mathbb{R}$ define

$$t\infty = \begin{cases} -\infty & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ \infty & \text{if } t > 0, \end{cases} \qquad t(-\infty) = \begin{cases} \infty & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ -\infty & \text{if } t > 0, \end{cases}$$

and

$$\begin{aligned} t + \infty = \infty + t = \infty + \infty = \infty, \\ t + (-\infty) = (-\infty) + t = (-\infty) + (-\infty) = -\infty, \\ \infty + (-\infty) = (-\infty) + \infty = 0. \end{aligned}$$

With these operations of addition and scalar multiplication, is $\mathbb{R} \cup \{\infty, -\infty\}$ a vector space over $\mathbb{R}$? Explain.

Solution. No. Write $W = \mathbb{R} \cup \{\infty, -\infty\}$. The set W with these operations is not a vector space over $\mathbb{R}$, because addition on W is not associative.

Indeed, take $u = -\infty$, $v = \infty$, and $w = 1$. Then

$$(u + v) + w = \big((-\infty) + \infty\big) + 1 = 0 + 1 = 1,$$

while

$$u + (v + w) = (-\infty) + (\infty + 1) = (-\infty) + \infty = 0,$$

using $\infty + 1 = \infty$ from the definition. Since $1 \neq 0$, the associativity requirement of 1.20 fails, so W is not a vector space over $\mathbb{R}$.

A distributive property fails as well, which gives a second independent reason. Take $a = 2$, $b = -1$, and $v = \infty$. Then

$$(a + b)v = 1 \cdot \infty = \infty,$$

whereas

$$av + bv = 2\infty + (-1)\infty = \infty + (-\infty) = 0.$$

Since $\infty \neq 0$, the requirement $(a + b)v = av + bv$ fails.

It is worth noting that these are the only failures: all the remaining requirements of 1.20 do hold for W.

- Commutativity of addition holds because every clause in the definition of addition is stated symmetrically in its two arguments, and addition of real numbers is commutative.
- The additive identity is the real number 0: for $t \in \mathbb{R}$ we have $t + 0 = t$ as usual, and $\infty + 0 = \infty$ and $(-\infty) + 0 = -\infty$ by the displayed rules with the real number $t = 0$.
- Every element of W has an additive inverse: for $t \in \mathbb{R}$ take $-t$; for ∞ take $-\infty$, since $\infty + (-\infty) = 0$; for $-\infty$ take ∞, since $(-\infty) + \infty = 0$.
- Associativity of scalar multiplication, $(ab)v = a(bv)$, holds. For $v \in \mathbb{R}$ this is associativity of multiplication of real numbers. For $v = \pm\infty$, both sides equal 0 whenever $a = 0$ or $b = 0$ (note $0 \cdot (\pm\infty) = 0$, and $a \cdot 0 = 0$ for every $a \in \mathbb{R}$ since $0 \in \mathbb{R}$); and when $a, b \neq 0$ both sides equal ∞ or $-\infty$ according to the sign of ab, since scalar multiplication by a nonzero t preserves $\pm\infty$ when $t > 0$ and swaps them when $t < 0$.
- The multiplicative identity condition $1v = v$ holds: it is clear for $v \in \mathbb{R}$, and $1 \cdot \infty = \infty$, $1 \cdot (-\infty) = -\infty$ because $1 > 0$.
- The other distributive property, $a(u + v) = au + av$, does hold. If $a = 0$ both sides equal 0. If $a > 0$, then multiplication by a fixes ∞ and $-\infty$ and acts as usual on $\mathbb{R}$; checking the cases (both of u, v real; one real and one infinite; both infinite of the same sign; both infinite of opposite signs, where each side equals 0) gives equality in every case. If $a < 0$, multiplication by a interchanges ∞ and $-\infty$ and negates the sign of real numbers, and the same case check again gives equality, the opposite-signs case once more giving 0 on both sides.

So W satisfies commutativity, the existence of an additive identity, the existence of additive inverses, associativity of scalar multiplication, the multiplicative identity condition, and the distributive property $a(u+v) = au+av$, but it fails associativity of addition and the distributive property $(a+b)v = av+bv$. One failure is enough: W is not a vector space over $\mathbb{R}$.

Problem 1B.7. Suppose S is a nonempty set. Let V^S denote the set of functions from S to V. Define a natural addition and scalar multiplication on V^S, and show that V^S is a vector space with these definitions.

Solution. Throughout, V is a vector space over $\mathbb{F}$ (1.29), and we use repeatedly that two functions $f, g\colon S \to V$ are equal if and only if $f(x) = g(x)$ for every $x \in S$.

Definition of the operations. Imitating 1.24, define addition and scalar multiplication on V^S pointwise: for $f, g \in V^S$ let $f + g \in V^S$ be the function given by

$$(f + g)(x) = f(x) + g(x) \quad \text{for all } x \in S,$$

and for $\lambda \in \mathbb{F}$ and $f \in V^S$ let $\lambda f \in V^S$ be the function given by

$$(\lambda f)(x) = \lambda\big(f(x)\big) \quad \text{for all } x \in S.$$

The right sides make sense because $f(x), g(x) \in V$ and V has an addition and a scalar multiplication; and the results are again functions from S to V, so these are indeed an addition and a scalar multiplication on V^S in the sense of 1.19.
We now verify each requirement of 1.20; in every case the verification is the corresponding property of V applied at each point $x \in S$.
Commutativity. Let $f, g \in V^S$. For every $x \in S$,

$$(f+g)(x) = f(x) + g(x) = g(x) + f(x) = (g+f)(x),$$

the middle equality by commutativity in V. Hence $f + g = g + f$.
Associativity. Let $f, g, h \in V^S$. For every $x \in S$,

$$\big((f+g)+h\big)(x) = \big(f(x) + g(x)\big) + h(x) = f(x) + \big(g(x) + h(x)\big) = \big(f + (g+h)\big)(x),$$

using associativity of addition in V. Hence $(f+g)+h = f+(g+h)$. Similarly, for $a, b \in \mathbb{F}$ and $f \in V^S$, for every $x \in S$,

$$\big((ab)f\big)(x) = (ab)\big(f(x)\big) = a\big(bf(x)\big) = a\big((bf)(x)\big) = \big(a(bf)\big)(x),$$

using associativity of scalar multiplication in V. Hence $(ab)f = a(bf)$.
Additive identity. Let $\mathbf{0} \in V^S$ be the function defined by $\mathbf{0}(x) = 0$ for all $x \in S$, where 0 is the additive identity of V. For $f \in V^S$ and every $x \in S$,

$$(f + \mathbf{0})(x) = f(x) + \mathbf{0}(x) = f(x) + 0 = f(x),$$

so $f + \mathbf{0} = f$. Thus $\mathbf{0}$ is an additive identity for V^S.
Additive inverse. Let $f \in V^S$ and define $-f \in V^S$ by $(-f)(x) = -\big(f(x)\big)$ for all $x \in S$, where $-\big(f(x)\big)$ is the additive inverse of $f(x)$ in V (which exists and is unique by 1.27). Then for every $x \in S$,

$$\big(f + (-f)\big)(x) = f(x) + \big(-f(x)\big) = 0 = \mathbf{0}(x),$$

so $f + (-f) = \mathbf{0}$. Thus every element of V^S has an additive inverse.
Multiplicative identity. For $f \in V^S$ and every $x \in S$,

$$(1f)(x) = 1\big(f(x)\big) = f(x),$$

by the multiplicative identity property in V. Hence $1f = f$.
Distributive properties. Let $a, b \in \mathbb{F}$ and $f, g \in V^S$. For every $x \in S$,

$$\big(a(f+g)\big)(x) = a\big(f(x) + g(x)\big) = af(x) + ag(x) = (af)(x) + (ag)(x) = (af + ag)(x),$$

so $a(f+g) = af + ag$; and

$$\big((a+b)f\big)(x) = (a+b)\big(f(x)\big) = af(x) + bf(x) = (af)(x) + (bf)(x) = (af + bf)(x),$$

so $(a+b)f = af + bf$. Both computations use the distributive properties in V.
All requirements of 1.20 are satisfied, so V^S with these operations is a vector space over $\mathbb{F}$. Its additive identity is the function $\mathbf{0}$ that sends every element of S to $0 \in V$, and the additive inverse of f is the function $x \mapsto -f(x)$.
Two remarks. First, taking $V = \mathbb{F}$ recovers exactly the example $\mathbb{F}^S$ of 1.24 and 1.25. Second, the hypothesis that S is nonempty was never used: if $S = \emptyset$ then V^S consists of the single (empty) function, and the argument above still applies, exhibiting V^S as the zero vector space $\{\mathbf{0}\}$. Nonemptiness of S merely matches the convention used in 1.25.

Problem 1B.8. Suppose V is a real vector space.

- The *complexification* of V, denoted by $V_{\mathbf{C}}$, equals $V \times V$. An element of $V_{\mathbf{C}}$ is an ordered pair (u, v), where $u, v \in V$, but we write this as $u + iv$.
- Addition on $V_{\mathbf{C}}$ is defined by
$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$
for all $u_1, v_1, u_2, v_2 \in V$.
- Complex scalar multiplication on $V_{\mathbf{C}}$ is defined by
$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$
for all $a, b \in \mathbf{R}$ and all $u, v \in V$.

Prove that with the definitions of addition and scalar multiplication as above, $V_{\mathbf{C}}$ is a complex vector space.

[Think of V as a subset of $V_{\mathbf{C}}$ by identifying $u \in V$ with $u + i0$. The construction of $V_{\mathbf{C}}$ from V can then be thought of as generalizing the construction of $\mathbf{C}^n$ from $\mathbf{R}^n$.]

Solution. Throughout, V is a vector space over $\mathbf{R}$, so all eight properties in the definition of a vector space (1.20) are available for V with real scalars. Since an element of $V_{\mathbf{C}}$ is by definition the ordered pair (u, v) merely written as $u + iv$, we have the equality criterion

$$u_1 + iv_1 = u_2 + iv_2 \quad \Longleftrightarrow \quad u_1 = u_2 \ \text{ and } \ v_1 = v_2.$$

Both operations defined above produce an expression of the form (element of V) $+\, i\,$(element of V), so addition is a function from $V_{\mathbf{C}} \times V_{\mathbf{C}}$ to $V_{\mathbf{C}}$ and complex scalar multiplication is a function from $\mathbf{C} \times V_{\mathbf{C}}$ to $V_{\mathbf{C}}$, as required. Note also that every complex number has a unique representation $a + bi$ with $a, b \in \mathbf{R}$, so the formula for scalar multiplication is unambiguous.

Preliminary remark on the notation $au - bv$. By 1.28, $au - bv$ means $au + \big(-(bv)\big)$. We first check that $-(bv) = (-b)v$ for all $b \in \mathbf{R}$ and $v \in V$. Indeed,

$$bv + (-b)v = \big(b + (-b)\big)v = 0v = 0,$$

where the first equality is the distributive property in V and the last is 1.30. Hence $(-b)v$ is an additive inverse of bv, and additive inverses are unique (1.27), so $-(bv) = (-b)v$. Consequently every expression below of the form $\alpha u + \beta v$ with $\alpha, \beta \in \mathbf{R}$ may be manipulated using only the vector space properties of V over $\mathbf{R}$, with subtraction of real coefficients handled as addition of negatives. We use this silently from now on.

We now verify the eight properties of 1.20 for $V_{\mathbf{C}}$ over $\mathbf{F} = \mathbf{C}$. Fix arbitrary $u, v, u_1, v_1, u_2, v_2, u_3, v_3 \in V$ and arbitrary $\lambda = a + bi$, $\mu = c + di$ in $\mathbf{C}$, where $a, b, c, d \in \mathbf{R}$.

Commutativity. Using commutativity of addition in V,

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2) = (u_2 + u_1) + i(v_2 + v_1) = (u_2 + iv_2) + (u_1 + iv_1).$$

Associativity of addition. Using associativity of addition in V,

$$\begin{aligned}
\big[(u_1 + iv_1) + (u_2 + iv_2)\big] + (u_3 + iv_3) &= \big[(u_1 + u_2) + i(v_1 + v_2)\big] + (u_3 + iv_3) \\
&= \big((u_1 + u_2) + u_3\big) + i\big((v_1 + v_2) + v_3\big) \\
&= \big(u_1 + (u_2 + u_3)\big) + i\big(v_1 + (v_2 + v_3)\big) \\
&= (u_1 + iv_1) + \big[(u_2 + u_3) + i(v_2 + v_3)\big] \\
&= (u_1 + iv_1) + \big[(u_2 + iv_2) + (u_3 + iv_3)\big].
\end{aligned}$$

Associativity of scalar multiplication. We must show $(\lambda\mu)(u + iv) = \lambda\big(\mu(u + iv)\big)$. In $\mathbf{C}$ we have $\lambda\mu = (ac - bd) + (ad + bc)i$, so by definition

$$(\lambda\mu)(u + iv) = \big((ac - bd)u - (ad + bc)v\big) + i\big((ac - bd)v + (ad + bc)u\big).$$

On the other hand $\mu(u+iv) = (cu-dv)+i(cv+du)$, and therefore

$$\begin{aligned}\lambda\big(\mu(u+iv)\big) &= \Big(a(cu-dv)-b(cv+du)\Big)+i\Big(a(cv+du)+b(cu-dv)\Big)\\ &= \Big((ac)u-(ad)v-(bc)v-(bd)u\Big)+i\Big((ac)v+(ad)u+(bc)u-(bd)v\Big)\\ &= \big((ac-bd)u-(ad+bc)v\big)+i\big((ac-bd)v+(ad+bc)u\big),\end{aligned}$$

where the second equality uses the distributive property $a(x+y) = ax+ay$ and associativity of scalar multiplication $a(cu) = (ac)u$ in V, and the third uses the distributive property $(\alpha+\beta)x = \alpha x+\beta x$ in V together with commutativity and associativity of addition in V. The two displayed expressions agree, as desired.

Additive identity. Let 0 denote the additive identity of V and set $\mathbf{0} = 0+i0 \in V_{\mathbf{C}}$. Then

$$(u+iv)+(0+i0) = (u+0)+i(v+0) = u+iv,$$

so $\mathbf{0}$ is an additive identity for $V_{\mathbf{C}}$ (and it is the only one, by 1.26).

Additive inverse. Given $u+iv \in V_{\mathbf{C}}$, put $w = (-u)+i(-v)$, which lies in $V_{\mathbf{C}}$ because $-u, -v \in V$. Then

$$(u+iv)+w = \big(u+(-u)\big)+i\big(v+(-v)\big) = 0+i0 = \mathbf{0}.$$

Thus every element of $V_{\mathbf{C}}$ has an additive inverse.

Multiplicative identity. The complex number 1 is $1+0i$, so

$$1(u+iv) = (1u-0v)+i(1v+0u) = (u-0)+i(v+0) = u+iv,$$

using $1u = u$ in V and $0v = 0u = 0$ by 1.30. Here $u-0v$ means $u+\big(-(0v)\big)$, and by the preliminary remark $-(0v) = (-0)v = 0v = 0$, so $u-0v = u+0 = u$; similarly $v+0u = v+0 = v$.

First distributive property. We show $\lambda(x+y) = \lambda x+\lambda y$. With $x = u_1+iv_1$ and $y = u_2+iv_2$,

$$\begin{aligned}\lambda(x+y) &= (a+bi)\big((u_1+u_2)+i(v_1+v_2)\big)\\ &= \Big(a(u_1+u_2)-b(v_1+v_2)\Big)+i\Big(a(v_1+v_2)+b(u_1+u_2)\Big)\\ &= \big(au_1+au_2-bv_1-bv_2\big)+i\big(av_1+av_2+bu_1+bu_2\big)\\ &= \big((au_1-bv_1)+(au_2-bv_2)\big)+i\big((av_1+bu_1)+(av_2+bu_2)\big)\\ &= \big((au_1-bv_1)+i(av_1+bu_1)\big)+\big((au_2-bv_2)+i(av_2+bu_2)\big)\\ &= \lambda x+\lambda y,\end{aligned}$$

where the third equality uses the distributive property $\alpha(x+y) = \alpha x+\alpha y$ in V (applied with the real scalars a and $-b$), and the fourth uses commutativity and associativity of addition in V to regroup the four summands.

Second distributive property. We show $(\lambda+\mu)x = \lambda x+\mu x$. Since $\lambda+\mu = (a+c)+(b+d)i$, with $x = u+iv$ we get

$$\begin{aligned}(\lambda+\mu)x &= \big((a+c)u-(b+d)v\big)+i\big((a+c)v+(b+d)u\big)\\ &= \big(au+cu-bv-dv\big)+i\big(av+cv+bu+du\big)\\ &= \big((au-bv)+(cu-dv)\big)+i\big((av+bu)+(cv+du)\big)\\ &= \big((au-bv)+i(av+bu)\big)+\big((cu-dv)+i(cv+du)\big)\\ &= \lambda x+\mu x,\end{aligned}$$

where the second equality uses the distributive property $(\alpha+\beta)x = \alpha x+\beta x$ in V (with the real scalars a, c and $-b, -d$, noting $-(b+d) = (-b)+(-d)$), and the third uses commutativity and associativity of addition in V.

All eight properties of 1.20 hold, so $V_{\mathbf{C}}$ is a vector space over $\mathbf{C}$.

Remark on the identification. Identify $u \in V$ with $u + i0 \in V_{\mathbf{C}}$. This identification respects the operations inherited from V: for $u_1, u_2 \in V$ we have $(u_1 + i0) + (u_2 + i0) = (u_1 + u_2) + i0$, and for a real scalar a, viewed as the complex number $a + 0i$, applying the definition with $b = 0$ and $v = 0$ gives

$$(a + 0i)(u + i0) = \big(au - 0 \cdot 0\big) + i\big(a0 + 0u\big) = au + i0,$$

since $0 \cdot 0 = 0$ and $0u = 0$ by 1.30 while $a0 = 0$ by 1.31 (and $-0 = 0$, as noted above). So the copy of V inside $V_{\mathbf{C}}$ carries exactly its original real vector space structure. Taking $V = \mathbf{R}^n$ recovers $\mathbf{C}^n$: the pair (x, y) with $x, y \in \mathbf{R}^n$ corresponds to $x + iy = (x_1 + iy_1, \ldots, x_n + iy_n) \in \mathbf{C}^n$, and the addition and scalar multiplication defined above become the usual coordinatewise operations on $\mathbf{C}^n$.

1.3 Exercises 1C

Problem 1C.1. For each of the following subsets of $\mathbf{F}^3$, determine whether it is a subspace of $\mathbf{F}^3$.
(a) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 + 2x_2 + 3x_3 = 0\}$
(b) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 + 2x_2 + 3x_3 = 4\}$
(c) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 x_2 x_3 = 0\}$
(d) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 = 5x_3\}$

Solution. Throughout we use 1.34: a subset U of a vector space V is a subspace if and only if $0 \in U$, U is closed under addition, and U is closed under scalar multiplication.
(a) This is a subspace. Write $U = \{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 + 2x_2 + 3x_3 = 0\}$.
Additive identity: $0 + 2 \cdot 0 + 3 \cdot 0 = 0$, so $(0, 0, 0) \in U$.
Closed under addition: suppose (x_1, x_2, x_3) and (y_1, y_2, y_3) belong to U. Then

$$(x_1 + y_1) + 2(x_2 + y_2) + 3(x_3 + y_3) = (x_1 + 2x_2 + 3x_3) + (y_1 + 2y_2 + 3y_3) = 0 + 0 = 0,$$

so $(x_1, x_2, x_3) + (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3) \in U$.
Closed under scalar multiplication: if $\lambda \in \mathbf{F}$ and $(x_1, x_2, x_3) \in U$, then

$$\lambda x_1 + 2\lambda x_2 + 3\lambda x_3 = \lambda(x_1 + 2x_2 + 3x_3) = \lambda \cdot 0 = 0,$$

so $\lambda(x_1, x_2, x_3) \in U$. Hence U is a subspace of $\mathbf{F}^3$.
(b) This is not a subspace, because it does not contain the additive identity: $0 + 2 \cdot 0 + 3 \cdot 0 = 0 \neq 4$, so $(0, 0, 0)$ is not in the set. By 1.34 the set is not a subspace of $\mathbf{F}^3$. (It also fails the other two conditions: $(4, 0, 0)$ lies in the set, but $(4, 0, 0) + (4, 0, 0) = (8, 0, 0)$ does not, since $8 \neq 4$; and $2(4, 0, 0) = (8, 0, 0)$ does not either.)
(c) This is not a subspace. The set contains $(1, 1, 0)$ and $(0, 0, 1)$, since $1 \cdot 1 \cdot 0 = 0$ and $0 \cdot 0 \cdot 1 = 0$. But

$$(1, 1, 0) + (0, 0, 1) = (1, 1, 1)$$

and $1 \cdot 1 \cdot 1 = 1 \neq 0$, so the sum is not in the set. Thus the set is not closed under addition and hence is not a subspace of $\mathbf{F}^3$. (Note that 0 does lie in the set and the set is closed under scalar multiplication, since $(\lambda x_1)(\lambda x_2)(\lambda x_3) = \lambda^3 x_1 x_2 x_3$; only closure under addition fails.)
(d) This is a subspace. Write $U = \{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 = 5x_3\}$.
Additive identity: $0 = 5 \cdot 0$, so $(0, 0, 0) \in U$.
Closed under addition: if $(x_1, x_2, x_3), (y_1, y_2, y_3) \in U$, then $x_1 = 5x_3$ and $y_1 = 5y_3$, so

$$x_1 + y_1 = 5x_3 + 5y_3 = 5(x_3 + y_3),$$

which says exactly that $(x_1 + y_1, x_2 + y_2, x_3 + y_3) \in U$.
Closed under scalar multiplication: if $\lambda \in \mathbf{F}$ and $(x_1, x_2, x_3) \in U$, then $\lambda x_1 = \lambda(5x_3) = 5(\lambda x_3)$, so $\lambda(x_1, x_2, x_3) \in U$.
Hence U is a subspace of $\mathbf{F}^3$.

Problem 1C.2. Verify all assertions about subspaces in Example 1.35.

Solution. Example 1.35 makes five assertions. We verify each in turn, using the criterion 1.34 throughout.

(a) If $b \in \mathbf{F}$, then $U = \{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_3 = 5x_4 + b\}$ is a subspace of $\mathbf{F}^4$ if and only if $b = 0$.

Suppose first that $b \neq 0$. Then $(0, 0, 0, 0) \notin U$, because $0 = 5 \cdot 0 + b$ would force $b = 0$. By 1.34, U is not a subspace of $\mathbf{F}^4$.

Now suppose $b = 0$, so $U = \{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_3 = 5x_4\}$. Since $0 = 5 \cdot 0$, we have $(0, 0, 0, 0) \in U$. If (x_1, x_2, x_3, x_4) and (y_1, y_2, y_3, y_4) lie in U, then $x_3 = 5x_4$ and $y_3 = 5y_4$, hence $x_3 + y_3 = 5(x_4 + y_4)$, so the sum lies in U. If $\lambda \in \mathbf{F}$ and $(x_1, x_2, x_3, x_4) \in U$, then $\lambda x_3 = \lambda(5x_4) = 5(\lambda x_4)$, so $\lambda(x_1, x_2, x_3, x_4) \in U$. Thus U is a subspace of $\mathbf{F}^4$.

(b) The set $C[0, 1]$ of continuous real-valued functions on $[0, 1]$ is a subspace of $\mathbf{R}^{[0,1]}$.

The zero function is constant, hence continuous, so $0 \in C[0, 1]$. If f, g are continuous on $[0, 1]$, then $f + g$ is continuous on $[0, 1]$ (the sum of two continuous functions is continuous), so $C[0, 1]$ is closed under addition. If $\lambda \in \mathbf{R}$ and f is continuous, then λf is continuous, so $C[0, 1]$ is closed under scalar multiplication. Hence $C[0, 1]$ is a subspace of $\mathbf{R}^{[0,1]}$.

(c) The set D of differentiable real-valued functions on $\mathbf{R}$ is a subspace of $\mathbf{R}^{\mathbf{R}}$.

The zero function is differentiable with derivative identically 0, so $0 \in D$. If $f, g \in D$, then $f + g$ is differentiable with $(f + g)' = f' + g'$, so $f + g \in D$. If $\lambda \in \mathbf{R}$ and $f \in D$, then λf is differentiable with $(\lambda f)' = \lambda f'$, so $\lambda f \in D$. Hence D is a subspace of $\mathbf{R}^{\mathbf{R}}$.

(d) The set $U = \{f \in \mathbf{R}^{(0,3)} : f \text{ is differentiable and } f'(2) = b\}$ is a subspace of $\mathbf{R}^{(0,3)}$ if and only if $b = 0$.

If $b \neq 0$, then the zero function, whose derivative at 2 equals 0, does not lie in U; by 1.34, U is not a subspace.

If $b = 0$: the zero function is differentiable and its derivative at 2 is 0, so $0 \in U$. If $f, g \in U$, then $f + g$ is differentiable and $(f + g)'(2) = f'(2) + g'(2) = 0 + 0 = 0$, so $f + g \in U$. If $\lambda \in \mathbf{R}$ and $f \in U$, then λf is differentiable and $(\lambda f)'(2) = \lambda f'(2) = \lambda \cdot 0 = 0$, so $\lambda f \in U$. Hence U is a subspace of $\mathbf{R}^{(0,3)}$.

(e) The set U of all sequences of complex numbers with limit 0 is a subspace of $\mathbf{C}^{\infty}$.

The sequence $(0, 0, \dots)$ has limit 0, so $0 \in U$. If $(x_1, x_2, \dots)$ and $(y_1, y_2, \dots)$ have limit 0, then by the limit law for sums $\lim_{n\to\infty}(x_n + y_n) = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n = 0$, so the sum lies in U. If $\lambda \in \mathbf{C}$ and $(x_1, x_2, \dots) \in U$, then $\lim_{n\to\infty} \lambda x_n = \lambda \lim_{n\to\infty} x_n = 0$, so $\lambda(x_1, x_2, \dots) \in U$. Hence U is a subspace of $\mathbf{C}^{\infty}$.

Problem 1C.3. Show that the set of differentiable real-valued functions f on the interval $(-4, 4)$ such that $f'(-1) = 3f(2)$ is a subspace of $\mathbf{R}^{(-4,4)}$.

Solution. Let

$$U = \{f \in \mathbf{R}^{(-4,4)} : f \text{ is differentiable on } (-4, 4) \text{ and } f'(-1) = 3f(2)\}.$$

Note that -1 and 2 both lie in $(-4, 4)$, so the condition makes sense. We verify the three conditions of 1.34.

Additive identity. The zero function 0 on $(-4, 4)$ is differentiable, with $0'(x) = 0$ for all x. Hence $0'(-1) = 0 = 3 \cdot 0 = 3 \cdot 0(2)$, so $0 \in U$.

Closed under addition. Suppose $f, g \in U$. The sum of two differentiable functions is differentiable, and $(f + g)' = f' + g'$. Therefore

$$(f + g)'(-1) = f'(-1) + g'(-1) = 3f(2) + 3g(2) = 3\big(f(2) + g(2)\big) = 3(f + g)(2),$$

where the second equality uses $f, g \in U$ and the last uses the definition of the sum of functions. Hence $f + g \in U$.

Closed under scalar multiplication. Suppose $\lambda \in \mathbf{R}$ and $f \in U$. Then λf is differentiable with $(\lambda f)' = \lambda f'$, so

$$(\lambda f)'(-1) = \lambda f'(-1) = \lambda \cdot 3f(2) = 3\big(\lambda f(2)\big) = 3(\lambda f)(2).$$

Hence $\lambda f \in U$.
By 1.34, U is a subspace of $\mathbf{R}^{(-4,4)}$.

Problem 1C.4. Suppose $b \in \mathbf{R}$. Show that the set of continuous real-valued functions f on the interval $[0,1]$ such that $\int_0^1 f = b$ is a subspace of $\mathbf{R}^{[0,1]}$ if and only if $b = 0$.

Solution. Let

$$U = \Big\{ f \in \mathbf{R}^{[0,1]} : f \text{ is continuous on } [0,1] \text{ and } \int_0^1 f = b \Big\}.$$

Every continuous function on the closed bounded interval $[0,1]$ is Riemann integrable there, so the defining condition makes sense for every continuous f.
First suppose $b \neq 0$. The zero function 0 on $[0,1]$ satisfies $\int_0^1 0 = 0 \neq b$, so $0 \notin U$. By 1.34, U is not a subspace of $\mathbf{R}^{[0,1]}$.
Now suppose $b = 0$, so that $U = \{f \in \mathbf{R}^{[0,1]} : f \text{ continuous}, \int_0^1 f = 0\}$. We check the three conditions of 1.34.
Additive identity. The zero function is continuous and $\int_0^1 0 = 0$, so $0 \in U$.
Closed under addition. Suppose $f, g \in U$. Then $f + g$ is continuous (the sum of continuous functions is continuous), and by additivity of the integral

$$\int_0^1 (f+g) = \int_0^1 f + \int_0^1 g = 0 + 0 = 0.$$

Hence $f + g \in U$.
Closed under scalar multiplication. Suppose $\lambda \in \mathbf{R}$ and $f \in U$. Then λf is continuous, and by homogeneity of the integral

$$\int_0^1 (\lambda f) = \lambda \int_0^1 f = \lambda \cdot 0 = 0.$$

Hence $\lambda f \in U$.
By 1.34, U is a subspace of $\mathbf{R}^{[0,1]}$ when $b = 0$. Combining the two directions, U is a subspace of $\mathbf{R}^{[0,1]}$ if and only if $b = 0$.

Problem 1C.5. Is $\mathbf{R}^2$ a subspace of the complex vector space $\mathbf{C}^2$?

Solution. No.
Here $\mathbf{R}^2$ is regarded as the subset $\{(x_1, x_2) \in \mathbf{C}^2 : x_1, x_2 \in \mathbf{R}\}$ of $\mathbf{C}^2$, and $\mathbf{C}^2$ is regarded as a vector space over $\mathbf{C}$, so that the scalars available are all complex numbers.
This subset does contain $(0,0)$ and is closed under addition, since the sum of two real numbers is real. However, it is not closed under scalar multiplication by complex scalars. Indeed, $(1,0) \in \mathbf{R}^2$ and $i \in \mathbf{C}$, but

$$i(1,0) = (i,0) \notin \mathbf{R}^2,$$

because i is not a real number. Since the third condition of 1.34 fails, $\mathbf{R}^2$ is not a subspace of the complex vector space $\mathbf{C}^2$.
It is worth noting where this leaves things: $\mathbf{R}^2$ is a vector space in its own right and it sits inside $\mathbf{C}^2$ with the same addition, but it fails to be a subspace of $\mathbf{C}^2$ precisely because the scalar field of $\mathbf{C}^2$ is $\mathbf{C}$ rather than $\mathbf{R}$. If instead we view $\mathbf{C}^2$ as a real vector space (that is, restrict the scalars to $\mathbf{R}$), then $\mathbf{R}^2$ is a subspace of it.

Problem 1C.6. (a) Is $\{(a,b,c) \in \mathbf{R}^3 : a^3 = b^3\}$ a subspace of $\mathbf{R}^3$?
(b) Is $\{(a,b,c) \in \mathbf{C}^3 : a^3 = b^3\}$ a subspace of $\mathbf{C}^3$?

Solution. (a) Yes.
The key point is that the function $t \mapsto t^3$ is injective on $\mathbf{R}$. To see this, suppose $a, b \in \mathbf{R}$ and $a^3 = b^3$. Then

$$0 = a^3 - b^3 = (a-b)(a^2 + ab + b^2).$$

Now

$$a^2 + ab + b^2 = \left(a + \tfrac{b}{2}\right)^2 + \tfrac{3}{4}b^2 \geq 0,$$

and this quantity equals 0 only if $b = 0$ and $a + \frac{b}{2} = 0$, that is, only if $a = b = 0$. So either $a = b = 0$, or $a^2 + ab + b^2 \neq 0$ and the displayed equation forces $a - b = 0$. In both cases $a = b$. The converse implication is trivial, so for real a, b we have

$$a^3 = b^3 \iff a = b.$$

(Note that this argument uses that $\mathbf{R}$ is an ordered field in which squares are nonnegative; it fails over $\mathbf{C}$, as part (b) shows.)
Consequently

$$U = \{(a,b,c) \in \mathbf{R}^3 : a^3 = b^3\} = \{(a,a,c) \in \mathbf{R}^3 : a, c \in \mathbf{R}\},$$

which is easily seen to be a subspace: $(0,0,0) \in U$; if (a,a,c) and (a',a',c') are in U then their sum $(a+a', a+a', c+c')$ has its first two coordinates equal, so it is in U; and for $\lambda \in \mathbf{R}$, $\lambda(a,a,c) = (\lambda a, \lambda a, \lambda c) \in U$. By 1.34, U is a subspace of $\mathbf{R}^3$.
(b) No.
Over $\mathbf{C}$ the cube function is not injective, and that breaks closure under addition. Let $\omega = e^{2\pi i/3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$, a primitive cube root of 1, so that $\omega^3 = 1$ and $1 + \omega + \omega^2 = 0$. Put

$$u = (1,1,0) \quad \text{and} \quad v = (1,\omega,0).$$

Both lie in the set: for u we have $1^3 = 1^3$, and for v we have $1^3 = 1 = \omega^3$.
Their sum is $u + v = (2, 1+\omega, 0)$. Now $2^3 = 8$, while, using $\omega^3 = 1$ and $\omega + \omega^2 = -1$,

$$(1+\omega)^3 = 1 + 3\omega + 3\omega^2 + \omega^3 = 1 + 3(\omega + \omega^2) + 1 = 2 + 3(-1) = -1.$$

Since $8 \neq -1$, the vector $u + v$ does not belong to the set. Thus the set is not closed under addition, and by 1.34 it is not a subspace of $\mathbf{C}^3$.

Problem 1C.7. Prove or give a counterexample: If U is a nonempty subset of $\mathbf{R}^2$ such that U is closed under addition and under taking additive inverses (meaning $-u \in U$ whenever $u \in U$), then U is a subspace of $\mathbf{R}^2$.

Solution. The statement is false. Here is a counterexample.
Let

$$U = \mathbf{Z}^2 = \{(m,n) \in \mathbf{R}^2 : m, n \text{ are integers}\}.$$

U is nonempty: for instance $(0,0) \in U$.
U is closed under addition: if (m,n) and (m',n') are in U, then $m + m'$ and $n + n'$ are integers, so $(m,n) + (m',n') = (m+m', n+n') \in U$.
U is closed under taking additive inverses: if $(m,n) \in U$, then $-m$ and $-n$ are integers, so $-(m,n) = (-m,-n) \in U$.
But U is not a subspace of $\mathbf{R}^2$, because it is not closed under scalar multiplication: $(1,0) \in U$ and $\frac{1}{2} \in \mathbf{R}$, yet

$$\tfrac{1}{2}(1,0) = \left(\tfrac{1}{2}, 0\right) \notin U,$$

since $\frac{1}{2}$ is not an integer. By 1.34, U is not a subspace of $\mathbf{R}^2$.
Remark on why the hypotheses are not enough. Closure under addition and additive inverses (together with nonemptiness) makes U a subgroup of the additive group of $\mathbf{R}^2$; in particular $0 \in U$, since picking any $u \in U$ gives $0 = u + (-u) \in U$. Iterating addition and inverses yields $nu \in U$ for every integer

n, so U is automatically closed under multiplication by integer scalars. What the hypotheses cannot deliver is closure under multiplication by non-integer real scalars, and that is exactly the gap the counterexample exploits. The set $\mathbf{Q}^2 = \{(x, y) \in \mathbf{R}^2 : x, y \in \mathbf{Q}\}$ is another counterexample, for the same reason: it is closed under addition and negation, but $\sqrt{2}\,(1, 0) = (\sqrt{2}, 0) \notin \mathbf{Q}^2$.
Note that over some other scalar fields the analogous statement can be true. For example, if $\mathbf{F} = \mathbf{Z}/2$ then every scalar is 0 or 1, and a nonempty subset closed under addition automatically contains 0 and is closed under scalar multiplication. The failure above is specific to scalar fields, such as $\mathbf{R}$, that contain scalars not generated from 1 by repeated addition and negation.

Problem 1C.8. Give an example of a nonempty subset U of $\mathbb{R}^2$ such that U is closed under scalar multiplication, but U is not a subspace of $\mathbb{R}^2$.

Solution. Let U be the union of the two coordinate axes:

$$U = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 = 0\}.$$

U is nonempty, since $(0, 0) \in U$.
U is closed under scalar multiplication. Indeed, suppose $\lambda \in \mathbb{R}$ and $(x_1, x_2) \in U$, so that $x_1 x_2 = 0$. Then

$$(\lambda x_1)(\lambda x_2) = \lambda^2 (x_1 x_2) = \lambda^2 \cdot 0 = 0,$$

so $\lambda(x_1, x_2) = (\lambda x_1, \lambda x_2) \in U$.
However U is not a subspace of $\mathbb{R}^2$, because it is not closed under addition: we have $(1, 0) \in U$ and $(0, 1) \in U$ (both products of coordinates are 0), but

$$(1, 0) + (0, 1) = (1, 1) \notin U,$$

since $1 \cdot 1 = 1 \neq 0$. Because closure under addition is one of the three conditions in 1.34, U is not a subspace of $\mathbb{R}^2$.
Geometrically, U is the union of two lines through the origin. Each line is itself a subspace, and scaling never moves a point off the line it lies on; but adding a nonzero point of one axis to a nonzero point of the other produces a point on neither axis. This example is the two-subspace case of Exercise 12: a union of two subspaces, neither contained in the other, is never a subspace.

Problem 1C.9. A function $f\colon \mathbb{R} \to \mathbb{R}$ is called periodic if there exists a positive number p such that $f(x) = f(x + p)$ for all $x \in \mathbb{R}$. Is the set of periodic functions from $\mathbb{R}$ to $\mathbb{R}$ a subspace of $\mathbb{R}^{\mathbb{R}}$? Explain.

Solution. No. Let $P \subseteq \mathbb{R}^{\mathbb{R}}$ denote the set of periodic functions from $\mathbb{R}$ to $\mathbb{R}$. Two of the three conditions of 1.34 hold, but closure under addition fails.
First the two conditions that do hold. The zero function is periodic (every positive number is a period for it), so $0 \in P$. And if $f \in P$ has period p and $\lambda \in \mathbb{R}$, then for all $x \in \mathbb{R}$ we have $(\lambda f)(x + p) = \lambda f(x + p) = \lambda f(x) = (\lambda f)(x)$, so λf is periodic with the same period p; thus P is closed under scalar multiplication.
Now we show P is not closed under addition. Define $f, g\colon \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \cos x, \qquad g(x) = \cos(\sqrt{2}\,x).$$

Then f is periodic with period 2π and g is periodic with period $2\pi/\sqrt{2} = \sqrt{2}\,\pi$, so $f, g \in P$. We claim $h = f + g$ is not periodic.
Because $\cos t \leq 1$ for all $t \in \mathbb{R}$, we have $h(x) \leq 2$ for all x, and $h(0) = 1 + 1 = 2$. Hence 2 is the maximum value of h, and $h(x) = 2$ forces $\cos x = 1$ and $\cos(\sqrt{2}\,x) = 1$ (if either summand were less than 1, the sum would be less than 2). Now $\cos x = 1$ if and only if $x = 2\pi m$ for some $m \in \mathbb{Z}$, and $\cos(\sqrt{2}\,x) = 1$ if and only if $\sqrt{2}\,x = 2\pi n$ for some $n \in \mathbb{Z}$. Suppose both hold. Substituting $x = 2\pi m$

into the second equation gives $2\sqrt{2}\,\pi m = 2\pi n$, hence

$$\sqrt{2}\,m = n.$$

If $m \neq 0$, this would give $\sqrt{2} = n/m \in \mathbb{Q}$, contradicting the irrationality of $\sqrt{2}$. Thus $m = 0$, and then $n = 0$ and $x = 0$. In other words,

$$\{x \in \mathbb{R} : h(x) = 2\} = \{0\}.$$

Suppose now, seeking a contradiction, that h were periodic, say $h(x) = h(x + p)$ for all $x \in \mathbb{R}$, with $p > 0$. Taking $x = 0$ gives $h(p) = h(0) = 2$, so by the displayed equation above $p = 0$, contradicting $p > 0$. Hence $h = f + g$ is not periodic.
So $f, g \in P$ but $f + g \notin P$, and therefore P is not a subspace of $\mathbb{R}^{\mathbb{R}}$.
Remark. The obstruction is exactly incommensurability of periods. If f has period p and g has period q with p/q rational, say $p/q = a/b$ with a, b positive integers, then $bp = aq$ is a common period and $f + g$ is periodic. The set of functions periodic with a fixed period p (i.e. satisfying $f(x) = f(x + p)$ for all x) is a subspace of $\mathbb{R}^{\mathbb{R}}$; it is the union of these subspaces over all $p > 0$ that fails to be one.

Problem 1C.10. Suppose V_1 and V_2 are subspaces of V. Prove that the intersection $V_1 \cap V_2$ is a subspace of V.

Solution. We verify the three conditions of 1.34 for the subset $V_1 \cap V_2$ of V.
Additive identity. Because V_1 is a subspace of V, we have $0 \in V_1$; because V_2 is a subspace of V, we have $0 \in V_2$. Hence $0 \in V_1 \cap V_2$.
Closed under addition. Suppose $u, w \in V_1 \cap V_2$. Then $u, w \in V_1$, and V_1 is closed under addition, so $u + w \in V_1$. Likewise $u, w \in V_2$ and V_2 is closed under addition, so $u + w \in V_2$. Hence $u + w \in V_1 \cap V_2$.
Closed under scalar multiplication. Suppose $a \in \mathbb{F}$ and $u \in V_1 \cap V_2$. Then $u \in V_1$, and V_1 is closed under scalar multiplication, so $au \in V_1$. Likewise $u \in V_2$ gives $au \in V_2$. Hence $au \in V_1 \cap V_2$.
All three conditions of 1.34 hold, so $V_1 \cap V_2$ is a subspace of V.

Problem 1C.11. Prove that the intersection of every collection of subspaces of V is a subspace of V.

Solution. Let $\mathcal{A}$ be a nonempty collection of subspaces of V, indexed as $\{V_i\}_{i \in I}$ with $I \neq \varnothing$, and set

$$W = \bigcap_{i \in I} V_i = \{v \in V : v \in V_i \text{ for every } i \in I\}.$$

Note $W \subseteq V$, since each $V_i \subseteq V$ and I is nonempty. We verify the three conditions of 1.34.
Additive identity. Each V_i is a subspace of V, so $0 \in V_i$ for every $i \in I$. Hence $0 \in W$.
Closed under addition. Suppose $u, w \in W$. Let $i \in I$. Then $u \in V_i$ and $w \in V_i$, and V_i is closed under addition, so $u + w \in V_i$. Since $i \in I$ was arbitrary, $u + w$ belongs to every V_i, that is, $u + w \in W$.
Closed under scalar multiplication. Suppose $a \in \mathbb{F}$ and $u \in W$. Let $i \in I$. Then $u \in V_i$, and V_i is closed under scalar multiplication, so $au \in V_i$. Since $i \in I$ was arbitrary, $au \in W$.
By 1.34, W is a subspace of V. Taking I to have two elements recovers Exercise 10.
Remark on the degenerate case. If the collection is empty, then the intersection, interpreted as usual within the ambient space V (the set of $v \in V$ lying in every member of the empty collection), is all of V, which is again a subspace of V. So the assertion holds for every collection, with that standard convention.
Remark. This result is what makes the notion of the smallest subspace containing a given subset $S \subseteq V$ well defined: the intersection of all subspaces of V that contain S is itself a subspace containing S, and it is contained in every such subspace. That intersection is the span of S; compare 1.40, which identifies $V_1 + \cdots + V_m$ as the smallest subspace containing $V_1, \ldots, V_m$. Note the contrast with unions, which by Exercises 12 and 13 are almost never subspaces.

Problem 1C.12. Prove that the union of two subspaces of V is a subspace of V if and only if one of the subspaces is contained in the other.

Solution. Let U and W be subspaces of V. We prove both implications.
First suppose one of the subspaces is contained in the other. Without loss of generality $W \subseteq U$ (otherwise interchange the names of U and W). Then $U \cup W = U$, which is a subspace of V by hypothesis. Thus $U \cup W$ is a subspace of V.
Now suppose $U \cup W$ is a subspace of V; we must show $U \subseteq W$ or $W \subseteq U$. Suppose neither inclusion holds. Then there exists $u \in U$ with $u \notin W$, and there exists $w \in W$ with $w \notin U$. Since $u, w \in U \cup W$ and $U \cup W$ is closed under addition, we have

$$u + w \in U \cup W.$$

Thus $u + w \in U$ or $u + w \in W$. We rule out both.
Case 1: $u + w \in U$. Since U is a subspace, $-u \in U$ and hence

$$w = (u + w) + (-u) \in U,$$

contradicting $w \notin U$.
Case 2: $u + w \in W$. Since W is a subspace, $-w \in W$ and hence

$$u = (u + w) + (-w) \in W,$$

contradicting $u \notin W$.
Both cases are impossible, so our supposition was false: one of U, W is contained in the other, completing the proof.
Note that only the closure of $U \cup W$ under addition was used; the conclusion holds under that weaker hypothesis alone. The set $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 = 0\}$ of Exercise 8 is the union of the two coordinate axes of $\mathbb{R}^2$, neither of which contains the other, and indeed it is not a subspace.

Problem 1C.13. Prove that the union of three subspaces of V is a subspace of V if and only if one of the subspaces contains the other two.
This exercise is surprisingly harder than Exercise 12, possibly because this exercise is not true if we replace $\mathbb{F}$ with a field containing only two elements.

Solution. Let V_1, V_2, V_3 be subspaces of V and set $W = V_1 \cup V_2 \cup V_3$.
The easy direction. Suppose one of the subspaces contains the other two; without loss of generality suppose $V_2 \subseteq V_1$ and $V_3 \subseteq V_1$. Then $W = V_1$, which is a subspace of V.
The other direction. Suppose $W = V_1 \cup V_2 \cup V_3$ is a subspace of V. Assume, seeking a contradiction, that no one of V_1, V_2, V_3 contains the other two.
Step 1: no V_j is contained in the union of the other two. Suppose it were, say (after relabelling) $V_3 \subseteq V_1 \cup V_2$. Then

$$W = V_1 \cup V_2 \cup V_3 = V_1 \cup V_2,$$

so $V_1 \cup V_2$ is a subspace of V. By Exercise 12, one of V_1, V_2 contains the other; without loss of generality $V_2 \subseteq V_1$. Then $W = V_1$. But $V_3 \subseteq W = V_1$ and $V_2 \subseteq V_1$, so V_1 contains the other two subspaces, contrary to our assumption. The same argument applies with the roles of the indices permuted. Hence for each j the subspace V_j is not contained in the union of the other two.
Step 2: choose suitable vectors. By Step 1 we may choose

$$u \in V_1 \setminus (V_2 \cup V_3) \qquad \text{and} \qquad w \in V_2 \setminus (V_1 \cup V_3).$$

Step 3: the line $u + \mathbb{F}w$ lands in V_3. Let $\lambda \in \mathbb{F}$ with $\lambda \neq 0$. Since $u, w \in W$ and W is a subspace, $u + \lambda w \in W = V_1 \cup V_2 \cup V_3$. We rule out the first two possibilities.
If $u + \lambda w \in V_1$, then, since $u \in V_1$ and V_1 is a subspace, $\lambda w = (u + \lambda w) - u \in V_1$; multiplying by λ^{-1} (available because $\lambda \neq 0$) gives $w \in V_1$, contradicting the choice of w.

If $u + \lambda w \in V_2$, then, since $w \in V_2$ and V_2 is a subspace, $\lambda w \in V_2$ and so $u = (u + \lambda w) - \lambda w \in V_2$, contradicting the choice of u.
Therefore

$$u + \lambda w \in V_3 \qquad \text{for every nonzero } \lambda \in \mathbb{F}.$$

Step 4: the contradiction. Here we use that $\mathbb{F}$ (which is $\mathbb{R}$ or $\mathbb{C}$) has at least three elements: the scalars 1 and $2 = 1 + 1$ are both nonzero and are distinct from each other. By Step 3,

$$u + 1 \cdot w \in V_3 \qquad \text{and} \qquad u + 2w \in V_3.$$

Since V_3 is a subspace, it is closed under subtraction, so

$$w = (u + 2w) - (u + w) \in V_3,$$

contradicting $w \notin V_3$.
This contradiction shows that our assumption was false. Hence one of V_1, V_2, V_3 contains the other two, completing the proof.
Remark on the two-element field. The hypothesis that $\mathbb{F}$ has more than two elements is genuinely needed, and Step 4 is exactly where it enters: over a field with only two elements the only nonzero scalar is 1, so Step 3 produces just one vector and yields no contradiction. Concretely, let $\mathbb{F}_2$ denote the field with two elements and take $V = \mathbb{F}_2^2$ with

$$V_1 = \{(0,0), (1,0)\}, \qquad V_2 = \{(0,0), (0,1)\}, \qquad V_3 = \{(0,0), (1,1)\}.$$

Each V_j is a subspace of V, no one of them contains another, yet $V_1 \cup V_2 \cup V_3$ contains all four vectors of $\mathbb{F}_2^2$ and so equals V, a subspace.
Remark. The same argument generalizes: if $\mathbb{F}$ has more than m elements (in particular if $\mathbb{F}$ is infinite, as $\mathbb{R}$ and $\mathbb{C}$ are), then a vector space over $\mathbb{F}$ is never the union of m proper subspaces.

Problem 1C.14. Suppose

$$U = \{(x, -x, 2x) \in \mathbb{F}^3 : x \in \mathbb{F}\} \qquad \text{and} \qquad W = \{(x, x, 2x) \in \mathbb{F}^3 : x \in \mathbb{F}\}.$$

Describe $U + W$ using symbols, and also give a description of $U + W$ that uses no symbols.

Solution. The answer is

$$U + W = \{(x, y, 2x) \in \mathbb{F}^3 : x, y \in \mathbb{F}\},$$

which in words is: $U+W$ is the set of elements of $\mathbb{F}^3$ whose third coordinate is twice the first coordinate. By 1.36, $U + W = \{u + w : u \in U,\ w \in W\}$. We prove the displayed equality by proving both inclusions.
First inclusion. A typical element of U is $(a, -a, 2a)$ with $a \in \mathbb{F}$, and a typical element of W is $(b, b, 2b)$ with $b \in \mathbb{F}$. Their sum is

$$(a, -a, 2a) + (b, b, 2b) = (a + b,\ b - a,\ 2a + 2b) = \big(a + b,\ b - a,\ 2(a + b)\big).$$

Setting $x = a+b$ and $y = b-a$, this is $(x, y, 2x)$. Hence every element of $U+W$ has the form $(x, y, 2x)$, that is,

$$U + W \subseteq \{(x, y, 2x) \in \mathbb{F}^3 : x, y \in \mathbb{F}\}.$$

Second inclusion. Suppose $x, y \in \mathbb{F}$. Because $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, the scalar 2 is invertible in $\mathbb{F}$, so we may set

$$a = \frac{x - y}{2}, \qquad b = \frac{x + y}{2}.$$

Then $a + b = x$ and $b - a = y$, so by the computation above

$$(a, -a, 2a) + (b, b, 2b) = \big(a + b,\ b - a,\ 2(a + b)\big) = (x, y, 2x),$$

with $(a, -a, 2a) \in U$ and $(b, b, 2b) \in W$. Hence $(x, y, 2x) \in U + W$, giving

$$\{(x, y, 2x) \in \mathbb{F}^3 : x, y \in \mathbb{F}\} \subseteq U + W.$$

The two inclusions together prove the claimed equality.
Remark. This sum is in fact a direct sum. If $v \in U \cap W$, then $v = (a, -a, 2a) = (b, b, 2b)$ for some $a, b \in \mathbb{F}$; comparing first coordinates gives $a = b$, and comparing second coordinates gives $-a = b$, so $2a = 0$ and hence $a = 0$. Thus $U \cap W = \{0\}$, and $U + W$ is a direct sum by 1.46. Geometrically (taking $\mathbb{F} = \mathbb{R}$), U and W are two distinct lines through the origin in $\mathbb{R}^3$, and $U \oplus W$ is the plane through the origin that contains them both, namely the plane with equation $x_3 = 2x_1$.

Problem 1C.15. Suppose U is a subspace of V. What is $U + U$?

Solution. We claim $U + U = U$.
By the definition of the sum of subspaces (1.36),

$$U + U = \{u_1 + u_2 : u_1 \in U,\ u_2 \in U\}.$$

First we show $U + U \subseteq U$. If $u_1, u_2 \in U$, then $u_1 + u_2 \in U$ because U, being a subspace of V, is closed under addition. Hence every element of $U + U$ lies in U.
Next we show $U \subseteq U + U$. Let $u \in U$. Because U is a subspace, $0 \in U$, and therefore

$$u = u + 0$$

exhibits u as a sum of an element of U and an element of U. Thus $u \in U + U$.
Combining the two inclusions gives $U + U = U$.
(Alternatively, this follows at once from 1.40: $U + U$ is the smallest subspace of V containing U and U, and the smallest subspace of V containing U is U itself, since U is already a subspace.)

Problem 1C.16. Is the operation of addition on the subspaces of V commutative? In other words, if U and W are subspaces of V, is $U + W = W + U$?

Solution. Yes, addition of subspaces is commutative.
Let U and W be subspaces of V. By 1.36,

$$U + W = \{u + w : u \in U,\ w \in W\}, \qquad W + U = \{w + u : w \in W,\ u \in U\}.$$

Suppose $v \in U + W$. Then $v = u + w$ for some $u \in U$ and some $w \in W$. Addition in the vector space V is commutative (this is one of the axioms in the definition 1.20 of a vector space), so

$$v = u + w = w + u,$$

and the right side exhibits v as the sum of an element of W and an element of U. Hence $v \in W + U$, and therefore $U + W \subseteq W + U$.
Interchanging the roles of U and W in the argument above gives $W + U \subseteq U + W$.
Thus $U + W = W + U$, as desired.
The same reasoning applies to sums of more than two subspaces: because addition in V is commutative and associative, $V_1 + \cdots + V_m$ is unchanged by permuting $V_1, \ldots, V_m$.

Problem 1C.17. Is the operation of addition on the subspaces of V associative? In other words, if V_1, V_2, V_3 are subspaces of V, is

$$(V_1 + V_2) + V_3 = V_1 + (V_2 + V_3)?$$

Solution. Yes, addition of subspaces is associative. We show that both sides equal

$$S = \{v_1 + v_2 + v_3 : v_1 \in V_1,\ v_2 \in V_2,\ v_3 \in V_3\},$$

which is the set $V_1 + V_2 + V_3$ of 1.36. (Note that S is well defined without parentheses, because addition in V is associative by the vector space axioms in 1.20.)
First we show $(V_1 + V_2) + V_3 = S$.
Suppose $v \in (V_1 + V_2) + V_3$. By 1.36 applied to the two subspaces $V_1 + V_2$ and V_3 (recall that $V_1 + V_2$ is a subspace of V, by 1.40), we can write $v = w + v_3$ with $w \in V_1 + V_2$ and $v_3 \in V_3$. Applying 1.36 again to w, we can write $w = v_1 + v_2$ with $v_1 \in V_1$ and $v_2 \in V_2$. Hence

$$v = (v_1 + v_2) + v_3 = v_1 + v_2 + v_3 \in S.$$

Thus $(V_1 + V_2) + V_3 \subseteq S$.
Conversely, suppose $v = v_1 + v_2 + v_3$ with $v_k \in V_k$ for each k. Then $v_1 + v_2 \in V_1 + V_2$, and so

$$v = (v_1 + v_2) + v_3 \in (V_1 + V_2) + V_3.$$

Thus $S \subseteq (V_1 + V_2) + V_3$, and therefore $(V_1 + V_2) + V_3 = S$.
The proof that $V_1 + (V_2 + V_3) = S$ is the same, with the parentheses moved: an element of $V_1 + (V_2 + V_3)$ has the form $v_1 + w$ with $v_1 \in V_1$ and $w = v_2 + v_3$ for some $v_2 \in V_2$, $v_3 \in V_3$, giving $v = v_1 + (v_2 + v_3) = v_1 + v_2 + v_3 \in S$; and conversely any $v_1 + v_2 + v_3$ with $v_k \in V_k$ equals $v_1 + (v_2 + v_3) \in V_1 + (V_2 + V_3)$. Hence

$$(V_1 + V_2) + V_3 = S = V_1 + (V_2 + V_3),$$

so addition of subspaces is associative.

Problem 1C.18. Does the operation of addition on the subspaces of V have an additive identity? Which subspaces have additive inverses?

Solution. Yes: the subspace $\{0\}$ is an additive identity, and it is the only one. The only subspace with an additive inverse is $\{0\}$ itself.
The subspace $\{0\}$ is an additive identity. Let U be a subspace of V. If $u \in U$, then $u = u + 0$ with $u \in U$ and $0 \in \{0\}$, so $u \in U + \{0\}$. Conversely, every element of $U + \{0\}$ has the form $u + 0 = u$ for some $u \in U$, hence lies in U. Therefore

$$U + \{0\} = U$$

for every subspace U of V. By 1C.16 also $\{0\} + U = U$, so $\{0\}$ is a two-sided identity.
It is the only additive identity. Suppose E is a subspace of V with $U + E = U$ for every subspace U of V. Taking $U = \{0\}$ gives $\{0\} + E = \{0\}$. On the other hand, the previous paragraph applied with $U = E$ gives $E + \{0\} = E$, and hence $\{0\} + E = E$ by commutativity (1C.16). Comparing the two computations of $\{0\} + E$ yields $E = \{0\}$.
Now consider additive inverses. Suppose U and W are subspaces of V with

$$U + W = \{0\}.$$

By 1.40 the sum $U + W$ contains U; explicitly, if $u \in U$ then $u = u + 0 \in U + W$. So $U \subseteq U + W = \{0\}$, forcing $U = \{0\}$. The same argument gives $W = \{0\}$.
Consequently a subspace U has an additive inverse if and only if $U = \{0\}$, and in that case its additive inverse is $\{0\}$ itself (indeed $\{0\} + \{0\} = \{0\}$).
Thus addition of subspaces of V is commutative (1C.16), associative (1C.17), and has an identity, but inverses fail badly: the collection of subspaces of V under addition is a commutative monoid, not a group, unless $V = \{0\}$.

Problem 1C.19. Prove or give a counterexample: If V_1, V_2, U are subspaces of V such that

$$V_1 + U = V_2 + U,$$

then $V_1 = V_2$.

Solution. The statement is false; in other words, addition of subspaces does not satisfy a cancellation law. Here is a counterexample.
Take $V = \mathbf{F}^2$ and set

$$V_1 = \{(x, 0) \in \mathbf{F}^2 : x \in \mathbf{F}\}, \qquad V_2 = \{(0, y) \in \mathbf{F}^2 : y \in \mathbf{F}\}, \qquad U = \mathbf{F}^2.$$

Each of V_1, V_2, U is a subspace of $\mathbf{F}^2$: each contains $(0, 0)$ and is clearly closed under addition and scalar multiplication, so 1.34 applies.
Now

$$V_1 + U = U = \mathbf{F}^2 \quad \text{and} \quad V_2 + U = U = \mathbf{F}^2.$$

Here each equality $V_k + U = U$ holds because $V_k \subseteq U$. Indeed, in that case every element $v + u$ with $v \in V_k$ and $u \in U$ lies in U, since U is closed under addition; this gives $V_k + U \subseteq U$. Conversely $U \subseteq V_k + U$, because each $u \in U$ can be written as $u = 0 + u$ with $0 \in V_k$. (Alternatively: by 1.40 the subspace $V_k + U$ is the smallest subspace of V containing both V_k and U, and when $V_k \subseteq U$ that smallest subspace is U itself.)
Hence $V_1 + U = V_2 + U$. But

$$(1, 0) \in V_1 \quad \text{and} \quad (1, 0) \notin V_2,$$

so $V_1 \neq V_2$. This is a counterexample.
The failure has nothing to do with the particular choice above. The obstruction is exactly the failure of additive inverses found in 1C.18: in a commutative monoid, cancellation would follow from the existence of inverses, and here the only invertible element is $\{0\}$. Indeed, the same phenomenon occurs whenever U is large enough to swallow both V_1 and V_2: if $V_1 \subseteq U$ and $V_2 \subseteq U$, then $V_1 + U = U = V_2 + U$ regardless of how different V_1 and V_2 are.
Cancellation can also fail with U a proper subspace of V. In $V = \mathbf{F}^3$, take

$$V_1 = \{(x, 0, 0) : x \in \mathbf{F}\}, \quad V_2 = \{(0, y, 0) : y \in \mathbf{F}\}, \quad U = \{(x, y, 0) : x, y \in \mathbf{F}\}.$$

Then $V_1 \subseteq U$ and $V_2 \subseteq U$, so $V_1 + U = U = V_2 + U$, while $V_1 \neq V_2$ and $U \neq V$.

Problem 1C.20. Suppose

$$U = \{(x, x, y, y) \in \mathbf{F}^4 : x, y \in \mathbf{F}\}.$$

Find a subspace W of $\mathbf{F}^4$ such that $\mathbf{F}^4 = U \oplus W$.

Solution. Take

$$W = \{(x, 0, y, 0) \in \mathbf{F}^4 : x, y \in \mathbf{F}\}.$$

First, W is a subspace of $\mathbf{F}^4$. It contains $(0, 0, 0, 0)$, it is closed under addition since

$$(x, 0, y, 0) + (x', 0, y', 0) = (x + x', 0, y + y', 0),$$

and it is closed under scalar multiplication since $\lambda(x, 0, y, 0) = (\lambda x, 0, \lambda y, 0)$. Thus 1.34 shows W is a subspace. (The same three checks show U is a subspace of $\mathbf{F}^4$.)
Next we show $\mathbf{F}^4 = U + W$. Let $(a, b, c, d) \in \mathbf{F}^4$. Then

$$(a, b, c, d) = (b, b, d, d) + (a - b, 0, c - d, 0),$$

where $(b, b, d, d) \in U$ and $(a - b, 0, c - d, 0) \in W$. Hence $(a, b, c, d) \in U + W$. Since $U + W \subseteq \mathbf{F}^4$ trivially, we get $\mathbf{F}^4 = U + W$.

Finally we show the sum is direct. By 1.46 it suffices to prove $U \cap W = \{0\}$. Suppose $v \in U \cap W$. Because $v \in U$, we have $v = (x, x, y, y)$ for some $x, y \in \mathbf{F}$. Because $v \in W$, the second and fourth coordinates of v are 0; that is, $x = 0$ and $y = 0$. Hence $v = (0, 0, 0, 0)$. Therefore $U \cap W = \{0\}$, and 1.46 gives that $U + W$ is a direct sum.
Combining the last two paragraphs,

$$\mathbf{F}^4 = U \oplus W.$$

This choice of W is far from unique. For instance $W' = \{(0, x, 0, y) : x, y \in \mathbf{F}\}$ works by the symmetric argument, as does $W'' = \{(0, x, y, 0) : x, y \in \mathbf{F}\}$: for the latter, $(a, b, c, d) = (a, a, d, d)+(0, b-a, c-d, 0)$, and a vector (x, x, y, y) in W'' must have $x = 0$ (first coordinate) and $y = 0$ (fourth coordinate).

Problem 1C.21. Suppose

$$U = \{(x, y, x + y, x - y, 2x) \in \mathbf{F}^5 : x, y \in \mathbf{F}\}.$$

Find a subspace W of $\mathbf{F}^5$ such that $\mathbf{F}^5 = U \oplus W$.

Solution. Take

$$W = \{(0, 0, z_1, z_2, z_3) \in \mathbf{F}^5 : z_1, z_2, z_3 \in \mathbf{F}\}.$$

First, W is a subspace of $\mathbf{F}^5$: it contains $(0, 0, 0, 0, 0)$, and

$$(0, 0, z_1, z_2, z_3)+(0, 0, z_1', z_2', z_3') = (0, 0, z_1+z_1', z_2+z_2', z_3+z_3'), \qquad \lambda(0, 0, z_1, z_2, z_3) = (0, 0, \lambda z_1, \lambda z_2, \lambda z_3),$$

so W is closed under addition and scalar multiplication; now apply 1.34.
(Similarly U is a subspace: it contains the vector obtained by taking $x = y = 0$, and writing $u(x, y) = (x, y, x + y, x - y, 2x)$ we have $u(x, y) + u(x', y') = u(x + x', y + y')$ and $\lambda\, u(x, y) = u(\lambda x, \lambda y)$.)
Next we show $\mathbf{F}^5 = U + W$. Let $(a_1, a_2, a_3, a_4, a_5) \in \mathbf{F}^5$. Put $x = a_1$ and $y = a_2$, so that

$$u = (a_1,\ a_2,\ a_1 + a_2,\ a_1 - a_2,\ 2a_1) \in U.$$

Then

$$(a_1, a_2, a_3, a_4, a_5) - u = \big(0,\ 0,\ a_3 - a_1 - a_2,\ a_4 - a_1 + a_2,\ a_5 - 2a_1\big),$$

whose first two coordinates are 0, so this difference lies in W. Calling it w, we have $(a_1, a_2, a_3, a_4, a_5) = u + w \in U + W$. Hence $\mathbf{F}^5 = U + W$.
Finally we show the sum is direct, using 1.46: we verify $U \cap W = \{0\}$. Suppose $v \in U \cap W$. Since $v \in U$, we may write

$$v = (x, y, x + y, x - y, 2x)$$

for some $x, y \in \mathbf{F}$. Since $v \in W$, its first two coordinates vanish, so $x = 0$ and $y = 0$. Then all five coordinates of v are 0, that is, $v = 0$. Hence $U \cap W = \{0\}$, and 1.46 shows that $U + W$ is a direct sum. Therefore

$$\mathbf{F}^5 = U \oplus W.$$

Note that the argument used no division and no property of $\mathbf{F}$ beyond the field axioms; in particular it is uniform in $\mathbf{F}$, which in this book denotes $\mathbf{R}$ or $\mathbf{C}$ (1.6). The underlying idea is general: U is the set of vectors whose last three coordinates are determined by the first two, so complementing it with the subspace of vectors whose first two coordinates vanish separates the free parameters from the determined ones.

Problem 1C.22. Suppose

$$U = \{(x, y, x + y, x - y, 2x) \in \mathbb{F}^5 : x, y \in \mathbb{F}\}.$$

Find three subspaces W_1, W_2, W_3 of $\mathbb{F}^5$, none of which equals $\{0\}$, such that $\mathbb{F}^5 = U \oplus W_1 \oplus W_2 \oplus W_3$.

Solution. Take

$$W_1 = \{(0,0,z,0,0) : z \in \mathbb{F}\}, \qquad W_2 = \{(0,0,0,z,0) : z \in \mathbb{F}\}, \qquad W_3 = \{(0,0,0,0,z) : z \in \mathbb{F}\}.$$

Each of these is a subspace of $\mathbb{F}^5$ by 1.34: each contains 0 (take $z = 0$) and each is closed under addition and scalar multiplication, since adding two such vectors or multiplying one by a scalar only adds or scales the single free coordinate and leaves all other coordinates equal to 0. None of them equals $\{0\}$, since each contains the vector with 1 in its free slot and 0 elsewhere.
The set U is also a subspace of $\mathbb{F}^5$, again by 1.34: taking $x = y = 0$ shows $0 \in U$, and

$$(x,y,x+y,x-y,2x)+(x',y',x'+y',x'-y',2x') = (x+x',y+y',(x+x')+(y+y'),(x+x')-(y+y'),2(x+x')),$$

$$\lambda(x,y,x+y,x-y,2x) = (\lambda x,\lambda y,\lambda x+\lambda y,\lambda x-\lambda y,2\lambda x),$$

so U is closed under addition and scalar multiplication. (We need this because the results on direct sums quoted below apply to subspaces.)
First we check that the sum is all of $\mathbb{F}^5$. Let $(a_1,a_2,a_3,a_4,a_5) \in \mathbb{F}^5$. Put $u = (a_1,a_2,a_1+a_2,a_1-a_2,2a_1)$, which lies in U (take $x = a_1$ and $y = a_2$ in the definition of U). Then

$$(a_1,a_2,a_3,a_4,a_5) - u = \big(0,0,a_3-a_1-a_2,a_4-a_1+a_2,a_5-2a_1\big),$$

and this vector is the sum

$$(0,0,a_3-a_1-a_2,0,0)+(0,0,0,a_4-a_1+a_2,0)+(0,0,0,0,a_5-2a_1)$$

of a vector in W_1, a vector in W_2, and a vector in W_3. Hence

$$(a_1,a_2,a_3,a_4,a_5) \in U+W_1+W_2+W_3,$$

so $\mathbb{F}^5 = U+W_1+W_2+W_3$ (the reverse inclusion is automatic, because U, W_1, W_2, W_3 are all subsets of $\mathbb{F}^5$).
Next we check that the sum is direct. By 1.45 it suffices to show that the only way to write 0 as $u+w_1+w_2+w_3$ with $u \in U$ and $w_k \in W_k$ is with all four vectors equal to 0. So suppose

$$(0,0,0,0,0) = (x,y,x+y,x-y,2x)+(0,0,c,0,0)+(0,0,0,d,0)+(0,0,0,0,e)$$

for some $x,y,c,d,e \in \mathbb{F}$. Comparing the first coordinates gives $x = 0$, and comparing the second coordinates gives $y = 0$. Therefore the U-term is $(0,0,0,0,0)$, and now comparing the third, fourth, and fifth coordinates gives $c = 0$, $d = 0$, and $e = 0$. Thus all four summands equal 0, and by 1.45 the sum $U+W_1+W_2+W_3$ is a direct sum.
Combining the two paragraphs, $\mathbb{F}^5 = U \oplus W_1 \oplus W_2 \oplus W_3$.

Problem 1C.23. Prove or give a counterexample: If V_1, V_2, U are subspaces of V such that

$$V = V_1 \oplus U \quad \text{and} \quad V = V_2 \oplus U,$$

then $V_1 = V_2$.
[Hint: When trying to discover whether a conjecture in linear algebra is true or false, it is often useful to start by experimenting in $\mathbb{F}^2$.]

Solution. The statement is false. Following the hint, we produce a counterexample in $V = \mathbb{F}^2$. Let

$$U = \{(x,0) : x \in \mathbb{F}\}, \qquad V_1 = \{(0,y) : y \in \mathbb{F}\}, \qquad V_2 = \{(y,y) : y \in \mathbb{F}\}.$$

Each of these three sets is a subspace of $\mathbb{F}^2$: each contains $(0,0)$, and each is visibly closed under addition and under scalar multiplication, since in each case the defining condition is a linear condition on the coordinates.

We claim $\mathbb{F}^2 = V_1 \oplus U$. For the sum, if $(a, b) \in \mathbb{F}^2$ then

$$(a, b) = (0, b) + (a, 0),$$

with $(0, b) \in V_1$ and $(a, 0) \in U$; hence $\mathbb{F}^2 = V_1 + U$. For directness, suppose $(a, b) \in V_1 \cap U$. Membership in V_1 forces $a = 0$ and membership in U forces $b = 0$, so $V_1 \cap U = \{(0, 0)\}$. By 1.46 the sum $V_1 + U$ is a direct sum. Thus $\mathbb{F}^2 = V_1 \oplus U$.
We claim also that $\mathbb{F}^2 = V_2 \oplus U$. For the sum, if $(a, b) \in \mathbb{F}^2$ then

$$(a, b) = (b, b) + (a - b, 0),$$

with $(b, b) \in V_2$ and $(a - b, 0) \in U$; hence $\mathbb{F}^2 = V_2 + U$. For directness, suppose $(a, b) \in V_2 \cap U$. Membership in U gives $b = 0$, and membership in V_2 gives $a = b$; hence $a = b = 0$ and $V_2 \cap U = \{(0, 0)\}$. By 1.46 the sum $V_2 + U$ is a direct sum. Thus $\mathbb{F}^2 = V_2 \oplus U$.
Finally, $V_1 \neq V_2$, because $(1, 1) \in V_2$ but $(1, 1) \notin V_1$ (every element of V_1 has first coordinate 0).
So V_1 and V_2 are subspaces of $V = \mathbb{F}^2$ with $V = V_1 \oplus U = V_2 \oplus U$ and $V_1 \neq V_2$, which refutes the conjecture.
Geometrically (thinking of $\mathbb{F} = \mathbb{R}$), U is the horizontal axis and V_1, V_2 are two different lines through the origin, neither of them horizontal; each of them, together with the horizontal axis, spans the plane, and meets the horizontal axis only at the origin. So a subspace has many different complements in general, not a unique one. What is true is a weaker statement: V_1 and V_2 must have the same dimension, since by 2.43 both equal $\dim V - \dim U$ when V is finite-dimensional.

Problem 1C.24. A function $f\colon \mathbb{R} \to \mathbb{R}$ is called even if

$$f(-x) = f(x)$$

for all $x \in \mathbb{R}$. A function $f\colon \mathbb{R} \to \mathbb{R}$ is called odd if

$$f(-x) = -f(x)$$

for all $x \in \mathbb{R}$. Let V_e denote the set of real-valued even functions on $\mathbb{R}$ and let V_o denote the set of real-valued odd functions on $\mathbb{R}$. Show that $\mathbb{R}^\mathbb{R} = V_\mathrm{e} \oplus V_\mathrm{o}$.

Solution. Here $\mathbb{R}^\mathbb{R}$ denotes the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$, with addition and scalar multiplication defined pointwise. There are three things to prove: that V_e and V_o are subspaces of $\mathbb{R}^\mathbb{R}$, that their sum is all of $\mathbb{R}^\mathbb{R}$, and that the sum is direct.
Step 1: V_e and V_o are subspaces. We use 1.34. The zero function 0 satisfies $0(-x) = 0 = 0(x)$ and also $0(-x) = 0 = -0(x)$ for all $x \in \mathbb{R}$, so the zero function belongs to both V_e and V_o. If $f, g \in V_\mathrm{e}$ and $x \in \mathbb{R}$, then

$$(f + g)(-x) = f(-x) + g(-x) = f(x) + g(x) = (f + g)(x),$$

so $f + g \in V_\mathrm{e}$; and if $\lambda \in \mathbb{R}$ then

$$(\lambda f)(-x) = \lambda f(-x) = \lambda f(x) = (\lambda f)(x),$$

so $\lambda f \in V_\mathrm{e}$. Hence V_e is a subspace of $\mathbb{R}^\mathbb{R}$. The same computations with the sign flipped work for V_o: if $f, g \in V_\mathrm{o}$ and $\lambda \in \mathbb{R}$, then for all $x \in \mathbb{R}$,

$$\begin{aligned}(f + g)(-x) &= f(-x) + g(-x) = -f(x) - g(x) = -(f + g)(x),\\ (\lambda f)(-x) &= \lambda f(-x) = -\lambda f(x) = -(\lambda f)(x),\end{aligned}$$

so V_o is a subspace of $\mathbb{R}^\mathbb{R}$ as well.
Step 2: $\mathbb{R}^\mathbb{R} = V_\mathrm{e} + V_\mathrm{o}$. Let $f \in \mathbb{R}^\mathbb{R}$. Define $g, h\colon \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \frac{f(x) + f(-x)}{2}, \qquad h(x) = \frac{f(x) - f(-x)}{2}$$

for $x \in \mathbb{R}$ (this makes sense because 2 is invertible in $\mathbb{R}$). Then for every $x \in \mathbb{R}$,

$$g(-x) = \frac{f(-x) + f(x)}{2} = g(x),$$

so $g \in V_{\mathrm{e}}$, and

$$h(-x) = \frac{f(-x) - f(x)}{2} = -\frac{f(x) - f(-x)}{2} = -h(x),$$

so $h \in V_{\mathrm{o}}$. Moreover, for every $x \in \mathbb{R}$,

$$g(x) + h(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2} = f(x),$$

so $f = g + h \in V_{\mathrm{e}} + V_{\mathrm{o}}$. Since f was arbitrary, $\mathbb{R}^{\mathbb{R}} = V_{\mathrm{e}} + V_{\mathrm{o}}$ (the inclusion $V_{\mathrm{e}} + V_{\mathrm{o}} \subseteq \mathbb{R}^{\mathbb{R}}$ is clear, as both are subspaces of $\mathbb{R}^{\mathbb{R}}$).

Step 3: the sum is direct. By 1.46 it suffices to show $V_{\mathrm{e}} \cap V_{\mathrm{o}} = \{0\}$. Suppose $f \in V_{\mathrm{e}} \cap V_{\mathrm{o}}$ and let $x \in \mathbb{R}$. Because f is even, $f(-x) = f(x)$; because f is odd, $f(-x) = -f(x)$. Hence $f(x) = -f(x)$, so $2f(x) = 0$ and therefore $f(x) = 0$. As $x \in \mathbb{R}$ was arbitrary, f is the zero function. Thus $V_{\mathrm{e}} \cap V_{\mathrm{o}} = \{0\}$, and by 1.46 the sum $V_{\mathrm{e}} + V_{\mathrm{o}}$ is a direct sum.

Combining the three steps, $\mathbb{R}^{\mathbb{R}} = V_{\mathrm{e}} \oplus V_{\mathrm{o}}$.

2 Finite-Dimensional Vector Spaces

Contents

2.1 Exercises 2A 36
2.2 Exercises 2B 48
2.3 Exercises 2C 56

2.1 Exercises 2A

Problem 2A.1. Find a list of four distinct vectors in $\mathbf{F}^3$ whose span equals

$$\{(x, y, z) \in \mathbf{F}^3 : x + y + z = 0\}.$$

Solution. Write $U = \{(x, y, z) \in \mathbf{F}^3 : x + y + z = 0\}$, and consider the list

$$(1, 0, -1), \quad (0, 1, -1), \quad (1, -1, 0), \quad (1, 1, -2).$$

These four vectors are pairwise distinct (compare first coordinates to separate the second vector from the others, and second coordinates to separate the remaining three). We claim that their span equals U.
First, $\operatorname{span}\big((1, 0, -1), (0, 1, -1), (1, -1, 0), (1, 1, -2)\big) \subseteq U$. Indeed, U is a subspace of $\mathbf{F}^3$: it contains $(0, 0, 0)$, and if $(x, y, z), (x', y', z') \in U$ and $\lambda \in \mathbf{F}$ then

$$(x + x') + (y + y') + (z + z') = (x + y + z) + (x' + y' + z') = 0$$

and $\lambda x + \lambda y + \lambda z = \lambda(x + y + z) = 0$, so U is closed under addition and scalar multiplication. Each of the four listed vectors lies in U, because the coordinates of each sum to 0:

$$1 + 0 + (-1) = 0, \qquad 0 + 1 + (-1) = 0, \qquad 1 + (-1) + 0 = 0, \qquad 1 + 1 + (-2) = 0.$$

By 2.6 the span of a list is the smallest subspace containing all the vectors in the list, so the span of our list is contained in U.
Conversely, $U \subseteq \operatorname{span}\big((1, 0, -1), (0, 1, -1), (1, -1, 0), (1, 1, -2)\big)$. Suppose $(x, y, z) \in U$. Then $z = -x - y$, so

$$(x, y, z) = (x,\, y,\, -x - y) = x(1, 0, -1) + y(0, 1, -1) + 0\,(1, -1, 0) + 0\,(1, 1, -2),$$

which exhibits (x, y, z) as a linear combination of the four vectors in the list.
The two inclusions give

$$\operatorname{span}\big((1, 0, -1), (0, 1, -1), (1, -1, 0), (1, 1, -2)\big) = U,$$

as required. (Of course the answer is far from unique: any four distinct vectors of U that include two vectors spanning the plane U will do. Note also that a list of four vectors spanning this 2-dimensional set must be linearly dependent, which is why the last two entries were free to be redundant.)

Problem 2A.2. Prove or give a counterexample: If v_1, v_2, v_3, v_4 spans V, then the list

$$v_1 - v_2, \quad v_2 - v_3, \quad v_3 - v_4, \quad v_4$$

also spans V.

Solution. The statement is true; we prove it.
Suppose v_1, v_2, v_3, v_4 spans V, and write

$$W = \operatorname{span}(v_1 - v_2,\, v_2 - v_3,\, v_3 - v_4,\, v_4).$$

By 2.6, W is a subspace of V. We show that v_1, v_2, v_3, v_4 all lie in W.

Working from the last vector backwards:

$$\begin{aligned} v_4 &= v_4, \\ v_3 &= (v_3 - v_4) + v_4, \\ v_2 &= (v_2 - v_3) + (v_3 - v_4) + v_4, \\ v_1 &= (v_1 - v_2) + (v_2 - v_3) + (v_3 - v_4) + v_4. \end{aligned}$$

Each right-hand side is a linear combination (with all coefficients equal to 1 or 0) of the vectors $v_1 - v_2,\ v_2 - v_3,\ v_3 - v_4,\ v_4$, so $v_1, v_2, v_3, v_4 \in W$.
Now W is a subspace of V containing v_1, v_2, v_3, v_4. By 2.6, $\operatorname{span}(v_1, v_2, v_3, v_4)$ is the smallest such subspace, so

$$V = \operatorname{span}(v_1, v_2, v_3, v_4) \subseteq W.$$

Since $W \subseteq V$ as well, we conclude $W = V$; that is, the list $v_1 - v_2,\ v_2 - v_3,\ v_3 - v_4,\ v_4$ spans V.
(Alternatively, one can verify the inclusion directly: given $u \in V$, write $u = a_1v_1 + a_2v_2 + a_3v_3 + a_4v_4$ and substitute the displayed expressions for v_1, v_2, v_3, v_4 to obtain

$$u = a_1(v_1 - v_2) + (a_1 + a_2)(v_2 - v_3) + (a_1 + a_2 + a_3)(v_3 - v_4) + (a_1 + a_2 + a_3 + a_4)v_4,$$

which displays u as a linear combination of the new list.)

Problem 2A.3. Suppose $v_1, \dots, v_m$ is a list of vectors in V. For $k \in \{1, \dots, m\}$, let

$$w_k = v_1 + \cdots + v_k.$$

Show that $\operatorname{span}(v_1, \dots, v_m) = \operatorname{span}(w_1, \dots, w_m)$.

Solution. Write $A = \operatorname{span}(v_1, \dots, v_m)$ and $B = \operatorname{span}(w_1, \dots, w_m)$. By 2.6 both A and B are subspaces of V, and each is the smallest subspace containing the vectors of its defining list. So it suffices to show that every w_k lies in A and that every v_k lies in B.
Each w_k lies in A. Indeed, by definition

$$w_k = v_1 + \cdots + v_k = 1\,v_1 + \cdots + 1\,v_k + 0\,v_{k+1} + \cdots + 0\,v_m,$$

which is a linear combination of $v_1, \dots, v_m$. Hence $w_k \in A$ for each $k \in \{1, \dots, m\}$. Since A is a subspace containing $w_1, \dots, w_m$, the minimality in 2.6 gives $B \subseteq A$.
Each v_k lies in B. For $k = 1$ we have $v_1 = w_1 \in B$. For $k \in \{2, \dots, m\}$,

$$w_k - w_{k-1} = (v_1 + \cdots + v_k) - (v_1 + \cdots + v_{k-1}) = v_k,$$

so $v_k = w_k - w_{k-1}$ is a linear combination of $w_1, \dots, w_m$ (take coefficient 1 on w_k, coefficient -1 on w_{k-1}, and 0 elsewhere). Hence $v_k \in B$ for each k. Since B is a subspace containing $v_1, \dots, v_m$, minimality in 2.6 gives $A \subseteq B$.
Combining the two inclusions, $A = B$, that is,

$$\operatorname{span}(v_1, \dots, v_m) = \operatorname{span}(w_1, \dots, w_m).$$

(If one prefers an explicit computation rather than an appeal to 2.6: given scalars $a_1, \dots, a_m$, the identity $v_k = w_k - w_{k-1}$ with the convention $w_0 = 0$ yields

$$\sum_{k=1}^{m} a_k v_k = \sum_{k=1}^{m} a_k(w_k - w_{k-1}) = \sum_{k=1}^{m-1} (a_k - a_{k+1})w_k + a_m w_m,$$

after summation by parts, and conversely $\sum_{k=1}^{m} b_k w_k = \sum_{k=1}^{m} (b_k + b_{k+1} + \cdots + b_m)v_k$. Each display converts a linear combination of one list into a linear combination of the other.)

Problem 2A.4. (a) Show that a list of length one in a vector space is linearly independent if and only if the vector in the list is not 0.
(b) Show that a list of length two in a vector space is linearly independent if and only if neither of the two vectors in the list is a scalar multiple of the other.

Solution. Throughout, V is a vector space over $\mathbf{F}$. Recall the definition 2.15: a list $v_1, \dots, v_m$ is linearly independent if the only choice of $a_1, \dots, a_m \in \mathbf{F}$ with $a_1 v_1 + \cdots + a_m v_m = 0$ is $a_1 = \cdots = a_m = 0$.
(a) Let $v \in V$ and consider the list of length one consisting of v.
Suppose first that $v \neq 0$. Let $a \in \mathbf{F}$ satisfy $av = 0$. If $a \neq 0$, then a is invertible in the field $\mathbf{F}$, and

$$v = 1\,v = \big(a^{-1}a\big)v = a^{-1}(av) = a^{-1}\,0 = 0$$

by 1.31, contradicting $v \neq 0$. Hence $a = 0$, and the list v is linearly independent.
Conversely, suppose $v = 0$. Then $1 \cdot v = 1 \cdot 0 = 0$ by 1.31, and the scalar 1 is not 0 (in a field $1 \neq 0$). So there is a choice of scalar that is not 0 producing the zero vector, and the list v is linearly dependent (2.17).
Thus the list v is linearly independent if and only if $v \neq 0$.
(b) Let $v, w \in V$ and consider the list v, w.
Suppose the list v, w is linearly dependent. By 2.17 there exist $a, b \in \mathbf{F}$, not both 0, with

$$av + bw = 0.$$

If $a \neq 0$, then multiplying by a^{-1} gives $v + a^{-1}b\,w = 0$, hence

$$v = -\big(a^{-1}b\big)w,$$

so v is a scalar multiple of w. If instead $a = 0$, then $b \neq 0$, and $bw = 0$ forces $w = b^{-1}(bw) = b^{-1}0 = 0$; then $w = 0 = 0\,v$ by 1.30, so w is a scalar multiple of v. In either case one of the two vectors is a scalar multiple of the other.
Conversely, suppose one of the two vectors is a scalar multiple of the other. If $v = \lambda w$ for some $\lambda \in \mathbf{F}$, then

$$1 \cdot v + (-\lambda)w = v - \lambda w = 0,$$

and the coefficients $1, -\lambda$ are not both 0 (since $1 \neq 0$); hence the list v, w is linearly dependent. Symmetrically, if $w = \mu v$ for some $\mu \in \mathbf{F}$, then $(-\mu)v + 1 \cdot w = 0$ with coefficients not both 0, so again the list is linearly dependent.
Taking the contrapositive of both implications: the list v, w is linearly independent if and only if neither of v, w is a scalar multiple of the other.
(These are exactly the assertions recorded without proof in 2.16(c) and 2.16(d). One direction of part (a) can also be read off the linear dependence lemma 2.19: if the list v of length one is linearly dependent, then 2.19 supplies $k = 1$ with $v \in \operatorname{span}() = \{0\}$, so $v = 0$. The converse direction still has to be checked by hand, as above.)

Problem 2A.5. Find a number t such that

$$(3, 1, 4), \quad (2, -3, 5), \quad (5, 9, t)$$

is not linearly independent in $\mathbf{R}^3$.

Solution. The answer is $t = 2$.
Verification. With $t = 2$ we exhibit an explicit nontrivial linear combination equal to 0:

$$(-3)(3, 1, 4) + 2(2, -3, 5) + 1\,(5, 9, 2) = (-9 + 4 + 5,\ -3 - 6 + 9,\ -12 + 10 + 2) = (0, 0, 0).$$

Since the scalars $-3, 2, 1$ are not all 0, the list $(3, 1, 4), (2, -3, 5), (5, 9, 2)$ is linearly dependent, i.e. not linearly independent (2.17). This settles the exercise.

How the value was found. Suppose $a, b, c \in \mathbf{R}$ satisfy

$$a(3,1,4) + b(2,-3,5) + c(5,9,t) = (0,0,0),$$

which is the system

$$\begin{aligned} 3a + 2b + 5c &= 0, \\ a - 3b + 9c &= 0, \\ 4a + 5b + tc &= 0. \end{aligned}$$

The list fails to be linearly independent exactly when this system has a solution with $(a,b,c) \neq (0,0,0)$. First note that $(3,1,4), (2,-3,5)$ is linearly independent. Both vectors are nonzero, so a scalar relating one to the other would be nonzero and hence invertible; it is enough to rule out $(2,-3,5) = \lambda(3,1,4)$. Comparing second coordinates forces $\lambda = -3$, but then the first coordinate would be $3\lambda = -9 \neq 2$. So neither vector is a scalar multiple of the other, and Exercise 2A.4(b) (recorded in the book as 2.16(d)) shows that the list of length two is linearly independent. Consequently, in any nontrivial solution we must have $c \neq 0$; otherwise $a(3,1,4) + b(2,-3,5) = 0$ would force $a = b = 0$ too. So we may scale to $c = 1$ and ask when $(5,9,t)$ is a linear combination of the first two vectors.
Solving the first two equations with $c = 1$: from $a = 3b - 9$ (the second equation) substituted into the first,

$$3(3b - 9) + 2b + 5 = 11b - 22 = 0,$$

so $b = 2$ and $a = -3$. The third equation then reads $4(-3) + 5(2) + t = t - 2 = 0$, i.e. $t = 2$. Thus $t = 2$ is in fact the only value of t for which the list fails to be linearly independent; for every other t the list is linearly independent.
(Equivalently, the 3×3 determinant with these three vectors as rows equals

$$3\big((-3)t - 45\big) - 1\big(2t - 25\big) + 4\big(18 + 15\big) = -11t + 22,$$

which vanishes precisely when $t = 2$. Determinants are not needed for the solution above, and are only introduced in Chapter 9 of the book.)

Problem 2A.6. Show that the list $(2,3,1), (1,-1,2), (7,3,c)$ is linearly dependent in $\mathbf{F}^3$ if and only if $c = 8$.

Solution. We prove both implications.
Suppose $c = 8$. Then

$$2(2,3,1) + 3(1,-1,2) + (-1)(7,3,8) = (4 + 3 - 7,\ 6 - 3 - 3,\ 2 + 6 - 8) = (0,0,0),$$

and the scalars $2, 3, -1$ are not all 0. By 2.17 the list $(2,3,1), (1,-1,2), (7,3,8)$ is linearly dependent. (This is exactly the computation displayed in 2.18.)
Conversely, suppose the list $(2,3,1), (1,-1,2), (7,3,c)$ is linearly dependent. By 2.17 there exist $a, b, d \in \mathbf{F}$, not all 0, with

$$a(2,3,1) + b(1,-1,2) + d(7,3,c) = (0,0,0).$$

We first observe that the list $(2,3,1), (1,-1,2)$ is linearly independent. Both vectors are nonzero, so if either were a scalar multiple of the other, the scalar would be nonzero and could be inverted; it therefore suffices to rule out $(1,-1,2) = \lambda(2,3,1)$ for $\lambda \in \mathbf{F}$. Comparing third coordinates gives $\lambda = 2$, while comparing first coordinates gives $2\lambda = 1$; these are incompatible, since $\lambda = 2$ yields $2\lambda = 4 \neq 1$. Hence neither vector is a scalar multiple of the other, and by Exercise 2A.4(b) (recorded in the book as 2.16(d)) the list of length two is linearly independent.
Now $d \neq 0$. For if $d = 0$, the displayed equation becomes $a(2,3,1) + b(1,-1,2) = (0,0,0)$, and linear independence of $(2,3,1), (1,-1,2)$ forces $a = b = 0$; then a, b, d would all be 0, contrary to assumption. Since $d \neq 0$ we may divide by d and set $\alpha = -a/d$ and $\beta = -b/d$, obtaining

$$(7,3,c) = \alpha(2,3,1) + \beta(1,-1,2),$$

that is,

$$2\alpha + \beta = 7, \qquad 3\alpha - \beta = 3, \qquad \alpha + 2\beta = c.$$

Adding the first two equations gives $5\alpha = 10$, so $\alpha = 2$, and then $\beta = 7 - 2\alpha = 3$. Substituting into the third equation,

$$c = \alpha + 2\beta = 2 + 6 = 8.$$

Therefore the list $(2, 3, 1), (1, -1, 2), (7, 3, c)$ is linearly dependent in $\mathbf{F}^3$ if and only if $c = 8$, verifying the claim asserted in 2.18.

Problem 2A.7. (a) Show that if we think of $\mathbf{C}$ as a vector space over $\mathbf{R}$, then the list $1 + i,\ 1 - i$ is linearly independent.
(b) Show that if we think of $\mathbf{C}$ as a vector space over $\mathbf{C}$, then the list $1+i,\ 1-i$ is linearly dependent.

Solution. The point of the exercise is that linear independence depends on the scalar field, not just on the vectors: the same list of two complex numbers is independent over $\mathbf{R}$ and dependent over $\mathbf{C}$.
(a) Regard $\mathbf{C}$ as a vector space over $\mathbf{R}$, so the allowed scalars are real numbers. Suppose $a, b \in \mathbf{R}$ satisfy

$$a(1 + i) + b(1 - i) = 0.$$

Expanding and collecting real and imaginary parts,

$$(a + b) + (a - b)i = 0.$$

Since $a + b$ and $a - b$ are real numbers, and a complex number is 0 if and only if its real and imaginary parts are both 0, we get the system

$$a + b = 0, \qquad a - b = 0.$$

Adding these gives $2a = 0$, so $a = 0$; subtracting gives $2b = 0$, so $b = 0$. Thus the only real scalars producing 0 are $a = b = 0$, and by 2.15 the list $1 + i,\ 1 - i$ is linearly independent in $\mathbf{C}$ viewed as a vector space over $\mathbf{R}$.
(b) Now regard $\mathbf{C}$ as a vector space over $\mathbf{C}$, so complex scalars are allowed. Observe that

$$(-i)(1 + i) = -i - i^2 = -i + 1 = 1 - i,$$

so $1 - i$ is a scalar multiple of $1 + i$, with the scalar $-i \in \mathbf{C}$. By Exercise 2A.4(b) (recorded in the book as 2.16(d)) the list $1 + i,\ 1 - i$ is therefore linearly dependent.
Explicitly, taking the scalars $-i$ and 1, which are not both 0, we have

$$(-i)(1 + i) + (-1)(1 - i) = (1 - i) - (1 - i) = 0,$$

a nontrivial linear combination equal to 0, so the list is linearly dependent by 2.17.
The contrast is instructive: the scalar $-i$ that witnesses the dependence in (b) is not available in (a), and indeed $1 + i,\ 1 - i$ is a spanning list (in fact a basis) of the two-dimensional real vector space $\mathbf{C}$, whereas over $\mathbf{C}$ the space $\mathbf{C}$ is one-dimensional, so by 2.22 no list of length two in it can be linearly independent.

Problem 2A.8. Suppose v_1, v_2, v_3, v_4 is linearly independent in V. Prove that the list

$$v_1 - v_2, \quad v_2 - v_3, \quad v_3 - v_4, \quad v_4$$

is also linearly independent.

Solution. Suppose $a_1, a_2, a_3, a_4 \in \mathbb{F}$ satisfy

$$a_1(v_1 - v_2) + a_2(v_2 - v_3) + a_3(v_3 - v_4) + a_4 v_4 = 0.$$

Expanding and collecting the coefficient of each v_j gives

$$a_1v_1 + (a_2 - a_1)v_2 + (a_3 - a_2)v_3 + (a_4 - a_3)v_4 = 0.$$

Because v_1, v_2, v_3, v_4 is linearly independent, every coefficient in the equation above must equal 0. Hence

$$a_1 = 0, \qquad a_2 - a_1 = 0, \qquad a_3 - a_2 = 0, \qquad a_4 - a_3 = 0.$$

Reading these in order, $a_1 = 0$, then $a_2 = a_1 = 0$, then $a_3 = a_2 = 0$, then $a_4 = a_3 = 0$. Thus the only choice of a_1, a_2, a_3, a_4 making the original linear combination equal 0 is $a_1 = a_2 = a_3 = a_4 = 0$, which is exactly the statement that

$$v_1 - v_2, \quad v_2 - v_3, \quad v_3 - v_4, \quad v_4$$

is linearly independent (see the definition of linear independence, 2.15).

Problem 2A.9. Prove or give a counterexample: If $v_1, v_2, \ldots, v_m$ is a linearly independent list of vectors in V, then

$$5v_1 - 4v_2, \quad v_2, \quad v_3, \quad \ldots, \quad v_m$$

is linearly independent.

Solution. The statement is true; here is a proof. (Implicitly $m \geq 2$, so that the list makes sense.)
Suppose $a_1, a_2, \ldots, a_m \in \mathbb{F}$ satisfy

$$a_1(5v_1 - 4v_2) + a_2v_2 + a_3v_3 + \cdots + a_mv_m = 0.$$

Collecting the coefficient of each v_j turns this into

$$(5a_1)v_1 + (a_2 - 4a_1)v_2 + a_3v_3 + \cdots + a_mv_m = 0.$$

Because $v_1, \ldots, v_m$ is linearly independent, all of these coefficients equal 0 (2.15). Thus

$$5a_1 = 0, \qquad a_2 - 4a_1 = 0, \qquad a_3 = \cdots = a_m = 0.$$

Here $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, so $5 \neq 0$ in $\mathbb{F}$ and hence $5a_1 = 0$ forces $a_1 = 0$. Then $a_2 = 4a_1 = 0$. Therefore $a_1 = a_2 = \cdots = a_m = 0$, and the list

$$5v_1 - 4v_2, \quad v_2, \quad v_3, \quad \ldots, \quad v_m$$

is linearly independent.
Remark. Nothing is special about the numbers 5 and -4 beyond $5 \neq 0$: the same computation shows that if $b, c \in \mathbb{F}$ with $b \neq 0$, then $bv_1 + cv_2, v_2, v_3, \ldots, v_m$ is linearly independent. If instead the first coefficient were 0, the claim would fail, since the list $-4v_2, v_2, v_3, \ldots, v_m$ is linearly dependent.

Problem 2A.10. Prove or give a counterexample: If $v_1, v_2, \ldots, v_m$ is a linearly independent list of vectors in V and $\lambda \in \mathbb{F}$ with $\lambda \neq 0$, then $\lambda v_1, \lambda v_2, \ldots, \lambda v_m$ is linearly independent.

Solution. The statement is true; here is a proof.
Suppose $a_1, \ldots, a_m \in \mathbb{F}$ satisfy

$$a_1(\lambda v_1) + a_2(\lambda v_2) + \cdots + a_m(\lambda v_m) = 0.$$

By the commutativity and associativity of scalar multiplication, the left side equals

$$(a_1\lambda)v_1 + (a_2\lambda)v_2 + \cdots + (a_m\lambda)v_m.$$

Thus $(a_1\lambda)v_1 + \cdots + (a_m\lambda)v_m = 0$. Because $v_1, \ldots, v_m$ is linearly independent, the definition 2.15 forces

$$a_1\lambda = a_2\lambda = \cdots = a_m\lambda = 0.$$

Since $\lambda \neq 0$, it has a multiplicative inverse in $\mathbb{F}$; multiplying each equation by λ^{-1} gives $a_k = 0$ for every $k \in \{1, \dots, m\}$.
Hence the only scalars $a_1, \dots, a_m$ with $a_1(\lambda v_1) + \cdots + a_m(\lambda v_m) = 0$ are $a_1 = \cdots = a_m = 0$, so $\lambda v_1, \dots, \lambda v_m$ is linearly independent.
Remark. The hypothesis $\lambda \neq 0$ cannot be dropped when $m \geq 1$: if $\lambda = 0$, then $\lambda v_1, \dots, \lambda v_m$ is the list $0, \dots, 0$, which is linearly dependent because every list containing the vector 0 is linearly dependent (2.18).

Problem 2A.11. Prove or give a counterexample: If $v_1, \dots, v_m$ and $w_1, \dots, w_m$ are linearly independent lists of vectors in V, then the list $v_1 + w_1, \dots, v_m + w_m$ is linearly independent.

Solution. The statement is false. Here is a counterexample.
Take $V = \mathbb{R}^2$ and $m = 2$, with

$$v_1 = (1, 0), \quad v_2 = (0, 1), \qquad w_1 = (-1, 0), \quad w_2 = (0, -1).$$

The list v_1, v_2 is linearly independent: if $a_1(1, 0) + a_2(0, 1) = (0, 0)$, then $(a_1, a_2) = (0, 0)$, so $a_1 = a_2 = 0$. The list w_1, w_2 is linearly independent by the same computation, since $b_1(-1, 0) + b_2(0, -1) = (-b_1, -b_2)$, which equals $(0, 0)$ only when $b_1 = b_2 = 0$.
However

$$v_1 + w_1 = (0, 0), \qquad v_2 + w_2 = (0, 0),$$

so the list $v_1 + w_1, v_2 + w_2$ is the list $(0, 0), (0, 0)$. This list is linearly dependent, because

$$1 \cdot (0, 0) + 0 \cdot (0, 0) = (0, 0)$$

exhibits a linear combination equal to 0 with scalars not all 0 (indeed, every list containing the vector 0 is linearly dependent, by 2.18).
Thus linear independence of $v_1, \dots, v_m$ and of $w_1, \dots, w_m$ does not imply linear independence of $v_1 + w_1, \dots, v_m + w_m$.
Remark. The same idea works whenever V actually contains a linearly independent list $v_1, \dots, v_m$ of the given length $m \geq 1$ (for instance in $\mathbb{F}^m$, or in any V with $\dim V \geq m$): set $w_k = -v_k$. Then $w_1, \dots, w_m$ is linearly independent (by Exercise 2A.10 with $\lambda = -1$), while each $v_k + w_k = 0$, so the summed list is linearly dependent. In particular no hypothesis on V or m can rescue the statement, apart from the degenerate cases $m = 0$ and $V = \{0\}$, where there is no linearly independent list of length $m \geq 1$ to begin with.

Problem 2A.12. Suppose $v_1, \dots, v_m$ is linearly independent in V and $w \in V$. Prove that if $v_1 + w, \dots, v_m + w$ is linearly dependent, then $w \in \operatorname{span}(v_1, \dots, v_m)$.

Solution. Suppose $v_1 + w, \dots, v_m + w$ is linearly dependent. By the definition of linear dependence (2.17), there exist $a_1, \dots, a_m \in \mathbb{F}$, not all 0, such that

$$a_1(v_1 + w) + a_2(v_2 + w) + \cdots + a_m(v_m + w) = 0.$$

Expanding and separating the terms involving w gives

$$a_1 v_1 + \cdots + a_m v_m + (a_1 + \cdots + a_m)w = 0.$$

Put $s = a_1 + \cdots + a_m$, so the equation above reads

$$a_1 v_1 + \cdots + a_m v_m + sw = 0.$$

First suppose $s = 0$. Then the equation becomes $a_1 v_1 + \cdots + a_m v_m = 0$ with $a_1, \dots, a_m$ not all 0, contradicting the linear independence of $v_1, \dots, v_m$ (2.15). Hence $s \neq 0$.

Because $s \neq 0$, we may divide by s: solving the displayed equation for w gives

$$w = -\frac{a_1}{s}v_1 - \frac{a_2}{s}v_2 - \cdots - \frac{a_m}{s}v_m.$$

The right side is a linear combination of $v_1, \ldots, v_m$ with coefficients in $\mathbb{F}$. Therefore $w \in \operatorname{span}(v_1, \ldots, v_m)$, as desired.

Problem 2A.13. Suppose $v_1, \ldots, v_m$ is linearly independent in V and $w \in V$. Show that

$$v_1, \ldots, v_m, w \text{ is linearly independent} \iff w \notin \operatorname{span}(v_1, \ldots, v_m).$$

Solution. We prove the two implications separately.
First suppose $v_1, \ldots, v_m, w$ is linearly independent; we show $w \notin \operatorname{span}(v_1, \ldots, v_m)$. Suppose instead that $w \in \operatorname{span}(v_1, \ldots, v_m)$. Then there exist $c_1, \ldots, c_m \in \mathbb{F}$ with

$$w = c_1v_1 + \cdots + c_mv_m.$$

Moving w to the other side gives

$$c_1v_1 + \cdots + c_mv_m + (-1)w = 0.$$

This is a linear combination of $v_1, \ldots, v_m, w$ equal to 0 in which the last scalar is $-1 \neq 0$, so the scalars are not all 0. Hence $v_1, \ldots, v_m, w$ is linearly dependent (2.17), contradicting our assumption. Therefore $w \notin \operatorname{span}(v_1, \ldots, v_m)$.
Conversely, suppose $w \notin \operatorname{span}(v_1, \ldots, v_m)$; we show $v_1, \ldots, v_m, w$ is linearly independent. Let $a_1, \ldots, a_m, b \in \mathbb{F}$ satisfy

$$a_1v_1 + \cdots + a_mv_m + bw = 0.$$

If $b \neq 0$, then b is invertible in $\mathbb{F}$ and we may solve for w:

$$w = -\frac{a_1}{b}v_1 - \cdots - \frac{a_m}{b}v_m,$$

which exhibits w as a linear combination of $v_1, \ldots, v_m$ and hence puts w in $\operatorname{span}(v_1, \ldots, v_m)$, contrary to hypothesis. Thus $b = 0$, and the displayed equation becomes

$$a_1v_1 + \cdots + a_mv_m = 0.$$

Because $v_1, \ldots, v_m$ is linearly independent, this forces $a_1 = \cdots = a_m = 0$ (2.15). Hence $a_1 = \cdots = a_m = b = 0$, which is exactly what is required for $v_1, \ldots, v_m, w$ to be linearly independent.
Combining the two implications gives the stated equivalence.
Remark. The case $m = 0$ is included: then $v_1, \ldots, v_m$ is the empty list, $\operatorname{span}(\,) = \{0\}$, and the equivalence says that the one-term list w is linearly independent if and only if $w \neq 0$, which agrees with 2.16(c).

Problem 2A.14. Suppose $v_1, \ldots, v_m$ is a list of vectors in V. For $k \in \{1, \ldots, m\}$, let

$$w_k = v_1 + \cdots + v_k.$$

Show that the list $v_1, \ldots, v_m$ is linearly independent if and only if the list $w_1, \ldots, w_m$ is linearly independent.

Solution. Note first the two ways of passing between the lists:

$$w_k = v_1 + \cdots + v_k \quad (1 \leq k \leq m), \qquad v_1 = w_1, \qquad v_k = w_k - w_{k-1} \quad (2 \leq k \leq m).$$

The second set of formulas follows immediately from the first, since $w_k - w_{k-1} = (v_1 + \cdots + v_k) - (v_1 +$

$\cdots + v_{k-1}) = v_k$.
Suppose first that $v_1, \ldots, v_m$ is linearly independent, and let $a_1, \ldots, a_m \in \mathbb{F}$ satisfy

$$a_1 w_1 + a_2 w_2 + \cdots + a_m w_m = 0.$$

Substituting $w_k = v_1 + \cdots + v_k$ and collecting the terms involving each v_j, we see that v_j occurs in w_k exactly when $k \geq j$, so the left side equals

$$\sum_{j=1}^{m} \Big(\sum_{k=j}^{m} a_k \Big) v_j.$$

Since $v_1, \ldots, v_m$ is linearly independent, every coefficient vanishes (2.15):

$$\sum_{k=j}^{m} a_k = 0 \qquad \text{for each } j \in \{1, \ldots, m\}.$$

Taking $j = m$ gives $a_m = 0$. For $j < m$, subtracting the equation for $j+1$ from the equation for j gives

$$a_j = \sum_{k=j}^{m} a_k - \sum_{k=j+1}^{m} a_k = 0 - 0 = 0.$$

Hence $a_1 = \cdots = a_m = 0$, so $w_1, \ldots, w_m$ is linearly independent.
Conversely, suppose $w_1, \ldots, w_m$ is linearly independent, and let $b_1, \ldots, b_m \in \mathbb{F}$ satisfy

$$b_1 v_1 + b_2 v_2 + \cdots + b_m v_m = 0.$$

Substituting $v_1 = w_1$ and $v_k = w_k - w_{k-1}$ for $k \geq 2$, the left side equals

$$\sum_{k=1}^{m} b_k w_k - \sum_{k=2}^{m} b_k w_{k-1} = \sum_{k=1}^{m} b_k w_k - \sum_{k=1}^{m-1} b_{k+1} w_k = \sum_{k=1}^{m-1} (b_k - b_{k+1}) w_k + b_m w_m.$$

Since $w_1, \ldots, w_m$ is linearly independent, all these coefficients vanish:

$$b_k - b_{k+1} = 0 \quad \text{for } k \in \{1, \ldots, m-1\}, \qquad b_m = 0.$$

The first family of equations says $b_1 = b_2 = \cdots = b_m$, and the last says $b_m = 0$; hence $b_1 = \cdots = b_m = 0$. Thus $v_1, \ldots, v_m$ is linearly independent.
This proves both implications, so $v_1, \ldots, v_m$ is linearly independent if and only if $w_1, \ldots, w_m$ is.
Remark. Exercise 2A.3 shows that these two lists always have the same span. The present exercise is the companion statement for linear independence; together they say that passing from $v_1, \ldots, v_m$ to its list of partial sums preserves both properties, and hence carries bases to bases.

Problem 2A.15. Explain why there does not exist a list of six polynomials that is linearly independent in $\mathcal{P}_4(\mathbf{F})$.

Solution. Recall from the discussion following 2.12 that

$$\mathcal{P}_4(\mathbf{F}) = \operatorname{span}(1, z, z^2, z^3, z^4),$$

where z^k denotes the function $z \mapsto z^k$ from $\mathbf{F}$ to $\mathbf{F}$. Indeed, if $p \in \mathcal{P}_4(\mathbf{F})$, then $\deg p \leq 4$ by 2.12, so by 2.10 and 2.11 there exist $a_0, a_1, a_2, a_3, a_4 \in \mathbf{F}$ with $p(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4$ for every $z \in \mathbf{F}$ (take all the coefficients to be 0 if $p = 0$, whose degree is $-\infty$; otherwise take the coefficients supplied by 2.11 and pad with zeros). This exhibits p as a linear combination of $1, z, z^2, z^3, z^4$; conversely every such linear combination is a polynomial of degree at most 4.
Thus $1, z, z^2, z^3, z^4$ is a list of length 5 that spans $\mathcal{P}_4(\mathbf{F})$. In particular $\mathcal{P}_4(\mathbf{F})$ is finite-dimensional

(2.9).
Now 2.22 states that in a finite-dimensional vector space the length of every linearly independent list is less than or equal to the length of every spanning list. Hence every linearly independent list in $\mathcal{P}_4(\mathbf{F})$ has length at most 5.
A list of six polynomials that is linearly independent in $\mathcal{P}_4(\mathbf{F})$ would have length $6 > 5$, which is impossible. Therefore no such list exists.

Problem 2A.16. Explain why no list of four polynomials spans $\mathcal{P}_4(\mathbf{F})$.

Solution. Consider the list $1, z, z^2, z^3, z^4$ of elements of $\mathcal{P}_4(\mathbf{F})$, where z^k denotes the function $z \mapsto z^k$. This list has length 5, and it is linearly independent by 2.16(b): if $a_0, a_1, a_2, a_3, a_4 \in \mathbf{F}$ satisfy

$$a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 = 0$$

as an element of $\mathcal{P}(\mathbf{F})$, meaning that $a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 = 0$ for every $z \in \mathbf{F}$, then $a_0 = a_1 = a_2 = a_3 = a_4 = 0$ (this is the uniqueness of the coefficients of a polynomial, which follows from 4.8). Although 2.16(b) is stated for $\mathcal{P}(\mathbf{F})$, the displayed condition defining linear independence involves only the vectors in the list and the scalar field $\mathbf{F}$, not the ambient vector space; since $1, z, z^2, z^3, z^4$ all lie in the subspace $\mathcal{P}_4(\mathbf{F})$, this list is linearly independent in $\mathcal{P}_4(\mathbf{F})$ as well.
Suppose now that some list q_1, q_2, q_3, q_4 of four polynomials spanned $\mathcal{P}_4(\mathbf{F})$. Then $\mathcal{P}_4(\mathbf{F})$ would be a finite-dimensional vector space (2.9) with a spanning list of length 4. By 2.22, the length of every linearly independent list in $\mathcal{P}_4(\mathbf{F})$ would then be at most 4. But $1, z, z^2, z^3, z^4$ is a linearly independent list in $\mathcal{P}_4(\mathbf{F})$ of length $5 > 4$, a contradiction.
Hence no list of four polynomials spans $\mathcal{P}_4(\mathbf{F})$.

Problem 2A.17. Prove that V is infinite-dimensional if and only if there is a sequence $v_1, v_2, \ldots$ of vectors in V such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m.

Solution. We prove the two implications separately.
First suppose there is a sequence $v_1, v_2, \ldots$ of vectors in V such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m. We must show V is infinite-dimensional, that is (2.13), that V is not finite-dimensional. Suppose instead that V is finite-dimensional. Then by 2.9 there is a list $w_1, \ldots, w_n$ of vectors in V that spans V. Take $m = n+1$. By hypothesis the list $v_1, \ldots, v_{n+1}$ is linearly independent in V, and it has length $n+1$. But 2.22 says that in a finite-dimensional vector space the length of every linearly independent list is at most the length of every spanning list, so $n+1 \le n$, which is false. This contradiction shows V is not finite-dimensional; that is, V is infinite-dimensional.
Conversely, suppose V is infinite-dimensional. We construct the required sequence recursively, choosing v_m at step m.
Step 1. The list of length 0, namely the empty list, has span $\{0\}$ (by the convention that $\operatorname{span}() = \{0\}$). If $V = \{0\}$, then the empty list spans V and V would be finite-dimensional, contrary to hypothesis. Hence $V \neq \{0\}$, so we may choose $v_1 \in V$ with $v_1 \neq 0$. By 2.16(c), the list v_1 is linearly independent.
Step $m+1$. Suppose $m \ge 1$ and vectors $v_1, \ldots, v_m \in V$ have been chosen so that $v_1, \ldots, v_m$ is linearly independent. Because V is infinite-dimensional, no list of vectors in V spans V; in particular $\operatorname{span}(v_1, \ldots, v_m) \neq V$. Hence we may choose

$$v_{m+1} \in V \setminus \operatorname{span}(v_1, \ldots, v_m).$$

By Exercise 13 of Section 2A (a linearly independent list stays linearly independent when a vector outside its span is adjoined), the list $v_1, \ldots, v_m, v_{m+1}$ is linearly independent.
For completeness, here is the verification of that step. Suppose $a_1, \ldots, a_{m+1} \in \mathbf{F}$ satisfy $a_1 v_1 + \cdots + a_m v_m + a_{m+1} v_{m+1} = 0$. If $a_{m+1} \neq 0$, then

$$v_{m+1} = -\frac{a_1}{a_{m+1}} v_1 - \cdots - \frac{a_m}{a_{m+1}} v_m \in \operatorname{span}(v_1, \ldots, v_m),$$

contradicting the choice of v_{m+1}. Hence $a_{m+1} = 0$, and then $a_1v_1 + \cdots + a_mv_m = 0$, which forces $a_1 = \cdots = a_m = 0$ because $v_1, \ldots, v_m$ is linearly independent. Thus all the scalars are 0, proving linear independence.
This recursion never halts, so it produces a sequence $v_1, v_2, \ldots$ of vectors in V. By construction $v_1, \ldots, v_m$ is linearly independent for every positive integer m, as desired.

Problem 2A.18. Prove that $\mathbf{F}^\infty$ is infinite-dimensional.

Solution. Recall from 1.23 that $\mathbf{F}^\infty$ is the vector space of all sequences of elements of $\mathbf{F}$,

$$\mathbf{F}^\infty = \{(x_1, x_2, \ldots) : x_k \in \mathbf{F} \text{ for } k = 1, 2, \ldots\},$$

with addition and scalar multiplication defined coordinatewise.
For each positive integer k let $e_k \in \mathbf{F}^\infty$ be the sequence whose k-th coordinate is 1 and whose other coordinates are all 0:

$$e_k = (0, \ldots, 0, 1, 0, 0, \ldots),$$

with the 1 in position k.
Claim: for every positive integer m, the list $e_1, \ldots, e_m$ is linearly independent in $\mathbf{F}^\infty$. To see this, suppose $a_1, \ldots, a_m \in \mathbf{F}$ and

$$a_1e_1 + \cdots + a_me_m = 0,$$

where 0 here denotes the sequence all of whose coordinates are 0. Computing the left side coordinatewise gives

$$a_1e_1 + \cdots + a_me_m = (a_1, a_2, \ldots, a_m, 0, 0, \ldots).$$

Two sequences are equal exactly when all their corresponding coordinates are equal, so $a_1 = a_2 = \cdots = a_m = 0$. Thus $e_1, \ldots, e_m$ is linearly independent, proving the claim.
Hence $e_1, e_2, \ldots$ is a sequence of vectors in $\mathbf{F}^\infty$ such that $e_1, \ldots, e_m$ is linearly independent for every positive integer m. By Exercise 17 of Section 2A, $\mathbf{F}^\infty$ is infinite-dimensional.
Alternatively, one can argue directly without invoking Exercise 17. Suppose $\mathbf{F}^\infty$ were finite-dimensional. Then by 2.9 some list $w_1, \ldots, w_n$ spans $\mathbf{F}^\infty$, and by 2.22 every linearly independent list in $\mathbf{F}^\infty$ has length at most n. But the claim above provides the linearly independent list $e_1, \ldots, e_{n+1}$, of length $n + 1 > n$. This contradiction shows that $\mathbf{F}^\infty$ is not finite-dimensional, that is, $\mathbf{F}^\infty$ is infinite-dimensional (2.13).

Problem 2A.19. Prove that the real vector space of all continuous real-valued functions on the interval $[0, 1]$ is infinite-dimensional.

Solution. Write $C[0, 1]$ for the set of continuous functions from $[0, 1]$ to $\mathbf{R}$. This is a subspace of the vector space $\mathbf{R}^{[0,1]}$ of all functions from $[0, 1]$ to $\mathbf{R}$ (see 1.24 and 1.34): the zero function is continuous, and sums and real scalar multiples of continuous functions are continuous. So $C[0, 1]$ is a real vector space, with the usual pointwise operations.
For each nonnegative integer k define $f_k \in C[0, 1]$ by

$$f_k(x) = x^k \qquad \text{for } x \in [0, 1],$$

with the convention $f_0(x) = 1$. Each f_k is continuous on $[0, 1]$.
Claim: for every nonnegative integer m, the list $f_0, f_1, \ldots, f_m$ is linearly independent in $C[0, 1]$. To prove this, suppose $a_0, a_1, \ldots, a_m \in \mathbf{R}$ and

$$a_0f_0 + a_1f_1 + \cdots + a_mf_m = 0$$

in $C[0, 1]$, which means

$$a_0 + a_1x + \cdots + a_mx^m = 0 \qquad \text{for every } x \in [0, 1].$$

Let $p \in \mathcal{P}(\mathbf{R})$ be the polynomial $p(z) = a_0 + a_1 z + \cdots + a_m z^m$ for $z \in \mathbf{R}$. Then every point of $[0,1]$ is a zero of p, so p has infinitely many zeros in $\mathbf{R}$. If p were not the zero polynomial, then p would have some degree d with $0 \le d \le m$, and p would have at most d zeros in $\mathbf{R}$: for $d \ge 1$ this is 4.8, and for $d = 0$ the polynomial p is a nonzero constant and so has no zeros at all. Either way p would have only finitely many zeros, a contradiction. Hence p is the zero polynomial, and therefore, by the uniqueness of the coefficients of a polynomial (see the discussion preceding 2.11, which rests on 4.8), we get

$$a_0 = a_1 = \cdots = a_m = 0.$$

This proves the claim.
Note carefully that the claim really is a statement about $C[0,1]$, not about $\mathcal{P}(\mathbf{R})$: two polynomials that agree at every point of $[0,1]$ determine the same element of $C[0,1]$, and the argument above shows that this forces their coefficients to agree.
Now set $g_k = f_{k-1}$ for each positive integer k, so that $g_1, g_2, \ldots$ is a sequence in $C[0,1]$ with $g_1, \ldots, g_m = f_0, \ldots, f_{m-1}$ linearly independent for every positive integer m, by the claim. By Exercise 17 of Section 2A, $C[0,1]$ is infinite-dimensional.
Directly: if $C[0,1]$ were finite-dimensional, then by 2.9 some list $w_1, \ldots, w_n$ of continuous functions would span $C[0,1]$, and by 2.22 every linearly independent list in $C[0,1]$ would have length at most n. But $f_0, f_1, \ldots, f_n$ is linearly independent of length $n+1 > n$, a contradiction. Hence $C[0,1]$ is infinite-dimensional (2.13).

Problem 2A.20. Suppose $p_0, p_1, \ldots, p_m$ are polynomials in $\mathcal{P}_m(\mathbf{F})$ such that $p_k(2) = 0$ for each $k \in \{0, \ldots, m\}$. Prove that $p_0, p_1, \ldots, p_m$ is not linearly independent in $\mathcal{P}_m(\mathbf{F})$.

Solution. Let

$$U = \{p \in \mathcal{P}_m(\mathbf{F}) : p(2) = 0\}.$$

By hypothesis $p_0, p_1, \ldots, p_m$ is a list of $m+1$ vectors in U.
First, U is a subspace of $\mathcal{P}_m(\mathbf{F})$. Indeed the zero polynomial lies in U; if $p, q \in U$ then $(p+q)(2) = p(2) + q(2) = 0 + 0 = 0$, so $p + q \in U$; and if $p \in U$ and $\lambda \in \mathbf{F}$ then $(\lambda p)(2) = \lambda p(2) = 0$, so $\lambda p \in U$. Thus U is a subspace by 1.34, and in particular U is itself a vector space over $\mathbf{F}$.
Next we exhibit a spanning list for U of length m. For $j \in \{1, \ldots, m\}$ define $q_j \in \mathcal{P}_m(\mathbf{F})$ by

$$q_j(z) = z^j - 2^j \qquad \text{for } z \in \mathbf{F}.$$

Each q_j has degree $j \le m$ and satisfies $q_j(2) = 2^j - 2^j = 0$, so $q_j \in U$. (When $m = 0$ this is the empty list.)
Claim: $U = \operatorname{span}(q_1, \ldots, q_m)$.
Since each $q_j \in U$ and U is a subspace, every linear combination of $q_1, \ldots, q_m$ lies in U; hence $\operatorname{span}(q_1, \ldots, q_m) \subseteq U$. (For $m = 0$ this reads $\operatorname{span}() = \{0\} \subseteq U$, which also holds.)
Conversely, let $p \in U$. By 2.11 and 2.12 there exist $a_0, a_1, \ldots, a_m \in \mathbf{F}$ with

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m \qquad \text{for every } z \in \mathbf{F}.$$

Because $p(2) = 0$, for every $z \in \mathbf{F}$ we have

$$\begin{aligned} p(z) &= p(z) - p(2) \\ &= (a_0 - a_0) + a_1(z-2) + a_2(z^2 - 2^2) + \cdots + a_m(z^m - 2^m) \\ &= a_1 q_1(z) + a_2 q_2(z) + \cdots + a_m q_m(z). \end{aligned}$$

Thus $p = a_1 q_1 + \cdots + a_m q_m \in \operatorname{span}(q_1, \ldots, q_m)$. (This is the same manipulation recorded in 4.7, inside the proof of 4.6.) When $m = 0$ the computation degenerates to $p = p - p(2) = 0$, so $U = \{0\} = \operatorname{span}()$, again as claimed.
This proves the claim, so U is spanned by a list of length m; in particular U is finite-dimensional (2.9).
Now suppose, for contradiction, that $p_0, p_1, \ldots, p_m$ were linearly independent in $\mathcal{P}_m(\mathbf{F})$. Linear independence of a list depends only on the vectors themselves and the scalar field, not on which vector

space we regard them as sitting in; since $p_0, \ldots, p_m$ all lie in the subspace U, the list $p_0, \ldots, p_m$ would then be a linearly independent list in the vector space U. It has length $m+1$, while $q_1, \ldots, q_m$ spans U and has length m. This contradicts 2.22, which asserts that in the finite-dimensional vector space U the length of every linearly independent list is at most the length of every spanning list, i.e. $m+1 \leq m$. Therefore $p_0, p_1, \ldots, p_m$ is not linearly independent in $\mathcal{P}_m(\mathbf{F})$.

2.2 Exercises 2B

Problem 2B.1. Find all vector spaces that have exactly one basis.

Solution. Claim: the only vector space with exactly one basis is the zero vector space $\{0\}$ (whose unique basis is the empty list).
First, $\{0\}$ does have exactly one basis. The empty list is linearly independent, since 2.15 explicitly declares the empty list to be linearly independent, and its span is $\{0\}$ by the convention in 2.4 that the span of the empty list is $\{0\}$; hence the empty list is a basis of $\{0\}$ by 2.26. Any nonempty list of vectors in $\{0\}$ has the form $0, \ldots, 0$, which is not linearly independent (take the first coefficient to be 1 and the rest 0), so no other list is a basis. Thus $\{0\}$ has exactly one basis.
Conversely, suppose V is a vector space with exactly one basis; call it $v_1, \ldots, v_n$. We must show $V = \{0\}$, i.e. that $n = 0$. Suppose instead $n \geq 1$. Since $v_1, \ldots, v_n$ is linearly independent, $v_1 \neq 0$: otherwise $1 \cdot v_1 + 0v_2 + \cdots + 0v_n = 0$ would be a nontrivial vanishing linear combination.
Consider the list

$$2v_1,\ v_2, \ldots, v_n.$$

It spans V, because $v_1 = \frac{1}{2}(2v_1) \in \operatorname{span}(2v_1, v_2, \ldots, v_n)$ and hence this span contains $\operatorname{span}(v_1, \ldots, v_n) = V$; note $\frac{1}{2} \in \mathbf{F}$ since $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$. It is linearly independent: if $a_1(2v_1) + a_2v_2 + \cdots + a_nv_n = 0$, then $(2a_1)v_1 + a_2v_2 + \cdots + a_nv_n = 0$, so linear independence of $v_1, \ldots, v_n$ gives $2a_1 = a_2 = \cdots = a_n = 0$, hence $a_1 = \cdots = a_n = 0$. So $2v_1, v_2, \ldots, v_n$ is a basis of V by 2.26.
But this basis differs from $v_1, \ldots, v_n$, since $2v_1 - v_1 = v_1 \neq 0$ and so $2v_1 \neq v_1$. This contradicts the assumption that V has exactly one basis. Hence $n = 0$, and therefore $V = \operatorname{span}() = \{0\}$.
(Note that a vector space with exactly one basis in particular has a basis, which in this book is by definition a finite list; so such a space is automatically finite-dimensional and the argument above applies to it.)

Problem 2B.2. Verify all assertions in Example 2.27.

Solution. Throughout, a list is a basis exactly when it is linearly independent and spans (2.26).
(a) The standard basis of $\mathbf{F}^n$ is $e_1, \ldots, e_n$, where e_k is the n-tuple with 1 in slot k and 0 elsewhere. For $a_1, \ldots, a_n \in \mathbf{F}$ we have $a_1e_1 + \cdots + a_ne_n = (a_1, \ldots, a_n)$. Hence every $(x_1, \ldots, x_n) \in \mathbf{F}^n$ equals $x_1e_1 + \cdots + x_ne_n$, so the list spans $\mathbf{F}^n$; and if $a_1e_1 + \cdots + a_ne_n = (0, \ldots, 0)$ then $(a_1, \ldots, a_n) = (0, \ldots, 0)$, so $a_1 = \cdots = a_n = 0$ and the list is linearly independent. Thus $e_1, \ldots, e_n$ is a basis of $\mathbf{F}^n$.
(b) For $(1,2), (3,5)$ in $\mathbf{F}^2$: given $(x, y) \in \mathbf{F}^2$, the equation $a(1,2) + b(3,5) = (x, y)$ says $a + 3b = x$ and $2a + 5b = y$. Doubling the first equation and subtracting the second gives $b = 2x - y$, and then $a = x - 3b = -5x + 3y$. Conversely $(-5x+3y)(1,2) + (2x-y)(3,5) = (-5x+3y+6x-3y,\ -10x+6y+10x-5y) = (x, y)$. So every (x, y) is uniquely of the form $a(1,2) + b(3,5)$, and $(1,2), (3,5)$ is a basis of $\mathbf{F}^2$ by 2.28. (In particular, taking $(x, y) = (0, 0)$ gives $a = b = 0$, which is the linear independence.) This list has length two, as does the standard basis of $\mathbf{F}^2$ from (a).
(c) For $(1, 2, -4), (7, -5, 6)$ in $\mathbf{F}^3$: if $a(1, 2, -4) + b(7, -5, 6) = (0, 0, 0)$, then $a + 7b = 0$ and $2a - 5b = 0$. The first gives $a = -7b$; substituting into the second gives $-14b - 5b = -19b = 0$, so $b = 0$ and then $a = 0$. Hence the list is linearly independent. It does not span $\mathbf{F}^3$: if $a(1, 2, -4) + b(7, -5, 6) = (0, 0, 1)$, the first two coordinates force $a = b = 0$ as above, and then the third coordinate is $0 \neq 1$. So $(0, 0, 1) \notin \operatorname{span}\big((1, 2, -4), (7, -5, 6)\big)$, and the list is not a basis of $\mathbf{F}^3$.
(d) For $(1,2), (3,5), (4,13)$ in $\mathbf{F}^2$: it spans $\mathbf{F}^2$ because already the sublist $(1,2), (3,5)$ spans $\mathbf{F}^2$ by (b), and enlarging a list only enlarges (weakly) its span. It is not linearly independent: applying the

formulas of (b) with $(x, y) = (4, 13)$ gives $b = 2 \cdot 4 - 13 = -5$ and $a = -5 \cdot 4 + 3 \cdot 13 = 19$, so $(4, 13) = 19(1, 2) - 5(3, 5)$, i.e.

$$19(1, 2) - 5(3, 5) - 1 \cdot (4, 13) = (0, 0)$$

with coefficients not all 0. Hence the list is not a basis.
(e) Let $U = \{(x, x, y) \in \mathbf{F}^3 : x, y \in \mathbf{F}\}$. Every linear combination $a(1, 1, 0) + b(0, 0, 1) = (a, a, b)$ has its first two coordinates equal, hence lies in U; thus $\operatorname{span}\big((1, 1, 0), (0, 0, 1)\big) \subseteq U$. Conversely, every element of U satisfies $(x, x, y) = x(1, 1, 0) + y(0, 0, 1)$, so $U \subseteq \operatorname{span}\big((1, 1, 0), (0, 0, 1)\big)$. Hence U equals that span; in particular U is a subspace of $\mathbf{F}^3$ by 2.6 (so it is a vector space and the assertion makes sense), and the list spans U. If $a(1, 1, 0) + b(0, 0, 1) = (a, a, b) = (0, 0, 0)$, then $a = b = 0$, so the list is linearly independent. Hence $(1, 1, 0), (0, 0, 1)$ is a basis of U.
(f) Let $U = \{(x, y, z) \in \mathbf{F}^3 : x + y + z = 0\}$. Both $(1, -1, 0)$ and $(1, 0, -1)$ lie in U, since $1 - 1 + 0 = 0$ and $1 + 0 - 1 = 0$. Every linear combination $a(1, -1, 0) + b(1, 0, -1) = (a + b, -a, -b)$ satisfies $(a + b) + (-a) + (-b) = 0$ and so lies in U; combined with the spanning computation below this shows $U = \operatorname{span}\big((1, -1, 0), (1, 0, -1)\big)$, so U is a subspace of $\mathbf{F}^3$ by 2.6 and the assertion makes sense. If $a(1, -1, 0) + b(1, 0, -1) = (a + b, -a, -b) = (0, 0, 0)$, then $a = b = 0$; so the list is linearly independent. For spanning, let $(x, y, z) \in U$, so $x = -y - z$. Then

$$(-y)(1, -1, 0) + (-z)(1, 0, -1) = (-y - z,\ y,\ z) = (x, y, z).$$

Hence $(1, -1, 0), (1, 0, -1)$ is a basis of U.
(g) By definition $\mathcal{P}_m(\mathbf{F})$ is the set of polynomial functions on $\mathbf{F}$ of the form $p(z) = a_0 + a_1 z + \cdots + a_m z^m$ with $a_0, \ldots, a_m \in \mathbf{F}$; that is exactly the statement that $1, z, \ldots, z^m$ spans $\mathcal{P}_m(\mathbf{F})$. This list is linearly independent by 2.16(b): if $a_0 + a_1 z + \cdots + a_m z^m = 0$ as an element of $\mathcal{P}(\mathbf{F})$, i.e. for every $z \in \mathbf{F}$, then $a_0 = a_1 = \cdots = a_m = 0$ (this follows from 4.8). Hence $1, z, \ldots, z^m$ is a basis of $\mathcal{P}_m(\mathbf{F})$.
Finally, the example asserts that $(7, 5), (-4, 9)$ and $(1, 2), (3, 5)$ are both bases of $\mathbf{F}^2$. The second was verified in (b). For the first, let $(x, y) \in \mathbf{F}^2$ and consider $a(7, 5) + b(-4, 9) = (x, y)$, i.e. $7a - 4b = x$ and $5a + 9b = y$. Multiplying the first by 9, the second by 4, and adding gives $83a = 9x + 4y$; multiplying the first by 5, the second by 7, and subtracting gives $-83b = 5x - 7y$. Since $83 \neq 0$ in $\mathbf{F}$, the unique solution is

$$a = \frac{9x + 4y}{83}, \qquad b = \frac{7y - 5x}{83},$$

and one checks directly that these values work: $7a - 4b = \frac{63x + 28y - 28y + 20x}{83} = x$ and $5a + 9b = \frac{45x + 20y + 63y - 45x}{83} = y$. So every vector of $\mathbf{F}^2$ has a unique representation as a linear combination of $(7, 5), (-4, 9)$, which by 2.28 makes that list a basis of $\mathbf{F}^2$.

Problem 2B.3. (a) Let U be the subspace of $\mathbf{R}^5$ defined by

$$U = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbf{R}^5 : x_1 = 3x_2 \text{ and } x_3 = 7x_4\}.$$

Find a basis of U.
(b) Extend the basis in (a) to a basis of $\mathbf{R}^5$.
(c) Find a subspace W of $\mathbf{R}^5$ such that $\mathbf{R}^5 = U \oplus W$.

Solution. (a) A vector $(x_1, x_2, x_3, x_4, x_5)$ lies in U precisely when $x_1 = 3x_2$ and $x_3 = 7x_4$; so x_2, x_4, x_5 may be arbitrary and then x_1, x_3 are determined. Writing $x_2 = t$, $x_4 = s$, $x_5 = r$, the general element of U is

$$(3t, t, 7s, s, r) = t(3, 1, 0, 0, 0) + s(0, 0, 7, 1, 0) + r(0, 0, 0, 0, 1).$$

Hence the list

$$(3, 1, 0, 0, 0), \quad (0, 0, 7, 1, 0), \quad (0, 0, 0, 0, 1)$$

spans U (and each of these three vectors indeed lies in U: $3 = 3 \cdot 1$ and $0 = 7 \cdot 0$ for the first, etc.). It

is linearly independent: if

$$a(3,1,0,0,0)+b(0,0,7,1,0)+c(0,0,0,0,1)=(3a,a,7b,b,c)=(0,0,0,0,0),$$

then reading off the second, fourth, and fifth coordinates gives $a=b=c=0$. So this list is a basis of U by 2.26.
(b) Adjoin $(1,0,0,0,0)$ and $(0,0,1,0,0)$, giving the list

$$(3,1,0,0,0),\ (0,0,7,1,0),\ (0,0,0,0,1),\ (1,0,0,0,0),\ (0,0,1,0,0).$$

This spans $\mathbf{R}^5$: given $(x_1,\dots,x_5)\in\mathbf{R}^5$, take $a=x_2$, $b=x_4$, $c=x_5$, $d=x_1-3x_2$, $e=x_3-7x_4$; then

$$a(3,1,0,0,0)+b(0,0,7,1,0)+c(0,0,0,0,1)+d(1,0,0,0,0)+e(0,0,1,0,0)$$

$$=(3x_2+x_1-3x_2,\ x_2,\ 7x_4+x_3-7x_4,\ x_4,\ x_5)=(x_1,x_2,x_3,x_4,x_5).$$

It is linearly independent: if the displayed combination equals $(0,0,0,0,0)$, then its coordinates are $(3a+d,\ a,\ 7b+e,\ b,\ c)=(0,0,0,0,0)$, so the second, fourth, and fifth coordinates give $a=b=c=0$, and then the first and third give $d=0$ and $e=0$. Hence this list of length five is a basis of $\mathbf{R}^5$ extending the basis from (a).
(c) Take

$$W=\operatorname{span}\big((1,0,0,0,0),(0,0,1,0,0)\big)=\{(x_1,0,x_3,0,0):x_1,x_3\in\mathbf{R}\},$$

which is a subspace of $\mathbf{R}^5$ because every span is a subspace (2.6).
$U+W=\mathbf{R}^5$: by (b) every vector of $\mathbf{R}^5$ is a linear combination of the five listed vectors; the part coming from the first three vectors lies in U (they span U) and the part coming from the last two lies in W.
$U\cap W=\{0\}$: a vector of W has the form $(x_1,0,x_3,0,0)$, and if it also lies in U then $x_1=3\cdot 0=0$ and $x_3=7\cdot 0=0$, so the vector is 0.
By 1.46 these two facts give $\mathbf{R}^5=U\oplus W$.

Problem 2B.4. (a) Let U be the subspace of $\mathbf{C}^5$ defined by

$$U=\{(z_1,z_2,z_3,z_4,z_5)\in\mathbf{C}^5:6z_1=z_2 \text{ and } z_3+2z_4+3z_5=0\}.$$

Find a basis of U.
(b) Extend the basis in (a) to a basis of $\mathbf{C}^5$.
(c) Find a subspace W of $\mathbf{C}^5$ such that $\mathbf{C}^5=U\oplus W$.

Solution. (a) A vector lies in U exactly when $z_2=6z_1$ and $z_3=-2z_4-3z_5$; thus z_1,z_4,z_5 are free and z_2,z_3 are determined by them. Writing $z_1=t$, $z_4=s$, $z_5=r$, the general element of U is

$$(t,6t,-2s-3r,s,r)=t(1,6,0,0,0)+s(0,0,-2,1,0)+r(0,0,-3,0,1).$$

So the list

$$(1,6,0,0,0),\quad (0,0,-2,1,0),\quad (0,0,-3,0,1)$$

spans U, and each of its vectors lies in U (for the first, $6\cdot 1=6$ and $0+0+0=0$; for the second, $6\cdot 0=0$ and $-2+2\cdot 1+0=0$; for the third, $6\cdot 0=0$ and $-3+0+3\cdot 1=0$). It is linearly independent: if

$$a(1,6,0,0,0)+b(0,0,-2,1,0)+c(0,0,-3,0,1)=(a,\ 6a,\ -2b-3c,\ b,\ c)=0,$$

the first, fourth, and fifth coordinates give $a=b=c=0$. Hence this list is a basis of U by 2.26.
(b) Adjoin $(0,1,0,0,0)$ and $(0,0,1,0,0)$, giving

$$(1,6,0,0,0),\ (0,0,-2,1,0),\ (0,0,-3,0,1),\ (0,1,0,0,0),\ (0,0,1,0,0).$$

For scalars $a,b,c,d,e\in\mathbf{C}$,

$$a(1,6,0,0,0)+b(0,0,-2,1,0)+c(0,0,-3,0,1)+d(0,1,0,0,0)+e(0,0,1,0,0)$$

$$= (a,\ 6a + d,\ -2b - 3c + e,\ b,\ c).$$

This list spans $\mathbf{C}^5$: given $(z_1, \ldots, z_5) \in \mathbf{C}^5$, choose $a = z_1$, $b = z_4$, $c = z_5$, $d = z_2 - 6z_1$, and $e = z_3 + 2z_4 + 3z_5$; then the displayed vector equals $(z_1, z_2, z_3, z_4, z_5)$. It is linearly independent: if the displayed vector is 0, then the first, fourth, and fifth coordinates give $a = b = c = 0$, and then the second and third coordinates give $d = 0$ and $e = 0$. Hence it is a basis of $\mathbf{C}^5$ extending the basis of (a).
(c) Take

$$W = \operatorname{span}\big((0,1,0,0,0), (0,0,1,0,0)\big) = \{(0, z_2, z_3, 0, 0) : z_2, z_3 \in \mathbf{C}\},$$

a subspace of $\mathbf{C}^5$ by 2.6.
$U+W = \mathbf{C}^5$: by (b) every vector of $\mathbf{C}^5$ is a linear combination of the five listed vectors; the contribution of the first three lies in U and the contribution of the last two lies in W.
$U \cap W = \{0\}$: if $(0, z_2, z_3, 0, 0) \in U$, then $z_2 = 6 \cdot 0 = 0$ and $z_3 + 2 \cdot 0 + 3 \cdot 0 = 0$ forces $z_3 = 0$; so the vector is 0.
By 1.46, $\mathbf{C}^5 = U \oplus W$.

Problem 2B.5. Suppose V is finite-dimensional and U, W are subspaces of V such that $V = U + W$. Prove that there exists a basis of V consisting of vectors in $U \cup W$.

Solution. Because V is finite-dimensional, so are its subspaces U and W (2.25). Hence U has a basis $u_1, \ldots, u_m$ and W has a basis $w_1, \ldots, w_n$ (2.31); here m or n may be 0, in which case the corresponding list is empty.
The combined list

$$u_1, \ldots, u_m, w_1, \ldots, w_n$$

spans V. Indeed, let $v \in V$. Since $V = U + W$, there are $u \in U$ and $w \in W$ with $v = u + w$. Since $u_1, \ldots, u_m$ spans U, there are scalars $a_1, \ldots, a_m$ with $u = a_1 u_1 + \cdots + a_m u_m$; similarly $w = b_1 w_1 + \cdots + b_n w_n$ for some scalars $b_1, \ldots, b_n$. Thus

$$v = a_1 u_1 + \cdots + a_m u_m + b_1 w_1 + \cdots + b_n w_n,$$

so $v \in \operatorname{span}(u_1, \ldots, u_m, w_1, \ldots, w_n)$.
By 2.30, every spanning list in a vector space can be reduced to a basis of that space by deleting some of its vectors. Applying this to the spanning list above produces a basis of V each of whose vectors is one of $u_1, \ldots, u_m, w_1, \ldots, w_n$. Every u_j lies in U and every w_k lies in W, so every vector of this basis lies in $U \cup W$. This is the desired basis of V.

Problem 2B.6. Prove or give a counterexample: If p_0, p_1, p_2, p_3 is a list in $\mathcal{P}_3(\mathbf{F})$ such that none of the polynomials p_0, p_1, p_2, p_3 has degree 2, then p_0, p_1, p_2, p_3 is not a basis of $\mathcal{P}_3(\mathbf{F})$.

Solution. The assertion is false. Here is a counterexample. Define

$$p_0(z) = 1, \qquad p_1(z) = z, \qquad p_2(z) = z^2 + z^3, \qquad p_3(z) = z^3.$$

These all lie in $\mathcal{P}_3(\mathbf{F})$, and their degrees are $0, 1, 3, 3$ respectively, so none of them has degree 2. Nevertheless p_0, p_1, p_2, p_3 is a basis of $\mathcal{P}_3(\mathbf{F})$.
Spanning: since $p_2 - p_3$ is the polynomial $z \mapsto z^2$, the span of p_0, p_1, p_2, p_3 contains the four polynomials $1, z, z^2, z^3$, hence contains their span, which is all of $\mathcal{P}_3(\mathbf{F})$ (Example 2.27(g)). Concretely, for $a, b, c, d \in \mathbf{F}$,

$$a + bz + cz^2 + dz^3 = a\,p_0 + b\,p_1 + c\,p_2 + (d - c)\,p_3.$$

Linear independence: suppose $a, b, c, d \in \mathbf{F}$ and $a\,p_0 + b\,p_1 + c\,p_2 + d\,p_3 = 0$. Collecting terms, this says

$$a + bz + cz^2 + (c + d)z^3 = 0$$

for all $z \in \mathbf{F}$. By 2.16(b) (which rests on 4.8), all coefficients vanish: $a = b = c = 0$ and $c + d = 0$; the last two give $d = 0$. Hence $a = b = c = d = 0$.

So p_0, p_1, p_2, p_3 is linearly independent and spans $\mathcal{P}_3(\mathbf{F})$, hence is a basis of $\mathcal{P}_3(\mathbf{F})$ by 2.26, even though none of these polynomials has degree 2. The claim in the exercise is therefore false.

Problem 2B.7. Suppose v_1, v_2, v_3, v_4 is a basis of V. Prove that

$$v_1 + v_2,\ v_2 + v_3,\ v_3 + v_4,\ v_4$$

is also a basis of V.

Solution. Write L for the list $v_1 + v_2,\ v_2 + v_3,\ v_3 + v_4,\ v_4$. By 2.26 we must show that L is linearly independent and spans V.
Linear independence. Suppose $a, b, c, d \in \mathbf{F}$ satisfy

$$a(v_1 + v_2) + b(v_2 + v_3) + c(v_3 + v_4) + dv_4 = 0.$$

Regrouping by v_1, v_2, v_3, v_4 gives

$$av_1 + (a + b)v_2 + (b + c)v_3 + (c + d)v_4 = 0.$$

Since v_1, v_2, v_3, v_4 is a basis of V, it is linearly independent, so all four coefficients are 0:

$$a = 0, \qquad a + b = 0, \qquad b + c = 0, \qquad c + d = 0.$$

Reading these in order gives $a = 0$, then $b = -a = 0$, then $c = -b = 0$, then $d = -c = 0$. Hence $a = b = c = d = 0$, so L is linearly independent.
Spanning. Let $S = \operatorname{span}(v_1 + v_2,\ v_2 + v_3,\ v_3 + v_4,\ v_4)$. Working from the end of the list backwards:

$$v_4 \in S, \qquad v_3 = (v_3 + v_4) - v_4 \in S, \qquad v_2 = (v_2 + v_3) - v_3 \in S, \qquad v_1 = (v_1 + v_2) - v_2 \in S,$$

where at each step we use that S is a subspace (2.6), hence closed under subtraction, together with the membership established at the previous step. Thus $v_1, v_2, v_3, v_4 \in S$, and since S is a subspace containing these four vectors, it contains their span:

$$V = \operatorname{span}(v_1, v_2, v_3, v_4) \subseteq S \subseteq V.$$

Therefore $S = V$, i.e. L spans V.
Being linearly independent and spanning V, the list $v_1 + v_2,\ v_2 + v_3,\ v_3 + v_4,\ v_4$ is a basis of V.

Problem 2B.8. Prove or give a counterexample: If v_1, v_2, v_3, v_4 is a basis of V and U is a subspace of V such that $v_1, v_2 \in U$ and $v_3 \notin U$ and $v_4 \notin U$, then v_1, v_2 is a basis of U.

Solution. The statement is false. Here is a counterexample.
Take $V = \mathbf{F}^4$ and let v_1, v_2, v_3, v_4 be the standard basis e_1, e_2, e_3, e_4 of $\mathbf{F}^4$, where e_k is the list of length 4 with 1 in the k^{th} slot and 0 elsewhere. Let

$$U = \operatorname{span}(e_1,\ e_2,\ e_3 + e_4) = \{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_3 = x_4\}.$$

To see that the two descriptions of U agree, note that $e_1, e_2, e_3 + e_4$ all lie in the set on the right, which is a subspace of $\mathbf{F}^4$, so the span is contained in it; conversely if $x_3 = x_4$ then $(x_1, x_2, x_3, x_4) = x_1e_1 + x_2e_2 + x_3(e_3 + e_4)$.
Now check the three hypotheses.
First, $v_1 = e_1 \in U$ and $v_2 = e_2 \in U$, because the third and fourth coordinates of each are both 0.
Second, $v_3 = e_3 \notin U$, because the third coordinate of e_3 is 1 and its fourth coordinate is 0, and $1 \neq 0$.
Third, $v_4 = e_4 \notin U$, for the same reason with the roles of the third and fourth coordinates reversed.
Thus all the hypotheses hold. But v_1, v_2 is not a basis of U: the vector $e_3 + e_4$ belongs to U, and it is not in $\operatorname{span}(e_1, e_2)$, since every element of $\operatorname{span}(e_1, e_2)$ has third coordinate 0 while $e_3 + e_4$ has third

coordinate 1. Hence v_1, v_2 does not span U and so is not a basis of U (see 2.26).
It is worth isolating exactly which half of the conclusion can fail. The list v_1, v_2 is a sublist of the linearly independent list v_1, v_2, v_3, v_4, and any sublist of a linearly independent list is linearly independent: a nontrivial vanishing linear combination of v_1, v_2 becomes a nontrivial vanishing linear combination of v_1, v_2, v_3, v_4 upon appending the coefficients $0, 0$. So the linear independence half of the definition of a basis always holds under these hypotheses. What can fail, as the counterexample shows, is the spanning half: knowing merely that v_3 and v_4 individually avoid U says nothing about whether some combination of them, such as $v_3 + v_4$, lies in U.

Problem 2B.9. Suppose $v_1, \dots, v_m$ is a list of vectors in V. For $k \in \{1, \dots, m\}$, let

$$w_k = v_1 + \cdots + v_k.$$

Show that $v_1, \dots, v_m$ is a basis of V if and only if $w_1, \dots, w_m$ is a basis of V.

Solution. Throughout, adopt the convention $w_0 = 0$. Then the defining relations can be inverted:

$$v_k = w_k - w_{k-1} \qquad \text{for each } k \in \{1, \dots, m\},$$

since $w_k - w_{k-1} = (v_1 + \cdots + v_k) - (v_1 + \cdots + v_{k-1}) = v_k$, and for $k = 1$ this reads $v_1 = w_1 - 0 = w_1$. Because a basis is precisely a linearly independent spanning list (2.26), it suffices to prove two things: that the two lists have the same span, and that one list is linearly independent if and only if the other is.
Step 1: the two lists have the same span.
Each $w_k = v_1 + \cdots + v_k$ is a linear combination of $v_1, \dots, v_m$, so $w_k \in \operatorname{span}(v_1, \dots, v_m)$ for every k. Since the span of a list is the smallest subspace containing all vectors in the list, this gives

$$\operatorname{span}(w_1, \dots, w_m) \subseteq \operatorname{span}(v_1, \dots, v_m).$$

Conversely, each $v_k = w_k - w_{k-1}$ is a linear combination of $w_1, \dots, w_m$ (with $w_0 = 0$ contributing nothing), so $v_k \in \operatorname{span}(w_1, \dots, w_m)$ for every k, giving the reverse inclusion. Hence

$$\operatorname{span}(v_1, \dots, v_m) = \operatorname{span}(w_1, \dots, w_m).$$

In particular, one list spans V if and only if the other does.
Step 2: if $v_1, \dots, v_m$ is linearly independent, then so is $w_1, \dots, w_m$.
Suppose $a_1, \dots, a_m \in \mathbf{F}$ satisfy $a_1 w_1 + \cdots + a_m w_m = 0$. Substituting $w_k = v_1 + \cdots + v_k$ and collecting the coefficient of each v_j (the vector v_j appears in w_k exactly when $k \geq j$) gives

$$0 = \sum_{k=1}^{m} a_k \sum_{j=1}^{k} v_j = \sum_{j=1}^{m} \Big(\sum_{k=j}^{m} a_k \Big) v_j.$$

Since $v_1, \dots, v_m$ is linearly independent, every coefficient vanishes:

$$\sum_{k=j}^{m} a_k = 0 \qquad \text{for each } j \in \{1, \dots, m\}.$$

Taking $j = m$ gives $a_m = 0$. For $j < m$, subtracting the equation for $j + 1$ from the equation for j gives

$$a_j = \sum_{k=j}^{m} a_k - \sum_{k=j+1}^{m} a_k = 0 - 0 = 0.$$

Thus $a_1 = \cdots = a_m = 0$, so $w_1, \dots, w_m$ is linearly independent.
Step 3: if $w_1, \dots, w_m$ is linearly independent, then so is $v_1, \dots, v_m$.

Suppose $b_1, \dots, b_m \in \mathbf{F}$ satisfy $b_1 v_1 + \cdots + b_m v_m = 0$. Using $v_k = w_k - w_{k-1}$ and the convention $w_0 = 0$, and also setting $b_{m+1} = 0$, we compute

$$\begin{aligned} 0 &= \sum_{k=1}^{m} b_k (w_k - w_{k-1}) \\ &= \sum_{k=1}^{m} b_k w_k - \sum_{k=1}^{m-1} b_{k+1} w_k \\ &= \sum_{k=1}^{m} (b_k - b_{k+1}) w_k, \end{aligned}$$

where the last equality uses $b_{m+1} = 0$ so that the $k = m$ term is $b_m w_m$. Since $w_1, \dots, w_m$ is linearly independent, every coefficient vanishes: $b_k = b_{k+1}$ for each $k \in \{1, \dots, m\}$. Reading these backwards from $b_{m+1} = 0$ gives $b_m = 0$, then $b_{m-1} = 0$, and so on, yielding $b_1 = \cdots = b_m = 0$. Thus $v_1, \dots, v_m$ is linearly independent.
Conclusion.
Suppose $v_1, \dots, v_m$ is a basis of V. Then it spans V and is linearly independent. By Step 1, $w_1, \dots, w_m$ spans V; by Step 2, it is linearly independent. Hence $w_1, \dots, w_m$ is a basis of V.
Conversely, suppose $w_1, \dots, w_m$ is a basis of V. By Step 1, $v_1, \dots, v_m$ spans V; by Step 3, it is linearly independent. Hence $v_1, \dots, v_m$ is a basis of V.
Note that no finite-dimensionality hypothesis on V was used anywhere; the argument is a purely formal change of coordinates, the linear map sending $(a_1, \dots, a_m)$ to its list of partial sums being invertible over any field, with the successive-differences map as its inverse.

Problem 2B.10. Suppose U and W are subspaces of V such that $V = U \oplus W$. Suppose also that $u_1, \dots, u_m$ is a basis of U and $w_1, \dots, w_n$ is a basis of W. Prove that

$$u_1, \dots, u_m, w_1, \dots, w_n$$

is a basis of V.

Solution. We must show that the concatenated list spans V and is linearly independent (2.26).
The list spans V.
Let $v \in V$. Since $V = U \oplus W$, in particular $V = U + W$, so there exist $u \in U$ and $w \in W$ with

$$v = u + w.$$

Because $u_1, \dots, u_m$ is a basis of U, it spans U, so there exist $a_1, \dots, a_m \in \mathbf{F}$ with $u = a_1 u_1 + \cdots + a_m u_m$. Because $w_1, \dots, w_n$ is a basis of W, it spans W, so there exist $b_1, \dots, b_n \in \mathbf{F}$ with $w = b_1 w_1 + \cdots + b_n w_n$. Hence

$$v = a_1 u_1 + \cdots + a_m u_m + b_1 w_1 + \cdots + b_n w_n,$$

exhibiting v as a linear combination of $u_1, \dots, u_m, w_1, \dots, w_n$. Since $v \in V$ was arbitrary, the list spans V.
The list is linearly independent.
Suppose $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbf{F}$ are such that

$$a_1 u_1 + \cdots + a_m u_m + b_1 w_1 + \cdots + b_n w_n = 0.$$

Set

$$u = a_1 u_1 + \cdots + a_m u_m \quad \text{and} \quad w = b_1 w_1 + \cdots + b_n w_n.$$

Then $u \in U$ and $w \in W$ (each is a linear combination of vectors in the relevant subspace, and subspaces are closed under addition and scalar multiplication), and $u + w = 0$. Because the sum $U + W$ is a direct sum, 1.45 tells us that the only way to write 0 as a sum of a vector in U and a vector in W is with both summands equal to 0. Hence

$$u = 0 \quad \text{and} \quad w = 0.$$

Now $a_1u_1 + \cdots + a_mu_m = 0$ and $u_1, \ldots, u_m$ is linearly independent (being a basis of U), so $a_1 = \cdots = a_m = 0$. Likewise $b_1w_1 + \cdots + b_nw_n = 0$ and $w_1, \ldots, w_n$ is linearly independent, so $b_1 = \cdots = b_n = 0$. Thus all the scalars are 0, and the list $u_1, \ldots, u_m, w_1, \ldots, w_n$ is linearly independent.
Having shown that the list spans V and is linearly independent, we conclude that it is a basis of V.
Two remarks. First, no finite-dimensionality hypothesis on V was needed: the hypotheses that U and W have the finite bases named already force $V = U \oplus W$ to be finite-dimensional, and the proof itself uses only the definitions. Second, in place of 1.45 one could invoke 1.46 with a slightly longer argument: from $u + w = 0$ we get $u = -w \in U \cap W = \{0\}$, so $u = 0$ and then $w = -u = 0$.

Problem 2B.11. Suppose V is a real vector space. Show that if $v_1, \ldots, v_n$ is a basis of V (as a real vector space), then $v_1, \ldots, v_n$ is also a basis of the complexification $V_{\mathbf{C}}$ (as a complex vector space). See Exercise 8 in Section 1B for the definition of the complexification $V_{\mathbf{C}}$.

Solution. Recall the definitions from Exercise 8 in Section 1B. As a set, $V_{\mathbf{C}} = V \times V$, and the ordered pair (u, v) is written $u + iv$. Thus two elements $u + iv$ and $u' + iv'$ of $V_{\mathbf{C}}$ are equal precisely when $u = u'$ and $v = v'$. Addition is defined by

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$

and complex scalar multiplication by

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

for all $a, b \in \mathbf{R}$ and all $u, v \in V$. The vector $v \in V$ is regarded as the element $v + i0$ of $V_{\mathbf{C}}$, and this is how the list $v_1, \ldots, v_n$ is to be read below. The zero vector of $V_{\mathbf{C}}$ is $0 + i0$.
A single computation drives everything. For $a, b \in \mathbf{R}$ and $v \in V$, the scalar multiplication rule with the second component equal to 0 gives

$$(a + bi)(v + i0) = (av - b \cdot 0) + i(a \cdot 0 + bv) = (av) + i(bv).$$

Consequently, for real numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$, repeated use of the addition rule yields

$$\sum_{k=1}^{n}(a_k + b_ki)(v_k + i0) = \sum_{k=1}^{n}\big((a_kv_k) + i(b_kv_k)\big) = \Big(\sum_{k=1}^{n} a_kv_k\Big) + i\Big(\sum_{k=1}^{n} b_kv_k\Big).$$

Call this identity $(*)$. Note that every complex number has the form $a + bi$ with $a, b \in \mathbf{R}$, so $(*)$ covers all complex linear combinations of $v_1 + i0, \ldots, v_n + i0$.
The list spans $V_{\mathbf{C}}$ over $\mathbf{C}$.
Let $u + iv$ be an arbitrary element of $V_{\mathbf{C}}$, where $u, v \in V$. Since $v_1, \ldots, v_n$ spans the real vector space V, there exist real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ with

$$u = a_1v_1 + \cdots + a_nv_n \quad \text{and} \quad v = b_1v_1 + \cdots + b_nv_n.$$

By $(*)$,

$$\sum_{k=1}^{n}(a_k + b_ki)(v_k + i0) = u + iv.$$

Thus $u + iv$ lies in the complex span of $v_1, \ldots, v_n$. Since $u + iv \in V_{\mathbf{C}}$ was arbitrary, the list spans $V_{\mathbf{C}}$.
The list is linearly independent over $\mathbf{C}$.
Suppose $\lambda_1, \ldots, \lambda_n \in \mathbf{C}$ satisfy

$$\lambda_1(v_1 + i0) + \cdots + \lambda_n(v_n + i0) = 0 + i0.$$

Write $\lambda_k = a_k + b_ki$ with $a_k, b_k \in \mathbf{R}$. By $(*)$ the left side equals

$$\Big(\sum_{k=1}^{n} a_kv_k\Big) + i\Big(\sum_{k=1}^{n} b_kv_k\Big).$$

Two elements of $V_{\mathbf{C}} = V \times V$ are equal exactly when their two components agree, so

$$\sum_{k=1}^{n} a_k v_k = 0 \quad \text{and} \quad \sum_{k=1}^{n} b_k v_k = 0,$$

these being equations in V. Because $v_1, \ldots, v_n$ is a basis of the real vector space V, it is linearly independent over $\mathbf{R}$, and the a_k and b_k are real. Hence

$$a_1 = \cdots = a_n = 0 \quad \text{and} \quad b_1 = \cdots = b_n = 0,$$

so $\lambda_k = a_k + b_k i = 0$ for every k. Thus $v_1, \ldots, v_n$ is linearly independent in $V_{\mathbf{C}}$ over $\mathbf{C}$.
Since the list spans $V_{\mathbf{C}}$ and is linearly independent there, it is a basis of the complex vector space $V_{\mathbf{C}}$ (2.26).
The moral is that complexification does not change the number of basis vectors, only the field of scalars: the real degrees of freedom that were split between the real and imaginary components of $u + iv$ get recombined into the real and imaginary parts of the complex coefficients. Anticipating the language of Section 2C, this says $\dim_{\mathbf{C}} V_{\mathbf{C}} = \dim_{\mathbf{R}} V$. The model case is $V = \mathbf{R}^n$, whose complexification is $\mathbf{C}^n$: the standard basis of $\mathbf{R}^n$ over $\mathbf{R}$ is also the standard basis of $\mathbf{C}^n$ over $\mathbf{C}$.

2.3 Exercises 2C

Problem 2C.1. Show that the subspaces of $\mathbb{R}^2$ are precisely $\{0\}$, all lines in $\mathbb{R}^2$ containing the origin, and $\mathbb{R}^2$.

Solution. Recall that a *line* in $\mathbb{R}^2$ is a set of the form

$$L = \{w + \lambda v : \lambda \in \mathbb{R}\}$$

for some $w \in \mathbb{R}^2$ and some $v \in \mathbb{R}^2$ with $v \neq 0$.
First observe that a line containing the origin is exactly the span of a single nonzero vector. Indeed, suppose $L = \{w + \lambda v : \lambda \in \mathbb{R}\}$ with $v \neq 0$ and $0 \in L$. Then $w = -\lambda_0 v$ for some $\lambda_0 \in \mathbb{R}$, so

$$L = \{(\lambda - \lambda_0)v : \lambda \in \mathbb{R}\} = \{\mu v : \mu \in \mathbb{R}\} = \operatorname{span}(v).$$

Conversely, if $v \neq 0$ then $\operatorname{span}(v) = \{0 + \lambda v : \lambda \in \mathbb{R}\}$ is a line containing the origin.
Each listed set is a subspace. The set $\{0\}$ and the whole space $\mathbb{R}^2$ are subspaces. A line containing the origin equals $\operatorname{span}(v)$ for some $v \neq 0$, and the span of a list of vectors in $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$ (by 2.6).
Each subspace is one of the listed sets. Suppose U is a subspace of $\mathbb{R}^2$. Since $\mathbb{R}^2$ is finite-dimensional, U is finite-dimensional (by 2.25), and $\dim U \leq \dim \mathbb{R}^2 = 2$ by 2.37 and 2.36. So $\dim U \in \{0, 1, 2\}$.
If $\dim U = 0$, then the empty list is a basis of U; the span of the empty list is $\{0\}$ by 2.4, so $U = \{0\}$.
If $\dim U = 1$, then U has a basis consisting of a single vector v, and $v \neq 0$ because a list containing 0 is not linearly independent. Hence $U = \operatorname{span}(v)$, which by the observation above is a line in $\mathbb{R}^2$ containing the origin.
If $\dim U = 2 = \dim \mathbb{R}^2$, then $U = \mathbb{R}^2$ by 2.39.
Thus the subspaces of $\mathbb{R}^2$ are precisely $\{0\}$, the lines through the origin, and $\mathbb{R}^2$.

Problem 2C.2. Show that the subspaces of $\mathbb{R}^3$ are precisely $\{0\}$, all lines in $\mathbb{R}^3$ containing the origin, all planes in $\mathbb{R}^3$ containing the origin, and $\mathbb{R}^3$.

Solution. Recall the definitions: a *line* in $\mathbb{R}^3$ is a set of the form $\{w + \lambda v : \lambda \in \mathbb{R}\}$ with $v \neq 0$, and a *plane* in $\mathbb{R}^3$ is a set of the form $\{w + \lambda_1 v_1 + \lambda_2 v_2 : \lambda_1, \lambda_2 \in \mathbb{R}\}$ where v_1, v_2 is a linearly independent list in $\mathbb{R}^3$.
Step 1: lines through the origin are exactly the 1*-dimensional subspaces.* If $L = \{w + \lambda v : \lambda \in \mathbb{R}\}$ with

$v \neq 0$ and $0 \in L$, then $w = -\lambda_0 v$ for some λ_0, so $L = \{\mu v : \mu \in \mathbb{R}\} = \operatorname{span}(v)$, a subspace of dimension 1 (the list v is linearly independent because $v \neq 0$, and it spans L). Conversely, a 1-dimensional subspace has a basis consisting of one vector v, necessarily nonzero, and equals $\operatorname{span}(v) = \{0 + \lambda v : \lambda \in \mathbb{R}\}$, a line containing the origin.

Step 2: planes through the origin are exactly the 2-dimensional subspaces. Suppose $P = \{w + \lambda_1 v_1 + \lambda_2 v_2 : \lambda_1, \lambda_2 \in \mathbb{R}\}$ with v_1, v_2 linearly independent, and $0 \in P$. Then $w = -(\mu_1 v_1 + \mu_2 v_2)$ for some $\mu_1, \mu_2 \in \mathbb{R}$, and hence

$$P = \{(\lambda_1 - \mu_1)v_1 + (\lambda_2 - \mu_2)v_2 : \lambda_1, \lambda_2 \in \mathbb{R}\} = \operatorname{span}(v_1, v_2).$$

This is a subspace (by 2.6) with basis v_1, v_2, so $\dim P = 2$. Conversely, a 2-dimensional subspace U has a basis v_1, v_2; this list is linearly independent and $U = \operatorname{span}(v_1, v_2) = \{0 + \lambda_1 v_1 + \lambda_2 v_2 : \lambda_1, \lambda_2 \in \mathbb{R}\}$, a plane containing the origin.

Step 3: the classification. Each of the four listed kinds of set is a subspace: $\{0\}$ and $\mathbb{R}^3$ obviously, and lines and planes through the origin by Steps 1 and 2.

Conversely, let U be a subspace of $\mathbb{R}^3$. Since $\mathbb{R}^3$ is finite-dimensional, so is U (by 2.25), and $\dim U \leq \dim \mathbb{R}^3 = 3$ by 2.37 and 2.36. Hence $\dim U \in \{0, 1, 2, 3\}$.

If $\dim U = 0$, then $U = \{0\}$, since the empty list is a basis of U and its span is $\{0\}$.

If $\dim U = 1$, then U is a line containing the origin, by Step 1.

If $\dim U = 2$, then U is a plane containing the origin, by Step 2.

If $\dim U = 3 = \dim \mathbb{R}^3$, then $U = \mathbb{R}^3$ by 2.39.

Thus the subspaces of $\mathbb{R}^3$ are precisely $\{0\}$, the lines through the origin, the planes through the origin, and $\mathbb{R}^3$.

Problem 2C.3. (a) Let $U = \{p \in \mathcal{P}_4(\mathbb{F}) : p(6) = 0\}$. Find a basis of U.
(b) Extend the basis in (a) to a basis of $\mathcal{P}_4(\mathbb{F})$.
(c) Find a subspace W of $\mathcal{P}_4(\mathbb{F})$ such that $\mathcal{P}_4(\mathbb{F}) = U \oplus W$.

Solution. Throughout, write $\mathcal{P}_4 = \mathcal{P}_4(\mathbb{F})$; recall $\dim \mathcal{P}_4 = 5$ by 2.36. Note first that U is a subspace of $\mathcal{P}_4$: the zero polynomial vanishes at 6, and if $p(6) = q(6) = 0$ and $\lambda \in \mathbb{F}$ then $(p + q)(6) = 0$ and $(\lambda p)(6) = 0$.

A convenient basis of $\mathcal{P}_4$. Consider the list

$$1,\ (x-6),\ (x-6)^2,\ (x-6)^3,\ (x-6)^4.$$

It is linearly independent: suppose $c_0 + c_1(x-6) + c_2(x-6)^2 + c_3(x-6)^3 + c_4(x-6)^4 = 0$ as a polynomial. Expanding, the coefficient of x^4 is c_4, so $c_4 = 0$; then the coefficient of x^3 is c_3, so $c_3 = 0$; continuing downward gives $c_2 = c_1 = c_0 = 0$. This list has length $5 = \dim \mathcal{P}_4$, so it is a basis of $\mathcal{P}_4$ by 2.38.

(a) Every $p \in \mathcal{P}_4$ therefore has a unique expansion

$$p = c_0 + c_1(x-6) + c_2(x-6)^2 + c_3(x-6)^3 + c_4(x-6)^4,$$

and evaluating at 6 gives $p(6) = c_0$. Hence $p \in U$ if and only if $c_0 = 0$, i.e. if and only if p lies in the span of

$$(x-6),\ (x-6)^2,\ (x-6)^3,\ (x-6)^4.$$

These four polynomials lie in U and span U, and they are linearly independent (being a sublist of a linearly independent list). So they form a basis of U, and $\dim U = 4$.

(b) Adjoining the constant polynomial 1 gives

$$(x-6),\ (x-6)^2,\ (x-6)^3,\ (x-6)^4,\ 1,$$

which is (after reordering) the basis of $\mathcal{P}_4$ established above.

(c) Take $W = \operatorname{span}(1)$, the subspace of constant polynomials. Then $\mathcal{P}_4 = U \oplus W$.

Indeed, for any $p \in \mathcal{P}_4$ let c denote the constant polynomial whose value is $p(6)$, and write

$$p = (p - c) + c.$$

Here $(p - c)(6) = p(6) - p(6) = 0$, so $p - c \in U$, while $c \in W$. Hence $\mathcal{P}_4 = U + W$. Moreover, if $q \in U \cap W$, then q is a constant polynomial, say $q = c$, with $q(6) = 0$; thus $c = 0$ and $q = 0$. So $U \cap W = \{0\}$, and therefore the sum is direct by 1.46. Hence $\mathcal{P}_4(\mathbb{F}) = U \oplus W$.

Problem 2C.4. (a) Let $U = \{p \in \mathcal{P}_4(\mathbb{R}) : p''(6) = 0\}$. Find a basis of U.
(b) Extend the basis in (a) to a basis of $\mathcal{P}_4(\mathbb{R})$.
(c) Find a subspace W of $\mathcal{P}_4(\mathbb{R})$ such that $\mathcal{P}_4(\mathbb{R}) = U \oplus W$.

Solution. Write $\mathcal{P}_4 = \mathcal{P}_4(\mathbb{R})$; recall $\dim \mathcal{P}_4 = 5$ by 2.36. The set U is a subspace of $\mathcal{P}_4$: the second derivative of the zero polynomial is 0, so the zero polynomial lies in U, and differentiation is additive and commutes with scalar multiplication, so $(p + q)''(6) = p''(6) + q''(6)$ and $(\lambda p)''(6) = \lambda p''(6)$.
A convenient basis of $\mathcal{P}_4$. The list

$$1,\ (x - 6),\ (x - 6)^2,\ (x - 6)^3,\ (x - 6)^4$$

is linearly independent: if $c_0 + c_1(x - 6) + \cdots + c_4(x - 6)^4 = 0$, then expanding and reading off the coefficient of x^4 gives $c_4 = 0$, then the coefficient of x^3 gives $c_3 = 0$, and continuing downward gives $c_2 = c_1 = c_0 = 0$. Having length $5 = \dim \mathcal{P}_4$, it is a basis of $\mathcal{P}_4$ by 2.38.
(a) Write $p \in \mathcal{P}_4$ in this basis:

$$p = c_0 + c_1(x - 6) + c_2(x - 6)^2 + c_3(x - 6)^3 + c_4(x - 6)^4.$$

Differentiating twice,

$$p'' = 2c_2 + 6c_3(x - 6) + 12c_4(x - 6)^2,$$

so $p''(6) = 2c_2$. Hence $p \in U$ if and only if $c_2 = 0$, i.e. if and only if p lies in the span of

$$1,\ (x - 6),\ (x - 6)^3,\ (x - 6)^4.$$

Each of these four polynomials lies in U (their second derivatives are 0, 0, $6(x - 6)$, and $12(x - 6)^2$, all of which vanish at 6), they span U by the computation above, and they are linearly independent as a sublist of a linearly independent list. So they form a basis of U, and $\dim U = 4$.
(b) Adjoin $(x - 6)^2$. The resulting list

$$1,\ (x - 6),\ (x - 6)^2,\ (x - 6)^3,\ (x - 6)^4$$

is the basis of $\mathcal{P}_4$ exhibited above.
(c) Take $W = \operatorname{span}\big((x - 6)^2\big)$. For any $p \in \mathcal{P}_4$ set $\lambda = \frac{1}{2}p''(6)$ and write

$$p = \Big(p - \lambda(x - 6)^2\Big) + \lambda(x - 6)^2.$$

The second summand lies in W. The first lies in U, since the second derivative of $\lambda(x - 6)^2$ is the constant $2\lambda = p''(6)$, so

$$\big(p - \lambda(x - 6)^2\big)''(6) = p''(6) - p''(6) = 0.$$

Hence $\mathcal{P}_4 = U + W$. If $q \in U \cap W$, then $q = c(x - 6)^2$ for some $c \in \mathbb{R}$ and $q''(6) = 2c = 0$, so $c = 0$ and $q = 0$. Thus $U \cap W = \{0\}$, and by 1.46 the sum is direct: $\mathcal{P}_4(\mathbb{R}) = U \oplus W$.

Problem 2C.5. (a) Let $U = \{p \in \mathcal{P}_4(\mathbb{F}) : p(2) = p(5)\}$. Find a basis of U.
(b) Extend the basis in (a) to a basis of $\mathcal{P}_4(\mathbb{F})$.
(c) Find a subspace W of $\mathcal{P}_4(\mathbb{F})$ such that $\mathcal{P}_4(\mathbb{F}) = U \oplus W$.

Solution. Write $\mathcal{P}_4 = \mathcal{P}_4(\mathbb{F})$, so $\dim \mathcal{P}_4 = 5$ by 2.36. The set U is a subspace: the zero polynomial satisfies the condition, and if $p(2) = p(5)$ and $q(2) = q(5)$ then $(p+q)(2) = (p+q)(5)$ and $(\lambda p)(2) = (\lambda p)(5)$ for every $\lambda \in \mathbb{F}$.
A convenient basis of $\mathcal{P}_4$. Consider

$$1, \quad x-2, \quad (x-2)(x-5), \quad x(x-2)(x-5), \quad x^2(x-2)(x-5).$$

These are monic polynomials of degrees $0, 1, 2, 3, 4$. If a linear combination $c_0 \cdot 1 + c_1(x-2) + c_2(x-2)(x-5) + c_3x(x-2)(x-5) + c_4x^2(x-2)(x-5)$ equals the zero polynomial, then reading off the coefficient of x^4 gives $c_4 = 0$; with that, the coefficient of x^3 gives $c_3 = 0$; then x^2 gives $c_2 = 0$, then x^1 gives $c_1 = 0$, and finally $c_0 = 0$. So the list is linearly independent; having length $5 = \dim \mathcal{P}_4$, it is a basis of $\mathcal{P}_4$ by 2.38.
(a) Write $p \in \mathcal{P}_4$ in this basis:

$$p = c_0 + c_1(x-2) + c_2(x-2)(x-5) + c_3x(x-2)(x-5) + c_4x^2(x-2)(x-5).$$

Every term after the second vanishes at both 2 and 5, so

$$p(2) = c_0, \qquad p(5) = c_0 + 3c_1.$$

Hence $p(2) = p(5)$ if and only if $3c_1 = 0$, i.e. (since $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$) if and only if $c_1 = 0$. Therefore U is exactly the span of

$$1, \quad (x-2)(x-5), \quad x(x-2)(x-5), \quad x^2(x-2)(x-5),$$

each of which does lie in U. Being a sublist of a linearly independent list, this list is linearly independent, so it is a basis of U and $\dim U = 4$.
(b) Adjoin $x - 2$. The resulting list

$$1, \quad x-2, \quad (x-2)(x-5), \quad x(x-2)(x-5), \quad x^2(x-2)(x-5)$$

is the basis of $\mathcal{P}_4$ exhibited above.
(c) Take $W = \operatorname{span}(x-2)$. Given $p \in \mathcal{P}_4$, set $\lambda = \frac{1}{3}\big(p(5) - p(2)\big)$ and write

$$p = \big(p - \lambda(x-2)\big) + \lambda(x-2).$$

The second summand lies in W. For the first, put $q = p - \lambda(x-2)$; then $q(2) = p(2)$ and

$$q(5) = p(5) - 3\lambda = p(5) - \big(p(5) - p(2)\big) = p(2) = q(2),$$

so $q \in U$. Hence $\mathcal{P}_4 = U + W$. If $r \in U \cap W$, then $r = c(x-2)$ for some $c \in \mathbb{F}$, and $r(2) = r(5)$ gives $0 = 3c$, so $c = 0$ and $r = 0$. Thus $U \cap W = \{0\}$, and by 1.46 the sum is direct: $\mathcal{P}_4(\mathbb{F}) = U \oplus W$.

Problem 2C.6. (a) Let $U = \{p \in \mathcal{P}_4(\mathbb{F}) : p(2) = p(5) = p(6)\}$. Find a basis of U.
(b) Extend the basis in (a) to a basis of $\mathcal{P}_4(\mathbb{F})$.
(c) Find a subspace W of $\mathcal{P}_4(\mathbb{F})$ such that $\mathcal{P}_4(\mathbb{F}) = U \oplus W$.

Solution. Write $\mathcal{P}_4 = \mathcal{P}_4(\mathbb{F})$, so $\dim \mathcal{P}_4 = 5$ by 2.36. As in the previous exercises, U is a subspace of $\mathcal{P}_4$: it contains 0, and the conditions $p(2) = p(5)$ and $p(5) = p(6)$ are preserved under addition and scalar multiplication.
A convenient basis of $\mathcal{P}_4$. Consider

$$1, \quad x-2, \quad (x-2)(x-5), \quad (x-2)(x-5)(x-6), \quad x(x-2)(x-5)(x-6).$$

These are monic of degrees $0, 1, 2, 3, 4$. If a linear combination with coefficients $c_0, \dots, c_4$ is the zero polynomial, then the coefficient of x^4 gives $c_4 = 0$, then the coefficient of x^3 gives $c_3 = 0$, then x^2 gives $c_2 = 0$, then x gives $c_1 = 0$, and finally $c_0 = 0$. So the list is linearly independent, and having length

$5 = \dim \mathcal{P}_4$ it is a basis of $\mathcal{P}_4$ by 2.38.
(a) Write $p \in \mathcal{P}_4$ in this basis:

$$p = c_0 + c_1(x-2) + c_2(x-2)(x-5) + c_3(x-2)(x-5)(x-6) + c_4 x(x-2)(x-5)(x-6).$$

The last two terms vanish at each of $2, 5, 6$. Evaluating,

$$p(2) = c_0, \qquad p(5) = c_0 + 3c_1, \qquad p(6) = c_0 + 4c_1 + 4c_2.$$

Thus $p \in U$ if and only if $3c_1 = 0$ and $4c_1 + 4c_2 = 0$, which (since 3 and 4 are invertible in $\mathbb{F}$) holds if and only if $c_1 = c_2 = 0$. Hence U is precisely the span of

$$1, \quad (x-2)(x-5)(x-6), \quad x(x-2)(x-5)(x-6),$$

all three of which do belong to U (the constant 1 takes the same value everywhere; the other two vanish at 2, 5, and 6). This list is linearly independent as a sublist of a linearly independent list, so it is a basis of U and $\dim U = 3$.
(b) Adjoin $x - 2$ and $(x-2)(x-5)$. The resulting list

$$1, \quad (x-2)(x-5)(x-6), \quad x(x-2)(x-5)(x-6), \quad x-2, \quad (x-2)(x-5)$$

is, after reordering, the basis of $\mathcal{P}_4$ exhibited above.
(c) Take $W = \operatorname{span}\big(x-2,\ (x-2)(x-5)\big)$.
Since the basis of U from (a) together with the two adjoined polynomials spans $\mathcal{P}_4$, and every vector in that spanning list lies in U or in W, we get $\mathcal{P}_4 = U + W$.
Now suppose $q \in U \cap W$. Because $q \in U$, part (a) gives scalars a_0, a_3, a_4 with

$$q = a_0 + a_3(x-2)(x-5)(x-6) + a_4\, x(x-2)(x-5)(x-6);$$

because $q \in W$, there are scalars b_1, b_2 with $q = b_1(x-2) + b_2(x-2)(x-5)$. Subtracting these two expressions expresses the zero polynomial as a linear combination of the basis of $\mathcal{P}_4$ from (b); by the uniqueness in the criterion for a basis (2.28), all coefficients are 0. In particular $a_0 = a_3 = a_4 = 0$, so $q = 0$.
Hence $U \cap W = \{0\}$, and by 1.46 the sum is direct: $\mathcal{P}_4(\mathbb{F}) = U \oplus W$.
(As a check on dimensions, $\dim U + \dim W = 3 + 2 = 5 = \dim \mathcal{P}_4$, consistent with 2.43.)

Problem 2C.7. (a) Let $U = \left\{p \in \mathcal{P}_4(\mathbb{R}) : \int_{-1}^{1} p = 0\right\}$. Find a basis of U.
(b) Extend the basis in (a) to a basis of $\mathcal{P}_4(\mathbb{R})$.
(c) Find a subspace W of $\mathcal{P}_4(\mathbb{R})$ such that $\mathcal{P}_4(\mathbb{R}) = U \oplus W$.

Solution. Write $\mathcal{P}_4 = \mathcal{P}_4(\mathbb{R})$, so $\dim \mathcal{P}_4 = 5$ by 2.36. Because the integral is linear, U is a subspace of $\mathcal{P}_4$: the zero polynomial has integral 0, and

$$\int_{-1}^{1} (p+q) = \int_{-1}^{1} p + \int_{-1}^{1} q, \qquad \int_{-1}^{1} \lambda p = \lambda \int_{-1}^{1} p.$$

The basic computation we need is, for an integer $k \geq 0$,

$$\int_{-1}^{1} x^k\, dx = \frac{1 - (-1)^{k+1}}{k+1} = \begin{cases} \dfrac{2}{k+1}, & k \text{ even}, \\ 0, & k \text{ odd}. \end{cases}$$

In particular $\int_{-1}^{1} 1 = 2$, $\int_{-1}^{1} x = 0$, $\int_{-1}^{1} x^2 = \frac{2}{3}$, $\int_{-1}^{1} x^3 = 0$, and $\int_{-1}^{1} x^4 = \frac{2}{5}$.
A convenient basis of $\mathcal{P}_4$. Consider

$$1, \quad x, \quad x^2 - \tfrac{1}{3}, \quad x^3, \quad x^4 - \tfrac{1}{5}.$$

These are monic of degrees $0, 1, 2, 3, 4$, so the list is linearly independent: if a linear combination with coefficients $c_0, \dots, c_4$ is the zero polynomial, the coefficient of x^4 forces $c_4 = 0$, then the coefficient of x^3 forces $c_3 = 0$, then x^2 forces $c_2 = 0$, then x forces $c_1 = 0$, and finally $c_0 = 0$. Having length $5 = \dim \mathcal{P}_4$, it is a basis of $\mathcal{P}_4$ by 2.38.
(a) By the computation above,

$$\int_{-1}^{1} \Big(x^2 - \tfrac{1}{3}\Big) = \tfrac{2}{3} - \tfrac{2}{3} = 0, \qquad \int_{-1}^{1} \Big(x^4 - \tfrac{1}{5}\Big) = \tfrac{2}{5} - \tfrac{2}{5} = 0,$$

and $\int_{-1}^{1} x = \int_{-1}^{1} x^3 = 0$, while $\int_{-1}^{1} 1 = 2$. So if

$$p = c_0 + c_1 x + c_2\Big(x^2 - \tfrac{1}{3}\Big) + c_3 x^3 + c_4\Big(x^4 - \tfrac{1}{5}\Big),$$

then $\int_{-1}^{1} p = 2c_0$. Hence $p \in U$ if and only if $c_0 = 0$, i.e. if and only if p lies in the span of

$$x, \quad x^2 - \tfrac{1}{3}, \quad x^3, \quad x^4 - \tfrac{1}{5}.$$

These four polynomials lie in U and span U, and they are linearly independent as a sublist of a linearly independent list. So they form a basis of U, and $\dim U = 4$.
(b) Adjoin the constant polynomial 1:

$$1, \quad x, \quad x^2 - \tfrac{1}{3}, \quad x^3, \quad x^4 - \tfrac{1}{5}$$

is the basis of $\mathcal{P}_4$ exhibited above.
(c) Take $W = \operatorname{span}(1)$, the subspace of constant polynomials. Given $p \in \mathcal{P}_4$, set $c = \frac{1}{2}\int_{-1}^{1} p$ and write

$$p = (p - c) + c.$$

The second summand is a constant, hence in W, and

$$\int_{-1}^{1} (p - c) = \int_{-1}^{1} p - 2c = 0,$$

so $p - c \in U$. Hence $\mathcal{P}_4 = U + W$. If $q \in U \cap W$, then q is a constant, say $q = a$, and $0 = \int_{-1}^{1} a = 2a$, so $a = 0$ and $q = 0$. Thus $U \cap W = \{0\}$, and by 1.46 the sum is direct: $\mathcal{P}_4(\mathbb{R}) = U \oplus W$.

Problem 2C.8. Suppose $v_1, \dots, v_m$ is linearly independent in V and $w \in V$. Prove that

$$\dim \operatorname{span}(v_1 + w, \dots, v_m + w) \geq m - 1.$$

Solution. Write $U = \operatorname{span}(v_1 + w, \dots, v_m + w)$. Since U is spanned by a finite list, U is finite-dimensional and $\dim U$ makes sense.
If $m = 1$ the claim reads $\dim U \geq 0$, which is true because dimensions are nonnegative. So assume $m \geq 2$.
Consider the list of $m - 1$ vectors

$$v_1 - v_m, \ v_2 - v_m, \ \dots, \ v_{m-1} - v_m.$$

First, every vector in this list belongs to U: for each $k \in \{1, \dots, m-1\}$,

$$v_k - v_m = (v_k + w) - (v_m + w),$$

which is a linear combination of $v_1 + w, \dots, v_m + w$ and hence lies in U.
Second, this list is linearly independent. Suppose $a_1, \dots, a_{m-1} \in \mathbb{F}$ satisfy

$$a_1(v_1 - v_m) + \cdots + a_{m-1}(v_{m-1} - v_m) = 0.$$

Regrouping, this says

$$a_1v_1 + \cdots + a_{m-1}v_{m-1} + \big(-(a_1 + \cdots + a_{m-1})\big)v_m = 0.$$

Because $v_1, \ldots, v_m$ is linearly independent, every coefficient in this equation is 0; in particular $a_1 = \cdots = a_{m-1} = 0$. Hence $v_1 - v_m, \ldots, v_{m-1} - v_m$ is linearly independent.
So U contains a linearly independent list of length $m-1$. A basis of U is a spanning list of U of length $\dim U$, and by 2.22 the length of a linearly independent list in U is at most the length of a spanning list of U. Therefore

$$m - 1 \le \dim U = \dim \operatorname{span}(v_1 + w, \ldots, v_m + w),$$

as desired.
Two remarks. The bound is sharp: if $w = -v_1$ and $m \ge 2$, then the list $v_1 + w, \ldots, v_m + w$ begins with 0 and one checks (by the same regrouping computation) that the remaining $m-1$ vectors $v_2 - v_1, \ldots, v_m - v_1$ are linearly independent, so the span has dimension exactly $m-1$. On the other hand, if $w \notin \operatorname{span}(v_1, \ldots, v_m)$, or more generally if w is not of the form $-\sum b_k v_k$ with $\sum b_k = 1$, then the dimension is m.

Problem 2C.9. Suppose m is a positive integer and $p_0, p_1, \ldots, p_m \in \mathcal{P}(\mathbb{F})$ are such that each p_k has degree k. Prove that $p_0, p_1, \ldots, p_m$ is a basis of $\mathcal{P}_m(\mathbb{F})$.

Solution. Each p_k has degree $k \le m$, so $p_0, p_1, \ldots, p_m$ is a list of vectors in $\mathcal{P}_m(\mathbb{F})$. This list has length $m+1$, and $\dim \mathcal{P}_m(\mathbb{F}) = m+1$ by 2.36. Hence by 2.38 (a linearly independent list of the right length is a basis), it suffices to prove that $p_0, p_1, \ldots, p_m$ is linearly independent.
Suppose $a_0, a_1, \ldots, a_m \in \mathbb{F}$ are such that

$$a_0p_0 + a_1p_1 + \cdots + a_mp_m = 0,$$

meaning that the left side is the zero function on $\mathbb{F}$. Suppose, for the sake of contradiction, that not all the a_k equal 0. Let k be the largest index in $\{0, 1, \ldots, m\}$ with $a_k \ne 0$. Then

$$a_0p_0 + a_1p_1 + \cdots + a_kp_k = 0.$$

For each j, write $p_j(z) = c_{j,0} + c_{j,1}z + \cdots + c_{j,j}z^j$ with $c_{j,j} \ne 0$ (possible because p_j has degree j). Collecting coefficients, the left side of the displayed equation is the polynomial whose coefficient of z^k equals

$$a_kc_{k,k},$$

because for $j < k$ the polynomial p_j has degree $j < k$ and so contributes nothing to the coefficient of z^k. Since $a_k \ne 0$ and $c_{k,k} \ne 0$, we get $a_kc_{k,k} \ne 0$.
But the left side of the displayed equation is the zero function, all of whose coefficients are 0. As noted immediately after 4.8, the coefficients of a polynomial are uniquely determined (if a polynomial had two different lists of coefficients, subtracting the two representations would produce a polynomial with a nonzero coefficient and yet infinitely many zeros, contradicting 4.8 together with the fact that $\mathbb{F}$, being $\mathbb{R}$ or $\mathbb{C}$, is infinite). Hence the coefficient of z^k on the left side must be 0, i.e. $a_kc_{k,k} = 0$, a contradiction.
Therefore all the a_k equal 0, so $p_0, p_1, \ldots, p_m$ is linearly independent, and therefore it is a basis of $\mathcal{P}_m(\mathbb{F})$.
Note that this argument covers the case $k = 0$ as well: there the equation reads $a_0p_0 = 0$ with p_0 the nonzero constant $c_{0,0}$, so $a_0c_{0,0} = 0$ forces $a_0 = 0$.

Problem 2C.10. Suppose m is a positive integer. For $0 \le k \le m$, let

$$p_k(x) = x^k(1-x)^{m-k}.$$

Show that $p_0, \ldots, p_m$ is a basis of $\mathcal{P}_m(\mathbb{F})$.
[The basis in this exercise leads to what are called Bernstein polynomials. You can do a web search to learn how Bernstein polynomials are used to approximate continuous functions on $[0, 1]$.]

Solution. Each p_k is a product of a polynomial of degree k and a polynomial of degree $m-k$, hence is a polynomial of degree m; in particular $p_k \in \mathcal{P}_m(\mathbb{F})$. Thus $p_0, \ldots, p_m$ is a list of $m+1$ vectors in $\mathcal{P}_m(\mathbb{F})$, and $\dim \mathcal{P}_m(\mathbb{F}) = m+1$ by 2.36. So by 2.38 it suffices to show that this list is linearly independent.
Suppose $a_0, \ldots, a_m \in \mathbb{F}$ are such that

$$a_0 p_0 + a_1 p_1 + \cdots + a_m p_m = 0,$$

where 0 denotes the zero function. Suppose, for the sake of contradiction, that not all the a_k equal 0, and let j be the smallest index with $a_j \neq 0$. Then

$$\sum_{k=j}^{m} a_k x^k (1-x)^{m-k} = 0$$

for every $x \in \mathbb{F}$.
Now expand the left side as a polynomial and look at the coefficient of x^j. By the binomial theorem,

$$x^k (1-x)^{m-k} = \sum_{i=0}^{m-k} \binom{m-k}{i} (-1)^i x^{k+i},$$

so every monomial occurring in $x^k(1-x)^{m-k}$ has degree at least k. Consequently the terms with $k > j$ contribute nothing to the coefficient of x^j, and the term with $k = j$ contributes $a_j \binom{m-j}{0}(-1)^0 = a_j$. Hence the left side, written in standard form, has x^j-coefficient equal to $a_j \neq 0$.
But the left side of the displayed equation is the zero function, all of whose coefficients are 0. As noted immediately after 4.8, the coefficients of a polynomial are uniquely determined: if a polynomial had two different lists of coefficients, subtracting the two representations would give a polynomial with a nonzero coefficient and yet infinitely many zeros, which is impossible because such a polynomial has some degree $d \geq 0$, and a polynomial of degree $d \geq 1$ has at most d zeros by 4.8 while a polynomial of degree 0 is a nonzero constant and so has none, whereas $\mathbb{F}$ (being $\mathbb{R}$ or $\mathbb{C}$) is infinite. Hence the x^j-coefficient of the left side is 0, i.e. $a_j = 0$, contradicting the choice of j.
Therefore all the a_k equal 0, so $p_0, \ldots, p_m$ is linearly independent. Being a linearly independent list of length $\dim \mathcal{P}_m(\mathbb{F})$, it is a basis of $\mathcal{P}_m(\mathbb{F})$ by 2.38.
Remark: the actual Bernstein basis polynomials of degree m are $\binom{m}{k} x^k (1-x)^{m-k}$; multiplying each p_k by the nonzero scalar $\binom{m}{k}$ clearly does not affect linear independence, so those also form a basis of $\mathcal{P}_m(\mathbb{F})$.

Problem 2C.11. Suppose U and W are both four-dimensional subspaces of $\mathbb{C}^6$. Prove that there exist two vectors in $U \cap W$ such that neither of these vectors is a scalar multiple of the other.

Solution. Here $\mathbb{C}^6$ is regarded as a vector space over $\mathbb{C}$, so $\dim \mathbb{C}^6 = 6$ by 2.36 and in particular $\mathbb{C}^6$ is finite-dimensional. Every subspace of a finite-dimensional vector space is finite-dimensional (2.25), and its dimension is at most the dimension of the whole space (2.37). Since $U \cap W$ and $U + W$ are subspaces of $\mathbb{C}^6$ (the intersection by Exercise 10 in Section 1C, the sum by 1.40), both are finite-dimensional and $\dim(U+W) \leq 6$.
By 2.43,

$$\dim(U \cap W) = \dim U + \dim W - \dim(U+W) \geq 4 + 4 - 6 = 2.$$

Let $n = \dim(U \cap W) \geq 2$ and let $u_1, \ldots, u_n$ be a basis of $U \cap W$. Then u_1, u_2 are two vectors in $U \cap W$, and the list u_1, u_2 is linearly independent, because any sublist of a linearly independent list is linearly independent (if a nontrivial linear combination of u_1, u_2 equalled 0, then padding it with zero

coefficients on $u_3, \dots, u_n$ would give a nontrivial linear combination of $u_1, \dots, u_n$ equal to 0).
Finally, a linearly independent list of length two consists of two vectors neither of which is a scalar multiple of the other. Indeed, if $u_2 = \lambda u_1$ for some $\lambda \in \mathbb{C}$, then

$$\lambda u_1 + (-1)u_2 = 0$$

is a linear combination equal to 0 whose coefficients are not both 0 (the coefficient of u_2 is -1), contradicting linear independence. Symmetrically, if $u_1 = \lambda u_2$ then $(-1)u_1 + \lambda u_2 = 0$ gives the same contradiction.
Thus u_1 and u_2 are two vectors in $U \cap W$ such that neither is a scalar multiple of the other.

Problem 2C.12. Suppose that U and W are subspaces of $\mathbb{R}^8$ such that $\dim U = 3$, $\dim W = 5$, and $U + W = \mathbb{R}^8$. Prove that $\mathbb{R}^8 = U \oplus W$.

Solution. The space $\mathbb{R}^8$ has dimension eight (by 2.36), so it is finite-dimensional and hence all of its subspaces are finite-dimensional (by 2.25); in particular $U \cap W$, which is a subspace of $\mathbb{R}^8$ by Exercise 10 in Section 1C, is finite-dimensional.
Apply 2.43 to U and W:

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W).$$

By hypothesis $U + W = \mathbb{R}^8$, so $\dim(U + W) = 8$, and $\dim U = 3$, $\dim W = 5$. Hence

$$8 = 3 + 5 - \dim(U \cap W),$$

which gives $\dim(U \cap W) = 0$.
A vector space of dimension 0 equals $\{0\}$: a basis of $U \cap W$ is then a list of length 0, i.e. the empty list, and the span of the empty list is $\{0\}$, so $U \cap W = \{0\}$.
By 1.46, $U + W$ is a direct sum. Since $U + W = \mathbb{R}^8$, we conclude

$$\mathbb{R}^8 = U \oplus W,$$

as desired.
(Note that the numerical hypothesis $\dim U + \dim W = 8 = \dim \mathbb{R}^8$ is exactly what makes the argument work; without it, $U + W = \mathbb{R}^8$ would not force the sum to be direct.)

Problem 2C.13. Suppose U and W are both five-dimensional subspaces of $\mathbb{R}^9$. Prove that $U \cap W \neq \{0\}$.

Solution. The space $\mathbb{R}^9$ has dimension nine (by 2.36), so it is finite-dimensional. Both $U + W$ and $U \cap W$ are subspaces of $\mathbb{R}^9$ (by 1.40 and by Exercise 10 in Section 1C), hence are finite-dimensional (by 2.25) with dimension at most nine (by 2.37). In particular

$$\dim(U + W) \leq 9.$$

By 2.43,

$$\dim(U \cap W) = \dim U + \dim W - \dim(U + W) \geq 5 + 5 - 9 = 1.$$

Thus $\dim(U \cap W) \geq 1$. If we had $U \cap W = \{0\}$, then the empty list would be a basis of $U \cap W$ and so $\dim(U \cap W) = 0$, contradicting the inequality above. Hence $U \cap W \neq \{0\}$.
Equivalently: since $\dim(U \cap W) \geq 1$, the subspace $U \cap W$ has a basis of length at least one, and the vectors in a basis are nonzero (a list containing 0 is linearly dependent), so $U \cap W$ contains a nonzero vector.

Problem 2C.14. Suppose V is a ten-dimensional vector space and V_1, V_2, V_3 are subspaces of V with $\dim V_1 = \dim V_2 = \dim V_3 = 7$. Prove that $V_1 \cap V_2 \cap V_3 \neq \{0\}$.

Solution. Every subspace of the finite-dimensional space V is finite-dimensional (by 2.25) with dimension at most $\dim V = 10$ (by 2.37). Moreover the intersection of two subspaces of V is again a subspace of V (Exercise 10 in Section 1C), and the sum of two subspaces of V is again a subspace of V (by 1.40). So all the dimensions below are defined, and every sum appearing below has dimension at most 10.
Step 1: apply 2.43 to V_1 and V_2. Since $V_1 + V_2$ is a subspace of V, we have $\dim(V_1 + V_2) \leq 10$, so

$$\dim(V_1 \cap V_2) = \dim V_1 + \dim V_2 - \dim(V_1 + V_2) \geq 7 + 7 - 10 = 4.$$

Step 2: apply 2.43 to the subspaces $V_1 \cap V_2$ and V_3. Note that

$$(V_1 \cap V_2) \cap V_3 = V_1 \cap V_2 \cap V_3,$$

and $(V_1 \cap V_2) + V_3$ is a subspace of V, so its dimension is at most 10. Therefore

$$\begin{aligned} \dim(V_1 \cap V_2 \cap V_3) &= \dim(V_1 \cap V_2) + \dim V_3 - \dim\big((V_1 \cap V_2) + V_3\big) \\ &\geq 4 + 7 - 10 \\ &= 1. \end{aligned}$$

Hence $\dim(V_1 \cap V_2 \cap V_3) \geq 1$. If $V_1 \cap V_2 \cap V_3$ were $\{0\}$, its dimension would be 0 (the empty list is a basis of $\{0\}$), contradicting the inequality above. Therefore

$$V_1 \cap V_2 \cap V_3 \neq \{0\},$$

as desired.
Remark: the same telescoping argument shows in general that if $V_1, \ldots, V_n$ are subspaces of an N-dimensional space V, then

$$\dim(V_1 \cap \cdots \cap V_n) \geq \dim V_1 + \cdots + \dim V_n - (n-1)N,$$

which here gives $7 + 7 + 7 - 2 \cdot 10 = 1 > 0$.

Problem 2C.15. Suppose V is finite-dimensional and V_1, V_2, V_3 are subspaces of V with $\dim V_1 + \dim V_2 + \dim V_3 > 2 \dim V$. Prove that $V_1 \cap V_2 \cap V_3 \neq \{0\}$.

Solution. The key is the following observation: if A and B are subspaces of the finite-dimensional space V, then

$$\dim(A \cap B) \geq \dim A + \dim B - \dim V.$$

Indeed, $A + B$ is a subspace of V (by 1.40), so $\dim(A + B) \leq \dim V$ by 2.37; combining this with the formula for the dimension of a sum (2.43),

$$\dim(A \cap B) = \dim A + \dim B - \dim(A + B) \geq \dim A + \dim B - \dim V.$$

(All the subspaces appearing here are finite-dimensional by 2.25, so every dimension written is a well-defined nonnegative integer and 2.43 applies.)
Now apply this twice. First, with $A = V_1$ and $B = V_2$,

$$\dim(V_1 \cap V_2) \geq \dim V_1 + \dim V_2 - \dim V.$$

Next, $V_1 \cap V_2$ is a subspace of V, so we may apply the observation again with $A = V_1 \cap V_2$ and $B = V_3$:

$$\begin{aligned}\dim(V_1 \cap V_2 \cap V_3) &= \dim\big((V_1 \cap V_2) \cap V_3\big) \\ &\geq \dim(V_1 \cap V_2) + \dim V_3 - \dim V \\ &\geq \big(\dim V_1 + \dim V_2 - \dim V\big) + \dim V_3 - \dim V \\ &= \dim V_1 + \dim V_2 + \dim V_3 - 2\dim V.\end{aligned}$$

By hypothesis the last quantity is strictly positive. Hence $\dim(V_1 \cap V_2 \cap V_3) \geq 1$, and a subspace of dimension at least 1 contains a nonzero vector (the subspace $\{0\}$ has dimension 0, its basis being the empty list). Therefore $V_1 \cap V_2 \cap V_3 \neq \{0\}$.

Problem 2C.16. Suppose V is finite-dimensional and U is a subspace of V with $U \neq V$. Let $n = \dim V$ and $m = \dim U$. Prove that there exist $n - m$ subspaces of V, each of dimension $n - 1$, whose intersection equals U.

Solution. First note that $m < n$. Indeed $m \leq n$ by 2.37, and if $m = n$ then $U = V$ by 2.39, contrary to hypothesis. Thus $n - m \geq 1$, and in particular $n \geq 1$.
Let $u_1, \ldots, u_m$ be a basis of U (this is the empty list when $m = 0$, i.e. when $U = \{0\}$). A basis of U is a linearly independent list in V, so by 2.32 it extends to a basis

$$u_1, \ldots, u_m, v_1, \ldots, v_{n-m}$$

of V; the extended list has length n because every basis of V has length $\dim V = n$ (by 2.34).
For each $k \in \{1, \ldots, n - m\}$ define

$$W_k = \operatorname{span}(u_1, \ldots, u_m, v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_{n-m}),$$

that is, the span of the whole basis with v_k deleted. This produces $n - m$ subspaces of V.
Each W_k has dimension $n - 1$. The list spanning W_k has length $n - 1$, and it is a sublist of a linearly independent list, hence is itself linearly independent. A linearly independent list that spans W_k is a basis of W_k, so $\dim W_k = n - 1$.
The intersection contains U. Each of $u_1, \ldots, u_m$ appears in the spanning list of W_k, so $U = \operatorname{span}(u_1, \ldots, u_m) \subseteq W_k$ for every k. Hence $U \subseteq W_1 \cap \cdots \cap W_{n-m}$.
The intersection is contained in U. Suppose $v \in W_1 \cap \cdots \cap W_{n-m}$. Because $u_1, \ldots, u_m, v_1, \ldots, v_{n-m}$ is a basis of V, we may write

$$v = a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_{n-m} v_{n-m}$$

for scalars $a_1, \ldots, a_m, b_1, \ldots, b_{n-m}$, and by 2.28 this representation is *unique*. Fix k. Since $v \in W_k$, we can also write v as a linear combination of $u_1, \ldots, u_m, v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_{n-m}$, which is a representation of v in the basis of V whose coefficient of v_k equals 0. Uniqueness now forces $b_k = 0$. As k was arbitrary, $b_1 = \cdots = b_{n-m} = 0$, so

$$v = a_1 u_1 + \cdots + a_m u_m \in U.$$

Combining the two inclusions gives $W_1 \cap \cdots \cap W_{n-m} = U$, with $W_1, \ldots, W_{n-m}$ being $n - m$ subspaces of V each of dimension $n - 1$, as required.

Problem 2C.17. Suppose that $V_1, \ldots, V_m$ are finite-dimensional subspaces of V. Prove that $V_1 + \cdots + V_m$ is finite-dimensional and

$$\dim(V_1 + \cdots + V_m) \leq \dim V_1 + \cdots + \dim V_m.$$

[The inequality above is an equality if and only if $V_1 + \cdots + V_m$ is a direct sum, as will be shown in 3.94.]

Solution. For each $k \in \{1, \dots, m\}$ put $n_k = \dim V_k$ and choose a basis

$$v_{k,1}, \dots, v_{k,n_k}$$

of V_k; such a basis exists because V_k is finite-dimensional (by 2.31), and it has length n_k by the definition of dimension. (If $V_k = \{0\}$ then $n_k = 0$ and the list is empty; nothing below is affected.)
Let L denote the list obtained by concatenating these m lists:

$$v_{1,1}, \dots, v_{1,n_1},\ v_{2,1}, \dots, v_{2,n_2},\ \dots,\ v_{m,1}, \dots, v_{m,n_m}.$$

This is a finite list of vectors of V, of length $n_1 + \cdots + n_m$.
Claim: L spans $V_1 + \cdots + V_m$. By definition of the sum of subspaces (1.36), every element of $V_1 + \cdots + V_m$ has the form $u_1 + \cdots + u_m$ with $u_k \in V_k$ for each k. Since $v_{k,1}, \dots, v_{k,n_k}$ spans V_k, we may write

$$u_k = a_{k,1} v_{k,1} + \cdots + a_{k,n_k} v_{k,n_k}$$

for suitable scalars. Summing over k exhibits $u_1 + \cdots + u_m$ as a linear combination of the vectors in L. Hence $V_1 + \cdots + V_m \subseteq \operatorname{span} L$. Conversely each vector of L lies in some V_k and hence in the subspace $V_1 + \cdots + V_m$ (a sum of subspaces is a subspace, by 1.40), so $\operatorname{span} L \subseteq V_1 + \cdots + V_m$. Therefore

$$\operatorname{span} L = V_1 + \cdots + V_m.$$

Because $V_1 + \cdots + V_m$ is a vector space with a finite spanning list, it is finite-dimensional by definition. Finally, L is a spanning list of $V_1 + \cdots + V_m$ and any basis of $V_1 + \cdots + V_m$ is a linearly independent list in $V_1 + \cdots + V_m$; hence by 2.22 the length of that basis is at most the length of L. In other words

$$\dim(V_1 + \cdots + V_m) \le n_1 + \cdots + n_m = \dim V_1 + \cdots + \dim V_m,$$

as desired. (Equivalently, one can reduce L to a basis of $V_1 + \cdots + V_m$ using 2.30; the reduced list has length at most that of L.)

Problem 2C.18. Suppose V is finite-dimensional, with $\dim V = n \ge 1$. Prove that there exist one-dimensional subspaces $V_1, \dots, V_n$ of V such that

$$V = V_1 \oplus \cdots \oplus V_n.$$

Solution. Because V is finite-dimensional it has a basis (by 2.31), say $v_1, \dots, v_n$; this list has length $n = \dim V$ by the definition of dimension. For each $k \in \{1, \dots, n\}$ set

$$V_k = \operatorname{span}(v_k).$$

Each V_k is a subspace of V, since a span is always a subspace.
Each V_k is one-dimensional. The list $v_1, \dots, v_n$ is linearly independent, so no v_k equals 0: if $v_k = 0$ then $0v_1 + \cdots + 1v_k + \cdots + 0v_n = 0$ would be a nontrivial vanishing linear combination. A one-vector list v_k with $v_k \neq 0$ is linearly independent, because $av_k = 0$ with $a \neq 0$ would give $v_k = a^{-1}(av_k) = 0$. Thus v_k is a linearly independent list spanning V_k, hence a basis of V_k, and $\dim V_k = 1$.
The sum is all of V. Given $v \in V$, write $v = a_1 v_1 + \cdots + a_n v_n$ with $a_1, \dots, a_n \in \mathbf{F}$, which is possible because $v_1, \dots, v_n$ spans V. Each $a_k v_k \in V_k$, so $v \in V_1 + \cdots + V_n$. Conversely $V_1 + \cdots + V_n \subseteq V$ because each V_k is a subspace of V. Hence $V = V_1 + \cdots + V_n$.
The sum is direct. By the condition for a direct sum (1.45), it suffices to show that the only way to write $0 = w_1 + \cdots + w_n$ with each $w_k \in V_k$ is by taking every $w_k = 0$. So suppose $w_1 + \cdots + w_n = 0$ with $w_k \in V_k$. Each w_k has the form $w_k = a_k v_k$ for some $a_k \in \mathbf{F}$, so

$$a_1 v_1 + \cdots + a_n v_n = 0.$$

Since $v_1, \dots, v_n$ is linearly independent, $a_1 = \cdots = a_n = 0$, and therefore $w_1 = \cdots = w_n = 0$.
Hence $V = V_1 \oplus \cdots \oplus V_n$ with each V_k one-dimensional, as required.

Problem 2C.19. Explain why you might guess, motivated by analogy with the formula for the number of elements in the union of three finite sets, that if V_1, V_2, V_3 are subspaces of a finite-dimensional vector space, then

$$\begin{aligned}\dim(V_1 + V_2 + V_3) = {} & \dim V_1 + \dim V_2 + \dim V_3 \\ & - \dim(V_1 \cap V_2) - \dim(V_1 \cap V_3) - \dim(V_2 \cap V_3) \\ & + \dim(V_1 \cap V_2 \cap V_3).\end{aligned}$$

Then either prove the formula above or give a counterexample.

Solution. Why one might guess it. The table preceding these exercises sets up a dictionary between finite sets and finite-dimensional vector spaces: the set S corresponds to the space V, the number of elements $\#S$ corresponds to $\dim V$, the union $S_1 \cup S_2$ (the smallest subset containing S_1 and S_2) corresponds to the sum $V_1 + V_2$ (the smallest subspace containing V_1 and V_2, by 1.40), and intersection corresponds to intersection. Under this dictionary the inclusion-exclusion formula for two sets,

$$\#(S_1 \cup S_2) = \#S_1 + \#S_2 - \#(S_1 \cap S_2),$$

becomes exactly 2.43:

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2),$$

which is a true theorem. For three finite sets the counting formula reads

$$\begin{aligned}\#(S_1 \cup S_2 \cup S_3) = {} & \#S_1 + \#S_2 + \#S_3 \\ & - \#(S_1 \cap S_2) - \#(S_1 \cap S_3) - \#(S_2 \cap S_3) \\ & + \#(S_1 \cap S_2 \cap S_3),\end{aligned}$$

so translating term by term through the same dictionary produces the displayed guess. Since the dictionary gave the correct answer in the two-subspace case, it is natural to expect it to work again.
The guess is false. Take $V = \mathbf{R}^2$ and

$$V_1 = \operatorname{span}\big((1,0)\big), \qquad V_2 = \operatorname{span}\big((0,1)\big), \qquad V_3 = \operatorname{span}\big((1,1)\big).$$

Each of these is a one-dimensional subspace of $\mathbf{R}^2$ (each is the span of a single nonzero vector), so $\dim V_1 = \dim V_2 = \dim V_3 = 1$.
The sum contains $(1,0)$ and $(0,1)$, which span $\mathbf{R}^2$, so $V_1 + V_2 + V_3 = \mathbf{R}^2$ and

$$\dim(V_1 + V_2 + V_3) = 2.$$

Each pairwise intersection is $\{0\}$. For instance, a vector in $V_1 \cap V_3$ has the form $(a, 0)$ and also the form (b, b); equating second coordinates gives $b = 0$, so the vector is $(0, 0)$. The same argument (comparing the coordinate that vanishes) handles $V_1 \cap V_2$ and $V_2 \cap V_3$: an element of $V_1 \cap V_2$ is $(a, 0) = (0, b)$, forcing $a = b = 0$, and an element of $V_2 \cap V_3$ is $(0, a) = (b, b)$, forcing $b = 0$. Hence all three pairwise intersections, and therefore also $V_1 \cap V_2 \cap V_3$, equal $\{0\}$, and all four of these intersections have dimension 0.
The right side of the proposed formula therefore equals

$$1 + 1 + 1 - 0 - 0 - 0 + 0 = 3,$$

while the left side equals 2. Since $2 \neq 3$, the formula is false, and these three subspaces of $\mathbf{R}^2$ form a counterexample.
The analogy breaks because dim does not behave like a measure on subspaces: three distinct lines through the origin in the plane pairwise overlap only in $\{0\}$, yet they cannot be independent, and the "overlap" responsible for this is invisible to the pairwise intersections. Exercise 2C.20 shows what the correct three-subspace identity looks like: it involves the spaces $(V_1 + V_2) \cap V_3$ and its permutations, not merely the pairwise intersections.

Problem 2C.20. Prove that if V_1, V_2, and V_3 are subspaces of a finite-dimensional vector space, then

$$\begin{aligned}\dim(V_1+V_2+V_3) = {} & \dim V_1 + \dim V_2 + \dim V_3 \\ & - \frac{\dim(V_1\cap V_2)+\dim(V_1\cap V_3)+\dim(V_2\cap V_3)}{3} \\ & - \frac{\dim\big((V_1+V_2)\cap V_3\big)+\dim\big((V_1+V_3)\cap V_2\big)+\dim\big((V_2+V_3)\cap V_1\big)}{3}.\end{aligned}$$

[The formula above may seem strange because the right side does not look like an integer.]

Solution. Let V be the ambient finite-dimensional vector space. Every subspace appearing below is a subspace of V (sums of subspaces are subspaces by 1.40, and an intersection of subspaces is a subspace, as is immediate from the conditions in 1.34), hence is finite-dimensional by 2.25. So all the dimensions written are well-defined nonnegative integers and 2.43 may be applied freely.
Step 1: a two-term grouping. Apply 2.43 to the two subspaces V_1+V_2 and V_3, whose sum is $V_1+V_2+V_3$:

$$\dim(V_1+V_2+V_3) = \dim(V_1+V_2)+\dim V_3 - \dim\big((V_1+V_2)\cap V_3\big).$$

Apply 2.43 again, this time to V_1 and V_2:

$$\dim(V_1+V_2) = \dim V_1 + \dim V_2 - \dim(V_1\cap V_2).$$

Substituting gives

$$\dim(V_1+V_2+V_3) = \dim V_1+\dim V_2+\dim V_3 - \dim(V_1\cap V_2) - \dim\big((V_1+V_2)\cap V_3\big).$$

Call this identity (1).
Step 2: the other two groupings. Addition of subspaces is commutative and associative, so $V_1+V_2+V_3$ equals $(V_1+V_3)+V_2$ and also $(V_2+V_3)+V_1$. Repeating the computation of Step 1 with these groupings yields

$$\dim(V_1+V_2+V_3) = \dim V_1+\dim V_2+\dim V_3 - \dim(V_1\cap V_3) - \dim\big((V_1+V_3)\cap V_2\big)$$

(call this identity (2)) and

$$\dim(V_1+V_2+V_3) = \dim V_1+\dim V_2+\dim V_3 - \dim(V_2\cap V_3) - \dim\big((V_2+V_3)\cap V_1\big)$$

(call this identity (3)). Explicitly, for (2): $\dim(V_1+V_2+V_3) = \dim(V_1+V_3)+\dim V_2-\dim((V_1+V_3)\cap V_2)$ and $\dim(V_1+V_3) = \dim V_1+\dim V_3-\dim(V_1\cap V_3)$; and similarly for (3).
Step 3: average the three identities. Each of (1), (2), (3) has left side $\dim(V_1+V_2+V_3)$. Adding them and dividing by 3 leaves the left side unchanged, and gives

$$\begin{aligned}\dim(V_1+V_2+V_3) = {} & \dim V_1 + \dim V_2 + \dim V_3 \\ & - \frac{\dim(V_1\cap V_2)+\dim(V_1\cap V_3)+\dim(V_2\cap V_3)}{3} \\ & - \frac{\dim\big((V_1+V_2)\cap V_3\big)+\dim\big((V_1+V_3)\cap V_2\big)+\dim\big((V_2+V_3)\cap V_1\big)}{3},\end{aligned}$$

which is precisely the claimed formula.
Remark on the apparent non-integrality. Nothing is amiss. Write

$$c_1 = \dim(V_1\cap V_2)+\dim\big((V_1+V_2)\cap V_3\big),$$

and let c_2, c_3 be the analogous quantities from (2) and (3). Identities (1)-(3) say that $c_1 = c_2 = c_3$, each being equal to the integer $\dim V_1+\dim V_2+\dim V_3-\dim(V_1+V_2+V_3)$. The two fractions in the statement together form $(c_1+c_2+c_3)/3$, the average of three equal integers, which is that same integer. The formula only looks non-integral because the pairing of each intersection term with its matching sum-intersection term has been broken up in the display.

3 Linear Maps

Contents

3.1 Exercises 3A 71
3.2 Exercises 3B 84
3.3 Exercises 3C 105
3.4 Exercises 3D 121
3.5 Exercises 3E 139
3.6 Exercises 3F 155

3.1 Exercises 3A

Problem 3A.1. Suppose $b, c \in \mathbb{R}$. Define $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ by

$$T(x, y, z) = (2x - 4y + 3z + b,\ 6x + cxyz).$$

Show that T is linear if and only if $b = c = 0$.

Solution. First suppose $b = c = 0$, so that

$$T(x, y, z) = (2x - 4y + 3z,\ 6x).$$

To verify additivity, let $u = (x_1, y_1, z_1)$ and $v = (x_2, y_2, z_2)$ be elements of $\mathbb{R}^3$. Then

$$\begin{aligned} T(u + v) &= T(x_1 + x_2,\ y_1 + y_2,\ z_1 + z_2) \\ &= \big(2(x_1 + x_2) - 4(y_1 + y_2) + 3(z_1 + z_2),\ 6(x_1 + x_2)\big) \\ &= (2x_1 - 4y_1 + 3z_1,\ 6x_1) + (2x_2 - 4y_2 + 3z_2,\ 6x_2) \\ &= Tu + Tv, \end{aligned}$$

where the third equality holds by expanding each product and regrouping the resulting real numbers, using that addition in $\mathbb{R}^2$ is coordinatewise. To verify homogeneity, let $\lambda \in \mathbb{R}$ and $v = (x, y, z) \in \mathbb{R}^3$. Then

$$\begin{aligned} T(\lambda v) &= T(\lambda x, \lambda y, \lambda z) \\ &= (2\lambda x - 4\lambda y + 3\lambda z,\ 6\lambda x) \\ &= \lambda(2x - 4y + 3z,\ 6x) \\ &= \lambda(Tv). \end{aligned}$$

Hence T satisfies the two conditions in 3.1, so T is linear.
Now suppose, conversely, that T is linear; we must show $b = c = 0$.
Because T is linear, 3.10 gives $T(0, 0, 0) = (0, 0)$. On the other hand, directly from the formula defining T,

$$T(0, 0, 0) = (2 \cdot 0 - 4 \cdot 0 + 3 \cdot 0 + b,\ 6 \cdot 0 + c \cdot 0 \cdot 0 \cdot 0) = (b, 0).$$

Comparing the first coordinates of these two expressions gives $b = 0$.
To show that $c = 0$, apply homogeneity with $\lambda = 2$ and $v = (1, 1, 1)$. From the formula defining T,

$$T(1, 1, 1) = (2 - 4 + 3 + b,\ 6 + c) = (1 + b,\ 6 + c)$$

and

$$T(2, 2, 2) = (4 - 8 + 6 + b,\ 12 + 8c) = (2 + b,\ 12 + 8c),$$

since $c \cdot 2 \cdot 2 \cdot 2 = 8c$. Homogeneity requires $T(2, 2, 2) = 2\,T(1, 1, 1)$, and the second coordinate of $2\,T(1, 1, 1)$ equals $12 + 2c$. Thus

$$12 + 8c = 12 + 2c,$$

which gives $6c = 0$ and hence $c = 0$.
Therefore T is linear if and only if $b = c = 0$.

Problem 3A.2. Suppose $b, c \in \mathbb{R}$. Define $T\colon \mathcal{P}(\mathbb{R}) \to \mathbb{R}^2$ by

$$Tp = \left(3p(4) + 5p'(6) + b\,p(1)p(2),\ \int_{-1}^{2} x^3 p(x)\,dx + c \sin p(0)\right).$$

Show that T is linear if and only if $b = c = 0$.

Solution. First suppose $b = c = 0$, so that

$$Tp = \left(3p(4) + 5p'(6),\ \int_{-1}^{2} x^3 p(x)\,dx\right).$$

Let $p, q \in \mathcal{P}(\mathbb{R})$. Evaluation at a point is additive, and differentiation is additive (that is, $(p+q)' = p' + q'$), so the first coordinate of $T(p+q)$ is

$$3(p+q)(4) + 5(p+q)'(6) = \big(3p(4) + 5p'(6)\big) + \big(3q(4) + 5q'(6)\big).$$

The integral of a sum is the sum of the integrals, so the second coordinate of $T(p+q)$ is

$$\int_{-1}^{2} x^3\big(p(x) + q(x)\big)dx = \int_{-1}^{2} x^3 p(x)\,dx + \int_{-1}^{2} x^3 q(x)\,dx.$$

Since addition in $\mathbb{R}^2$ is coordinatewise, these two displays give $T(p+q) = Tp + Tq$.
Similarly, if $\lambda \in \mathbb{R}$ and $p \in \mathcal{P}(\mathbb{R})$, then $(\lambda p)(4) = \lambda p(4)$, $(\lambda p)' = \lambda p'$, and $\int_{-1}^{2} x^3 \lambda p(x)\,dx = \lambda \int_{-1}^{2} x^3 p(x)\,dx$. Hence each coordinate of $T(\lambda p)$ equals λ times the corresponding coordinate of Tp, so $T(\lambda p) = \lambda(Tp)$. Thus T is linear.
Now suppose, conversely, that T is linear; we must show $b = c = 0$.
To show $b = 0$, let p be the constant polynomial 1, so $p' = 0$. Then

$$3p(4) + 5p'(6) + b\,p(1)p(2) = 3 + b,$$

while $2p$ is the constant polynomial 2, whose first coordinate under T is

$$3 \cdot 2 + 5 \cdot 0 + b \cdot 2 \cdot 2 = 6 + 4b.$$

Homogeneity of T requires $T(2p) = 2\,(Tp)$; comparing first coordinates gives

$$6 + 4b = 2(3 + b) = 6 + 2b,$$

so $2b = 0$ and hence $b = 0$.
To show $c = 0$, let q be the constant polynomial $\pi/2$. First note that

$$\int_{-1}^{2} x^3\,dx = \left[\frac{x^4}{4}\right]_{-1}^{2} = \frac{16}{4} - \frac{1}{4} = \frac{15}{4}.$$

Since $q(0) = \pi/2$ and $\sin(\pi/2) = 1$, the second coordinate of Tq is

$$\int_{-1}^{2} x^3 \cdot \frac{\pi}{2}\,dx + c\sin\frac{\pi}{2} = \frac{15\pi}{8} + c.$$

Since $2q$ is the constant polynomial π and $\sin\pi = 0$, the second coordinate of $T(2q)$ is

$$\int_{-1}^{2} x^3 \cdot \pi\,dx + c\sin\pi = \frac{15\pi}{4}.$$

Homogeneity of T requires $T(2q) = 2\,(Tq)$; comparing second coordinates gives

$$\frac{15\pi}{4} = 2\left(\frac{15\pi}{8} + c\right) = \frac{15\pi}{4} + 2c,$$

so $2c = 0$ and hence $c = 0$.
Therefore T is linear if and only if $b = c = 0$.

Problem 3A.3. Suppose that $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$. Show that there exist scalars $A_{j,k} \in \mathbb{F}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$ such that

$$T(x_1, \ldots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \ \ldots, \ A_{m,1}x_1 + \cdots + A_{m,n}x_n)$$

for every $(x_1, \ldots, x_n) \in \mathbb{F}^n$.
[This exercise shows that the linear map T has the form promised in the second to last item of Example 3.3.]

Solution. Let $e_1, \ldots, e_n$ denote the standard basis of $\mathbb{F}^n$ (see 2.27); thus e_k is the element of $\mathbb{F}^n$ whose k^{th} coordinate equals 1 and whose other coordinates equal 0.
For each $k \in \{1, \ldots, n\}$ the vector Te_k lies in $\mathbb{F}^m$, so it has m coordinates; define $A_{j,k} \in \mathbb{F}$, for $j = 1, \ldots, m$, to be the j^{th} coordinate of Te_k. In other words, the scalars $A_{j,k}$ are defined by

$$Te_k = (A_{1,k}, \ A_{2,k}, \ \ldots, \ A_{m,k})$$

for each $k = 1, \ldots, n$. This defines $A_{j,k}$ for all $j = 1, \ldots, m$ and all $k = 1, \ldots, n$, as required.
Now let $(x_1, \ldots, x_n) \in \mathbb{F}^n$ be arbitrary. Because addition and scalar multiplication in $\mathbb{F}^n$ are defined coordinatewise, we have

$$(x_1, \ldots, x_n) = x_1e_1 + \cdots + x_ne_n.$$

The additivity of T, applied repeatedly (formally, by induction on the number of summands), shows that T takes a sum of n vectors to the sum of their images; combining this with the homogeneity of T gives

$$T(x_1, \ldots, x_n) = T(x_1e_1 + \cdots + x_ne_n) = x_1(Te_1) + \cdots + x_n(Te_n).$$

Substituting the coordinates of each Te_k and again using that the vector space operations on $\mathbb{F}^m$ are coordinatewise, we obtain

$$\begin{aligned} T(x_1, \ldots, x_n) &= x_1(A_{1,1}, \ldots, A_{m,1}) + \cdots + x_n(A_{1,n}, \ldots, A_{m,n}) \\ &= (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \ \ldots, \ A_{m,1}x_1 + \cdots + A_{m,n}x_n), \end{aligned}$$

because for each j the j^{th} coordinate of the middle expression equals $A_{j,1}x_1 + \cdots + A_{j,n}x_n$.
Since $(x_1, \ldots, x_n)$ was an arbitrary element of $\mathbb{F}^n$, the scalars $A_{j,k}$ constructed above have the required property.
Conversely, every choice of scalars $A_{j,k}$ produces a linear map by the formula above (this is the second to last item of Example 3.3), so the maps of this form are exactly the elements of $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$.

Problem 3A.4. Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \ldots, v_m$ is a list of vectors in V such that $Tv_1, \ldots, Tv_m$ is a linearly independent list in W. Prove that $v_1, \ldots, v_m$ is linearly independent.

Solution. Suppose $a_1, \ldots, a_m \in \mathbb{F}$ are such that

$$a_1v_1 + \cdots + a_mv_m = 0.$$

We must show that $a_1 = \cdots = a_m = 0$.
Apply T to both sides of the equation above. The homogeneity of T gives $T(a_kv_k) = a_k(Tv_k)$ for each k, and the additivity of T, applied repeatedly, shows that T takes a sum of m vectors to the sum of their images. Hence

$$T(a_1v_1 + \cdots + a_mv_m) = a_1(Tv_1) + \cdots + a_m(Tv_m).$$

On the other hand, the left side equals $T(0)$, which equals 0 by 3.10. Therefore

$$a_1(Tv_1) + \cdots + a_m(Tv_m) = 0.$$

Because $Tv_1, \ldots, Tv_m$ is a linearly independent list in W, the only choice of scalars making such a linear combination equal 0 is the choice in which all the scalars are 0. Thus $a_1 = \cdots = a_m = 0$.

We have shown that the only way to write 0 as a linear combination of $v_1, \dots, v_m$ is with all coefficients equal to 0. Hence $v_1, \dots, v_m$ is linearly independent, as desired.
Note that the converse of this result is false: if $T = 0$ and $v_1, \dots, v_m$ is any linearly independent list with $m \geq 1$, then $Tv_1, \dots, Tv_m$ is the list $0, \dots, 0$, which is not linearly independent.

Problem 3A.5. Prove that $\mathcal{L}(V, W)$ is a vector space, as was asserted in 3.6.

Solution. Recall the operations, defined in 3.5: for $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$,

$$(S + T)(v) = Sv + Tv \quad \text{and} \quad (\lambda T)(v) = \lambda(Tv)$$

for all $v \in V$. Throughout, two functions from V to W are equal precisely when they agree at every $v \in V$, so every identity below is verified by evaluating both sides at an arbitrary $v \in V$. All manipulations on the right-hand sides take place in the vector space W, whose vector space properties we use freely.
Step 1: the operations produce elements of $\mathcal{L}(V, W)$.
Let $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$. For $u, v \in V$,

$$\begin{aligned}(S + T)(u + v) &= S(u + v) + T(u + v) = (Su + Sv) + (Tu + Tv) \\ &= (Su + Tu) + (Sv + Tv) = (S + T)(u) + (S + T)(v),\end{aligned}$$

using the additivity of S and T and then the commutativity and associativity of addition in W. For $\mu \in \mathbb{F}$ and $v \in V$,

$$(S + T)(\mu v) = S(\mu v) + T(\mu v) = \mu(Sv) + \mu(Tv) = \mu(Sv + Tv) = \mu\big((S + T)(v)\big),$$

using homogeneity of S and T and a distributive property in W. Hence $S + T \in \mathcal{L}(V, W)$.
Similarly, for $u, v \in V$,

$$(\lambda T)(u + v) = \lambda\big(T(u + v)\big) = \lambda(Tu + Tv) = \lambda(Tu) + \lambda(Tv) = (\lambda T)(u) + (\lambda T)(v),$$

and for $\mu \in \mathbb{F}$ and $v \in V$,

$$(\lambda T)(\mu v) = \lambda\big(T(\mu v)\big) = \lambda\big(\mu(Tv)\big) = (\lambda\mu)(Tv) = (\mu\lambda)(Tv) = \mu\big(\lambda(Tv)\big) = \mu\big((\lambda T)(v)\big),$$

using associativity of scalar multiplication in W and commutativity of multiplication in $\mathbb{F}$. Hence $\lambda T \in \mathcal{L}(V, W)$. So $\mathcal{L}(V, W)$ is closed under both operations, and they are genuine operations on $\mathcal{L}(V, W)$.
Step 2: commutativity. For $S, T \in \mathcal{L}(V, W)$ and $v \in V$,

$$(S + T)(v) = Sv + Tv = Tv + Sv = (T + S)(v),$$

so $S + T = T + S$.
Step 3: associativity. For $R, S, T \in \mathcal{L}(V, W)$ and $v \in V$,

$$\big((R + S) + T\big)(v) = (Rv + Sv) + Tv = Rv + (Sv + Tv) = \big(R + (S + T)\big)(v),$$

so $(R + S) + T = R + (S + T)$. For $\lambda, \mu \in \mathbb{F}$, $T \in \mathcal{L}(V, W)$, and $v \in V$,

$$\big((\lambda\mu)T\big)(v) = (\lambda\mu)(Tv) = \lambda\big(\mu(Tv)\big) = \lambda\big((\mu T)(v)\big) = \big(\lambda(\mu T)\big)(v),$$

so $(\lambda\mu)T = \lambda(\mu T)$.
Step 4: additive identity. Let 0 denote the zero linear map from V to W defined in Example 3.3, so $0v = 0$ for all $v \in V$; it belongs to $\mathcal{L}(V, W)$. For every $T \in \mathcal{L}(V, W)$ and $v \in V$,

$$(T + 0)(v) = Tv + 0 = Tv,$$

so $T + 0 = T$. Thus $\mathcal{L}(V, W)$ has an additive identity.

Step 5: additive inverses. Let $T \in \mathcal{L}(V,W)$ and set $-T = (-1)T$, which lies in $\mathcal{L}(V,W)$ by Step 1. For every $v \in V$,

$$\big(T + (-T)\big)(v) = Tv + (-1)(Tv) = 0,$$

using 1.32 in W (namely that $(-1)w$ is the additive inverse of w). Hence $T + (-T) = 0$, so every element of $\mathcal{L}(V,W)$ has an additive inverse.
Step 6: multiplicative identity. For $T \in \mathcal{L}(V,W)$ and $v \in V$,

$$(1T)(v) = 1(Tv) = Tv,$$

so $1T = T$.
Step 7: distributive properties. For $\lambda \in \mathbb{F}$, $S, T \in \mathcal{L}(V,W)$, and $v \in V$,

$$\big(\lambda(S+T)\big)(v) = \lambda(Sv + Tv) = \lambda(Sv) + \lambda(Tv) = (\lambda S)(v) + (\lambda T)(v) = (\lambda S + \lambda T)(v),$$

so $\lambda(S+T) = \lambda S + \lambda T$. For $\lambda, \mu \in \mathbb{F}$, $T \in \mathcal{L}(V,W)$, and $v \in V$,

$$\big((\lambda+\mu)T\big)(v) = (\lambda+\mu)(Tv) = \lambda(Tv) + \mu(Tv) = (\lambda T)(v) + (\mu T)(v) = (\lambda T + \mu T)(v),$$

so $(\lambda+\mu)T = \lambda T + \mu T$.
All the conditions in the definition of a vector space (1.20) have now been verified. Hence $\mathcal{L}(V,W)$, with the operations of 3.5, is a vector space over $\mathbb{F}$, and its additive identity is the zero linear map.

Problem 3A.6. Prove that multiplication of linear maps has the associative, identity, and distributive properties asserted in 3.8.

Solution. Recall from 3.7 that if $T \in \mathcal{L}(U,V)$ and $S \in \mathcal{L}(V,W)$, then the product $ST \in \mathcal{L}(U,W)$ is defined by $(ST)(u) = S(Tu)$ for all $u \in U$. In each part below, the two sides of the claimed equation are functions with the same domain, so it suffices to show that they agree at every point of that domain.
Associativity. Suppose T_1, T_2, T_3 are linear maps for which the products below make sense; that is, suppose there are vector spaces V_1, V_2, V_3, V_4 with

$$T_3 \in \mathcal{L}(V_1, V_2), \qquad T_2 \in \mathcal{L}(V_2, V_3), \qquad T_1 \in \mathcal{L}(V_3, V_4).$$

Then $T_2T_3 \in \mathcal{L}(V_1, V_3)$ and $T_1T_2 \in \mathcal{L}(V_2, V_4)$, so both $(T_1T_2)T_3$ and $T_1(T_2T_3)$ are maps from V_1 to V_4. Let $u \in V_1$. Applying the definition of the product twice to each side,

$$\big((T_1T_2)T_3\big)(u) = (T_1T_2)(T_3u) = T_1\big(T_2(T_3u)\big)$$

and

$$\big(T_1(T_2T_3)\big)(u) = T_1\big((T_2T_3)(u)\big) = T_1\big(T_2(T_3u)\big).$$

The right sides agree, so $(T_1T_2)T_3 = T_1(T_2T_3)$. (Note that no linearity was used here; this is just associativity of composition of functions.)
Identity. Suppose $T \in \mathcal{L}(V,W)$. Let I denote the identity operator on V in the product TI, and the identity operator on W in the product IT; both products then make sense and are maps from V to W. For every $v \in V$,

$$(TI)(v) = T(Iv) = Tv \qquad \text{and} \qquad (IT)(v) = I(Tv) = Tv,$$

since the identity operator takes each vector to itself. Hence $TI = T$ and $IT = T$, so $TI = IT = T$.
Distributive properties. Suppose $T, T_1, T_2 \in \mathcal{L}(U,V)$ and $S, S_1, S_2 \in \mathcal{L}(V,W)$. All four products appearing below are maps from U to W, and by 3.5 so are the sums $S_1T + S_2T$ and $ST_1 + ST_2$.
For the first identity, let $u \in U$. Then

$$\begin{aligned}\big((S_1+S_2)T\big)(u) &= (S_1+S_2)(Tu)\\ &= S_1(Tu) + S_2(Tu)\\ &= (S_1T)(u) + (S_2T)(u)\\ &= (S_1T + S_2T)(u),\end{aligned}$$

where the first and third equalities use the definition of a product of linear maps and the second and fourth use the definition of the sum of two linear maps. Hence $(S_1 + S_2)T = S_1T + S_2T$.
For the second identity, let $u \in U$. Then

$$\begin{aligned}\big(S(T_1 + T_2)\big)(u) &= S\big((T_1 + T_2)(u)\big)\\ &= S(T_1u + T_2u)\\ &= S(T_1u) + S(T_2u)\\ &= (ST_1)(u) + (ST_2)(u)\\ &= (ST_1 + ST_2)(u),\end{aligned}$$

where the third equality uses the additivity of S (this is the one place where linearity, rather than just the definitions of sum and composition, is needed). Hence $S(T_1 + T_2) = ST_1 + ST_2$.
This establishes all three properties asserted in 3.8.

Problem 3A.7. Show that every linear map from a one-dimensional vector space to itself is multiplication by some scalar. More precisely, prove that if $\dim V = 1$ and $T \in \mathcal{L}(V)$, then there exists $\lambda \in \mathbb{F}$ such that $Tv = \lambda v$ for all $v \in V$.

Solution. Suppose $\dim V = 1$ and $T \in \mathcal{L}(V)$.
Because $\dim V = 1$, the space V has a basis consisting of exactly one vector; call it u. Thus

$$V = \operatorname{span}(u) = \{au : a \in \mathbb{F}\},$$

and $u \neq 0$ (a list containing 0 is not linearly independent, so a basis vector cannot be 0).
Since T maps V into V, we have $Tu \in V = \operatorname{span}(u)$. Hence there exists a scalar $\lambda \in \mathbb{F}$ such that

$$Tu = \lambda u.$$

We claim this λ works for every vector in V. Indeed, let $v \in V$ be arbitrary. Because u spans V, there exists $a \in \mathbb{F}$ with $v = au$. Then, using the homogeneity of T and then the associativity and commutativity of scalar multiplication,

$$Tv = T(au) = a(Tu) = a(\lambda u) = (a\lambda)u = (\lambda a)u = \lambda(au) = \lambda v.$$

Thus $Tv = \lambda v$ for all $v \in V$, which is exactly the statement that T is multiplication by the scalar λ.
Finally, note that λ is uniquely determined by T. If $\mu \in \mathbb{F}$ also satisfies $Tv = \mu v$ for all $v \in V$, then in particular $\lambda u = \mu u$, so $(\lambda - \mu)u = 0$. If $\lambda - \mu \neq 0$, multiplying both sides by $(\lambda - \mu)^{-1}$ would give $u = 0$, contradicting $u \neq 0$. Hence $\lambda = \mu$.
Conversely, for each $\lambda \in \mathbb{F}$ the map $v \mapsto \lambda v$ is a linear map from V to V (its additivity and homogeneity are immediate from the vector space properties of V). Combining this with the above, when $\dim V = 1$ the map $\lambda \mapsto \lambda I$ is a one-to-one correspondence between $\mathbb{F}$ and $\mathcal{L}(V)$.

Problem 3A.8. Give an example of a function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ such that

$$\varphi(av) = a\varphi(v)$$

for all $a \in \mathbb{R}$ and all $v \in \mathbb{R}^2$ but φ is not linear.
[This exercise and the next exercise show that neither homogeneity nor additivity alone is enough to imply that a function is a linear map.]

Solution. Define $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$\varphi(x,y) = \begin{cases} \dfrac{x^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

This is a well-defined function, since the denominator x^2+y^2 vanishes only at $(0,0)$, where the second clause applies instead.
Homogeneity. Let $a \in \mathbb{R}$ and $v = (x,y) \in \mathbb{R}^2$. If $a = 0$ or $v = (0,0)$, then $av = (0,0)$, so $\varphi(av) = 0$ and also $a\varphi(v) = 0$ (in the first case because the scalar is 0, in the second because $\varphi(0,0) = 0$); hence $\varphi(av) = a\varphi(v)$. Now suppose $a \neq 0$ and $(x,y) \neq (0,0)$. Then $av = (ax, ay) \neq (0,0)$, so

$$\varphi(av) = \frac{(ax)^3}{(ax)^2+(ay)^2} = \frac{a^3x^3}{a^2(x^2+y^2)} = a \cdot \frac{x^3}{x^2+y^2} = a\varphi(v).$$

Thus $\varphi(av) = a\varphi(v)$ for all $a \in \mathbb{R}$ and all $v \in \mathbb{R}^2$.
Failure of additivity. Take $u = (1,0)$ and $v = (0,1)$. Then

$$\varphi(u) = \frac{1}{1} = 1, \qquad \varphi(v) = \frac{0}{1} = 0, \qquad \varphi(u+v) = \varphi(1,1) = \frac{1}{2}.$$

Hence $\varphi(u+v) = \frac{1}{2} \neq 1 = \varphi(u) + \varphi(v)$, so φ is not additive and therefore is not a linear map (see 3.1). Thus φ satisfies the homogeneity condition of 3.1 for every scalar and every vector, yet φ is not linear: homogeneity alone does not imply linearity.

Problem 3A.9. Give an example of a function $\varphi\colon \mathbb{C} \to \mathbb{C}$ such that

$$\varphi(w+z) = \varphi(w) + \varphi(z)$$

for all $w, z \in \mathbb{C}$ but φ is not linear. (Here $\mathbb{C}$ is thought of as a complex vector space.)
[There also exists a function $\varphi\colon \mathbb{R} \to \mathbb{R}$ such that φ satisfies the additivity condition above but φ is not linear. However, showing the existence of such a function involves considerably more advanced tools.]

Solution. Define $\varphi\colon \mathbb{C} \to \mathbb{C}$ to be complex conjugation:

$$\varphi(z) = \overline{z},$$

so that $\varphi(a+bi) = a - bi$ for all $a, b \in \mathbb{R}$.
Additivity. Let $w = a+bi$ and $z = c+di$ with $a, b, c, d \in \mathbb{R}$. Then $w + z = (a+c) + (b+d)i$, so

$$\varphi(w+z) = (a+c) - (b+d)i = (a-bi) + (c-di) = \varphi(w) + \varphi(z).$$

Thus φ is additive.
Failure of homogeneity. Here $\mathbb{C}$ is regarded as a vector space over $\mathbb{C}$, so the scalars include i. Take $\lambda = i$ and $z = 1$. Then

$$\varphi(\lambda z) = \varphi(i) = \overline{i} = -i, \qquad \lambda\varphi(z) = i \cdot \overline{1} = i.$$

Since $-i \neq i$, the homogeneity condition of 3.1 fails, so φ is not a linear map from the complex vector space $\mathbb{C}$ to itself.
More generally, for any $z \neq 0$ we have $\varphi(iz) = \overline{iz} = \overline{i}\,\overline{z} = -i\overline{z}$ while $i\varphi(z) = i\overline{z}$. Their difference is $i\overline{z} - (-i\overline{z}) = 2i\overline{z}$, which is nonzero because $z \neq 0$ implies $\overline{z} \neq 0$. So homogeneity fails for every nonzero z when $\lambda = i$.
Together with the previous exercise, this shows that neither of the two conditions in the definition 3.1 of a linear map follows from the other. Note that the parenthetical remark in the statement is essential: if $\mathbb{C}$ were regarded as a vector space over $\mathbb{R}$, then conjugation would satisfy $\varphi(av) = a\varphi(v)$ for all real a and would indeed be linear.

Problem 3A.10. Prove or give a counterexample: If $q \in \mathcal{P}(\mathbb{R})$ and $T\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ is defined by $Tp = q \circ p$, then T is a linear map.
[The function T defined here differs from the function T defined in the last bullet point of 3.3 by the order of the functions in the compositions.]

Solution. The statement is false. Here is a counterexample.
Let $q \in \mathcal{P}(\mathbb{R})$ be given by $q(x) = x^2$, and let $T\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ be defined by $Tp = q \circ p$, so that

$$(Tp)(x) = \big(p(x)\big)^2 \qquad \text{for all } x \in \mathbb{R}.$$

(This really does map $\mathcal{P}(\mathbb{R})$ into $\mathcal{P}(\mathbb{R})$: the square of a polynomial is a polynomial.)
Homogeneity fails. Let p be the constant polynomial $p(x) = 1$ and take $\lambda = 2$. Then $(2p)(x) = 2$ for all x, so

$$\big(T(2p)\big)(x) = \big((2p)(x)\big)^2 = 4,$$

whereas

$$\big(2(Tp)\big)(x) = 2\big(p(x)\big)^2 = 2.$$

Thus $T(2p)$ is the constant polynomial 4 and $2(Tp)$ is the constant polynomial 2; these are different elements of $\mathcal{P}(\mathbb{R})$, so $T(2p) \neq 2(Tp)$ and the homogeneity condition of 3.1 fails.
Additivity fails too. With the same p, we have $p + p = 2p$, so

$$T(p + p) = T(2p) = 4 \neq 2 = Tp + Tp.$$

Hence T is not a linear map, and the assertion of the exercise is false.
Remark. Contrast this with the last bullet point of 3.3, where T is defined by $(Tp)(x) = p(q(x))$, that is, $Tp = p \circ q$. That map *is* linear, because the composition happens on the inside: for all $x \in \mathbb{R}$,

$$\big(T(p_1 + p_2)\big)(x) = (p_1 + p_2)(q(x)) = p_1(q(x)) + p_2(q(x)) = (Tp_1)(x) + (Tp_2)(x),$$

and $\big(T(\lambda p)\big)(x) = (\lambda p)(q(x)) = \lambda\, p(q(x)) = \lambda(Tp)(x)$. The failure in the present exercise comes from the fact that q is applied *after* p, so the vector-space operations of $\mathcal{P}(\mathbb{R})$ get fed into the nonlinear function q.
Indeed, the map $p \mapsto q \circ p$ is linear precisely when $q(x) = cx$ for some $c \in \mathbb{R}$. If $q(x) = cx$, then $q \circ p = cp$ for every p, which is linear. Conversely, suppose q is not of this form. If $q(0) \neq 0$, then T sends the zero polynomial to the nonzero constant polynomial $q(0)$, so T is not linear by 3.10. If $q(0) = 0$ but $q(x) \neq cx$ for every c, then in particular q is not the polynomial $q(1)x$, so there is a real number λ with $q(\lambda) \neq \lambda\, q(1)$; taking p to be the constant polynomial 1 gives $T(\lambda p) = q(\lambda) \neq \lambda\, q(1) = \lambda(Tp)$ as constant polynomials, so again T is not linear.

Problem 3A.11. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T is a scalar multiple of the identity if and only if $ST = TS$ for every $S \in \mathcal{L}(V)$.

Solution. If $V = \{0\}$ then $\mathcal{L}(V)$ contains only the zero map, which equals $0 \cdot I$, and both conditions hold trivially. So assume $V \neq \{0\}$.
The easy direction. Suppose $T = \lambda I$ for some $\lambda \in \mathbf{F}$. Let $S \in \mathcal{L}(V)$ and $v \in V$. Then

$$(ST)v = S(\lambda v) = \lambda(Sv) = (\lambda I)(Sv) = (TS)v,$$

where the second equality is the homogeneity of S. Since $v \in V$ was arbitrary, $ST = TS$.
The interesting direction. Now suppose $ST = TS$ for every $S \in \mathcal{L}(V)$. We first show that every nonzero vector of V is mapped by T into its own span.
Claim: if $v \in V$ and $v \neq 0$, then $Tv \in \operatorname{span}(v)$.
Suppose not. Then $Tv \notin \operatorname{span}(v)$, and hence the list v, Tv is linearly independent: if $av + b(Tv) = 0$ with $b \neq 0$ then $Tv = -(a/b)v \in \operatorname{span}(v)$, a contradiction, so $b = 0$; then $av = 0$ with $v \neq 0$ forces $a = 0$.

Because V is finite-dimensional, the linearly independent list v, Tv extends to a basis $v, Tv, u_3, \ldots, u_n$ of V by 2.32. By the linear map lemma 3.4 there exists $S \in \mathcal{L}(V)$ with

$$Sv = v, \qquad S(Tv) = 0, \qquad Su_k = 0 \text{ for } k = 3, \ldots, n.$$

Then

$$(ST)v = S(Tv) = 0, \qquad (TS)v = T(Sv) = Tv.$$

But $Tv \neq 0$, since Tv belongs to a linearly independent list. Hence $(ST)v \neq (TS)v$, contradicting $ST = TS$. This proves the claim.
Thus for each nonzero $v \in V$ there is a scalar $a_v \in \mathbf{F}$ with

$$Tv = a_v v.$$

(The scalar a_v is unique, because $v \neq 0$.)
It remains to show that a_v does not depend on v. Let $v, w \in V$ be nonzero.
Case 1: v, w is linearly dependent. Since both are nonzero, there is $c \in \mathbf{F}$ with $c \neq 0$ and $w = cv$. Then

$$a_w w = Tw = T(cv) = cTv = c\,a_v v = a_v(cv) = a_v w.$$

Because $w \neq 0$, this gives $a_w = a_v$.
Case 2: v, w is linearly independent. Then $v + w \neq 0$, and

$$a_{v+w}v + a_{v+w}w = a_{v+w}(v + w) = T(v + w) = Tv + Tw = a_v v + a_w w.$$

Rearranging,

$$(a_{v+w} - a_v)v + (a_{v+w} - a_w)w = 0.$$

Linear independence of v, w forces $a_{v+w} = a_v$ and $a_{v+w} = a_w$, so $a_v = a_w$.
In both cases $a_v = a_w$. Hence there is a single scalar $\lambda \in \mathbf{F}$ with $a_v = \lambda$ for every nonzero $v \in V$ (such a v exists because $V \neq \{0\}$). Therefore $Tv = \lambda v$ for every nonzero $v \in V$, and also $T0 = 0 = \lambda \cdot 0$ by 3.10. Thus $T = \lambda I$, so T is a scalar multiple of the identity.

Problem 3A.12. Suppose U is a subspace of V with $U \neq V$. Suppose $S \in \mathcal{L}(U, W)$ and $S \neq 0$ (which means that $Su \neq 0$ for some $u \in U$). Define $T\colon V \to W$ by

$$Tv = \begin{cases} Sv & \text{if } v \in U, \\ 0 & \text{if } v \in V \text{ and } v \notin U. \end{cases}$$

Prove that T is not a linear map on V.

Solution. First note that T is a well-defined function from V to W: each $v \in V$ satisfies exactly one of the conditions $v \in U$ and $v \notin U$, so exactly one clause of the definition applies.
Because $S \neq 0$, there exists $u \in U$ with

$$Su \neq 0.$$

Because $U \neq V$ and $U \subseteq V$, there exists

$$w \in V \text{ with } w \notin U.$$

Consider the vector $u + w \in V$. We claim $u + w \notin U$. Indeed, if $u + w \in U$, then since $u \in U$ and U is a subspace (hence closed under additive inverses and addition),

$$w = (u + w) - u \in U,$$

contradicting $w \notin U$. So $u + w \notin U$, and therefore, by the definition of T,

$$T(u + w) = 0.$$

On the other hand, $u \in U$ gives $Tu = Su$, and $w \notin U$ gives $Tw = 0$, so

$$Tu + Tw = Su + 0 = Su \neq 0.$$

Hence

$$T(u + w) = 0 \neq Su = Tu + Tw,$$

so T fails the additivity condition in the definition 3.1 of a linear map. Therefore T is not a linear map from V to W.

Remark. The point of this exercise is that "extend by zero" is not a legitimate way to extend a linear map from a proper subspace U to all of V: the set $V \setminus U$ is not closed under addition with elements of U, so the two clauses of the definition interfere with each other. The next exercise shows that when V is finite-dimensional a genuine linear extension does always exist; it is obtained by extending a basis of U to a basis of V and defining the extension to be 0 on the new basis vectors only, not on all of $V \setminus U$.

Problem 3A.13. Suppose V is finite-dimensional. Prove that every linear map on a subspace of V can be extended to a linear map on V. In other words, show that if U is a subspace of V and $S \in \mathcal{L}(U, W)$, then there exists $T \in \mathcal{L}(V, W)$ such that $Tu = Su$ for all $u \in U$.
[The result in this exercise is used in the proof of 3.125.]

Solution. Suppose U is a subspace of V and $S \in \mathcal{L}(U, W)$.
Since V is finite-dimensional, so is U, by 2.25; hence U has a basis

$$u_1, \ldots, u_m$$

by 2.31 (if $U = \{0\}$ this is the empty list, so $m = 0$). In particular $u_1, \ldots, u_m$ is linearly independent in U, hence also linearly independent as a list in V. Because V is finite-dimensional, 2.32 says this list extends to a basis

$$u_1, \ldots, u_m, v_1, \ldots, v_n$$

of V.
By the linear map lemma 3.4 applied to this basis of V and to the vectors $Su_1, \ldots, Su_m, 0, \ldots, 0 \in W$, there exists a (unique) linear map $T \in \mathcal{L}(V, W)$ such that

$$Tu_k = Su_k \quad \text{for } k = 1, \ldots, m, \qquad Tv_j = 0 \quad \text{for } j = 1, \ldots, n.$$

We check that T extends S. Let $u \in U$. Because $u_1, \ldots, u_m$ is a basis of U, there are scalars $c_1, \ldots, c_m \in \mathbf{F}$ with

$$u = c_1 u_1 + \cdots + c_m u_m.$$

Using the linearity of T and then the linearity of S,

$$\begin{aligned} Tu &= T(c_1 u_1 + \cdots + c_m u_m) \\ &= c_1(Tu_1) + \cdots + c_m(Tu_m) \\ &= c_1(Su_1) + \cdots + c_m(Su_m) \\ &= S(c_1 u_1 + \cdots + c_m u_m) \\ &= Su. \end{aligned}$$

(When $m = 0$ this reads $T0 = 0 = S0$, which holds by 3.10.)
Thus $T \in \mathcal{L}(V, W)$ and $Tu = Su$ for all $u \in U$, so S has been extended to a linear map on all of V, as desired.

Remark. The extension is generally far from unique: any choice of vectors $w_1, \ldots, w_n \in W$ in place of the zeros above yields, via 3.4, another linear extension of S. Compare exercise 3A.12, which shows that the naive extension "$Tv = 0$ for every $v \notin U$" never works when $U \neq V$ and $S \neq 0$.

Problem 3A.14. Suppose V is finite-dimensional with $\dim V > 0$, and suppose W is infinite-dimensional. Prove that $\mathcal{L}(V, W)$ is infinite-dimensional.

Solution. Recall that $\mathcal{L}(V, W)$ is a vector space with the operations of 3.5, by 3.6. Recall also that a vector space is called infinite-dimensional (2.13) when it is not finite-dimensional, that is, when no finite list of its vectors spans it.

We prove the result by exhibiting, for every positive integer m, a linearly independent list of length m in $\mathcal{L}(V, W)$.

Step 1: W contains linearly independent lists of every length.

We show by induction on k that for every nonnegative integer k there is a linearly independent list $w_1, \dots, w_k$ of vectors in W. For $k = 0$ the empty list is linearly independent. Suppose $w_1, \dots, w_k$ is a linearly independent list in W. Then $\operatorname{span}(w_1, \dots, w_k) \neq W$, for otherwise the finite list $w_1, \dots, w_k$ would span W, making W finite-dimensional (2.9) and contradicting our hypothesis. Choose

$$w_{k+1} \in W \setminus \operatorname{span}(w_1, \dots, w_k).$$

Then $w_1, \dots, w_{k+1}$ is linearly independent: suppose $a_1, \dots, a_{k+1} \in \mathbf{F}$ satisfy

$$a_1 w_1 + \cdots + a_k w_k + a_{k+1} w_{k+1} = 0.$$

If $a_{k+1} \neq 0$, then

$$w_{k+1} = -\frac{a_1}{a_{k+1}} w_1 - \cdots - \frac{a_k}{a_{k+1}} w_k \in \operatorname{span}(w_1, \dots, w_k),$$

a contradiction. Hence $a_{k+1} = 0$, and then $a_1 w_1 + \cdots + a_k w_k = 0$ forces $a_1 = \cdots = a_k = 0$ by the linear independence of $w_1, \dots, w_k$. This completes the induction.

Step 2: transporting these lists into $\mathcal{L}(V, W)$.

Since V is finite-dimensional with $\dim V > 0$, it has a basis $v_1, \dots, v_p$ with $p \geq 1$ (2.31). Fix a positive integer m, and by Step 1 choose a linearly independent list $w_1, \dots, w_m$ in W.

For each $j = 1, \dots, m$, the linear map lemma 3.4 provides a linear map $T_j \in \mathcal{L}(V, W)$ with

$$T_j v_1 = w_j \qquad \text{and} \qquad T_j v_k = 0 \text{ for } k = 2, \dots, p.$$

We claim $T_1, \dots, T_m$ is linearly independent in $\mathcal{L}(V, W)$. Suppose $a_1, \dots, a_m \in \mathbf{F}$ and

$$a_1 T_1 + \cdots + a_m T_m = 0,$$

where the 0 on the right is the zero map from V to W. Evaluating both sides at v_1 and using the definition 3.5 of addition and scalar multiplication in $\mathcal{L}(V, W)$, we get

$$a_1(T_1 v_1) + \cdots + a_m(T_m v_1) = 0, \qquad \text{that is,} \qquad a_1 w_1 + \cdots + a_m w_m = 0.$$

Since $w_1, \dots, w_m$ is linearly independent, $a_1 = \cdots = a_m = 0$. Hence $T_1, \dots, T_m$ is linearly independent.

Step 3: conclusion.

Suppose, for a contradiction, that $\mathcal{L}(V, W)$ were finite-dimensional. Then by 2.9 there would be a finite list $R_1, \dots, R_d$ spanning $\mathcal{L}(V, W)$. By Step 2 there is a linearly independent list in $\mathcal{L}(V, W)$ of length $d + 1$. But 2.22 says the length of a linearly independent list is at most the length of a spanning list, so $d + 1 \leq d$, a contradiction.

Therefore $\mathcal{L}(V, W)$ is not finite-dimensional; that is, $\mathcal{L}(V, W)$ is infinite-dimensional.

Remark. The hypothesis $\dim V > 0$ is needed: if $V = \{0\}$, then the only linear map from V to W is the zero map, so $\mathcal{L}(V, W) = \{0\}$ is finite-dimensional (of dimension 0) no matter how large W is. Intuitively, Step 2 says that a single nonzero vector v_1 of V suffices to encode an arbitrary vector of W inside a linear map, and W has room for arbitrarily long independent lists.

Problem 3A.15. Suppose $v_1, \dots, v_m$ is a linearly dependent list of vectors in V. Suppose also that $W \neq \{0\}$. Prove that there exist $w_1, \dots, w_m \in W$ such that no $T \in \mathcal{L}(V, W)$ satisfies $Tv_k = w_k$ for each $k = 1, \dots, m$.

Solution. Because $v_1, \dots, v_m$ is linearly dependent, there exist scalars $a_1, \dots, a_m \in \mathbf{F}$, not all zero, such that

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Fix an index j with $a_j \neq 0$. Because $W \neq \{0\}$, we may also fix a vector $w \in W$ with $w \neq 0$.
Now define the list $w_1, \dots, w_m$ in W by

$$w_j = w \quad \text{and} \quad w_k = 0 \ \text{ for } k \neq j.$$

We claim no $T \in \mathcal{L}(V, W)$ satisfies $Tv_k = w_k$ for each $k = 1, \dots, m$. Suppose, seeking a contradiction, that some $T \in \mathcal{L}(V, W)$ did satisfy these m equations. Applying T to the dependence relation above and using additivity and homogeneity of T, we get

$$T(a_1 v_1 + \cdots + a_m v_m) = a_1 T v_1 + \cdots + a_m T v_m = a_1 w_1 + \cdots + a_m w_m = a_j w,$$

where the last equality holds because $w_k = 0$ for every $k \neq j$. On the other hand, $a_1 v_1 + \cdots + a_m v_m = 0$, and every linear map takes 0 to 0 (by 3.10), so the left side equals 0. Hence

$$a_j w = 0.$$

Since $a_j \neq 0$, we may multiply by $1/a_j$ to conclude $w = 0$, contradicting the choice of $w \neq 0$.
Thus no $T \in \mathcal{L}(V, W)$ satisfies $Tv_k = w_k$ for each $k = 1, \dots, m$, as desired.
Remark: this exercise is the exact converse of the linear map lemma 3.4 in the following sense. If $v_1, \dots, v_m$ is a basis of V, then 3.4 says that every choice of $w_1, \dots, w_m \in W$ is achieved by some (indeed exactly one) $T \in \mathcal{L}(V, W)$. This exercise shows that if the list $v_1, \dots, v_m$ is linearly dependent, then — provided W contains a nonzero vector — some choice of $w_1, \dots, w_m$ is achieved by no $T \in \mathcal{L}(V, W)$ at all. So linear independence of $v_1, \dots, v_m$ is exactly what is needed for arbitrary prescription of the values Tv_k to be possible.

Problem 3A.16. Suppose V is finite-dimensional with $\dim V > 1$. Prove that there exist $S, T \in \mathcal{L}(V)$ such that $ST \neq TS$.

Solution. Let $n = \dim V$; by hypothesis $n \geq 2$ and $n < \infty$. Choose a basis $v_1, \dots, v_n$ of V.
By the linear map lemma 3.4, a linear map on V may be defined by specifying its values arbitrarily on the basis $v_1, \dots, v_n$. Use this to define $S, T \in \mathcal{L}(V)$ by

$$\begin{aligned} Sv_1 &= v_1, \quad Sv_k = 0 \quad \text{for } k = 2, \dots, n, \\ Tv_1 &= v_2, \quad Tv_k = 0 \quad \text{for } k = 2, \dots, n. \end{aligned}$$

Note that these definitions make sense precisely because $n \geq 2$: we need a second basis vector v_2 for T to aim at.
Now evaluate both compositions at v_1. First,

$$(ST)v_1 = S(Tv_1) = Sv_2 = 0,$$

since $2 \leq n$ means v_2 is one of the basis vectors sent to 0 by S. Second,

$$(TS)v_1 = T(Sv_1) = Tv_1 = v_2.$$

Since $v_1, \dots, v_n$ is a basis, it is in particular linearly independent, so $v_2 \neq 0$. Therefore $(ST)v_1 = 0 \neq v_2 = (TS)v_1$, and the two operators ST and TS disagree at the vector v_1. Hence $ST \neq TS$, as desired.
Remark: the hypothesis $\dim V > 1$ cannot be dropped. If $\dim V \leq 1$ then every operator on V is a scalar multiple of the identity (this is trivial when $V = \{0\}$, and is Exercise 7 of this section when

$\dim V = 1$), and any two scalar multiples of the identity commute. Thus $\mathcal{L}(V)$ is a commutative algebra exactly when $\dim V \leq 1$.

Problem 3A.17. Suppose V is finite-dimensional. Show that the only two-sided ideals of $\mathcal{L}(V)$ are $\{0\}$ and $\mathcal{L}(V)$.
[A subspace $\mathcal{E}$ of $\mathcal{L}(V)$ is called a **two-sided ideal** of $\mathcal{L}(V)$ if $TE \in \mathcal{E}$ and $ET \in \mathcal{E}$ for all $E \in \mathcal{E}$ and all $T \in \mathcal{L}(V)$.]

Solution. First note that $\{0\}$ and $\mathcal{L}(V)$ really are two-sided ideals: both are subspaces of $\mathcal{L}(V)$, the products $T0$ and $0T$ equal 0 for every $T \in \mathcal{L}(V)$, and $\mathcal{L}(V)$ trivially contains all products of its elements. So the content of the exercise is that there are no others.
Suppose then that $\mathcal{E}$ is a two-sided ideal of $\mathcal{L}(V)$ with $\mathcal{E} \neq \{0\}$. We show $\mathcal{E} = \mathcal{L}(V)$. The strategy is to produce the identity operator I inside $\mathcal{E}$; once $I \in \mathcal{E}$, every $T \in \mathcal{L}(V)$ satisfies $T = TI$ by the identity property in 3.8, and $TI \in \mathcal{E}$ by the ideal property applied to the element I of $\mathcal{E}$ and the operator T; so $\mathcal{L}(V) \subseteq \mathcal{E}$ and hence $\mathcal{E} = \mathcal{L}(V)$.
Because $\mathcal{E} \neq \{0\}$, there is some $E \in \mathcal{E}$ with $E \neq 0$. Since E is not the zero operator, there exists $u \in V$ with

$$Eu \neq 0.$$

In particular $V \neq \{0\}$, so $n := \dim V \geq 1$; and $n < \infty$ because V is finite-dimensional. Fix a basis $v_1, \ldots, v_n$ of V.
Two families of auxiliary operators. For each $j \in \{1, \ldots, n\}$, the linear map lemma 3.4 lets us define $A_j \in \mathcal{L}(V)$ by its values on the basis:

$$A_j v_j = u \quad \text{and} \quad A_j v_i = 0 \ \text{ for } i \neq j.$$

Next, since $Eu \neq 0$, the one-term list Eu is linearly independent in V. By 2.32 it extends to a basis

$$u_1, u_2, \ldots, u_n \quad \text{of } V, \qquad \text{where } u_1 = Eu.$$

(The extended basis has n terms because every basis of V has length $\dim V = n$.) For each $j \in \{1, \ldots, n\}$, use 3.4 again to define $B_j \in \mathcal{L}(V)$ by

$$B_j u_1 = v_j \quad \text{and} \quad B_j u_i = 0 \ \text{ for } i = 2, \ldots, n.$$

The products lie in $\mathcal{E}$. Fix j. Since $E \in \mathcal{E}$ and $A_j \in \mathcal{L}(V)$, the ideal property gives $EA_j \in \mathcal{E}$. Since $EA_j \in \mathcal{E}$ and $B_j \in \mathcal{L}(V)$, the ideal property gives $B_j(EA_j) \in \mathcal{E}$. By associativity (3.8) we may write this operator unambiguously as $B_j E A_j$. Thus

$$B_j E A_j \in \mathcal{E} \qquad \text{for each } j = 1, \ldots, n.$$

Computing these products on the basis. Fix j and let $i \in \{1, \ldots, n\}$. If $i \neq j$ then $A_j v_i = 0$, so

$$(B_j E A_j) v_i = B_j E(0) = 0,$$

using 3.10 (linear maps take 0 to 0). If $i = j$ then

$$(B_j E A_j) v_j = B_j E(u) = B_j(Eu) = B_j u_1 = v_j.$$

Assembling the identity. Because $\mathcal{E}$ is a subspace of $\mathcal{L}(V)$, it is closed under addition, so the finite sum

$$R := B_1 E A_1 + B_2 E A_2 + \cdots + B_n E A_n$$

belongs to $\mathcal{E}$. (Finite-dimensionality of V is what makes this legitimate: the sum has only finitely many terms. It was also used earlier, to produce the finite basis $v_1, \ldots, v_n$ and to invoke 2.32.) For each $i \in \{1, \ldots, n\}$, the computation above gives

$$R v_i = \sum_{j=1}^{n} (B_j E A_j) v_i = (B_i E A_i) v_i = v_i,$$

since all terms with $j \neq i$ vanish. So the operators R and I agree on the basis $v_1, \dots, v_n$; by the uniqueness half of the linear map lemma 3.4 (or simply because every $v \in V$ is a linear combination of the v_i and both maps are linear), $R = I$.
Therefore $I \in \mathcal{E}$, and by the first paragraph $\mathcal{E} = \mathcal{L}(V)$.
Conclusion: every two-sided ideal of $\mathcal{L}(V)$ is either $\{0\}$ or all of $\mathcal{L}(V)$. (If $V = \{0\}$ then $\mathcal{L}(V) = \{0\}$ and the two listed ideals coincide, so the statement holds vacuously in that degenerate case.)
Remark: the hypothesis that V is finite-dimensional is essential. If V is infinite-dimensional, then the set of operators $T \in \mathcal{L}(V)$ whose range is finite-dimensional is a two-sided ideal — it is a subspace, and composing a finite-rank operator with any operator on either side again has finite-dimensional range — and it is neither $\{0\}$ (it contains, for instance, any nonzero operator with one-dimensional range) nor all of $\mathcal{L}(V)$ (it does not contain I). The argument above breaks down precisely at the assembling step, where the sum $\sum_j B_j E A_j$ would have to be infinite.

3.2 Exercises 3B

Problem 3B.1. Give an example of a linear map T with $\dim \operatorname{null} T = 3$ and $\dim \operatorname{range} T = 2$.

Solution. The fundamental theorem of linear maps (3.21) tells us where to look: if T is defined on a finite-dimensional space V, then $\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = 3 + 2 = 5$. So take $V = \mathbf{F}^5$ and $W = \mathbf{F}^2$, and define $T \in \mathcal{L}(\mathbf{F}^5, \mathbf{F}^2)$ by

$$T(z_1, z_2, z_3, z_4, z_5) = (z_1, z_2).$$

This map is linear: each coordinate of the output is a coordinate of the input, so additivity and homogeneity are immediate from the coordinatewise definitions of addition and scalar multiplication on $\mathbf{F}^5$ and $\mathbf{F}^2$.
Null space. We have $T(z_1, \dots, z_5) = (0, 0)$ if and only if $z_1 = z_2 = 0$. Hence

$$\operatorname{null} T = \{(0, 0, z_3, z_4, z_5) : z_3, z_4, z_5 \in \mathbf{F}\}.$$

The list $(0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1)$ spans this set and is linearly independent (a linear combination with coefficients z_3, z_4, z_5 equals $(0, 0, z_3, z_4, z_5)$, which is 0 only when all three scalars are 0). Thus it is a basis of $\operatorname{null} T$ and $\dim \operatorname{null} T = 3$.
Range. For any $(w_1, w_2) \in \mathbf{F}^2$ we have $T(w_1, w_2, 0, 0, 0) = (w_1, w_2)$, so $\operatorname{range} T = \mathbf{F}^2$ and therefore $\dim \operatorname{range} T = 2$.
As a check, $\dim \mathbf{F}^5 = 5 = 3 + 2 = \dim \operatorname{null} T + \dim \operatorname{range} T$, in agreement with 3.21.

Problem 3B.2. Suppose $S, T \in \mathcal{L}(V)$ are such that $\operatorname{range} S \subseteq \operatorname{null} T$. Prove that $(ST)^2 = 0$.

Solution. Recall that $(ST)^2$ denotes the composition $ST \circ ST = STST$, and that the zero element of $\mathcal{L}(V)$ is the map sending every vector to 0. So we must show $(STST)v = 0$ for every $v \in V$.
Let $v \in V$. Put $u = S(Tv)$. Then $u \in \operatorname{range} S$, and since $\operatorname{range} S \subseteq \operatorname{null} T$ we get $u \in \operatorname{null} T$, i.e.

$$T\big(S(Tv)\big) = Tu = 0.$$

Therefore

$$(ST)^2 v = S\Big(T\big(S(Tv)\big)\Big) = S(0) = 0,$$

where the last equality holds because every linear map sends 0 to 0: indeed $S(0) = S(0+0) = S(0)+S(0)$ by additivity, and adding $-S(0)$ to both sides gives $S(0) = 0$.
Since $v \in V$ was arbitrary, $(ST)^2$ is the zero map, that is, $(ST)^2 = 0$.

Problem 3B.3. Suppose $v_1, \dots, v_m$ is a list of vectors in V. Define $T \in \mathcal{L}(\mathbf{F}^m, V)$ by

$$T(z_1, \dots, z_m) = z_1 v_1 + \cdots + z_m v_m.$$

(a) What property of T corresponds to $v_1, \dots, v_m$ spanning V?
(b) What property of T corresponds to the list $v_1, \dots, v_m$ being linearly independent?

Solution. (a) The list $v_1, \dots, v_m$ spans V if and only if T is surjective.
To see this, compute the range of T. By definition,

$$\operatorname{range} T = \{z_1 v_1 + \cdots + z_m v_m : z_1, \dots, z_m \in \mathbf{F}\},$$

which is exactly the set of all linear combinations of $v_1, \dots, v_m$, i.e. $\operatorname{range} T = \operatorname{span}(v_1, \dots, v_m)$. Now T is surjective means $\operatorname{range} T = V$ (3.19), and $v_1, \dots, v_m$ spans V means $\operatorname{span}(v_1, \dots, v_m) = V$. Since the two sets in question are the same set, the two conditions are equivalent.
(b) The list $v_1, \dots, v_m$ is linearly independent if and only if T is injective.
Compute the null space:

$$\operatorname{null} T = \{(z_1, \dots, z_m) \in \mathbf{F}^m : z_1 v_1 + \cdots + z_m v_m = 0\}.$$

By definition, $v_1, \dots, v_m$ is linearly independent exactly when the only choice of scalars $z_1, \dots, z_m \in \mathbf{F}$ making $z_1 v_1 + \cdots + z_m v_m = 0$ is $z_1 = \cdots = z_m = 0$; that is, exactly when $\operatorname{null} T = \{0\}$. By 3.15, $\operatorname{null} T = \{0\}$ if and only if T is injective. Hence linear independence of the list is equivalent to injectivity of T.
Remark. Combining the two parts: $v_1, \dots, v_m$ is a basis of V if and only if T is both injective and surjective, i.e. if and only if T is an isomorphism from $\mathbf{F}^m$ onto V.

Problem 3B.4. Show that $\{T \in \mathcal{L}(\mathbf{R}^5, \mathbf{R}^4) : \dim \operatorname{null} T > 2\}$ is not a subspace of $\mathcal{L}(\mathbf{R}^5, \mathbf{R}^4)$.

Solution. Write $E = \{T \in \mathcal{L}(\mathbf{R}^5, \mathbf{R}^4) : \dim \operatorname{null} T > 2\}$. We show E is not closed under addition, which is enough to conclude that E is not a subspace.
Define $S, T \in \mathcal{L}(\mathbf{R}^5, \mathbf{R}^4)$ by

$$S(x_1, x_2, x_3, x_4, x_5) = (x_1, x_2, 0, 0), \qquad T(x_1, x_2, x_3, x_4, x_5) = (0, 0, x_3, x_4).$$

Both are linear, since each output coordinate is either 0 or one of the input coordinates.
S lies in E: we have $S(x_1, \dots, x_5) = 0$ if and only if $x_1 = x_2 = 0$, so

$$\operatorname{null} S = \{(0, 0, x_3, x_4, x_5) : x_3, x_4, x_5 \in \mathbf{R}\},$$

which has basis $(0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1)$ and hence $\dim \operatorname{null} S = 3 > 2$.
T lies in E: similarly $T(x_1, \dots, x_5) = 0$ if and only if $x_3 = x_4 = 0$, so

$$\operatorname{null} T = \{(x_1, x_2, 0, 0, x_5) : x_1, x_2, x_5 \in \mathbf{R}\},$$

which has basis $(1,0,0,0,0), (0,1,0,0,0), (0,0,0,0,1)$ and hence $\dim \operatorname{null} T = 3 > 2$.
But $S + T$ is given by

$$(S + T)(x_1, x_2, x_3, x_4, x_5) = (x_1, x_2, x_3, x_4),$$

so $(S + T)(x_1, \dots, x_5) = 0$ if and only if $x_1 = x_2 = x_3 = x_4 = 0$. Thus

$$\operatorname{null}(S + T) = \{(0, 0, 0, 0, x_5) : x_5 \in \mathbf{R}\},$$

which is spanned by the single nonzero vector $(0,0,0,0,1)$, so $\dim \operatorname{null}(S + T) = 1$. Since 1 is not greater than 2, we have $S + T \notin E$.
Hence E contains S and T but not $S + T$, so E is not closed under addition and therefore is not a subspace of $\mathcal{L}(\mathbf{R}^5, \mathbf{R}^4)$.

Remark. E does contain the zero map (whose null space is all of $\mathbf{R}^5$, of dimension 5) and E is closed under scalar multiplication (for $\lambda \neq 0$, $\operatorname{null}(\lambda T) = \operatorname{null} T$; and $0T = 0 \in E$). So failure of closure under addition is the only way this set can fail to be a subspace, and the example above exhibits that failure.

Problem 3B.5. Give an example of $T \in \mathcal{L}(\mathbf{R}^4)$ such that $\operatorname{range} T = \operatorname{null} T$.

Solution. Define $T \in \mathcal{L}(\mathbf{R}^4)$ by

$$T(x_1, x_2, x_3, x_4) = (0, 0, x_1, x_2).$$

This map is linear, since each output coordinate is either the constant 0 or one of the input coordinates.
Null space. We have $T(x_1, x_2, x_3, x_4) = (0, 0, 0, 0)$ if and only if $x_1 = x_2 = 0$. Hence

$$\operatorname{null} T = \{(0, 0, x_3, x_4) : x_3, x_4 \in \mathbf{R}\}.$$

Range. As (x_1, x_2, x_3, x_4) ranges over $\mathbf{R}^4$, the vector $(0, 0, x_1, x_2)$ ranges over all vectors whose first two coordinates are 0: given any $a, b \in \mathbf{R}$ we have $T(a, b, 0, 0) = (0, 0, a, b)$. Hence

$$\operatorname{range} T = \{(0, 0, a, b) : a, b \in \mathbf{R}\}.$$

The two sets described are literally the same subset of $\mathbf{R}^4$, so $\operatorname{range} T = \operatorname{null} T$, as desired.
Sanity check with 3.21: $\dim \operatorname{null} T = \dim \operatorname{range} T = 2$, and indeed $\dim \mathbf{R}^4 = 4 = 2 + 2$. This is why such a T can exist on $\mathbf{R}^4$ but, as the next exercise shows, not on $\mathbf{R}^5$.

Problem 3B.6. Prove that there does not exist $T \in \mathcal{L}(\mathbf{R}^5)$ such that $\operatorname{range} T = \operatorname{null} T$.

Solution. Suppose, for contradiction, that $T \in \mathcal{L}(\mathbf{R}^5)$ satisfies $\operatorname{range} T = \operatorname{null} T$.
The space $\mathbf{R}^5$ is finite-dimensional with $\dim \mathbf{R}^5 = 5$, so the fundamental theorem of linear maps (3.21) applies and gives

$$5 = \dim \mathbf{R}^5 = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Since $\operatorname{range} T$ and $\operatorname{null} T$ are the same subspace of $\mathbf{R}^5$, they have the same dimension; write $k = \dim \operatorname{null} T = \dim \operatorname{range} T$, a nonnegative integer. The displayed equation becomes

$$5 = k + k = 2k.$$

Thus 5 would be an even integer, which is false; equivalently, $k = 5/2$ is not an integer, contradicting the fact that the dimension of a finite-dimensional vector space is a nonnegative integer.
This contradiction shows no such T exists.
Remark. The same argument shows more generally that if V is finite-dimensional with $\dim V$ odd, then no $T \in \mathcal{L}(V)$ satisfies $\operatorname{range} T = \operatorname{null} T$. When $\dim V$ is even, say $\dim V = 2k$, such a map does exist: take a basis $v_1, \ldots, v_{2k}$ of V and use the linear map lemma (3.4) to define T by $Tv_j = v_{k+j}$ for $j = 1, \ldots, k$ and $Tv_j = 0$ for $j = k+1, \ldots, 2k$; then both $\operatorname{range} T$ and $\operatorname{null} T$ equal $\operatorname{span}(v_{k+1}, \ldots, v_{2k})$.
Exercise 3B.5 is the case $\dim V = 4$.

Problem 3B.7. Suppose V and W are finite-dimensional with $2 \leq \dim V \leq \dim W$. Show that $\{T \in \mathcal{L}(V, W) : T \text{ is not injective}\}$ is not a subspace of $\mathcal{L}(V, W)$.

Solution. Write $E = \{T \in \mathcal{L}(V, W) : T \text{ is not injective}\}$. We produce $S, T \in E$ with $S + T \notin E$; since a subspace must be closed under addition, this shows E is not a subspace of $\mathcal{L}(V, W)$.
Set $n = \dim V$ and $m = \dim W$, so $2 \leq n \leq m$. Choose a basis $v_1, \ldots, v_n$ of V and a basis $w_1, \ldots, w_m$ of W. Since $n \leq m$, the vectors $w_1, \ldots, w_n$ are among the basis vectors of W, and in particular $w_1, \ldots, w_n$ is a linearly independent list in W.

By the linear map lemma (3.4), there are unique $S, T \in \mathcal{L}(V, W)$ satisfying

$$\begin{aligned} Sv_1 &= 0, \quad & Sv_j &= w_j \quad (j = 2, \dots, n), \\ Tv_1 &= w_1, \quad & Tv_j &= 0 \quad (j = 2, \dots, n). \end{aligned}$$

S is not injective. The vector v_1 is nonzero (it belongs to a basis, and a list containing 0 is linearly dependent) and $Sv_1 = 0$, so $\operatorname{null} S \neq \{0\}$; by 3.15, S is not injective. Hence $S \in E$.
T is not injective. Here we use the hypothesis $n = \dim V \geq 2$, which guarantees that the basis vector v_2 exists. It is nonzero and $Tv_2 = 0$, so $\operatorname{null} T \neq \{0\}$ and T is not injective by 3.15. Hence $T \in E$.
$S + T$ is injective. For each $j \in \{1, \dots, n\}$ we have

$$(S + T)v_j = Sv_j + Tv_j = w_j,$$

since exactly one of the two summands is w_j and the other is 0. Now suppose $v \in \operatorname{null}(S + T)$. Write $v = a_1 v_1 + \cdots + a_n v_n$ with $a_1, \dots, a_n \in \mathbf{F}$, which is possible because $v_1, \dots, v_n$ is a basis of V. Then, by linearity,

$$0 = (S + T)v = a_1(S + T)v_1 + \cdots + a_n(S + T)v_n = a_1 w_1 + \cdots + a_n w_n.$$

Because $w_1, \dots, w_n$ is linearly independent, $a_1 = \cdots = a_n = 0$, hence $v = 0$. Thus $\operatorname{null}(S + T) = \{0\}$, so $S + T$ is injective by 3.15, and therefore $S + T \notin E$.
We have exhibited $S, T \in E$ whose sum is not in E, so E is not closed under addition and consequently is not a subspace of $\mathcal{L}(V, W)$.
Remark on the hypotheses. The condition $\dim V \leq \dim W$ is needed: if $\dim V > \dim W$ then by 3.22 no linear map from V to W is injective, so E would be all of $\mathcal{L}(V, W)$, which is a subspace of itself. The condition $\dim V \geq 2$ is also needed: if $\dim V = 1$, then a linear map T on V fails to be injective exactly when $T = 0$ (for v_1 a basis of V, T is determined by Tv_1, and $\operatorname{null} T \neq \{0\}$ forces $Tv_1 = 0$), so $E = \{0\}$, again a subspace. If $\dim V = 0$ then every map is injective and E is empty, hence not a subspace, but the argument above would not apply.

Problem 3B.8. Suppose V and W are finite-dimensional with $\dim V \geq \dim W \geq 2$. Show that $\{T \in \mathcal{L}(V, W) : T \text{ is not surjective}\}$ is not a subspace of $\mathcal{L}(V, W)$.

Solution. Write $n = \dim V$ and $m = \dim W$, so that $n \geq m \geq 2$. Let $v_1, \dots, v_n$ be a basis of V and let $w_1, \dots, w_m$ be a basis of W. Denote by

$$E = \{T \in \mathcal{L}(V, W) : T \text{ is not surjective}\}.$$

We exhibit two elements of E whose sum is not in E; since a subspace must be closed under addition, this shows E is not a subspace of $\mathcal{L}(V, W)$.
By the linear map lemma (3.4) there is a unique $S \in \mathcal{L}(V, W)$ with

$$Sv_k = \begin{cases} w_k & \text{if } 1 \leq k \leq m - 1, \\ 0 & \text{if } m \leq k \leq n, \end{cases}$$

and a unique $T \in \mathcal{L}(V, W)$ with

$$Tv_k = \begin{cases} w_m & \text{if } k = m, \\ 0 & \text{otherwise.} \end{cases}$$

Note that both definitions make sense: $m - 1 \geq 1$ because $m \geq 2$, and the index m is a legitimate index for the basis of V because $m \leq n$.
Neither S nor T is surjective. Indeed, every $v \in V$ has the form $v = a_1 v_1 + \cdots + a_n v_n$ for some scalars $a_1, \dots, a_n$, so by linearity $Sv = a_1 w_1 + \cdots + a_{m-1} w_{m-1}$. Hence

$$\operatorname{range} S \subseteq \operatorname{span}(w_1, \dots, w_{m-1}),$$

and the reverse inclusion holds too because $w_k = Sv_k$ for $1 \le k \le m-1$, so in fact $\operatorname{range} S = \operatorname{span}(w_1, \dots, w_{m-1})$. As $w_1, \dots, w_{m-1}$ is linearly independent (being a sublist of a basis), this subspace has dimension $m-1 < m = \dim W$; since equal subspaces have equal dimensions, $\operatorname{range} S \neq W$. The same computation gives

$$\operatorname{range} T = \operatorname{span}(w_m),$$

of dimension $1 < 2 \le m = \dim W$, so $\operatorname{range} T \neq W$. Thus $S \in E$ and $T \in E$.
Now consider $S+T$. For $1 \le k \le m-1$ we have $(S+T)v_k = w_k + 0 = w_k$, and $(S+T)v_m = 0 + w_m = w_m$. Hence $w_1, \dots, w_m$ all lie in $\operatorname{range}(S+T)$. Since $\operatorname{range}(S+T)$ is a subspace of W (by 3.18) containing the basis $w_1, \dots, w_m$ of W, we get $\operatorname{range}(S+T) = W$. Therefore $S+T$ is surjective, i.e. $S+T \notin E$.
So E is not closed under addition, and consequently E is not a subspace of $\mathcal{L}(V, W)$.
Both hypotheses are genuinely used. If $\dim W = 1$, then a map $T \in \mathcal{L}(V, W)$ fails to be surjective exactly when $\operatorname{range} T = \{0\}$, i.e. exactly when $T = 0$; so in that case $E = \{0\}$, which is a subspace. If $\dim V < \dim W$, then no linear map from V to W is surjective (by 3.24), so $E = \mathcal{L}(V, W)$, again a subspace.

Problem 3B.9. Suppose $T \in \mathcal{L}(V, W)$ is injective and $v_1, \dots, v_n$ is linearly independent in V. Prove that $Tv_1, \dots, Tv_n$ is linearly independent in W.

Solution. Suppose $a_1, \dots, a_n \in \mathbf{F}$ satisfy

$$a_1 Tv_1 + \cdots + a_n Tv_n = 0.$$

Because T is linear (it commutes with scalar multiplication and preserves sums), the left side equals $T(a_1 v_1 + \cdots + a_n v_n)$. Thus

$$T(a_1 v_1 + \cdots + a_n v_n) = 0,$$

which says that $a_1 v_1 + \cdots + a_n v_n \in \operatorname{null} T$.
Since T is injective, 3.15 gives $\operatorname{null} T = \{0\}$. Hence

$$a_1 v_1 + \cdots + a_n v_n = 0.$$

Now the linear independence of $v_1, \dots, v_n$ forces $a_1 = \cdots = a_n = 0$.
We have shown that the only choice of scalars making $a_1 Tv_1 + \cdots + a_n Tv_n = 0$ is $a_1 = \cdots = a_n = 0$. By definition, $Tv_1, \dots, Tv_n$ is linearly independent in W.

Problem 3B.10. Suppose $v_1, \dots, v_n$ spans V and $T \in \mathcal{L}(V, W)$. Show that $Tv_1, \dots, Tv_n$ spans $\operatorname{range} T$.

Solution. We prove the two inclusions

$$\operatorname{span}(Tv_1, \dots, Tv_n) \subseteq \operatorname{range} T \quad \text{and} \quad \operatorname{range} T \subseteq \operatorname{span}(Tv_1, \dots, Tv_n).$$

First inclusion. Each Tv_k lies in $\operatorname{range} T$ by definition of the range. By 3.18, $\operatorname{range} T$ is a subspace of W, hence is closed under scalar multiplication and addition; therefore every linear combination $a_1 Tv_1 + \cdots + a_n Tv_n$ lies in $\operatorname{range} T$. Since the span of a list is precisely the set of its linear combinations, $\operatorname{span}(Tv_1, \dots, Tv_n) \subseteq \operatorname{range} T$.
Second inclusion. Let $w \in \operatorname{range} T$. Then $w = Tv$ for some $v \in V$. Because $v_1, \dots, v_n$ spans V, there are scalars $a_1, \dots, a_n \in \mathbf{F}$ with

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Applying T and using linearity,

$$w = Tv = T(a_1 v_1 + \cdots + a_n v_n) = a_1 Tv_1 + \cdots + a_n Tv_n,$$

which lies in $\operatorname{span}(Tv_1, \dots, Tv_n)$.

The two inclusions give $\operatorname{span}(Tv_1, \dots, Tv_n) = \operatorname{range} T$; that is, the list $Tv_1, \dots, Tv_n$ spans $\operatorname{range} T$. Note in particular that if V is finite-dimensional then $\operatorname{range} T$ has a finite spanning list, so $\operatorname{range} T$ is finite-dimensional and $\dim \operatorname{range} T \le \dim V$.

Problem 3B.11. Suppose that V is finite-dimensional and that $T \in \mathcal{L}(V, W)$. Prove that there exists a subspace U of V such that

$$U \cap \operatorname{null} T = \{0\} \quad \text{and} \quad \operatorname{range} T = \{Tu : u \in U\}.$$

Solution. By 3.13, $\operatorname{null} T$ is a subspace of V; since V is finite-dimensional, $\operatorname{null} T$ is finite-dimensional by 2.25, so it has a basis $u_1, \dots, u_m$ (by 2.31). This list is linearly independent in V, so by 2.32 it extends to a basis

$$u_1, \dots, u_m, w_1, \dots, w_n$$

of V. (Either m or n may be 0; the argument below is unaffected.) Define

$$U = \operatorname{span}(w_1, \dots, w_n),$$

which is a subspace of V because a span is always a subspace.
Step 1: $U \cap \operatorname{null} T = \{0\}$.
Certainly 0 lies in both U and $\operatorname{null} T$. Conversely, suppose $v \in U \cap \operatorname{null} T$. Since $v \in U$ there are scalars $b_1, \dots, b_n$ with $v = b_1 w_1 + \cdots + b_n w_n$, and since $v \in \operatorname{null} T$ and $u_1, \dots, u_m$ is a basis of $\operatorname{null} T$ there are scalars $a_1, \dots, a_m$ with $v = a_1 u_1 + \cdots + a_m u_m$. Subtracting,

$$a_1 u_1 + \cdots + a_m u_m - b_1 w_1 - \cdots - b_n w_n = 0.$$

The list $u_1, \dots, u_m, w_1, \dots, w_n$ is linearly independent, so every coefficient here vanishes: $a_1 = \cdots = a_m = 0$ and $b_1 = \cdots = b_n = 0$. Hence $v = 0$, proving $U \cap \operatorname{null} T = \{0\}$.
Step 2: $\operatorname{range} T = \{Tu : u \in U\}$.
The inclusion $\{Tu : u \in U\} \subseteq \operatorname{range} T$ is immediate, since every Tu with $u \in U \subseteq V$ belongs to $\operatorname{range} T$.
For the reverse inclusion, let $x \in \operatorname{range} T$, say $x = Tv$ with $v \in V$. Write v in the basis of V:

$$v = a_1 u_1 + \cdots + a_m u_m + b_1 w_1 + \cdots + b_n w_n$$

for some scalars a_j, b_k. Put $u = b_1 w_1 + \cdots + b_n w_n$, so $u \in U$. Since $u_1, \dots, u_m \in \operatorname{null} T$, linearity of T gives

$$x = Tv = a_1 Tu_1 + \cdots + a_m Tu_m + Tu = 0 + \cdots + 0 + Tu = Tu.$$

Thus $x \in \{Tu : u \in U\}$.
Combining the two inclusions, $\operatorname{range} T = \{Tu : u \in U\}$, and with Step 1 this completes the proof.
Remark. The two conclusions say exactly that the restriction $T|_U$ is injective (its null space is $U \cap \operatorname{null} T = \{0\}$, by 3.15) and has the same range as T. Counting dimensions, the list $w_1, \dots, w_n$ is linearly independent (a sublist of a basis) and spans U, hence is a basis of U and $\dim U = n$; since $n = \dim V - m = \dim V - \dim \operatorname{null} T$, the fundamental theorem of linear maps (3.21) gives $\dim U = \dim \operatorname{range} T$, which is consistent with $T|_U$ being an injective map of U onto $\operatorname{range} T$.

Problem 3B.12. Suppose T is a linear map from $\mathbf{F}^4$ to $\mathbf{F}^2$ such that

$$\operatorname{null} T = \{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_1 = 5x_2 \text{ and } x_3 = 7x_4\}.$$

Prove that T is surjective.

Solution. The strategy is to compute $\dim \operatorname{null} T$, apply the fundamental theorem of linear maps (3.21) to get $\dim \operatorname{range} T = 2$, and then conclude $\operatorname{range} T = \mathbf{F}^2$ from 2.39.

Step 1: $\dim \operatorname{null} T = 2$.
A vector $(x_1, x_2, x_3, x_4) \in \mathbf{F}^4$ lies in $\operatorname{null} T$ precisely when $x_1 = 5x_2$ and $x_3 = 7x_4$, that is, precisely when it has the form

$$(5x_2, x_2, 7x_4, x_4) = x_2(5, 1, 0, 0) + x_4(0, 0, 7, 1)$$

for some $x_2, x_4 \in \mathbf{F}$. Conversely every such combination satisfies the two equations. Hence

$$\operatorname{null} T = \operatorname{span}\big((5, 1, 0, 0), (0, 0, 7, 1)\big).$$

The list $(5, 1, 0, 0), (0, 0, 7, 1)$ is linearly independent: if

$$a(5, 1, 0, 0) + b(0, 0, 7, 1) = (5a, a, 7b, b) = (0, 0, 0, 0),$$

then looking at the second coordinate gives $a = 0$ and at the fourth gives $b = 0$. So this list is a basis of $\operatorname{null} T$ and $\dim \operatorname{null} T = 2$.
Step 2: $\dim \operatorname{range} T = 2$.
Since $\mathbf{F}^4$ is finite-dimensional with $\dim \mathbf{F}^4 = 4$, the fundamental theorem of linear maps (3.21) gives

$$4 = \dim \mathbf{F}^4 = \dim \operatorname{null} T + \dim \operatorname{range} T = 2 + \dim \operatorname{range} T,$$

so $\dim \operatorname{range} T = 2$.
Step 3: T is surjective.
By 3.18, $\operatorname{range} T$ is a subspace of $\mathbf{F}^2$, and $\dim \mathbf{F}^2 = 2 = \dim \operatorname{range} T$. A subspace of full dimension equals the whole space (2.39), so $\operatorname{range} T = \mathbf{F}^2$. That is, T is surjective.

Problem 3B.13. Suppose U is a three-dimensional subspace of $\mathbf{R}^8$ and that T is a linear map from $\mathbf{R}^8$ to $\mathbf{R}^5$ such that $\operatorname{null} T = U$. Prove that T is surjective.

Solution. The space $\mathbf{R}^8$ is finite-dimensional with $\dim \mathbf{R}^8 = 8$, and by hypothesis

$$\dim \operatorname{null} T = \dim U = 3.$$

The fundamental theorem of linear maps (3.21) applied to $T \in \mathcal{L}(\mathbf{R}^8, \mathbf{R}^5)$ gives

$$8 = \dim \mathbf{R}^8 = \dim \operatorname{null} T + \dim \operatorname{range} T = 3 + \dim \operatorname{range} T,$$

so

$$\dim \operatorname{range} T = 5.$$

By 3.18, $\operatorname{range} T$ is a subspace of $\mathbf{R}^5$, and $\dim \mathbf{R}^5 = 5 = \dim \operatorname{range} T$. Since a subspace whose dimension equals the dimension of the whole (finite-dimensional) space is the whole space (2.39), we conclude

$$\operatorname{range} T = \mathbf{R}^5.$$

Hence T is surjective.

Problem 3B.14. Prove that there does not exist a linear map from $\mathbf{F}^5$ to $\mathbf{F}^2$ whose null space equals $\{(x_1, x_2, x_3, x_4, x_5) \in \mathbf{F}^5 : x_1 = 3x_2 \text{ and } x_3 = x_4 = x_5\}$.

Solution. Write

$$N = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbf{F}^5 : x_1 = 3x_2 \text{ and } x_3 = x_4 = x_5\}.$$

Step 1: $\dim N = 2$.
A vector of $\mathbf{F}^5$ lies in N exactly when it has the form

$$(3x_2, x_2, x_3, x_3, x_3) = x_2(3, 1, 0, 0, 0) + x_3(0, 0, 1, 1, 1)$$

for some $x_2, x_3 \in \mathbf{F}$; conversely every such combination satisfies $x_1 = 3x_2$ and $x_3 = x_4 = x_5$. Hence

$$N = \operatorname{span}\big((3,1,0,0,0),(0,0,1,1,1)\big).$$

This spanning list is linearly independent: if

$$a(3,1,0,0,0) + b(0,0,1,1,1) = (3a, a, b, b, b) = 0,$$

then the second coordinate gives $a = 0$ and the third gives $b = 0$. So $(3,1,0,0,0),(0,0,1,1,1)$ is a basis of N and $\dim N = 2$.
Step 2: no linear map can have N as its null space.
Suppose, for contradiction, that $T \in \mathcal{L}(\mathbf{F}^5, \mathbf{F}^2)$ satisfies $\operatorname{null} T = N$. Since $\mathbf{F}^5$ is finite-dimensional with $\dim \mathbf{F}^5 = 5$, the fundamental theorem of linear maps (3.21) gives

$$5 = \dim \mathbf{F}^5 = \dim \operatorname{null} T + \dim \operatorname{range} T = 2 + \dim \operatorname{range} T,$$

so $\dim \operatorname{range} T = 3$.
But $\operatorname{range} T$ is a subspace of $\mathbf{F}^2$ (by 3.18), and $\mathbf{F}^2$ is finite-dimensional of dimension 2, so 2.37 yields

$$\dim \operatorname{range} T \le \dim \mathbf{F}^2 = 2.$$

This contradicts $\dim \operatorname{range} T = 3$.
Therefore no such linear map exists.
Remark. The same counting shows more generally that a subspace $N \subseteq \mathbf{F}^5$ can be the null space of some $T \in \mathcal{L}(\mathbf{F}^5, \mathbf{F}^2)$ only when $\dim N \ge 3$. Conversely, if $\dim N = k \ge 3$ then such a T does exist: extend a basis $n_1, \dots, n_k$ of N to a basis $n_1, \dots, n_k, w_1, \dots, w_{5-k}$ of $\mathbf{F}^5$, and since $5 - k \le 2$ we may pick a linearly independent list $e_1, \dots, e_{5-k}$ in $\mathbf{F}^2$. The linear map lemma (3.4) supplies T with $Tn_i = 0$ and $Tw_j = e_j$, and one checks from the linear independence of $e_1, \dots, e_{5-k}$ that $\operatorname{null} T = N$.

Problem 3B.15. Suppose there exists a linear map on V whose null space and range are both finite-dimensional. Prove that V is finite-dimensional.

Solution. Let T be such a map. (The argument below uses nothing about the target space, so we may as well suppose $T \in \mathcal{L}(V, W)$ for some vector space W; the case of an operator $T \in \mathcal{L}(V)$ is the case $W = V$.) By hypothesis $\operatorname{null} T$ and $\operatorname{range} T$ are both finite-dimensional. We produce a finite list that spans V, which by definition makes V finite-dimensional.
Let $u_1, \dots, u_m$ be a basis of $\operatorname{null} T$ and let $w_1, \dots, w_n$ be a basis of $\operatorname{range} T$ (these exist because both spaces are finite-dimensional, by 2.31). For each $k \in \{1, \dots, n\}$ we have $w_k \in \operatorname{range} T$, so we may choose $v_k \in V$ with

$$Tv_k = w_k.$$

We claim that the list

$$u_1, \dots, u_m, v_1, \dots, v_n$$

spans V. To see this, let $v \in V$. Then $Tv \in \operatorname{range} T$, so there exist $b_1, \dots, b_n \in \mathbf{F}$ with

$$Tv = b_1 w_1 + \cdots + b_n w_n = b_1 Tv_1 + \cdots + b_n Tv_n = T(b_1 v_1 + \cdots + b_n v_n),$$

where the last equality uses the linearity of T. Hence

$$T\big(v - (b_1 v_1 + \cdots + b_n v_n)\big) = 0,$$

so $v - (b_1 v_1 + \cdots + b_n v_n) \in \operatorname{null} T$. Because $u_1, \dots, u_m$ spans $\operatorname{null} T$, there exist $a_1, \dots, a_m \in \mathbf{F}$ with

$$v - (b_1 v_1 + \cdots + b_n v_n) = a_1 u_1 + \cdots + a_m u_m.$$

Therefore

$$v = a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_n v_n,$$

which exhibits v as a linear combination of $u_1, \dots, u_m, v_1, \dots, v_n$.
Thus the finite list $u_1, \dots, u_m, v_1, \dots, v_n$ spans V, and hence V is finite-dimensional.
Note that the conclusion is exactly what one expects from the fundamental theorem of linear maps (3.21): once V is known to be finite-dimensional, that theorem gives $\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$, and indeed the spanning list constructed above has length $m + n$. (In fact that list is a basis of V: if $a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n = 0$, applying T gives $b_1w_1 + \cdots + b_nw_n = 0$, so all the b_k are 0 because $w_1, \dots, w_n$ is linearly independent, and then all the a_j are 0 because $u_1, \dots, u_m$ is linearly independent.)

Problem 3B.16. Suppose V and W are both finite-dimensional. Prove that there exists an injective linear map from V to W if and only if $\dim V \le \dim W$.

Solution. First suppose there exists an injective $T \in \mathcal{L}(V, W)$. Because T is injective, $\operatorname{null} T = \{0\}$ (by 3.15), so $\dim \operatorname{null} T = 0$. The fundamental theorem of linear maps (3.21) now gives

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = \dim \operatorname{range} T.$$

Since $\operatorname{range} T$ is a subspace of W (by 3.18) and W is finite-dimensional, we have $\dim \operatorname{range} T \le \dim W$ (by 2.37). Combining the two displays gives $\dim V \le \dim W$.
Conversely, suppose $\dim V \le \dim W$. Write $n = \dim V$ and $m = \dim W$, so $n \le m$. Let $v_1, \dots, v_n$ be a basis of V and let $w_1, \dots, w_m$ be a basis of W (by 2.31). Because $n \le m$, the vectors $w_1, \dots, w_n$ are among $w_1, \dots, w_m$; in particular $w_1, \dots, w_n$ is a linearly independent list in W, being a sublist of a linearly independent list.
By the linear map lemma (3.4) there exists $T \in \mathcal{L}(V, W)$ with

$$Tv_k = w_k \quad \text{for each } k = 1, \dots, n.$$

We claim T is injective. By 3.15 it suffices to show $\operatorname{null} T = \{0\}$. Suppose $v \in \operatorname{null} T$. Because $v_1, \dots, v_n$ is a basis of V, we may write $v = c_1v_1 + \cdots + c_nv_n$ with $c_1, \dots, c_n \in \mathbf{F}$. Then

$$0 = Tv = c_1Tv_1 + \cdots + c_nTv_n = c_1w_1 + \cdots + c_nw_n.$$

Because $w_1, \dots, w_n$ is linearly independent, $c_1 = \cdots = c_n = 0$, and hence $v = 0$. Thus $\operatorname{null} T = \{0\}$ and T is injective, as desired.
(If $n = 0$, so that $V = \{0\}$, the lists above are empty and the map T produced is the zero map, which is indeed injective on $\{0\}$; the argument as written covers this degenerate case.)

Problem 3B.17. Suppose V and W are both finite-dimensional. Prove that there exists a surjective linear map from V onto W if and only if $\dim V \ge \dim W$.

Solution. First suppose there exists a surjective $T \in \mathcal{L}(V, W)$. Surjectivity means $\operatorname{range} T = W$ (by 3.19). The fundamental theorem of linear maps (3.21) gives

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = \dim \operatorname{null} T + \dim W \ge \dim W,$$

where the last inequality holds because $\dim \operatorname{null} T \ge 0$. Thus $\dim V \ge \dim W$.
Conversely, suppose $\dim V \ge \dim W$. Write $n = \dim V$ and $m = \dim W$, so $n \ge m$. Let $v_1, \dots, v_n$ be a basis of V and let $w_1, \dots, w_m$ be a basis of W (by 2.31). By the linear map lemma (3.4) there exists $T \in \mathcal{L}(V, W)$ with

$$Tv_k = \begin{cases} w_k & \text{if } 1 \le k \le m, \\ 0 & \text{if } m < k \le n. \end{cases}$$

(This is legitimate precisely because $m \le n$: we have prescribed a value in W for each of the n basis vectors of V.)

We claim T is surjective. Indeed, $\operatorname{range} T$ is a subspace of W (by 3.18) that contains $w_1, \dots, w_m$, so it contains $\operatorname{span}(w_1, \dots, w_m) = W$ (using 2.6, that the span is the smallest subspace containing the list). Hence $\operatorname{range} T = W$, so T is surjective.
Alternatively, one can see surjectivity directly: given $w \in W$, write $w = c_1 w_1 + \cdots + c_m w_m$; then $w = T(c_1 v_1 + \cdots + c_m v_m)$.
(If $m = 0$, so that $W = \{0\}$, the map T is the zero map, whose range is $\{0\} = W$; so T is surjective and the argument covers this degenerate case as well.)

Problem 3B.18. Suppose V and W are finite-dimensional and that U is a subspace of V. Prove that there exists $T \in \mathcal{L}(V, W)$ such that $\operatorname{null} T = U$ if and only if $\dim U \geq \dim V - \dim W$.

Solution. First suppose there exists $T \in \mathcal{L}(V, W)$ with $\operatorname{null} T = U$. The fundamental theorem of linear maps (3.21) gives

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = \dim U + \dim \operatorname{range} T.$$

Because $\operatorname{range} T$ is a subspace of the finite-dimensional space W (by 3.18), we have $\dim \operatorname{range} T \leq \dim W$ (by 2.37). Hence

$$\dim V \leq \dim U + \dim W,$$

which rearranges to $\dim U \geq \dim V - \dim W$, as desired.
Conversely, suppose $\dim U \geq \dim V - \dim W$. Let $u_1, \dots, u_j$ be a basis of U, where $j = \dim U$; note U is finite-dimensional because it is a subspace of the finite-dimensional space V (by 2.25). Extend this linearly independent list to a basis

$$u_1, \dots, u_j, v_1, \dots, v_n$$

of V (by 2.32). Then $\dim V = j + n$, so

$$n = \dim V - \dim U \leq \dim W,$$

where the inequality is exactly the hypothesis rearranged. Let $w_1, \dots, w_m$ be a basis of W; since $n \leq m$, the list $w_1, \dots, w_n$ is a well-defined sublist of it and hence is linearly independent in W.
By the linear map lemma (3.4) there exists $T \in \mathcal{L}(V, W)$ with

$$Tu_i = 0 \quad \text{for } i = 1, \dots, j, \qquad Tv_k = w_k \quad \text{for } k = 1, \dots, n.$$

We show $\operatorname{null} T = U$.
First, $U \subseteq \operatorname{null} T$: each u_i lies in $\operatorname{null} T$, and $\operatorname{null} T$ is a subspace of V (by 3.13), so it contains $\operatorname{span}(u_1, \dots, u_j) = U$.
Second, $\operatorname{null} T \subseteq U$: suppose $v \in \operatorname{null} T$ and write

$$v = a_1 u_1 + \cdots + a_j u_j + b_1 v_1 + \cdots + b_n v_n$$

with all coefficients in $\mathbf{F}$, which is possible because $u_1, \dots, u_j, v_1, \dots, v_n$ is a basis of V. Applying T gives

$$0 = Tv = b_1 w_1 + \cdots + b_n w_n.$$

Because $w_1, \dots, w_n$ is linearly independent, $b_1 = \cdots = b_n = 0$. Hence $v = a_1 u_1 + \cdots + a_j u_j \in U$.
Thus $\operatorname{null} T = U$, completing the proof.

Problem 3B.19. Suppose W is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that T is injective if and only if there exists $S \in \mathcal{L}(W, V)$ such that ST is the identity operator on V.

Solution. First suppose there exists $S \in \mathcal{L}(W, V)$ with $ST = I$, the identity operator on V. If $u, v \in V$ satisfy $Tu = Tv$, then applying S gives

$$u = (ST)u = S(Tu) = S(Tv) = (ST)v = v.$$

Hence T is injective. (This direction uses no finite-dimensionality at all.)
Now suppose T is injective; we construct S. Let $w_1, \dots, w_n$ be a basis of $\operatorname{range} T$; this exists because $\operatorname{range} T$ is a subspace of the finite-dimensional space W (by 3.18 and 2.25). Extend it to a basis

$$w_1, \dots, w_n, w_{n+1}, \dots, w_m$$

of W (by 2.32; here is where finite-dimensionality of W is used). For each $k \in \{1, \dots, n\}$, the vector w_k lies in $\operatorname{range} T$, so we may choose $v_k \in V$ with

$$Tv_k = w_k.$$

By the linear map lemma (3.4) there exists $S \in \mathcal{L}(W, V)$ with

$$Sw_k = \begin{cases} v_k & \text{if } 1 \le k \le n, \\ 0 & \text{if } n < k \le m. \end{cases}$$

We claim $ST = I$. Let $v \in V$. Then $Tv \in \operatorname{range} T$, so there exist $c_1, \dots, c_n \in \mathbf{F}$ with

$$Tv = c_1 w_1 + \cdots + c_n w_n.$$

Applying S and using linearity gives

$$(ST)v = S(Tv) = c_1 Sw_1 + \cdots + c_n Sw_n = c_1 v_1 + \cdots + c_n v_n.$$

On the other hand,

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 Tv_1 + \cdots + c_n Tv_n = c_1 w_1 + \cdots + c_n w_n = Tv.$$

Because T is injective, the equation $T(c_1 v_1 + \cdots + c_n v_n) = Tv$ forces

$$c_1 v_1 + \cdots + c_n v_n = v.$$

Combining the last two displays gives $(ST)v = v$. Since $v \in V$ was arbitrary, ST is the identity operator on V, as desired.
Remark: it is worth noting where each hypothesis enters. Finite-dimensionality of W was used only to extend a basis of $\operatorname{range} T$ to a basis of W, i.e. to split off a complement of $\operatorname{range} T$ inside W. (Injectivity of T in fact forces V to be finite-dimensional as well: an injective linear map sends a linearly independent list in V to a linearly independent list in W, so no linearly independent list in V has length greater than $\dim W$; so we may choose a linearly independent list $u_1, \dots, u_k$ in V of maximal length. That list spans V: if some $v \in V$ were not in $\operatorname{span}(u_1, \dots, u_k)$, then $u_1, \dots, u_k, v$ would be linearly independent, contradicting maximality. Indeed, if $u_1, \dots, u_k, v$ were linearly dependent, the linear dependence lemma (2.19) would supply a term lying in the span of the terms preceding it; that term cannot be v, since $v \notin \operatorname{span}(u_1, \dots, u_k)$, and it cannot be some u_j, since $u_j \in \operatorname{span}(u_1, \dots, u_{j-1})$ would make $u_1, \dots, u_k$ linearly dependent. Hence V is spanned by a finite list, so V is finite-dimensional. We did not need this fact, and the proof above deliberately avoids it.)

Problem 3B.20. Suppose W is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that T is surjective if and only if there exists $S \in \mathcal{L}(W, V)$ such that TS is the identity operator on W.

Solution. First suppose there exists $S \in \mathcal{L}(W, V)$ with $TS = I$, the identity operator on W. Let $w \in W$. Then

$$w = (TS)w = T(Sw),$$

so $w \in \operatorname{range} T$. Since $w \in W$ was arbitrary, $\operatorname{range} T = W$, i.e. T is surjective. (This direction uses no finite-dimensionality.)
Now suppose T is surjective; we construct S. Because W is finite-dimensional it has a basis $w_1, \dots, w_m$

(by 2.31). Since range $T = W$, for each $k \in \{1, \dots, m\}$ we may choose $v_k \in V$ with

$$Tv_k = w_k.$$

By the linear map lemma (3.4) there exists $S \in \mathcal{L}(W, V)$ with

$$Sw_k = v_k \quad \text{for each } k = 1, \dots, m.$$

(Finite-dimensionality of W is used exactly here, to have a finite basis on which to prescribe S.)
We claim $TS = I$. For each k,

$$(TS)w_k = T(Sw_k) = Tv_k = w_k,$$

so the linear maps TS and I (both in $\mathcal{L}(W)$) agree on the basis $w_1, \dots, w_m$. Hence they are equal: explicitly, given $w \in W$, write $w = c_1 w_1 + \cdots + c_m w_m$ and compute

$$(TS)w = c_1(TS)w_1 + \cdots + c_m(TS)w_m = c_1 w_1 + \cdots + c_m w_m = w.$$

(One can also invoke the uniqueness assertion of 3.4: a linear map on W is determined by its values on a basis.)
Thus TS is the identity operator on W, as desired.

Problem 3B.21. Suppose V is finite-dimensional, $T \in \mathcal{L}(V, W)$, and U is a subspace of W. Prove that $\{v \in V : Tv \in U\}$ is a subspace of V and

$$\dim\{v \in V : Tv \in U\} = \dim \operatorname{null} T + \dim(U \cap \operatorname{range} T).$$

Solution. Write

$$E = \{v \in V : Tv \in U\}.$$

Step 1: E is a subspace of V. Because T is linear, $T0 = 0$ (by 3.10), and $0 \in U$ because U is a subspace of W; hence $0 \in E$. If $v_1, v_2 \in E$, then $Tv_1, Tv_2 \in U$, so

$$T(v_1 + v_2) = Tv_1 + Tv_2 \in U$$

because U is closed under addition; hence $v_1 + v_2 \in E$, so E is closed under addition. If $v \in E$ and $\lambda \in \mathbf{F}$, then

$$T(\lambda v) = \lambda Tv \in U$$

because U is closed under scalar multiplication; hence $\lambda v \in E$, so E is closed under scalar multiplication. Since E contains 0 and is closed under addition and scalar multiplication, E is a subspace of V (by 1.34).
Step 2: the dimension formula. Because V is finite-dimensional and E is a subspace of V, the space E is finite-dimensional (by 2.25). Consider the restriction of T to E,

$$T|_E : E \to W, \qquad (T|_E)v = Tv \quad \text{for } v \in E.$$

This is a linear map from E to W, since it satisfies additivity and homogeneity as inherited from T; that is, $T|_E \in \mathcal{L}(E, W)$. Applying the fundamental theorem of linear maps (3.21) to $T|_E$, whose domain E is finite-dimensional, gives

$$\dim E = \dim \operatorname{null}(T|_E) + \dim \operatorname{range}(T|_E).$$

It remains to identify the two spaces on the right.
First we show $\operatorname{null}(T|_E) = \operatorname{null} T$. If $v \in \operatorname{null}(T|_E)$, then $v \in E \subseteq V$ and $Tv = 0$, so $v \in \operatorname{null} T$. Conversely, if $v \in \operatorname{null} T$, then $Tv = 0 \in U$, so $v \in E$, and therefore $(T|_E)v = Tv = 0$, giving $v \in \operatorname{null}(T|_E)$. Hence the two null spaces are equal, and in particular

$$\dim \operatorname{null}(T|_E) = \dim \operatorname{null} T.$$

Next we show $\operatorname{range}(T|_E) = U \cap \operatorname{range} T$. If $w \in \operatorname{range}(T|_E)$, then $w = Tv$ for some $v \in E$; then $w \in \operatorname{range} T$ by definition, and $w = Tv \in U$ by the definition of E, so $w \in U \cap \operatorname{range} T$. Conversely, suppose $w \in U \cap \operatorname{range} T$. Since $w \in \operatorname{range} T$, there exists $v \in V$ with $Tv = w$. Since $w \in U$, we have $Tv \in U$, so $v \in E$; hence $w = (T|_E)v \in \operatorname{range}(T|_E)$. Hence the two sets are equal, and in particular

$$\dim \operatorname{range}(T|_E) = \dim(U \cap \operatorname{range} T).$$

Substituting these two identifications into the displayed application of 3.21 yields

$$\dim\{v \in V : Tv \in U\} = \dim E = \dim \operatorname{null} T + \dim(U \cap \operatorname{range} T),$$

as desired.
Remark: taking $U = \{0\}$ recovers $E = \operatorname{null} T$ and the trivial identity $\dim \operatorname{null} T = \dim \operatorname{null} T$, while taking $U = W$ recovers $E = V$ and the fundamental theorem of linear maps itself, since then $U \cap \operatorname{range} T = \operatorname{range} T$.

Problem 3B.22. Suppose U and V are finite-dimensional vector spaces and $S \in \mathcal{L}(V, W)$ and $T \in \mathcal{L}(U, V)$. Prove that

$$\dim \operatorname{null} ST \leq \dim \operatorname{null} S + \dim \operatorname{null} T.$$

Solution. Here $ST \in \mathcal{L}(U, W)$, so $\operatorname{null} ST$ is a subspace of U (by 3.13). Since U is finite-dimensional, $\operatorname{null} ST$ is finite-dimensional (by 2.25).
The idea is to apply the fundamental theorem of linear maps to T restricted to $\operatorname{null} ST$. Define

$$R = T|_{\operatorname{null} ST} \in \mathcal{L}(\operatorname{null} ST,\ V),$$

so $Ru = Tu$ for every $u \in \operatorname{null} ST$. This is a linear map because T is linear and $\operatorname{null} ST$ is a subspace of the domain U of T.
Step 1: $\operatorname{null} R \subseteq \operatorname{null} T$. If $u \in \operatorname{null} R$ then $u \in \operatorname{null} ST$ and $Tu = Ru = 0$, so $u \in \operatorname{null} T$. Both $\operatorname{null} R$ and $\operatorname{null} T$ are subspaces of U (by 3.13), and $\operatorname{null} T$ is finite-dimensional (by 2.25, as U is finite-dimensional); since $\operatorname{null} R$ is contained in $\operatorname{null} T$, it is a subspace of the vector space $\operatorname{null} T$, so 2.37 gives

$$\dim \operatorname{null} R \leq \dim \operatorname{null} T.$$

Step 2: $\operatorname{range} R \subseteq \operatorname{null} S$. Suppose $v \in \operatorname{range} R$, say $v = Ru = Tu$ with $u \in \operatorname{null} ST$. Then

$$Sv = S(Tu) = (ST)u = 0,$$

so $v \in \operatorname{null} S$. Both $\operatorname{range} R$ and $\operatorname{null} S$ are subspaces of the finite-dimensional space V (by 3.18 and 3.13), hence both are finite-dimensional (by 2.25); as $\operatorname{range} R \subseteq \operatorname{null} S$, it is a subspace of $\operatorname{null} S$, so 2.37 gives

$$\dim \operatorname{range} R \leq \dim \operatorname{null} S.$$

Step 3: combine. Applying the fundamental theorem of linear maps (3.21) to R, whose domain $\operatorname{null} ST$ is finite-dimensional, we get

$$\dim \operatorname{null} ST = \dim \operatorname{null} R + \dim \operatorname{range} R.$$

Substituting the two inequalities from Steps 1 and 2 yields

$$\dim \operatorname{null} ST \leq \dim \operatorname{null} T + \dim \operatorname{null} S,$$

as desired.
Intuitively: a vector killed by ST is either killed already by T, or else T sends it into $\operatorname{null} S$; the first possibility accounts for at most $\dim \operatorname{null} T$ dimensions and the second for at most $\dim \operatorname{null} S$.

Problem 3B.23. Suppose U and V are finite-dimensional vector spaces and $S \in \mathcal{L}(V, W)$ and $T \in \mathcal{L}(U, V)$. Prove that

$$\dim \operatorname{range} ST \leq \min\{\dim \operatorname{range} S, \dim \operatorname{range} T\}.$$

Solution. Again $ST \in \mathcal{L}(U, W)$. Because U and V are finite-dimensional, the fundamental theorem of linear maps (3.21) tells us that $\operatorname{range} T$, $\operatorname{range} S$, and $\operatorname{range} ST$ are all finite-dimensional, so all three dimensions appearing above make sense. We prove the two inequalities separately.
Claim 1: $\dim \operatorname{range} ST \leq \dim \operatorname{range} S$. Suppose $w \in \operatorname{range} ST$. Then $w = (ST)u = S(Tu)$ for some $u \in U$, and $Tu \in V$, so $w \in \operatorname{range} S$. Thus

$$\operatorname{range} ST \subseteq \operatorname{range} S.$$

Both are subspaces of W (by 3.18), and $\operatorname{range} S$ is finite-dimensional, so 2.37 applied inside the vector space $\operatorname{range} S$ gives $\dim \operatorname{range} ST \leq \dim \operatorname{range} S$.
Claim 2: $\dim \operatorname{range} ST \leq \dim \operatorname{range} T$. Restrict S to $\operatorname{range} T$: define

$$R = S|_{\operatorname{range} T} \in \mathcal{L}(\operatorname{range} T,\ W),$$

which makes sense because $\operatorname{range} T$ is a subspace of V (by 3.18), and which is linear because S is. We compute the range of R:

$$\operatorname{range} R = \{Sv : v \in \operatorname{range} T\} = \{S(Tu) : u \in U\} = \{(ST)u : u \in U\} = \operatorname{range} ST,$$

where the second equality holds because v ranges over $\operatorname{range} T$ exactly when $v = Tu$ for some $u \in U$. Now $\operatorname{range} T$ is finite-dimensional, so applying 3.21 to R gives

$$\dim \operatorname{range} T = \dim \operatorname{null} R + \dim \operatorname{range} R \geq \dim \operatorname{range} R = \dim \operatorname{range} ST.$$

Since $\dim \operatorname{range} ST$ is less than or equal to each of $\dim \operatorname{range} S$ and $\dim \operatorname{range} T$, it is less than or equal to their minimum, as desired.

Problem 3B.24. (a) Suppose $\dim V = 5$ and $S, T \in \mathcal{L}(V)$ are such that $ST = 0$. Prove that $\dim \operatorname{range} TS \leq 2$.
(b) Give an example of $S, T \in \mathcal{L}(\mathbf{F}^5)$ with $ST = 0$ and $\dim \operatorname{range} TS = 2$.

Solution. (a) First we turn the hypothesis $ST = 0$ into a statement about ranges and null spaces. Let $v \in V$. Then $S(Tv) = (ST)v = 0$, so $Tv \in \operatorname{null} S$. As v was arbitrary,

$$\operatorname{range} T \subseteq \operatorname{null} S.$$

Both $\operatorname{range} T$ and $\operatorname{null} S$ are subspaces of the five-dimensional space V (by 3.18 and 3.13), so both are finite-dimensional (by 2.25); since $\operatorname{range} T \subseteq \operatorname{null} S$, it is a subspace of $\operatorname{null} S$ and 2.37 gives $\dim \operatorname{range} T \leq \dim \operatorname{null} S$. Combining this with the fundamental theorem of linear maps (3.21) applied to S, namely $\dim \operatorname{null} S = 5 - \dim \operatorname{range} S$, we obtain

$$\dim \operatorname{range} S + \dim \operatorname{range} T \leq 5.$$

Call this the *rank bound*.
Now apply Exercise 23 of this section to the composition TS (with S as the inner map and T as the outer map; its hypotheses hold because V is finite-dimensional and both maps act on V):

$$\dim \operatorname{range} TS \leq \min\{\dim \operatorname{range} T, \dim \operatorname{range} S\}.$$

If we had $\min\{\dim \operatorname{range} T, \dim \operatorname{range} S\} \geq 3$, then both $\dim \operatorname{range} S \geq 3$ and $\dim \operatorname{range} T \geq 3$, giving $\dim \operatorname{range} S + \dim \operatorname{range} T \geq 6$, which contradicts the rank bound. Hence the minimum is at most 2, and therefore

$$\dim \operatorname{range} TS \leq 2,$$

as desired.
(b) Define $S, T \in \mathcal{L}(\mathbf{F}^5)$ by

$$S(x_1, x_2, x_3, x_4, x_5) = (x_1, x_2, 0, 0, 0), \qquad T(x_1, x_2, x_3, x_4, x_5) = (0, 0, x_1, x_2, 0).$$

Both maps are linear, since each coordinate of the output is a linear function of the input coordinates. For every $(x_1, \dots, x_5) \in \mathbf{F}^5$ we have

$$(ST)(x_1, \dots, x_5) = S(0, 0, x_1, x_2, 0) = (0, 0, 0, 0, 0),$$

because the first two coordinates of $T(x_1, \dots, x_5)$ are 0. Thus $ST = 0$.
On the other hand,

$$(TS)(x_1, \dots, x_5) = T(x_1, x_2, 0, 0, 0) = (0, 0, x_1, x_2, 0),$$

so

$$\operatorname{range} TS = \{(0, 0, x_1, x_2, 0) : x_1, x_2 \in \mathbf{F}\} = \operatorname{span}\big((0, 0, 1, 0, 0),\ (0, 0, 0, 1, 0)\big).$$

The two spanning vectors listed above are linearly independent, so $\dim \operatorname{range} TS = 2$.
(Consistently with the rank bound, here $\dim \operatorname{range} S = \dim \operatorname{range} T = 2$ and $2 + 2 \le 5$.)

Problem 3B.25. Suppose that W is finite-dimensional and $S, T \in \mathcal{L}(V, W)$. Prove that $\operatorname{null} S \subseteq \operatorname{null} T$ if and only if there exists $E \in \mathcal{L}(W)$ such that $T = ES$.

Solution. The easy direction. Suppose there exists $E \in \mathcal{L}(W)$ with $T = ES$. Let $v \in \operatorname{null} S$. Then

$$Tv = (ES)v = E(Sv) = E0 = 0,$$

using $E0 = 0$ (by 3.10). Hence $v \in \operatorname{null} T$. Thus $\operatorname{null} S \subseteq \operatorname{null} T$. Note that this direction does not use the finite-dimensionality of W.
The other direction. Now suppose $\operatorname{null} S \subseteq \operatorname{null} T$; we must manufacture $E \in \mathcal{L}(W)$ with $ES = T$. The point is that the hypothesis makes the assignment sending Sv to Tv unambiguous on $\operatorname{range} S$; finite-dimensionality of W then lets us extend that assignment to all of W.
By 3.18, $\operatorname{range} S$ is a subspace of W, hence is finite-dimensional (by 2.25). Let

$$w_1, \dots, w_m$$

be a basis of $\operatorname{range} S$ (the empty list if $\operatorname{range} S = \{0\}$). For each $k \in \{1, \dots, m\}$, since $w_k \in \operatorname{range} S$ we may choose $v_k \in V$ with

$$Sv_k = w_k.$$

The list $w_1, \dots, w_m$ is linearly independent in W, so by 2.32 it extends to a basis

$$w_1, \dots, w_m, w_{m+1}, \dots, w_n$$

of W. By the linear map lemma (3.4) there is a unique $E \in \mathcal{L}(W)$ with

$$Ew_k = Tv_k \quad \text{for } k = 1, \dots, m, \qquad Ew_k = 0 \quad \text{for } k = m+1, \dots, n.$$

We claim $ES = T$. Let $v \in V$. Since $Sv \in \operatorname{range} S$ and $w_1, \dots, w_m$ is a basis of $\operatorname{range} S$, there are scalars $a_1, \dots, a_m \in \mathbf{F}$ with

$$Sv = a_1 w_1 + \cdots + a_m w_m.$$

Put $u = a_1 v_1 + \cdots + a_m v_m \in V$. Then by linearity of S,

$$Su = a_1 Sv_1 + \cdots + a_m Sv_m = a_1 w_1 + \cdots + a_m w_m = Sv,$$

so $S(v - u) = 0$, that is, $v - u \in \operatorname{null} S$. By hypothesis $v - u \in \operatorname{null} T$, so $Tv = Tu$. Therefore

$$(ES)v = E(Sv) = E(a_1 w_1 + \cdots + a_m w_m) = a_1 Ew_1 + \cdots + a_m Ew_m = a_1 Tv_1 + \cdots + a_m Tv_m = Tu = Tv.$$

Since $v \in V$ was arbitrary, $ES = T$, as desired.
(If $\operatorname{range} S = \{0\}$ the argument still works, read with empty sums: then $S = 0$, so $\operatorname{null} S = V \subseteq \operatorname{null} T$ forces $T = 0$, and $E = 0$ works.)

Problem 3B.26. Suppose that V is finite-dimensional and $S, T \in \mathcal{L}(V, W)$. Prove that $\operatorname{range} S \subseteq \operatorname{range} T$ if and only if there exists $E \in \mathcal{L}(V)$ such that $S = TE$.

Solution. The easy direction. Suppose there exists $E \in \mathcal{L}(V)$ with $S = TE$. Let $w \in \operatorname{range} S$, say $w = Sv$ with $v \in V$. Then

$$w = Sv = (TE)v = T(Ev),$$

and $Ev \in V$, so $w \in \operatorname{range} T$. Hence $\operatorname{range} S \subseteq \operatorname{range} T$. This direction does not use finite-dimensionality of V.
The other direction. Suppose $\operatorname{range} S \subseteq \operatorname{range} T$. Because V is finite-dimensional, it has a basis

$$v_1, \ldots, v_n.$$

For each $k \in \{1, \ldots, n\}$ we have $Sv_k \in \operatorname{range} S \subseteq \operatorname{range} T$, so we may choose $u_k \in V$ with

$$Tu_k = Sv_k.$$

(This is exactly where the containment of ranges is used: it guarantees that each Sv_k is hit by T.)
By the linear map lemma (3.4), there exists a (unique) linear map $E \in \mathcal{L}(V)$ with

$$Ev_k = u_k \quad \text{for } k = 1, \ldots, n.$$

We claim $TE = S$. Both TE and S are linear maps from V to W, and for each k,

$$(TE)v_k = T(Ev_k) = Tu_k = Sv_k.$$

Thus TE and S agree on the basis $v_1, \ldots, v_n$. Now let $v \in V$ be arbitrary and write $v = a_1v_1 + \cdots + a_nv_n$ with $a_1, \ldots, a_n \in \mathbf{F}$. By linearity,

$$(TE)v = a_1(TE)v_1 + \cdots + a_n(TE)v_n = a_1Sv_1 + \cdots + a_nSv_n = Sv.$$

Hence $S = TE$, as desired.
(Equivalently, one may invoke the uniqueness clause of 3.4: two linear maps on V that agree on a basis of V are equal.)

Problem 3B.27. Suppose $P \in \mathcal{L}(V)$ and $P^2 = P$. Prove that $V = \operatorname{null} P \oplus \operatorname{range} P$.

Solution. Recall that P^2 means $P \circ P$, so the hypothesis says $P(Pv) = Pv$ for every $v \in V$. By 3.13 and 3.18, $\operatorname{null} P$ and $\operatorname{range} P$ are subspaces of V, so the sum $\operatorname{null} P + \operatorname{range} P$ makes sense. Note that no finite-dimensionality is assumed or needed.
Step 1: $V = \operatorname{null} P + \operatorname{range} P$. Let $v \in V$. Write

$$v = (v - Pv) + Pv.$$

Clearly $Pv \in \operatorname{range} P$. For the first summand, linearity of P and the hypothesis give

$$P(v - Pv) = Pv - P(Pv) = Pv - P^2v = Pv - Pv = 0,$$

so $v - Pv \in \operatorname{null} P$. Hence $v \in \operatorname{null} P + \operatorname{range} P$. Since v was arbitrary and both summands are subspaces of V, we conclude $V = \operatorname{null} P + \operatorname{range} P$.
Step 2: $\operatorname{null} P \cap \operatorname{range} P = \{0\}$. Suppose $v \in \operatorname{null} P \cap \operatorname{range} P$. Because $v \in \operatorname{range} P$, there exists $u \in V$ with $v = Pu$. Because $v \in \operatorname{null} P$, we have $Pv = 0$. Therefore

$$v = Pu = P^2u = P(Pu) = Pv = 0.$$

Hence the intersection contains only 0; the reverse containment $\{0\} \subseteq \operatorname{null} P \cap \operatorname{range} P$ holds because both are subspaces.

Conclusion. By 1.46, a sum of two subspaces is a direct sum if and only if their intersection is $\{0\}$. Steps 1 and 2 therefore give

$$V = \operatorname{null} P \oplus \operatorname{range} P,$$

as desired.
Concretely, Step 1 exhibits the decomposition explicitly: the $\operatorname{range} P$ component of v is Pv and the $\operatorname{null} P$ component is $v - Pv$, and by directness this is the only such decomposition. Such a P is called a projection onto $\operatorname{range} P$ along $\operatorname{null} P$; note also that P acts as the identity on $\operatorname{range} P$, since $Pu = v$ implies $Pv = P^2u = Pu = v$.

Problem 3B.28. Suppose $D \in \mathcal{L}(\mathcal{P}(\mathbf{R}))$ is such that $\deg Dp = (\deg p) - 1$ for every nonconstant polynomial $p \in \mathcal{P}(\mathbf{R})$. Prove that D is surjective.
The notation D is used above to remind you of the differentiation map that sends a polynomial p to p'.

Solution. Throughout, $\mathcal{P}_m(\mathbf{R})$ denotes the space of polynomials with real coefficients of degree at most m; the list $1, x, \ldots, x^m$ is a basis of it, so $\dim \mathcal{P}_m(\mathbf{R}) = m + 1$. We use the standard convention $\deg 0 = -\infty$, and the word nonconstant means $\deg p \geq 1$. The strategy is to show that D maps $\mathcal{P}_m(\mathbf{R})$ **onto** $\mathcal{P}_{m-1}(\mathbf{R})$ for every $m \geq 1$; since every polynomial lies in some $\mathcal{P}_{m-1}(\mathbf{R})$, surjectivity of D follows.
Step 1: $D1$ is a constant polynomial. The hypothesis says nothing directly about constant polynomials, so we must extract this from linearity. The polynomial x is nonconstant with $\deg x = 1$, so $\deg Dx = 0$; thus $Dx = c$ for some nonzero $c \in \mathbf{R}$. Likewise $x + 1$ is nonconstant with $\deg(x + 1) = 1$, so $\deg D(x + 1) = 0$; thus $D(x + 1) = c'$ for some nonzero $c' \in \mathbf{R}$. By linearity $D(x + 1) = Dx + D1$, hence

$$D1 = D(x + 1) - Dx = c' - c,$$

a constant polynomial. Consequently, for every $a \in \mathbf{R}$ the polynomial $D(a \cdot 1) = a\, D1$ is constant.
Step 2: D maps $\mathcal{P}_m(\mathbf{R})$ into $\mathcal{P}_{m-1}(\mathbf{R})$ for every $m \geq 1$. Let $p \in \mathcal{P}_m(\mathbf{R})$. If p is nonconstant, then

$$\deg Dp = (\deg p) - 1 \leq m - 1,$$

so $Dp \in \mathcal{P}_{m-1}(\mathbf{R})$. If p is constant, then Dp is constant by Step 1, so $Dp \in \mathcal{P}_0(\mathbf{R}) \subseteq \mathcal{P}_{m-1}(\mathbf{R})$ because $m - 1 \geq 0$. Either way $Dp \in \mathcal{P}_{m-1}(\mathbf{R})$.
Step 3: $\operatorname{null} D \subseteq \mathcal{P}_0(\mathbf{R})$. Suppose $p \in \mathcal{P}(\mathbf{R})$ is nonconstant. Then $\deg Dp = (\deg p) - 1 \geq 0$, so Dp has a well-defined nonnegative degree and hence $Dp \neq 0$ (the zero polynomial has degree $-\infty$). Thus no nonconstant polynomial lies in $\operatorname{null} D$, that is, $\operatorname{null} D \subseteq \mathcal{P}_0(\mathbf{R})$.
Step 4: the restriction of D is surjective onto $\mathcal{P}_{m-1}(\mathbf{R})$. Fix an integer $m \geq 1$ and let

$$D_m = D|_{\mathcal{P}_m(\mathbf{R})}.$$

By Step 2 we may regard $D_m \in \mathcal{L}\big(\mathcal{P}_m(\mathbf{R}),\ \mathcal{P}_{m-1}(\mathbf{R})\big)$; it is linear because D is and $\mathcal{P}_m(\mathbf{R})$ is a subspace of $\mathcal{P}(\mathbf{R})$. Its null space is

$$\operatorname{null} D_m = \operatorname{null} D \cap \mathcal{P}_m(\mathbf{R}) \subseteq \mathcal{P}_0(\mathbf{R})$$

by Step 3, so $\dim \operatorname{null} D_m \leq \dim \mathcal{P}_0(\mathbf{R}) = 1$ by 2.37. Since $\mathcal{P}_m(\mathbf{R})$ is finite-dimensional, the fundamental theorem of linear maps (3.21) gives

$$\dim \operatorname{range} D_m = \dim \mathcal{P}_m(\mathbf{R}) - \dim \operatorname{null} D_m \geq (m + 1) - 1 = m = \dim \mathcal{P}_{m-1}(\mathbf{R}).$$

But $\operatorname{range} D_m$ is a subspace of $\mathcal{P}_{m-1}(\mathbf{R})$ (by 3.18), so $\dim \operatorname{range} D_m \leq \dim \mathcal{P}_{m-1}(\mathbf{R})$ by 2.37; hence equality holds. A subspace of full dimension equals the whole space (2.39), so

$$\operatorname{range} D_m = \mathcal{P}_{m-1}(\mathbf{R}).$$

Step 5: conclusion. Let $q \in \mathcal{P}(\mathbf{R})$ be arbitrary. Set $m = 1$ if $q = 0$, and $m = (\deg q) + 1$ otherwise; in either case $m \geq 1$ and $q \in \mathcal{P}_{m-1}(\mathbf{R})$. By Step 4 there exists $p \in \mathcal{P}_m(\mathbf{R})$ with

$$Dp = D_m p = q.$$

Hence $q \in \operatorname{range} D$. Since q was arbitrary, $\operatorname{range} D = \mathcal{P}(\mathbf{R})$, that is, D is surjective (by 3.19).
Note that D need not be the differentiation map: for instance $Dp = p' + p'''$ also satisfies the degree hypothesis, and the proof above shows it too is surjective. What drives the argument is only that lowering the degree by exactly one forces the restriction D_m to have at most a one-dimensional null space while its target is one dimension smaller than its domain.

Problem 3B.29. Suppose $p \in \mathcal{P}(\mathbb{R})$. Prove that there exists a polynomial $q \in \mathcal{P}(\mathbb{R})$ such that $5q'' + 3q' = p$.
This exercise can be done without linear algebra, but it's more fun to do it using linear algebra.

Solution. Choose $m \in \{0, 1, 2, \dots\}$ with $p \in \mathcal{P}_m(\mathbb{R})$; such an m exists because p is a polynomial (take $m = \deg p$ if $p \neq 0$, and $m = 0$ if $p = 0$).
Define

$$T \colon \mathcal{P}_{m+1}(\mathbb{R}) \to \mathcal{P}_m(\mathbb{R}), \qquad Tq = 5q'' + 3q'.$$

This makes sense: if $q \in \mathcal{P}_{m+1}(\mathbb{R})$ then q' and q'' both lie in $\mathcal{P}_m(\mathbb{R})$ (differentiation lowers degree by at least one), and $\mathcal{P}_m(\mathbb{R})$ is closed under sums and scalar multiples. Moreover T is linear, because differentiation is a linear operation on polynomials: for $q, r \in \mathcal{P}_{m+1}(\mathbb{R})$ and $\lambda \in \mathbb{R}$,

$$T(q + r) = 5(q + r)'' + 3(q + r)' = (5q'' + 3q') + (5r'' + 3r') = Tq + Tr,$$

and $T(\lambda q) = 5\lambda q'' + 3\lambda q' = \lambda Tq$.
Step 1: compute $\operatorname{null} T$.
Suppose $q \in \mathcal{P}_{m+1}(\mathbb{R})$ satisfies $5q'' + 3q' = 0$. Put $r = q'$, a polynomial, so that

$$5r' + 3r = 0, \qquad \text{i.e.} \qquad r = -\tfrac{5}{3}r'.$$

If $r \neq 0$, let $d = \deg r \geq 0$. Then $\deg r' < d$ (indeed $r' = 0$ if $d = 0$, and $\deg r' = d - 1$ if $d \geq 1$), so $\deg(-\frac{5}{3}r') < d = \deg r$, contradicting $r = -\frac{5}{3}r'$. Hence $r = 0$, i.e. $q' = 0$, so q is a constant polynomial. Conversely every constant polynomial q satisfies $5q'' + 3q' = 0$. Therefore

$$\operatorname{null} T = \{\text{constant polynomials}\} = \mathcal{P}_0(\mathbb{R}),$$

which has dimension 1.
Step 2: conclude surjectivity.
By the fundamental theorem of linear maps (3.21),

$$\dim \operatorname{range} T = \dim \mathcal{P}_{m+1}(\mathbb{R}) - \dim \operatorname{null} T = (m + 2) - 1 = m + 1,$$

using $\dim \mathcal{P}_k(\mathbb{R}) = k + 1$ (2.36). But $\operatorname{range} T$ is a subspace of $\mathcal{P}_m(\mathbb{R})$ (3.18), and $\dim \mathcal{P}_m(\mathbb{R}) = m + 1$ as well. A subspace whose dimension equals the dimension of the whole (finite-dimensional) space is the whole space (2.39), so

$$\operatorname{range} T = \mathcal{P}_m(\mathbb{R}).$$

Since $p \in \mathcal{P}_m(\mathbb{R}) = \operatorname{range} T$, there exists $q \in \mathcal{P}_{m+1}(\mathbb{R}) \subseteq \mathcal{P}(\mathbb{R})$ with $Tq = p$, that is, $5q'' + 3q' = p$, as desired.

Problem 3B.30. Suppose $\varphi \in \mathcal{L}(V, \mathbb{F})$ and $\varphi \neq 0$. Suppose $u \in V$ is not in $\operatorname{null} \varphi$. Prove that

$$V = \operatorname{null} \varphi \oplus \{au : a \in \mathbb{F}\}.$$

Solution. Write $U = \{au : a \in \mathbb{F}\} = \operatorname{span}(u)$. Then U is a subspace of V (the span of a list of vectors in V is a subspace of V, by 2.6), and $\operatorname{null} \varphi$ is a subspace of V (3.13). By 1.46 it therefore suffices to prove two things: that $\operatorname{null} \varphi + U = V$, and that $(\operatorname{null} \varphi) \cap U = \{0\}$.

The hypothesis $u \notin \operatorname{null}\varphi$ says precisely that

$$\varphi(u) \neq 0.$$

Note that this already forces $\varphi \neq 0$, so the hypothesis $\varphi \neq 0$ is only there to guarantee that such a u can be found; it is not needed again below.
The sum is all of V. Let $v \in V$ and set

$$a = \frac{\varphi(v)}{\varphi(u)} \in \mathbb{F},$$

which makes sense because $\varphi(u) \neq 0$. Then, by linearity of φ,

$$\varphi(v - au) = \varphi(v) - a\,\varphi(u) = \varphi(v) - \frac{\varphi(v)}{\varphi(u)}\,\varphi(u) = 0,$$

so $v - au \in \operatorname{null}\varphi$. Hence

$$v = (v - au) + au \in \operatorname{null}\varphi + U.$$

As $v \in V$ was arbitrary, $V = \operatorname{null}\varphi + U$ (the reverse inclusion is automatic, since $\operatorname{null}\varphi$ and U are subsets of V).
The intersection is trivial. Suppose $w \in (\operatorname{null}\varphi) \cap U$. Since $w \in U$, we have $w = au$ for some $a \in \mathbb{F}$. Since $w \in \operatorname{null}\varphi$,

$$0 = \varphi(w) = \varphi(au) = a\,\varphi(u).$$

Because $\varphi(u) \neq 0$ and $\mathbb{F}$ is a field, $a = 0$, and hence $w = 0u = 0$. Thus $(\operatorname{null}\varphi) \cap U = \{0\}$.
By 1.46 the sum $\operatorname{null}\varphi + U$ is a direct sum, and combined with the first part we conclude

$$V = \operatorname{null}\varphi \oplus \{au : a \in \mathbb{F}\},$$

as desired.

Problem 3B.31. Suppose V is finite-dimensional, X is a subspace of V, and Y is a finite-dimensional subspace of W. Prove that there exists $T \in \mathcal{L}(V, W)$ such that $\operatorname{null} T = X$ and $\operatorname{range} T = Y$ if and only if $\dim X + \dim Y = \dim V$.

Solution. Note first that X is automatically finite-dimensional, being a subspace of the finite-dimensional space V (2.25), so $\dim X$ makes sense; $\dim Y$ makes sense by hypothesis. Write $n = \dim V$ and $m = \dim X$, so $0 \leq m \leq n$ by 2.37.
Suppose first that such a T exists, say $T \in \mathcal{L}(V, W)$ with $\operatorname{null} T = X$ and $\operatorname{range} T = Y$. Since V is finite-dimensional, the fundamental theorem of linear maps (3.21) gives

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = \dim X + \dim Y.$$

This is the desired equality.
Conversely, suppose $\dim X + \dim Y = \dim V$, i.e. $\dim Y = n - m$.
Choose a basis $x_1, \ldots, x_m$ of X. This is a linearly independent list in V, so it extends to a basis

$$x_1, \ldots, x_m, v_1, \ldots, v_{n-m}$$

of V (2.32); the extension has exactly $n - m$ extra vectors because every basis of V has length n (2.34).
Choose also a basis $y_1, \ldots, y_{n-m}$ of Y, which has length $n - m$ since $\dim Y = n - m$.
By the linear map lemma (3.4) there is a unique $T \in \mathcal{L}(V, W)$ with

$$Tx_k = 0 \quad (k = 1, \ldots, m), \qquad Tv_j = y_j \quad (j = 1, \ldots, n - m),$$

where we regard $y_1, \ldots, y_{n-m}$ as vectors in W (they lie in the subspace Y of W). If $m = n$, so that $Y = \{0\}$ and there are no v_j's, this simply defines $T = 0$; if $m = 0$, there are no x_k's. Both degenerate cases are covered by what follows.

Range of T. Since $x_1, \dots, x_m, v_1, \dots, v_{n-m}$ spans V and T is linear, every $v \in V$ can be written $v = \sum_k a_k x_k + \sum_j b_j v_j$, and then

$$Tv = \sum_{j=1}^{n-m} b_j y_j.$$

Hence $\operatorname{range} T = \operatorname{span}(y_1, \dots, y_{n-m}) = Y$: the displayed formula shows $\operatorname{range} T \subseteq \operatorname{span}(y_1, \dots, y_{n-m})$, and each $y_j = Tv_j$ lies in $\operatorname{range} T$, which is a subspace (3.18), so the reverse inclusion holds too.
Null space of T. Each x_k satisfies $Tx_k = 0$, so $x_k \in \operatorname{null} T$; since $\operatorname{null} T$ is a subspace (3.13) containing $x_1, \dots, x_m$, it contains their span, namely X. Thus $X \subseteq \operatorname{null} T$. On the other hand, by 3.21 together with the computation of the range just made,

$$\dim \operatorname{null} T = \dim V - \dim \operatorname{range} T = n - (n-m) = m = \dim X.$$

So X is a subspace of the finite-dimensional space $\operatorname{null} T$ with $\dim X = \dim \operatorname{null} T$, whence $X = \operatorname{null} T$ by 2.39.
Therefore this T satisfies $\operatorname{null} T = X$ and $\operatorname{range} T = Y$, completing the proof of the converse and of the equivalence.

Problem 3B.32. Suppose V is finite-dimensional with $\dim V > 1$. Show that if $\varphi\colon \mathcal{L}(V) \to \mathbb{F}$ is a linear map such that $\varphi(ST) = \varphi(S)\varphi(T)$ for all $S, T \in \mathcal{L}(V)$, then $\varphi = 0$.
Hint: The description of the two-sided ideals of $\mathcal{L}(V)$ given by Exercise 17 in Section 3A might be useful.

Solution. Recall the definition from Exercise 17 in Section 3A: a subspace $\mathcal{E}$ of $\mathcal{L}(V)$ is a two-sided ideal of $\mathcal{L}(V)$ if $TE \in \mathcal{E}$ and $ET \in \mathcal{E}$ for all $E \in \mathcal{E}$ and all $T \in \mathcal{L}(V)$. That exercise asserts that when V is finite-dimensional the only two-sided ideals of $\mathcal{L}(V)$ are $\{0\}$ and $\mathcal{L}(V)$. For completeness we prove this first.
Lemma (Exercise 3A.17). If V is finite-dimensional and $\mathcal{E}$ is a two-sided ideal of $\mathcal{L}(V)$ with $\mathcal{E} \neq \{0\}$, then $\mathcal{E} = \mathcal{L}(V)$.
Proof of the lemma. We may assume $V \neq \{0\}$ (otherwise $\mathcal{L}(V) = \{0\}$ and there is nothing to prove). Let $n = \dim V \geq 1$ and fix a basis $v_1, \dots, v_n$ of V. Pick $E \in \mathcal{E}$ with $E \neq 0$; then there is $u \in V$ with $Eu \neq 0$.
For each $k \in \{1, \dots, n\}$, the linear map lemma (3.4) provides $A_k \in \mathcal{L}(V)$ with

$$A_k v_k = u, \qquad A_k v_j = 0 \ \text{ for } j \neq k.$$

Because $Eu \neq 0$, the one-element list Eu is linearly independent, so it extends to a basis $Eu, w_2, \dots, w_n$ of V (2.32). Again by 3.4, for each k there is $B_k \in \mathcal{L}(V)$ with

$$B_k(Eu) = v_k, \qquad B_k w_i = 0 \ \text{ for } i = 2, \dots, n.$$

Now compute $B_k E A_k$ on the basis: $B_k E A_k v_k = B_k E u = v_k$, while $B_k E A_k v_j = B_k E 0 = 0$ for $j \neq k$. Summing over k, the operator $\sum_{k=1}^{n} B_k E A_k$ sends each v_j to v_j, so by 3.4 (uniqueness) it equals the identity I.
Since $\mathcal{E}$ is a two-sided ideal, $B_k E A_k \in \mathcal{E}$ for each k (multiply $E \in \mathcal{E}$ on the right by A_k, then on the left by B_k), and $\mathcal{E}$ is a subspace, so $I = \sum_k B_k E A_k \in \mathcal{E}$. Finally, for any $T \in \mathcal{L}(V)$ we get $T = TI \in \mathcal{E}$. Hence $\mathcal{E} = \mathcal{L}(V)$, proving the lemma.
Now let $\varphi\colon \mathcal{L}(V) \to \mathbb{F}$ be linear with $\varphi(ST) = \varphi(S)\varphi(T)$ for all $S, T \in \mathcal{L}(V)$.
Claim: $\operatorname{null} \varphi$ is a two-sided ideal of $\mathcal{L}(V)$. Indeed, $\operatorname{null} \varphi$ is a subspace of $\mathcal{L}(V)$ because φ is linear (3.13). If $E \in \operatorname{null} \varphi$ and $T \in \mathcal{L}(V)$, then

$$\varphi(TE) = \varphi(T)\varphi(E) = \varphi(T) \cdot 0 = 0, \qquad \varphi(ET) = \varphi(E)\varphi(T) = 0 \cdot \varphi(T) = 0,$$

so $TE \in \operatorname{null} \varphi$ and $ET \in \operatorname{null} \varphi$. This proves the claim.
By the lemma, $\operatorname{null} \varphi = \{0\}$ or $\operatorname{null} \varphi = \mathcal{L}(V)$.

Suppose, for contradiction, that $\operatorname{null}\varphi = \{0\}$. Then φ is injective (3.15). But $\mathcal{L}(V)$ is finite-dimensional with

$$\dim \mathcal{L}(V) = (\dim V)^2 \geq 2^2 = 4 > 1 = \dim \mathbb{F},$$

using $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$ (3.72 applied with $W = V$) and $\dim V > 1$. Since $\dim \mathcal{L}(V) > \dim \mathbb{F}$, no linear map from $\mathcal{L}(V)$ to $\mathbb{F}$ is injective (3.22). This contradicts the injectivity of φ.
(If one prefers to avoid the dimension formula 3.72, here is a direct version of the same contradiction. Fix a basis $v_1, \ldots, v_n$ of V with $n \geq 2$, and let $P \in \mathcal{L}(V)$ be defined by $Pv_1 = v_1$ and $Pv_j = 0$ for $j \geq 2$. The operators I and P are linearly independent in $\mathcal{L}(V)$: if $aI + bP = 0$, applying this to v_2 gives $av_2 = 0$, so $a = 0$, and then applying bP to v_1 gives $bv_1 = 0$, so $b = 0$. But $\varphi(I)$ and $\varphi(P)$ lie in the one-dimensional space $\mathbb{F}$, so they are linearly dependent: there exist $a, b \in \mathbb{F}$, not both 0, with $a\varphi(I) + b\varphi(P) = 0$, i.e. $\varphi(aI + bP) = 0$. Injectivity of φ then forces $aI + bP = 0$, contradicting the linear independence of I and P.)
Hence $\operatorname{null}\varphi = \mathcal{L}(V)$, which says exactly that $\varphi(T) = 0$ for every $T \in \mathcal{L}(V)$, i.e. $\varphi = 0$.
Remark. The hypothesis $\dim V > 1$ cannot be dropped: if $\dim V = 1$ then every $T \in \mathcal{L}(V)$ is $T = \lambda_T I$ for a unique scalar λ_T, and $\varphi(T) = \lambda_T$ is a nonzero linear map satisfying $\varphi(ST) = \varphi(S)\varphi(T)$.

Problem 3B.33. Suppose that V and W are real vector spaces and $T \in \mathcal{L}(V, W)$. Define $T_{\mathbb{C}} \colon V_{\mathbb{C}} \to W_{\mathbb{C}}$ by

$$T_{\mathbb{C}}(u + iv) = Tu + iTv$$

for all $u, v \in V$.
(a) Show that $T_{\mathbb{C}}$ is a (complex) linear map from $V_{\mathbb{C}}$ to $W_{\mathbb{C}}$.
(b) Show that $T_{\mathbb{C}}$ is injective if and only if T is injective.
(c) Show that $\operatorname{range} T_{\mathbb{C}} = W_{\mathbb{C}}$ if and only if $\operatorname{range} T = W$.
See Exercise 8 in Section 1B for the definition of the complexification $V_{\mathbb{C}}$. The linear map $T_{\mathbb{C}}$ is called the complexification of the linear map T.

Solution. Throughout we use the definitions from Exercise 8 in Section 1B. As a set, $V_{\mathbb{C}} = V \times V$, and the ordered pair (u, v) is written $u + iv$. Thus two elements $u_1 + iv_1$ and $u_2 + iv_2$ of $V_{\mathbb{C}}$ are equal if and only if $u_1 = u_2$ and $v_1 = v_2$; in particular $u + iv = 0$ if and only if $u = 0$ and $v = 0$. Addition and complex scalar multiplication are given by

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2),$$

$$(a + bi)(u + iv) = (au - bv) + i(av + bu) \qquad (a, b \in \mathbb{R}).$$

The same formulas define the complex vector space $W_{\mathbb{C}}$. Note also that $T_{\mathbb{C}}$ is a genuinely well-defined function: every element of $V_{\mathbb{C}}$ has exactly one expression as $u + iv$ with $u, v \in V$, so the formula $T_{\mathbb{C}}(u + iv) = Tu + iTv$ assigns exactly one value to each input.
(a) We check additivity and complex homogeneity.
Additivity: for $u_1, v_1, u_2, v_2 \in V$,

$$\begin{aligned} T_{\mathbb{C}}\big((u_1 + iv_1) + (u_2 + iv_2)\big) &= T_{\mathbb{C}}\big((u_1 + u_2) + i(v_1 + v_2)\big) \\ &= T(u_1 + u_2) + i\,T(v_1 + v_2) \\ &= (Tu_1 + Tu_2) + i(Tv_1 + Tv_2) \\ &= (Tu_1 + iTv_1) + (Tu_2 + iTv_2) \\ &= T_{\mathbb{C}}(u_1 + iv_1) + T_{\mathbb{C}}(u_2 + iv_2), \end{aligned}$$

where the third equality uses the additivity of T and the fourth uses the definition of addition in $W_{\mathbb{C}}$.

Homogeneity over $\mathbb{C}$: let $a, b \in \mathbb{R}$ and $u, v \in V$. Then

$$\begin{aligned} T_{\mathbb{C}}\big((a+bi)(u+iv)\big) &= T_{\mathbb{C}}\big((au-bv)+i(av+bu)\big) \\ &= T(au-bv)+i\,T(av+bu) \\ &= (a\,Tu-b\,Tv)+i(a\,Tv+b\,Tu) \\ &= (a+bi)(Tu+iTv) \\ &= (a+bi)\,T_{\mathbb{C}}(u+iv), \end{aligned}$$

where the third equality uses that T is linear over $\mathbb{R}$ and the fourth is the definition of complex scalar multiplication in $W_{\mathbb{C}}$. Hence $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}}, W_{\mathbb{C}})$ as a complex linear map.
(b) By (a) and 3.15, $T_{\mathbb{C}}$ is injective if and only if $\operatorname{null} T_{\mathbb{C}} = \{0\}$, and T is injective if and only if $\operatorname{null} T = \{0\}$. We first identify the null space of $T_{\mathbb{C}}$:

$$u+iv \in \operatorname{null} T_{\mathbb{C}} \iff Tu+iTv = 0 \iff Tu = 0 \text{ and } Tv = 0 \iff u, v \in \operatorname{null} T,$$

the middle equivalence because an element $w_1 + iw_2$ of $W_{\mathbb{C}}$ is 0 exactly when $w_1 = 0$ and $w_2 = 0$. In other words $\operatorname{null} T_{\mathbb{C}} = (\operatorname{null} T)_{\mathbb{C}}$.
Now suppose T is injective, so $\operatorname{null} T = \{0\}$. If $u + iv \in \operatorname{null} T_{\mathbb{C}}$, then by the display $u \in \operatorname{null} T$ and $v \in \operatorname{null} T$, so $u = v = 0$ and hence $u + iv = 0$. Thus $\operatorname{null} T_{\mathbb{C}} = \{0\}$ and $T_{\mathbb{C}}$ is injective.
Conversely suppose $T_{\mathbb{C}}$ is injective, and let $u \in \operatorname{null} T$. Then $T_{\mathbb{C}}(u + i0) = Tu + i\,T0 = 0 + i0 = 0$, so $u + i0 = 0$, which gives $u = 0$. Thus $\operatorname{null} T = \{0\}$ and T is injective.
(c) Suppose first that $\operatorname{range} T = W$. Let $w_1 + iw_2 \in W_{\mathbb{C}}$ be arbitrary, with $w_1, w_2 \in W$. Choose $u, v \in V$ with $Tu = w_1$ and $Tv = w_2$. Then

$$T_{\mathbb{C}}(u+iv) = Tu + iTv = w_1 + iw_2.$$

Hence every element of $W_{\mathbb{C}}$ is in $\operatorname{range} T_{\mathbb{C}}$, so $\operatorname{range} T_{\mathbb{C}} = W_{\mathbb{C}}$.
Conversely suppose $\operatorname{range} T_{\mathbb{C}} = W_{\mathbb{C}}$. Let $w \in W$. Then $w + i0 \in W_{\mathbb{C}}$, so there exist $u, v \in V$ with

$$Tu + iTv = T_{\mathbb{C}}(u+iv) = w + i0.$$

Comparing the two coordinates gives $Tu = w$ (and $Tv = 0$). Hence $w \in \operatorname{range} T$. As $w \in W$ was arbitrary, $\operatorname{range} T = W$.
(The two computations in (c) together show more precisely that $\operatorname{range} T_{\mathbb{C}} = (\operatorname{range} T)_{\mathbb{C}}$, viewing $(\operatorname{range} T)_{\mathbb{C}}$ as the set of $w_1 + iw_2 \in W_{\mathbb{C}}$ with $w_1, w_2 \in \operatorname{range} T$.)

3.3 Exercises 3C

Problem 3C.1. Suppose $T \in \mathcal{L}(V, W)$. Show that with respect to each choice of bases of V and W, the matrix of T has at least $\dim \operatorname{range} T$ nonzero entries.

Solution. Fix a basis $v_1, \ldots, v_n$ of V and a basis $w_1, \ldots, w_m$ of W, and let $A = \mathcal{M}(T)$ be the matrix of T with respect to these bases. By the definition of the matrix of a linear map,

$$Tv_k = \sum_{j=1}^{m} A_{j,k}\, w_j \qquad \text{for each } k \in \{1, \ldots, n\}.$$

Let

$$K = \{\, k \in \{1, \ldots, n\} : \text{column } k \text{ of } A \text{ has at least one nonzero entry} \,\}.$$

First we bound $\dim \operatorname{range} T$ by the size of K. Because $v_1, \ldots, v_n$ spans V and T is linear, $\operatorname{range} T = \operatorname{span}(Tv_1, \ldots, Tv_n)$. If $k \notin K$, then every entry $A_{j,k}$ is 0, and the displayed formula gives $Tv_k = 0$; such a vector contributes nothing to a span. Hence

$$\operatorname{range} T = \operatorname{span}(Tv_k : k \in K),$$

a span of a list of length $|K|$. The dimension of the span of a list is at most the length of that list (a spanning list of $\operatorname{span}(Tv_k : k \in K)$ can be reduced to a basis of it, by 2.30, and that basis is no longer than the original list), so

$$\dim \operatorname{range} T \leq |K|.$$

Now we bound $|K|$ by the number of nonzero entries of A. For each $k \in K$ choose one entry of column k that is nonzero; distinct indices k give entries in distinct columns of A, hence distinct entries of A. This exhibits $|K|$ distinct nonzero entries of A, so

$$|K| \leq \#\{\text{nonzero entries of } A\}.$$

Combining the two inequalities,

$$\dim \operatorname{range} T \leq |K| \leq \#\{\text{nonzero entries of } \mathcal{M}(T)\},$$

which is exactly the assertion: the matrix of T with respect to any choice of bases has at least $\dim \operatorname{range} T$ nonzero entries.
(Note that Exercise 3C.5 shows that this bound is attained: there always exist bases for which $\mathcal{M}(T)$ has exactly $\dim \operatorname{range} T$ nonzero entries, all of them equal to 1 and sitting on the diagonal.)

Problem 3C.2. Suppose $T \in \mathcal{L}(V, W)$, where V and W are finite-dimensional and nonzero. Prove that $\dim \operatorname{range} T = 1$ if and only if there exist a basis of V and a basis of W such that with respect to these bases, all entries of $\mathcal{M}(T)$ equal 1.

Solution. Write $n = \dim V$ and $m = \dim W$; because V and W are nonzero we have $n \geq 1$ and $m \geq 1$.
First suppose there are bases $v_1, \ldots, v_n$ of V and $w_1, \ldots, w_m$ of W with respect to which every entry of $\mathcal{M}(T)$ equals 1. Put $w = w_1 + \cdots + w_m$. By the definition of the matrix of a linear map, for each $k \in \{1, \ldots, n\}$ we have

$$Tv_k = \sum_{j=1}^{m} 1 \cdot w_j = w.$$

The vector w is nonzero, because $w_1, \ldots, w_m$ is linearly independent and the coefficients $1, \ldots, 1$ are not all 0. Since $v_1, \ldots, v_n$ spans V,

$$\operatorname{range} T = \operatorname{span}(Tv_1, \ldots, Tv_n) = \operatorname{span}(w),$$

and $\operatorname{span}(w)$ has dimension 1 because $w \neq 0$. Thus $\dim \operatorname{range} T = 1$.
Conversely, suppose $\dim \operatorname{range} T = 1$. Choose $w \in W$ with $w \neq 0$ and $\operatorname{range} T = \operatorname{span}(w)$.
We first build the basis of W. The list w is linearly independent in W (as $w \neq 0$), so it extends to a basis $u_1, \ldots, u_m$ of W with $u_1 = w$ (by 2.32). Define

$$w_1 = u_1 - u_2 - \cdots - u_m, \qquad w_j = u_j \ \text{ for } j \in \{2, \ldots, m\}.$$

(When $m = 1$ this simply says $w_1 = u_1$.) Then $\operatorname{span}(w_1, \ldots, w_m)$ contains $u_2, \ldots, u_m$ and also contains $w_1 + u_2 + \cdots + u_m = u_1$; hence it contains $\operatorname{span}(u_1, \ldots, u_m) = W$. So $w_1, \ldots, w_m$ is a spanning list of W of length $m = \dim W$, and therefore a basis of W (by 2.42, a spanning list of the right length is a basis). By construction

$$w_1 + w_2 + \cdots + w_m = u_1 = w.$$

Now we build the basis of V. Since $w \in \operatorname{range} T$, there is $v_1 \in V$ with $Tv_1 = w$. By the fundamental theorem of linear maps (3.21),

$$\dim \operatorname{null} T = \dim V - \dim \operatorname{range} T = n - 1.$$

Let $z_1, \ldots, z_{n-1}$ be a basis of $\operatorname{null} T$ (an empty list if $n = 1$); the letters z are used to avoid any clash with the vectors $u_1, \ldots, u_m$ of W above. The list $v_1, z_1, \ldots, z_{n-1}$ is linearly independent: if $av_1 + c_1 z_1 + \cdots + c_{n-1} z_{n-1} = 0$, applying T gives $aw = 0$, hence $a = 0$ since $w \neq 0$, and then

$c_1 = \cdots = c_{n-1} = 0$ by the linear independence of $z_1, \ldots, z_{n-1}$. Having length $n = \dim V$, this list is a basis of V (by 2.38).
Define

$$v_k = v_1 + z_{k-1} \qquad \text{for } k \in \{2, \ldots, n\}.$$

Then $\operatorname{span}(v_1, \ldots, v_n)$ contains v_1 and contains $v_k - v_1 = z_{k-1}$ for each $k \geq 2$, hence contains $\operatorname{span}(v_1, z_1, \ldots, z_{n-1}) = V$. So $v_1, \ldots, v_n$ spans V and has length $n = \dim V$, and is therefore a basis of V (by 2.42).
Finally, $Tv_1 = w$, and for $k \in \{2, \ldots, n\}$ we have $Tv_k = Tv_1 + Tz_{k-1} = w + 0 = w$. Thus for every k,

$$Tv_k = w = w_1 + w_2 + \cdots + w_m = \sum_{j=1}^{m} 1 \cdot w_j,$$

so every entry of column k of $\mathcal{M}(T)$ equals 1. As this holds for each k, all entries of $\mathcal{M}(T)$ with respect to the bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ equal 1.

Problem 3C.3. Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W.
- (a) Show that if $S, T \in \mathcal{L}(V, W)$, then $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$.
- (b) Show that if $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$, then $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.

[This exercise asks you to verify 3.35 and 3.38.]

Solution. Throughout, all matrices are taken with respect to the fixed bases $v_1, \ldots, v_n$ of V and $w_1, \ldots, w_m$ of W. Recall the defining property of the matrix of a linear map $R \in \mathcal{L}(V, W)$: $\mathcal{M}(R)$ is the m-by-n matrix whose entries $\mathcal{M}(R)_{j,k}$ are the unique scalars satisfying

$$Rv_k = \sum_{j=1}^{m} \mathcal{M}(R)_{j,k}\, w_j \qquad \text{for each } k \in \{1, \ldots, n\}.$$

Uniqueness of these scalars is exactly the statement that $w_1, \ldots, w_m$, being a basis of W, represents each vector of W as a linear combination of $w_1, \ldots, w_m$ in only one way (the criterion for basis, 2.28). This uniqueness is what makes the arguments below work: to identify the entries of a matrix $\mathcal{M}(R)$, it suffices to exhibit any one expression of Rv_k as a linear combination of $w_1, \ldots, w_m$ and read off its coefficients.
(a) Let $S, T \in \mathcal{L}(V, W)$, and set $A = \mathcal{M}(S)$ and $C = \mathcal{M}(T)$; both are m-by-n matrices. Fix $k \in \{1, \ldots, n\}$. Using the definition of the sum $S + T$ of two linear maps and then the two displayed formulas above,

$$\begin{aligned}(S + T)v_k &= Sv_k + Tv_k \\ &= \sum_{j=1}^{m} A_{j,k}\, w_j + \sum_{j=1}^{m} C_{j,k}\, w_j \\ &= \sum_{j=1}^{m} (A_{j,k} + C_{j,k})\, w_j.\end{aligned}$$

Because $S + T \in \mathcal{L}(V, W)$ (by 3.5 and 3.6) and the representation of $(S + T)v_k$ in the basis $w_1, \ldots, w_m$ is unique, we conclude that

$$\mathcal{M}(S + T)_{j,k} = A_{j,k} + C_{j,k} \qquad \text{for all } j \in \{1, \ldots, m\},\ k \in \{1, \ldots, n\}.$$

By the definition of matrix addition (3.34), the right side is the entry in row j, column k, of $A + C$. Since $\mathcal{M}(S + T)$ and $A + C$ are both m-by-n and agree in every entry, $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$. This proves 3.35.
(b) Let $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$, and set $C = \mathcal{M}(T)$. Fix $k \in \{1, \ldots, n\}$. Using the definition of the scalar multiple λT and then the defining formula for C,

$$(\lambda T)v_k = \lambda(Tv_k) = \lambda \sum_{j=1}^{m} C_{j,k}\, w_j = \sum_{j=1}^{m} (\lambda C_{j,k})\, w_j,$$

where the last equality uses distributivity and associativity of scalar multiplication in W. Again $\lambda T \in \mathcal{L}(V,W)$ (by 3.5 and 3.6), and uniqueness of the representation in the basis $w_1,\dots,w_m$ gives

$$\mathcal{M}(\lambda T)_{j,k} = \lambda C_{j,k} \qquad \text{for all } j,k.$$

By the definition of scalar multiplication of a matrix (3.36), the right side is the entry in row j, column k, of λC. Hence $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$. This proves 3.38.

Problem 3C.4. Suppose that $D \in \mathcal{L}(\mathcal{P}_3(\mathbf{R}), \mathcal{P}_2(\mathbf{R}))$ is the differentiation map defined by $Dp = p'$. Find a basis of $\mathcal{P}_3(\mathbf{R})$ and a basis of $\mathcal{P}_2(\mathbf{R})$ such that the matrix of D with respect to these bases is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

[Compare with Example 3.33. The next exercise generalizes this exercise.]

Solution. Take the basis of $\mathcal{P}_3(\mathbf{R})$ to be

$$v_1 = x, \qquad v_2 = \tfrac{x^2}{2}, \qquad v_3 = \tfrac{x^3}{3}, \qquad v_4 = 1,$$

and the basis of $\mathcal{P}_2(\mathbf{R})$ to be the standard one,

$$w_1 = 1, \qquad w_2 = x, \qquad w_3 = x^2.$$

First we check that these lists really are bases.
The list w_1, w_2, w_3 is the standard basis $1, x, x^2$ of $\mathcal{P}_2(\mathbf{R})$, so it is a basis; in particular $\dim \mathcal{P}_2(\mathbf{R}) = 3$.
For v_1, v_2, v_3, v_4: the span of this list contains 1, x, $\frac{x^2}{2}$, $\frac{x^3}{3}$, hence (being closed under scalar multiplication) contains $1, x, x^2, x^3$, and therefore contains $\operatorname{span}(1, x, x^2, x^3) = \mathcal{P}_3(\mathbf{R})$. So v_1, v_2, v_3, v_4 spans $\mathcal{P}_3(\mathbf{R})$, and its length 4 equals $\dim \mathcal{P}_3(\mathbf{R}) = 4$. By 2.42, a spanning list of the right length is a basis, so v_1, v_2, v_3, v_4 is a basis of $\mathcal{P}_3(\mathbf{R})$.
Now compute the matrix. Differentiating,

$$\begin{aligned} Dv_1 &= (x)' = 1 && = 1 \cdot w_1 + 0 \cdot w_2 + 0 \cdot w_3, \\ Dv_2 &= \left(\tfrac{x^2}{2}\right)' = x && = 0 \cdot w_1 + 1 \cdot w_2 + 0 \cdot w_3, \\ Dv_3 &= \left(\tfrac{x^3}{3}\right)' = x^2 && = 0 \cdot w_1 + 0 \cdot w_2 + 1 \cdot w_3, \\ Dv_4 &= (1)' = 0 && = 0 \cdot w_1 + 0 \cdot w_2 + 0 \cdot w_3. \end{aligned}$$

The k-th column of $\mathcal{M}(D)$ consists of the coefficients needed to write Dv_k as a linear combination of w_1, w_2, w_3. Reading off the four lines above,

$$\mathcal{M}(D) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

as required.
Two remarks. In Example 3.33 the same map D, computed with respect to the standard bases $1, x, x^2, x^3$ and $1, x, x^2$, has matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix},$$

whose only nonzero entries are the $1, 2, 3$ just above the diagonal. Two changes produce the matrix asked for here: replacing the domain basis vectors x, x^2, x^3 by $x, \frac{x^2}{2}, \frac{x^3}{3}$ divides out the factors $1, 2, 3$, and moving the constant polynomial 1 (which spans $\operatorname{null} D$) from the first to the last position shifts the zero column from the front to the back, which is what turns the superdiagonal into the diagonal. This is exactly the recipe of Exercise 3C.5 applied to D: here $\dim \operatorname{range} D = 3$, since D is surjective onto $\mathcal{P}_2(\mathbf{R})$, and indeed the matrix above has 1 in row k, column k, for $1 \le k \le 3$ and 0 elsewhere.

Problem 3C.5. Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that there exist a basis of V and a basis of W such that with respect to these bases, all entries of $\mathcal{M}(T)$ are 0 except that the entries in row k, column k, equal 1 if $1 \le k \le \dim \operatorname{range} T$.

Solution. Write $n = \dim V$, $m = \dim W$, and $r = \dim \operatorname{range} T$. Note $r \le m$, since $\operatorname{range} T$ is a subspace of W (by 2.37), and $r \le n$ by the fundamental theorem of linear maps.
Step 1: a convenient basis of V.
Since $\operatorname{null} T$ is a subspace of the finite-dimensional space V, it is finite-dimensional and has a basis $u_1, \dots, u_d$, where $d = \dim \operatorname{null} T$. By the fundamental theorem of linear maps (3.21),

$$n = \dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = d + r, \qquad \text{so } d = n - r.$$

A basis of $\operatorname{null} T$ is a linearly independent list in V, so it extends to a basis of V (by 2.32); write that extension as

$$u_1, \dots, u_{n-r},\ v_1, \dots, v_r,$$

which necessarily has r new vectors because every basis of V has length n.
Step 2: the images $Tv_1, \dots, Tv_r$ form a basis of $\operatorname{range} T$.
They span $\operatorname{range} T$: for $v \in V$ write $v = a_1u_1 + \cdots + a_{n-r}u_{n-r} + b_1v_1 + \cdots + b_rv_r$; applying T and using $Tu_i = 0$ gives $Tv = b_1Tv_1 + \cdots + b_rTv_r$. They are linearly independent: if $c_1Tv_1 + \cdots + c_rTv_r = 0$, then $T(c_1v_1 + \cdots + c_rv_r) = 0$, so $c_1v_1 + \cdots + c_rv_r \in \operatorname{null} T$ and hence

$$c_1v_1 + \cdots + c_rv_r = a_1u_1 + \cdots + a_{n-r}u_{n-r}$$

for some scalars a_i; linear independence of the basis $u_1, \dots, u_{n-r}, v_1, \dots, v_r$ of V then forces all $c_k = 0$. (This is precisely the argument inside the proof of 3.21.)
Step 3: the bases.
Define

$$w_k = Tv_k \qquad \text{for } k \in \{1, \dots, r\}.$$

By Step 2 the list $w_1, \dots, w_r$ is linearly independent in W, so by 2.32 it extends to a basis

$$w_1, \dots, w_r, w_{r+1}, \dots, w_m$$

of W (again the length must be $m = \dim W$).
For the domain, reorder the basis found in Step 1 by putting the v's first: set

$$\tilde{v}_k = v_k \ \text{ for } 1 \le k \le r, \qquad \tilde{v}_{r+i} = u_i \ \text{ for } 1 \le i \le n - r.$$

Reordering a basis yields a basis (spanning and linear independence do not depend on the order of the list), so $\tilde{v}_1, \dots, \tilde{v}_n$ is a basis of V.
Step 4: computing $\mathcal{M}(T)$.
Let $A = \mathcal{M}(T)$ with respect to $\tilde{v}_1, \dots, \tilde{v}_n$ and $w_1, \dots, w_m$. For $1 \le k \le r$,

$$T\tilde{v}_k = Tv_k = w_k = 0 \cdot w_1 + \cdots + 0 \cdot w_{k-1} + 1 \cdot w_k + 0 \cdot w_{k+1} + \cdots + 0 \cdot w_m,$$

so column k of A has a 1 in row k and 0 everywhere else. For $r < k \le n$ we have $\tilde{v}_k = u_{k-r} \in \operatorname{null} T$, so

$$T\tilde{v}_k = 0 = 0 \cdot w_1 + \cdots + 0 \cdot w_m,$$

and column k of A is entirely 0.
Therefore $A_{k,k} = 1$ for $1 \le k \le r = \dim \operatorname{range} T$, and all other entries of A are 0, which is what was to be proved.
(In particular $\mathcal{M}(T)$ then has exactly $\dim \operatorname{range} T$ nonzero entries, showing that the lower bound of Exercise 3C.1 is sharp.)

Problem 3C.6. Suppose $v_1, \dots, v_m$ is a basis of V and W is finite-dimensional. Suppose $T \in \mathcal{L}(V, W)$. Prove that there exists a basis $w_1, \dots, w_n$ of W such that all entries in the first column of $\mathcal{M}(T)$ [with respect to the bases $v_1, \dots, v_m$ and $w_1, \dots, w_n$] are 0 except for possibly a 1 in the first row, first column.
[In this exercise, unlike Exercise 5, you are given the basis of V instead of being able to choose a basis of V.]

Solution. Let $n = \dim W$. Recall that if $w_1, \dots, w_n$ is a basis of W and $A = \mathcal{M}(T)$ with respect to $v_1, \dots, v_m$ and $w_1, \dots, w_n$, then the first column of A is determined by

$$Tv_1 = A_{1,1}w_1 + A_{2,1}w_2 + \cdots + A_{n,1}w_n.$$

So what must be arranged is precisely this: either $Tv_1 = 0$ (first column all 0), or $Tv_1 = w_1$ for the first vector w_1 of the chosen basis (first column equal to $1, 0, \dots, 0$).
Case 1: $Tv_1 = 0$.
Since W is finite-dimensional it has a basis (by 2.31); let $w_1, \dots, w_n$ be any basis of W. Then

$$Tv_1 = 0 = 0 \cdot w_1 + 0 \cdot w_2 + \cdots + 0 \cdot w_n,$$

so by the uniqueness of the representation of a vector in a basis (2.28), $A_{j,1} = 0$ for every $j \in \{1, \dots, n\}$. Every entry in the first column is 0, which is allowed by the statement, since the 1 in the upper-left corner is only possibly there.
Case 2: $Tv_1 \neq 0$.
Set $w_1 = Tv_1$. The one-element list w_1 is linearly independent in W, because $w_1 \neq 0$. Since W is finite-dimensional, this list extends to a basis

$$w_1, w_2, \dots, w_n$$

of W (by 2.32); the extension has length $n = \dim W$ because all bases of W have that length. Now

$$Tv_1 = w_1 = 1 \cdot w_1 + 0 \cdot w_2 + \cdots + 0 \cdot w_n,$$

so again by uniqueness of coefficients (2.28) we get $A_{1,1} = 1$ and $A_{j,1} = 0$ for $j \in \{2, \dots, n\}$. Thus the first column of $\mathcal{M}(T)$ is 0 except for a 1 in row 1.
In both cases a basis $w_1, \dots, w_n$ of W has been produced with the required property, so the result holds. (The case distinction is genuinely necessary: if $Tv_1 = 0$, then no choice of basis of W can put a 1 in the upper-left corner, since $Tv_1 = 0$ forces every entry of the first column to vanish. Conversely if $Tv_1 \neq 0$ then the first column cannot be all 0, no matter which basis of W is used.)
Note that nothing was assumed about $v_2, \dots, v_m$ or about the remaining columns of $\mathcal{M}(T)$; only the first basis vector of V enters the argument, and the freedom used is entirely the freedom to choose the basis of W.

Problem 3C.7. Suppose $w_1, \dots, w_n$ is a basis of W and V is finite-dimensional. Suppose $T \in \mathcal{L}(V, W)$. Prove that there exists a basis $v_1, \dots, v_m$ of V such that all entries in the first row of $\mathcal{M}(T)$ [with respect to the bases $v_1, \dots, v_m$ and $w_1, \dots, w_n$] are 0 except for possibly a 1 in the first row, first column.
[In this exercise, unlike Exercise 5, you are given the basis of W instead of being able to choose a basis of W.]

Solution. Let $m = \dim V$. If $m = 0$ the matrix $\mathcal{M}(T)$ has no columns and the assertion is vacuous, so assume $m \geq 1$.
The key observation is that the first row of $\mathcal{M}(T)$ records, for each domain basis vector, the coefficient of w_1 in its image. Make this precise as follows. Since $w_1, \dots, w_n$ is a basis of W, every $w \in W$ can be

written uniquely as $w = a_1w_1 + \cdots + a_nw_n$ with $a_1, \ldots, a_n \in \mathbf{F}$ (by 2.28). Define $\pi\colon W \to \mathbf{F}$ by

$$\pi(a_1w_1 + \cdots + a_nw_n) = a_1.$$

Uniqueness of the coefficients makes π a well-defined function, and it is linear: if $w = a_1w_1 + \cdots + a_nw_n$ and $w' = a'_1w_1 + \cdots + a'_nw_n$ and $\lambda \in \mathbf{F}$, then $w + w' = (a_1 + a'_1)w_1 + \cdots + (a_n + a'_n)w_n$ and $\lambda w = (\lambda a_1)w_1 + \cdots + (\lambda a_n)w_n$, whence $\pi(w + w') = a_1 + a'_1 = \pi(w) + \pi(w')$ and $\pi(\lambda w) = \lambda a_1 = \lambda\pi(w)$. Now set

$$\varphi = \pi \circ T \in \mathcal{L}(V, \mathbf{F}),$$

which is linear as a composition (indeed a product) of linear maps (3.7). By construction, for any basis $v_1, \ldots, v_m$ of V, if $A = \mathcal{M}(T)$ with respect to $v_1, \ldots, v_m$ and $w_1, \ldots, w_n$, then $Tv_k = \sum_{j=1}^{n} A_{j,k}w_j$ gives

$$\varphi(v_k) = \pi(Tv_k) = A_{1,k}.$$

So the first row of $\mathcal{M}(T)$ is exactly the list $\varphi(v_1), \ldots, \varphi(v_m)$, and the task is to find a basis $v_1, \ldots, v_m$ of V with $\varphi(v_1) \in \{0, 1\}$ and $\varphi(v_2) = \cdots = \varphi(v_m) = 0$.
Case 1: $\varphi = 0$.
Take $v_1, \ldots, v_m$ to be any basis of V (one exists by 2.31). Then $A_{1,k} = \varphi(v_k) = 0$ for every k, so the entire first row of $\mathcal{M}(T)$ is 0, which is permitted.
Case 2: $\varphi \neq 0$.
Then $\operatorname{range}\varphi$ is a nonzero subspace of $\mathbf{F}$, and $\dim \mathbf{F} = 1$, so $\operatorname{range}\varphi = \mathbf{F}$ and $\dim \operatorname{range}\varphi = 1$ (by 2.37 and 2.39). By the fundamental theorem of linear maps (3.21),

$$\dim \operatorname{null}\varphi = \dim V - \dim \operatorname{range}\varphi = m - 1.$$

Because $1 \in \operatorname{range}\varphi$, choose $v_1 \in V$ with $\varphi(v_1) = 1$. Let $v_2, \ldots, v_m$ be a basis of $\operatorname{null}\varphi$, which is a list of the correct length $m - 1$ (an empty list if $m = 1$).
The list $v_1, v_2, \ldots, v_m$ is linearly independent: suppose

$$c_1v_1 + c_2v_2 + \cdots + c_mv_m = 0.$$

Applying φ and using $\varphi(v_k) = 0$ for $k \geq 2$ gives $c_1\varphi(v_1) = 0$, that is, $c_1 = 0$. The equation then reads $c_2v_2 + \cdots + c_mv_m = 0$, and linear independence of the basis $v_2, \ldots, v_m$ of $\operatorname{null}\varphi$ forces $c_2 = \cdots = c_m = 0$.
Since this linearly independent list has length $m = \dim V$, it is a basis of V (by 2.38).
With respect to this basis $v_1, \ldots, v_m$ of V and the given basis $w_1, \ldots, w_n$ of W, the first row of $\mathcal{M}(T)$ is

$$A_{1,1} = \varphi(v_1) = 1, \qquad A_{1,k} = \varphi(v_k) = 0 \ \text{ for } k \in \{2, \ldots, m\},$$

so all entries in the first row are 0 except for a 1 in row 1, column 1.
In both cases the required basis of V exists, completing the proof. (As in Exercise 3C.6, the two cases are forced on us: $\varphi = 0$ says that $\operatorname{range} T$ lies inside $\operatorname{span}(w_2, \ldots, w_n)$, and then no basis of V can produce a nonzero entry in the first row; otherwise the first row cannot be identically 0.)

Problem 3C.8. Suppose A is an m-by-n matrix and B is an n-by-p matrix. Prove that

$$(AB)_{j,\cdot} = A_{j,\cdot}\, B$$

for each $1 \leq j \leq m$. In other words, show that row j of AB equals (row j of A) times B.
[This exercise gives the row version of 3.48.]

Solution. Fix j with $1 \leq j \leq m$.
First we check that both sides of the claimed equation are matrices of the same size. The matrix AB is m-by-p (by 3.41), so its row j, namely $(AB)_{j,\cdot}$, is a 1-by-p matrix (by 3.44). On the other side, $A_{j,\cdot}$ is a 1-by-n matrix (by 3.44), and B is n-by-p; hence the product $A_{j,\cdot}B$ is defined and is a 1-by-p matrix (by 3.41). Thus both sides are 1-by-p matrices, and to prove they are equal it suffices to show that

they have the same entry in column k for each $k \in \{1, \dots, p\}$.
So fix k with $1 \le k \le p$. The entry in row 1, column k, of $(AB)_{j,\cdot}$ is by definition the entry in row j, column k, of AB, which by the definition of matrix multiplication (3.41) equals

$$(AB)_{j,k} = \sum_{r=1}^{n} A_{j,r} B_{r,k}.$$

Now consider the right side. The 1-by-n matrix $A_{j,\cdot}$ has entries $(A_{j,\cdot})_{1,r} = A_{j,r}$ for $r \in \{1, \dots, n\}$. Hence, again by 3.41, the entry in row 1, column k, of the product $A_{j,\cdot}B$ equals

$$\big(A_{j,\cdot}B\big)_{1,k} = \sum_{r=1}^{n} \big(A_{j,\cdot}\big)_{1,r} B_{r,k} = \sum_{r=1}^{n} A_{j,r} B_{r,k}.$$

The two displayed sums are identical. Thus $(AB)_{j,\cdot}$ and $A_{j,\cdot}B$ are 1-by-p matrices with the same entry in column k for every $k \in \{1, \dots, p\}$, and therefore

$$(AB)_{j,\cdot} = A_{j,\cdot}\, B,$$

as desired. In words: row j of AB equals (row j of A) times B.
This is exactly the row analogue of 3.48, and the proof above is the mirror image of the proof of 3.48: there one fixes a column index k and varies the row index, here one fixes a row index j and varies the column index.

Problem 3C.9. Suppose $a = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$ is a 1-by-n matrix and B is an n-by-p matrix. Prove that

$$aB = a_1 B_{1,\cdot} + \cdots + a_n B_{n,\cdot}.$$

In other words, show that aB is a linear combination of the rows of B, with the scalars that multiply the rows coming from a.
[This exercise gives the row version of 3.50.]

Solution. First we check sizes. Since a is 1-by-n and B is n-by-p, the product aB is defined and is a 1-by-p matrix (by 3.41). For each $r \in \{1, \dots, n\}$, the row $B_{r,\cdot}$ is a 1-by-p matrix (by 3.44), so each scalar multiple $a_r B_{r,\cdot}$ is 1-by-p (by 3.36) and the sum $a_1 B_{1,\cdot} + \cdots + a_n B_{n,\cdot}$ is a 1-by-p matrix (by 3.34). Thus both sides of the claimed equation are 1-by-p matrices, and it suffices to prove that for each $k \in \{1, \dots, p\}$ they have the same entry in column k.
Fix k with $1 \le k \le p$. Writing $a_{1,r} = a_r$ for the entries of a, the definition of matrix multiplication (3.41) gives

$$(aB)_{1,k} = \sum_{r=1}^{n} a_{1,r} B_{r,k} = \sum_{r=1}^{n} a_r B_{r,k}.$$

Now compute the entry in column k of the right side. For each r, the entry in row 1, column k, of $B_{r,\cdot}$ is $B_{r,k}$, so by the definitions of scalar multiplication of a matrix (3.36) and matrix addition (3.34),

$$\big(a_1 B_{1,\cdot} + \cdots + a_n B_{n,\cdot}\big)_{1,k} = a_1 B_{1,k} + \cdots + a_n B_{n,k} = \sum_{r=1}^{n} a_r B_{r,k}.$$

The two displayed expressions agree. Because aB and $a_1 B_{1,\cdot} + \cdots + a_n B_{n,\cdot}$ are 1-by-p matrices whose entries in column k agree for every $k \in \{1, \dots, p\}$, we conclude that

$$aB = a_1 B_{1,\cdot} + \cdots + a_n B_{n,\cdot}.$$

Thus aB is the linear combination of the rows $B_{1,\cdot}, \dots, B_{n,\cdot}$ of B whose coefficients $a_1, \dots, a_n$ are the entries of a. This is the row version of 3.50, whose proof it mirrors: 3.50 fixes a row index and expands over the columns of A, whereas here we fix a column index and expand over the rows of B.

Together with Exercise 3C.8, this yields 3.51(b): if C is m-by-c and R is c-by-n and $j \in \{1, \dots, m\}$, then row j of CR equals $C_{j,\cdot}R$ (by 3C.8), which by the present exercise is the linear combination of the rows of R with coefficients coming from row j of C.

Problem 3C.10. Give an example of 2-by-2 matrices A and B such that $AB \neq BA$.

Solution. Take

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Using the definition of matrix multiplication (3.41), the entry in row j, column k, of AB is $\sum_{r=1}^{2} A_{j,r}B_{r,k}$. Since the only nonzero entry of A is $A_{1,2} = 1$ and the only nonzero entry of B is $B_{2,1} = 1$, the only nonzero term arises from $j = 1$, $r = 2$, $k = 1$. Hence

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Similarly, the entry in row j, column k, of BA is $\sum_{r=1}^{2} B_{j,r}A_{r,k}$, and the only nonzero term arises from $j = 2$, $r = 1$, $k = 2$. Hence

$$BA = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The matrices AB and BA differ, for example, in the entry in row 1, column 1: it is 1 for AB and 0 for BA. Therefore $AB \neq BA$, showing that matrix multiplication is not commutative.
It is worth noting why such an example must exist: if A and B are the matrices (with respect to the standard bases of $\mathbf{F}^2$) of $S, T \in \mathcal{L}(\mathbf{F}^2)$, then $AB = \mathcal{M}(ST)$ and $BA = \mathcal{M}(TS)$ by 3.43, so a pair of noncommuting matrices is precisely a pair of linear maps on $\mathbf{F}^2$ with $ST \neq TS$. For the matrices above, S and T are the maps $S(x, y) = (y, 0)$ and $T(x, y) = (0, x)$, and indeed $ST(x, y) = (x, 0)$ while $TS(x, y) = (0, y)$.

Problem 3C.11. Prove that the distributive property holds for matrix addition and matrix multiplication. In other words, suppose A, B, C, D, E, and F are matrices whose sizes are such that $A(B + C)$ and $(D + E)F$ make sense. Explain why $AB + AC$ and $DF + EF$ both make sense and prove that

$$A(B + C) = AB + AC \qquad \text{and} \qquad (D + E)F = DF + EF.$$

Solution. The sizes. Suppose $A(B + C)$ makes sense. Then $B + C$ makes sense, so by the definition of matrix addition (3.34) the matrices B and C have the same size; say both are n-by-p. Then $B + C$ is also n-by-p. Since the product $A(B + C)$ is defined, the number of columns of A equals the number of rows of $B + C$, so A has n columns; say A is m-by-n. Now A is m-by-n and each of B, C is n-by-p, so both products AB and AC make sense and both are m-by-p matrices (by 3.41). Being of the same size, they can be added, so $AB + AC$ makes sense and is m-by-p, the same size as $A(B + C)$.
Similarly, suppose $(D + E)F$ makes sense. Then D and E have the same size, say both m-by-n, and $D + E$ is m-by-n; since $(D + E)F$ is defined, F has n rows, say F is n-by-p. Hence DF and EF both make sense and are both m-by-p, so $DF + EF$ makes sense and is m-by-p, the same size as $(D + E)F$.
The first identity. Both $A(B + C)$ and $AB + AC$ are m-by-p matrices, so it suffices to show that their entries in row j, column k, agree for all $j \in \{1, \dots, m\}$ and $k \in \{1, \dots, p\}$. Fix such j and k. Using the definition of matrix multiplication (3.41), then the definition of matrix addition (3.34) applied to $B + C$, then the distributive property of $\mathbf{F}$ together with the fact that a finite sum may be split, we

get

$$\begin{aligned}\big(A(B+C)\big)_{j,k} &= \sum_{r=1}^{n} A_{j,r}\,(B+C)_{r,k}\\ &= \sum_{r=1}^{n} A_{j,r}\,\big(B_{r,k}+C_{r,k}\big)\\ &= \sum_{r=1}^{n} \big(A_{j,r}B_{r,k}+A_{j,r}C_{r,k}\big)\\ &= \sum_{r=1}^{n} A_{j,r}B_{r,k} + \sum_{r=1}^{n} A_{j,r}C_{r,k}\\ &= (AB)_{j,k}+(AC)_{j,k}\\ &= (AB+AC)_{j,k},\end{aligned}$$

where the last two equalities use 3.41 and 3.34 again. Since j and k were arbitrary, $A(B+C) = AB+AC$. The second identity. Both $(D+E)F$ and $DF+EF$ are m-by-p matrices. Fix $j \in \{1,\dots,m\}$ and $k \in \{1,\dots,p\}$. Then

$$\begin{aligned}\big((D+E)F\big)_{j,k} &= \sum_{r=1}^{n} (D+E)_{j,r}\,F_{r,k}\\ &= \sum_{r=1}^{n} \big(D_{j,r}+E_{j,r}\big)F_{r,k}\\ &= \sum_{r=1}^{n} \big(D_{j,r}F_{r,k}+E_{j,r}F_{r,k}\big)\\ &= \sum_{r=1}^{n} D_{j,r}F_{r,k} + \sum_{r=1}^{n} E_{j,r}F_{r,k}\\ &= (DF)_{j,k}+(EF)_{j,k}\\ &= (DF+EF)_{j,k}.\end{aligned}$$

Since j and k were arbitrary, $(D+E)F = DF+EF$.
Note that both computations rest only on the distributive property of multiplication over addition in the field $\mathbf{F}$; that the two one-sided distributive laws must be proved separately reflects the fact that matrix multiplication is not commutative (Exercise 3C.10).

Problem 3C.12. Prove that matrix multiplication is associative. In other words, suppose A, B, and C are matrices whose sizes are such that $(AB)C$ makes sense. Explain why $A(BC)$ makes sense and prove that

$$(AB)C = A(BC).$$

[Try to find a clean proof that illustrates the following quote from Emil Artin: "It is my experience that proofs involving matrices can be shortened by 50% if one throws the matrices out."]

Solution. The sizes. Suppose $(AB)C$ makes sense. Then AB makes sense, so the number of columns of A equals the number of rows of B; say A is m-by-n and B is n-by-p, in which case AB is m-by-p (by 3.41). Since $(AB)C$ makes sense, C has as many rows as AB has columns, so C is p-by-q for some positive integer q, and $(AB)C$ is m-by-q. Now B is n-by-p and C is p-by-q, so BC makes sense and is n-by-p times p-by-q, hence n-by-q. Since A is m-by-n, the product $A(BC)$ makes sense and is m-by-q. Thus both $(AB)C$ and $A(BC)$ are defined and have the same size.
A proof with the matrices thrown out. Let $\mathbf{F}$ be the field containing the entries of A, B, C. Recall that a matrix determines a linear map between coordinate spaces: given an m-by-n matrix A, define

$T_A\colon \mathbf{F}^n \to \mathbf{F}^m$ by

$$T_A(x_1,\dots,x_n) = \Big(\sum_{r=1}^{n} A_{1,r}x_r,\ \dots,\ \sum_{r=1}^{n} A_{m,r}x_r\Big).$$

Each coordinate of T_A is a sum of scalar multiples of the coordinates of the input, so additivity and homogeneity of T_A follow at once from the corresponding identities $\sum_r A_{j,r}(x_r + y_r) = \sum_r A_{j,r}x_r + \sum_r A_{j,r}y_r$ and $\sum_r A_{j,r}(\lambda x_r) = \lambda \sum_r A_{j,r}x_r$ in $\mathbf{F}$; hence $T_A \in \mathcal{L}(\mathbf{F}^n, \mathbf{F}^m)$. Moreover, if $e_1,\dots,e_n$ is the standard basis of $\mathbf{F}^n$ and $f_1,\dots,f_m$ is the standard basis of $\mathbf{F}^m$, then

$$T_A e_k = (A_{1,k},\dots,A_{m,k}) = \sum_{j=1}^{m} A_{j,k} f_j,$$

so the kth column of $\mathcal{M}(T_A)$ is the kth column of A for each k, and therefore, by the definition of the matrix of a linear map (3.31),

$$\mathcal{M}(T_A) = A,$$

the matrix being taken with respect to the standard bases.
Now form $T_C \in \mathcal{L}(\mathbf{F}^q, \mathbf{F}^p)$, $T_B \in \mathcal{L}(\mathbf{F}^p, \mathbf{F}^n)$, and $T_A \in \mathcal{L}(\mathbf{F}^n, \mathbf{F}^m)$ as above, so that $\mathcal{M}(T_A) = A$, $\mathcal{M}(T_B) = B$, $\mathcal{M}(T_C) = C$, all with respect to standard bases. Applying 3.43 (the matrix of a product of linear maps is the product of the matrices) twice, and using standard bases throughout so that 3.43 applies as stated, we get

$$\mathcal{M}\big((T_A T_B)T_C\big) = \mathcal{M}(T_A T_B)\,\mathcal{M}(T_C) = \big(\mathcal{M}(T_A)\mathcal{M}(T_B)\big)C = (AB)C$$

and

$$\mathcal{M}\big(T_A(T_B T_C)\big) = \mathcal{M}(T_A)\,\mathcal{M}(T_B T_C) = A\big(\mathcal{M}(T_B)\mathcal{M}(T_C)\big) = A(BC).$$

But composition of functions is associative: for every $u \in \mathbf{F}^q$,

$$\big((T_A T_B)T_C\big)(u) = (T_A T_B)(T_C u) = T_A\big(T_B(T_C u)\big) = T_A\big((T_B T_C)(u)\big) = \big(T_A(T_B T_C)\big)(u).$$

Hence $(T_A T_B)T_C = T_A(T_B T_C)$ as linear maps from $\mathbf{F}^q$ to $\mathbf{F}^m$. Equal linear maps have equal matrices, so comparing the two displayed computations of matrices gives

$$(AB)C = A(BC).$$

This is the promised "throw the matrices out" argument: associativity of matrix multiplication is nothing but associativity of composition of functions, transported through 3.43.
A direct verification. For completeness, here is the computation with the matrices left in. Both $(AB)C$ and $A(BC)$ are m-by-q, so fix $j \in \{1,\dots,m\}$ and $k \in \{1,\dots,q\}$. By 3.41 applied twice, and then interchanging the order of the two finite sums (legitimate for finite sums in $\mathbf{F}$),

$$\begin{aligned}
\big((AB)C\big)_{j,k} &= \sum_{s=1}^{p} (AB)_{j,s}C_{s,k} = \sum_{s=1}^{p}\Big(\sum_{r=1}^{n} A_{j,r}B_{r,s}\Big)C_{s,k} = \sum_{s=1}^{p}\sum_{r=1}^{n} A_{j,r}B_{r,s}C_{s,k} \\
&= \sum_{r=1}^{n}\sum_{s=1}^{p} A_{j,r}B_{r,s}C_{s,k} = \sum_{r=1}^{n} A_{j,r}\Big(\sum_{s=1}^{p} B_{r,s}C_{s,k}\Big) = \sum_{r=1}^{n} A_{j,r}(BC)_{r,k} \\
&= \big(A(BC)\big)_{j,k}.
\end{aligned}$$

Since j and k were arbitrary, $(AB)C = A(BC)$.

Problem 3C.13. Suppose A is an n-by-n matrix and $1 \le j, k \le n$. Show that the entry in row j,

column k, of A^3 (which is defined to mean AAA) is

$$\sum_{p=1}^{n}\sum_{r=1}^{n} A_{j,p}A_{p,r}A_{r,k}.$$

Solution. Because A is n-by-n, all the products below are defined and every matrix appearing is n-by-n (by 3.41). Moreover, matrix multiplication is associative (Exercise 3C.12), so the expression AAA is unambiguous: $(AA)A = A(AA)$. We compute using the grouping $A^3 = (AA)A = A^2A$.
Fix j and k with $1 \le j, k \le n$. By the definition of matrix multiplication (3.41) applied to the product A^2A, with r as the summation index,

$$(A^3)_{j,k} = (A^2A)_{j,k} = \sum_{r=1}^{n} (A^2)_{j,r}\, A_{r,k}.$$

Applying 3.41 again to $A^2 = AA$, this time with p as the summation index, gives, for each $r \in \{1, \dots, n\}$,

$$(A^2)_{j,r} = \sum_{p=1}^{n} A_{j,p}A_{p,r}.$$

Substituting this into the previous display and using the distributive property in $\mathbf{F}$ to move the factor $A_{r,k}$ inside the inner sum, we obtain

$$(A^3)_{j,k} = \sum_{r=1}^{n}\Big(\sum_{p=1}^{n} A_{j,p}A_{p,r}\Big)A_{r,k} = \sum_{r=1}^{n}\sum_{p=1}^{n} A_{j,p}A_{p,r}A_{r,k}.$$

Finally, the two sums are finite sums of elements of $\mathbf{F}$, so their order of summation may be interchanged, giving

$$(A^3)_{j,k} = \sum_{p=1}^{n}\sum_{r=1}^{n} A_{j,p}A_{p,r}A_{r,k},$$

which is the desired formula.
The pattern generalizes in the obvious way: for a positive integer t and $1 \le j, k \le n$,

$$(A^t)_{j,k} = \sum_{r_1=1}^{n} \cdots \sum_{r_{t-1}=1}^{n} A_{j,r_1}A_{r_1,r_2}\cdots A_{r_{t-1},k},$$

as follows by induction on t using the same computation.

Problem 3C.14. Suppose m and n are positive integers. Prove that the function $A \mapsto A^{\mathrm{t}}$ is a linear map from $\mathbf{F}^{m,n}$ to $\mathbf{F}^{n,m}$.

Solution. Let $\Phi\colon \mathbf{F}^{m,n} \to \mathbf{F}^{n,m}$ be defined by $\Phi(A) = A^{\mathrm{t}}$.
First, Φ really does map $\mathbf{F}^{m,n}$ into $\mathbf{F}^{n,m}$: if A is an m-by-n matrix with entries in $\mathbf{F}$, then by the definition of the transpose (3.54) the matrix A^{t} is the n-by-m matrix with entries in $\mathbf{F}$ determined by

$$(A^{\mathrm{t}})_{k,j} = A_{j,k} \quad \text{for } 1 \le j \le m,\ 1 \le k \le n.$$

So $\Phi(A) \in \mathbf{F}^{n,m}$, and Φ is a well-defined function (the equation above prescribes each entry of A^{t} exactly once).
To show Φ is linear we verify additivity and homogeneity, using the definition of a linear map (3.1) and recalling that addition and scalar multiplication on $\mathbf{F}^{m,n}$ and $\mathbf{F}^{n,m}$ are entrywise (3.34 and 3.36).
Additivity. Suppose $A, B \in \mathbf{F}^{m,n}$. Then $A + B \in \mathbf{F}^{m,n}$, so $(A + B)^{\mathrm{t}}$ and $A^{\mathrm{t}} + B^{\mathrm{t}}$ are both n-by-m

matrices. Fix $j \in \{1, \dots, m\}$ and $k \in \{1, \dots, n\}$. Then

$$\big((A+B)^{\mathrm{t}}\big)_{k,j} = (A+B)_{j,k} = A_{j,k} + B_{j,k} = (A^{\mathrm{t}})_{k,j} + (B^{\mathrm{t}})_{k,j} = \big(A^{\mathrm{t}} + B^{\mathrm{t}}\big)_{k,j},$$

where the first and third equalities use 3.54 and the second and fourth use 3.34. Since every entry of an n-by-m matrix is of the form (row k, column j) with $1 \le k \le n$ and $1 \le j \le m$, this shows

$$\Phi(A+B) = (A+B)^{\mathrm{t}} = A^{\mathrm{t}} + B^{\mathrm{t}} = \Phi(A) + \Phi(B).$$

Homogeneity. Suppose $\lambda \in \mathbf{F}$ and $A \in \mathbf{F}^{m,n}$. Then $\lambda A \in \mathbf{F}^{m,n}$, and $(\lambda A)^{\mathrm{t}}$ and λA^{t} are both n-by-m matrices. Fix $j \in \{1, \dots, m\}$ and $k \in \{1, \dots, n\}$. Then

$$\big((\lambda A)^{\mathrm{t}}\big)_{k,j} = (\lambda A)_{j,k} = \lambda A_{j,k} = \lambda (A^{\mathrm{t}})_{k,j} = \big(\lambda A^{\mathrm{t}}\big)_{k,j},$$

where the first and third equalities use 3.54 and the second and fourth use 3.36. Hence

$$\Phi(\lambda A) = (\lambda A)^{\mathrm{t}} = \lambda A^{\mathrm{t}} = \lambda\, \Phi(A).$$

Since Φ is additive and homogeneous, $\Phi \in \mathcal{L}(\mathbf{F}^{m,n}, \mathbf{F}^{n,m})$, which is what we wanted to prove.
Two remarks. First, this establishes the identities $(A+B)^{\mathrm{t}} = A^{\mathrm{t}} + B^{\mathrm{t}}$ and $(\lambda A)^{\mathrm{t}} = \lambda A^{\mathrm{t}}$ quoted in the text just after 3.55. Second, Φ is in fact an isomorphism: applying 3.54 twice gives $\big((A^{\mathrm{t}})^{\mathrm{t}}\big)_{j,k} = (A^{\mathrm{t}})_{k,j} = A_{j,k}$, so $(A^{\mathrm{t}})^{\mathrm{t}} = A$ for every $A \in \mathbf{F}^{m,n}$. Writing $\Psi \colon \mathbf{F}^{n,m} \to \mathbf{F}^{m,n}$ for the transpose map in the other direction (linear by the same argument with m and n exchanged), we get $\Psi\Phi = I$ on $\mathbf{F}^{m,n}$ and $\Phi\Psi = I$ on $\mathbf{F}^{n,m}$, so Φ is invertible with $\Phi^{-1} = \Psi$.

Problem 3C.15. Prove that if A is an m-by-n matrix and C is an n-by-p matrix, then

$$(AC)^t = C^t A^t.$$

This exercise shows that the transpose of the product of two matrices is the product of the transposes in the opposite order.

Solution. First we check that both sides of the asserted equation are matrices of the same size, so that the equation has a chance of making sense. Because A is m-by-n and C is n-by-p, the product AC is defined and is an m-by-p matrix (by 3.41). Hence $(AC)^t$ is a p-by-m matrix (by 3.54). On the other side, C^t is a p-by-n matrix and A^t is an n-by-m matrix (both by 3.54); the number of columns of C^t equals the number of rows of A^t, so the product $C^t A^t$ is defined and is a p-by-m matrix. Thus $(AC)^t$ and $C^t A^t$ are both p-by-m matrices, and to prove they are equal we need only show that their entries agree in every position.
So suppose $k \in \{1, \dots, p\}$ and $j \in \{1, \dots, m\}$; we compute the entry in row k, column j, of each side. For the left side, the definition of the transpose (3.54) gives

$$\big((AC)^t\big)_{k,j} = (AC)_{j,k},$$

and the definition of matrix multiplication (3.41) gives

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

Hence

$$\big((AC)^t\big)_{k,j} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

For the right side, the definition of matrix multiplication (3.41), applied to the p-by-n matrix C^t and the n-by-m matrix A^t, gives

$$(C^t A^t)_{k,j} = \sum_{r=1}^{n} (C^t)_{k,r} (A^t)_{r,j}.$$

The definition of the transpose (3.54) tells us that $(C^t)_{k,r} = C_{r,k}$ and $(A^t)_{r,j} = A_{j,r}$. Substituting these into the equation above yields

$$(C^t A^t)_{k,j} = \sum_{r=1}^{n} C_{r,k} A_{j,r} = \sum_{r=1}^{n} A_{j,r} C_{r,k},$$

where the last equality holds because multiplication in $\mathbf{F}$ is commutative.
The two displayed computations produce the same number. Thus

$$\big((AC)^t\big)_{k,j} = (C^t A^t)_{k,j}$$

for every $k \in \{1, \dots, p\}$ and every $j \in \{1, \dots, m\}$. Because $(AC)^t$ and $C^t A^t$ have the same size and the same entry in every position, we conclude that $(AC)^t = C^t A^t$, as desired.
Note the role played by the order reversal: the sum defining $(AC)_{j,k}$ pairs row j of A with column k of C, and transposing turns these into column j of A^t and row k of C^t. Producing that same sum as a matrix product therefore requires taking row k of C^t times column j of A^t, which is exactly the entry in row k, column j, of $C^t A^t$. In general $A^t C^t$ is not even defined (its inner dimensions are m and p), which further explains why the order must be reversed.

Problem 3C.16. Suppose A is an m-by-n matrix with $A \neq 0$. Prove that the rank of A is 1 if and only if there exist $(c_1, \dots, c_m) \in \mathbf{F}^m$ and $(d_1, \dots, d_n) \in \mathbf{F}^n$ such that

$$A_{j,k} = c_j d_k$$

for every $j = 1, \dots, m$ and every $k = 1, \dots, n$.

Solution. Throughout, recall that the rank of A is by definition the column rank of A (by 3.58), which is the dimension of the span of the columns $A_{\cdot,1}, \dots, A_{\cdot,n}$ of A in $\mathbf{F}^{m,1}$ (by 3.52). Write

$$U = \operatorname{span}(A_{\cdot,1}, \dots, A_{\cdot,n}) \subseteq \mathbf{F}^{m,1},$$

so that the rank of A equals $\dim U$.
First suppose the rank of A is 1, so $\dim U = 1$. Every one-dimensional vector space has a basis consisting of a single nonzero vector, so there exists $c \in \mathbf{F}^{m,1}$ with $c \neq 0$ and $U = \operatorname{span}(c)$. Write

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix},$$

so $(c_1, \dots, c_m) \in \mathbf{F}^m$. Now let $k \in \{1, \dots, n\}$. The column $A_{\cdot,k}$ lies in $U = \operatorname{span}(c)$, so there exists $d_k \in \mathbf{F}$ with

$$A_{\cdot,k} = d_k c.$$

This defines $(d_1, \dots, d_n) \in \mathbf{F}^n$. Comparing the entry in row j of the two sides of the equation above, and recalling that the entry in row j of the m-by-1 matrix $A_{\cdot,k}$ is $A_{j,k}$ (by 3.44), we obtain

$$A_{j,k} = d_k c_j = c_j d_k$$

for every $j = 1, \dots, m$ and every $k = 1, \dots, n$, as desired.
Conversely, suppose there exist $(c_1, \dots, c_m) \in \mathbf{F}^m$ and $(d_1, \dots, d_n) \in \mathbf{F}^n$ with $A_{j,k} = c_j d_k$ for all j and k. Let $c \in \mathbf{F}^{m,1}$ be the m-by-1 matrix whose entry in row j is c_j. Fix $k \in \{1, \dots, n\}$. For each $j \in \{1, \dots, m\}$, the entry in row j of the column $A_{\cdot,k}$ equals $A_{j,k} = c_j d_k$, which is the entry in row j of $d_k c$. Because these two m-by-1 matrices agree in every entry,

$$A_{\cdot,k} = d_k c.$$

Thus every column of A lies in $\operatorname{span}(c)$, and hence

$$U = \operatorname{span}(A_{\cdot,1}, \dots, A_{\cdot,n}) \subseteq \operatorname{span}(c).$$

The list consisting of the single vector c spans $\operatorname{span}(c)$, and every spanning list can be reduced to a basis (by 2.30), so $\operatorname{span}(c)$ has a basis of length at most 1; hence $\dim \operatorname{span}(c) \leq 1$. Because U is a subspace of the finite-dimensional space $\operatorname{span}(c)$, we conclude that $\dim U \leq \dim \operatorname{span}(c) \leq 1$ (by 2.37). On the other hand, $A \neq 0$, so some entry $A_{j,k}$ of A is nonzero, and hence the column $A_{\cdot,k}$ is a nonzero element of U; thus $U \neq \{0\}$ and so $\dim U \geq 1$. Combining the two inequalities gives $\dim U = 1$, which says exactly that the rank of A is 1.
Two remarks. First, the hypothesis $A \neq 0$ is genuinely needed for the direction just proved: taking $c_1 = \cdots = c_m = 0$ and $d_1 = \cdots = d_n = 0$ produces $A_{j,k} = c_j d_k$ for the zero matrix, whose rank is 0 rather than 1.
Second, the condition $A_{j,k} = c_j d_k$ can be restated as a factorization. Let C be the m-by-1 matrix with $C_{j,1} = c_j$ and let R be the 1-by-n matrix with $R_{1,k} = d_k$. Then the definition of matrix multiplication (3.41) gives

$$(CR)_{j,k} = \sum_{r=1}^{1} C_{j,r} R_{r,k} = c_j d_k,$$

so the condition in the exercise says precisely that $A = CR$ for some m-by-1 matrix C and some 1-by-n matrix R. From this point of view the forward direction is exactly the column–row factorization 3.56 applied with column rank $c = 1$, and the reverse direction says that any nonzero matrix of the form CR, with C an m-by-1 matrix and R a 1-by-n matrix, has all of its columns inside the single one-dimensional subspace spanned by C.

Problem 3C.17. Suppose $T \in \mathcal{L}(V)$, and $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Prove that the following are equivalent.
(a) T is injective.
(b) The columns of $\mathcal{M}(T)$ are linearly independent in $\mathbf{F}^{n,1}$.
(c) The columns of $\mathcal{M}(T)$ span $\mathbf{F}^{n,1}$.
(d) The rows of $\mathcal{M}(T)$ span $\mathbf{F}^{1,n}$.
(e) The rows of $\mathcal{M}(T)$ are linearly independent in $\mathbf{F}^{1,n}$.
Here $\mathcal{M}(T)$ means $\mathcal{M}(T, (u_1, \dots, u_n), (v_1, \dots, v_n))$.

Solution. Write $A = \mathcal{M}(T) = \mathcal{M}(T, (u_1, \dots, u_n), (v_1, \dots, v_n))$. Because $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are both bases of V, we have $\dim V = n$ and A is an n-by-n matrix. By the definition of the matrix of a linear map (3.31), the entries of A are determined by

$$Tu_k = \sum_{j=1}^{n} A_{j,k} v_j \quad \text{for each } k = 1, \dots, n.$$

We will prove (a) $\iff$ (b), then (b) $\iff$ (c), then (c) $\iff$ (d), then (d) $\iff$ (e); this chain shows that all five statements are equivalent.
Step 0: a dictionary between V and $\mathbf{F}^{n,1}$.
Suppose

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{F}^{n,1}, \qquad \text{and set} \qquad u = x_1 u_1 + \cdots + x_n u_n \in V.$$

Using the linearity of T and the displayed formula for Tu_k, we compute

$$\begin{aligned} Tu &= \sum_{k=1}^{n} x_k \, Tu_k \\ &= \sum_{k=1}^{n} x_k \sum_{j=1}^{n} A_{j,k} v_j \\ &= \sum_{j=1}^{n} \Big(\sum_{k=1}^{n} A_{j,k} x_k \Big) v_j. \end{aligned}$$

Because A is n-by-n and x is n-by-1, the product Ax is an n-by-1 matrix whose entry in row j is $\sum_{k=1}^{n} A_{j,k} x_k$ (by the definition of matrix multiplication, 3.41); moreover $Ax = x_1 A_{\cdot,1} + \cdots + x_n A_{\cdot,n}$ (by 3.50). Hence the computation above says

$$Tu = \sum_{j=1}^{n} (Ax)_{j,1} \, v_j.$$

Because $v_1, \ldots, v_n$ is a basis of V and hence linearly independent, the vector Tu equals 0 if and only if all the coefficients $(Ax)_{j,1}$ equal 0, that is, if and only if

$$x_1 A_{\cdot,1} + \cdots + x_n A_{\cdot,n} = Ax = 0.$$

Similarly, because $u_1, \ldots, u_n$ is a basis of V and hence linearly independent, we have $u = 0$ if and only if $x = 0$. Finally, because $u_1, \ldots, u_n$ spans V, every $u \in V$ arises from some $x \in \mathbf{F}^{n,1}$ in this way. To summarize Step 0: the assignment $x \mapsto u = x_1 u_1 + \cdots + x_n u_n$ maps $\mathbf{F}^{n,1}$ onto V; for corresponding x and u we have $Tu = 0$ if and only if $Ax = 0$, and $u = 0$ if and only if $x = 0$.
Step 1: proof that (a) is equivalent to (b).
First suppose T is injective. Suppose $x_1, \ldots, x_n \in \mathbf{F}$ satisfy

$$x_1 A_{\cdot,1} + \cdots + x_n A_{\cdot,n} = 0.$$

Let $x \in \mathbf{F}^{n,1}$ be the column with entries $x_1, \ldots, x_n$ and let $u = x_1 u_1 + \cdots + x_n u_n$. By Step 0, $Tu = 0$, so $u \in \operatorname{null} T$. Because T is injective, $\operatorname{null} T = \{0\}$ (by 3.15), so $u = 0$; hence $x = 0$ by Step 0, meaning $x_1 = \cdots = x_n = 0$. Thus the only choice of scalars making the displayed linear combination of the columns of A equal 0 is the trivial one, so the columns of A are linearly independent in $\mathbf{F}^{n,1}$. This proves (b).
Conversely, suppose the columns of A are linearly independent. Let $u \in \operatorname{null} T$. Because $u_1, \ldots, u_n$ is a basis of V, we may write $u = x_1 u_1 + \cdots + x_n u_n$ for some $x_1, \ldots, x_n \in \mathbf{F}$; let $x \in \mathbf{F}^{n,1}$ be the corresponding column. By Step 0, $Tu = 0$ implies

$$x_1 A_{\cdot,1} + \cdots + x_n A_{\cdot,n} = 0.$$

Linear independence of the columns now forces $x_1 = \cdots = x_n = 0$, so $u = 0$. Hence $\operatorname{null} T = \{0\}$, and therefore T is injective (by 3.15). This proves (a).
Step 2: proof that (b) is equivalent to (c).
The columns $A_{\cdot,1}, \ldots, A_{\cdot,n}$ of A form a list of length n of vectors in $\mathbf{F}^{n,1}$, and $\dim \mathbf{F}^{n,1} = n$ (by 3.40). If this list is linearly independent, then because its length equals $\dim \mathbf{F}^{n,1}$ it is a basis of $\mathbf{F}^{n,1}$ (by 2.38); in particular it spans $\mathbf{F}^{n,1}$, giving (c). If instead this list spans $\mathbf{F}^{n,1}$, then because its length equals $\dim \mathbf{F}^{n,1}$ it is a basis of $\mathbf{F}^{n,1}$ (by 2.42); in particular it is linearly independent, giving (b). Thus (b) and (c) are equivalent, each being equivalent to the statement that the columns of A form a basis of $\mathbf{F}^{n,1}$.
Step 3: proof that (c) is equivalent to (d).
We show that each of (c) and (d) is equivalent to the statement that the rank of A equals n.
Let U denote the span of the columns of A in $\mathbf{F}^{n,1}$; by 3.52 and 3.58, $\dim U$ is the column rank of A, which is the rank of A. If the columns span $\mathbf{F}^{n,1}$, then $U = \mathbf{F}^{n,1}$, so $\dim U = n$ by 3.40. Conversely,

if $\dim U = n$, then U is a subspace of $\mathbf{F}^{n,1}$ whose dimension equals $\dim \mathbf{F}^{n,1}$, so $U = \mathbf{F}^{n,1}$ (by 2.39), meaning the columns span $\mathbf{F}^{n,1}$. Hence (c) holds if and only if the column rank of A equals n.
The identical argument with rows in place of columns, using $\dim \mathbf{F}^{1,n} = n$ (by 3.40), shows that (d) holds if and only if the row rank of A equals n: writing W for the span of the rows of A in $\mathbf{F}^{1,n}$, we have $\dim W$ equal to the row rank of A (by 3.52), and $W = \mathbf{F}^{1,n}$ if and only if $\dim W = n$ (by 2.39 for one direction and 3.40 for the other).
By 3.57, the column rank of A equals the row rank of A. Hence the column rank equals n if and only if the row rank equals n, and therefore (c) and (d) are equivalent.
Step 4: proof that (d) is equivalent to (e).
The rows $A_{1,\cdot}, \ldots, A_{n,\cdot}$ of A form a list of length n of vectors in $\mathbf{F}^{1,n}$, and $\dim \mathbf{F}^{1,n} = n$ (by 3.40). If this list spans $\mathbf{F}^{1,n}$, then it is a basis of $\mathbf{F}^{1,n}$ (by 2.42) and hence is linearly independent, giving (e). If this list is linearly independent, then it is a basis of $\mathbf{F}^{1,n}$ (by 2.38) and hence spans $\mathbf{F}^{1,n}$, giving (d). Thus (d) and (e) are equivalent.
Having proved (a) $\iff$ (b) $\iff$ (c) $\iff$ (d) $\iff$ (e), we conclude that all five statements are equivalent, completing the proof.
Two comments are worth making. First, nothing above required $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ to be the same basis: the two bases enter only through the formula $Tu_k = \sum_j A_{j,k} v_j$, and Step 0 used the first basis to convert vectors of V into columns and the second basis merely to detect when $Tu = 0$. Second, because T is an operator on the finite-dimensional space V, injectivity is equivalent to surjectivity (by 3.65), so each of (a) through (e) is also equivalent to the statement that T is surjective, and hence to the statement that T is invertible.

3.4 Exercises 3D

Problem 3D.1. Suppose $T \in \mathcal{L}(V, W)$ is invertible. Show that T^{-1} is invertible and

$$(T^{-1})^{-1} = T.$$

Solution. Recall the definition of invertibility (3.59): a linear map $T \in \mathcal{L}(V, W)$ is invertible if there exists $S \in \mathcal{L}(W, V)$ with $ST = I$ (the identity operator on V) and $TS = I$ (the identity operator on W); such an S is unique (3.60) and is denoted T^{-1}.
Because T is invertible, the map $T^{-1} \in \mathcal{L}(W, V)$ exists and satisfies

$$T^{-1}T = I_V \qquad \text{and} \qquad TT^{-1} = I_W.$$

Now apply the definition of invertibility to the linear map $T^{-1} \in \mathcal{L}(W, V)$. To show T^{-1} is invertible we must produce a linear map $R \in \mathcal{L}(V, W)$ with $RT^{-1} = I_W$ and $T^{-1}R = I_V$. Take $R = T$. Then $T \in \mathcal{L}(V, W)$ is linear, and the two displayed equations above say exactly that

$$TT^{-1} = I_W \qquad \text{and} \qquad T^{-1}T = I_V,$$

which are the required identities. Hence T^{-1} is invertible, and T is an inverse of T^{-1}.
Finally, by the uniqueness of the inverse (3.60), the inverse of T^{-1} is the unique linear map with these two properties, so

$$(T^{-1})^{-1} = T,$$

as desired.

Problem 3D.2. Suppose $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$ are both invertible linear maps. Prove that $ST \in \mathcal{L}(U, W)$ is invertible and that $(ST)^{-1} = T^{-1}S^{-1}$.

Solution. First, $ST \in \mathcal{L}(U, W)$: the composition of linear maps is linear (this is noted in 3.8 and is immediate from the definitions), and ST maps U into W because T maps U into V and S maps V into W.

Since S and T are invertible, we have $S^{-1} \in \mathcal{L}(W,V)$ and $T^{-1} \in \mathcal{L}(V,U)$, so the composition $T^{-1}S^{-1}$ is a linear map from W to U; that is, $T^{-1}S^{-1} \in \mathcal{L}(W,U)$, which is the right type of object to be an inverse of ST.
Now compute, using associativity of composition of functions. For the composition in one order,

$$(T^{-1}S^{-1})(ST) = T^{-1}(S^{-1}S)T = T^{-1}I_V T = T^{-1}T = I_U,$$

where we used $S^{-1}S = I_V$ and $T^{-1}T = I_U$. For the other order,

$$(ST)(T^{-1}S^{-1}) = S(TT^{-1})S^{-1} = SI_V S^{-1} = SS^{-1} = I_W,$$

where we used $TT^{-1} = I_V$ and $SS^{-1} = I_W$.
Thus $T^{-1}S^{-1}$ is a linear map from W to U whose composition with ST is the identity in both orders. By the definition of invertibility (3.59), ST is invertible, and by the uniqueness of the inverse (3.60),

$$(ST)^{-1} = T^{-1}S^{-1},$$

as desired.

Problem 3D.3. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that the following are equivalent.
(a) T is invertible.
(b) $Tv_1, \dots, Tv_n$ is a basis of V for every basis $v_1, \dots, v_n$ of V.
(c) $Tv_1, \dots, Tv_n$ is a basis of V for some basis $v_1, \dots, v_n$ of V.

Solution. Throughout, let $n = \dim V$. We prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a).
(a) $\Rightarrow$ (b). Suppose T is invertible and let $v_1, \dots, v_n$ be any basis of V. We show $Tv_1, \dots, Tv_n$ is linearly independent and spans V.
Linear independence: suppose $a_1, \dots, a_n \in \mathbf{F}$ satisfy $a_1Tv_1 + \cdots + a_nTv_n = 0$. By linearity this says $T(a_1v_1 + \cdots + a_nv_n) = 0$. Because T is invertible it is injective (3.63), so $\operatorname{null} T = \{0\}$ (3.15) and hence $a_1v_1 + \cdots + a_nv_n = 0$. Since $v_1, \dots, v_n$ is linearly independent, $a_1 = \cdots = a_n = 0$.
Spanning: let $v \in V$. Because T is invertible, $T^{-1}v \in V$, so there exist $a_1, \dots, a_n \in \mathbf{F}$ with $T^{-1}v = a_1v_1 + \cdots + a_nv_n$. Applying T and using linearity,

$$v = T(T^{-1}v) = a_1Tv_1 + \cdots + a_nTv_n.$$

Thus $\operatorname{span}(Tv_1, \dots, Tv_n) = V$.
Hence $Tv_1, \dots, Tv_n$ is a basis of V.
(b) $\Rightarrow$ (c). Because V is finite-dimensional, V has a basis $v_1, \dots, v_n$ (2.31). Applying (b) to this particular basis shows that $Tv_1, \dots, Tv_n$ is a basis of V, which is exactly statement (c). (If $V = \{0\}$, the basis is the empty list and the statement holds trivially.)
(c) $\Rightarrow$ (a). Suppose $v_1, \dots, v_n$ is a basis of V such that $Tv_1, \dots, Tv_n$ is a basis of V. Since $v_1, \dots, v_n$ spans V, every $v \in V$ has the form $v = a_1v_1 + \cdots + a_nv_n$, so

$$Tv = a_1Tv_1 + \cdots + a_nTv_n \in \operatorname{span}(Tv_1, \dots, Tv_n).$$

Conversely each Tv_k lies in $\operatorname{range} T$, so

$$\operatorname{range} T = \operatorname{span}(Tv_1, \dots, Tv_n)$$

(this is Exercise 10 in Section 3B). Because $Tv_1, \dots, Tv_n$ is a basis of V, its span is V, so $\operatorname{range} T = V$ and T is surjective. Since V is finite-dimensional and T maps V to itself, the domain and target have the same finite dimension, so injectivity is equivalent to surjectivity (3.65); thus T is also injective. Being both injective and surjective, T is invertible (3.63).
Having shown (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a), the three statements are equivalent.

Problem 3D.4. Suppose V is finite-dimensional and $\dim V > 1$. Prove that the set of noninvertible linear maps from V to itself is not a subspace of $\mathcal{L}(V)$.

Solution. Let $n = \dim V$, so $n \geq 2$, and let

$$\mathcal{N} = \{T \in \mathcal{L}(V) : T \text{ is not invertible}\}.$$

We show $\mathcal{N}$ is not closed under addition, hence is not a subspace (1.34).
Because V is finite-dimensional with $\dim V = n \geq 2$, there is a basis $v_1, \ldots, v_n$ of V (2.31), and it has at least two entries. By the linear map lemma (3.4) there exist unique $S, T \in \mathcal{L}(V)$ satisfying

$$Sv_1 = v_1, \quad Sv_k = 0 \text{ for } k = 2, \ldots, n,$$

$$Tv_1 = 0, \quad Tv_k = v_k \text{ for } k = 2, \ldots, n.$$

Neither S nor T is invertible. Indeed v_2 exists (because $n \geq 2$), $v_2 \neq 0$ (an entry of a basis is nonzero, since a list containing 0 is linearly dependent), and $Sv_2 = 0$; thus $\operatorname{null} S \neq \{0\}$, so S is not injective (3.15) and therefore not invertible (3.63). Likewise $v_1 \neq 0$ and $Tv_1 = 0$, so T is not invertible. Hence $S, T \in \mathcal{N}$.
However, $S + T$ agrees with the identity operator on the basis $v_1, \ldots, v_n$:

$$(S+T)v_1 = v_1 + 0 = v_1, \qquad (S+T)v_k = 0 + v_k = v_k \text{ for } k = 2, \ldots, n.$$

Two linear maps that agree on a basis are equal (each $v \in V$ is a linear combination of $v_1, \ldots, v_n$, and both maps are determined on such combinations by linearity). Therefore $S + T = I$, the identity operator on V, which is invertible (it is its own inverse). Thus $S + T \notin \mathcal{N}$.
So $\mathcal{N}$ contains S and T but not $S+T$, i.e. $\mathcal{N}$ is not closed under addition. A subspace must be closed under addition, so $\mathcal{N}$ is not a subspace of $\mathcal{L}(V)$.
Remark: the hypothesis $\dim V > 1$ is needed. If $\dim V = 1$, then a linear map $T \in \mathcal{L}(V)$ has the form $Tv = \lambda v$ for a fixed $\lambda \in \mathbf{F}$, and T fails to be invertible exactly when $\lambda = 0$; thus the noninvertible maps form $\{0\}$, which is a subspace. If $\dim V = 0$, then $\mathcal{L}(V) = \{0\}$ and 0 is invertible, so the set of noninvertible maps is empty and hence not a subspace either; but the intended content of the exercise is the case $\dim V \geq 2$ treated above.

Problem 3D.5. Suppose V is finite-dimensional, U is a subspace of V, and $S \in \mathcal{L}(U, V)$. Prove that there exists an invertible linear map T from V to itself such that $Tu = Su$ for every $u \in U$ if and only if S is injective.

Solution. Let $n = \dim V$.
First suppose such an invertible $T \in \mathcal{L}(V)$ exists, with $Tu = Su$ for every $u \in U$. Let $u \in U$ satisfy $Su = 0$. Then $Tu = Su = 0$. Because T is invertible it is injective (3.63), so $\operatorname{null} T = \{0\}$ (3.15) and hence $u = 0$. Thus $\operatorname{null} S = \{0\}$, so S is injective (3.15).
Conversely, suppose S is injective. Because U is a subspace of the finite-dimensional space V, U is finite-dimensional (2.25); let $m = \dim U$ and let $u_1, \ldots, u_m$ be a basis of U (2.31). Extend it to a basis

$$u_1, \ldots, u_m, v_1, \ldots, v_p$$

of V (2.32), where $m + p = n$.
Next, $Su_1, \ldots, Su_m$ is a linearly independent list in V: if $a_1, \ldots, a_m \in \mathbf{F}$ satisfy $a_1 Su_1 + \cdots + a_m Su_m = 0$, then $S(a_1 u_1 + \cdots + a_m u_m) = 0$ by linearity, so $a_1 u_1 + \cdots + a_m u_m \in \operatorname{null} S = \{0\}$ (using injectivity of S and 3.15), and linear independence of $u_1, \ldots, u_m$ forces $a_1 = \cdots = a_m = 0$.
Extend this linearly independent list to a basis of V (2.32):

$$Su_1, \ldots, Su_m, w_1, \ldots, w_q.$$

Every basis of V has length n (2.34), so $m+q=n=m+p$, giving $q=p$. Thus the two bases

$$u_1,\dots,u_m,v_1,\dots,v_p \qquad \text{and} \qquad Su_1,\dots,Su_m,w_1,\dots,w_p$$

have the same length, and we may define $T\in\mathcal{L}(V)$ by the linear map lemma (3.4) via

$$Tu_k=Su_k \ \ (k=1,\dots,m), \qquad Tv_j=w_j \ \ (j=1,\dots,p).$$

The map T sends the basis $u_1,\dots,u_m,v_1,\dots,v_p$ of V to the list $Su_1,\dots,Su_m,w_1,\dots,w_p$, which is a basis of V. By the equivalence (c) $\Rightarrow$ (a) of Exercise 3 in this section, T is invertible. (Directly: by Exercise 10 in Section 3B, $\operatorname{range}T$ is the span of the images of a basis, which here is the span of a basis of V, hence equals V; so T is surjective, hence injective by 3.65 since V is finite-dimensional, hence invertible by 3.63.)
Finally, T agrees with S on U. Let $u\in U$ and write $u=a_1u_1+\cdots+a_mu_m$ with $a_1,\dots,a_m\in\mathbf{F}$, possible because $u_1,\dots,u_m$ is a basis of U. Then

$$Tu=a_1Tu_1+\cdots+a_mTu_m=a_1Su_1+\cdots+a_mSu_m=S(a_1u_1+\cdots+a_mu_m)=Su,$$

using linearity of T and of S. Hence $Tu=Su$ for every $u\in U$, completing the proof.

Problem 3D.6. Suppose that W is finite-dimensional and $S,T\in\mathcal{L}(V,W)$. Prove that $\operatorname{null}S=\operatorname{null}T$ if and only if there exists an invertible $E\in\mathcal{L}(W)$ such that $S=ET$.

Solution. Note that V is not assumed to be finite-dimensional; only W is.
First suppose there is an invertible $E\in\mathcal{L}(W)$ with $S=ET$. If $v\in\operatorname{null}T$, then $Sv=E(Tv)=E0=0$, so $v\in\operatorname{null}S$. Conversely, if $v\in\operatorname{null}S$, then $E(Tv)=Sv=0$; since E is invertible it is injective (3.63), so $\operatorname{null}E=\{0\}$ (3.15) and hence $Tv=0$, i.e. $v\in\operatorname{null}T$. Therefore $\operatorname{null}S=\operatorname{null}T$.
Now suppose $\operatorname{null}S=\operatorname{null}T$. We build E.
Step 1: an isomorphism from $\operatorname{range}T$ onto $\operatorname{range}S$. Define $\varphi\colon \operatorname{range}T\to\operatorname{range}S$ by

$$\varphi(Tv)=Sv \qquad \text{for } v\in V.$$

This is well defined: every element of $\operatorname{range}T$ has the form Tv for some $v\in V$, and if $Tv_1=Tv_2$ then $v_1-v_2\in\operatorname{null}T=\operatorname{null}S$, so $Sv_1-Sv_2=S(v_1-v_2)=0$, i.e. $Sv_1=Sv_2$. Thus the value assigned to a point of $\operatorname{range}T$ does not depend on the choice of v representing it.
φ is linear: for $v_1,v_2\in V$ and $\lambda\in\mathbf{F}$,

$$\varphi(Tv_1+Tv_2)=\varphi(T(v_1+v_2))=S(v_1+v_2)=Sv_1+Sv_2=\varphi(Tv_1)+\varphi(Tv_2),$$

$$\varphi(\lambda Tv_1)=\varphi(T(\lambda v_1))=S(\lambda v_1)=\lambda Sv_1=\lambda\varphi(Tv_1).$$

φ is surjective onto $\operatorname{range}S$, since every element of $\operatorname{range}S$ is $Sv=\varphi(Tv)$ for some $v\in V$. And φ is injective: if $\varphi(Tv)=0$ then $Sv=0$, so $v\in\operatorname{null}S=\operatorname{null}T$, so $Tv=0$. Hence $\operatorname{null}\varphi=\{0\}$ and φ is injective (3.15).
Therefore φ is an isomorphism from $\operatorname{range}T$ onto $\operatorname{range}S$ (3.63, 3.69). Both spaces are subspaces of the finite-dimensional space W, hence finite-dimensional (2.25), and $\dim\operatorname{range}T=\dim\operatorname{range}S$ (3.70).
Step 2: construction of E. Let $m=\dim\operatorname{range}T$ and $n=\dim W$, and let $u_1,\dots,u_m$ be a basis of $\operatorname{range}T$ (2.31). Extend it to a basis

$$u_1,\dots,u_m,u_{m+1},\dots,u_n$$

of W (2.32).
The list $\varphi u_1,\dots,\varphi u_m$ is a basis of $\operatorname{range}S$: it spans $\operatorname{range}S$ because φ is surjective onto $\operatorname{range}S$ and $u_1,\dots,u_m$ spans $\operatorname{range}T$ (so every element of $\operatorname{range}S$ is φ of a linear combination of the u_k, hence a linear combination of the φu_k); and it is linearly independent because if $\sum_{k=1}^m a_k\varphi u_k=0$ then $\varphi\big(\sum_k a_ku_k\big)=0$, so $\sum_k a_ku_k=0$ by injectivity of φ, so all $a_k=0$.

Extend $\varphi u_1, \dots, \varphi u_m$ to a basis of W (2.32):

$$\varphi u_1, \dots, \varphi u_m, y_{m+1}, \dots, y_n,$$

where the number of adjoined vectors is $n - m$ because every basis of W has length n (2.34).
By the linear map lemma (3.4) there is a unique $E \in \mathcal{L}(W)$ with

$$Eu_k = \varphi u_k \ \ (k = 1, \dots, m), \qquad Eu_k = y_k \ \ (k = m+1, \dots, n).$$

Then E maps the basis $u_1, \dots, u_n$ of W onto the basis $\varphi u_1, \dots, \varphi u_m, y_{m+1}, \dots, y_n$ of W, so E is invertible by Exercise 3 in this section, part (c) $\Rightarrow$ (a). (Directly: $\operatorname{range} E$ is the span of the images of a basis (Exercise 10 in Section 3B), which is the span of a basis of W, hence equals W; so E is surjective, hence injective by 3.65 since W is finite-dimensional, hence invertible by 3.63.)
Step 3: $S = ET$. Let $v \in V$. Then $Tv \in \operatorname{range} T$, so $Tv = c_1 u_1 + \cdots + c_m u_m$ for some $c_1, \dots, c_m \in \mathbf{F}$. Hence

$$E(Tv) = c_1 Eu_1 + \cdots + c_m Eu_m = c_1 \varphi u_1 + \cdots + c_m \varphi u_m = \varphi(c_1 u_1 + \cdots + c_m u_m) = \varphi(Tv) = Sv,$$

using linearity of E, the definition of E on $u_1, \dots, u_m$, linearity of φ, and the definition of φ. Since $v \in V$ was arbitrary, $S = ET$ with $E \in \mathcal{L}(W)$ invertible, as required.

Problem 3D.7. Suppose that V is finite-dimensional and $S, T \in \mathcal{L}(V, W)$. Prove that $\operatorname{range} S = \operatorname{range} T$ if and only if there exists an invertible $E \in \mathcal{L}(V)$ such that $S = TE$.

Solution. Here V is finite-dimensional; W need not be. Let $n = \dim V$.
First suppose there is an invertible $E \in \mathcal{L}(V)$ with $S = TE$. Then

$$\operatorname{range} S = \{T(Ev) : v \in V\} = \{Tu : u \in \operatorname{range} E\} = \{Tu : u \in V\} = \operatorname{range} T,$$

where the third equality holds because E is invertible, hence surjective (3.63), so $\operatorname{range} E = V$.
Conversely, suppose $\operatorname{range} S = \operatorname{range} T$.
Step 1: the two null spaces have the same dimension. Since V is finite-dimensional, the fundamental theorem of linear maps (3.21) applied to S and to T gives

$$\dim \operatorname{null} S = \dim V - \dim \operatorname{range} S = \dim V - \dim \operatorname{range} T = \dim \operatorname{null} T.$$

Call this common dimension m, and set $r = n - m$. (All the spaces involved are finite-dimensional: $\operatorname{null} S$ and $\operatorname{null} T$ are subspaces of V, and $\operatorname{range} S$, $\operatorname{range} T$ are finite-dimensional by 3.21.)
Step 2: adapted bases. Let $u_1, \dots, u_m$ be a basis of $\operatorname{null} S$ (2.31), and extend it to a basis

$$u_1, \dots, u_m, v_1, \dots, v_r$$

of V (2.32); the number of adjoined vectors is $n - m = r$ because every basis of V has length n (2.34).
Claim: $Sv_1, \dots, Sv_r$ is a basis of $\operatorname{range} S$. Spanning: for $v \in V$ write $v = \sum_{k=1}^{m} b_k u_k + \sum_{j=1}^{r} a_j v_j$; then, since $Su_k = 0$ for each k,

$$Sv = a_1 Sv_1 + \cdots + a_r Sv_r,$$

so $\operatorname{range} S = \operatorname{span}(Sv_1, \dots, Sv_r)$. Linear independence: if $\sum_{j=1}^{r} a_j Sv_j = 0$, then $\sum_j a_j v_j \in \operatorname{null} S = \operatorname{span}(u_1, \dots, u_m)$, so $\sum_j a_j v_j = \sum_k b_k u_k$ for some scalars b_k; that is,

$$b_1 u_1 + \cdots + b_m u_m - a_1 v_1 - \cdots - a_r v_r = 0.$$

Since $u_1, \dots, u_m, v_1, \dots, v_r$ is linearly independent, all coefficients vanish; in particular $a_1 = \cdots = a_r = 0$. This proves the claim.
Now let $p_1, \dots, p_m$ be a basis of $\operatorname{null} T$; such a basis exists because $\operatorname{null} T$ is a subspace of the finite-dimensional space V and hence finite-dimensional (2.25, 2.31), and by Step 1 it indeed has length $\dim \operatorname{null} T = m$.

Step 3: choice of preimages. For each $j \in \{1, \dots, r\}$ we have $Sv_j \in \operatorname{range} S = \operatorname{range} T$, so there exists $q_j \in V$ with

$$Tq_j = Sv_j.$$

Step 4: the list $p_1, \dots, p_m, q_1, \dots, q_r$ is a basis of V. Its length is $m + r = n = \dim V$, so by 2.38 it suffices to prove linear independence. Suppose

$$b_1 p_1 + \cdots + b_m p_m + a_1 q_1 + \cdots + a_r q_r = 0.$$

Apply T. Since $Tp_k = 0$ for each k (the p_k lie in $\operatorname{null} T$) and $Tq_j = Sv_j$, we get

$$a_1 Sv_1 + \cdots + a_r Sv_r = 0.$$

By the claim in Step 2, $Sv_1, \dots, Sv_r$ is linearly independent, so $a_1 = \cdots = a_r = 0$. The displayed equation then reduces to $b_1 p_1 + \cdots + b_m p_m = 0$, and linear independence of the basis $p_1, \dots, p_m$ of $\operatorname{null} T$ gives $b_1 = \cdots = b_m = 0$. Hence the list is linearly independent and therefore a basis of V.
Step 5: construction of E and verification. By the linear map lemma (3.4) there is a unique $E \in \mathcal{L}(V)$ with

$$Eu_k = p_k \quad (k = 1, \dots, m), \qquad Ev_j = q_j \quad (j = 1, \dots, r),$$

since $u_1, \dots, u_m, v_1, \dots, v_r$ is a basis of V. This E carries that basis onto the list $p_1, \dots, p_m, q_1, \dots, q_r$, which by Step 4 is a basis of V; hence E is invertible by Exercise 3 in this section, part (c) $\Rightarrow$ (a). (Directly: $\operatorname{range} E$ is the span of the images of a basis (Exercise 10 in Section 3B), which is the span of a basis of V, hence equals V; so E is surjective, hence injective by 3.65 since V is finite-dimensional, hence invertible by 3.63.)
Finally, TE and S agree on the basis $u_1, \dots, u_m, v_1, \dots, v_r$ of V:

$$(TE)u_k = Tp_k = 0 = Su_k \quad (k = 1, \dots, m), \qquad (TE)v_j = Tq_j = Sv_j \quad (j = 1, \dots, r),$$

using $p_k \in \operatorname{null} T$, $u_k \in \operatorname{null} S$, and Step 3. Two linear maps that agree on a basis are equal, because every vector of V is a linear combination of that basis and both maps are linear. Therefore $S = TE$ with $E \in \mathcal{L}(V)$ invertible, completing the proof.

Problem 3D.8. Suppose V and W are finite-dimensional and $S, T \in \mathcal{L}(V, W)$. Prove that there exist invertible $E_1 \in \mathcal{L}(V)$ and $E_2 \in \mathcal{L}(W)$ such that $S = E_2 T E_1$ if and only if $\dim \operatorname{null} S = \dim \operatorname{null} T$.

Solution. First suppose there exist invertible $E_1 \in \mathcal{L}(V)$ and $E_2 \in \mathcal{L}(W)$ with $S = E_2 T E_1$. If $v \in V$, then

$$Sv = 0 \iff E_2(TE_1 v) = 0 \iff TE_1 v = 0 \iff E_1 v \in \operatorname{null} T,$$

where the middle equivalence holds because E_2 is injective (by 3.63) and $E_2 0 = 0$. Thus

$$\operatorname{null} S = \{v \in V : E_1 v \in \operatorname{null} T\} = \{E_1^{-1} u : u \in \operatorname{null} T\}.$$

Hence the restriction of E_1^{-1} to $\operatorname{null} T$ is a linear map from $\operatorname{null} T$ onto $\operatorname{null} S$; it is injective because E_1^{-1} is injective. So this restriction is an isomorphism from $\operatorname{null} T$ onto $\operatorname{null} S$ (by 3.63). Both $\operatorname{null} S$ and $\operatorname{null} T$ are subspaces of the finite-dimensional vector space V and hence are finite-dimensional (by 2.25), so 3.70 applies and gives $\dim \operatorname{null} S = \dim \operatorname{null} T$.
Now suppose $\dim \operatorname{null} S = \dim \operatorname{null} T$. Write $n = \dim \operatorname{null} S = \dim \operatorname{null} T$ and $m = \dim V$. By the fundamental theorem of linear maps (3.21),

$$\dim \operatorname{range} S = m - n = \dim \operatorname{range} T.$$

Set $k = m - n$.
Choose a basis $u_1, \dots, u_n$ of $\operatorname{null} S$ and extend it (by 2.32) to a basis

$$u_1, \dots, u_n, x_1, \dots, x_k$$

of V. Similarly choose a basis $p_1, \dots, p_n$ of $\operatorname{null} T$ and extend it to a basis $p_1, \dots, p_n, y_1, \dots, y_k$ of V. The list $Sx_1, \dots, Sx_k$ spans $\operatorname{range} S$, because $\operatorname{range} S$ is spanned by $Su_1, \dots, Su_n, Sx_1, \dots, Sx_k$ and each $Su_j = 0$. Since this spanning list has length $k = \dim \operatorname{range} S$, it is a basis of $\operatorname{range} S$ (by 2.42). For the same reason $Ty_1, \dots, Ty_k$ is a basis of $\operatorname{range} T$.
Extend $Sx_1, \dots, Sx_k$ to a basis

$$Sx_1, \dots, Sx_k, w_1, \dots, w_r$$

of W, where $r = \dim W - k$, and extend $Ty_1, \dots, Ty_k$ to a basis $Ty_1, \dots, Ty_k, z_1, \dots, z_r$ of W (both extensions use 2.32, and both lists have the same length because both start with k vectors).
By the linear map lemma (3.4) there is a unique $E_1 \in \mathcal{L}(V)$ with

$$E_1 u_j = p_j \quad (j = 1, \dots, n), \qquad E_1 x_i = y_i \quad (i = 1, \dots, k),$$

and a unique $E_2 \in \mathcal{L}(W)$ with

$$E_2(Ty_i) = Sx_i \quad (i = 1, \dots, k), \qquad E_2 z_j = w_j \quad (j = 1, \dots, r).$$

Each of E_1, E_2 maps a basis onto a basis, so each is surjective (its range contains a spanning list of the target space) and hence invertible by 3.65.
Finally, S and $E_2 T E_1$ agree on the basis $u_1, \dots, u_n, x_1, \dots, x_k$ of V: for each j,

$$E_2 T E_1 u_j = E_2 T p_j = E_2 0 = 0 = S u_j$$

since $p_j \in \operatorname{null} T$ and $u_j \in \operatorname{null} S$; and for each i,

$$E_2 T E_1 x_i = E_2 T y_i = S x_i.$$

Two linear maps that agree on a basis of V are equal, so $S = E_2 T E_1$, as desired.

Problem 3D.9. Suppose V is finite-dimensional and $T\colon V \to W$ is a surjective linear map of V onto W. Prove that there is a subspace U of V such that $T|_U$ is an isomorphism of U onto W.
Here $T|_U$ means the function T restricted to U. Thus $T|_U$ is the function whose domain is U, with $T|_U$ defined by $T|_U(u) = Tu$ for every $u \in U$.

Solution. Because V is finite-dimensional and $\operatorname{null} T$ is a subspace of V, there is a subspace U of V such that

$$V = \operatorname{null} T \oplus U$$

by 2.33.
The function $T|_U$ is a linear map from U to W: it satisfies additivity and homogeneity because T does, and U is closed under addition and scalar multiplication.
$T|_U$ is injective. Suppose $u \in U$ and $T|_U(u) = 0$. Then $Tu = 0$, so $u \in \operatorname{null} T$. Thus $u \in \operatorname{null} T \cap U = \{0\}$, where the last equality holds because the sum $\operatorname{null} T + U$ is a direct sum (by 1.46). Hence $u = 0$, and so $\operatorname{null}(T|_U) = \{0\}$, which means $T|_U$ is injective (by 3.15).
$T|_U$ is surjective. Suppose $w \in W$. Because T is surjective, there exists $v \in V$ with $Tv = w$. Write $v = n + u$ with $n \in \operatorname{null} T$ and $u \in U$. Then

$$w = Tv = Tn + Tu = 0 + Tu = T|_U(u).$$

Thus $\operatorname{range}(T|_U) = W$.
Because $T|_U$ is injective and surjective, it is invertible (by 3.63), which is to say $T|_U$ is an isomorphism of U onto W.

Problem 3D.10. Suppose V and W are finite-dimensional and U is a subspace of V. Let

$$\mathcal{E} = \{T \in \mathcal{L}(V, W) : U \subseteq \operatorname{null} T\}.$$

(a) Show that $\mathcal{E}$ is a subspace of $\mathcal{L}(V, W)$.
(b) Find a formula for $\dim \mathcal{E}$ in terms of $\dim V$, $\dim W$, and $\dim U$.
Hint: Define $\Phi\colon \mathcal{L}(V, W) \to \mathcal{L}(U, W)$ by $\Phi(T) = T|_U$. What is $\operatorname{null} \Phi$? What is $\operatorname{range} \Phi$?

Solution. (a) The zero map $0 \in \mathcal{L}(V, W)$ satisfies $\operatorname{null} 0 = V \supseteq U$, so $0 \in \mathcal{E}$; in particular $\mathcal{E}$ is nonempty.
Suppose $S, T \in \mathcal{E}$. If $u \in U$, then $(S + T)u = Su + Tu = 0 + 0 = 0$. Thus $U \subseteq \operatorname{null}(S + T)$, so $S + T \in \mathcal{E}$.
Suppose $T \in \mathcal{E}$ and $\lambda \in \mathbf{F}$. If $u \in U$, then $(\lambda T)u = \lambda(Tu) = \lambda 0 = 0$. Thus $\lambda T \in \mathcal{E}$.
Because $\mathcal{E}$ contains the additive identity of $\mathcal{L}(V, W)$ and is closed under addition and scalar multiplication, $\mathcal{E}$ is a subspace of $\mathcal{L}(V, W)$ (by 1.34).
(b) The formula is

$$\dim \mathcal{E} = (\dim V - \dim U)(\dim W).$$

To prove this, note first that U is finite-dimensional (by 2.25, since U is a subspace of the finite-dimensional vector space V). Following the hint, define

$$\Phi\colon \mathcal{L}(V, W) \to \mathcal{L}(U, W) \quad \text{by} \quad \Phi(T) = T|_U.$$

This makes sense: the restriction of a linear map $V \to W$ to the subspace U is a linear map $U \to W$. Furthermore Φ is linear, because for all $S, T \in \mathcal{L}(V, W)$, all $\lambda \in \mathbf{F}$, and all $u \in U$ we have

$$(S + T)|_U(u) = Su + Tu = \big(S|_U + T|_U\big)(u), \qquad (\lambda T)|_U(u) = \lambda\big(T|_U(u)\big),$$

so $\Phi(S + T) = \Phi(S) + \Phi(T)$ and $\Phi(\lambda T) = \lambda\Phi(T)$.
The null space of Φ is $\mathcal{E}$. Indeed, $\Phi(T) = 0$ means $Tu = 0$ for every $u \in U$, which says exactly that $U \subseteq \operatorname{null} T$. Thus $\operatorname{null} \Phi = \mathcal{E}$.
The range of Φ is all of $\mathcal{L}(U, W)$. To see this, let $R \in \mathcal{L}(U, W)$. Choose a basis $u_1, \dots, u_m$ of U and extend it to a basis $u_1, \dots, u_m, v_1, \dots, v_n$ of V (by 2.32). By the linear map lemma (3.4) there exists $T \in \mathcal{L}(V, W)$ with

$$Tu_j = Ru_j \quad (j = 1, \dots, m), \qquad Tv_k = 0 \quad (k = 1, \dots, n).$$

The linear maps $T|_U$ and R agree on the basis $u_1, \dots, u_m$ of U, hence they are equal. Thus $\Phi(T) = R$, proving that Φ is surjective. (The degenerate cases cause no trouble: if $U = \{0\}$ then $m = 0$ and the list $u_1, \dots, u_m$ is empty, and $T = 0$ works; if $U = V$ then $n = 0$ and $T = R$ works.)
Because $\mathcal{L}(V, W)$ is finite-dimensional (by 3.72), the fundamental theorem of linear maps (3.21) applied to Φ gives

$$\dim \mathcal{L}(V, W) = \dim \operatorname{null} \Phi + \dim \operatorname{range} \Phi = \dim \mathcal{E} + \dim \mathcal{L}(U, W).$$

Now 3.72 evaluates the two outer terms: $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$ and $\dim \mathcal{L}(U, W) = (\dim U)(\dim W)$. Therefore

$$\dim \mathcal{E} = (\dim V)(\dim W) - (\dim U)(\dim W) = (\dim V - \dim U)(\dim W),$$

as claimed.

Problem 3D.11. Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that

$$ST \text{ is invertible} \iff S \text{ and } T \text{ are invertible.}$$

Solution. First suppose S and T are invertible. Then ST is invertible with $(ST)^{-1} = T^{-1}S^{-1}$, because

$$(ST)(T^{-1}S^{-1}) = S(TT^{-1})S^{-1} = SS^{-1} = I$$

and

$$(T^{-1}S^{-1})(ST) = T^{-1}(S^{-1}S)T = T^{-1}T = I.$$

(This is Exercise 2 in this section.) Note that this direction does not require V to be finite-dimensional. Now suppose ST is invertible. We show separately that S is surjective and T is injective, and then invoke 3.65.
S is surjective. By 3.63 the invertible map ST is surjective, so $\operatorname{range}(ST) = V$. But every vector of the form $(ST)v = S(Tv)$ lies in $\operatorname{range} S$, so

$$V = \operatorname{range}(ST) \subseteq \operatorname{range} S \subseteq V.$$

Hence $\operatorname{range} S = V$, so S is surjective. Because V is finite-dimensional and $S \in \mathcal{L}(V)$, surjectivity implies invertibility (by 3.65). Thus S is invertible.
T is injective. By 3.63 the invertible map ST is injective. Suppose $v \in V$ and $Tv = 0$. Then $(ST)v = S(Tv) = S0 = 0$, so $v = 0$ by the injectivity of ST (using 3.15). Hence $\operatorname{null} T = \{0\}$, so T is injective (by 3.15). Because V is finite-dimensional and $T \in \mathcal{L}(V)$, injectivity implies invertibility (by 3.65). Thus T is invertible.
This completes the proof of both implications.
Finite-dimensionality is genuinely needed for the implication ST invertible $\implies$ S and T invertible: on $V = \mathbf{F}^\infty$, let S be the backward shift $S(x_1, x_2, \dots) = (x_2, x_3, \dots)$ and let T be the forward shift $T(x_1, x_2, \dots) = (0, x_1, x_2, \dots)$. Then $ST = I$ is invertible, but neither S (not injective) nor T (not surjective) is invertible.

Problem 3D.12. Suppose V is finite-dimensional and $S, T, U \in \mathcal{L}(V)$ and $STU = I$. Show that T is invertible and that $T^{-1} = US$.

Solution. First we show that U is invertible. Suppose $v \in V$ and $Uv = 0$. Then

$$v = Iv = (STU)v = S\big(T(Uv)\big) = S(T0) = 0.$$

Thus $\operatorname{null} U = \{0\}$, so U is injective (by 3.15). Because V is finite-dimensional and $U \in \mathcal{L}(V)$, this implies U is invertible (by 3.65).
Next we show that S is invertible. Every vector of the form $(STU)v = S\big((TU)v\big)$ lies in $\operatorname{range} S$, so

$$V = \operatorname{range} I = \operatorname{range}(STU) \subseteq \operatorname{range} S \subseteq V,$$

whence $\operatorname{range} S = V$. Thus S is surjective, and therefore invertible (by 3.65).
Now multiply the equation $STU = I$ on the left by S^{-1} and on the right by U^{-1}:

$$T = S^{-1}(STU)U^{-1} = S^{-1}IU^{-1} = S^{-1}U^{-1}.$$

Because S^{-1} and U^{-1} are invertible (Exercise 1 in this section), their product $S^{-1}U^{-1}$ is invertible (Exercise 2 in this section). Hence T is invertible, and

$$T^{-1} = (S^{-1}U^{-1})^{-1} = (U^{-1})^{-1}(S^{-1})^{-1} = US,$$

using Exercise 2 for the inverse of a product and Exercise 1 for $(A^{-1})^{-1} = A$.
As a direct check: $T(US) = S^{-1}U^{-1}US = S^{-1}S = I$ and $(US)T = USS^{-1}U^{-1} = UU^{-1} = I$.

Problem 3D.13. Show that the result in Exercise 12 can fail without the hypothesis that V is finite-dimensional.

Solution. Take $V = \mathbf{F}^\infty$, which is not finite-dimensional. Let $B \in \mathcal{L}(V)$ be the backward shift

$$B(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots)$$

and let $F \in \mathcal{L}(V)$ be the forward shift

$$F(x_1, x_2, x_3, \dots) = (0, x_1, x_2, \dots).$$

Both are linear (see Example 3.3).
Now set

$$S = B, \qquad T = B, \qquad U = F^2.$$

Then $STU = B^2F^2$, and for every $(x_1, x_2, \dots) \in \mathbf{F}^\infty$,

$$B^2F^2(x_1, x_2, \dots) = B^2(0, 0, x_1, x_2, \dots) = (x_1, x_2, \dots).$$

Thus $STU = I$.
However $T = B$ is not invertible: $B(1, 0, 0, \dots) = (0, 0, \dots)$, so $\operatorname{null} B \neq \{0\}$ and B is not injective (by 3.15), hence not invertible (by 3.63). So the conclusion of Exercise 12 fails badly here.
The formula $T^{-1} = US$ fails too, in the sense that US is not a two-sided inverse of T: we have

$$US = F^2B \quad \text{and} \quad T(US) = BF^2B = FB,$$

because $BF = I$. But $FB(1, 0, 0, \dots) = F(0, 0, \dots) = (0, 0, \dots) \neq (1, 0, 0, \dots)$, so $T(US) \neq I$.
For a version in which T is invertible but the formula still fails, take $S = B$, $T = I$, $U = F$. Then $STU = BIF = BF = I$ and $T^{-1} = I$, yet

$$US(x_1, x_2, x_3, \dots) = F(x_2, x_3, \dots) = (0, x_2, x_3, \dots),$$

so $US \neq I = T^{-1}$.

Problem 3D.14. Prove or give a counterexample: If V is a finite-dimensional vector space and $R, S, T \in \mathcal{L}(V)$ are such that RST is surjective, then S is injective.

Solution. The statement is true; here is a proof.
Suppose RST is surjective, so $\operatorname{range}(RST) = V$.
Step 1: R is invertible. Every vector of the form $(RST)v = R\big((ST)v\big)$ lies in $\operatorname{range} R$, so

$$V = \operatorname{range}(RST) \subseteq \operatorname{range} R \subseteq V.$$

Hence $\operatorname{range} R = V$, so R is surjective. Because V is finite-dimensional and $R \in \mathcal{L}(V)$, this implies R is invertible (by 3.65); in particular R is injective (by 3.63).
Step 2: ST is surjective. Let $v \in V$. Because RST is surjective, there exists $u \in V$ with

$$(RST)u = Rv.$$

That is, $R\big((ST)u\big) = Rv$. Since R is injective, $(ST)u = v$. Thus $v \in \operatorname{range}(ST)$, and so $\operatorname{range}(ST) = V$.
Step 3: S is surjective. Every vector of the form $(ST)u = S(Tu)$ lies in $\operatorname{range} S$, so

$$V = \operatorname{range}(ST) \subseteq \operatorname{range} S \subseteq V,$$

giving $\operatorname{range} S = V$.
Step 4: S is injective. Because V is finite-dimensional and $S \in \mathcal{L}(V)$ is surjective, 3.65 gives that S is invertible and hence injective.
Therefore S is injective, as claimed.
Remark: the finite-dimensional hypothesis is essential. On $V = \mathbf{F}^\infty$ let B be the backward shift and F the forward shift. Take $R = S = B$ and $T = F^2$. Then $RST = B^2F^2 = I$, which is surjective, but $S = B$ is not injective, since $B(1, 0, 0, \dots) = 0$.

Problem 3D.15. Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_m$ is a list in V such that $Tv_1, \dots, Tv_m$ spans V. Prove that $v_1, \dots, v_m$ spans V.

Solution. First note that V is finite-dimensional: the list $Tv_1, \dots, Tv_m$ is a finite list of vectors in V that spans V, which is exactly the definition of V being finite-dimensional.
Next, T is surjective. Indeed, each Tv_k lies in $\operatorname{range} T$, and $\operatorname{range} T$ is a subspace of V (by 3.18), so

$$V = \operatorname{span}(Tv_1, \dots, Tv_m) \subseteq \operatorname{range} T \subseteq V.$$

Hence $\operatorname{range} T = V$.
Now apply 3.65 with $W = V$ (the hypothesis $\dim V = \dim W < \infty$ holds trivially): since T is surjective, T is injective.
Finally we show $\operatorname{span}(v_1, \dots, v_m) = V$. The inclusion $\operatorname{span}(v_1, \dots, v_m) \subseteq V$ is immediate. For the reverse inclusion, let $v \in V$. Then $Tv \in V$, so because $Tv_1, \dots, Tv_m$ spans V there exist $a_1, \dots, a_m \in \mathbf{F}$ with

$$Tv = a_1 Tv_1 + \cdots + a_m Tv_m = T(a_1 v_1 + \cdots + a_m v_m),$$

where the second equality uses the linearity of T. Because T is injective, this gives

$$v = a_1 v_1 + \cdots + a_m v_m \in \operatorname{span}(v_1, \dots, v_m).$$

Thus $V \subseteq \operatorname{span}(v_1, \dots, v_m)$, and therefore $v_1, \dots, v_m$ spans V.

Problem 3D.16. Prove that every linear map from $\mathbf{F}^{n,1}$ to $\mathbf{F}^{m,1}$ is given by a matrix multiplication. In other words, prove that if $T \in \mathcal{L}(\mathbf{F}^{n,1}, \mathbf{F}^{m,1})$, then there exists an m-by-n matrix A such that $Tx = Ax$ for every $x \in \mathbf{F}^{n,1}$.

Solution. For $k = 1, \dots, n$ let $e_k \in \mathbf{F}^{n,1}$ denote the n-by-1 column vector whose k^{th} entry is 1 and whose other entries are 0. The list $e_1, \dots, e_n$ is a basis of $\mathbf{F}^{n,1}$; indeed, for

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{F}^{n,1}$$

we have $x = x_1 e_1 + \cdots + x_n e_n$, and this representation is clearly unique because the coefficients are recovered as the entries of x.
Now suppose $T \in \mathcal{L}(\mathbf{F}^{n,1}, \mathbf{F}^{m,1})$. Each Te_k is an element of $\mathbf{F}^{m,1}$, that is, an m-by-1 column vector. Define an m-by-n matrix A by declaring its k^{th} column to be Te_k; explicitly,

$$A_{j,k} = (Te_k)_j \qquad \text{for } j = 1, \dots, m \text{ and } k = 1, \dots, n,$$

where $(Te_k)_j$ denotes the entry in row j of the column vector Te_k. Thus $A_{\cdot,k} = Te_k$ for each k.
Let $x \in \mathbf{F}^{n,1}$ have entries $x_1, \dots, x_n$. Using $x = x_1 e_1 + \cdots + x_n e_n$ and the linearity of T,

$$Tx = x_1 Te_1 + \cdots + x_n Te_n = x_1 A_{\cdot,1} + \cdots + x_n A_{\cdot,n}.$$

On the other hand, 3.50 (linear combination of columns) states precisely that

$$Ax = x_1 A_{\cdot,1} + \cdots + x_n A_{\cdot,n}.$$

Comparing the two displayed equations gives $Tx = Ax$ for every $x \in \mathbf{F}^{n,1}$, as desired.
Remark: this is the concrete instance of 3.76, which says $\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v)$. Here we use the standard bases $e_1, \dots, e_n$ of $\mathbf{F}^{n,1}$ and the analogous standard basis of $\mathbf{F}^{m,1}$; with respect to these bases $\mathcal{M}(v) = v$ for every column vector v, so 3.76 reads $Tx = \mathcal{M}(T)x$, and $A = \mathcal{M}(T)$ is the matrix constructed above.

Problem 3D.17. Suppose V is finite-dimensional and $S \in \mathcal{L}(V)$. Define $\mathcal{A} \in \mathcal{L}(\mathcal{L}(V))$ by

$$\mathcal{A}(T) = ST$$

for $T \in \mathcal{L}(V)$.
(a) Show that $\dim \operatorname{null} \mathcal{A} = (\dim V)(\dim \operatorname{null} S)$.
(b) Show that $\dim \operatorname{range} \mathcal{A} = (\dim V)(\dim \operatorname{range} S)$.

Solution. Write $n = \dim V$. First note that $\mathcal{A}$ really is a linear map from $\mathcal{L}(V)$ to $\mathcal{L}(V)$: if $T_1, T_2 \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then for every $v \in V$ we have $\big(S(T_1 + T_2)\big)v = S(T_1 v + T_2 v) = ST_1 v + ST_2 v$ and $\big(S(\lambda T_1)\big)v = S(\lambda T_1 v) = \lambda S T_1 v$, using the linearity of S; hence $\mathcal{A}(T_1 + T_2) = \mathcal{A}(T_1) + \mathcal{A}(T_2)$ and $\mathcal{A}(\lambda T_1) = \lambda \mathcal{A}(T_1)$. Also, by 3.72 the space $\mathcal{L}(V)$ is finite-dimensional with

$$\dim \mathcal{L}(V) = (\dim V)(\dim V) = n^2.$$

(a) Let $U = \operatorname{null} S$, a subspace of V; because V is finite-dimensional, so is U (by 2.25). I claim

$$\operatorname{null} \mathcal{A} = \{T \in \mathcal{L}(V) : \operatorname{range} T \subseteq U\}.$$

Indeed, $T \in \operatorname{null} \mathcal{A}$ means $ST = 0$, which means $S(Tv) = 0$ for every $v \in V$, which means $Tv \in \operatorname{null} S = U$ for every $v \in V$, which means $\operatorname{range} T \subseteq U$.
Now define a map

$$\Phi : \mathcal{L}(V, U) \to \mathcal{L}(V)$$

by letting $\Phi(R)$ be the linear map from V to V given by $\big(\Phi(R)\big)(v) = Rv$ for $v \in V$; in other words, $\Phi(R)$ is R with its target space enlarged from U to V. Since $U \subseteq V$, this makes sense, and $\Phi(R)$ is linear because R is. The map Φ is linear, since $\Phi(R_1 + R_2)$ and $\Phi(R_1) + \Phi(R_2)$ agree at every $v \in V$, and similarly for scalar multiplication.
Φ is injective: if $\Phi(R) = 0$, then $Rv = 0$ for every $v \in V$, so $R = 0$.
$\operatorname{range} \Phi = \operatorname{null} \mathcal{A}$: for $R \in \mathcal{L}(V, U)$ we have $\operatorname{range} \Phi(R) \subseteq U$, so $\Phi(R) \in \operatorname{null} \mathcal{A}$ by the claim above. Conversely, if $T \in \operatorname{null} \mathcal{A}$, then $\operatorname{range} T \subseteq U$, so the formula $Rv = Tv$ defines an element $R \in \mathcal{L}(V, U)$, and $\Phi(R) = T$.
Thus Φ is an isomorphism from $\mathcal{L}(V, U)$ onto $\operatorname{null} \mathcal{A}$, so these two spaces have the same dimension (by 3.70, or directly because an isomorphism carries a basis to a basis). Applying 3.72 to $\mathcal{L}(V, U)$, which is legitimate because V and U are both finite-dimensional, gives

$$\dim \operatorname{null} \mathcal{A} = \dim \mathcal{L}(V, U) = (\dim V)(\dim U) = (\dim V)(\dim \operatorname{null} S),$$

as desired.
(b) Apply the fundamental theorem of linear maps (3.21) to $\mathcal{A} \in \mathcal{L}(\mathcal{L}(V))$, which is legitimate because $\mathcal{L}(V)$ is finite-dimensional:

$$\dim \mathcal{L}(V) = \dim \operatorname{null} \mathcal{A} + \dim \operatorname{range} \mathcal{A}.$$

Using $\dim \mathcal{L}(V) = n^2$ and part (a),

$$\dim \operatorname{range} \mathcal{A} = n^2 - n\,(\dim \operatorname{null} S) = n\big(n - \dim \operatorname{null} S\big).$$

Applying 3.21 to $S \in \mathcal{L}(V)$ gives $n = \dim \operatorname{null} S + \dim \operatorname{range} S$, so $n - \dim \operatorname{null} S = \dim \operatorname{range} S$. Substituting,

$$\dim \operatorname{range} \mathcal{A} = (\dim V)(\dim \operatorname{range} S),$$

as desired.

Problem 3D.18. Show that V and $\mathcal{L}(\mathbf{F}, V)$ are isomorphic vector spaces.

Solution. Here $\mathbf{F}$ is regarded as a vector space over $\mathbf{F}$ in the usual way. We exhibit an explicit isomorphism; no finite-dimensionality is needed.
Define $\Phi : \mathcal{L}(\mathbf{F}, V) \to V$ by

$$\Phi(T) = T1$$

for $T \in \mathcal{L}(\mathbf{F}, V)$, where 1 denotes the multiplicative identity of $\mathbf{F}$, viewed as a vector in the vector space $\mathbf{F}$.
Φ is linear. If $T, R \in \mathcal{L}(\mathbf{F}, V)$ then, by the definition of addition in $\mathcal{L}(\mathbf{F}, V)$,

$$\Phi(T + R) = (T + R)(1) = T1 + R1 = \Phi(T) + \Phi(R),$$

and if $\lambda \in \mathbf{F}$ then, by the definition of scalar multiplication in $\mathcal{L}(\mathbf{F}, V)$,

$$\Phi(\lambda T) = (\lambda T)(1) = \lambda(T1) = \lambda\Phi(T).$$

Φ is injective. Suppose $\Phi(T) = 0$, that is, $T1 = 0$. Then for every $\lambda \in \mathbf{F}$,

$$T\lambda = T(\lambda \cdot 1) = \lambda\, T1 = \lambda \cdot 0 = 0,$$

using the homogeneity of T. Hence $T = 0$, so $\operatorname{null}\Phi = \{0\}$ and Φ is injective by 3.15.
Φ is surjective. Let $v \in V$ and define $T_v : \mathbf{F} \to V$ by $T_v\lambda = \lambda v$. Then T_v is linear, because for $\lambda, \mu, \alpha \in \mathbf{F}$,

$$T_v(\lambda + \mu) = (\lambda + \mu)v = \lambda v + \mu v = T_v\lambda + T_v\mu$$

and

$$T_v(\alpha\lambda) = (\alpha\lambda)v = \alpha(\lambda v) = \alpha\, T_v\lambda,$$

where we used the vector space axioms for V. Thus $T_v \in \mathcal{L}(\mathbf{F}, V)$ and $\Phi(T_v) = T_v1 = 1 \cdot v = v$. Hence $\operatorname{range}\Phi = V$.
Therefore Φ is an invertible linear map from $\mathcal{L}(\mathbf{F}, V)$ onto V (by 3.63, a linear map is invertible if and only if it is injective and surjective), that is, Φ is an isomorphism in the sense of 3.69. Consequently V and $\mathcal{L}(\mathbf{F}, V)$ are isomorphic vector spaces.
Remark: when V is finite-dimensional one can argue more quickly. By 3.72,

$$\dim \mathcal{L}(\mathbf{F}, V) = (\dim \mathbf{F})(\dim V) = 1 \cdot \dim V = \dim V,$$

so 3.70 gives the isomorphism. The argument above has the advantage of also covering infinite-dimensional V, and of producing an explicit isomorphism.

Problem 3D.19. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T has the same matrix with respect to every basis of V if and only if T is a scalar multiple of the identity operator.

Solution. Throughout, $\mathcal{M}(T, (v_1, \ldots, v_n))$ denotes the matrix of T with respect to the basis $v_1, \ldots, v_n$ used for both the domain and the target, as in Section 3D. Recall that this matrix A is determined by

$$Tv_k = A_{1,k}v_1 + \cdots + A_{n,k}v_n \qquad \text{for } k = 1, \ldots, n.$$

Let $n = \dim V$. If $n = 0$ then $\mathcal{L}(V) = \{0\}$, so $T = 0 = 0 \cdot I$ and T has the same (empty) matrix with respect to the only basis of V; both sides of the equivalence hold. So assume $n \geq 1$.
Suppose first that $T = \lambda I$ for some $\lambda \in \mathbf{F}$. Let $v_1, \ldots, v_n$ be any basis of V. Then

$$Tv_k = \lambda v_k = 0 \cdot v_1 + \cdots + 0 \cdot v_{k-1} + \lambda v_k + 0 \cdot v_{k+1} + \cdots + 0 \cdot v_n,$$

so the k^{th} column of $\mathcal{M}(T)$ has λ in row k and 0 elsewhere. Hence $\mathcal{M}(T, (v_1, \ldots, v_n)) = \lambda I$, the n-by-n diagonal matrix with λ on the diagonal. This matrix does not depend on the chosen basis, so T has the same matrix with respect to every basis of V.
Conversely, suppose there is an n-by-n matrix A with $\mathcal{M}(T, (w_1, \ldots, w_n)) = A$ for every basis $w_1, \ldots, w_n$ of V.

Step 1: A is a diagonal matrix. Fix a basis $v_1, \dots, v_n$ of V and fix $k \in \{1, \dots, n\}$. Let

$$u_k = 2v_k, \qquad u_i = v_i \ \text{ for } i \neq k.$$

Then $u_1, \dots, u_n$ is again a basis of V: it spans V because $v_k = \frac{1}{2}u_k$, and it is linearly independent because a dependence relation among the u_i becomes one among the v_i after replacing u_k by $2v_k$. (Here we use that $2 \neq 0$ in $\mathbf{F}$, which holds since $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$.)
Reading off column k of A with respect to the basis $v_1, \dots, v_n$ gives

$$Tv_k = \sum_{j=1}^{n} A_{j,k} v_j,$$

while reading off column k of A with respect to the basis $u_1, \dots, u_n$ gives

$$2\,Tv_k = Tu_k = \sum_{j=1}^{n} A_{j,k} u_j = 2A_{k,k} v_k + \sum_{j \neq k} A_{j,k} v_j.$$

Multiplying the first display by 2 and subtracting the second display yields

$$\sum_{j \neq k} A_{j,k} v_j = 0.$$

Since $v_1, \dots, v_n$ is linearly independent, $A_{j,k} = 0$ for every $j \neq k$. As k was arbitrary, A is diagonal; write $A_{k,k} = \lambda_k$, so $A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$.
Note the consequence: because A is the matrix of T with respect to every basis, for every basis $w_1, \dots, w_n$ of V and every k we have

$$Tw_k = \lambda_k w_k.$$

Step 2: $\lambda_1 = \lambda_2 = \cdots = \lambda_n$. If $n = 1$ there is nothing to prove, so suppose $n \geq 2$ and fix $k \in \{2, \dots, n\}$. Keep the basis $v_1, \dots, v_n$ and let $w_1, \dots, w_n$ be the list obtained from it by interchanging v_1 and v_k; reordering a basis clearly yields a basis, since spanning and linear independence do not depend on the order of the list. Since $w_1 = v_k$, the observation at the end of Step 1 applied to the basis $w_1, \dots, w_n$ gives

$$Tv_k = Tw_1 = \lambda_1 w_1 = \lambda_1 v_k,$$

while the same observation applied to the basis $v_1, \dots, v_n$ gives $Tv_k = \lambda_k v_k$. Subtracting,

$$(\lambda_1 - \lambda_k) v_k = 0.$$

A vector in a linearly independent list is nonzero, so $v_k \neq 0$ and therefore $\lambda_k = \lambda_1$.
Set $\lambda = \lambda_1$. By Steps 1 and 2, $Tv_k = \lambda v_k = (\lambda I) v_k$ for $k = 1, \dots, n$. Two linear maps that agree on each vector of a basis agree everywhere: if $v = a_1 v_1 + \cdots + a_n v_n$, then

$$Tv = a_1 Tv_1 + \cdots + a_n Tv_n = a_1 \lambda v_1 + \cdots + a_n \lambda v_n = \lambda v = (\lambda I) v.$$

Hence $T = \lambda I$, a scalar multiple of the identity operator.

Problem 3D.20. Suppose $q \in \mathcal{P}(\mathbf{R})$. Prove that there exists a polynomial $p \in \mathcal{P}(\mathbf{R})$ such that

$$q(x) = (x^2 + x) p''(x) + 2x\, p'(x) + p(3)$$

for all $x \in \mathbf{R}$.

Solution. This follows the pattern of Example 3.67: we produce an operator on a suitable $\mathcal{P}_m(\mathbf{R})$, show it is injective, and then invoke 3.65 to conclude that it is surjective.

Fix a nonnegative integer m and define $T : \mathcal{P}_m(\mathbf{R}) \to \mathcal{P}_m(\mathbf{R})$ by

$$(Tp)(x) = (x^2 + x)p''(x) + 2x\,p'(x) + p(3).$$

T maps $\mathcal{P}_m(\mathbf{R})$ into $\mathcal{P}_m(\mathbf{R})$. If $\deg p \le m$, then $\deg p'' \le m - 2$ (interpreting this as $p'' = 0$ when $m \le 1$), so $(x^2 + x)p''$ has degree at most m; likewise $\deg p' \le m - 1$, so $2xp'$ has degree at most m; and $p(3)$ is a constant. Hence $Tp \in \mathcal{P}_m(\mathbf{R})$.
T is linear. Differentiation is linear, evaluation at 3 is linear, and multiplication by a fixed polynomial is linear; so for $p, r \in \mathcal{P}_m(\mathbf{R})$ and $\lambda \in \mathbf{R}$,

$$\big(T(p + r)\big)(x) = (x^2 + x)\big(p''(x) + r''(x)\big) + 2x\big(p'(x) + r'(x)\big) + p(3) + r(3) = (Tp)(x) + (Tr)(x),$$

and similarly $T(\lambda p) = \lambda Tp$. Thus $T \in \mathcal{L}(\mathcal{P}_m(\mathbf{R}))$.
T is injective. Suppose $p \in \mathcal{P}_m(\mathbf{R})$ and $p \ne 0$; we show $Tp \ne 0$. Let $d = \deg p$ and write

$$p(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_0, \qquad a_d \ne 0.$$

If $d = 0$, then p is the nonzero constant a_0, so $p' = p'' = 0$ and $(Tp)(x) = p(3) = a_0 \ne 0$.
Now suppose $d \ge 1$. We compute the coefficient of x^d in Tp. First,

$$(x^2 + x)p''(x) = d(d - 1)a_d x^d + (\text{terms of degree} < d).$$

When $d = 1$ this is the assertion that $(x^2 + x)p''(x) = 0$, which is true because $p'' = 0$ for a polynomial of degree 1, and indeed $d(d - 1) = 0$ in that case. When $d \ge 2$ we have $p''(x) = d(d - 1)a_d x^{d-2} +$ (terms of degree $< d - 2$); multiplying by x^2 produces $d(d - 1)a_d x^d$ plus terms of degree less than d, and multiplying by x produces only terms of degree at most $d - 1$. Similarly

$$p'(x) = d\,a_d x^{d-1} + (\text{terms of degree} < d - 1), \quad \text{so} \quad 2x\,p'(x) = 2d\,a_d x^d + (\text{terms of degree} < d).$$

Finally $p(3)$ contributes only to the constant term, and $d \ge 1$. Hence the coefficient of x^d in Tp equals

$$d(d - 1)a_d + 2d\,a_d = a_d\big(d^2 - d + 2d\big) = a_d\,d(d + 1).$$

Because $a_d \ne 0$ and $d \ge 1$, we have $a_d\,d(d + 1) \ne 0$, so $Tp \ne 0$.
Thus $\operatorname{null} T = \{0\}$, so T is injective by 3.15.
T is surjective. The space $\mathcal{P}_m(\mathbf{R})$ is finite-dimensional (the list $1, x, \ldots, x^m$ is a basis of it), and T is an operator on it, so trivially the domain and target have equal finite dimension. By 3.65, injectivity of T implies that T is surjective.
Conclusion. Let $q \in \mathcal{P}(\mathbf{R})$. Choose a nonnegative integer m with $q \in \mathcal{P}_m(\mathbf{R})$ (take $m = \deg q$ if $q \ne 0$, and $m = 0$ if $q = 0$). By the surjectivity of the corresponding T, there exists $p \in \mathcal{P}_m(\mathbf{R}) \subseteq \mathcal{P}(\mathbf{R})$ with $Tp = q$, that is,

$$q(x) = (x^2 + x)p''(x) + 2x\,p'(x) + p(3)$$

for all $x \in \mathbf{R}$, as desired.

Problem 3D.21. Suppose n is a positive integer and $A_{j,k} \in \mathbf{F}$ for all $j, k = 1, \ldots, n$. Prove that the following are equivalent (note that in both parts below, the number of equations equals the number of variables).
(a) The trivial solution $x_1 = \cdots = x_n = 0$ is the only solution to the homogeneous system of

equations

$$\sum_{k=1}^{n} A_{1,k}x_k = 0$$
$$\vdots$$
$$\sum_{k=1}^{n} A_{n,k}x_k = 0.$$

(b) For every $c_1, \dots, c_n \in \mathbf{F}$, there exists a solution to the system of equations

$$\sum_{k=1}^{n} A_{1,k}x_k = c_1$$
$$\vdots$$
$$\sum_{k=1}^{n} A_{n,k}x_k = c_n.$$

Solution. Define $T : \mathbf{F}^n \to \mathbf{F}^n$ by

$$T(x_1, \dots, x_n) = \Big(\sum_{k=1}^{n} A_{1,k}x_k,\ \dots,\ \sum_{k=1}^{n} A_{n,k}x_k \Big).$$

T is linear. For $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbf{F}^n$ and $j \in \{1, \dots, n\}$, the j^{th} coordinate of $T(x + y)$ is

$$\sum_{k=1}^{n} A_{j,k}(x_k + y_k) = \sum_{k=1}^{n} A_{j,k}x_k + \sum_{k=1}^{n} A_{j,k}y_k,$$

which is the j^{th} coordinate of $Tx + Ty$; hence $T(x + y) = Tx + Ty$. Similarly, for $\lambda \in \mathbf{F}$ the j^{th} coordinate of $T(\lambda x)$ is $\sum_{k=1}^{n} A_{j,k}(\lambda x_k) = \lambda \sum_{k=1}^{n} A_{j,k}x_k$, so $T(\lambda x) = \lambda Tx$. Thus $T \in \mathcal{L}(\mathbf{F}^n)$.
Reformulation of (a). A list $x_1, \dots, x_n$ of scalars solves the homogeneous system in (a) exactly when $T(x_1, \dots, x_n) = (0, \dots, 0)$, that is, exactly when $(x_1, \dots, x_n) \in \operatorname{null} T$. Hence (a) says $\operatorname{null} T = \{0\}$, which by 3.15 is equivalent to T being injective.
Reformulation of (b). Given $c_1, \dots, c_n \in \mathbf{F}$, the system in (b) has a solution exactly when there is some $x \in \mathbf{F}^n$ with $Tx = (c_1, \dots, c_n)$, that is, exactly when $(c_1, \dots, c_n) \in \operatorname{range} T$. Hence (b) says $\operatorname{range} T = \mathbf{F}^n$, which is exactly the statement that T is surjective (3.19).
Now $\mathbf{F}^n$ is finite-dimensional with $\dim \mathbf{F}^n = n$, and T maps $\mathbf{F}^n$ to $\mathbf{F}^n$, so the hypotheses of 3.65 are satisfied (with $V = W = \mathbf{F}^n$). That result states

$$T \text{ is invertible} \iff T \text{ is injective} \iff T \text{ is surjective.}$$

In particular T is injective if and only if T is surjective. Combining this with the two reformulations above, (a) holds if and only if (b) holds.
Remark: the hypothesis that the number of equations equals the number of variables is essential; it is what makes the domain and target of T have the same dimension, which is required for 3.65. For example, the single equation $x_1 + x_2 = 0$ in two variables has nontrivial solutions yet every right-hand side is attained, while the system $x_1 = c_1$, $x_1 = c_2$ in one variable has only the trivial solution when $c_1 = c_2 = 0$ but is unsolvable when $c_1 \neq c_2$.

Problem 3D.22. Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of V. Prove that

$$\mathcal{M}\big(T, (v_1, \dots, v_n)\big) \text{ is invertible} \iff T \text{ is invertible.}$$

Solution. Because V has a basis of length n, the space V is finite-dimensional with $\dim V = n$. Throughout this solution every matrix of an operator on V is taken with respect to the single basis $v_1, \dots, v_n$, so that we may write $\mathcal{M}(S)$ for $\mathcal{M}(S, (v_1, \dots, v_n))$ when $S \in \mathcal{L}(V)$. Write

$$A = \mathcal{M}(T) = \mathcal{M}\big(T, (v_1, \dots, v_n)\big).$$

Two facts will be used repeatedly. First, the matrix of the identity operator with respect to a single basis is the identity matrix: indeed $Iv_k = v_k$ for each k, so column k of $\mathcal{M}(I)$ has a 1 in row k and 0 elsewhere, which says exactly that $\mathcal{M}(I) = I$ (see 3.79 and the discussion following it). Second, applying 3.81 with $U = V = W$ and with the basis $v_1, \dots, v_n$ used in all three roles gives

$$\mathcal{M}(SR) = \mathcal{M}(S)\,\mathcal{M}(R) \qquad \text{for all } S, R \in \mathcal{L}(V).$$

Suppose first that T is invertible. Then $T^{-1} \in \mathcal{L}(V)$ and $TT^{-1} = T^{-1}T = I$. Taking matrices and using the two facts above,

$$A\,\mathcal{M}(T^{-1}) = \mathcal{M}(T)\,\mathcal{M}(T^{-1}) = \mathcal{M}(TT^{-1}) = \mathcal{M}(I) = I$$

and likewise

$$\mathcal{M}(T^{-1})\,A = \mathcal{M}(T^{-1}T) = \mathcal{M}(I) = I.$$

Because $\mathcal{M}(T^{-1})$ is a square matrix of the same size as A, the definition 3.80 shows that A is invertible, with $A^{-1} = \mathcal{M}(T^{-1})$. (Incidentally, this is exactly the statement 3.86, whose proof the book leaves as an exercise.)

Conversely, suppose A is invertible, and let $B = A^{-1}$, so that B is an n-by-n matrix with $AB = BA = I$. By 3.71 the map

$$\mathcal{M}\colon \mathcal{L}(V) \to \mathbf{F}^{n,n}$$

is an isomorphism (apply 3.71 with $W = V$ and with $v_1, \dots, v_n$ used as the basis of both the domain and the target). In particular $\mathcal{M}$ is surjective, so there exists $S \in \mathcal{L}(V)$ with $\mathcal{M}(S) = B$. Now

$$\mathcal{M}(ST) = \mathcal{M}(S)\,\mathcal{M}(T) = BA = I = \mathcal{M}(I)$$

and

$$\mathcal{M}(TS) = \mathcal{M}(T)\,\mathcal{M}(S) = AB = I = \mathcal{M}(I).$$

Because $\mathcal{M}$ is also injective (again by 3.71), the two displayed equations give $ST = I$ and $TS = I$. Hence T is invertible (with $T^{-1} = S$), by the definition of invertibility.

If one prefers not to invoke injectivity of $\mathcal{M}$ twice, note that only one of the two products is needed: from $\mathcal{M}(ST) = \mathcal{M}(I)$ we get $ST = I$, and then 3.68 (applied with $W = V$, legitimate since V is finite-dimensional) yields $TS = I$ as well.

Combining the two directions gives $\mathcal{M}(T, (v_1, \dots, v_n))$ invertible $\iff$ T invertible, as desired.

Problem 3D.23. Suppose that $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Let $T \in \mathcal{L}(V)$ be such that $Tv_k = u_k$ for each $k = 1, \dots, n$. Prove that

$$\mathcal{M}\big(T, (v_1, \dots, v_n)\big) = \mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big).$$

Solution. First note that such a T exists and is unique: by the linear map lemma (3.4) there is exactly one $T \in \mathcal{L}(V)$ with $Tv_k = u_k$ for each k, since $v_1, \dots, v_n$ is a basis of V. So the statement to be proved concerns a well-defined operator.

Both sides of the asserted equation are n-by-n matrices, so it suffices to show that corresponding entries agree. Let

$$A = \mathcal{M}\big(T, (v_1, \dots, v_n)\big) = \mathcal{M}\big(T, (v_1, \dots, v_n), (v_1, \dots, v_n)\big),$$
$$C = \mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big).$$

By the definition of the matrix of a linear map (3.31), applied to T with the basis $v_1, \dots, v_n$ of the

domain and the same basis $v_1, \dots, v_n$ of the target, the entries $A_{j,k}$ of A are the unique scalars satisfying

$$Tv_k = A_{1,k}v_1 + \cdots + A_{n,k}v_n \qquad \text{for each } k = 1, \dots, n.$$

Again by 3.31, applied now to the identity operator $I \in \mathcal{L}(V)$ with the basis $u_1, \dots, u_n$ of the domain and the basis $v_1, \dots, v_n$ of the target, the entries $C_{j,k}$ of C are the unique scalars satisfying

$$u_k = Iu_k = C_{1,k}v_1 + \cdots + C_{n,k}v_n \qquad \text{for each } k = 1, \dots, n.$$

Fix $k \in \{1, \dots, n\}$. By hypothesis $Tv_k = u_k$, so the two displayed equations give

$$A_{1,k}v_1 + \cdots + A_{n,k}v_n = Tv_k = u_k = C_{1,k}v_1 + \cdots + C_{n,k}v_n.$$

Because $v_1, \dots, v_n$ is a basis of V, every vector of V, in particular u_k, has a unique representation as a linear combination of $v_1, \dots, v_n$ (this is the criterion for basis, 2.28). Hence

$$A_{j,k} = C_{j,k} \qquad \text{for each } j = 1, \dots, n.$$

Since k was arbitrary, all entries of A and C agree, so $A = C$, which is the desired equation

$$\mathcal{M}\big(T, (v_1, \dots, v_n)\big) = \mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big).$$

Two remarks tie this to the surrounding theory. As the book notes just before 3.82, column k of $\mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n))$ consists precisely of the scalars needed to write u_k in terms of $v_1, \dots, v_n$; the content of this exercise is that these are also the scalars needed to write Tv_k in terms of $v_1, \dots, v_n$, which is immediate once $Tv_k = u_k$. Moreover, since 3.82 tells us that $\mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n))$ is invertible, this exercise combined with Exercise 22 of this section shows that the operator T carrying the basis $v_1, \dots, v_n$ to the basis $u_1, \dots, u_n$ is invertible, as one also sees directly from Exercise 3 of this section.

Problem 3D.24. Suppose A and B are square matrices of the same size and $AB = I$. Prove that $BA = I$.

Solution. Let n be the common size of A and B, so $A, B \in \mathbf{F}^{n,n}$. The strategy is to convert the matrix hypothesis into a statement about operators on an n-dimensional vector space, where 3.68 applies, and then convert back.

Let $V = \mathbf{F}^n$, with standard basis $e_1, \dots, e_n$; thus $\dim V = n$ and V is finite-dimensional. All matrices of operators on V below are taken with respect to this single basis, used both as the basis of the domain and as the basis of the target, so we write $\mathcal{M}(R)$ for $\mathcal{M}(R, (e_1, \dots, e_n), (e_1, \dots, e_n))$ when $R \in \mathcal{L}(V)$. By 3.71 (applied with $W = V$ and with $e_1, \dots, e_n$ as the basis in both roles), the map

$$\mathcal{M} \colon \mathcal{L}(\mathbf{F}^n) \to \mathbf{F}^{n,n}$$

is an isomorphism; in particular it is injective and surjective. By surjectivity there exist $S, T \in \mathcal{L}(\mathbf{F}^n)$ with

$$\mathcal{M}(S) = A \qquad \text{and} \qquad \mathcal{M}(T) = B.$$

(Explicitly, by the linear map lemma 3.4 one may take S to be the unique operator with $Se_k = \sum_{j=1}^{n} A_{j,k}e_j$ for each k, and similarly for T with the entries of B.)

Recall also that $\mathcal{M}(I) = I$: the identity operator sends each e_k to itself, so by 3.31 its matrix with respect to a single basis has 1's on the diagonal and 0's elsewhere, which is the identity matrix 3.79.

Now apply 3.81 with $U = V = W = \mathbf{F}^n$ and with $e_1, \dots, e_n$ used as all three bases. This gives $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$, hence

$$\mathcal{M}(ST) = AB = I = \mathcal{M}(I).$$

Because $\mathcal{M}$ is injective, we conclude that

$$ST = I,$$

where now I denotes the identity operator on $\mathbf{F}^n$.
Since $\mathbf{F}^n$ is finite-dimensional and $S, T \in \mathcal{L}(\mathbf{F}^n)$ (so the two spaces involved trivially have the same dimension), 3.68 applies and yields

$$TS = I.$$

For completeness, here is that argument in this special case: if $v \in \mathbf{F}^n$ and $Tv = 0$, then $v = Iv = (ST)v = S(Tv) = S(0) = 0$, so $\operatorname{null} T = \{0\}$ and T is injective by 3.15; because T is an operator on the finite-dimensional space $\mathbf{F}^n$, injectivity forces invertibility by 3.65; multiplying $ST = I$ on the right by T^{-1} gives $S = T^{-1}$, and therefore $TS = TT^{-1} = I$.
Finally convert back to matrices. Using 3.81 once more,

$$BA = \mathcal{M}(T)\,\mathcal{M}(S) = \mathcal{M}(TS) = \mathcal{M}(I) = I.$$

Thus $BA = I$, as desired. In particular $AB = BA = I$, so A is an invertible matrix with $A^{-1} = B$ in the sense of 3.80: the hypothesis $AB = I$ alone already forces B to be a two-sided inverse of A.
The finiteness of the size is essential to this argument, and not merely an artifact of the proof. For linear maps on infinite-dimensional spaces the analogous statement fails: with $V = \mathbf{F}^\infty$, let T be the forward shift $T(x_1, x_2, \dots) = (0, x_1, x_2, \dots)$ and let S be the backward shift $S(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots)$. Then $ST = I$ while $TS \neq I$, since $TS(1, 0, 0, \dots) = 0$. This is the phenomenon behind Exercise 13 of this section; for square matrices it cannot happen because $\mathbf{F}^n$ has finite dimension.

3.5 Exercises 3E

Problem 3E.1. Suppose T is a function from V to W. The graph of T is the subset of $V \times W$ defined by

$$\text{graph of } T = \{(v, Tv) \in V \times W : v \in V\}.$$

Prove that T is a linear map if and only if the graph of T is a subspace of $V \times W$.
[Formally, a function T from V to W is a subset T of $V \times W$ such that for each $v \in V$, there exists exactly one element $(v, w) \in T$. In other words, formally a function is what is called above its graph. We do not usually think of functions in this formal manner. However, if we do become formal, then this exercise could be rephrased as follows: Prove that a function T from V to W is a linear map if and only if T is a subspace of $V \times W$.]

Solution. Write $G = \{(v, Tv) : v \in V\}$ for the graph of T. The key observation, used repeatedly below, is that for $v \in V$ and $w \in W$ we have

$$(v, w) \in G \iff w = Tv.$$

Indeed, every element of G has the form (u, Tu) for some $u \in V$, and $(u, Tu) = (v, w)$ forces $u = v$ and hence $w = Tv$; conversely, if $w = Tv$ then $(v, w) = (v, Tv) \in G$. So the first slot of an element of G determines the second slot. Note that this uses only that T is a function from V to W, not that it is linear.
Recall (1.34) that a subset of a vector space is a subspace exactly when it contains the additive identity and is closed under addition and under scalar multiplication. The additive identity of $V \times W$ is $(0, 0)$, by 3.89 and the remark following it.
First suppose T is a linear map. Then $T0 = 0$ by 3.10, so $(0, 0) = (0, T0) \in G$. If (u, Tu) and (v, Tv) are elements of G, then by additivity of T,

$$(u, Tu) + (v, Tv) = (u + v, Tu + Tv) = (u + v, T(u + v)) \in G,$$

so G is closed under addition. If $\lambda \in \mathbf{F}$ and $(v, Tv) \in G$, then by homogeneity of T,

$$\lambda(v, Tv) = (\lambda v, \lambda Tv) = (\lambda v, T(\lambda v)) \in G,$$

so G is closed under scalar multiplication. Hence G is a subspace of $V \times W$.

Conversely, suppose G is a subspace of $V \times W$. Let $u, v \in V$. Then $(u, Tu) \in G$ and $(v, Tv) \in G$, so closure of G under addition gives

$$(u + v,\ Tu + Tv) = (u, Tu) + (v, Tv) \in G.$$

By the key observation applied to the vector $u + v$ in the first slot, this says exactly that $Tu + Tv = T(u + v)$. Thus T is additive.
Now let $\lambda \in \mathbf{F}$ and $v \in V$. Then $(v, Tv) \in G$, so closure of G under scalar multiplication gives

$$(\lambda v,\ \lambda Tv) = \lambda(v, Tv) \in G.$$

By the key observation applied to the vector λv in the first slot, this says exactly that $\lambda Tv = T(\lambda v)$. Thus T is homogeneous.
Hence T satisfies both conditions in the definition of a linear map (3.1), so T is linear.
(Note that in the reverse direction we never needed the hypothesis $(0, 0) \in G$; closure under addition and under scalar multiplication already forced additivity and homogeneity of T.)

Problem 3E.2. Suppose that $V_1, \dots, V_m$ are vector spaces such that $V_1 \times \cdots \times V_m$ is finite-dimensional. Prove that V_k is finite-dimensional for each $k = 1, \dots, m$.

Solution. Fix $k \in \{1, \dots, m\}$ and define the k-th coordinate projection

$$P_k : V_1 \times \cdots \times V_m \to V_k, \qquad P_k(v_1, \dots, v_m) = v_k.$$

First, P_k is linear. Addition and scalar multiplication on the product are defined slotwise (3.87), so for $(u_1, \dots, u_m)$ and $(v_1, \dots, v_m)$ in the product and $\lambda \in \mathbf{F}$,

$$\begin{aligned} P_k\big((u_1, \dots, u_m) + (v_1, \dots, v_m)\big) &= P_k(u_1 + v_1, \dots, u_m + v_m) = u_k + v_k \\ &= P_k(u_1, \dots, u_m) + P_k(v_1, \dots, v_m), \end{aligned}$$

and

$$P_k\big(\lambda(v_1, \dots, v_m)\big) = P_k(\lambda v_1, \dots, \lambda v_m) = \lambda v_k = \lambda P_k(v_1, \dots, v_m).$$

So $P_k \in \mathcal{L}(V_1 \times \cdots \times V_m,\ V_k)$.
Second, P_k is surjective: given $v \in V_k$, the element

$$(0, \dots, 0, v, 0, \dots, 0) \in V_1 \times \cdots \times V_m$$

(with v in slot k and the additive identity of V_j in slot j for $j \neq k$) is mapped by P_k to v. Hence $\operatorname{range} P_k = V_k$.
Now apply the fundamental theorem of linear maps (3.21) to P_k. Its hypothesis is satisfied because the domain $V_1 \times \cdots \times V_m$ is finite-dimensional by assumption. The theorem asserts in particular that $\operatorname{range} P_k$ is finite-dimensional. Since $\operatorname{range} P_k = V_k$, we conclude that V_k is finite-dimensional.
As k was arbitrary, V_k is finite-dimensional for each $k = 1, \dots, m$.
(Once each V_k is known to be finite-dimensional, 3.92 applies and gives the sharper statement $\dim V_1 + \cdots + \dim V_m = \dim(V_1 \times \cdots \times V_m)$; in particular each $\dim V_k$ is at most $\dim(V_1 \times \cdots \times V_m)$.)

Problem 3E.3. Suppose $V_1, \dots, V_m$ are vector spaces. Prove that $\mathcal{L}(V_1 \times \cdots \times V_m, W)$ and $\mathcal{L}(V_1, W) \times \cdots \times \mathcal{L}(V_m, W)$ are isomorphic vector spaces.
[There is no assumption in the exercise above or in the two following exercises that the vector spaces are finite-dimensional.]

Solution. Both objects are vector spaces: each $\mathcal{L}(V_k, W)$ and $\mathcal{L}(V_1 \times \cdots \times V_m, W)$ is a vector space by 3.6, and a product of vector spaces is a vector space by 3.89 (the product $V_1 \times \cdots \times V_m$ is a vector

space by 3.89 as well). Write

$$\mathcal{V} = V_1 \times \cdots \times V_m.$$

For $k = 1, \ldots, m$ define the k-th inclusion

$$\iota_k : V_k \to \mathcal{V}, \qquad \iota_k(v) = (0, \ldots, 0, v, 0, \ldots, 0),$$

with v in slot k and zeros elsewhere. Because the operations on $\mathcal{V}$ are slotwise (3.87), each ι_k is linear. Note also the identity

$$(v_1, \ldots, v_m) = \iota_1(v_1) + \cdots + \iota_m(v_m) \qquad \text{for all } (v_1, \ldots, v_m) \in \mathcal{V},$$

which is immediate from slotwise addition.
Define

$$\Phi : \mathcal{L}(\mathcal{V}, W) \to \mathcal{L}(V_1, W) \times \cdots \times \mathcal{L}(V_m, W), \qquad \Phi(T) = (T \circ \iota_1, \ldots, T \circ \iota_m).$$

This makes sense: each $T \circ \iota_k$ is a composition of linear maps, hence lies in $\mathcal{L}(V_k, W)$ by 3.7 (or directly from the definitions).
Φ is linear. Let $S, T \in \mathcal{L}(\mathcal{V}, W)$ and $\lambda \in \mathbf{F}$. For each k and each $v \in V_k$, the definition of addition on $\mathcal{L}(\mathcal{V}, W)$ (3.5) gives

$$\big((S+T) \circ \iota_k\big)(v) = (S+T)(\iota_k v) = S(\iota_k v) + T(\iota_k v) = (S \circ \iota_k)(v) + (T \circ \iota_k)(v),$$

so $(S+T) \circ \iota_k = S \circ \iota_k + T \circ \iota_k$. Since addition on the product space is slotwise, $\Phi(S+T) = \Phi(S) + \Phi(T)$. Similarly $\big((\lambda T) \circ \iota_k\big)(v) = \lambda T(\iota_k v) = \lambda (T \circ \iota_k)(v)$, so $\Phi(\lambda T) = \lambda \Phi(T)$.
Φ is injective. Suppose $\Phi(T) = (0, \ldots, 0)$, that is, $T \circ \iota_k = 0$ for every k. Then for any $(v_1, \ldots, v_m) \in \mathcal{V}$, using the identity above and linearity of T,

$$T(v_1, \ldots, v_m) = T\big(\iota_1(v_1) + \cdots + \iota_m(v_m)\big) = (T \circ \iota_1)(v_1) + \cdots + (T \circ \iota_m)(v_m) = 0.$$

Hence $T = 0$, so $\operatorname{null} \Phi = \{0\}$ and Φ is injective by 3.15.
Φ is surjective. Let $(T_1, \ldots, T_m) \in \mathcal{L}(V_1, W) \times \cdots \times \mathcal{L}(V_m, W)$. Define $T : \mathcal{V} \to W$ by

$$T(v_1, \ldots, v_m) = T_1 v_1 + \cdots + T_m v_m.$$

Then T is linear: for $(u_1, \ldots, u_m), (v_1, \ldots, v_m) \in \mathcal{V}$,

$$\begin{aligned} T\big((u_1, \ldots, u_m) + (v_1, \ldots, v_m)\big) &= T_1(u_1 + v_1) + \cdots + T_m(u_m + v_m) \\ &= (T_1 u_1 + \cdots + T_m u_m) + (T_1 v_1 + \cdots + T_m v_m), \end{aligned}$$

using additivity of each T_k and commutativity of addition in W; and for $\lambda \in \mathbf{F}$,

$$T\big(\lambda(v_1, \ldots, v_m)\big) = T_1(\lambda v_1) + \cdots + T_m(\lambda v_m) = \lambda\big(T_1 v_1 + \cdots + T_m v_m\big).$$

So $T \in \mathcal{L}(\mathcal{V}, W)$. Finally, for each k and each $v \in V_k$,

$$(T \circ \iota_k)(v) = T(0, \ldots, 0, v, 0, \ldots, 0) = T_1 0 + \cdots + T_k v + \cdots + T_m 0 = T_k v,$$

where we used $T_j 0 = 0$ (3.10). Thus $T \circ \iota_k = T_k$ for every k, i.e. $\Phi(T) = (T_1, \ldots, T_m)$.
Being linear, injective, and surjective, Φ is invertible by 3.63, hence an isomorphism (3.69). Therefore $\mathcal{L}(V_1 \times \cdots \times V_m, W)$ and $\mathcal{L}(V_1, W) \times \cdots \times \mathcal{L}(V_m, W)$ are isomorphic vector spaces. No finite-dimensionality was used anywhere.

Problem 3E.4. Suppose $W_1, \ldots, W_m$ are vector spaces. Prove that $\mathcal{L}(V, W_1 \times \cdots \times W_m)$ and $\mathcal{L}(V, W_1) \times \cdots \times \mathcal{L}(V, W_m)$ are isomorphic vector spaces.

Solution. Write $\mathcal{W} = W_1 \times \cdots \times W_m$, a vector space by 3.89. Each $\mathcal{L}(V, W_k)$ and $\mathcal{L}(V, \mathcal{W})$ is a vector space by 3.6, and so is the product $\mathcal{L}(V, W_1) \times \cdots \times \mathcal{L}(V, W_m)$ by 3.89.
For $k = 1, \ldots, m$ let

$$P_k : \mathcal{W} \to W_k, \qquad P_k(w_1, \ldots, w_m) = w_k$$

be the k-th coordinate projection; it is linear because the operations on $\mathcal{W}$ are slotwise (3.87), as verified in Exercise 2 of this section. Note that

$$w = \big(P_1 w, \ldots, P_m w\big) \qquad \text{for all } w \in \mathcal{W}.$$

Call this identity the reconstruction identity.
Define

$$\Psi : \mathcal{L}(V, \mathcal{W}) \to \mathcal{L}(V, W_1) \times \cdots \times \mathcal{L}(V, W_m), \qquad \Psi(T) = (P_1 \circ T, \ldots, P_m \circ T).$$

Each $P_k \circ T$ is a composition of linear maps and hence lies in $\mathcal{L}(V, W_k)$ by 3.7, so Ψ indeed maps into the stated space. Concretely, if $Tv = (T_1 v, \ldots, T_m v)$ then $\Psi(T) = (T_1, \ldots, T_m)$: the map Ψ just records the m component maps of T.
Ψ is linear. For $S, T \in \mathcal{L}(V, \mathcal{W})$, $\lambda \in \mathbf{F}$, $v \in V$, and each k, the definition of addition and scalar multiplication on $\mathcal{L}(V, \mathcal{W})$ (3.5) together with linearity of P_k gives

$$\big(P_k \circ (S + T)\big)(v) = P_k(Sv + Tv) = P_k(Sv) + P_k(Tv) = (P_k \circ S)(v) + (P_k \circ T)(v)$$

and

$$\big(P_k \circ (\lambda T)\big)(v) = P_k(\lambda\, Tv) = \lambda\, P_k(Tv) = \lambda (P_k \circ T)(v).$$

Since the operations on the product of the spaces $\mathcal{L}(V, W_k)$ are slotwise, these identities say $\Psi(S+T) = \Psi(S) + \Psi(T)$ and $\Psi(\lambda T) = \lambda \Psi(T)$.
Ψ is injective. Suppose $\Psi(T) = (0, \ldots, 0)$, so $P_k(Tv) = 0$ for every k and every $v \in V$. By the reconstruction identity applied to $w = Tv$,

$$Tv = \big(P_1(Tv), \ldots, P_m(Tv)\big) = (0, \ldots, 0),$$

which is the additive identity of $\mathcal{W}$. Hence $T = 0$, so $\operatorname{null} \Psi = \{0\}$ and Ψ is injective by 3.15.
Ψ is surjective. Let $(T_1, \ldots, T_m) \in \mathcal{L}(V, W_1) \times \cdots \times \mathcal{L}(V, W_m)$ and define $T : V \to \mathcal{W}$ by

$$Tv = (T_1 v, \ldots, T_m v).$$

Then T is linear: for $u, v \in V$, using additivity of each T_k and slotwise addition in $\mathcal{W}$,

$$T(u + v) = \big(T_1(u + v), \ldots, T_m(u + v)\big) = (T_1 u + T_1 v, \ldots, T_m u + T_m v) = Tu + Tv,$$

and for $\lambda \in \mathbf{F}$,

$$T(\lambda v) = (T_1(\lambda v), \ldots, T_m(\lambda v)) = (\lambda T_1 v, \ldots, \lambda T_m v) = \lambda\, Tv.$$

So $T \in \mathcal{L}(V, \mathcal{W})$, and by construction $(P_k \circ T)(v) = T_k v$ for all v and all k, i.e. $\Psi(T) = (T_1, \ldots, T_m)$.
Thus Ψ is a linear map that is injective and surjective, hence invertible by 3.63 and therefore an isomorphism (3.69). Consequently $\mathcal{L}(V, W_1 \times \cdots \times W_m)$ and $\mathcal{L}(V, W_1) \times \cdots \times \mathcal{L}(V, W_m)$ are isomorphic vector spaces. Again no finite-dimensionality hypothesis was needed.

Problem 3E.5. For m a positive integer, define V^m by

$$V^m = \underbrace{V \times \cdots \times V}_{m \text{ times}}.$$

Prove that V^m and $\mathcal{L}(\mathbf{F}^m, V)$ are isomorphic vector spaces.

Solution. Both V^m (by 3.89) and $\mathcal{L}(\mathbf{F}^m, V)$ (by 3.6) are vector spaces over $\mathbf{F}$.
Let $e_1, \dots, e_m$ be the standard basis of $\mathbf{F}^m$, so e_k is the list of length m with 1 in slot k and 0 elsewhere; thus

$$(x_1, \dots, x_m) = x_1 e_1 + \cdots + x_m e_m \qquad \text{for all } (x_1, \dots, x_m) \in \mathbf{F}^m.$$

Define

$$\Phi : \mathcal{L}(\mathbf{F}^m, V) \to V^m, \qquad \Phi(T) = (Te_1, \dots, Te_m).$$

Φ is linear. If $S, T \in \mathcal{L}(\mathbf{F}^m, V)$ and $\lambda \in \mathbf{F}$, then by the definition of the operations on $\mathcal{L}(\mathbf{F}^m, V)$ (3.5) and slotwise operations on V^m (3.87),

$$\Phi(S+T) = \big((S+T)e_1, \dots, (S+T)e_m\big) = (Se_1 + Te_1, \dots, Se_m + Te_m) = \Phi(S) + \Phi(T),$$

$$\Phi(\lambda T) = \big((\lambda T)e_1, \dots, (\lambda T)e_m\big) = (\lambda Te_1, \dots, \lambda Te_m) = \lambda\, \Phi(T).$$

Φ is injective. Suppose $\Phi(T) = (0, \dots, 0)$, i.e. $Te_k = 0$ for each k. Then for any $(x_1, \dots, x_m) \in \mathbf{F}^m$, linearity of T gives

$$T(x_1, \dots, x_m) = T(x_1 e_1 + \cdots + x_m e_m) = x_1 Te_1 + \cdots + x_m Te_m = 0.$$

Hence $T = 0$, so $\operatorname{null} \Phi = \{0\}$ and Φ is injective by 3.15.
Φ is surjective. Let $(v_1, \dots, v_m) \in V^m$. Because $e_1, \dots, e_m$ is a basis of $\mathbf{F}^m$, the linear map lemma (3.4) provides a (unique) $T \in \mathcal{L}(\mathbf{F}^m, V)$ with $Te_k = v_k$ for each $k = 1, \dots, m$; explicitly, $T(x_1, \dots, x_m) = x_1 v_1 + \cdots + x_m v_m$. Then $\Phi(T) = (v_1, \dots, v_m)$.
Since Φ is linear, injective, and surjective, it is invertible by 3.63, hence an isomorphism (3.69). Therefore $\mathcal{L}(\mathbf{F}^m, V)$ and V^m are isomorphic vector spaces.
(Nothing here requires V to be finite-dimensional. When V is finite-dimensional, both sides have dimension $m \dim V$, by 3.92 on the one hand and by 3.72 on the other, which is consistent.)

Problem 3E.6. Suppose that v, x are vectors in V and that U, W are subspaces of V such that $v + U = x + W$. Prove that $U = W$.

Solution. Recall the notation $v + U = \{v + u : u \in U\}$ from 3.95.
First we show $U \subseteq W$. Because U is a subspace, $0 \in U$, so

$$v = v + 0 \in v + U = x + W,$$

which means there exists $w_0 \in W$ with

$$v = x + w_0.$$

Now let $u \in U$ be arbitrary. Then $v + u \in v + U = x + W$, so there exists $w \in W$ with $v + u = x + w$, and therefore

$$u = (x + w) - v = (x + w) - (x + w_0) = w - w_0.$$

Since W is a subspace it is closed under additive inverses and addition, so $w - w_0 \in W$. Hence $u \in W$. As $u \in U$ was arbitrary, $U \subseteq W$.
The reverse inclusion is the same argument with the roles of (v, U) and (x, W) interchanged, which is legitimate because the hypothesis $v + U = x + W$ is symmetric in the two pairs. Explicitly: since $0 \in W$ we have $x \in x + W = v + U$, so $x = v + u_0$ for some $u_0 \in U$. Let $w \in W$ be arbitrary. Then $x + w \in x + W = v + U$, so $x + w = v + u$ for some $u \in U$, and therefore

$$w = (v + u) - x = (v + u) - (v + u_0) = u - u_0 \in U,$$

because U is a subspace. As $w \in W$ was arbitrary, $W \subseteq U$.
Combining the two inclusions gives $U = W$.
(Remark: this shows that a translate determines the subspace it translates, so it is unambiguous to speak of the subspace of which a given set is a translate. The vector v is by contrast far from unique: by 3.101, $v + U = x + U$ precisely when $v - x \in U$. Combining the exercise with 3.101 gives the complete description: $v + U = x + W$ if and only if $U = W$ and $v - x \in U$.)

Problem 3E.7. Let $U = \{(x, y, z) \in \mathbf{R}^3 : 2x + 3y + 5z = 0\}$. Suppose $A \subseteq \mathbf{R}^3$. Prove that A is a translate of U if and only if there exists $c \in \mathbf{R}$ such that

$$A = \{(x, y, z) \in \mathbf{R}^3 : 2x + 3y + 5z = c\}.$$

Solution. Define $\varphi : \mathbf{R}^3 \to \mathbf{R}$ by

$$\varphi(x, y, z) = 2x + 3y + 5z.$$

Then φ is linear: for $(x, y, z), (x', y', z') \in \mathbf{R}^3$ and $\lambda \in \mathbf{R}$,

$$\varphi(x + x', y + y', z + z') = 2(x + x') + 3(y + y') + 5(z + z') = \varphi(x, y, z) + \varphi(x', y', z')$$

and $\varphi(\lambda x, \lambda y, \lambda z) = \lambda(2x + 3y + 5z) = \lambda\,\varphi(x, y, z)$. By definition $U = \operatorname{null}\varphi$, so U is a subspace of $\mathbf{R}^3$ by 3.13. For $c \in \mathbf{R}$ write

$$A_c = \{p \in \mathbf{R}^3 : \varphi(p) = c\},$$

so the set displayed in the exercise is exactly A_c, and $A_0 = U$.
The whole proof rests on the following observation. Let $v \in \mathbf{R}^3$ and put $c = \varphi(v)$. Then for every $p \in \mathbf{R}^3$,

$$p \in v + U \iff p - v \in U \iff \varphi(p - v) = 0 \iff \varphi(p) = \varphi(v) = c \iff p \in A_c.$$

Here the first equivalence holds because $p \in v + U$ means $p = v + u$ for some $u \in U$, which is the same as saying $p - v \in U$; the second is the definition of $U = \operatorname{null}\varphi$; and the third uses linearity of φ, namely $\varphi(p - v) = \varphi(p) - \varphi(v)$. Hence

$$v + U = A_{\varphi(v)} \qquad \text{for every } v \in \mathbf{R}^3.$$

Call this the key identity.
Now suppose A is a translate of U, so $A = v + U$ for some $v \in \mathbf{R}^3$ (3.97). Set $c = \varphi(v) \in \mathbf{R}$. By the key identity, $A = v + U = A_c$, which is precisely the desired conclusion: there exists $c \in \mathbf{R}$ with

$$A = \{(x, y, z) \in \mathbf{R}^3 : 2x + 3y + 5z = c\}.$$

Conversely, suppose there exists $c \in \mathbf{R}$ with $A = A_c$. Put

$$v = \left(\tfrac{c}{2},\, 0,\, 0\right) \in \mathbf{R}^3,$$

so that $\varphi(v) = 2 \cdot \frac{c}{2} + 3 \cdot 0 + 5 \cdot 0 = c$. By the key identity, $v + U = A_{\varphi(v)} = A_c = A$. Thus A is a translate of U.
This proves both implications, so A is a translate of U if and only if $A = A_c$ for some $c \in \mathbf{R}$.
(Geometrically: U is the plane through the origin with normal vector $(2, 3, 5)$, and the translates of U are exactly the planes parallel to it, which are the level sets of φ. This is the general fact recorded in the last bullet of 3.98. Note also that the scalar c is uniquely determined by A, since c is the single value φ takes on the nonempty set A; and by Exercise 6, U itself is determined by A.)

Problem 3E.8. (a) Suppose $T \in \mathcal{L}(V, W)$ and $c \in W$. Prove that $\{x \in V : Tx = c\}$ is either the empty set or is a translate of $\operatorname{null} T$.
(b) Explain why the set of solutions to a system of linear equations such as 3.27 is either the empty set or is a translate of some subspace of $\mathbf{F}^n$.

Solution. (a) Write

$$A = \{x \in V : Tx = c\}.$$

If $A = \varnothing$ there is nothing to prove, so assume $A \neq \varnothing$ and fix some $x_0 \in A$, meaning $Tx_0 = c$. We

claim that

$$A = x_0 + \operatorname{null} T,$$

which exhibits A as a translate of $\operatorname{null} T$ (recall that $\operatorname{null} T$ is a subspace of V by 3.13, and see 3.97 for the definition of translate).
First suppose $x \in A$. Then $Tx = c$, so by linearity

$$T(x - x_0) = Tx - Tx_0 = c - c = 0.$$

Hence $x - x_0 \in \operatorname{null} T$, and therefore

$$x = x_0 + (x - x_0) \in x_0 + \operatorname{null} T.$$

Thus $A \subseteq x_0 + \operatorname{null} T$.
Conversely, suppose $u \in \operatorname{null} T$. Then

$$T(x_0 + u) = Tx_0 + Tu = c + 0 = c,$$

so $x_0 + u \in A$. Thus $x_0 + \operatorname{null} T \subseteq A$.
The two inclusions give $A = x_0 + \operatorname{null} T$, as claimed. Hence A is either empty or a translate of $\operatorname{null} T$.
(b) The system 3.27 consists of the m equations

$$\sum_{k=1}^{n} A_{j,k} x_k = c_j \qquad (j = 1, \dots, m)$$

in the unknowns $x_1, \dots, x_n \in \mathbf{F}$, where the $A_{j,k}$ and the c_j are given scalars. Define $T\colon \mathbf{F}^n \to \mathbf{F}^m$ as in 3.25 by

$$T(x_1, \dots, x_n) = \Big(\sum_{k=1}^{n} A_{1,k} x_k,\ \dots,\ \sum_{k=1}^{n} A_{m,k} x_k \Big);$$

this T is a linear map. Set $c = (c_1, \dots, c_m) \in \mathbf{F}^m$. Then an n-tuple $(x_1, \dots, x_n)$ satisfies the system 3.27 if and only if $T(x_1, \dots, x_n) = c$. In other words, the solution set of the system is exactly

$$\{x \in \mathbf{F}^n : Tx = c\}.$$

By part (a), this set is either empty or a translate of $\operatorname{null} T$. Since $\operatorname{null} T$ is a subspace of $\mathbf{F}^n$ (by 3.13), the solution set of a system of linear equations is either the empty set or a translate of some subspace of $\mathbf{F}^n$, as desired.

Problem 3E.9. Prove that a nonempty subset A of V is a translate of some subspace of V if and only if $\lambda v + (1 - \lambda) w \in A$ for all $v, w \in A$ and all $\lambda \in \mathbf{F}$.

Solution. First suppose A is a translate of a subspace of V, say

$$A = x + U$$

for some $x \in V$ and some subspace U of V. Let $v, w \in A$ and let $\lambda \in \mathbf{F}$. Then $v = x + u_1$ and $w = x + u_2$ for some $u_1, u_2 \in U$. Hence

$$\begin{aligned} \lambda v + (1 - \lambda) w &= \lambda(x + u_1) + (1 - \lambda)(x + u_2) \\ &= \big(\lambda + (1 - \lambda)\big) x + \lambda u_1 + (1 - \lambda) u_2 \\ &= x + \big(\lambda u_1 + (1 - \lambda) u_2\big). \end{aligned}$$

Because U is a subspace, it is closed under scalar multiplication and addition, so $\lambda u_1 + (1 - \lambda) u_2 \in U$. Thus $\lambda v + (1 - \lambda) w \in x + U = A$, as required.
Now suppose conversely that A is a nonempty subset of V such that

$$\lambda v + (1 - \lambda) w \in A \quad \text{for all } v, w \in A \text{ and all } \lambda \in \mathbf{F}.$$

Call this condition $(*)$. Because A is nonempty we may fix some $x \in A$. Define

$$U = A - x = \{a - x : a \in A\}.$$

Then $A = x + U$, because $a \mapsto a - x$ is a bijection from A onto U whose inverse is $u \mapsto x + u$. So it suffices to show that U is a subspace of V; we verify the three conditions of 1.34.
Additive identity. Since $x \in A$, we have $0 = x - x \in U$.
Closure under scalar multiplication. Let $u \in U$ and $\lambda \in \mathbf{F}$. Write $u = a - x$ with $a \in A$. Applying $(*)$ with $v = a$ and $w = x$ (both in A) gives

$$\lambda a + (1 - \lambda)x \in A.$$

Since

$$\lambda a + (1 - \lambda)x = x + \lambda(a - x) = x + \lambda u,$$

we conclude that $x + \lambda u \in A$, which means $\lambda u \in A - x = U$.
Closure under addition. Let $u_1, u_2 \in U$, say $u_1 = a - x$ and $u_2 = b - x$ with $a, b \in A$. Applying $(*)$ with $v = a$, $w = b$, and $\lambda = \frac{1}{2}$ (legitimate since $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$, so $\frac{1}{2} \in \mathbf{F}$) gives

$$\tfrac{1}{2}a + \tfrac{1}{2}b \in A.$$

Hence

$$\tfrac{1}{2}u_1 + \tfrac{1}{2}u_2 = \tfrac{1}{2}(a - x) + \tfrac{1}{2}(b - x) = \big(\tfrac{1}{2}a + \tfrac{1}{2}b\big) - x \in A - x = U.$$

Now the closure under scalar multiplication proved above, applied with the scalar 2, gives

$$u_1 + u_2 = 2\big(\tfrac{1}{2}u_1 + \tfrac{1}{2}u_2\big) \in U.$$

Thus U is a subspace of V and $A = x + U$ is a translate of U, completing the proof.

Problem 3E.10. Suppose $A_1 = v + U_1$ and $A_2 = w + U_2$ for some $v, w \in V$ and some subspaces U_1, U_2 of V. Prove that the intersection $A_1 \cap A_2$ is either a translate of some subspace of V or is the empty set.

Solution. If $A_1 \cap A_2 = \varnothing$, then the conclusion holds. So assume $A_1 \cap A_2 \neq \varnothing$ and fix

$$x \in A_1 \cap A_2.$$

We will show that

$$A_1 \cap A_2 = x + (U_1 \cap U_2),$$

which finishes the proof because $U_1 \cap U_2$ is a subspace of V (this is Exercise 1C.10; one checks directly that $U_1 \cap U_2$ contains 0 and is closed under addition and scalar multiplication, since U_1 and U_2 each are).
Because $x \in A_1 = v + U_1$, we have $x - v \in U_1$, and hence 3.101 gives

$$A_1 = v + U_1 = x + U_1.$$

Similarly $x - w \in U_2$, so 3.101 gives

$$A_2 = w + U_2 = x + U_2.$$

Therefore

$$A_1 \cap A_2 = (x + U_1) \cap (x + U_2).$$

Now let $y \in V$. Then

$$\begin{aligned} y \in (x + U_1) \cap (x + U_2) &\iff y - x \in U_1 \text{ and } y - x \in U_2 \\ &\iff y - x \in U_1 \cap U_2 \\ &\iff y \in x + (U_1 \cap U_2). \end{aligned}$$

Hence

$$A_1 \cap A_2 = x + (U_1 \cap U_2),$$

which is a translate of the subspace $U_1 \cap U_2$ of V. Thus $A_1 \cap A_2$ is either the empty set or a translate of a subspace of V, as desired.

Problem 3E.11. Suppose $U = \{(x_1, x_2, \ldots) \in \mathbf{F}^\infty : x_k \neq 0 \text{ for only finitely many } k\}$.
(a) Show that U is a subspace of $\mathbf{F}^\infty$.
(b) Prove that $\mathbf{F}^\infty/U$ is infinite-dimensional.

Solution. For a sequence $x = (x_1, x_2, \ldots) \in \mathbf{F}^\infty$, write

$$\operatorname{supp} x = \{k \in \mathbf{Z}^+ : x_k \neq 0\}.$$

Thus $U = \{x \in \mathbf{F}^\infty : \operatorname{supp} x \text{ is finite}\}$.
(a) We verify the three conditions of 1.34.
The zero sequence $0 = (0, 0, \ldots)$ has $\operatorname{supp} 0 = \varnothing$, which is finite, so $0 \in U$.
Suppose $x, y \in U$. If $k \notin \operatorname{supp} x \cup \operatorname{supp} y$ then $x_k = y_k = 0$, so $x_k + y_k = 0$. Hence

$$\operatorname{supp}(x + y) \subseteq \operatorname{supp} x \cup \operatorname{supp} y,$$

and the right side is a union of two finite sets, hence finite. Thus $\operatorname{supp}(x + y)$ is finite and $x + y \in U$.
Suppose $x \in U$ and $\lambda \in \mathbf{F}$. If $k \notin \operatorname{supp} x$ then $\lambda x_k = 0$, so $\operatorname{supp}(\lambda x) \subseteq \operatorname{supp} x$, which is finite. Thus $\lambda x \in U$.
Hence U is a subspace of $\mathbf{F}^\infty$.
(b) By 2.22, if a vector space is finite-dimensional then every linearly independent list in it has length at most the length of any spanning list, and in particular at most the dimension of the space. Hence to prove that $\mathbf{F}^\infty/U$ is infinite-dimensional it suffices to produce, for every positive integer n, a linearly independent list of length n in $\mathbf{F}^\infty/U$.
We first split $\mathbf{Z}^+$ into infinitely many pairwise disjoint infinite pieces. For each positive integer j set

$$S_j = \{2^{j-1}(2k - 1) : k \in \mathbf{Z}^+\},$$

so S_j is the set of positive integers whose factorization into a power of 2 times an odd number uses exactly 2^{j-1}. Every positive integer m can be written in exactly one way as $m = 2^a q$ with $a \geq 0$ an integer and q odd; consequently m lies in S_{a+1} and in no other S_j. Thus the sets $S_1, S_2, \ldots$ are pairwise disjoint, and each S_j is infinite (it contains $2^{j-1}(2k - 1)$ for every $k \in \mathbf{Z}^+$, and these are distinct).
Now for each positive integer j define $v_j \in \mathbf{F}^\infty$ to be the sequence whose k-th coordinate $(v_j)_k$ equals 1 if $k \in S_j$ and equals 0 if $k \notin S_j$.
Fix a positive integer n; we claim that

$$v_1 + U,\ \ldots,\ v_n + U$$

is a linearly independent list in $\mathbf{F}^\infty/U$. Suppose $a_1, \ldots, a_n \in \mathbf{F}$ satisfy

$$a_1(v_1 + U) + \cdots + a_n(v_n + U) = 0 + U.$$

By the definition 3.102 of the operations on $\mathbf{F}^\infty/U$, the left side equals $(a_1 v_1 + \cdots + a_n v_n) + U$. Hence by 3.101,

$$a_1 v_1 + \cdots + a_n v_n \in U.$$

Set $x = a_1 v_1 + \cdots + a_n v_n$. Because $S_1, \ldots, S_n$ are pairwise disjoint, for each $j \in \{1, \ldots, n\}$ and each $k \in S_j$ exactly one of $(v_1)_k, \ldots, (v_n)_k$ equals 1 (namely $(v_j)_k$) and the rest are 0; hence

$$x_k = a_j \qquad \text{for every } k \in S_j.$$

If some $a_j \neq 0$, then $x_k = a_j \neq 0$ for all k in the infinite set S_j, so $\operatorname{supp} x$ is infinite and $x \notin U$, a contradiction. Therefore $a_1 = \cdots = a_n = 0$, which proves that $v_1 + U, \ldots, v_n + U$ is linearly independent.
Since $\mathbf{F}^\infty/U$ contains linearly independent lists of every finite length, it cannot be finite-dimensional.
Hence $\mathbf{F}^\infty/U$ is infinite-dimensional.

Problem 3E.12. Suppose $v_1, \dots, v_m \in V$. Let

$$A = \{\lambda_1 v_1 + \cdots + \lambda_m v_m : \lambda_1, \dots, \lambda_m \in \mathbf{F} \text{ and } \lambda_1 + \cdots + \lambda_m = 1\}.$$

(a) Prove that A is a translate of some subspace of V.
(b) Prove that if B is a translate of some subspace of V and $\{v_1, \dots, v_m\} \subseteq B$, then $A \subseteq B$.
(c) Prove that A is a translate of some subspace of V of dimension less than m.

Solution. Here m is a positive integer, as the notation $v_1, \dots, v_m$ indicates. Throughout, put

$$U = \operatorname{span}(v_2 - v_1, \dots, v_m - v_1),$$

a subspace of V by 2.6 (when $m = 1$ this is the span of the empty list, namely $\{0\}$, and $A = \{v_1\}$).
(a) and (c). We prove the single statement $A = v_1 + U$; this proves (a), and since U is the span of a list of length $m - 1$, we have $\dim U \le m - 1 < m$ (a spanning list of length $m - 1$ contains a basis by 2.30, so the dimension is at most $m - 1$), which proves (c).
First suppose $a \in A$, say $a = \lambda_1 v_1 + \cdots + \lambda_m v_m$ with $\lambda_1 + \cdots + \lambda_m = 1$. Then

$$\begin{aligned} a - v_1 &= \sum_{j=1}^{m} \lambda_j v_j - \Big(\sum_{j=1}^{m} \lambda_j\Big) v_1 \\ &= \sum_{j=1}^{m} \lambda_j (v_j - v_1) \\ &= \sum_{j=2}^{m} \lambda_j (v_j - v_1), \end{aligned}$$

where the last equality holds because the $j = 1$ term is $\lambda_1(v_1 - v_1) = 0$. Hence $a - v_1 \in U$, so $a \in v_1 + U$. Thus $A \subseteq v_1 + U$.
Conversely, suppose $u \in U$, say $u = c_2(v_2 - v_1) + \cdots + c_m(v_m - v_1)$ with $c_2, \dots, c_m \in \mathbf{F}$. Then

$$v_1 + u = \Big(1 - \sum_{j=2}^{m} c_j\Big) v_1 + c_2 v_2 + \cdots + c_m v_m.$$

The coefficients of $v_1, \dots, v_m$ on the right side sum to

$$\Big(1 - \sum_{j=2}^{m} c_j\Big) + \sum_{j=2}^{m} c_j = 1,$$

so $v_1 + u \in A$. Thus $v_1 + U \subseteq A$.
Therefore $A = v_1 + U$, a translate of the subspace U of V, and $\dim U < m$.
(b) Suppose B is a translate of a subspace of V with $\{v_1, \dots, v_m\} \subseteq B$, say

$$B = x + W$$

for some $x \in V$ and some subspace W of V. Since $v_j \in B$ for each j, we have

$$v_j - x \in W \qquad (j = 1, \dots, m).$$

Let $a \in A$, say $a = \lambda_1 v_1 + \cdots + \lambda_m v_m$ with $\lambda_1 + \cdots + \lambda_m = 1$. Then, using $\sum_j \lambda_j = 1$,

$$a - x = \sum_{j=1}^{m} \lambda_j v_j - \Big(\sum_{j=1}^{m} \lambda_j\Big) x = \sum_{j=1}^{m} \lambda_j (v_j - x).$$

Each $v_j - x$ lies in the subspace W, so this linear combination lies in W. Hence $a - x \in W$, which means $a \in x + W = B$.
Therefore $A \subseteq B$, as desired.
(Remark: parts (a) and (b) together say that A is the smallest translate of a subspace containing $v_1, \dots, v_m$, i.e. A is the affine hull of $v_1, \dots, v_m$.)

Problem 3E.13. Suppose U is a subspace of V such that V/U is finite-dimensional. Prove that V is isomorphic to $U \times (V/U)$.

Solution. Let $n = \dim V/U$ and choose a basis of V/U. Every element of V/U is a translate $v + U$ for some $v \in V$ (by 3.99), so we may write this basis as

$$v_1 + U,\ \ldots,\ v_n + U$$

for some $v_1, \ldots, v_n \in V$. (If $n = 0$ this is the empty list; the argument below goes through with all sums over the empty index set interpreted as 0.)
By the linear map lemma 3.4 there is a unique linear map

$$\Psi \in \mathcal{L}(V/U, V) \quad \text{with} \quad \Psi(v_j + U) = v_j \ \text{ for } j = 1, \ldots, n.$$

Define $S\colon U \times (V/U) \to V$ by

$$S(u, x) = u + \Psi x.$$

We show that S is an isomorphism.
S is linear. If $u, u' \in U$ and $x, x' \in V/U$, then, using the definition 3.87 of the operations on the product $U \times (V/U)$ and the linearity of Ψ,

$$\begin{aligned} S\big((u, x) + (u', x')\big) &= S(u + u', x + x') \\ &= (u + u') + \Psi(x + x') \\ &= (u + \Psi x) + (u' + \Psi x') \\ &= S(u, x) + S(u', x'). \end{aligned}$$

Similarly, for $\lambda \in \mathbf{F}$,

$$S\big(\lambda(u, x)\big) = S(\lambda u, \lambda x) = \lambda u + \Psi(\lambda x) = \lambda u + \lambda \Psi x = \lambda\, S(u, x).$$

Thus $S \in \mathcal{L}\big(U \times (V/U), V\big)$.
S is surjective. Let $v \in V$. Since $v + U \in V/U$ and $v_1 + U, \ldots, v_n + U$ spans V/U, there exist $a_1, \ldots, a_n \in \mathbf{F}$ with

$$v + U = a_1(v_1 + U) + \cdots + a_n(v_n + U) = (a_1 v_1 + \cdots + a_n v_n) + U,$$

the last equality by 3.102. By 3.101 this gives

$$u := v - (a_1 v_1 + \cdots + a_n v_n) \in U.$$

Setting $x = v + U$ we get, by linearity of Ψ,

$$\Psi x = a_1 \Psi(v_1 + U) + \cdots + a_n \Psi(v_n + U) = a_1 v_1 + \cdots + a_n v_n,$$

and hence

$$S(u, x) = u + \Psi x = \big(v - (a_1 v_1 + \cdots + a_n v_n)\big) + (a_1 v_1 + \cdots + a_n v_n) = v.$$

Thus S is surjective.
S is injective. By 3.15 it suffices to show that $S(u, x) = 0$ forces $(u, x) = (0, 0)$. So suppose $u \in U$ and $x \in V/U$ satisfy $u + \Psi x = 0$. Write $x = a_1(v_1 + U) + \cdots + a_n(v_n + U)$ with $a_1, \ldots, a_n \in \mathbf{F}$, which is possible because $v_1 + U, \ldots, v_n + U$ spans V/U. As above, $\Psi x = a_1 v_1 + \cdots + a_n v_n$. Hence

$$a_1 v_1 + \cdots + a_n v_n = -u \in U.$$

By 3.101 this implies $(a_1 v_1 + \cdots + a_n v_n) + U = 0 + U$, that is (using 3.102),

$$a_1(v_1 + U) + \cdots + a_n(v_n + U) = 0 + U,$$

which is the zero vector of V/U. Because $v_1 + U, \ldots, v_n + U$ is linearly independent, $a_1 = \cdots = a_n = 0$. Hence $x = 0$ and $\Psi x = 0$, and therefore $u = -\Psi x = 0$. Thus $(u, x) = (0, 0)$ and S is injective.
Being linear, injective, and surjective, S is an isomorphism from $U \times (V/U)$ onto V (by 3.63, such a map is invertible, and an invertible linear map is an isomorphism by 3.69). Therefore V is isomorphic to $U \times (V/U)$, as desired.
Note that no assumption was made about U or V being finite-dimensional; only $\dim V/U < \infty$ was used, and it was used to obtain a finite basis of V/U.

Problem 3E.14. Suppose U and W are subspaces of V and $V = U \oplus W$. Suppose $w_1, \dots, w_m$ is a basis of W. Prove that $w_1 + U, \dots, w_m + U$ is a basis of V/U.

Solution. Each $w_k + U$ belongs to V/U by 3.99. We must show that the list $w_1 + U, \dots, w_m + U$ spans V/U and is linearly independent.
Spanning. Let $x \in V/U$. By 3.99 there is $v \in V$ with $x = v + U$. Because $V = U \oplus W$, in particular $V = U + W$, so there exist $u \in U$ and $w \in W$ with

$$v = u + w.$$

Because $w_1, \dots, w_m$ is a basis of W, there exist $a_1, \dots, a_m \in \mathbf{F}$ with

$$w = a_1 w_1 + \cdots + a_m w_m.$$

Then

$$v - (a_1 w_1 + \cdots + a_m w_m) = v - w = u \in U,$$

so 3.101 gives

$$v + U = (a_1 w_1 + \cdots + a_m w_m) + U.$$

By the definition 3.102 of the operations on V/U, the right side equals

$$a_1(w_1 + U) + \cdots + a_m(w_m + U).$$

Hence

$$x = v + U = a_1(w_1 + U) + \cdots + a_m(w_m + U),$$

so $x \in \operatorname{span}(w_1 + U, \dots, w_m + U)$. Since $x \in V/U$ was arbitrary, the list $w_1 + U, \dots, w_m + U$ spans V/U.
Linear independence. Suppose $a_1, \dots, a_m \in \mathbf{F}$ satisfy

$$a_1(w_1 + U) + \cdots + a_m(w_m + U) = 0 + U,$$

where $0 + U = U$ is the additive identity of V/U (see 3.103). By 3.102 the left side equals $(a_1 w_1 + \cdots + a_m w_m) + U$, so 3.101 gives

$$a_1 w_1 + \cdots + a_m w_m \in U.$$

On the other hand $a_1 w_1 + \cdots + a_m w_m \in W$, because $w_1, \dots, w_m \in W$ and W is a subspace. Hence

$$a_1 w_1 + \cdots + a_m w_m \in U \cap W.$$

Because $V = U \oplus W$, we have $U \cap W = \{0\}$ by 1.46. Therefore

$$a_1 w_1 + \cdots + a_m w_m = 0,$$

and the linear independence of the basis $w_1, \dots, w_m$ of W forces $a_1 = \cdots = a_m = 0$. Hence $w_1 + U, \dots, w_m + U$ is linearly independent.
Being a linearly independent spanning list of V/U, the list $w_1 + U, \dots, w_m + U$ is a basis of V/U, as desired.
(As a check on the result, note that it gives $\dim V/U = m = \dim W = \dim V - \dim U$ when V is finite-dimensional, in agreement with 3.105 and 1.46 together with 2.43.)

Problem 3E.15. Suppose U is a subspace of V and $v_1 + U, \dots, v_m + U$ is a basis of V/U and $u_1, \dots, u_n$ is a basis of U. Prove that $v_1, \dots, v_m, u_1, \dots, u_n$ is a basis of V.

Solution. Let $\pi \colon V \to V/U$ denote the quotient map, so $\pi(v) = v + U$ for each $v \in V$ (see 3.104).

Recall from the proof of 3.105 that

$$\operatorname{null}\pi = U,$$

because $\pi(v) = 0 + U$ means $v + U = 0 + U$, which by 3.101 holds if and only if $v - 0 = v \in U$.
First we show that the list spans V. Suppose $v \in V$. Then $v+U \in V/U$, so because $v_1+U, \dots, v_m+U$ spans V/U there exist $a_1, \dots, a_m \in \mathbf{F}$ with

$$v + U = a_1(v_1 + U) + \cdots + a_m(v_m + U) = (a_1v_1 + \cdots + a_mv_m) + U,$$

where the last equality uses the definition 3.102 of addition and scalar multiplication on V/U. By 3.101 this gives

$$v - (a_1v_1 + \cdots + a_mv_m) \in U.$$

Because $u_1, \dots, u_n$ spans U, there exist $b_1, \dots, b_n \in \mathbf{F}$ with $v-(a_1v_1+\cdots+a_mv_m) = b_1u_1+\cdots+b_nu_n$. Hence

$$v = a_1v_1 + \cdots + a_mv_m + b_1u_1 + \cdots + b_nu_n,$$

so $v \in \operatorname{span}(v_1, \dots, v_m, u_1, \dots, u_n)$. Because $v \in V$ was arbitrary, the list $v_1, \dots, v_m, u_1, \dots, u_n$ spans V.
Now we show that the list is linearly independent. Suppose $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbf{F}$ satisfy

$$a_1v_1 + \cdots + a_mv_m + b_1u_1 + \cdots + b_nu_n = 0.$$

Apply the linear map π to both sides. Since each $u_j \in U = \operatorname{null}\pi$, we get $\pi(u_j) = 0$, and therefore

$$a_1(v_1 + U) + \cdots + a_m(v_m + U) = 0 + U,$$

the additive identity of V/U. Because $v_1 + U, \dots, v_m + U$ is a basis of V/U and hence linearly independent, we conclude $a_1 = \cdots = a_m = 0$. The displayed equation above now reduces to

$$b_1u_1 + \cdots + b_nu_n = 0,$$

and the linear independence of the basis $u_1, \dots, u_n$ of U gives $b_1 = \cdots = b_n = 0$. Thus all the coefficients vanish, so $v_1, \dots, v_m, u_1, \dots, u_n$ is linearly independent.
Being a spanning list of V that is linearly independent, $v_1, \dots, v_m, u_1, \dots, u_n$ is a basis of V (by the definition 2.26 of basis).
Note that this exercise in particular shows that if U and V/U are both finite-dimensional, then so is V, and $\dim V = \dim V/U + \dim U$, in agreement with 3.105.

Problem 3E.16. Suppose $\varphi \in \mathcal{L}(V, \mathbf{F})$ and $\varphi \neq 0$. Prove that $\dim V/(\operatorname{null}\varphi) = 1$.

Solution. Note that V is not assumed to be finite-dimensional, so we cannot simply invoke 3.105; instead we use 3.107.
We first determine $\operatorname{range}\varphi$. It is a subspace of $\mathbf{F}$ (by 3.18). Regarded as a vector space over itself, $\mathbf{F}$ has the one-element list 1 as a basis, since every $\lambda \in \mathbf{F}$ equals $\lambda \cdot 1$ and $\lambda \cdot 1 = 0$ forces $\lambda = 0$. Hence $\dim \mathbf{F} = 1$.
Because $\varphi \neq 0$, there exists $v \in V$ with $\varphi(v) \neq 0$. Then for every $\lambda \in \mathbf{F}$ we have

$$\varphi\left(\frac{\lambda}{\varphi(v)}\, v\right) = \frac{\lambda}{\varphi(v)}\,\varphi(v) = \lambda,$$

where the division makes sense because $\varphi(v) \neq 0$. Thus every element of $\mathbf{F}$ is in $\operatorname{range}\varphi$, so

$$\operatorname{range}\varphi = \mathbf{F},$$

and therefore $\dim \operatorname{range}\varphi = 1$.
By 3.107(d), the map $\widetilde{\varphi}$ of 3.106 is an isomorphism from $V/(\operatorname{null}\varphi)$ onto $\operatorname{range}\varphi = \mathbf{F}$. An isomorphism is by definition an invertible linear map (3.69), so $\widetilde{\varphi}^{-1}$ is a linear map from $\mathbf{F}$ onto $V/(\operatorname{null}\varphi)$, and it too is invertible, hence injective and surjective by 3.63.

Now use the general fact that an isomorphism carries a basis to a basis: if Φ is an isomorphism from one space onto another and $e_1, \dots, e_k$ is a basis of the domain, then $\Phi e_1, \dots, \Phi e_k$ is a basis of the target (surjectivity of Φ gives spanning, and injectivity of Φ together with its linearity gives linear independence, since $a_1\Phi e_1 + \cdots + a_k \Phi e_k = 0$ implies $\Phi(a_1 e_1 + \cdots + a_k e_k) = 0$, hence $a_1 e_1 + \cdots + a_k e_k = 0$, hence every $a_j = 0$). It is essential here to apply this to $\widetilde{\varphi}^{-1}$ rather than to $\widetilde{\varphi}$: we possess a basis of $\mathbf{F}$, namely the one-vector list 1, whereas a basis of $V/(\operatorname{null}\varphi)$ is exactly what we are trying to produce, and V is not assumed finite-dimensional so we may not presume one exists. Transporting the basis 1 of $\mathbf{F}$ through $\widetilde{\varphi}^{-1}$ shows that the one-vector list $\widetilde{\varphi}^{-1}(1)$ is a basis of $V/(\operatorname{null}\varphi)$. Hence

$$\dim V/(\operatorname{null}\varphi) = 1,$$

as desired.
Here is a direct argument that avoids invoking 3.107 at all. With v as above, every $x \in V$ satisfies

$$x - \frac{\varphi(x)}{\varphi(v)}\, v \in \operatorname{null}\varphi,$$

since φ of the left side is $\varphi(x) - \varphi(x) = 0$. Hence by 3.101,

$$x + \operatorname{null}\varphi = \frac{\varphi(x)}{\varphi(v)}\,\big(v + \operatorname{null}\varphi\big),$$

so the single vector $v + \operatorname{null}\varphi$ spans $V/(\operatorname{null}\varphi)$. It is also linearly independent, because $\lambda(v + \operatorname{null}\varphi) = 0 + \operatorname{null}\varphi$ means $\lambda v \in \operatorname{null}\varphi$, so $\lambda\varphi(v) = 0$, so $\lambda = 0$. Thus $v + \operatorname{null}\varphi$ is a basis of $V/(\operatorname{null}\varphi)$, which therefore has dimension 1.

Problem 3E.17. Suppose U is a subspace of V such that $\dim V/U = 1$. Prove that there exists $\varphi \in \mathcal{L}(V, \mathbf{F})$ such that $\operatorname{null}\varphi = U$.

Solution. Let $\pi\colon V \to V/U$ be the quotient map, defined by $\pi(v) = v + U$ (see 3.104); it is a linear map. As noted in the proof of 3.105, 3.101 gives

$$\operatorname{null}\pi = U,$$

because $\pi(v) = 0 + U$ if and only if $v \in U$.
Because $\dim V/U = 1$, the space V/U has a basis consisting of one vector; write this basis as $v + U$ for some $v \in V$. Thus every element of V/U can be written in exactly one way as $\lambda(v + U)$ with $\lambda \in \mathbf{F}$: it can be so written because $v + U$ spans V/U, and uniquely so because $\lambda(v + U) = \mu(v + U)$ implies $(\lambda - \mu)(v + U) = 0$, which by linear independence forces $\lambda = \mu$.
Hence we may define a function $\alpha\colon V/U \to \mathbf{F}$ by

$$\alpha\big(\lambda(v + U)\big) = \lambda \quad \text{for each } \lambda \in \mathbf{F};$$

the uniqueness just established shows that α is well defined on all of V/U. (This is the map given by 3.4, the existence of a linear map taking prescribed values on a basis, applied to the basis $v + U$ of V/U and the value $1 \in \mathbf{F}$.)
The map α is linear: if $\lambda, \mu, c \in \mathbf{F}$ then

$$\alpha\big(\lambda(v + U) + \mu(v + U)\big) = \alpha\big((\lambda + \mu)(v + U)\big) = \lambda + \mu = \alpha\big(\lambda(v + U)\big) + \alpha\big(\mu(v + U)\big)$$

and

$$\alpha\big(c \cdot \lambda(v + U)\big) = \alpha\big((c\lambda)(v + U)\big) = c\lambda = c\,\alpha\big(\lambda(v + U)\big).$$

The map α is injective: if $\alpha(\lambda(v + U)) = 0$ then $\lambda = 0$, so $\lambda(v + U) = 0$; thus $\operatorname{null}\alpha = \{0\}$, and α is injective by 3.15.
Now define

$$\varphi = \alpha \circ \pi.$$

Then φ is a linear map from V to $\mathbf{F}$, because the composition (product) of linear maps is linear (see 3.7). Finally,

$$\operatorname{null}\varphi = \{x \in V : \alpha(\pi x) = 0\} = \{x \in V : \pi x = 0\} = \operatorname{null}\pi = U,$$

where the second equality holds because α is injective and $\alpha(0) = 0$, so $\alpha(\pi x) = 0$ if and only if $\pi x = 0$. Thus $\varphi \in \mathcal{L}(V, \mathbf{F})$ satisfies $\operatorname{null}\varphi = U$, as desired. Note that $\varphi \neq 0$, since $\varphi(v) = \alpha(v + U) = 1$; combined with the previous exercise, this shows that the subspaces U of V with $\dim V/U = 1$ are exactly the null spaces of the nonzero linear functionals on V.

Problem 3E.18. Suppose that U is a subspace of V such that V/U is finite-dimensional.
(a) Show that if W is a finite-dimensional subspace of V and $V = U + W$, then $\dim W \geq \dim V/U$.
(b) Prove that there exists a finite-dimensional subspace W of V such that $\dim W = \dim V/U$ and $V = U \oplus W$.

Solution. Throughout, let $\pi\colon V \to V/U$ denote the quotient map $\pi(v) = v + U$ (see 3.104), which is linear, and recall from 3.101 that $\pi(v) = 0$ if and only if $v \in U$; thus $\operatorname{null}\pi = U$.
(a) Suppose W is a finite-dimensional subspace of V with $V = U + W$. Let $\pi|_W$ denote the restriction of π to W; then $\pi|_W \in \mathcal{L}(W, V/U)$.
We claim $\pi|_W$ is surjective. Indeed, let $v+U$ be an arbitrary element of V/U. Since $V = U+W$, we may write $v = u+w$ with $u \in U$ and $w \in W$. Then $v - w = u \in U$, so 3.101 gives $v+U = w+U = \pi|_W(w)$. Hence

$$\operatorname{range}\pi|_W = V/U.$$

Because W is finite-dimensional, the fundamental theorem of linear maps (3.21) applied to $\pi|_W$ gives

$$\dim W = \dim\operatorname{null}\pi|_W + \dim\operatorname{range}\pi|_W = \dim(U \cap W) + \dim V/U \geq \dim V/U,$$

where we used $\operatorname{null}\pi|_W = W \cap \operatorname{null}\pi = U \cap W$ and the fact that dimensions are nonnegative. This proves (a).
(b) Let $n = \dim V/U$ and let $v_1 + U, \ldots, v_n + U$ be a basis of V/U; such a basis exists because V/U is finite-dimensional (by 2.31), and each of its vectors has the form $v_j + U$ for some $v_j \in V$ because the quotient map π is surjective. The case $n = 0$ is permitted throughout what follows, all lists then being empty: in that case $V/U = \{0\}$, so 3.101 forces $V = U$, and the construction below produces $W = \operatorname{span}() = \{0\}$, for which $\dim W = 0 = \dim V/U$ and $V = U \oplus \{0\}$. Define

$$W = \operatorname{span}(v_1, \ldots, v_n),$$

which is a finite-dimensional subspace of V (by 2.6 and the definition of finite-dimensional).
The list $v_1, \ldots, v_n$ is linearly independent in V: if $a_1v_1 + \cdots + a_nv_n = 0$ for some $a_1, \ldots, a_n \in \mathbf{F}$, then applying π gives

$$a_1(v_1 + U) + \cdots + a_n(v_n + U) = 0 + U,$$

and the linear independence of the basis $v_1 + U, \ldots, v_n + U$ of V/U forces $a_1 = \cdots = a_n = 0$. Hence $v_1, \ldots, v_n$ is a basis of W and

$$\dim W = n = \dim V/U.$$

Next, $V = U+W$. Indeed, let $v \in V$. Because $v_1+U, \ldots, v_n+U$ spans V/U, there exist $a_1, \ldots, a_n \in \mathbf{F}$ with

$$v + U = a_1(v_1 + U) + \cdots + a_n(v_n + U) = (a_1v_1 + \cdots + a_nv_n) + U.$$

By 3.101 this means $v - (a_1v_1 + \cdots + a_nv_n) \in U$, so

$$v = \underbrace{\big(v - (a_1v_1 + \cdots + a_nv_n)\big)}_{\in U} + \underbrace{(a_1v_1 + \cdots + a_nv_n)}_{\in W} \in U + W.$$

Finally, $U \cap W = \{0\}$. Suppose $w \in U \cap W$. Because $w \in W$, we can write $w = a_1v_1 + \cdots + a_nv_n$ for some $a_1, \ldots, a_n \in \mathbf{F}$. Because $w \in U$, we have $\pi(w) = 0$, that is,

$$a_1(v_1 + U) + \cdots + a_n(v_n + U) = 0 + U.$$

Linear independence of $v_1 + U, \dots, v_n + U$ gives $a_1 = \cdots = a_n = 0$, hence $w = 0$.
Because $V = U + W$ and $U \cap W = \{0\}$, the sum $U + W$ is a direct sum by 1.46, so

$$V = U \oplus W \quad \text{and} \quad \dim W = \dim V/U,$$

as desired.
Part (b) shows that the lower bound in part (a) is attained, and moreover it is attained by a subspace W that is a complement of U in V. Combining the two parts: among all finite-dimensional subspaces W of V with $V = U + W$, the minimum possible dimension is exactly $\dim V/U$.

Problem 3E.19. Suppose $T \in \mathcal{L}(V, W)$ and U is a subspace of V. Let π denote the quotient map from V onto V/U. Prove that there exists $S \in \mathcal{L}(V/U, W)$ such that $T = S \circ \pi$ if and only if $U \subseteq \operatorname{null} T$.

Solution. Recall that $\pi \in \mathcal{L}(V, V/U)$ is given by $\pi(v) = v + U$ (see 3.104), and that by 3.101 we have $\pi(v) = 0 + U$ if and only if $v \in U$; in other words $\operatorname{null} \pi = U$.
First suppose there exists $S \in \mathcal{L}(V/U, W)$ with $T = S \circ \pi$. Let $u \in U$. Then $\pi(u) = u + U = 0 + U$, which is the additive identity of V/U. Hence

$$Tu = (S \circ \pi)(u) = S(\pi(u)) = S(0) = 0,$$

the last equality because every linear map sends 0 to 0 (by 3.10). Thus $u \in \operatorname{null} T$. Since $u \in U$ was arbitrary, $U \subseteq \operatorname{null} T$.
Conversely, suppose $U \subseteq \operatorname{null} T$. We want to define $S \colon V/U \to W$ by

$$S(v + U) = Tv \quad \text{for each } v \in V.$$

This definition makes sense (that is, S is well defined). Suppose $v, v' \in V$ satisfy $v + U = v' + U$. By 3.101 this gives $v - v' \in U$, and since $U \subseteq \operatorname{null} T$ we get $T(v - v') = 0$, hence $Tv = Tv'$. So the value assigned to the element $v + U$ of V/U does not depend on which representative v is used, and every element of V/U has the form $v + U$ for some $v \in V$. Thus S is a well-defined function from V/U to W.
The function S is linear. For $v, v' \in V$, using the definition 3.102 of addition on V/U and the additivity of T,

$$S\big((v + U) + (v' + U)\big) = S\big((v + v') + U\big) = T(v + v') = Tv + Tv' = S(v + U) + S(v' + U).$$

For $\lambda \in \mathbf{F}$ and $v \in V$, using the definition 3.102 of scalar multiplication on V/U and the homogeneity of T,

$$S\big(\lambda(v + U)\big) = S\big((\lambda v) + U\big) = T(\lambda v) = \lambda\, Tv = \lambda\, S(v + U).$$

Hence $S \in \mathcal{L}(V/U, W)$.
Finally, for each $v \in V$,

$$(S \circ \pi)(v) = S(\pi(v)) = S(v + U) = Tv,$$

so $T = S \circ \pi$, completing the proof of the converse direction and hence of the equivalence.
Two remarks. First, the map S produced above is unique: if $S' \in \mathcal{L}(V/U, W)$ also satisfies $T = S' \circ \pi$, then for every $v \in V$ we have $S'(v + U) = (S' \circ \pi)(v) = Tv = S(v + U)$, and since every element of V/U has the form $v + U$, we get $S' = S$. Second, in the special case $U = \operatorname{null} T$ the hypothesis $U \subseteq \operatorname{null} T$ holds automatically, and the map S constructed here is exactly the map $\widetilde{T}$ of 3.106; the identity $T = S \circ \pi$ then becomes 3.107(a).

3.6 Exercises 3F

Problem 3F.1. Explain why each linear functional is surjective or is the zero map.

Solution. Let V be a vector space over $\mathbf{F}$ and let $\varphi \in V'$, so φ is a linear map from V to $\mathbf{F}$. By 3.18, $\operatorname{range}\varphi$ is a subspace of $\mathbf{F}$.
Now $\mathbf{F}$, regarded as a vector space over $\mathbf{F}$, has dimension 1 (the list consisting of the single vector 1 is a basis of $\mathbf{F}$). Hence if M is a subspace of $\mathbf{F}$, then $\dim M \le 1$ by 2.37, so $\dim M = 0$ or $\dim M = 1$. If $\dim M = 0$ then $M = \{0\}$, and if $\dim M = 1$ then $M = \mathbf{F}$ by 2.39. Thus the only subspaces of $\mathbf{F}$ are $\{0\}$ and $\mathbf{F}$.
Applying this to $M = \operatorname{range}\varphi$ gives two cases. If $\operatorname{range}\varphi = \{0\}$, then $\varphi(v) = 0$ for every $v \in V$, which means φ is the zero map. If $\operatorname{range}\varphi = \mathbf{F}$, then φ is surjective by the definition of surjectivity (3.19). Hence every linear functional is surjective or is the zero map.
Here is the same argument written without reference to dimension, which shows that no finite-dimensionality hypothesis of any kind is needed. Suppose $\varphi \in V'$ is not the zero map. Then there exists $u \in V$ with $\varphi(u) \neq 0$. Set $a = \varphi(u) \in \mathbf{F}$, so $a \neq 0$ and a has a multiplicative inverse in $\mathbf{F}$. Given any $\lambda \in \mathbf{F}$, the vector $v = (\lambda/a)u \in V$ satisfies

$$\varphi(v) = \varphi\left(\frac{\lambda}{a}\,u\right) = \frac{\lambda}{a}\,\varphi(u) = \frac{\lambda}{a}\cdot a = \lambda,$$

where the second equality uses the homogeneity of the linear map φ. Thus every $\lambda \in \mathbf{F}$ is in $\operatorname{range}\varphi$, so φ is surjective.
Note that the crucial feature of the situation is that the target space $\mathbf{F}$ is one-dimensional; a linear map into a space of dimension at least 2 can of course be neither surjective nor the zero map.

Problem 3F.2. Give three distinct examples of linear functionals on $\mathbf{R}^{[0,1]}$.

Solution. Recall that $\mathbf{R}^{[0,1]}$ denotes the vector space of all functions from the interval $[0,1]$ to $\mathbf{R}$, with addition and scalar multiplication defined pointwise:

$$(f+g)(x) = f(x) + g(x), \qquad (\lambda f)(x) = \lambda\, f(x)$$

for all $f, g \in \mathbf{R}^{[0,1]}$, all $\lambda \in \mathbf{R}$, and all $x \in [0,1]$. A linear functional on this space is a linear map from $\mathbf{R}^{[0,1]}$ to $\mathbf{R}$.
Define $\varphi_1, \varphi_2, \varphi_3 \colon \mathbf{R}^{[0,1]} \to \mathbf{R}$ by

$$\varphi_1(f) = f(0), \qquad \varphi_2(f) = f(1), \qquad \varphi_3(f) = f(0) + f(1).$$

Each is a linear functional. Indeed, let $f, g \in \mathbf{R}^{[0,1]}$ and $\lambda \in \mathbf{R}$, and fix $t \in [0,1]$. By the pointwise definitions above,

$$(f+g)(t) = f(t) + g(t) \qquad \text{and} \qquad (\lambda f)(t) = \lambda\, f(t).$$

So the evaluation map $f \mapsto f(t)$ is additive and homogeneous, hence a linear functional on $\mathbf{R}^{[0,1]}$. Taking $t = 0$ gives φ_1 and taking $t = 1$ gives φ_2. Finally $\varphi_3 = \varphi_1 + \varphi_2$, and a sum of linear maps is again a linear map (3.5, and 3.6, which says $\mathcal{L}(V, W)$ is a vector space), so φ_3 is a linear functional as well.
These three functionals are distinct. Let f be the function on $[0,1]$ defined by $f(0) = 1$ and $f(x) = 0$ for $x \in (0,1]$, and let g be defined by $g(1) = 1$ and $g(x) = 0$ for $x \in [0,1)$. Then

$$\varphi_1(f) = 1, \quad \varphi_2(f) = 0, \quad \varphi_3(f) = 1,$$

$$\varphi_1(g) = 0, \quad \varphi_2(g) = 1, \quad \varphi_3(g) = 1.$$

From the values at f we get $\varphi_1 \neq \varphi_2$ and $\varphi_2 \neq \varphi_3$; from the values at g we get $\varphi_1 \neq \varphi_3$. Hence the three examples are pairwise distinct.

Two remarks. First, the same argument shows that for any $t_1, \dots, t_k \in [0,1]$ and any $c_1, \dots, c_k \in \mathbf{R}$, the map $f \mapsto c_1 f(t_1) + \cdots + c_k f(t_k)$ is a linear functional on $\mathbf{R}^{[0,1]}$; this manufactures infinitely many distinct examples. Second, the familiar functional $f \mapsto \int_0^1 f$ is not available here, because $\mathbf{R}^{[0,1]}$ contains all functions from $[0,1]$ to $\mathbf{R}$, including functions that are not Riemann integrable (for instance the indicator function of the rationals in $[0,1]$), so that formula does not define a function on all of $\mathbf{R}^{[0,1]}$. It does define a linear functional on the subspace of continuous real-valued functions on $[0,1]$.

Problem 3F.3. Suppose V is finite-dimensional and $v \in V$ with $v \neq 0$. Prove that there exists $\varphi \in V'$ such that $\varphi(v) = 1$.

Solution. First note that the length-one list v is linearly independent. Indeed, if $a \in \mathbf{F}$ and $av = 0$ with $a \neq 0$, then $v = a^{-1}(av) = a^{-1}0 = 0$, contradicting the hypothesis $v \neq 0$; hence $av = 0$ forces $a = 0$.
Because V is finite-dimensional, 2.32 tells us that the linearly independent list v can be extended to a basis of V. Thus there is a basis $v_1, \dots, v_n$ of V with $v_1 = v$.
Now apply the linear map lemma (3.4): since $v_1, \dots, v_n$ is a basis of V, there exists a unique linear map $\varphi\colon V \to \mathbf{F}$ satisfying

$$\varphi(v_1) = 1 \qquad \text{and} \qquad \varphi(v_j) = 0 \quad \text{for } j = 2, \dots, n.$$

Since φ is a linear map from V to $\mathbf{F}$, it is a linear functional on V, that is, $\varphi \in V'$. And $\varphi(v) = \varphi(v_1) = 1$, as required.
Equivalently, one can phrase the last step using the dual basis: let $\varphi_1, \dots, \varphi_n \in V'$ be the dual basis of $v_1, \dots, v_n$ (3.112), which by definition satisfies $\varphi_j(v_k) = 1$ if $j = k$ and $\varphi_j(v_k) = 0$ if $j \neq k$. Then $\varphi = \varphi_1$ works, since $\varphi_1(v) = \varphi_1(v_1) = 1$.
Observe that once such a φ exists, it is far from unique when $\dim V > 1$: for example, in the notation above, $\varphi_1 + \varphi_2$ also sends v to 1. Note also that the conclusion fails for $v = 0$, since every linear map sends 0 to 0 (3.10). Finally, finite-dimensionality is used only to guarantee that the linearly independent list v extends to a basis; the result holds for arbitrary vector spaces if one is willing to use the fact (proved with the axiom of choice, and not needed in this book) that every vector space has a basis.

Problem 3F.4. Suppose V is finite-dimensional and U is a subspace of V such that $U \neq V$. Prove that there exists $\varphi \in V'$ such that $\varphi(u) = 0$ for every $u \in U$ but $\varphi \neq 0$.

Solution. Because V is finite-dimensional, so is its subspace U (2.25), and $\dim U \leq \dim V$ (2.37). Moreover $\dim U \neq \dim V$: if $\dim U = \dim V$, then $U = V$ by 2.39, contradicting the hypothesis. Hence

$$m := \dim U < \dim V =: n.$$

Let $u_1, \dots, u_m$ be a basis of U (if $U = \{0\}$ this is the empty list, and the argument below goes through unchanged). Since $u_1, \dots, u_m$ is a linearly independent list in V, it extends to a basis

$$u_1, \dots, u_m, v_{m+1}, \dots, v_n$$

of V by 2.32. Because $m < n$, the vector v_{m+1} really is present in this list.
By the linear map lemma (3.4), there is a (unique) linear map $\varphi\colon V \to \mathbf{F}$ with

$$\varphi(u_j) = 0 \ \text{ for } j = 1, \dots, m, \qquad \varphi(v_{m+1}) = 1, \qquad \varphi(v_k) = 0 \ \text{ for } k = m+2, \dots, n,$$

since the values of a linear map may be assigned arbitrarily on a basis. As a linear map from V to $\mathbf{F}$, we have $\varphi \in V'$.
This φ has the two required properties.
First, φ vanishes on U. Let $u \in U$. Since $u_1, \dots, u_m$ is a basis of U, there exist $a_1, \dots, a_m \in \mathbf{F}$ with $u = a_1 u_1 + \cdots + a_m u_m$. Then, by linearity of φ,

$$\varphi(u) = a_1 \varphi(u_1) + \cdots + a_m \varphi(u_m) = a_1 \cdot 0 + \cdots + a_m \cdot 0 = 0.$$

(In the degenerate case $m = 0$, where $U = \{0\}$ and the list $u_1, \dots, u_m$ is empty, this reads: the only $u \in U$ is $u = 0$, and $\varphi(0) = 0$ by 3.10. Note also that $m < n$ holds automatically here, since $U \neq V$ forces $V \neq \{0\}$ and hence $n \geq 1$.)
Second, $\varphi \neq 0$, because $\varphi(v_{m+1}) = 1 \neq 0$, whereas the zero linear functional takes the value 0 at every vector.
In the language of this section, the conclusion says precisely that $U^0 \neq \{0\}$, and the exercise is thus the implication $U \neq V \Rightarrow U^0 \neq \{0\}$, which is the contrapositive of one direction of 3.127(a). The proof above is a direct one, using only 2.32 and the linear map lemma, rather than the dimension formula 3.125.
Note also that the hypothesis $U \neq V$ cannot be dropped: if $U = V$, then a functional vanishing on all of U vanishes on all of V and is therefore the zero functional. Finally, the construction above yields slightly more than was asked: the functional φ produced satisfies $\varphi(v_{m+1}) = 1$ for a vector $v_{m+1} \notin U$, so it not only fails to be 0 but takes the value 1 somewhere outside U. In this respect the proof runs parallel to that of Exercise 3F.3, with the basis of U playing the role that the single vector v played there.

Problem 3F.5. Suppose $T \in \mathcal{L}(V, W)$ and $w_1, \dots, w_m$ is a basis of range T. Hence for each $v \in V$, there exist unique numbers $\varphi_1(v), \dots, \varphi_m(v)$ such that

$$Tv = \varphi_1(v)w_1 + \cdots + \varphi_m(v)w_m,$$

thus defining functions $\varphi_1, \dots, \varphi_m$ from V to $\mathbf{F}$. Show that each of the functions $\varphi_1, \dots, \varphi_m$ is a linear functional on V.

Solution. First let us be clear about why the functions $\varphi_1, \dots, \varphi_m$ are well defined. If $v \in V$, then $Tv \in \operatorname{range} T$, and $w_1, \dots, w_m$ is a basis of range T; by 2.28, every vector in range T has a unique representation as a linear combination of $w_1, \dots, w_m$. So there is exactly one m-tuple of scalars, which we name $\varphi_1(v), \dots, \varphi_m(v)$, with

$$Tv = \varphi_1(v)w_1 + \cdots + \varphi_m(v)w_m.$$

This assigns to each $v \in V$ exactly one scalar $\varphi_j(v)$ for each j, so each φ_j is a function from V to $\mathbf{F}$. This uniqueness is the engine of the whole proof: it lets us conclude equality of coefficients from equality of vectors.
Additivity. Suppose $u, v \in V$. Applying the defining equation to u, to v, and to $u + v$, and using the additivity of T, we get

$$\begin{aligned} \varphi_1(u+v)w_1 + \cdots + \varphi_m(u+v)w_m &= T(u+v) \\ &= Tu + Tv \\ &= \big(\varphi_1(u)w_1 + \cdots + \varphi_m(u)w_m\big) + \big(\varphi_1(v)w_1 + \cdots + \varphi_m(v)w_m\big) \\ &= \big(\varphi_1(u) + \varphi_1(v)\big)w_1 + \cdots + \big(\varphi_m(u) + \varphi_m(v)\big)w_m, \end{aligned}$$

where the last step just collects terms. Now the first and last lines exhibit the vector $T(u+v) \in \operatorname{range} T$ as two linear combinations of the basis $w_1, \dots, w_m$. By the uniqueness in 2.28, the corresponding coefficients agree:

$$\varphi_j(u+v) = \varphi_j(u) + \varphi_j(v) \qquad \text{for each } j = 1, \dots, m.$$

Homogeneity. Suppose $\lambda \in \mathbf{F}$ and $v \in V$. Using the homogeneity of T,

$$\begin{aligned} \varphi_1(\lambda v)w_1 + \cdots + \varphi_m(\lambda v)w_m &= T(\lambda v) \\ &= \lambda(Tv) \\ &= \lambda\big(\varphi_1(v)w_1 + \cdots + \varphi_m(v)w_m\big) \\ &= \big(\lambda\varphi_1(v)\big)w_1 + \cdots + \big(\lambda\varphi_m(v)\big)w_m. \end{aligned}$$

Again the first and last lines are two representations of the same vector of range T in terms of the basis $w_1, \dots, w_m$, so by uniqueness

$$\varphi_j(\lambda v) = \lambda\varphi_j(v) \qquad \text{for each } j = 1, \dots, m.$$

Thus each φ_j is a function from V to $\mathbf{F}$ that is additive and homogeneous, hence a linear map from V to $\mathbf{F}$; that is, each φ_j is a linear functional on V (3.108), so $\varphi_1, \dots, \varphi_m \in V'$.
A remark that identifies these functionals. Let $\psi_1, \dots, \psi_m$ be the dual basis of $w_1, \dots, w_m$ in $(\operatorname{range} T)'$. Then 3.114 says exactly that $\psi_j(w)$ is the j-th coefficient of w in the basis $w_1, \dots, w_m$, so $\varphi_j(v) = \psi_j(Tv)$, that is,

$$\varphi_j = \psi_j \circ T = S'(\psi_j),$$

where $S \in \mathcal{L}(V, \operatorname{range} T)$ is T with its target space narrowed to $\operatorname{range} T$. Since the product (composition) of two linear maps is again a linear map (3.7), this gives a second, one-line proof of linearity. Note that no finite-dimensionality of V or W was needed anywhere; all that was used is that $\operatorname{range} T$ has the finite basis $w_1, \dots, w_m$.

Problem 3F.6. Suppose $\varphi, \beta \in V'$. Prove that $\operatorname{null} \varphi \subseteq \operatorname{null} \beta$ if and only if there exists $c \in \mathbf{F}$ such that $\beta = c\varphi$.

Solution. We prove the two implications separately. Note that V is not assumed to be finite-dimensional, and none of what follows uses such an assumption.
First suppose there exists $c \in \mathbf{F}$ with $\beta = c\varphi$. Let $v \in \operatorname{null} \varphi$, so $\varphi(v) = 0$. Then

$$\beta(v) = (c\varphi)(v) = c\,\varphi(v) = c \cdot 0 = 0,$$

where the second equality is the definition of scalar multiplication in V' (3.5). Hence $v \in \operatorname{null} \beta$. This proves $\operatorname{null} \varphi \subseteq \operatorname{null} \beta$.
Now suppose $\operatorname{null} \varphi \subseteq \operatorname{null} \beta$; we must produce $c \in \mathbf{F}$ with $\beta = c\varphi$. There are two cases.
Case 1: $\varphi = 0$. Then $\operatorname{null} \varphi = V$, so the hypothesis gives $V = \operatorname{null} \varphi \subseteq \operatorname{null} \beta \subseteq V$, forcing $\operatorname{null} \beta = V$, i.e. $\beta = 0$. Thus $\beta = 0 = 0 \cdot \varphi$, and $c = 0$ works (indeed every $c \in \mathbf{F}$ works in this degenerate case).
Case 2: $\varphi \neq 0$. By Exercise 3F.1, a nonzero linear functional is surjective, so there exists $u \in V$ with

$$\varphi(u) = 1.$$

(Concretely: choose x with $\varphi(x) = a \neq 0$ and set $u = x/a$.) Define

$$c = \beta(u) \in \mathbf{F}.$$

We claim $\beta = c\varphi$. Let $v \in V$ be arbitrary and put

$$w = v - \varphi(v)\,u.$$

Then, using linearity of φ and $\varphi(u) = 1$,

$$\varphi(w) = \varphi(v) - \varphi(v)\,\varphi(u) = \varphi(v) - \varphi(v) = 0,$$

so $w \in \operatorname{null} \varphi$. By hypothesis $w \in \operatorname{null} \beta$, that is, $\beta(w) = 0$. Applying the linear functional β to $v = w + \varphi(v)u$ therefore yields

$$\beta(v) = \beta(w) + \varphi(v)\,\beta(u) = 0 + \varphi(v)\,c = c\,\varphi(v) = (c\varphi)(v).$$

Since $v \in V$ was arbitrary, $\beta = c\varphi$, completing Case 2 and the proof.
Two comments. First, in Case 2 the scalar c is unique: if $c_1\varphi = c_2\varphi$ with $\varphi \neq 0$, pick u with $\varphi(u) = 1$ to get $c_1 = c_2$. Second, the geometric content is that a nonzero linear functional is determined up to a scalar by its null space: by 3.21 (when V is finite-dimensional) the null space of a nonzero $\varphi \in V'$ has dimension $\dim V - 1$, i.e. it is a hyperplane, and the exercise says that β vanishes on that hyperplane exactly when β is a multiple of φ. The decomposition $v = w + \varphi(v)u$ used above is precisely the direct sum decomposition $V = \operatorname{null} \varphi \oplus \operatorname{span}(u)$.

Problem 3F.7. Suppose that $V_1, \dots, V_m$ are vector spaces. Prove that $(V_1 \times \cdots \times V_m)'$ and $V_1' \times \cdots \times V_m'$ are isomorphic vector spaces.

Solution. Write $V = V_1 \times \cdots \times V_m$, whose vector space operations are defined coordinatewise (3.87 and 3.89):

$$(u_1, \dots, u_m) + (v_1, \dots, v_m) = (u_1 + v_1, \dots, u_m + v_m), \qquad \lambda(v_1, \dots, v_m) = (\lambda v_1, \dots, \lambda v_m).$$

The space $V_1' \times \cdots \times V_m'$ is likewise a product of vector spaces, with coordinatewise operations. Since no finite-dimensionality is assumed, we cannot argue by comparing dimensions (3.70); instead we exhibit an explicit isomorphism, that is, an invertible linear map (3.69).
For each $j \in \{1, \dots, m\}$ define the inclusion $\iota_j \colon V_j \to V$ by

$$\iota_j(v) = (0, \dots, 0, v, 0, \dots, 0),$$

with v in the j-th slot and the zero vector of V_k in slot k for $k \neq j$. Each ι_j is linear, because the operations on V are coordinatewise: $\iota_j(u+v) = \iota_j(u) + \iota_j(v)$ and $\iota_j(\lambda v) = \lambda \iota_j(v)$, the zero coordinates behaving correctly since $0 + 0 = 0$ and $\lambda 0 = 0$. Note also the decomposition identity, valid for every $(v_1, \dots, v_m) \in V$:

$$(v_1, \dots, v_m) = \iota_1(v_1) + \cdots + \iota_m(v_m).$$

Define

$$\Gamma \colon V' \to V_1' \times \cdots \times V_m', \qquad \Gamma(\varphi) = (\varphi \circ \iota_1, \dots, \varphi \circ \iota_m).$$

This makes sense: if $\varphi \in V'$, then each $\varphi \circ \iota_j$ is a composition of linear maps $V_j \to V \to \mathbf{F}$, hence a linear map from V_j to $\mathbf{F}$ (3.7), i.e. an element of V_j'.
Γ is linear. Let $\varphi, \psi \in V'$ and $\lambda \in \mathbf{F}$. For each j and each $v \in V_j$,

$$\big((\varphi + \psi) \circ \iota_j\big)(v) = (\varphi + \psi)(\iota_j(v)) = \varphi(\iota_j(v)) + \psi(\iota_j(v)) = (\varphi \circ \iota_j)(v) + (\psi \circ \iota_j)(v),$$

using the definition of addition in V' (3.5); hence $(\varphi+\psi)\circ\iota_j = \varphi\circ\iota_j + \psi\circ\iota_j$. Similarly $(\lambda\varphi)\circ\iota_j = \lambda(\varphi\circ\iota_j)$. Since addition and scalar multiplication in $V_1' \times \cdots \times V_m'$ are coordinatewise, these identities say exactly that

$$\Gamma(\varphi + \psi) = \Gamma(\varphi) + \Gamma(\psi) \qquad \text{and} \qquad \Gamma(\lambda\varphi) = \lambda\,\Gamma(\varphi).$$

Γ is injective. Suppose $\Gamma(\varphi) = 0$, i.e. $\varphi \circ \iota_j = 0$ for each j. Let $(v_1, \dots, v_m) \in V$. By the decomposition identity above and linearity of φ,

$$\varphi(v_1, \dots, v_m) = \varphi(\iota_1(v_1)) + \cdots + \varphi(\iota_m(v_m)) = 0 + \cdots + 0 = 0.$$

Hence $\varphi = 0$, so $\operatorname{null}\Gamma = \{0\}$ and Γ is injective by 3.15.
Γ is surjective. Let $(\varphi_1, \dots, \varphi_m) \in V_1' \times \cdots \times V_m'$. Define $\varphi \colon V \to \mathbf{F}$ by

$$\varphi(v_1, \dots, v_m) = \varphi_1(v_1) + \cdots + \varphi_m(v_m).$$

Then φ is linear: for $(u_1, \dots, u_m), (v_1, \dots, v_m) \in V$ and $\lambda \in \mathbf{F}$,

$$\begin{aligned} \varphi\big((u_1, \dots, u_m) + (v_1, \dots, v_m)\big) &= \varphi_1(u_1 + v_1) + \cdots + \varphi_m(u_m + v_m) \\ &= \big(\varphi_1(u_1) + \cdots + \varphi_m(u_m)\big) + \big(\varphi_1(v_1) + \cdots + \varphi_m(v_m)\big) \\ &= \varphi(u_1, \dots, u_m) + \varphi(v_1, \dots, v_m), \end{aligned}$$

and

$$\varphi\big(\lambda(v_1, \dots, v_m)\big) = \varphi_1(\lambda v_1) + \cdots + \varphi_m(\lambda v_m) = \lambda\big(\varphi_1(v_1) + \cdots + \varphi_m(v_m)\big) = \lambda\,\varphi(v_1, \dots, v_m),$$

each step using linearity of the individual φ_j. Thus $\varphi \in V'$. Moreover, for each j and each $v \in V_j$,

$$(\varphi \circ \iota_j)(v) = \varphi(0, \dots, 0, v, 0, \dots, 0) = \varphi_1(0) + \cdots + \varphi_j(v) + \cdots + \varphi_m(0) = \varphi_j(v),$$

since a linear map sends 0 to 0 (3.10). Hence $\varphi \circ \iota_j = \varphi_j$ for each j, so $\Gamma(\varphi) = (\varphi_1, \dots, \varphi_m)$, proving surjectivity.
Being linear, injective, and surjective, Γ is invertible (3.63), hence an isomorphism (3.69). Therefore

$$(V_1 \times \cdots \times V_m)' \quad \text{and} \quad V_1' \times \cdots \times V_m'$$

are isomorphic vector spaces.
Two remarks. The surjectivity argument in fact exhibits the inverse map explicitly: $\Gamma^{-1}(\varphi_1, \dots, \varphi_m)$ is the functional $(v_1, \dots, v_m) \mapsto \varphi_1(v_1) + \cdots + \varphi_m(v_m)$. And in the finite-dimensional case the result is also visible through dimensions: $\dim(V_1 \times \cdots \times V_m)' = \dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m$ by 3.111 and 3.92, while $\dim(V_1' \times \cdots \times V_m') = \dim V_1' + \cdots + \dim V_m' = \dim V_1 + \cdots + \dim V_m$, so 3.70 gives the isomorphism. The proof above has the advantage of requiring no dimension hypothesis and of being natural, i.e. constructed without choosing bases.

Problem 3F.8. Suppose $v_1, \dots, v_n$ is a basis of V and $\varphi_1, \dots, \varphi_n$ is the dual basis of V'. Define $\Gamma\colon V \to \mathbf{F}^n$ and $\Lambda\colon \mathbf{F}^n \to V$ by

$$\Gamma(v) = (\varphi_1(v), \dots, \varphi_n(v)) \quad \text{and} \quad \Lambda(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n.$$

Explain why Γ and Λ are inverses of each other.

Solution. First note that both maps are linear. Each φ_j is a linear functional on V, so for $u, v \in V$ and $\lambda \in \mathbf{F}$ we have

$$\Gamma(u + \lambda v) = (\varphi_1(u) + \lambda\varphi_1(v), \dots, \varphi_n(u) + \lambda\varphi_n(v)) = \Gamma(u) + \lambda\Gamma(v),$$

because addition and scalar multiplication in $\mathbf{F}^n$ are defined coordinatewise. Similarly Λ is linear, since

$$\Lambda\big((a_1, \dots, a_n) + \lambda(b_1, \dots, b_n)\big) = \sum_{k=1}^{n} (a_k + \lambda b_k) v_k = \sum_{k=1}^{n} a_k v_k + \lambda \sum_{k=1}^{n} b_k v_k.$$

Thus $\Gamma \in \mathcal{L}(V, \mathbf{F}^n)$ and $\Lambda \in \mathcal{L}(\mathbf{F}^n, V)$, and it makes sense to ask whether they are inverses.
Now we show $\Lambda\Gamma$ is the identity on V. If $v \in V$, then

$$(\Lambda \circ \Gamma)(v) = \Lambda(\varphi_1(v), \dots, \varphi_n(v)) = \varphi_1(v) v_1 + \cdots + \varphi_n(v) v_n = v,$$

where the last equality is exactly 3.114 (the dual basis gives the coefficients for writing v as a linear combination of $v_1, \dots, v_n$). Hence $\Lambda\Gamma = I_V$.
Next we show $\Gamma\Lambda$ is the identity on $\mathbf{F}^n$. Let $(a_1, \dots, a_n) \in \mathbf{F}^n$ and put $v = a_1 v_1 + \cdots + a_n v_n$. For each $j \in \{1, \dots, n\}$, linearity of φ_j and the defining property $\varphi_j(v_k) = 1$ if $k = j$ and $\varphi_j(v_k) = 0$ if $k \neq j$ (see 3.112) give

$$\varphi_j(v) = \sum_{k=1}^{n} a_k \varphi_j(v_k) = a_j.$$

Therefore

$$(\Gamma \circ \Lambda)(a_1, \dots, a_n) = \Gamma(v) = (\varphi_1(v), \dots, \varphi_n(v)) = (a_1, \dots, a_n),$$

so $\Gamma\Lambda = I_{\mathbf{F}^n}$.
Because $\Lambda\Gamma = I_V$ and $\Gamma\Lambda = I_{\mathbf{F}^n}$, the map Λ is by definition (3.59) an inverse of Γ, and Γ is an inverse of Λ. Hence both maps are invertible, and since inverses are unique (3.60) we may write $\Gamma^{-1} = \Lambda$ and $\Lambda^{-1} = \Gamma$.
Informally: Λ turns a list of scalars into the vector having those scalars as coordinates with respect to $v_1, \dots, v_n$, and Γ reads off the coordinates of a vector with respect to that same basis. These operations undo each other, which is precisely the content of 3.114. In particular Γ is an invertible linear map from V onto $\mathbf{F}^n$, so it is an isomorphism, and V and $\mathbf{F}^n$ are isomorphic vector spaces whenever $\dim V = n$.

Problem 3F.9. Suppose m is a positive integer. Show that the dual basis of the basis $1, x, \ldots, x^m$ of $\mathcal{P}_m(\mathbf{R})$ is $\varphi_0, \varphi_1, \ldots, \varphi_m$, where

$$\varphi_k(p) = \frac{p^{(k)}(0)}{k!}.$$

Here $p^{(k)}$ denotes the k^{th} derivative of p, with the understanding that the 0^{th} derivative of p is p.

Solution. Write $v_j = x^j$ for $j = 0, 1, \ldots, m$, so that $v_0, v_1, \ldots, v_m$ is the given basis of $\mathcal{P}_m(\mathbf{R})$. By 3.112, the dual basis of this list is the unique list $\varphi_0, \ldots, \varphi_m$ of linear functionals on $\mathcal{P}_m(\mathbf{R})$ satisfying

$$\varphi_k(v_j) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

(Uniqueness holds because a linear map on $\mathcal{P}_m(\mathbf{R})$ is determined by its values on a basis; see the linear map lemma 3.4.) So there are two things to check for the functionals $\varphi_k(p) = p^{(k)}(0)/k!$: that each is a linear functional, and that they satisfy the displayed condition.
Step 1: each φ_k is a linear functional. Fix $k \in \{0, 1, \ldots, m\}$. If $p \in \mathcal{P}_m(\mathbf{R})$ then $p^{(k)}$ is again a polynomial function on $\mathbf{R}$, so $p^{(k)}(0)/k!$ is a well-defined real number; thus φ_k maps $\mathcal{P}_m(\mathbf{R})$ into $\mathbf{R}$. Differentiation is linear, hence so is its k-fold composition: $(p+\lambda q)^{(k)} = p^{(k)} + \lambda q^{(k)}$ for all $p, q \in \mathcal{P}_m(\mathbf{R})$ and $\lambda \in \mathbf{R}$. Evaluation at 0 and division by the fixed nonzero scalar $k!$ are also linear. Hence

$$\varphi_k(p + \lambda q) = \frac{(p + \lambda q)^{(k)}(0)}{k!} = \frac{p^{(k)}(0) + \lambda q^{(k)}(0)}{k!} = \varphi_k(p) + \lambda \varphi_k(q),$$

so $\varphi_k \in \big(\mathcal{P}_m(\mathbf{R})\big)'$.
Step 2: the duality condition. Fix $j, k \in \{0, 1, \ldots, m\}$ and compute $\varphi_k(x^j)$, considering three cases.
If $k > j$, then differentiating x^j more than j times gives the zero polynomial: indeed $(x^j)^{(j)}$ is the constant $j!$, and every further derivative is 0. Hence $(x^j)^{(k)} = 0$, so $\varphi_k(x^j) = 0$.
If $k < j$, then repeated differentiation gives

$$(x^j)^{(k)} = j(j-1)\cdots(j-k+1)\, x^{j-k},$$

and $j - k \geq 1$, so evaluating at 0 gives 0. Hence $\varphi_k(x^j) = 0$.
If $k = j$, then by the formula above with k replaced by j in the last factor, $(x^k)^{(k)} = k(k-1)\cdots 1 = k!$, the constant polynomial. Evaluating at 0 gives $k!$, so

$$\varphi_k(x^k) = \frac{k!}{k!} = 1.$$

Combining the three cases, $\varphi_k(v_j) = 1$ if $j = k$ and $\varphi_k(v_j) = 0$ if $j \neq k$, which is exactly the condition in 3.112. Therefore $\varphi_0, \varphi_1, \ldots, \varphi_m$ is the dual basis of $1, x, \ldots, x^m$, as claimed.
As a consistency check, 3.114 now says that every $p \in \mathcal{P}_m(\mathbf{R})$ satisfies

$$p = \varphi_0(p) \cdot 1 + \varphi_1(p) x + \cdots + \varphi_m(p) x^m = \sum_{k=0}^{m} \frac{p^{(k)}(0)}{k!} x^k,$$

which is the Taylor expansion of p at 0. This is a polynomial identity, not a limiting statement, because the sum terminates.

Problem 3F.10. Suppose m is a positive integer.
(a) Show that $1, x - 5, \ldots, (x-5)^m$ is a basis of $\mathcal{P}_m(\mathbf{R})$.
(b) What is the dual basis of the basis in (a)?

Solution. (a) For $j = 0, 1, \ldots, m$ set $v_j = (x-5)^j$, meaning the polynomial function $x \mapsto (x - 5)^j$. Expanding by the binomial theorem shows $v_j \in \mathcal{P}_m(\mathbf{R})$, since v_j has degree $j \leq m$. The list

$v_0, v_1, \ldots, v_m$ has length $m+1$, and $\dim \mathcal{P}_m(\mathbf{R}) = m+1$ (see 2.36, since the standard basis $1, x, \ldots, x^m$ has length $m+1$). So by 2.38, it suffices to prove that the list is linearly independent.
Suppose $a_0, a_1, \ldots, a_m \in \mathbf{R}$ satisfy

$$a_0 + a_1(x-5) + \cdots + a_m(x-5)^m = 0,$$

where 0 denotes the zero polynomial. Suppose, seeking a contradiction, that not all the a_j equal 0, and let k be the largest index with $a_k \neq 0$. Then the left side equals $\sum_{j=0}^{k} a_j(x-5)^j$. Each term $a_j(x-5)^j$ with $j < k$ is a polynomial of degree at most $k-1$, while $a_k(x-5)^k$ has degree exactly k with leading coefficient a_k (the coefficient of x^k in $(x-5)^k$ is 1). Hence the coefficient of x^k in the sum is $a_k \neq 0$, so the sum is a polynomial of degree k, and in particular is not the zero polynomial. This contradiction shows $a_0 = a_1 = \cdots = a_m = 0$.
Thus $1, x-5, \ldots, (x-5)^m$ is a linearly independent list of length $\dim \mathcal{P}_m(\mathbf{R})$ in $\mathcal{P}_m(\mathbf{R})$, and therefore is a basis of $\mathcal{P}_m(\mathbf{R})$ by 2.38.
(b) The dual basis is $\varphi_0, \varphi_1, \ldots, \varphi_m$, where

$$\varphi_k(p) = \frac{p^{(k)}(5)}{k!}.$$

That is, the dual basis is obtained from the one in Exercise 9 by evaluating derivatives at 5 instead of at 0.
To verify this, first note that each φ_k is a linear functional on $\mathcal{P}_m(\mathbf{R})$: differentiating k times, evaluating at the point 5, and dividing by the fixed nonzero scalar $k!$ are each linear operations, so

$$\varphi_k(p + \lambda q) = \frac{p^{(k)}(5) + \lambda q^{(k)}(5)}{k!} = \varphi_k(p) + \lambda \varphi_k(q)$$

for all $p, q \in \mathcal{P}_m(\mathbf{R})$ and $\lambda \in \mathbf{R}$.
Now fix $j, k \in \{0, 1, \ldots, m\}$ and compute $\varphi_k\big((x-5)^j\big)$. Repeated differentiation of $(x-5)^j$ gives, for $k \leq j$,

$$\big((x-5)^j\big)^{(k)} = j(j-1)\cdots(j-k+1)\,(x-5)^{j-k},$$

and $\big((x-5)^j\big)^{(k)} = 0$ for $k > j$ (once the exponent reaches 0 the polynomial is constant, and one more derivative kills it).
If $k > j$, we get $\varphi_k\big((x-5)^j\big) = 0$.
If $k < j$, then $j - k \geq 1$, so the displayed derivative vanishes at $x = 5$; hence $\varphi_k\big((x-5)^j\big) = 0$.
If $k = j$, then the displayed derivative is the constant $k(k-1)\cdots 1 = k!$, so

$$\varphi_k\big((x-5)^k\big) = \frac{k!}{k!} = 1.$$

Thus $\varphi_k(v_j) = 1$ if $j = k$ and $\varphi_k(v_j) = 0$ if $j \neq k$, which is precisely the condition in 3.112 defining the dual basis of $v_0, \ldots, v_m$. Hence $\varphi_0, \ldots, \varphi_m$ is the dual basis of $1, x-5, \ldots, (x-5)^m$.
By 3.114 this gives, for every $p \in \mathcal{P}_m(\mathbf{R})$, the identity

$$p = \sum_{k=0}^{m} \frac{p^{(k)}(5)}{k!}(x-5)^k,$$

the Taylor expansion of p centered at 5.

Problem 3F.11. Suppose $v_1, \ldots, v_n$ is a basis of V and $\varphi_1, \ldots, \varphi_n$ is the corresponding dual basis of V'. Suppose $\psi \in V'$. Prove that

$$\psi = \psi(v_1)\varphi_1 + \cdots + \psi(v_n)\varphi_n.$$

Solution. Because V has a basis of length n, the space V is finite-dimensional with $\dim V = n$. Set

$$\eta = \psi(v_1)\varphi_1 + \cdots + \psi(v_n)\varphi_n.$$

Each φ_k belongs to V' and each $\psi(v_k)$ is a scalar, so $\eta \in V'$ because V' is a vector space (3.110) and hence closed under linear combinations. Thus ψ and η are both linear functionals on V, and to prove they are equal it suffices to show they agree on the basis $v_1, \ldots, v_n$: two linear maps on V that agree on a basis agree on $\operatorname{span}(v_1, \ldots, v_n) = V$, since every $v \in V$ can be written $v = c_1 v_1 + \cdots + c_n v_n$ and then linearity gives $\psi(v) = \sum_k c_k \psi(v_k) = \sum_k c_k \eta(v_k) = \eta(v)$.
So let $j \in \{1, \ldots, n\}$ and evaluate η at v_j. Using the definition of the dual basis (3.112), namely $\varphi_k(v_j) = 1$ if $k = j$ and $\varphi_k(v_j) = 0$ if $k \neq j$, we get

$$\eta(v_j) = \sum_{k=1}^{n} \psi(v_k)\varphi_k(v_j) = \psi(v_j),$$

since every term with $k \neq j$ vanishes and the term with $k = j$ equals $\psi(v_j) \cdot 1$.
Hence $\eta(v_j) = \psi(v_j)$ for each $j \in \{1, \ldots, n\}$, and by the paragraph above $\eta = \psi$. That is,

$$\psi = \psi(v_1)\varphi_1 + \cdots + \psi(v_n)\varphi_n,$$

as desired.
Two remarks. First, this result is the dual counterpart of 3.114: whereas 3.114 says the dual basis supplies the coefficients for expanding a vector in the basis $v_1, \ldots, v_n$, this exercise says the original basis supplies the coefficients for expanding a linear functional in the dual basis $\varphi_1, \ldots, \varphi_n$.
Second, one can also argue via 3.116: that result tells us $\varphi_1, \ldots, \varphi_n$ is a basis of V', so there exist unique scalars $a_1, \ldots, a_n \in \mathbf{F}$ with $\psi = a_1\varphi_1 + \cdots + a_n\varphi_n$. Applying both sides to v_j and using $\varphi_k(v_j) = 0$ for $k \neq j$ gives $\psi(v_j) = a_j$, which yields the same conclusion. The proof given first is slightly more economical because it does not need 3.116.

Problem 3F.12. Suppose $S, T \in \mathcal{L}(V, W)$.
(a) Prove that $(S + T)' = S' + T'$.
(b) Prove that $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbf{F}$.
This exercise asks you to verify (a) and (b) in 3.120.

Solution. Throughout, recall the definition of the dual map (3.118): if $R \in \mathcal{L}(V, W)$, then $R' \in \mathcal{L}(W', V')$ is defined by $R'(\varphi) = \varphi \circ R$ for each $\varphi \in W'$. Both sides of each identity to be proved are functions from W' to V', so they are equal precisely when they take the same value at every $\varphi \in W'$; and those values are linear functionals on V, so they in turn are equal precisely when they agree at every $v \in V$.
(a) Let $\varphi \in W'$ and let $v \in V$. Then

$$\begin{aligned}
\big((S+T)'(\varphi)\big)(v) &= \big(\varphi \circ (S+T)\big)(v) \\
&= \varphi\big((S+T)v\big) \\
&= \varphi(Sv + Tv) \\
&= \varphi(Sv) + \varphi(Tv) \\
&= (\varphi \circ S)(v) + (\varphi \circ T)(v) \\
&= \big(S'(\varphi)\big)(v) + \big(T'(\varphi)\big)(v) \\
&= \big(S'(\varphi) + T'(\varphi)\big)(v).
\end{aligned}$$

Here the first and sixth equalities use the definition of the dual map, the third uses the definition of the sum $S + T$ of linear maps (3.5), the fourth uses the additivity of the linear functional φ, and the last uses the definition of addition in V' (pointwise addition of functions).
Since this holds for every $v \in V$, we conclude $(S+T)'(\varphi) = S'(\varphi) + T'(\varphi)$. By the definition of addition in $\mathcal{L}(W', V')$, the right side equals $(S' + T')(\varphi)$. Since $\varphi \in W'$ was arbitrary, $(S + T)' = S' + T'$.

(b) Let $\lambda \in \mathbf{F}$, let $\varphi \in W'$, and let $v \in V$. Then

$$\begin{aligned}\big((\lambda T)'(\varphi)\big)(v) &= \big(\varphi \circ (\lambda T)\big)(v)\\ &= \varphi\big((\lambda T)v\big)\\ &= \varphi(\lambda\, Tv)\\ &= \lambda\, \varphi(Tv)\\ &= \lambda\, \big(T'(\varphi)\big)(v)\\ &= \big(\lambda\, T'(\varphi)\big)(v).\end{aligned}$$

The first and fifth equalities use the definition of the dual map, the third uses the definition of the scalar multiple λT of a linear map (3.5), the fourth uses the homogeneity of the linear functional φ, and the last uses the definition of scalar multiplication in V' (pointwise).
Since this holds for every $v \in V$, we get $(\lambda T)'(\varphi) = \lambda\, T'(\varphi)$, and the right side equals $(\lambda T')(\varphi)$ by the definition of scalar multiplication in $\mathcal{L}(W', V')$. Since $\varphi \in W'$ was arbitrary, $(\lambda T)' = \lambda T'$.
Together, (a) and (b) say that the map $T \mapsto T'$ from $\mathcal{L}(V, W)$ to $\mathcal{L}(W', V')$ is itself linear, as remarked in the text just before 3.120.

Problem 3F.13. Show that the dual map of the identity operator on V is the identity operator on V'.

Solution. Let I denote the identity operator on V, so $I \in \mathcal{L}(V, V)$ and $Iv = v$ for every $v \in V$. Its dual map I' is then an element of $\mathcal{L}(V', V')$, so it is an operator on V'; we must show it is the identity operator on V'.
Let $\varphi \in V'$. By the definition of the dual map (3.118),

$$I'(\varphi) = \varphi \circ I.$$

To see that $\varphi \circ I = \varphi$, let $v \in V$; then

$$(\varphi \circ I)(v) = \varphi(Iv) = \varphi(v).$$

Since $\varphi \circ I$ and φ are functions on V taking the same value at every $v \in V$, they are equal. Hence $I'(\varphi) = \varphi$.
Because $\varphi \in V'$ was arbitrary, I' sends every element of V' to itself. Thus I' is the identity operator on V', as desired.
Remark: combined with 3.120(c), which states $(ST)' = T'S'$, this gives a quick proof of one direction of Exercise 17 of this section. Suppose $T \in \mathcal{L}(V, W)$ is invertible, with inverse $T^{-1} \in \mathcal{L}(W, V)$. Taking duals in $T^{-1}T = I_V$ and using 3.120(c) gives $T'(T^{-1})' = (I_V)' = I_{V'}$; taking duals in $TT^{-1} = I_W$ gives $(T^{-1})'T' = (I_W)' = I_{W'}$. Since $(T^{-1})' \in \mathcal{L}(V', W')$ and $T' \in \mathcal{L}(W', V')$, these two equations say exactly that T' is invertible with $(T')^{-1} = (T^{-1})'$. (The converse direction of Exercise 17 needs a separate argument, for instance via 3.129 and 3.131.)

Problem 3F.14. Define $T\colon \mathbf{R}^3 \to \mathbf{R}^2$ by

$$T(x, y, z) = (4x + 5y + 6z,\ 7x + 8y + 9z).$$

Suppose φ_1, φ_2 denotes the dual basis of the standard basis of $\mathbf{R}^2$ and ψ_1, ψ_2, ψ_3 denotes the dual basis of the standard basis of $\mathbf{R}^3$.
(a) Describe the linear functionals $T'(\varphi_1)$ and $T'(\varphi_2)$.
(b) Write $T'(\varphi_1)$ and $T'(\varphi_2)$ as linear combinations of ψ_1, ψ_2, ψ_3.

Solution. By 3.113, the dual basis of the standard basis consists of the coordinate-selection functionals.

Thus

$$\varphi_1(a,b) = a, \qquad \varphi_2(a,b) = b \qquad \text{for } (a,b) \in \mathbf{R}^2,$$

and

$$\psi_1(x,y,z) = x, \qquad \psi_2(x,y,z) = y, \qquad \psi_3(x,y,z) = z \qquad \text{for } (x,y,z) \in \mathbf{R}^3.$$

Note that $T' \in \mathcal{L}\big((\mathbf{R}^2)', (\mathbf{R}^3)'\big)$, so $T'(\varphi_1)$ and $T'(\varphi_2)$ are linear functionals on $\mathbf{R}^3$.
(a) By the definition of the dual map (3.118), $T'(\varphi_j) = \varphi_j \circ T$. Hence for $(x,y,z) \in \mathbf{R}^3$,

$$\big(T'(\varphi_1)\big)(x,y,z) = \varphi_1\big(T(x,y,z)\big) = \varphi_1(4x+5y+6z,\ 7x+8y+9z) = 4x+5y+6z,$$

since φ_1 selects the first coordinate. Similarly, since φ_2 selects the second coordinate,

$$\big(T'(\varphi_2)\big)(x,y,z) = \varphi_2\big(T(x,y,z)\big) = 7x+8y+9z.$$

So $T'(\varphi_1)$ is the linear functional on $\mathbf{R}^3$ taking (x,y,z) to $4x+5y+6z$, and $T'(\varphi_2)$ is the linear functional on $\mathbf{R}^3$ taking (x,y,z) to $7x+8y+9z$. In words: $T'(\varphi_j)$ reads off the j^{th} coordinate of the output of T, which is the corresponding row of the formula defining T.
(b) Using the formulas for ψ_1, ψ_2, ψ_3 above, for every $(x,y,z) \in \mathbf{R}^3$ we have

$$(4\psi_1 + 5\psi_2 + 6\psi_3)(x,y,z) = 4x+5y+6z = \big(T'(\varphi_1)\big)(x,y,z),$$

and

$$(7\psi_1 + 8\psi_2 + 9\psi_3)(x,y,z) = 7x+8y+9z = \big(T'(\varphi_2)\big)(x,y,z).$$

Since these are equalities of functions at every point of $\mathbf{R}^3$,

$$T'(\varphi_1) = 4\psi_1 + 5\psi_2 + 6\psi_3 \qquad \text{and} \qquad T'(\varphi_2) = 7\psi_1 + 8\psi_2 + 9\psi_3.$$

(Alternatively, part (b) follows from Exercise 11 of this section: the coefficient of ψ_k in the expansion of a functional η on $\mathbf{R}^3$ is $\eta(e_k)$, where e_1, e_2, e_3 is the standard basis. Here $\big(T'(\varphi_1)\big)(e_1) = 4$, $\big(T'(\varphi_1)\big)(e_2) = 5$, $\big(T'(\varphi_1)\big)(e_3) = 6$, and similarly $7, 8, 9$ for $T'(\varphi_2)$.)
This illustrates 3.132. With respect to the standard bases,

$$\mathcal{M}(T) = \begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix},$$

so 3.132 predicts that the matrix of T' with respect to the dual bases is

$$\mathcal{M}(T') = \big(\mathcal{M}(T)\big)^{\mathrm{t}} = \begin{pmatrix} 4 & 7 \\ 5 & 8 \\ 6 & 9 \end{pmatrix}.$$

The columns of this matrix are the coefficient lists $(4,5,6)$ and $(7,8,9)$ found above, exactly as the computation in (b) shows.

Problem 3F.15. Define $T\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ by

$$(Tp)(x) = x^2 p(x) + p''(x)$$

for each $x \in \mathbb{R}$.
(a) Suppose $\varphi \in \mathcal{P}(\mathbb{R})'$ is defined by $\varphi(p) = p'(4)$. Describe the linear functional $T'(\varphi)$ on $\mathcal{P}(\mathbb{R})$.
(b) Suppose $\varphi \in \mathcal{P}(\mathbb{R})'$ is defined by $\varphi(p) = \int_0^1 p$. Evaluate $(T'(\varphi))(x^3)$.

Solution. Recall the definition of the dual map (3.118): for $T \in \mathcal{L}(V,W)$ and $\varphi \in W'$, the linear functional $T'(\varphi) \in V'$ is $\varphi \circ T$, so that $(T'(\varphi))(v) = \varphi(Tv)$ for every $v \in V$. Here $V = W = \mathcal{P}(\mathbb{R})$, so $T'(\varphi)$ is again a linear functional on $\mathcal{P}(\mathbb{R})$.

(a) Let $p \in \mathcal{P}(\mathbb{R})$. By the product rule,

$$(Tp)'(x) = \frac{d}{dx}\big(x^2 p(x) + p''(x)\big) = 2x\,p(x) + x^2 p'(x) + p'''(x).$$

Hence

$$(T'(\varphi))(p) = \varphi(Tp) = (Tp)'(4) = 8p(4) + 16\,p'(4) + p'''(4).$$

Thus $T'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbb{R})$ that takes p to

$$8p(4) + 16\,p'(4) + p'''(4).$$

(b) Here x^3 denotes the polynomial p with $p(x) = x^3$. Then $p''(x) = 6x$, so

$$(Tp)(x) = x^2 \cdot x^3 + 6x = x^5 + 6x.$$

Therefore

$$(T'(\varphi))(x^3) = \varphi(Tp) = \int_0^1 (x^5 + 6x)\,dx = \left[\frac{x^6}{6} + 3x^2\right]_0^1 = \frac{1}{6} + 3 = \frac{19}{6}.$$

Problem 3F.16. Suppose W is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that

$$T' = 0 \iff T = 0.$$

Solution. First suppose $T = 0$. If $\varphi \in W'$, then for every $v \in V$ we have

$$\big(T'(\varphi)\big)(v) = \varphi(Tv) = \varphi(0) = 0.$$

Thus $T'(\varphi) = 0$ for every $\varphi \in W'$, which means $T' = 0$. Note that this direction does not use the hypothesis that W is finite-dimensional.

Now suppose $T \neq 0$; we show $T' \neq 0$. Because $T \neq 0$, there exists $v \in V$ with $Tv \neq 0$. The list consisting of the single vector Tv is linearly independent in W, so because W is finite-dimensional this list can be extended to a basis

$$w_1 = Tv,\ w_2,\ \ldots,\ w_m$$

of W (by 2.32). Let $\psi_1, \ldots, \psi_m \in W'$ be the dual basis of $w_1, \ldots, w_m$ (see 3.112), so that $\psi_1(w_1) = 1$. Then

$$\big(T'(\psi_1)\big)(v) = \psi_1(Tv) = \psi_1(w_1) = 1 \neq 0.$$

Hence $T'(\psi_1) \neq 0$, and therefore $T' \neq 0$. Taking the contrapositive, $T' = 0$ implies $T = 0$.

Combining the two paragraphs gives $T' = 0 \iff T = 0$.

Alternatively, the second direction follows quickly from the results in this section. Suppose $T' = 0$. Then $\operatorname{null} T' = W'$. By 3.128(a) (whose proof, as noted in the book, does not use finite-dimensionality), $\operatorname{null} T' = (\operatorname{range} T)^0$. Thus $(\operatorname{range} T)^0 = W'$, and since W is finite-dimensional and $\operatorname{range} T$ is a subspace of W, part (b) of 3.127 gives $\operatorname{range} T = \{0\}$. Hence $Tv = 0$ for every $v \in V$, so $T = 0$.

Problem 3F.17. Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that T is invertible if and only if $T' \in \mathcal{L}(W', V')$ is invertible.

Solution. By 3.63, a linear map is invertible if and only if it is both injective and surjective. Because V and W are finite-dimensional, so are V' and W' (indeed $\dim V' = \dim V$ and $\dim W' = \dim W$, by 3.111), so 3.63 applies to T' as well. Now

$$\begin{aligned} T \text{ is invertible} &\iff T \text{ is injective and } T \text{ is surjective} \\ &\iff T' \text{ is surjective and } T' \text{ is injective} \\ &\iff T' \text{ is invertible,} \end{aligned}$$

where the first and third equivalences come from 3.63 and the second equivalence comes from 3.131 (which says T is injective $\iff$ T' is surjective) together with 3.129 (which says T is surjective $\iff$ T' is injective).
It is worth recording the inverse explicitly. Suppose T is invertible, and let $S = T^{-1} \in \mathcal{L}(W, V)$, so that $ST = I_V$ and $TS = I_W$. Applying 3.120(c) and Exercise 13 in this section (which says that the dual of an identity operator is the identity operator on the dual space), we get

$$T'S' = (ST)' = (I_V)' = I_{V'} \qquad \text{and} \qquad S'T' = (TS)' = (I_W)' = I_{W'}.$$

Thus T' is invertible with $(T')^{-1} = (T^{-1})'$. Note that this half of the argument uses neither 3.129, 3.131, nor finite-dimensionality.

Problem 3F.18. Suppose V and W are finite-dimensional. Prove that the map that takes $T \in \mathcal{L}(V, W)$ to $T' \in \mathcal{L}(W', V')$ is an isomorphism of $\mathcal{L}(V, W)$ onto $\mathcal{L}(W', V')$.

Solution. Define

$$\Lambda \colon \mathcal{L}(V, W) \to \mathcal{L}(W', V') \qquad \text{by} \qquad \Lambda T = T'.$$

This makes sense: if $T \in \mathcal{L}(V, W)$ then T' is indeed a linear map from W' to V' (see 3.118). We must show that Λ is linear, injective, and surjective; by definition 3.69 that makes Λ an isomorphism onto $\mathcal{L}(W', V')$.
Step 1: Λ is linear. If $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$, then 3.120(a) and 3.120(b) give

$$\Lambda(S + T) = (S + T)' = S' + T' = \Lambda S + \Lambda T, \qquad \Lambda(\lambda T) = (\lambda T)' = \lambda T' = \lambda\, \Lambda T.$$

Thus $\Lambda \in \mathcal{L}\big(\mathcal{L}(V, W), \mathcal{L}(W', V')\big)$.
Step 2: Λ is injective. Because Λ is linear, it suffices to show that $\operatorname{null} \Lambda = \{0\}$ (by 3.15). So suppose $T \in \mathcal{L}(V, W)$ and $\Lambda T = 0$, that is, $T' = 0$. Since W is finite-dimensional, Exercise 16 in this section gives $T = 0$. (Explicitly: if $T \neq 0$, pick $v \in V$ with $Tv \neq 0$, extend the linearly independent list Tv to a basis $Tv, w_2, \ldots, w_m$ of W using 2.32, and let $\psi_1, \ldots, \psi_m$ be the dual basis; then $(T'(\psi_1))(v) = \psi_1(Tv) = 1 \neq 0$, so $T' \neq 0$.) Hence $\operatorname{null} \Lambda = \{0\}$ and Λ is injective.
Step 3: the two spaces have the same dimension. By 3.72,

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Also V' and W' are finite-dimensional with $\dim V' = \dim V$ and $\dim W' = \dim W$, by 3.111. Hence, using 3.72 again,

$$\dim \mathcal{L}(W', V') = (\dim W')(\dim V') = (\dim W)(\dim V) = \dim \mathcal{L}(V, W).$$

Step 4: Λ is surjective. By Steps 2 and 3, Λ is an injective linear map between finite-dimensional vector spaces of equal dimension, so 3.65 implies that Λ is surjective (and indeed invertible).
Therefore Λ is an invertible linear map from $\mathcal{L}(V, W)$ onto $\mathcal{L}(W', V')$, i.e., an isomorphism, as desired. In particular $\mathcal{L}(V, W)$ and $\mathcal{L}(W', V')$ are isomorphic vector spaces.

Problem 3F.19. Suppose $U \subseteq V$. Explain why

$$U^0 = \{\varphi \in V' : U \subseteq \operatorname{null} \varphi\}.$$

Solution. Both sides are subsets of V', so it suffices to check that a linear functional $\varphi \in V'$ belongs to one side exactly when it belongs to the other.
By the definition of the annihilator (3.121),

$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \text{ for all } u \in U\},$$

and by the definition of null space (3.11),

$$\operatorname{null}\varphi = \{v \in V : \varphi(v) = 0\}.$$

Now fix $\varphi \in V'$. Then

$$\begin{aligned} \varphi \in U^0 &\iff \varphi(u) = 0 \text{ for every } u \in U \\ &\iff u \in \operatorname{null}\varphi \text{ for every } u \in U \\ &\iff U \subseteq \operatorname{null}\varphi, \end{aligned}$$

where the second equivalence is just the definition of $\operatorname{null}\varphi$ (each $u \in U$ is in particular an element of V, since $U \subseteq V$), and the third equivalence is the definition of set inclusion.
Because $\varphi \in V'$ was arbitrary, the two sets have exactly the same elements, so

$$U^0 = \{\varphi \in V' : U \subseteq \operatorname{null}\varphi\}.$$

Note that nothing here required U to be a subspace of V; the statement holds for an arbitrary subset U of V.

Problem 3F.20. Suppose V is finite-dimensional and U is a subspace of V. Show that

$$U = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in U^0\}.$$

Solution. Write

$$X = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in U^0\}.$$

We prove $U \subseteq X$ and $X \subseteq U$.
The inclusion $U \subseteq X$. Suppose $u \in U$. If $\varphi \in U^0$, then $\varphi(u) = 0$ by the definition of the annihilator (3.121). Since this holds for every $\varphi \in U^0$, we have $u \in X$. Hence $U \subseteq X$. (This direction does not use finite-dimensionality.)
The inclusion $X \subseteq U$. We prove the contrapositive statement: if $v \in V$ and $v \notin U$, then there exists $\varphi \in U^0$ with $\varphi(v) \neq 0$, so that $v \notin X$.
So suppose $v \in V$ with $v \notin U$. Because V is finite-dimensional, so is its subspace U (by 2.25); let $u_1, \ldots, u_m$ be a basis of U (take the empty list if $U = \{0\}$). We claim the list $u_1, \ldots, u_m, v$ is linearly independent. Indeed, suppose $a_1, \ldots, a_m, b \in \mathbb{F}$ satisfy

$$a_1u_1 + \cdots + a_mu_m + bv = 0.$$

If $b \neq 0$, then $v = -\frac{1}{b}(a_1u_1 + \cdots + a_mu_m) \in U$, contradicting $v \notin U$. Hence $b = 0$, and then $a_1 = \cdots = a_m = 0$ because $u_1, \ldots, u_m$ is linearly independent. This proves the claim.
Because V is finite-dimensional, the linearly independent list $u_1, \ldots, u_m, v$ extends to a basis

$$u_1, \ldots, u_m,\ v,\ w_1, \ldots, w_k$$

of V (by 2.32). Let

$$\varphi_1, \ldots, \varphi_m,\ \varphi,\ \psi_1, \ldots, \psi_k$$

be the corresponding dual basis of V' (3.112), where φ denotes the functional dual to v. By the defining property of a dual basis,

$$\varphi(u_j) = 0 \text{ for } j = 1, \ldots, m, \qquad \varphi(v) = 1.$$

Since φ vanishes on the basis $u_1, \ldots, u_m$ of U, linearity gives $\varphi(u) = 0$ for every $u \in U$; that is, $\varphi \in U^0$. But $\varphi(v) = 1 \neq 0$, so $v \notin X$. This completes the proof of the contrapositive, hence $X \subseteq U$.
Combining the two inclusions gives $U = X$, as desired.
Remark (a dimension-counting alternative for the second inclusion). The set X is a subspace of V: it is the intersection of the subspaces $\operatorname{null}\varphi$ over $\varphi \in U^0$, or directly, $0 \in X$ and X is closed under addition and scalar multiplication because each condition $\varphi(v) = 0$ is linear in v. Under the canonical identification of V with V'' that sends v to the functional $\varphi \mapsto \varphi(v)$ (this map is injective by the

construction above, applied with $U = \{0\}$, and $\dim V'' = \dim V' = \dim V$ by 3.111, so it is an isomorphism by 3.65), the subspace X corresponds exactly to $(U^0)^0$. Hence, by 3.125 applied twice,

$$\dim X = \dim V' - \dim U^0 = \dim V - (\dim V - \dim U) = \dim U.$$

Since $U \subseteq X$ and $\dim U = \dim X$, we get $U = X$ by 2.39.

Problem 3F.21. Suppose V is finite-dimensional and U and W are subspaces of V.
(a) Prove that $W^0 \subseteq U^0$ if and only if $U \subseteq W$.
(b) Prove that $W^0 = U^0$ if and only if $U = W$.

Solution. (a) First suppose $U \subseteq W$; we show $W^0 \subseteq U^0$. Let $\varphi \in W^0$. Then $\varphi \in V'$ and $\varphi(w) = 0$ for every $w \in W$, by 3.121. If $u \in U$, then $u \in W$ (because $U \subseteq W$), and hence $\varphi(u) = 0$. Thus φ vanishes on all of U, so $\varphi \in U^0$. This shows $W^0 \subseteq U^0$. Note that this implication holds for arbitrary subsets $U \subseteq W$ of V and does not use finite-dimensionality: annihilation reverses inclusions.
Conversely, suppose $W^0 \subseteq U^0$; we show $U \subseteq W$. Let $u \in U$. Let $\varphi \in W^0$ be arbitrary. Because $W^0 \subseteq U^0$, we have $\varphi \in U^0$, and hence $\varphi(u) = 0$ since $u \in U$. Thus

$$\varphi(u) = 0 \quad \text{for every } \varphi \in W^0.$$

Because V is finite-dimensional and W is a subspace of V, Exercise 20 in this section gives

$$W = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in W^0\}.$$

The displayed condition therefore says exactly that $u \in W$. Since $u \in U$ was arbitrary, $U \subseteq W$.
This completes the proof of (a): $W^0 \subseteq U^0 \iff U \subseteq W$.
(b) First suppose $U = W$. Then U^0 and W^0 are annihilators of the same subspace, so trivially $W^0 = U^0$.
Conversely, suppose $W^0 = U^0$. Then in particular $W^0 \subseteq U^0$, so part (a) gives $U \subseteq W$. Also $U^0 \subseteq W^0$, so part (a) with the roles of U and W interchanged gives $W \subseteq U$. Combining the two inclusions yields $U = W$.
This completes the proof of (b): $W^0 = U^0 \iff U = W$.
Remark. Part (b) says that on the collection of subspaces of a finite-dimensional vector space V the map $U \mapsto U^0$ is injective, and part (a) says that it is an inclusion-reversing correspondence. In fact $U \mapsto U^0$ is an inclusion-reversing bijection from the subspaces of V onto the subspaces of V'. Only surjectivity remains to be checked: given a subspace Y of V', put

$$U = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in Y\},$$

which is a subspace of V (each condition $\varphi(v) = 0$ is linear in v). If $\varphi \in Y$ and $u \in U$, then $\varphi(u) = 0$ by the definition of U, so $Y \subseteq U^0$. Under the canonical identification of V with V'' the subspace U is exactly Y^0, so 3.125 applied to the subspace Y of V' gives $\dim U = \dim V' - \dim Y = \dim V - \dim Y$, using 3.111. Hence 3.125 applied to U gives

$$\dim U^0 = \dim V - \dim U = \dim Y.$$

Since $Y \subseteq U^0$ and these subspaces of V' have the same dimension, 2.39 gives $Y = U^0$.

Problem 3F.22. Suppose V is finite-dimensional and U and W are subspaces of V.
(a) Show that $(U + W)^0 = U^0 \cap W^0$.
(b) Show that $(U \cap W)^0 = U^0 + W^0$.

Solution. Throughout, recall the definition 3.121 of the annihilator: for $X \subseteq V$,

$$X^0 = \{\varphi \in V' : \varphi(x) = 0 \text{ for every } x \in X\},$$

and recall 3.124, which says that X^0 is a subspace of V'.
Part (a). This part does not use the hypothesis that V is finite-dimensional.
First suppose $\varphi \in (U+W)^0$. Because $0 \in W$, every $u \in U$ can be written $u = u + 0 \in U + W$; thus $U \subseteq U + W$, and similarly $W \subseteq U + W$. Hence φ vanishes on U and on W, which means $\varphi \in U^0$ and $\varphi \in W^0$. Therefore $(U+W)^0 \subseteq U^0 \cap W^0$.
Conversely, suppose $\varphi \in U^0 \cap W^0$. A typical element of $U + W$ has the form $u + w$ with $u \in U$ and $w \in W$, and

$$\varphi(u+w) = \varphi(u) + \varphi(w) = 0 + 0 = 0.$$

Thus $\varphi \in (U+W)^0$, giving $U^0 \cap W^0 \subseteq (U+W)^0$.
The two inclusions give $(U+W)^0 = U^0 \cap W^0$.
Part (b). First we prove the inclusion $U^0 + W^0 \subseteq (U \cap W)^0$. If $\varphi \in U^0$, then φ vanishes on U and hence on the smaller set $U \cap W$; thus $U^0 \subseteq (U \cap W)^0$. Similarly $W^0 \subseteq (U \cap W)^0$. Because $(U \cap W)^0$ is a subspace of V' (by 3.124), it is closed under addition, and therefore

$$U^0 + W^0 \subseteq (U \cap W)^0.$$

Now we show that these two subspaces of V' have the same dimension. Let $n = \dim V$. Because V is finite-dimensional, so is V', with $\dim V' = \dim V = n$ (by 3.111); hence 2.43 may legitimately be applied to the subspaces U^0 and W^0 of V', and every dimension written below is finite. Using the formula for the dimension of a sum of two subspaces (2.43) applied inside V', then part (a), then the formula 3.125 for the dimension of an annihilator, we get

$$\begin{aligned}
\dim(U^0 + W^0) &= \dim U^0 + \dim W^0 - \dim(U^0 \cap W^0) \\
&= \dim U^0 + \dim W^0 - \dim\big((U+W)^0\big) \\
&= (n - \dim U) + (n - \dim W) - \big(n - \dim(U+W)\big) \\
&= n - \dim U - \dim W + \dim(U+W).
\end{aligned}$$

Applying 2.43 once more, this time to U and W inside V, we have $\dim(U+W) = \dim U + \dim W - \dim(U \cap W)$. Substituting this into the last line above gives

$$\dim(U^0 + W^0) = n - \dim(U \cap W) = \dim\big((U \cap W)^0\big),$$

where the last equality is 3.125 applied to the subspace $U \cap W$ of V.
Thus $U^0 + W^0$ is a subspace of the finite-dimensional vector space $(U \cap W)^0$ with $\dim(U^0 + W^0) = \dim(U \cap W)^0$. By 2.39, this forces

$$(U \cap W)^0 = U^0 + W^0,$$

as desired.

Problem 3F.23. Suppose V is finite-dimensional and $\varphi_1, \dots, \varphi_m \in V'$. Prove that the following three sets are equal to each other.
(a) $\operatorname{span}(\varphi_1, \dots, \varphi_m)$
(b) $\big((\operatorname{null}\varphi_1) \cap \cdots \cap (\operatorname{null}\varphi_m)\big)^0$
(c) $\{\varphi \in V' : (\operatorname{null}\varphi_1) \cap \cdots \cap (\operatorname{null}\varphi_m) \subseteq \operatorname{null}\varphi\}$

Solution. Write

$$W = (\operatorname{null}\varphi_1) \cap \cdots \cap (\operatorname{null}\varphi_m),$$

which is a subspace of V, being an intersection of subspaces.
Step 1: the sets in (b) and (c) are equal. For $\varphi \in V'$ we have

$$\varphi \in W^0 \iff \varphi(v) = 0 \text{ for every } v \in W \iff W \subseteq \operatorname{null}\varphi,$$

where the first equivalence is the definition 3.121 of the annihilator and the second is the definition of null space. Thus W^0 is exactly the set in (c). (This is Exercise 19 in this section.)

Step 2: for a single linear functional $\psi \in V'$ we have $(\operatorname{null}\psi)^0 = \operatorname{span}(\psi)$.
If $\psi = 0$, then $\operatorname{null}\psi = V$ and $V^0 = \{0\} = \operatorname{span}(0) = \operatorname{span}(\psi)$; here $V^0 = \{0\}$ by 3.127(a) (or directly: a linear functional vanishing on all of V is the zero functional).
Now suppose $\psi \neq 0$. Then $\operatorname{range}\psi$ is a nonzero subspace of the one-dimensional space $\mathbf{F}$, so $\operatorname{range}\psi = \mathbf{F}$ and $\dim\operatorname{range}\psi = 1$. The fundamental theorem of linear maps (3.21) gives

$$\dim\operatorname{null}\psi = \dim V - 1.$$

Hence by 3.125,

$$\dim(\operatorname{null}\psi)^0 = \dim V - \dim\operatorname{null}\psi = 1.$$

Also ψ itself vanishes on $\operatorname{null}\psi$, so $\psi \in (\operatorname{null}\psi)^0$, and $\psi \neq 0$. Thus $\operatorname{span}(\psi)$ is a one-dimensional subspace of the one-dimensional space $(\operatorname{null}\psi)^0$, and 2.39 gives $(\operatorname{null}\psi)^0 = \operatorname{span}(\psi)$.
Step 3: the annihilator of an intersection is the sum of the annihilators. We claim that if $U_1, \dots, U_m$ are subspaces of V, then

$$(U_1 \cap \cdots \cap U_m)^0 = U_1^0 + \cdots + U_m^0.$$

This holds trivially for $m = 1$. Suppose $m > 1$ and the claim holds for $m - 1$ subspaces. Since $U_1 \cap \cdots \cap U_{m-1}$ is a subspace of V, Exercise 22(b) applied to the two subspaces $U_1 \cap \cdots \cap U_{m-1}$ and U_m gives

$$(U_1 \cap \cdots \cap U_m)^0 = \big((U_1 \cap \cdots \cap U_{m-1}) \cap U_m\big)^0 = (U_1 \cap \cdots \cap U_{m-1})^0 + U_m^0,$$

and the induction hypothesis turns the first summand into $U_1^0 + \cdots + U_{m-1}^0$. This completes the induction.
Step 4: conclusion. Applying Step 3 with $U_k = \operatorname{null}\varphi_k$ and then Step 2 to each term, we get

$$W^0 = (\operatorname{null}\varphi_1)^0 + \cdots + (\operatorname{null}\varphi_m)^0 = \operatorname{span}(\varphi_1) + \cdots + \operatorname{span}(\varphi_m) = \operatorname{span}(\varphi_1, \dots, \varphi_m),$$

where the last equality holds because a sum of spans of single vectors consists precisely of all sums $a_1\varphi_1 + \cdots + a_m\varphi_m$ with $a_1, \dots, a_m \in \mathbf{F}$, which is the span of the list.
Thus the set in (a) equals the set in (b), and by Step 1 the sets in (b) and (c) are equal. Hence all three sets coincide.

Problem 3F.24. Suppose V is finite-dimensional and $v_1, \dots, v_m \in V$. Define a linear map $\Gamma : V' \to \mathbf{F}^m$ by $\Gamma(\varphi) = (\varphi(v_1), \dots, \varphi(v_m))$.
(a) Prove that $v_1, \dots, v_m$ spans V if and only if Γ is injective.
(b) Prove that $v_1, \dots, v_m$ is linearly independent if and only if Γ is surjective.

Solution. Let $U = \operatorname{span}(v_1, \dots, v_m)$, a subspace of V.
The key computation, used in both parts, is

$$\operatorname{null}\Gamma = U^0.$$

To see this, first suppose $\varphi \in \operatorname{null}\Gamma$. Then $(\varphi(v_1), \dots, \varphi(v_m)) = (0, \dots, 0)$, so $\varphi(v_k) = 0$ for each k. If $u \in U$, then $u = a_1v_1 + \cdots + a_mv_m$ for some $a_1, \dots, a_m \in \mathbf{F}$, and linearity of φ gives

$$\varphi(u) = a_1\varphi(v_1) + \cdots + a_m\varphi(v_m) = 0.$$

Hence $\varphi \in U^0$. Conversely, if $\varphi \in U^0$, then φ vanishes on U and in particular $\varphi(v_k) = 0$ for each k (because $v_k \in U$), so $\Gamma(\varphi) = 0$. This proves the displayed equality.
Part (a). Using 3.15 (a linear map is injective if and only if its null space is $\{0\}$), the computation above, and 3.127(a), we have

$$\Gamma \text{ is injective} \iff \operatorname{null}\Gamma = \{0\} \iff U^0 = \{0\} \iff U = V.$$

Since $U = V$ says exactly that $v_1, \dots, v_m$ spans V, part (a) follows.

Part (b). We compute $\dim \operatorname{range} \Gamma$. By the fundamental theorem of linear maps (3.21) applied to $\Gamma \in \mathcal{L}(V', \mathbf{F}^m)$, together with $\dim V' = \dim V$ (by 3.111), the equality $\operatorname{null} \Gamma = U^0$, and 3.125,

$$\begin{aligned} \dim \operatorname{range} \Gamma &= \dim V' - \dim \operatorname{null} \Gamma \\ &= \dim V - \dim U^0 \\ &= \dim V - (\dim V - \dim U) \\ &= \dim U. \end{aligned}$$

Because $\operatorname{range} \Gamma$ is a subspace of $\mathbf{F}^m$, which has dimension m, the map Γ is surjective if and only if $\dim \operatorname{range} \Gamma = m$ (by 2.39). By the computation above, this happens if and only if $\dim U = m$.
It remains to note that $\dim U = m$ if and only if $v_1, \ldots, v_m$ is linearly independent. Indeed, $v_1, \ldots, v_m$ is by construction a list of length m spanning U. If the list is linearly independent, then it is a basis of U, so $\dim U = m$. Conversely, if $\dim U = m$, then $v_1, \ldots, v_m$ is a spanning list of U whose length equals $\dim U$, so it is a basis of U by 2.42, and in particular it is linearly independent.
Combining the last two paragraphs: Γ is surjective if and only if $v_1, \ldots, v_m$ is linearly independent.

Problem 3F.25. Suppose V is finite-dimensional and $\varphi_1, \ldots, \varphi_m \in V'$. Define a linear map $\Gamma : V \to \mathbf{F}^m$ by $\Gamma(v) = (\varphi_1(v), \ldots, \varphi_m(v))$.
(a) Prove that $\varphi_1, \ldots, \varphi_m$ spans V' if and only if Γ is injective.
(b) Prove that $\varphi_1, \ldots, \varphi_m$ is linearly independent if and only if Γ is surjective.

Solution. Let

$$W = (\operatorname{null} \varphi_1) \cap \cdots \cap (\operatorname{null} \varphi_m),$$

a subspace of V.
The key observation, used in both parts, is

$$\operatorname{null} \Gamma = W.$$

Indeed, for $v \in V$ we have $\Gamma(v) = 0$ if and only if $\varphi_k(v) = 0$ for every $k \in \{1, \ldots, m\}$, which says exactly that $v \in \operatorname{null} \varphi_k$ for every k, that is, $v \in W$.
We also use Exercise 23 of this section, which shows that

$$W^0 = \operatorname{span}(\varphi_1, \ldots, \varphi_m).$$

Part (a). By 3.15, the equality $\operatorname{null} \Gamma = W$, and 3.127(b) applied to the subspace W of V, we have

$$\Gamma \text{ is injective} \iff \operatorname{null} \Gamma = \{0\} \iff W = \{0\} \iff W^0 = V'.$$

By the displayed consequence of Exercise 23, the condition $W^0 = V'$ says precisely that $\operatorname{span}(\varphi_1, \ldots, \varphi_m) = V'$, that is, that $\varphi_1, \ldots, \varphi_m$ spans V'. This proves (a).
Part (b). The fundamental theorem of linear maps (3.21) applied to Γ, together with $\operatorname{null} \Gamma = W$ and 3.125, gives

$$\dim \operatorname{range} \Gamma = \dim V - \dim \operatorname{null} \Gamma = \dim V - \dim W = \dim W^0 = \dim \operatorname{span}(\varphi_1, \ldots, \varphi_m),$$

where the last equality again uses Exercise 23.
Because $\operatorname{range} \Gamma$ is a subspace of the m-dimensional space $\mathbf{F}^m$, the map Γ is surjective if and only if $\dim \operatorname{range} \Gamma = m$ (by 2.39). By the display above, this happens if and only if $\dim \operatorname{span}(\varphi_1, \ldots, \varphi_m) = m$.
Finally, $\dim \operatorname{span}(\varphi_1, \ldots, \varphi_m) = m$ if and only if $\varphi_1, \ldots, \varphi_m$ is linearly independent. Indeed, the list $\varphi_1, \ldots, \varphi_m$ has length m and spans the subspace $\operatorname{span}(\varphi_1, \ldots, \varphi_m)$ of V'. If it is linearly independent, it is a basis of its span, so that span has dimension m. Conversely, if the span has dimension m, then $\varphi_1, \ldots, \varphi_m$ is a spanning list of that span with length equal to its dimension, hence a basis of it by 2.42, and in particular linearly independent.
Combining the last two paragraphs proves (b).

Problem 3F.26. Suppose V is finite-dimensional and Ω is a subspace of V'. Prove that

$$\Omega = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in \Omega\}^0.$$

Solution. Set

$$W = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in \Omega\}.$$

We must show that $W^0 = \Omega$. Note first that W is indeed a subspace of V: it is the intersection of the subspaces $\operatorname{null}\varphi$ over all $\varphi \in \Omega$ (and $W = V$ if $\Omega = \{0\}$, since then the only condition imposed is $0(v) = 0$).
Case 1: $\Omega = \{0\}$. Then every $v \in V$ satisfies the defining condition of W, so $W = V$. Hence $W^0 = V^0 = \{0\} = \Omega$, where $V^0 = \{0\}$ by 3.127(a).
Case 2: $\Omega \neq \{0\}$. Since V is finite-dimensional, so is V' (indeed $\dim V' = \dim V$ by 3.111), hence so is its subspace Ω (by 2.25 and 2.37). Let $\varphi_1, \ldots, \varphi_m$ be a basis of Ω, where $m = \dim\Omega \geq 1$.
We claim that

$$W = (\operatorname{null}\varphi_1) \cap \cdots \cap (\operatorname{null}\varphi_m).$$

For one inclusion: if $v \in W$, then $\varphi(v) = 0$ for every $\varphi \in \Omega$, and in particular $\varphi_k(v) = 0$ for each k, since each $\varphi_k \in \Omega$; thus v lies in the intersection. For the other inclusion: suppose $\varphi_k(v) = 0$ for every k, and let $\varphi \in \Omega$ be arbitrary. Because $\varphi_1, \ldots, \varphi_m$ is a basis of Ω, there exist $a_1, \ldots, a_m \in \mathbf{F}$ with $\varphi = a_1\varphi_1 + \cdots + a_m\varphi_m$, and therefore

$$\varphi(v) = a_1\varphi_1(v) + \cdots + a_m\varphi_m(v) = 0.$$

Hence $v \in W$, proving the claim.
Now Exercise 23 of this section says that

$$\big((\operatorname{null}\varphi_1) \cap \cdots \cap (\operatorname{null}\varphi_m)\big)^0 = \operatorname{span}(\varphi_1, \ldots, \varphi_m).$$

Combining this with the claim, and using that $\varphi_1, \ldots, \varphi_m$ is a basis of Ω (so its span is Ω), we conclude

$$W^0 = \operatorname{span}(\varphi_1, \ldots, \varphi_m) = \Omega,$$

as desired.

Problem 3F.27. Suppose $T \in \mathcal{L}(\mathcal{P}_5(\mathbf{R}))$ and $\operatorname{null} T' = \operatorname{span}(\varphi)$, where φ is the linear functional on $\mathcal{P}_5(\mathbf{R})$ defined by $\varphi(p) = p(8)$. Prove that

$$\operatorname{range} T = \{p \in \mathcal{P}_5(\mathbf{R}) : p(8) = 0\}.$$

Solution. Write $V = \mathcal{P}_5(\mathbf{R})$, so $\dim V = 6$ (the list $1, x, x^2, x^3, x^4, x^5$ is a basis of V). The map $\varphi : V \to \mathbf{R}$ given by $\varphi(p) = p(8)$ is a linear functional on V, and the set in question is exactly $\operatorname{null}\varphi$:

$$\{p \in \mathcal{P}_5(\mathbf{R}) : p(8) = 0\} = \operatorname{null}\varphi.$$

First observe that $\varphi \neq 0$, because $\varphi(1) = 1 \neq 0$ for the constant polynomial 1. Hence $\operatorname{range}\varphi$ is a nonzero subspace of the one-dimensional vector space $\mathbf{R}$, so $\operatorname{range}\varphi = \mathbf{R}$ and $\dim\operatorname{range}\varphi = 1$. The fundamental theorem of linear maps (3.21) now gives

$$\dim\operatorname{null}\varphi = \dim V - \dim\operatorname{range}\varphi = 6 - 1 = 5.$$

Also, since $\varphi \neq 0$, the list consisting of the single vector φ is linearly independent and hence a basis of $\operatorname{span}(\varphi)$, so $\dim\operatorname{span}(\varphi) = 1$.
Next we show that $\operatorname{range} T \subseteq \operatorname{null}\varphi$. By 3.128(a) applied to $T \in \mathcal{L}(V, V)$,

$$(\operatorname{range} T)^0 = \operatorname{null} T' = \operatorname{span}(\varphi).$$

In particular $\varphi \in (\operatorname{range} T)^0$, which by the definition 3.121 of the annihilator means $\varphi(w) = 0$ for every $w \in \operatorname{range} T$. Hence every $w \in \operatorname{range} T$ lies in $\operatorname{null} \varphi$, that is, $\operatorname{range} T \subseteq \operatorname{null} \varphi$.
Finally we compare dimensions. Since $\operatorname{range} T$ is a subspace of the finite-dimensional space V, 3.125 gives

$$\dim \operatorname{range} T = \dim V - \dim(\operatorname{range} T)^0 = 6 - \dim \operatorname{span}(\varphi) = 6 - 1 = 5.$$

Thus $\operatorname{range} T$ is a subspace of $\operatorname{null} \varphi$ and

$$\dim \operatorname{range} T = 5 = \dim \operatorname{null} \varphi.$$

By 2.39 (a subspace of full dimension equals the whole space, applied inside the vector space $\operatorname{null} \varphi$), we conclude

$$\operatorname{range} T = \operatorname{null} \varphi = \{p \in \mathcal{P}_5(\mathbf{R}) : p(8) = 0\},$$

as desired.

Problem 3F.28. Suppose V is finite-dimensional and $\varphi_1, \dots, \varphi_m$ is a linearly independent list in V'. Prove that

$$\dim\big((\operatorname{null} \varphi_1) \cap \cdots \cap (\operatorname{null} \varphi_m)\big) = (\dim V) - m.$$

Solution. Let

$$W = (\operatorname{null} \varphi_1) \cap \cdots \cap (\operatorname{null} \varphi_m),$$

a subspace of V, being an intersection of subspaces.
Define $\Gamma : V \to \mathbf{F}^m$ by

$$\Gamma(v) = (\varphi_1(v), \dots, \varphi_m(v)),$$

as in Exercise 25 of this section. This map is linear: for $u, v \in V$ and $\lambda \in \mathbf{F}$, linearity of each φ_k gives $\Gamma(u + v) = \Gamma(u) + \Gamma(v)$ and $\Gamma(\lambda v) = \lambda \Gamma(v)$.
As computed in the solution to Exercise 25,

$$\operatorname{null} \Gamma = W,$$

because $\Gamma(v) = 0$ holds if and only if $\varphi_k(v) = 0$ for every $k \in \{1, \dots, m\}$, which says exactly that v belongs to each $\operatorname{null} \varphi_k$.
Because $\varphi_1, \dots, \varphi_m$ is linearly independent, Exercise 25(b) shows that Γ is surjective, so

$$\operatorname{range} \Gamma = \mathbf{F}^m \quad \text{and} \quad \dim \operatorname{range} \Gamma = m.$$

Now the fundamental theorem of linear maps (3.21) applied to $\Gamma \in \mathcal{L}(V, \mathbf{F}^m)$ gives

$$\dim V = \dim \operatorname{null} \Gamma + \dim \operatorname{range} \Gamma = \dim W + m.$$

Therefore

$$\dim\big((\operatorname{null} \varphi_1) \cap \cdots \cap (\operatorname{null} \varphi_m)\big) = \dim W = (\dim V) - m,$$

as desired.
An alternative route avoids Exercise 25: by Exercise 23 of this section, $W^0 = \operatorname{span}(\varphi_1, \dots, \varphi_m)$, and this span has dimension m because $\varphi_1, \dots, \varphi_m$ is linearly independent and hence a basis of its span. Then 3.125 gives

$$\dim W = \dim V - \dim W^0 = (\dim V) - m.$$

Note in passing that the hypothesis of linear independence is essential: for instance, if $\varphi_1 = \varphi_2 \neq 0$, then $W = \operatorname{null} \varphi_1$ has dimension $(\dim V) - 1$, not $(\dim V) - 2$.

Problem 3F.29. Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$.
(a) Prove that if $\varphi \in W'$ and $\operatorname{null} T' = \operatorname{span}(\varphi)$, then $\operatorname{range} T = \operatorname{null} \varphi$.
(b) Prove that if $\psi \in V'$ and $\operatorname{range} T' = \operatorname{span}(\psi)$, then $\operatorname{null} T = \operatorname{null} \psi$.

Solution. We use repeatedly the following elementary observation: if X is a finite-dimensional vector space and $\alpha \in X'$ is nonzero, then $\operatorname{range} \alpha$ is a nonzero subspace of $\mathbf{F}$, hence equals $\mathbf{F}$, so the fundamental theorem of linear maps (3.21) gives

$$\dim \operatorname{null} \alpha = \dim X - \dim \operatorname{range} \alpha = \dim X - 1.$$

Part (a). By 3.128(a) we have $\operatorname{null} T' = (\operatorname{range} T)^0$, so the hypothesis says

$$(\operatorname{range} T)^0 = \operatorname{span}(\varphi).$$

First suppose $\varphi = 0$. Then $(\operatorname{range} T)^0 = \{0\}$, so $\operatorname{range} T = W$ by 3.127(a). Also $\operatorname{null} \varphi = \operatorname{null} 0 = W$. Hence $\operatorname{range} T = W = \operatorname{null} \varphi$.
Now suppose $\varphi \neq 0$. Because $\varphi \in \operatorname{span}(\varphi) = (\operatorname{range} T)^0$, the definition of the annihilator (3.121) gives $\varphi(w) = 0$ for every $w \in \operatorname{range} T$; in other words,

$$\operatorname{range} T \subseteq \operatorname{null} \varphi.$$

Because $\varphi \neq 0$, the list φ is linearly independent, so $\dim \operatorname{span}(\varphi) = 1$. Thus $\dim(\operatorname{range} T)^0 = 1$, and 3.125 (applied to the subspace $\operatorname{range} T$ of W) gives

$$1 = \dim(\operatorname{range} T)^0 = \dim W - \dim \operatorname{range} T,$$

so $\dim \operatorname{range} T = \dim W - 1$. By the observation at the start of this solution, $\dim \operatorname{null} \varphi = \dim W - 1$ as well. Hence $\operatorname{range} T$ is a subspace of $\operatorname{null} \varphi$ of the same (finite) dimension, so 2.39 gives $\operatorname{range} T = \operatorname{null} \varphi$.
Part (b). By 3.130(b) we have $\operatorname{range} T' = (\operatorname{null} T)^0$, so the hypothesis says

$$(\operatorname{null} T)^0 = \operatorname{span}(\psi).$$

First suppose $\psi = 0$. Then $(\operatorname{null} T)^0 = \{0\}$, so $\operatorname{null} T = V$ by 3.127(a). Also $\operatorname{null} \psi = \operatorname{null} 0 = V$. Hence $\operatorname{null} T = V = \operatorname{null} \psi$.
Now suppose $\psi \neq 0$. Because $\psi \in \operatorname{span}(\psi) = (\operatorname{null} T)^0$, we have $\psi(v) = 0$ for every $v \in \operatorname{null} T$; that is,

$$\operatorname{null} T \subseteq \operatorname{null} \psi.$$

As in (a), $\dim \operatorname{span}(\psi) = 1$, so 3.125 (applied to the subspace $\operatorname{null} T$ of V) gives

$$1 = \dim(\operatorname{null} T)^0 = \dim V - \dim \operatorname{null} T,$$

hence $\dim \operatorname{null} T = \dim V - 1$. By the observation at the start, $\dim \operatorname{null} \psi = \dim V - 1$ too. Thus $\operatorname{null} T$ is a subspace of $\operatorname{null} \psi$ of the same dimension, and 2.39 gives $\operatorname{null} T = \operatorname{null} \psi$.

Problem 3F.30. Suppose V is finite-dimensional and $\varphi_1, \ldots, \varphi_n$ is a basis of V'. Show that there exists a basis of V whose dual basis is $\varphi_1, \ldots, \varphi_n$.

Solution. By 3.111 we have $\dim V = \dim V' = n$.
We first record a standard fact: if $v \in V$ and $v \neq 0$, then there exists $\varphi \in V'$ with $\varphi(v) \neq 0$. Indeed, the list v is linearly independent, so it extends to a basis $v, u_2, \ldots, u_n$ of V (by 2.32), and the linear map lemma (3.4) produces a linear map $\varphi\colon V \to \mathbf{F}$ with $\varphi(v) = 1$ and $\varphi(u_k) = 0$ for $k = 2, \ldots, n$. Then $\varphi \in V'$ and $\varphi(v) = 1 \neq 0$.
Now define $\Gamma\colon V \to \mathbf{F}^n$ by

$$\Gamma(v) = \big(\varphi_1(v), \ldots, \varphi_n(v)\big).$$

Each coordinate of Γ is a linear functional, so Γ is linear.
Γ is injective. Suppose $v \in \operatorname{null}\Gamma$, so $\varphi_1(v) = \cdots = \varphi_n(v) = 0$. If $\varphi \in V'$ is arbitrary, then, because $\varphi_1, \ldots, \varphi_n$ spans V', there exist $a_1, \ldots, a_n \in \mathbf{F}$ with $\varphi = a_1\varphi_1 + \cdots + a_n\varphi_n$, and hence

$$\varphi(v) = a_1\varphi_1(v) + \cdots + a_n\varphi_n(v) = 0.$$

Thus $\varphi(v) = 0$ for every $\varphi \in V'$. By the standard fact above, this forces $v = 0$. Hence $\operatorname{null}\Gamma = \{0\}$ and Γ is injective by 3.15.
Because $\dim V = n = \dim \mathbf{F}^n$ and Γ is injective, 3.65 shows that Γ is also surjective, so Γ is invertible (3.63).
Let $e_1, \ldots, e_n$ denote the standard basis of $\mathbf{F}^n$ and set

$$v_j = \Gamma^{-1}(e_j) \qquad \text{for } j = 1, \ldots, n.$$

The list $v_1, \ldots, v_n$ is a basis of V. To see that it is linearly independent, suppose $a_1, \ldots, a_n \in \mathbf{F}$ satisfy $a_1v_1 + \cdots + a_nv_n = 0$. Applying Γ and using linearity gives $a_1e_1 + \cdots + a_ne_n = 0$, and thus $a_1 = \cdots = a_n = 0$ because $e_1, \ldots, e_n$ is linearly independent. So $v_1, \ldots, v_n$ is a linearly independent list in V of length $n = \dim V$, and therefore is a basis of V by 2.38.
Finally, $\Gamma(v_j) = e_j$ means precisely that

$$\big(\varphi_1(v_j), \ldots, \varphi_n(v_j)\big) = e_j,$$

that is, $\varphi_k(v_j) = 1$ if $k = j$ and $\varphi_k(v_j) = 0$ if $k \neq j$. By the definition of the dual basis (3.112), this says exactly that $\varphi_1, \ldots, \varphi_n$ is the dual basis of $v_1, \ldots, v_n$, as desired.

Problem 3F.31. Suppose U is a subspace of V. Let $i\colon U \to V$ be the inclusion map defined by $i(u) = u$. Thus $i' \in \mathcal{L}(V', U')$.
(a) Show that $\operatorname{null} i' = U^0$.
(b) Prove that if V is finite-dimensional, then $\operatorname{range} i' = U'$.
(c) Prove that if V is finite-dimensional, then $\widetilde{i'}$ is an isomorphism from V'/U^0 onto U'.
[The isomorphism in (c) is natural in that it does not depend on a choice of basis in either vector space.]

Solution. Throughout, note that for $\varphi \in V'$ the definition of the dual map (3.118) gives $i'(\varphi) = \varphi \circ i$, and for $u \in U$ we have $(\varphi \circ i)(u) = \varphi(i(u)) = \varphi(u)$. Thus

$$i'(\varphi) = \varphi|_U,$$

the restriction of φ to U.
Part (a). Let $\varphi \in V'$. Then $\varphi \in \operatorname{null} i'$ if and only if $i'(\varphi)$ is the zero linear functional on U, which by the displayed formula holds if and only if $\varphi(u) = 0$ for every $u \in U$. By the definition of the annihilator (3.121), that last condition says exactly $\varphi \in U^0$. Hence $\operatorname{null} i' = U^0$. (No finite-dimensionality is used here.)
Part (b). Certainly $\operatorname{range} i' \subseteq U'$, since $i' \in \mathcal{L}(V', U')$. For the reverse inclusion, suppose $\varphi \in U'$. Because V is finite-dimensional, so is its subspace U (by 2.25); let $u_1, \ldots, u_m$ be a basis of U and extend it to a basis $u_1, \ldots, u_m, u_{m+1}, \ldots, u_n$ of V (by 2.32). The linear map lemma (3.4) gives a linear map $\psi\colon V \to \mathbf{F}$ with

$$\psi(u_j) = \begin{cases} \varphi(u_j) & \text{if } 1 \le j \le m, \\ 0 & \text{if } m < j \le n. \end{cases}$$

Then $\psi \in V'$, and the linear functionals $\psi|_U$ and φ on U agree on the basis $u_1, \ldots, u_m$ of U, hence agree on all of U. Thus $i'(\psi) = \psi|_U = \varphi$, so $\varphi \in \operatorname{range} i'$. Therefore $\operatorname{range} i' = U'$.
Part (c). By the notation of 3.106 applied to the linear map $i' \in \mathcal{L}(V', U')$, the map $\widetilde{i'}$ is the linear map

$$\widetilde{i'}\colon V'/(\operatorname{null} i') \to U', \qquad \widetilde{i'}\big(\varphi + \operatorname{null} i'\big) = i'(\varphi).$$

By part (a), $\operatorname{null} i' = U^0$, so the domain of $\widetilde{i'}$ is V'/U^0.
By 3.107(b), $\widetilde{i'}$ is injective. By 3.107(c), $\operatorname{range} \widetilde{i'} = \operatorname{range} i'$, which equals U' by part (b); hence $\widetilde{i'}$ is surjective. A linear map that is both injective and surjective is invertible (3.63), so $\widetilde{i'}$ is an isomorphism from V'/U^0 onto U'.

Problem 3F.32. The double dual space of V, denoted by V'', is defined to be the dual space of V'. In other words, $V'' = (V')'$. Define $\Lambda \colon V \to V''$ by

$$(\Lambda v)(\varphi) = \varphi(v)$$

for each $v \in V$ and each $\varphi \in V'$.
(a) Show that Λ is a linear map from V to V''.
(b) Show that if $T \in \mathcal{L}(V)$, then $T'' \circ \Lambda = \Lambda \circ T$, where $T'' = (T')'$.
(c) Show that if V is finite-dimensional, then Λ is an isomorphism from V onto V''.
[Suppose V is finite-dimensional. Then V and V' are isomorphic, but finding an isomorphism from V onto V' generally requires choosing a basis of V. In contrast, the isomorphism Λ from V onto V'' does not require a choice of basis and thus is considered more natural.]

Solution. Part (a). First we check that Λv really is an element of $V'' = (V')'$, that is, that Λv is a linear functional on V'. If $\varphi, \psi \in V'$, then

$$(\Lambda v)(\varphi + \psi) = (\varphi + \psi)(v) = \varphi(v) + \psi(v) = (\Lambda v)(\varphi) + (\Lambda v)(\psi),$$

where the middle equality is the definition of addition in $V' = \mathcal{L}(V, \mathbf{F})$. If $\lambda \in \mathbf{F}$ and $\varphi \in V'$, then

$$(\Lambda v)(\lambda\varphi) = (\lambda\varphi)(v) = \lambda\big(\varphi(v)\big) = \lambda(\Lambda v)(\varphi),$$

using the definition of scalar multiplication in V'. Thus $\Lambda v \colon V' \to \mathbf{F}$ is linear, so $\Lambda v \in V''$.
Now we check that Λ is linear. Let $u, v \in V$. For every $\varphi \in V'$,

$$\big(\Lambda(u + v)\big)(\varphi) = \varphi(u + v) = \varphi(u) + \varphi(v) = (\Lambda u)(\varphi) + (\Lambda v)(\varphi) = (\Lambda u + \Lambda v)(\varphi).$$

Because two elements of V'' that agree at every $\varphi \in V'$ are equal, $\Lambda(u + v) = \Lambda u + \Lambda v$. Similarly, if $\lambda \in \mathbf{F}$ and $v \in V$, then for every $\varphi \in V'$,

$$\big(\Lambda(\lambda v)\big)(\varphi) = \varphi(\lambda v) = \lambda\varphi(v) = \lambda(\Lambda v)(\varphi) = \big(\lambda(\Lambda v)\big)(\varphi),$$

so $\Lambda(\lambda v) = \lambda \Lambda v$. Hence $\Lambda \in \mathcal{L}(V, V'')$.
Part (b). Suppose $T \in \mathcal{L}(V)$. Then $T' \in \mathcal{L}(V', V')$ and $T'' = (T')' \in \mathcal{L}(V'', V'')$. Both $T'' \circ \Lambda$ and $\Lambda \circ T$ are maps from V to V'', so it suffices to show they agree at each $v \in V$, and for that it suffices to show that the two elements of V'' agree at each $\varphi \in V'$.
Fix $v \in V$ and $\varphi \in V'$. Using the definition of the dual map (3.118) twice,

$$\begin{aligned}
\big((T'' \circ \Lambda)(v)\big)(\varphi) &= \big(T''(\Lambda v)\big)(\varphi) \\
&= \big((\Lambda v) \circ T'\big)(\varphi) \\
&= (\Lambda v)\big(T'(\varphi)\big) \\
&= (\Lambda v)(\varphi \circ T) \\
&= (\varphi \circ T)(v) \\
&= \varphi(Tv) \\
&= \big(\Lambda(Tv)\big)(\varphi) \\
&= \big((\Lambda \circ T)(v)\big)(\varphi),
\end{aligned}$$

where the second equality uses $T''(\alpha) = \alpha \circ T'$ for $\alpha \in V''$, the fourth uses $T'(\varphi) = \varphi \circ T$, and the fifth and seventh use the definition of Λ. Since $\varphi \in V'$ was arbitrary, $(T'' \circ \Lambda)(v) = (\Lambda \circ T)(v)$; since $v \in V$ was arbitrary, $T'' \circ \Lambda = \Lambda \circ T$.

Part (c). Suppose V is finite-dimensional. Applying 3.111 twice gives

$$\dim V'' = \dim V' = \dim V,$$

so V and V'' are finite-dimensional of the same dimension.
We show Λ is injective. Suppose $v \in \operatorname{null}\Lambda$, so $\Lambda v = 0$, which means $\varphi(v) = 0$ for every $\varphi \in V'$. If $v \neq 0$, then the list v is linearly independent and so extends to a basis $v, u_2, \ldots, u_n$ of V (by 2.32); the linear map lemma (3.4) then provides $\varphi \in V'$ with $\varphi(v) = 1$ and $\varphi(u_k) = 0$ for $k \geq 2$, contradicting $\varphi(v) = 0$. Hence $v = 0$, so $\operatorname{null}\Lambda = \{0\}$ and Λ is injective by 3.15.
Because $\Lambda \in \mathcal{L}(V, V'')$ is injective and $\dim V = \dim V'' < \infty$, 3.65 shows that Λ is surjective as well. An injective and surjective linear map is invertible (3.63), so Λ is an isomorphism from V onto V''.

Problem 3F.33. Suppose U is a subspace of V. Let $\pi\colon V \to V/U$ be the usual quotient map. Thus $\pi' \in \mathcal{L}\big((V/U)', V'\big)$.
(a) Show that π' is injective.
(b) Show that $\operatorname{range}\pi' = U^0$.
(c) Conclude that π' is an isomorphism from $(V/U)'$ onto U^0.
[The isomorphism in (c) is natural in that it does not depend on a choice of basis in either vector space. In fact, there is no assumption here that any of these vector spaces are finite-dimensional.]

Solution. Recall that the quotient map is defined by $\pi(v) = v + U$ for $v \in V$, and that for $\varphi \in (V/U)'$ the dual map is given by $\pi'(\varphi) = \varphi \circ \pi$ (3.118), so

$$\big(\pi'(\varphi)\big)(v) = \varphi(v + U) \qquad \text{for all } v \in V.$$

No finite-dimensionality is assumed anywhere below.
Part (a). Suppose $\varphi \in \operatorname{null}\pi'$, so $\varphi \circ \pi = 0$, that is, $\varphi(v + U) = 0$ for every $v \in V$. Every element of V/U has the form $v + U$ for some $v \in V$ (equivalently, π is surjective), so φ vanishes on all of V/U; hence $\varphi = 0$. Thus $\operatorname{null}\pi' = \{0\}$, and π' is injective by 3.15.
Part (b). First we show $\operatorname{range}\pi' \subseteq U^0$. Suppose $\varphi \in (V/U)'$ and $u \in U$. Since $u \in U$, we have $u + U = 0 + U$ by 3.101, and therefore

$$\big(\pi'(\varphi)\big)(u) = \varphi(u + U) = \varphi(0 + U) = 0,$$

the last equality because $0 + U$ is the additive identity of V/U and φ is linear (3.10). As $u \in U$ was arbitrary, $\pi'(\varphi) \in U^0$ by the definition of the annihilator (3.121).
Now we show $U^0 \subseteq \operatorname{range}\pi'$. Suppose $\psi \in U^0$; thus $\psi \in V'$ and $\psi(u) = 0$ for all $u \in U$. Define

$$\varphi\colon V/U \to \mathbf{F} \qquad \text{by} \qquad \varphi(v + U) = \psi(v).$$

This is well defined: if $v + U = w + U$, then $v - w \in U$ (by 3.101), so $\psi(v) - \psi(w) = \psi(v - w) = 0$, giving $\psi(v) = \psi(w)$. It is linear: for $v, w \in V$ and $\lambda \in \mathbf{F}$,

$$\begin{aligned}\varphi\big((v + U) + (w + U)\big) &= \varphi\big((v + w) + U\big) = \psi(v + w) = \psi(v) + \psi(w) = \varphi(v + U) + \varphi(w + U),\\ \varphi\big(\lambda(v + U)\big) &= \varphi(\lambda v + U) = \psi(\lambda v) = \lambda\psi(v) = \lambda\varphi(v + U),\end{aligned}$$

using the definitions of addition and scalar multiplication on V/U. Hence $\varphi \in (V/U)'$. Finally, for every $v \in V$,

$$\big(\pi'(\varphi)\big)(v) = \varphi(v + U) = \psi(v),$$

so $\pi'(\varphi) = \psi$ and therefore $\psi \in \operatorname{range}\pi'$.
Combining the two inclusions gives $\operatorname{range}\pi' = U^0$.
Part (c). By 3.124, U^0 is a subspace of V', so we may regard π' as a linear map from $(V/U)'$ into U^0; call this map S, so $S(\varphi) = \pi'(\varphi)$ for $\varphi \in (V/U)'$, which is legitimate by part (b). Then S is injective, since $\operatorname{null} S = \operatorname{null}\pi' = \{0\}$ by part (a), and S is surjective onto U^0, since $\operatorname{range} S = \operatorname{range}\pi' = U^0$ by part (b). A linear map that is injective and surjective is invertible (3.63). Hence π' is an isomorphism from $(V/U)'$ onto U^0.

4 Polynomials

Contents

4.1 Exercises 180

4.1 Exercises

Problem 4.1. Suppose $w, z \in \mathbb{C}$. Verify the following equalities and inequalities.
(a) $z + \bar{z} = 2 \operatorname{Re} z$
(b) $z - \bar{z} = 2(\operatorname{Im} z)i$
(c) $z\bar{z} = |z|^2$
(d) $\overline{w+z} = \bar{w} + \bar{z}$ and $\overline{wz} = \bar{w}\,\bar{z}$
(e) $\bar{\bar{z}} = z$
(f) $|\operatorname{Re} z| \le |z|$ and $|\operatorname{Im} z| \le |z|$
(g) $|\bar{z}| = |z|$
(h) $|wz| = |w|\,|z|$
The results above are the parts of 4.4 that were left to the reader.

Solution. Throughout, write

$$z = a + bi, \qquad w = c + di,$$

where $a = \operatorname{Re} z$, $b = \operatorname{Im} z$, $c = \operatorname{Re} w$, $d = \operatorname{Im} w$ are real numbers. By the definitions in 4.2 we then have $\bar{z} = a - bi$, $\bar{w} = c - di$, and

$$|z| = \sqrt{a^2 + b^2}, \qquad |w| = \sqrt{c^2 + d^2}.$$

(a) $z + \bar{z} = (a + bi) + (a - bi) = 2a = 2 \operatorname{Re} z$.
(b) $z - \bar{z} = (a + bi) - (a - bi) = 2bi = 2(\operatorname{Im} z)i$.
(c) Using $i^2 = -1$,

$$z\bar{z} = (a + bi)(a - bi) = a^2 - abi + abi - b^2 i^2 = a^2 + b^2 = |z|^2.$$

(d) First, $w + z = (c + a) + (d + b)i$, and $c + a$ and $d + b$ are real, so

$$\overline{w+z} = (c + a) - (d + b)i = (c - di) + (a - bi) = \bar{w} + \bar{z}.$$

Second, $wz = (ca - db) + (cb + da)i$, and $ca - db$ and $cb + da$ are real, so

$$\overline{wz} = (ca - db) - (cb + da)i.$$

On the other hand,

$$\bar{w}\,\bar{z} = (c - di)(a - bi) = ca - cbi - dai + db\,i^2 = (ca - db) - (cb + da)i.$$

The two right sides agree, so $\overline{wz} = \bar{w}\,\bar{z}$.
(e) Because $\bar{z} = a + (-b)i$ with a and $-b$ real, we have $\operatorname{Re} \bar{z} = a$ and $\operatorname{Im} \bar{z} = -b$. Hence

$$\bar{\bar{z}} = a - (-b)i = a + bi = z.$$

(f) Because $b^2 \ge 0$ we have $a^2 \le a^2 + b^2$, and the square root function is increasing on $[0, \infty)$, so

$$|\operatorname{Re} z| = |a| = \sqrt{a^2} \le \sqrt{a^2 + b^2} = |z|.$$

Similarly $a^2 \ge 0$ gives $b^2 \le a^2 + b^2$ and therefore

$$|\operatorname{Im} z| = |b| = \sqrt{b^2} \le \sqrt{a^2 + b^2} = |z|.$$

(g) As noted in (e), $\operatorname{Re} \bar{z} = a$ and $\operatorname{Im} \bar{z} = -b$. Hence

$$|\bar{z}| = \sqrt{a^2 + (-b)^2} = \sqrt{a^2 + b^2} = |z|.$$

(h) Using (c) three times and (d) once, along with commutativity of complex multiplication,

$$|wz|^2 = (wz)\overline{(wz)} = wz\,\bar{w}\,\bar{z} = (w\bar{w})(z\bar{z}) = |w|^2|z|^2 = \big(|w|\,|z|\big)^2.$$

Both $|wz|$ and $|w|\,|z|$ are nonnegative real numbers with equal squares, so taking square roots gives $|wz| = |w|\,|z|$.

Problem 4.2. Prove that if $w, z \in \mathbb{C}$, then $\big|\,|w| - |z|\,\big| \le |w - z|$.
The inequality above is called the reverse triangle inequality.

Solution. First note that if $u \in \mathbb{C}$, then by the multiplicativity of absolute value (4.4, proved in Exercise 4.1(h)),

$$|-u| = |(-1)u| = |-1|\,|u| = |u|,$$

because $|-1| = \sqrt{(-1)^2 + 0^2} = 1$. In particular $|z - w| = |-(w - z)| = |w - z|$.
Now apply the triangle inequality (4.4) to the decomposition $w = (w - z) + z$:

$$|w| = \big|(w - z) + z\big| \le |w - z| + |z|,$$

and hence

$$|w| - |z| \le |w - z|.$$

Interchanging the roles of w and z in the display above gives

$$|z| - |w| \le |z - w| = |w - z|.$$

The real number $\big|\,|w| - |z|\,\big|$ equals either $|w| - |z|$ or $|z| - |w|$, and both of those quantities have just been shown to be at most $|w - z|$. Therefore

$$\big|\,|w| - |z|\,\big| \le |w - z|,$$

as desired.

Problem 4.3. Suppose V is a complex vector space and $\varphi \in V'$. Define $\sigma\colon V \to \mathbb{R}$ by $\sigma(v) = \operatorname{Re}\varphi(v)$ for each $v \in V$. Show that

$$\varphi(v) = \sigma(v) - i\sigma(iv)$$

for all $v \in V$.

Solution. Recall that V' denotes the dual space of V, so φ is a linear functional from V to $\mathbb{C}$; in particular $\varphi(\lambda v) = \lambda\varphi(v)$ for all $\lambda \in \mathbb{C}$ and all $v \in V$, because V is a complex vector space.
Fix $v \in V$ and write

$$\varphi(v) = \alpha + \beta i,$$

where $\alpha = \operatorname{Re}\varphi(v)$ and $\beta = \operatorname{Im}\varphi(v)$ are real numbers. Then $\sigma(v) = \alpha$.
By the homogeneity of φ with respect to the complex scalar i,

$$\varphi(iv) = i\varphi(v) = i(\alpha + \beta i) = -\beta + \alpha i.$$

Because $-\beta$ and α are real, the last expression exhibits the real and imaginary parts of $\varphi(iv)$; thus

$$\sigma(iv) = \operatorname{Re}\varphi(iv) = -\beta.$$

Therefore

$$\sigma(v) - i\sigma(iv) = \alpha - i(-\beta) = \alpha + \beta i = \varphi(v).$$

Because $v \in V$ was arbitrary, the desired equation holds for all $v \in V$.
Note that the point of this exercise is that a complex linear functional is completely determined by its real part: if $\varphi_1, \varphi_2 \in V'$ satisfy $\operatorname{Re}\varphi_1(v) = \operatorname{Re}\varphi_2(v)$ for all $v \in V$, then the formula just proved forces $\varphi_1 = \varphi_2$.

Problem 4.4. Suppose m is a positive integer. Is the set

$$\{0\} \cup \{p \in \mathcal{P}(\mathbb{F}) : \deg p = m\}$$

a subspace of $\mathcal{P}(\mathbb{F})$?

Solution. No. Call the set above U. Then U is not closed under addition, so U is not a subspace of $\mathcal{P}(\mathbb{F})$.
To see this, define $p, q \in \mathcal{P}(\mathbb{F})$ by

$$p(z) = z^m \qquad \text{and} \qquad q(z) = 1 - z^m$$

for $z \in \mathbb{F}$. The coefficient of z^m in p is $1 \neq 0$ and the coefficient of z^m in q is $-1 \neq 0$, and no higher powers appear in either polynomial; hence $\deg p = \deg q = m$ and so $p, q \in U$. However

$$(p + q)(z) = z^m + (1 - z^m) = 1$$

for all $z \in \mathbb{F}$. Thus $p + q$ is the constant polynomial 1, which is not the zero polynomial and has degree 0. Because m is a positive integer, $0 \neq m$, and therefore $p + q \notin U$.
Two remarks. First, the other two requirements for a subspace do hold: $0 \in U$ by the definition of U, and U is closed under scalar multiplication, because if $\lambda \in \mathbb{F}$ and $p \in U$, then either $\lambda = 0$ or $p = 0$ (in which case $\lambda p = 0 \in U$), or else $\lambda \neq 0$ and $\deg p = m$, in which case multiplying each coefficient of p by the nonzero scalar λ leaves the coefficient of z^m nonzero, so $\deg(\lambda p) = m$ and $\lambda p \in U$. Only additive closure fails.
Second, the hypothesis that m is positive is needed for the counterexample as phrased, but in fact it is needed for the conclusion as well: if $m = 0$, then $\{0\} \cup \{p : \deg p = 0\}$ is the set of constant polynomials, which is a subspace of $\mathcal{P}(\mathbb{F})$.

Problem 4.5. Is the set

$$\{0\} \cup \{p \in \mathcal{P}(\mathbb{F}) : \deg p \text{ is even}\}$$

a subspace of $\mathcal{P}(\mathbb{F})$?

Solution. No. Call the set above U. As in Exercise 4.4, the failure is closure under addition.
Define $p, q \in \mathcal{P}(\mathbb{F})$ by

$$p(z) = z^2 + z \qquad \text{and} \qquad q(z) = -z^2$$

for $z \in \mathbb{F}$. Both p and q have degree 2, which is even, so $p, q \in U$. But

$$(p + q)(z) = (z^2 + z) + (-z^2) = z$$

for all $z \in \mathbb{F}$. Thus $p + q$ is the polynomial $z \mapsto z$, which is not the zero polynomial and has degree 1, an odd number. Hence $p + q \notin U$, and so U is not a subspace of $\mathcal{P}(\mathbb{F})$.
As in Exercise 4.4, note that the other two subspace requirements are satisfied: $0 \in U$ by definition, and U is closed under scalar multiplication, because $\lambda p = 0 \in U$ when $\lambda = 0$, while for $\lambda \neq 0$ and $p \neq 0$ we have $\deg(\lambda p) = \deg p$, which is even when $p \in U$.
The point worth noticing is that cancellation of leading terms can change the degree by an odd amount: adding two polynomials of even degree can produce a polynomial of odd degree. Choosing the two polynomials to have leading coefficients that cancel while the degree-one terms do not, as above, exhibits exactly that.

Problem 4.6. Suppose that m and n are positive integers with $m \leq n$, and suppose $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$. Prove that there exists a polynomial $p \in \mathcal{P}(\mathbb{F})$ with $\deg p = n$ such that $0 = p(\lambda_1) = \cdots = p(\lambda_m)$ and such that p has no other zeros.

Solution. Define $p \in \mathcal{P}(\mathbb{F})$ by

$$p(z) = (z - \lambda_1)^{n-m+1}(z - \lambda_2)(z - \lambda_3) \cdots (z - \lambda_m)$$

for $z \in \mathbb{F}$; if $m = 1$ this means simply $p(z) = (z - \lambda_1)^n$. Note that $n - m + 1 \geq 1$ because $m \leq n$, so

the exponent is a positive integer and p is indeed a product of polynomials.
We verify the three required properties.
First, $\deg p = n$. Each factor $z - \lambda_k$ is a monic polynomial of degree 1, meaning its coefficient of z^1 equals 1. A product of monic polynomials is monic of degree equal to the sum of the degrees: indeed, if r and s are monic of degrees j and k, then multiplying out rs produces the term z^{j+k} with coefficient 1 and no higher power of z, so rs is monic of degree $j + k$. Applying this repeatedly to the $(n - m + 1) + (m - 1) = n$ degree-one factors in the definition of p shows that p is monic of degree

$$(n - m + 1) + \underbrace{1 + \cdots + 1}_{m-1 \text{ times}} = (n - m + 1) + (m - 1) = n.$$

In particular $p \neq 0$.
Second, $p(\lambda_k) = 0$ for each $k \in \{1, \ldots, m\}$. If $k = 1$, then the factor $(\lambda_1 - \lambda_1)^{n-m+1} = 0$ makes the whole product 0. If $k \geq 2$, then the factor $\lambda_k - \lambda_k = 0$ occurs in the product, again making the product 0.
Third, p has no zeros other than $\lambda_1, \ldots, \lambda_m$. Suppose $\lambda \in \mathbb{F}$ and $p(\lambda) = 0$, so

$$(\lambda - \lambda_1)^{n-m+1}(\lambda - \lambda_2) \cdots (\lambda - \lambda_m) = 0.$$

Because $\mathbb{F}$ (which is $\mathbb{R}$ or $\mathbb{C}$) has no zero divisors, a product of elements of $\mathbb{F}$ equals 0 only if one of the factors equals 0. Hence $\lambda - \lambda_k = 0$ for some $k \in \{1, \ldots, m\}$, that is, $\lambda \in \{\lambda_1, \ldots, \lambda_m\}$.
Thus p has degree n and its set of zeros is exactly $\{\lambda_1, \ldots, \lambda_m\}$, as required.
Two comments. The numbers $\lambda_1, \ldots, \lambda_m$ are not assumed to be distinct, and nothing above used distinctness; if some of them coincide, the set of zeros is just the (possibly smaller) set $\{\lambda_1, \ldots, \lambda_m\}$, and the conclusion reads the same. Also, the trick of piling the extra $n - m$ degrees onto the single factor $z - \lambda_1$ avoids any need to find a polynomial of degree $n - m$ with no zeros in $\mathbb{F}$; over $\mathbb{R}$ no such polynomial exists when $n - m$ is odd (see Exercise 4.9), and over $\mathbb{C}$ no such polynomial exists when $n - m \geq 1$, by the fundamental theorem of algebra 4.12.

Problem 4.7. Suppose that m is a nonnegative integer, $z_1, \ldots, z_{m+1}$ are distinct elements of $\mathbb{F}$, and $w_1, \ldots, w_{m+1} \in \mathbb{F}$. Prove that there exists a unique polynomial $p \in \mathcal{P}_m(\mathbb{F})$ such that

$$p(z_k) = w_k$$

for each $k = 1, \ldots, m + 1$.
This result can be proved without using linear algebra. However, try to find the clearer, shorter proof that uses some linear algebra.

Solution. Define $T\colon \mathcal{P}_m(\mathbb{F}) \to \mathbb{F}^{m+1}$ by

$$Tp = \big(p(z_1), \ldots, p(z_{m+1})\big).$$

T is linear. Indeed, if $p, q \in \mathcal{P}_m(\mathbb{F})$ and $\lambda \in \mathbb{F}$, then for each k we have $(p + q)(z_k) = p(z_k) + q(z_k)$ and $(\lambda p)(z_k) = \lambda\, p(z_k)$, by the definition of addition and scalar multiplication of functions; hence $T(p + q) = Tp + Tq$ and $T(\lambda p) = \lambda Tp$.
T is injective. Because T is linear, it suffices to show that $\operatorname{null} T = \{0\}$ (by 3.15). So suppose $p \in \mathcal{P}_m(\mathbb{F})$ and $Tp = 0$, meaning

$$p(z_1) = \cdots = p(z_{m+1}) = 0.$$

Because $z_1, \ldots, z_{m+1}$ are distinct, this says that p has at least $m + 1$ distinct zeros in $\mathbb{F}$. Suppose $p \neq 0$ and let $k = \deg p$, so $0 \leq k \leq m$. If $k = 0$, then p is a nonzero constant polynomial and thus has no zeros at all, contradicting the display above (there is at least one z_k, since $m + 1 \geq 1$). If $k \geq 1$, then 4.8 tells us that p has at most $k \leq m$ zeros in $\mathbb{F}$, again contradicting the fact that p has at least $m + 1$ zeros. Hence $p = 0$, and so $\operatorname{null} T = \{0\}$, proving that T is injective.
Now note that

$$\dim \mathcal{P}_m(\mathbb{F}) = m + 1 = \dim \mathbb{F}^{m+1},$$

where the first equality holds because $1, z, \ldots, z^m$ is a basis of $\mathcal{P}_m(\mathbb{F})$ (see 2.36). Both spaces are finite-dimensional and have the same dimension, so by 3.65 the injectivity of T implies that T is invertible; in particular T is surjective as well as injective.
Surjectivity of T applied to the vector $(w_1, \ldots, w_{m+1}) \in \mathbb{F}^{m+1}$ produces $p \in \mathcal{P}_m(\mathbb{F})$ with

$$\big(p(z_1), \ldots, p(z_{m+1})\big) = (w_1, \ldots, w_{m+1}),$$

that is, $p(z_k) = w_k$ for each $k = 1, \ldots, m+1$. This proves existence.
Injectivity of T proves uniqueness: if $p, \tilde{p} \in \mathcal{P}_m(\mathbb{F})$ both satisfy $p(z_k) = w_k = \tilde{p}(z_k)$ for each k, then $Tp = T\tilde{p}$, and hence $p = \tilde{p}$.
Thus there is exactly one $p \in \mathcal{P}_m(\mathbb{F})$ with $p(z_k) = w_k$ for $k = 1, \ldots, m+1$.
For the record, the polynomial can be written down explicitly (the Lagrange interpolation formula):

$$p(z) = \sum_{k=1}^{m+1} w_k \prod_{\substack{j=1 \\ j \neq k}}^{m+1} \frac{z - z_j}{z_k - z_j},$$

where each product is a well-defined polynomial of degree m because the z_j are distinct, so no denominator vanishes; evaluating at $z = z_k$ makes the k-th product equal 1 and every other product equal 0, giving $p(z_k) = w_k$. The linear algebra proof above has the advantage of delivering uniqueness for free, from injectivity, rather than by a separate argument.

Problem 4.8. Suppose $p \in \mathcal{P}(\mathbb{C})$ has degree m. Prove that p has m distinct zeros if and only if p and its derivative p' have no zeros in common.

Solution. Throughout, the derivative of $p(z) = a_0 + a_1 z + \cdots + a_m z^m$ is the polynomial

$$p'(z) = a_1 + 2a_2 z + \cdots + m a_m z^{m-1}.$$

The map $p \mapsto p'$ is linear, and the product rule

$$(fg)' = f'g + fg'$$

holds for all $f, g \in \mathcal{P}(\mathbb{C})$: both sides are bilinear in the pair (f, g), so it suffices to check it for monomials $f(z) = z^j$, $g(z) = z^k$, where it reads $(j+k)z^{j+k-1} = jz^{j-1}z^k + z^j k z^{k-1}$, which is true.
First dispose of the case $m = 0$. Then p is a nonzero constant, so p has no zeros at all; thus p has $m = 0$ distinct zeros. Here p' is the zero polynomial, every complex number being one of its zeros, but since p has no zeros the two polynomials still have no zeros in common. So both sides of the asserted equivalence hold, and the statement is true. Assume from now on that $m \geq 1$.
By the second version of the fundamental theorem of algebra (4.13) there are $c, \lambda_1, \ldots, \lambda_m \in \mathbb{C}$ with $c \neq 0$ such that

$$p(z) = c(z - \lambda_1) \cdots (z - \lambda_m)$$

for all $z \in \mathbb{C}$. The zeros of p are precisely $\lambda_1, \ldots, \lambda_m$, so p has m distinct zeros if and only if $\lambda_1, \ldots, \lambda_m$ are distinct.
Suppose first that $\lambda_1, \ldots, \lambda_m$ are distinct. Repeated application of the product rule to the factorization above gives

$$p'(z) = c \sum_{k=1}^{m} \prod_{j \neq k} (z - \lambda_j).$$

Fix $k \in \{1, \ldots, m\}$ and evaluate at $z = \lambda_k$. For every index $i \neq k$, the product $\prod_{j \neq i}(\lambda_k - \lambda_j)$ contains the factor $\lambda_k - \lambda_k = 0$ and hence vanishes. Only the k^{th} summand survives, giving

$$p'(\lambda_k) = c \prod_{j \neq k} (\lambda_k - \lambda_j),$$

which is a product of nonzero numbers (here we use that the λ's are distinct, so $\lambda_k - \lambda_j \neq 0$ for $j \neq k$) times $c \neq 0$. Hence $p'(\lambda_k) \neq 0$. Since $\lambda_1, \ldots, \lambda_m$ are all the zeros of p, no zero of p is a zero of p'; that is, p and p' have no zeros in common.
Conversely, suppose p does not have m distinct zeros. Then in the factorization above two of the λ's coincide; after relabeling, say $\lambda_1 = \lambda_2 =: \lambda$. Put $g(z) = c(z - \lambda_3) \cdots (z - \lambda_m)$, so that

$$p(z) = (z - \lambda)^2 g(z).$$

By the product rule,

$$p'(z) = 2(z - \lambda)g(z) + (z - \lambda)^2 g'(z),$$

and therefore $p'(\lambda) = 0$. Since also $p(\lambda) = 0$, the number λ is a zero common to p and p'. Taking the contrapositive: if p and p' have no zeros in common, then p has m distinct zeros.
Combining the two implications gives the desired equivalence.

Problem 4.9. Prove that every polynomial of odd degree with real coefficients has a real zero.

Solution. Suppose $p \in \mathcal{P}(\mathbb{R})$ has odd degree n. In particular $n \geq 1$, so p is nonconstant and the factorization theorem over $\mathbb{R}$ (4.16) applies: there exist $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbb{R}$ with $b_k^2 < 4c_k$ for each k, such that

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M)$$

for all $x \in \mathbb{R}$. Here $c \neq 0$: if c were 0, the right side would be the zero polynomial, contradicting $p \neq 0$. Thus c is the coefficient of x^n in p.
Comparing degrees on the two sides of this equation gives

$$n = m + 2M.$$

Because n is odd and $2M$ is even, m is odd. In particular $m \neq 0$, so $m \geq 1$ and the real number λ_1 occurs in the factorization. Evaluating at $x = \lambda_1$ makes the factor $x - \lambda_1$ equal to 0, so

$$p(\lambda_1) = 0.$$

Thus $\lambda_1 \in \mathbb{R}$ is a real zero of p, as desired.
A remark on the structure of the argument: the quadratic factors are exactly the ones that contribute no real zeros, since $b_k^2 < 4c_k$ forces $x^2 + b_k x + c_k = \left(x + \frac{b_k}{2}\right)^2 + \left(c_k - \frac{b_k^2}{4}\right) > 0$ for all $x \in \mathbb{R}$ (this completing-the-square computation is the one used in the proof of 4.15). Each such factor uses up two degrees. Hence a polynomial with no real zeros must have even degree, which is the contrapositive of what was to be proved.

Problem 4.10. For $p \in \mathcal{P}(\mathbb{R})$, define $Tp\colon \mathbb{R} \to \mathbb{R}$ by

$$(Tp)(x) = \begin{cases} \dfrac{p(x) - p(3)}{x - 3} & \text{if } x \neq 3, \\ p'(3) & \text{if } x = 3 \end{cases}$$

for each $x \in \mathbb{R}$. Show that $Tp \in \mathcal{P}(\mathbb{R})$ for every polynomial $p \in \mathcal{P}(\mathbb{R})$ and also show that $T\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ is a linear map.

Solution. Fix $p \in \mathcal{P}(\mathbb{R})$ and let $\tilde{p}$ denote the polynomial $\tilde{p}(x) = p(x) - p(3)$. Then $\tilde{p}(3) = 0$.
If p is constant, then $\tilde{p}$ is the zero polynomial and $p' = 0$; hence $(Tp)(x) = 0$ for $x \neq 3$ and $(Tp)(3) = p'(3) = 0$. So Tp is the zero polynomial, which lies in $\mathcal{P}(\mathbb{R})$.
Now suppose $\deg p = m \geq 1$. Subtracting the constant $p(3)$ changes only the constant coefficient, so $\deg \tilde{p} = m \geq 1$ as well, and 3 is a zero of $\tilde{p}$. By 4.6 there is a polynomial $q \in \mathcal{P}(\mathbb{R})$ of degree $m - 1$

such that

$$p(x) - p(3) = \tilde{p}(x) = (x-3)q(x)$$

for every $x \in \mathbb{R}$. Dividing by $x - 3$ shows that

$$(Tp)(x) = q(x) \qquad \text{for all } x \neq 3.$$

It remains to check that Tp and q also agree at $x = 3$. Differentiating the identity $p(x) - p(3) = (x-3)q(x)$ (the right side by the product rule) gives

$$p'(x) = q(x) + (x-3)q'(x)$$

for all $x \in \mathbb{R}$, and setting $x = 3$ yields $p'(3) = q(3)$. Since $(Tp)(3) = p'(3)$ by definition, we get $(Tp)(3) = q(3)$.
Therefore $Tp = q$ as functions on $\mathbb{R}$, so Tp is a polynomial with real coefficients; that is, $Tp \in \mathcal{P}(\mathbb{R})$. (Its degree is $\deg p - 1$ when p is nonconstant.) This shows T is a well-defined map from $\mathcal{P}(\mathbb{R})$ to $\mathcal{P}(\mathbb{R})$.
Now we verify linearity. Let $p_1, p_2 \in \mathcal{P}(\mathbb{R})$ and $\alpha \in \mathbb{R}$. Two polynomials are equal precisely when they are equal as functions on $\mathbb{R}$, so it suffices to compare values at each $x \in \mathbb{R}$.
For $x \neq 3$,

$$\big(T(p_1+p_2)\big)(x) = \frac{(p_1+p_2)(x) - (p_1+p_2)(3)}{x-3} = \frac{p_1(x)-p_1(3)}{x-3} + \frac{p_2(x)-p_2(3)}{x-3} = (Tp_1)(x) + (Tp_2)(x).$$

For $x = 3$, using that differentiation is additive,

$$\big(T(p_1+p_2)\big)(3) = (p_1+p_2)'(3) = p_1'(3) + p_2'(3) = (Tp_1)(3) + (Tp_2)(3).$$

Hence $T(p_1 + p_2) = Tp_1 + Tp_2$.
Similarly, for $x \neq 3$,

$$\big(T(\alpha p_1)\big)(x) = \frac{\alpha p_1(x) - \alpha p_1(3)}{x-3} = \alpha \cdot \frac{p_1(x)-p_1(3)}{x-3} = \alpha (Tp_1)(x),$$

and at $x = 3$, since $(\alpha p_1)' = \alpha p_1'$,

$$\big(T(\alpha p_1)\big)(3) = (\alpha p_1)'(3) = \alpha p_1'(3) = \alpha (Tp_1)(3).$$

Hence $T(\alpha p_1) = \alpha\, Tp_1$.
Thus T is additive and homogeneous, so $T\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ is a linear map.

Problem 4.11. Suppose $p \in \mathcal{P}(\mathbb{C})$. Define $q\colon \mathbb{C} \to \mathbb{C}$ by

$$q(z) = p(z)\,\overline{p(\bar{z})}.$$

Prove that q is a polynomial with real coefficients.

Solution. Write

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

with $a_0, \ldots, a_m \in \mathbb{C}$, and define a second polynomial $\bar{p} \in \mathcal{P}(\mathbb{C})$ by

$$\bar{p}(z) = \overline{a_0} + \overline{a_1} z + \cdots + \overline{a_m} z^m.$$

The first step is to identify $\overline{p(\bar{z})}$ with $\bar{p}(z)$. Using that conjugation is additive and multiplicative and that $\overline{\bar{z}} = z$ (see 4.4), we have, for every $z \in \mathbb{C}$,

$$\overline{p(\bar{z})} = \overline{\sum_{k=0}^{m} a_k (\bar{z})^k} = \sum_{k=0}^{m} \overline{a_k}\,\overline{(\bar{z})^k} = \sum_{k=0}^{m} \overline{a_k}\,\big(\overline{\bar{z}}\big)^k = \sum_{k=0}^{m} \overline{a_k} z^k = \bar{p}(z).$$

Therefore

$$q(z) = p(z)\,\bar{p}(z) \qquad \text{for all } z \in \mathbb{C},$$

so q is a product of two polynomials and hence is itself a polynomial: $q \in \mathcal{P}(\mathbb{C})$.
It remains to prove that the coefficients of q are real. Multiplying out the product and collecting powers of z gives

$$q(z) = \sum_{n=0}^{2m} c_n z^n, \qquad \text{where } c_n = \sum_{\substack{j+k=n \\ 0\le j,\,k\le m}} a_j \overline{a_k}.$$

(These are the coefficients of q, and by the uniqueness of the coefficients of a polynomial — a consequence of 4.8 — they are the only ones.)
Fix n and conjugate:

$$\overline{c_n} = \overline{\sum_{j+k=n} a_j \overline{a_k}} = \sum_{j+k=n} \overline{a_j}\,\overline{\overline{a_k}} = \sum_{j+k=n} \overline{a_j} a_k = \sum_{j+k=n} a_k \overline{a_j}.$$

The index set $\{(j,k) : j+k=n,\ 0 \le j,k \le m\}$ is carried to itself by the swap $(j,k) \mapsto (k,j)$, so relabeling the summation in the last expression by interchanging the names of j and k turns it back into $\sum_{j+k=n} a_j \overline{a_k} = c_n$. Hence

$$\overline{c_n} = c_n,$$

which by 4.4 means $c_n \in \mathbb{R}$ for every $n \in \{0, 1, \ldots, 2m\}$.
Thus q is a polynomial all of whose coefficients are real, as desired.
(An alternative way to see that each c_n is real: $\overline{q(\bar{z})} = \overline{p(\bar{z})}\,\overline{\bar{p}(\bar{z})} = \bar{p}(z)p(z) = q(z)$ for all z, since $\overline{\bar{p}(\bar{z})} = p(z)$ by the computation above applied to $\bar{p}$. Taking $z = x$ real gives $\overline{q(x)} = q(x)$, so q takes real values on $\mathbb{R}$; then the imaginary parts of the coefficients form a polynomial vanishing identically on $\mathbb{R}$, hence vanish by 4.8.)

Problem 4.12. Suppose m is a nonnegative integer and $p \in \mathcal{P}_m(\mathbb{C})$ is such that there are distinct real numbers $x_0, x_1, \ldots, x_m$ with $p(x_k) \in \mathbb{R}$ for each $k = 0, 1, \ldots, m$. Prove that all coefficients of p are real.

Solution. Write

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

with $a_0, \ldots, a_m \in \mathbb{C}$ (some of these coefficients may be 0, since p is only assumed to have degree at most m). Define

$$u(x) = (\operatorname{Im} a_0) + (\operatorname{Im} a_1)x + \cdots + (\operatorname{Im} a_m)x^m,$$

which is a polynomial with real coefficients of degree at most m.
Suppose $x \in \mathbb{R}$. Then $x^k \in \mathbb{R}$ for each k, and since Im is additive and satisfies $\operatorname{Im}(at) = t \operatorname{Im} a$ for $a \in \mathbb{C}$ and $t \in \mathbb{R}$, we get

$$\operatorname{Im} p(x) = \sum_{k=0}^{m} \operatorname{Im}\big(a_k x^k\big) = \sum_{k=0}^{m} (\operatorname{Im} a_k) x^k = u(x).$$

By hypothesis $p(x_k) \in \mathbb{R}$, so $\operatorname{Im} p(x_k) = 0$, for each $k = 0, 1, \ldots, m$. Hence

$$u(x_0) = u(x_1) = \cdots = u(x_m) = 0,$$

so the polynomial u has at least $m+1$ distinct zeros in $\mathbb{R}$ (the numbers $x_0, \ldots, x_m$ are distinct by assumption).
Now suppose, for the sake of contradiction, that u is not the zero polynomial, and let $d = \deg u$, so $0 \le d \le m$. If $d = 0$, then u is a nonzero constant and has no zeros at all, contradicting the display above (there is at least one x_k, since $m \ge 0$). If $d \ge 1$, then by 4.8 the polynomial u has at most $d \le m$ zeros, again contradicting the fact that it has at least $m+1$ distinct zeros. Therefore $u = 0$.

Since the coefficients of a polynomial are uniquely determined (a consequence of 4.8), $u = 0$ forces

$$\operatorname{Im} a_k = 0 \qquad \text{for each } k = 0, 1, \ldots, m,$$

that is, $a_k \in \mathbb{R}$ for every k. Thus all coefficients of p are real, as desired.
The hypothesis that there are $m+1$ such points cannot be weakened to m points: for instance $p(z) = iz$ lies in $\mathcal{P}_1(\mathbb{C})$ and satisfies $p(0) = 0 \in \mathbb{R}$, yet its coefficients are not all real.
Alternative argument via Exercise 4.7. Set $w_k = p(x_k) \in \mathbb{R}$ for $k = 0, \ldots, m$. Applying Exercise 4.7 over the field $\mathbb{R}$ produces a polynomial $\tilde{p} \in \mathcal{P}_m(\mathbb{R})$ with $\tilde{p}(x_k) = w_k$ for each k. Regarding $\tilde{p}$ as an element of $\mathcal{P}_m(\mathbb{C})$, both p and $\tilde{p}$ lie in $\mathcal{P}_m(\mathbb{C})$ and take the value w_k at x_k for each of the $m+1$ distinct points $x_0, \ldots, x_m$. The uniqueness assertion of Exercise 4.7 over $\mathbb{C}$ then gives $p = \tilde{p}$, so all coefficients of p are real.

Problem 4.13. Suppose $p \in \mathcal{P}(\mathbb{F})$ with $p \neq 0$. Let $U = \{pq : q \in \mathcal{P}(\mathbb{F})\}$.
(a) Show that $\dim \mathcal{P}(\mathbb{F})/U = \deg p$.
(b) Find a basis of $\mathcal{P}(\mathbb{F})/U$.

Solution. Let $m = \deg p$; since $p \neq 0$ we have $m \geq 0$.
First note that U is a subspace of $\mathcal{P}(\mathbb{F})$: it contains $0 = p \cdot 0$, and it is closed under addition and scalar multiplication because $pq_1 + pq_2 = p(q_1 + q_2)$ and $\lambda(pq) = p(\lambda q)$ for all $q_1, q_2, q \in \mathcal{P}(\mathbb{F})$ and $\lambda \in \mathbb{F}$. Hence the quotient space $\mathcal{P}(\mathbb{F})/U$ makes sense.
We prove (a) and (b) together by exhibiting a basis with m elements. Let z^k denote the polynomial $z \mapsto z^k$, and consider the list of cosets

$$1 + U,\ z + U,\ z^2 + U,\ \ldots,\ z^{m-1} + U$$

in $\mathcal{P}(\mathbb{F})/U$. (When $m = 0$ this is the empty list.)
Spanning. Let $r \in \mathcal{P}(\mathbb{F})$ be arbitrary. Because $p \neq 0$, the division algorithm for polynomials (4.9) provides $q, s \in \mathcal{P}(\mathbb{F})$ with

$$r = pq + s \quad \text{and} \quad \deg s < \deg p = m.$$

Since $pq \in U$, we have

$$r + U = s + U.$$

The condition $\deg s < m$ says exactly that $s \in \mathcal{P}_{m-1}(\mathbb{F})$, so $s = a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}$ for some $a_0, \ldots, a_{m-1} \in \mathbb{F}$. Because the quotient map $r \mapsto r + U$ is linear (3.104), this gives

$$r + U = s + U = a_0(1 + U) + a_1(z + U) + \cdots + a_{m-1}\big(z^{m-1} + U\big).$$

Hence the list above spans $\mathcal{P}(\mathbb{F})/U$. (When $m = 0$ the polynomial s must be 0, so every coset equals U and the empty list spans the zero space, consistent with the fact that a nonzero constant p gives $U = \mathcal{P}(\mathbb{F})$.)
Linear independence. Suppose $a_0, \ldots, a_{m-1} \in \mathbb{F}$ satisfy

$$a_0(1 + U) + a_1(z + U) + \cdots + a_{m-1}\big(z^{m-1} + U\big) = 0 + U.$$

Setting $s = a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}$, this says $s + U = 0 + U$, which by 3.101 means $s \in U$. So $s = pq$ for some $q \in \mathcal{P}(\mathbb{F})$. If $q \neq 0$, then

$$\deg s = \deg p + \deg q = m + \deg q \geq m,$$

contradicting $\deg s \leq m - 1$. Hence $q = 0$ and therefore $s = 0$. Since the coefficients of a polynomial are uniquely determined (a consequence of 4.8), $a_0 = a_1 = \cdots = a_{m-1} = 0$. Thus the list is linearly independent.
Conclusion for (b). The list

$$1 + U,\ z + U,\ \ldots,\ z^{m-1} + U$$

is a basis of $\mathcal{P}(\mathbb{F})/U$, where $m = \deg p$. Equivalently: the cosets $z^k + U$ for $k = 0, 1, \ldots, (\deg p) - 1$ form a basis.
Conclusion for (a). This basis has length m, so

$$\dim \mathcal{P}(\mathbb{F})/U = m = \deg p,$$

as claimed. Note in particular that although $\mathcal{P}(\mathbb{F})$ is infinite-dimensional, the quotient $\mathcal{P}(\mathbb{F})/U$ is finite-dimensional.

Problem 4.14. Suppose $p, q \in \mathcal{P}(\mathbb{C})$ are nonconstant polynomials with no zeros in common. Let $m = \deg p$ and $n = \deg q$. Use linear algebra as outlined below in (a)–(c) to prove that there exist $r \in \mathcal{P}_{n-1}(\mathbb{C})$ and $s \in \mathcal{P}_{m-1}(\mathbb{C})$ such that

$$rp + sq = 1.$$

(a) Define $T\colon \mathcal{P}_{n-1}(\mathbb{C}) \times \mathcal{P}_{m-1}(\mathbb{C}) \to \mathcal{P}_{m+n-1}(\mathbb{C})$ by

$$T(r, s) = rp + sq.$$

Show that the linear map T is injective.
(b) Show that the linear map T in (a) is surjective.
(c) Use (b) to conclude that there exist $r \in \mathcal{P}_{n-1}(\mathbb{C})$ and $s \in \mathcal{P}_{m-1}(\mathbb{C})$ such that $rp + sq = 1$.

Solution. Since p and q are nonconstant, $m \geq 1$ and $n \geq 1$.
Preliminaries. Recall that $\deg(uv) = \deg u + \deg v$ for all $u, v \in \mathcal{P}(\mathbb{C})$: if $u, v \neq 0$ the leading coefficient of uv is the product of the leading coefficients of u and v, which is nonzero because $\mathbb{C}$ has no zero divisors; and if one of them is 0 then both sides equal $-\infty$. Two consequences we shall use: (i) $uv = 0$ implies $u = 0$ or $v = 0$; (ii) cancellation holds, i.e. if $u \neq 0$ and $uv = uw$ then $u(v - w) = 0$ and hence $v = w$.
T is well defined and linear. If $\deg r \leq n-1$ and $\deg s \leq m-1$, then $\deg(rp) \leq (n-1)+m = m+n-1$ and $\deg(sq) \leq (m-1)+n = m+n-1$, so $rp + sq \in \mathcal{P}_{m+n-1}(\mathbb{C})$. Linearity is immediate from the distributive law:

$$T\big((r_1, s_1) + (r_2, s_2)\big) = (r_1 + r_2)p + (s_1 + s_2)q = T(r_1, s_1) + T(r_2, s_2),$$

and $T(\lambda r, \lambda s) = \lambda rp + \lambda sq = \lambda\, T(r, s)$ for $\lambda \in \mathbb{C}$.
(a) Injectivity. By 3.15 it suffices to show $\operatorname{null} T = \{(0, 0)\}$. So suppose $(r, s) \in \mathcal{P}_{n-1}(\mathbb{C}) \times \mathcal{P}_{m-1}(\mathbb{C})$ satisfies

$$rp + sq = 0, \qquad \text{i.e.} \qquad rp = -sq.$$

Because q is nonconstant, the second version of the fundamental theorem of algebra (4.13) gives $c \in \mathbb{C}$ with $c \neq 0$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ (not necessarily distinct) such that

$$q(z) = c(z - \lambda_1) \cdots (z - \lambda_n).$$

Each λ_k is a zero of q, hence is not a zero of p; that is,

$$p(\lambda_k) \neq 0 \qquad \text{for } k = 1, \ldots, n.$$

Assume, seeking a contradiction, that $r \neq 0$. We claim that for each $j \in \{0, 1, \ldots, n\}$ there is a polynomial $r_j \in \mathcal{P}(\mathbb{C})$ with $r_j \neq 0$ such that

$$r(z) = (z - \lambda_1) \cdots (z - \lambda_j)\, r_j(z).$$

We prove the claim by induction on j. For $j = 0$ take $r_0 = r$ (the empty product is 1), which is nonzero by assumption. Suppose the claim holds for some j with $0 \leq j < n$. Substituting $r = (z - \lambda_1) \cdots (z - \lambda_j) r_j$ into $rp = -sq$ and using the factorization of q gives

$$(z - \lambda_1) \cdots (z - \lambda_j)\, r_j(z) p(z) = -s(z)\, c\, (z - \lambda_1) \cdots (z - \lambda_j)(z - \lambda_{j+1}) \cdots (z - \lambda_n).$$

The polynomial $(z-\lambda_1)\cdots(z-\lambda_j)$ is nonzero, so cancellation (consequence (ii) above) yields

$$r_j(z)p(z) = -c\,s(z)\,(z-\lambda_{j+1})\cdots(z-\lambda_n)$$

for all $z \in \mathbb{C}$. Evaluating at $z = \lambda_{j+1}$ makes the right side 0, so

$$r_j(\lambda_{j+1})\,p(\lambda_{j+1}) = 0.$$

Since $p(\lambda_{j+1}) \neq 0$, we get $r_j(\lambda_{j+1}) = 0$. Now $r_j \neq 0$ and r_j has a zero, so r_j is not a nonzero constant; hence $\deg r_j \geq 1$ and 4.6 provides a polynomial r_{j+1} with $\deg r_{j+1} = \deg r_j - 1$ (in particular $r_{j+1} \neq 0$) such that $r_j(z) = (z-\lambda_{j+1})r_{j+1}(z)$. Substituting this into the formula for r completes the induction. Taking $j = n$ in the claim gives $r = (z-\lambda_1)\cdots(z-\lambda_n)r_n$ with $r_n \neq 0$, so

$$\deg r = n + \deg r_n \geq n,$$

contradicting $\deg r \leq n-1$. Therefore $r = 0$.
With $r = 0$ the equation $rp = -sq$ becomes $sq = 0$; since $q \neq 0$, consequence (i) gives $s = 0$. Hence $\operatorname{null} T = \{(0,0)\}$ and T is injective.
(b) Surjectivity. Because $1, z, \ldots, z^k$ is a basis of $\mathcal{P}_k(\mathbb{C})$, we have (see 2.36) $\dim \mathcal{P}_k(\mathbb{C}) = k+1$ for every nonnegative integer k. Hence, using the formula for the dimension of a product (3.92),

$$\dim\big(\mathcal{P}_{n-1}(\mathbb{C}) \times \mathcal{P}_{m-1}(\mathbb{C})\big) = \dim \mathcal{P}_{n-1}(\mathbb{C}) + \dim \mathcal{P}_{m-1}(\mathbb{C}) = n + m,$$

while

$$\dim \mathcal{P}_{m+n-1}(\mathbb{C}) = m + n.$$

So T is a linear map between finite-dimensional vector spaces of the same dimension. By 3.65, injectivity is equivalent to surjectivity for such a map, and T is injective by (a). Therefore T is surjective.
(c) Conclusion. Since $m + n - 1 \geq 1 \geq 0$, the constant polynomial 1 belongs to $\mathcal{P}_{m+n-1}(\mathbb{C})$. By (b) there exists a pair

$$(r, s) \in \mathcal{P}_{n-1}(\mathbb{C}) \times \mathcal{P}_{m-1}(\mathbb{C})$$

with $T(r, s) = 1$, that is,

$$rp + sq = 1,$$

which is exactly what was to be proved. (Injectivity of T shows moreover that this pair (r, s) is unique.)

5 Eigenvalues and Eigenvectors

Contents

5.1 Exercises 5A 192
5.2 Exercises 5B 221
5.3 Exercises 5C 244
5.4 Exercises 5D 257
5.5 Exercises 5E 277

5.1 Exercises 5A

Problem 5A.1. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V.
(a) Prove that if $U \subseteq \operatorname{null} T$, then U is invariant under T.
(b) Prove that if $\operatorname{range} T \subseteq U$, then U is invariant under T.

Solution. Recall the definition 5.2: a subspace U of V is invariant under T precisely when $Tu \in U$ for every $u \in U$. So in each part we must take an arbitrary $u \in U$ and show Tu lands back in U.
(a) Suppose $U \subseteq \operatorname{null} T$. Let $u \in U$. Then $u \in \operatorname{null} T$, so

$$Tu = 0.$$

Since U is a subspace of V, it contains 0. Hence $Tu = 0 \in U$. As $u \in U$ was arbitrary, U is invariant under T.
(b) Suppose $\operatorname{range} T \subseteq U$. Let $u \in U$. Then $u \in V$, so $Tu \in \operatorname{range} T$ by the definition of the range. Therefore

$$Tu \in \operatorname{range} T \subseteq U,$$

so $Tu \in U$. As $u \in U$ was arbitrary, U is invariant under T.
Note that the two hypotheses are extreme cases of a general principle: in (a) the operator crushes all of U to the single point 0, which every subspace contains, and in (b) the operator cannot escape U no matter where it starts, since everything in the image of T already lies in U. In particular, taking $U = \operatorname{null} T$ in (a) and $U = \operatorname{range} T$ in (b) recovers the observation of 5.4 that $\operatorname{null} T$ and $\operatorname{range} T$ are invariant under T.

Problem 5A.2. Suppose that $T \in \mathcal{L}(V)$ and $V_1, \dots, V_m$ are subspaces of V invariant under T. Prove that $V_1 + \cdots + V_m$ is invariant under T.

Solution. First recall that a sum of subspaces of V is again a subspace of V (1.40), so $V_1 + \cdots + V_m$ is a subspace of V and it makes sense to ask whether it is invariant under T.
Let $v \in V_1 + \cdots + V_m$. By the definition of the sum of subspaces, there exist vectors

$$v_1 \in V_1, \quad \dots, \quad v_m \in V_m$$

such that

$$v = v_1 + \cdots + v_m.$$

Applying T and using the additivity of T,

$$Tv = T(v_1 + \cdots + v_m) = Tv_1 + \cdots + Tv_m.$$

For each $k \in \{1, \dots, m\}$ we have $v_k \in V_k$ and V_k is invariant under T, so $Tv_k \in V_k$. Hence Tv is a sum of a vector in V_1, a vector in V_2, and so on through a vector in V_m, which means precisely that

$$Tv \in V_1 + \cdots + V_m.$$

Since $v \in V_1 + \cdots + V_m$ was arbitrary, $V_1 + \cdots + V_m$ is invariant under T.

Problem 5A.3. Suppose $T \in \mathcal{L}(V)$. Prove that the intersection of every collection of subspaces of V invariant under T is invariant under T.

Solution. Let $\{U_a\}_{a \in A}$ be a collection of subspaces of V, each invariant under T, indexed by a set A, and set

$$U = \bigcap_{a \in A} U_a.$$

(If A is empty the intersection is interpreted as all of V, which is invariant under T because T maps V into V; so assume A is nonempty.)
First, U is a subspace of V. Indeed, $0 \in U_a$ for every $a \in A$ because each U_a is a subspace, so $0 \in U$. If $u, w \in U$ and $\lambda \in \mathbf{F}$, then for each $a \in A$ we have $u, w \in U_a$, and since U_a is closed under addition and scalar multiplication we get $u + w \in U_a$ and $\lambda u \in U_a$. As this holds for every $a \in A$, we conclude $u + w \in U$ and $\lambda u \in U$. Thus U contains 0 and is closed under addition and scalar multiplication, so U is a subspace of V (1.34).
Now we show U is invariant under T. Let $u \in U$. Fix any $a \in A$. Then $u \in U_a$, and U_a is invariant under T, so

$$Tu \in U_a.$$

Since $a \in A$ was arbitrary, Tu belongs to U_a for every $a \in A$, and therefore

$$Tu \in \bigcap_{a \in A} U_a = U.$$

As $u \in U$ was arbitrary, U is invariant under T.

Problem 5A.4. Prove or give a counterexample: If V is finite-dimensional and U is a subspace of V that is invariant under every operator on V, then $U = \{0\}$ or $U = V$.

Solution. The statement is true; we prove it.
Suppose V is finite-dimensional and U is a subspace of V invariant under every operator on V. If $U = \{0\}$ we are done, so assume $U \neq \{0\}$. We show $U = V$.
Because $U \neq \{0\}$, we may choose $u \in U$ with $u \neq 0$. The list consisting of the single vector u is linearly independent (a one-vector list is linearly independent exactly when the vector is nonzero). Since V is finite-dimensional, 2.32 lets us extend this list to a basis

$$u, e_2, \ldots, e_n$$

of V, where $n = \dim V$.
Now let $v \in V$ be arbitrary. By the linear map lemma 3.4 there is a (unique) linear map $T\colon V \to V$ with

$$Tu = v, \qquad Te_2 = 0, \quad \ldots, \quad Te_n = 0,$$

since a linear map may be defined by prescribing its values arbitrarily on a basis. This T is an operator on V, so by hypothesis U is invariant under T. Since $u \in U$, invariance gives

$$v = Tu \in U.$$

As $v \in V$ was arbitrary, $V \subseteq U$. Combined with $U \subseteq V$, this gives $U = V$.
Hence $U = \{0\}$ or $U = V$, as claimed.
Two remarks. First, the hypothesis that U is invariant under every operator is a very strong assumption, and we used only a small part of it: for each target vector v we needed just the single operator T constructed above. Second, finite-dimensionality entered only through 2.32, the extension of a linearly independent list to a basis. If one is willing to use the existence of a basis of an arbitrary vector space (a fact whose proof requires the axiom of choice and which is not proved in this book), the same argument gives the same conclusion without the finite-dimensionality hypothesis.

Problem 5A.5. Suppose $T \in \mathcal{L}(\mathbf{R}^2)$ is defined by $T(x, y) = (-3y, x)$. Find the eigenvalues of T.

Solution. The answer: T has no eigenvalues.
By definition 5.5, a number $\lambda \in \mathbf{R}$ is an eigenvalue of T if and only if there is a nonzero $(x, y) \in \mathbf{R}^2$ with $T(x, y) = \lambda(x, y)$. Written out, this equation says

$$(-3y,\, x) = (\lambda x,\, \lambda y),$$

which is equivalent to the simultaneous equations

$$-3y = \lambda x, \qquad x = \lambda y.$$

Substituting the second equation into the first gives

$$-3y = \lambda(\lambda y) = \lambda^2 y,$$

that is,

$$(\lambda^2 + 3)y = 0.$$

Suppose $\lambda \in \mathbf{R}$ is an eigenvalue, with eigenvector $(x, y) \neq (0, 0)$. If $y = 0$, then the equation $x = \lambda y$ forces $x = 0$, so $(x, y) = (0, 0)$, a contradiction. Hence $y \neq 0$, and dividing the displayed equation by y gives

$$\lambda^2 = -3.$$

But $\lambda^2 \geq 0$ for every real number λ, so no real λ satisfies this. Therefore T has no eigenvalues (and hence no eigenvectors).

Geometrically this is unsurprising. Writing $S(x, y) = (x, 3y)$ for the stretch by a factor of 3 in the second coordinate and $R(x, y) = (-y, x)$ for the counterclockwise rotation by 90 degrees about the origin, we have

$$R\big(S(x, y)\big) = R(x, 3y) = (-3y,\, x) = T(x, y),$$

so $T = RS$. A stretch along an axis followed by a quarter turn moves every nonzero vector off the line through it, which is exactly the computation carried out above. Equivalently, by the comments following 5.5, T has no one-dimensional invariant subspace.

For contrast, if the same formula defined an operator on $\mathbf{C}^2$, then $\lambda^2 = -3$ would have the two solutions $\lambda = \sqrt{3}\, i$ and $\lambda = -\sqrt{3}\, i$, and both would be eigenvalues. This illustrates the point made in 5.9 that the field matters.

Problem 5A.6. Define $T \in \mathcal{L}(\mathbf{F}^2)$ by $T(w, z) = (z, w)$. Find all eigenvalues and eigenvectors of T.

Solution. The answer: the eigenvalues of T are 1 and -1. The eigenvectors corresponding to 1 are the vectors (w, w) with $w \in \mathbf{F}$ and $w \neq 0$; the eigenvectors corresponding to -1 are the vectors $(w, -w)$ with $w \in \mathbf{F}$ and $w \neq 0$.

A scalar $\lambda \in \mathbf{F}$ is an eigenvalue of T with eigenvector (w, z) exactly when $(w, z) \neq (0, 0)$ and $T(w, z) = \lambda(w, z)$, that is, when

$$(z,\, w) = (\lambda w,\, \lambda z).$$

This is equivalent to the pair of equations

$$z = \lambda w, \qquad w = \lambda z.$$

Substituting the first equation into the second gives

$$w = \lambda(\lambda w) = \lambda^2 w,$$

so

$$(\lambda^2 - 1)w = 0.$$

If $w = 0$, then $z = \lambda w = 0$ and (w, z) is the zero vector, which is not an eigenvector. Hence $w \neq 0$, and dividing by w gives $\lambda^2 = 1$, so

$$\lambda = 1 \quad \text{or} \quad \lambda = -1.$$

Thus 1 and -1 are the only candidates for eigenvalues. (This is consistent with 5.12: an operator on the two-dimensional space $\mathbf{F}^2$ has at most 2 distinct eigenvalues.) Note that these two candidates are distinct scalars because $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$.

Now we check each candidate and determine its eigenvectors.

Case $\lambda = 1$. The equations become $z = w$ and $w = z$, which together say exactly $z = w$. So the vectors satisfying $T(w, z) = (w, z)$ are precisely those of the form (w, w) with $w \in \mathbf{F}$. Indeed $T(w, w) = (w, w)$. Taking $w = 1$ shows $(1, 1)$ is a nonzero such vector, so 1 is an eigenvalue, and the eigenvectors corresponding to 1 are the vectors (w, w) with $w \neq 0$. Equivalently,

$$\operatorname{null}(T - I) = \{(w, w) : w \in \mathbf{F}\} = \operatorname{span}\big((1, 1)\big).$$

Case $\lambda = -1$. The equations become $z = -w$ and $w = -z$, which are the same condition $z = -w$. So the solutions are the vectors $(w, -w)$ with $w \in \mathbf{F}$, and indeed

$$T(w, -w) = (-w,\, w) = -(w, -w).$$

Taking $w = 1$ shows $(1, -1)$ is nonzero, so -1 is an eigenvalue, and the eigenvectors corresponding to -1 are the vectors $(w, -w)$ with $w \neq 0$. Equivalently,

$$\operatorname{null}(T + I) = \{(w, -w) : w \in \mathbf{F}\} = \operatorname{span}\big((1, -1)\big).$$

The answer is the same for $\mathbf{F} = \mathbf{R}$ and $\mathbf{F} = \mathbf{C}$, in contrast with the operator $(w, z) \mapsto (-z, w)$ of 5.9. Note also that $(1, 1)$ and $(1, -1)$ are eigenvectors corresponding to the distinct eigenvalues 1 and -1, so the list $(1, 1), (1, -1)$ is linearly independent by 5.11; since its length equals $\dim \mathbf{F}^2 = 2$, it is a basis of $\mathbf{F}^2$ by 2.38. Thus $\mathbf{F}^2$ has a basis consisting of eigenvectors of T, and the matrix of T with respect to that basis is the diagonal matrix with diagonal entries 1 and -1.

Problem 5A.7. Define $T \in \mathcal{L}(\mathbf{F}^3)$ by $T(z_1, z_2, z_3) = (2z_2, 0, 5z_3)$. Find all eigenvalues and eigenvectors of T.

Solution. The answer: the eigenvalues of T are 0 and 5. The eigenvectors corresponding to 0 are the vectors $(z_1, 0, 0)$ with $z_1 \in \mathbf{F}$ and $z_1 \neq 0$; the eigenvectors corresponding to 5 are the vectors $(0, 0, z_3)$ with $z_3 \in \mathbf{F}$ and $z_3 \neq 0$.
A scalar $\lambda \in \mathbf{F}$ is an eigenvalue with eigenvector $(z_1, z_2, z_3) \neq (0, 0, 0)$ exactly when

$$(2z_2,\, 0,\, 5z_3) = (\lambda z_1,\, \lambda z_2,\, \lambda z_3),$$

which is equivalent to the three equations

$$2z_2 = \lambda z_1, \qquad 0 = \lambda z_2, \qquad 5z_3 = \lambda z_3.$$

We analyze these by cases on λ.
Case $\lambda = 0$. The equations reduce to $2z_2 = 0$, $0 = 0$, and $5z_3 = 0$, that is, $z_2 = 0$ and $z_3 = 0$, with z_1 unrestricted. So the solutions are exactly the vectors $(z_1, 0, 0)$ with $z_1 \in \mathbf{F}$, and indeed

$$T(z_1, 0, 0) = (0, 0, 0) = 0 \cdot (z_1, 0, 0).$$

Taking $z_1 = 1$ gives the nonzero vector $(1, 0, 0)$, so 0 is an eigenvalue of T, and the corresponding eigenvectors are the vectors $(z_1, 0, 0)$ with $z_1 \neq 0$. In other words,

$$\operatorname{null} T = \{(z_1, 0, 0) : z_1 \in \mathbf{F}\} = \operatorname{span}\big((1, 0, 0)\big).$$

(That 0 is an eigenvalue is equivalent to T not being injective, by 5.7 with $\lambda = 0$; and indeed $T(1, 0, 0) = 0$.)
Case $\lambda \neq 0$. From the second equation $0 = \lambda z_2$ and $\lambda \neq 0$ we get $z_2 = 0$. Substituting into the first equation gives $\lambda z_1 = 2z_2 = 0$, and again since $\lambda \neq 0$ we get $z_1 = 0$. So any eigenvector for a nonzero λ has the form $(0, 0, z_3)$, and since an eigenvector is nonzero we must have $z_3 \neq 0$. The third equation $5z_3 = \lambda z_3$ then gives, after dividing by $z_3 \neq 0$,

$$\lambda = 5.$$

So $\lambda = 5$ is the only possible nonzero eigenvalue, and it really is one: for $z_3 \in \mathbf{F}$,

$$T(0, 0, z_3) = (0, 0, 5z_3) = 5 \cdot (0, 0, z_3),$$

so $(0, 0, 1)$ is an eigenvector corresponding to 5. The eigenvectors corresponding to 5 are exactly the vectors $(0, 0, z_3)$ with $z_3 \neq 0$, that is,

$$\operatorname{null}(T - 5I) = \operatorname{span}\big((0, 0, 1)\big).$$

Hence the complete list of eigenvalues is $0, 5$, with eigenvectors as described. Notice that T has only 2 distinct eigenvalues even though $\dim \mathbf{F}^3 = 3$; 5.12 gives an upper bound of 3, not an exact count. Indeed $\mathbf{F}^3$ has no basis consisting of eigenvectors of T: every eigenvector of T lies in $\operatorname{span}\big((1, 0, 0), (0, 0, 1)\big)$, which is a two-dimensional subspace, and a list of three vectors lying in a two-dimensional subspace cannot be linearly independent (by 2.22 applied inside that subspace). The obstruction is visible in the matrix of T with respect to the standard basis,

$$\begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{pmatrix},$$

whose upper-left 2×2 block is nilpotent but nonzero.

Problem 5A.8. Suppose $P \in \mathcal{L}(V)$ is such that $P^2 = P$. Prove that if λ is an eigenvalue of P, then $\lambda = 0$ or $\lambda = 1$.

Solution. Suppose $\lambda \in \mathbf{F}$ is an eigenvalue of P. By the definition of eigenvalue (5.5) there exists $v \in V$ with $v \neq 0$ and

$$Pv = \lambda v.$$

Apply P to both sides of this equation. On the left we get $P(Pv) = P^2 v = Pv = \lambda v$, using the hypothesis $P^2 = P$ and then the eigenvalue equation. On the right, linearity of P gives $P(\lambda v) = \lambda P v = \lambda(\lambda v) = \lambda^2 v$. Hence

$$\lambda v = \lambda^2 v,$$

which we rewrite as

$$(\lambda^2 - \lambda)v = 0.$$

Because $v \neq 0$, the scalar $\lambda^2 - \lambda$ must equal 0 (if a scalar c satisfies $cv = 0$ with $v \neq 0$, then $c = 0$, since otherwise multiplying by c^{-1} would give $v = 0$). Thus

$$\lambda(\lambda - 1) = 0.$$

Because $\mathbf{F}$ (which is $\mathbf{R}$ or $\mathbf{C}$) has no zero divisors, this forces $\lambda = 0$ or $\lambda = 1$, as desired.

Problem 5A.9. Define $T\colon \mathcal{P}(\mathbf{R}) \to \mathcal{P}(\mathbf{R})$ by $Tp = p'$. Find all eigenvalues and eigenvectors of T.

Solution. The only eigenvalue of T is 0, and the eigenvectors corresponding to 0 are exactly the nonzero constant polynomials.

First we verify that 0 is an eigenvalue. Let p be the constant polynomial $p(x) = 1$. Then $p \neq 0$ and $Tp = p' = 0 = 0 \cdot p$, so 0 is an eigenvalue of T and p is a corresponding eigenvector.

Now suppose $\lambda \in \mathbf{R}$ is an eigenvalue of T with $\lambda \neq 0$; we derive a contradiction. There exists $p \in \mathcal{P}(\mathbf{R})$ with $p \neq 0$ and $p' = \lambda p$. Because $p \neq 0$, it has a well-defined degree $m \geq 0$; write

$$p(x) = a_0 + a_1 x + \cdots + a_m x^m, \qquad a_m \neq 0.$$

If $m = 0$, then $p' = 0$ while λp is the nonzero constant λa_0 (nonzero because $\lambda \neq 0$ and $a_0 = a_m \neq 0$),

so $p' \neq \lambda p$, a contradiction. If $m \geq 1$, then

$$p'(x) = a_1 + 2a_2 x + \cdots + m a_m x^{m-1},$$

whose coefficient of x^{m-1} is $m a_m \neq 0$ (here we use that $\mathbf{R}$ has characteristic 0, so $m a_m \neq 0$). Thus p' has degree $m-1$, whereas λp has degree m because $\lambda a_m \neq 0$. Two polynomials of different degrees cannot be equal, so again $p' \neq \lambda p$, a contradiction.
Hence no nonzero scalar is an eigenvalue of T, so 0 is the only eigenvalue.
Finally we identify the eigenvectors corresponding to 0. A vector $p \neq 0$ is such an eigenvector if and only if $Tp = 0 \cdot p$, that is, $p' = 0$. Writing $p(x) = a_0 + a_1 x + \cdots + a_m x^m$, the condition $p' = 0$ says $a_1 = 2a_2 = \cdots = m a_m = 0$, hence $a_1 = \cdots = a_m = 0$, i.e. p is constant. So the eigenvectors of T are precisely the nonzero constant polynomials.

Problem 5A.10. Define $T \in \mathcal{L}(\mathcal{P}_4(\mathbf{R}))$ by $(Tp)(x) = x p'(x)$ for all $x \in \mathbf{R}$. Find all eigenvalues and eigenvectors of T.

Solution. The eigenvalues of T are exactly $0, 1, 2, 3, 4$. For each $j \in \{0, 1, 2, 3, 4\}$, the eigenvectors of T corresponding to j are exactly the nonzero scalar multiples of the polynomial x^j.
Work with the standard basis $1, x, x^2, x^3, x^4$ of $\mathcal{P}_4(\mathbf{R})$. For each $j \in \{0, 1, 2, 3, 4\}$, if $p(x) = x^j$ then $p'(x) = j x^{j-1}$ (interpreting this as 0 when $j = 0$), so

$$(Tp)(x) = x \cdot j x^{j-1} = j x^j = j\, p(x).$$

Thus $Tx^j = j x^j$, and since $x^j \neq 0$, each of $0, 1, 2, 3, 4$ is an eigenvalue of T with x^j a corresponding eigenvector.
Now suppose $\lambda \in \mathbf{R}$ and $p \in \mathcal{P}_4(\mathbf{R})$ satisfy $Tp = \lambda p$. Write

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$$

with $a_0, \ldots, a_4 \in \mathbf{R}$. Then

$$(Tp)(x) = x\big(a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3\big) = a_1 x + 2a_2 x^2 + 3a_3 x^3 + 4a_4 x^4,$$

so $Tp = \sum_{j=0}^{4} j\, a_j x^j$. The equation $Tp = \lambda p$ therefore says

$$\sum_{j=0}^{4} j\, a_j x^j = \sum_{j=0}^{4} \lambda\, a_j x^j.$$

Because $1, x, x^2, x^3, x^4$ is a basis of $\mathcal{P}_4(\mathbf{R})$, coefficients may be compared, giving

$$(j - \lambda) a_j = 0 \qquad \text{for } j = 0, 1, 2, 3, 4.$$

If $\lambda \notin \{0, 1, 2, 3, 4\}$, then $j - \lambda \neq 0$ for every such j, forcing $a_0 = a_1 = a_2 = a_3 = a_4 = 0$ and hence $p = 0$. Thus no such λ is an eigenvalue, and the list $0, 1, 2, 3, 4$ found above is the complete list of eigenvalues. (Consistently with 5.12, there are $5 = \dim \mathcal{P}_4(\mathbf{R})$ of them, the maximum possible.)
If $\lambda = k$ for some $k \in \{0, 1, 2, 3, 4\}$, then $(j - k) a_j = 0$ gives $a_j = 0$ for every $j \neq k$, so $p(x) = a_k x^k$. Conversely every such p satisfies $Tp = kp$. Hence the eigenvectors corresponding to the eigenvalue k are exactly the polynomials $a_k x^k$ with $a_k \neq 0$.

Problem 5A.11. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\alpha \in \mathbf{F}$. Prove that there exists $\delta > 0$ such that $T - \lambda I$ is invertible for all $\lambda \in \mathbf{F}$ such that $0 < |\alpha - \lambda| < \delta$.

Solution. The point is that T has only finitely many eigenvalues, so the eigenvalues other than α stay a positive distance away from α.

By 5.12, the operator T has at most $\dim V$ distinct eigenvalues. In particular the set

$$E = \{\mu \in \mathbf{F} : \mu \text{ is an eigenvalue of } T \text{ and } \mu \neq \alpha\}$$

is finite.
Case 1: $E = \emptyset$. Take $\delta = 1$.
Case 2: $E \neq \emptyset$. Write $E = \{\lambda_1, \ldots, \lambda_m\}$ and set

$$\delta = \min\{|\alpha - \lambda_1|, \ldots, |\alpha - \lambda_m|\}.$$

This is a minimum of finitely many numbers, each of which is strictly positive because $\lambda_j \neq \alpha$ for each j. Hence $\delta > 0$.
In either case, suppose $\lambda \in \mathbf{F}$ satisfies $0 < |\alpha - \lambda| < \delta$. We claim λ is not an eigenvalue of T. Indeed, $|\alpha - \lambda| > 0$ gives $\lambda \neq \alpha$, so if λ were an eigenvalue of T then $\lambda \in E$. In Case 1 this is impossible since E is empty. In Case 2 it would give $\lambda = \lambda_j$ for some j, whence

$$|\alpha - \lambda| = |\alpha - \lambda_j| \geq \delta,$$

contradicting $|\alpha - \lambda| < \delta$. So λ is not an eigenvalue of T.
Because V is finite-dimensional, 5.7 tells us that λ is an eigenvalue of T if and only if $T - \lambda I$ is not invertible. Since λ is not an eigenvalue of T, the operator $T - \lambda I$ is invertible. This holds for every $\lambda \in \mathbf{F}$ with $0 < |\alpha - \lambda| < \delta$, completing the proof.
(In the language of analysis: the set of eigenvalues of T is finite, hence every point of $\mathbf{F}$ has a punctured neighborhood containing no eigenvalues. Note that the hypothesis says nothing about α itself; α may well be an eigenvalue of T, which is exactly why the punctured condition $0 < |\alpha - \lambda|$ appears.)

Problem 5A.12. Suppose $V = U \oplus W$, where U and W are nonzero subspaces of V. Define $P \in \mathcal{L}(V)$ by $P(u + w) = u$ for each $u \in U$ and each $w \in W$. Find all eigenvalues and eigenvectors of P.

Solution. The eigenvalues of P are exactly 0 and 1. The eigenvectors corresponding to 1 are the nonzero vectors in U; the eigenvectors corresponding to 0 are the nonzero vectors in W.
First note that P is well defined: because $V = U \oplus W$, every $v \in V$ has a unique representation $v = u + w$ with $u \in U$ and $w \in W$ (this uniqueness is exactly the definition of direct sum, 1.41), so the formula assigns exactly one value to each v.
Only 0 and 1 can be eigenvalues. For $v = u + w$ as above we have $Pv = u$, and since $u = u + 0$ is the decomposition of u, we get $P(Pv) = Pu = u = Pv$. Thus $P^2 = P$, so by Exercise 5A.8 every eigenvalue of P equals 0 or 1.
Both 0 and 1 are eigenvalues. Because $U \neq \{0\}$ there exists $u \in U$ with $u \neq 0$, and $Pu = u = 1 \cdot u$, so 1 is an eigenvalue of P. Because $W \neq \{0\}$ there exists $w \in W$ with $w \neq 0$, and $Pw = 0 = 0 \cdot w$ (here $w = 0 + w$ is the decomposition of w), so 0 is an eigenvalue of P.
Eigenvectors corresponding to 1. Suppose $v \in V$ with $v \neq 0$ and $Pv = v$. Write $v = u + w$ with $u \in U$, $w \in W$. Then $u = Pv = v = u + w$, so $w = 0$ and hence $v = u \in U$. Conversely, if $v \in U$ and $v \neq 0$, then $Pv = v$ as computed above. So the eigenvectors of P corresponding to 1 are precisely the nonzero elements of U; equivalently $\operatorname{null}(P - I) = U$.
Eigenvectors corresponding to 0. Suppose $v \in V$ with $v \neq 0$ and $Pv = 0$. Write $v = u + w$ with $u \in U$, $w \in W$. Then $u = Pv = 0$, so $v = w \in W$. Conversely, if $v \in W$ and $v \neq 0$, then $Pv = 0$. So the eigenvectors of P corresponding to 0 are precisely the nonzero elements of W; equivalently $\operatorname{null} P = W$.

Problem 5A.13. Suppose $T \in \mathcal{L}(V)$. Suppose $S \in \mathcal{L}(V)$ is invertible.
- (a) Prove that T and $S^{-1}TS$ have the same eigenvalues.
- (b) What is the relationship between the eigenvectors of T and the eigenvectors of $S^{-1}TS$?

Solution. (a) Suppose $\lambda \in \mathbf{F}$ is an eigenvalue of T. Then there exists $v \in V$ with $v \neq 0$ and $Tv = \lambda v$. Put

$$w = S^{-1}v.$$

Then $w \neq 0$: if $w = 0$ then applying S gives $v = Sw = 0$, contradicting $v \neq 0$. Moreover, using $SS^{-1} = I$,

$$(S^{-1}TS)w = S^{-1}TSS^{-1}v = S^{-1}Tv = S^{-1}(\lambda v) = \lambda S^{-1}v = \lambda w,$$

where the second-to-last equality uses linearity of S^{-1}. Hence λ is an eigenvalue of $S^{-1}TS$, with eigenvector w.
For the reverse inclusion, apply what was just proved with T replaced by the operator $S^{-1}TS$ and with S replaced by the invertible operator S^{-1} (whose inverse is S). This shows that every eigenvalue of $S^{-1}TS$ is an eigenvalue of

$$(S^{-1})^{-1}(S^{-1}TS)S^{-1} = SS^{-1}TSS^{-1} = T.$$

Thus T and $S^{-1}TS$ have exactly the same eigenvalues.
(Alternative proof, valid in the same generality: for each $\lambda \in \mathbf{F}$ we have the identity

$$S^{-1}TS - \lambda I = S^{-1}(T - \lambda I)S,$$

because $S^{-1}(\lambda I)S = \lambda S^{-1}S = \lambda I$. Since S and S^{-1} are injective and surjective, $S^{-1}(T - \lambda I)S$ is injective if and only if $T - \lambda I$ is injective. By 5.5 and the remark following 5.8, λ is an eigenvalue of an operator exactly when that operator minus λI fails to be injective, so λ is an eigenvalue of $S^{-1}TS$ if and only if it is an eigenvalue of T.)
(b) The eigenvectors correspond under the invertible map S^{-1}, with the eigenvalue unchanged. Precisely: for $\lambda \in \mathbf{F}$ and $w \in V$ with $w \neq 0$,

w is an eigenvector of $S^{-1}TS$ corresponding to $\lambda \iff Sw$ is an eigenvector of T corresponding to λ.

To see this, note that $Sw \neq 0$ if and only if $w \neq 0$, since S is injective and $S0 = 0$. Now apply S to the equation $(S^{-1}TS)w = \lambda w$: this equation holds if and only if

$$S(S^{-1}TS)w = S(\lambda w),$$

that is, if and only if $T(Sw) = \lambda(Sw)$ (again because S is injective, applying S loses no information, and $S(\lambda w) = \lambda Sw$ by linearity). This is exactly the statement that Sw is an eigenvector of T corresponding to λ.
Equivalently, in terms of null spaces, for every $\lambda \in \mathbf{F}$,

$$\operatorname{null}\big(S^{-1}TS - \lambda I\big) = S^{-1}\big(\operatorname{null}(T - \lambda I)\big),$$

so S^{-1} maps the set of eigenvectors of T corresponding to λ bijectively onto the set of eigenvectors of $S^{-1}TS$ corresponding to λ, and S maps them back.

Problem 5A.14. Give an example of an operator on $\mathbf{R}^4$ that has no (real) eigenvalues.

Solution. Define $T \in \mathcal{L}(\mathbf{R}^4)$ by

$$T(x_1, x_2, x_3, x_4) = (-x_2,\ x_1,\ -x_4,\ x_3).$$

This map is clearly linear (each coordinate of the output is a linear function of the input coordinates). Geometrically, T is a counterclockwise rotation by 90° in each of the two coordinate planes $\{(x_1, x_2, 0, 0)\}$ and $\{(0, 0, x_3, x_4)\}$; compare 5.9, where the same idea is used on $\mathbf{R}^2$.
We claim T has no eigenvalues. Suppose $\lambda \in \mathbf{R}$ and $v = (x_1, x_2, x_3, x_4) \in \mathbf{R}^4$ satisfy $Tv = \lambda v$. Comparing coordinates gives the four equations

$$-x_2 = \lambda x_1, \qquad x_1 = \lambda x_2, \qquad -x_4 = \lambda x_3, \qquad x_3 = \lambda x_4.$$

Substituting the second equation into the first gives

$$-x_2 = \lambda(\lambda x_2) = \lambda^2 x_2,$$

hence $(\lambda^2 + 1)x_2 = 0$. Because λ is a real number, $\lambda^2 + 1 \geq 1 > 0$, so $x_2 = 0$. Then $x_1 = \lambda x_2 = 0$.
Substituting the fourth equation into the third gives, in the same way,

$$-x_4 = \lambda^2 x_4,$$

so $(\lambda^2 + 1)x_4 = 0$, forcing $x_4 = 0$, and then $x_3 = \lambda x_4 = 0$.
Thus $v = (0, 0, 0, 0)$. Hence there is no nonzero $v \in \mathbf{R}^4$ with $Tv = \lambda v$, for any $\lambda \in \mathbf{R}$. By the definition of eigenvalue (5.5), T has no eigenvalues.
Note that the analogous operator on $\mathbf{C}^4$ does have eigenvalues, namely i and $-i$: the obstruction above disappears once $\lambda^2 + 1 = 0$ has solutions. This is consistent with 5.19, which guarantees that every operator on a nonzero finite-dimensional complex vector space has an eigenvalue.

Problem 5A.15. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{F}$. Show that λ is an eigenvalue of T if and only if λ is an eigenvalue of the dual operator $T' \in \mathcal{L}(V')$.

Solution. Throughout, I denotes the identity operator on V and $I_{V'}$ denotes the identity operator on the dual space V'. Recall (3.118) that for $S \in \mathcal{L}(V)$ the dual map $S' \in \mathcal{L}(V')$ is defined by $S'(\varphi) = \varphi \circ S$ for $\varphi \in V'$.
First observe that the dual of the identity operator is the identity operator on V': for every $\varphi \in V'$ we have

$$I'(\varphi) = \varphi \circ I = \varphi,$$

so $I' = I_{V'}$.
Next, by 3.120(a) and 3.120(b) (the dual map operation is additive and homogeneous),

$$(T - \lambda I)' = T' + (-\lambda I)' = T' - \lambda I' = T' - \lambda I_{V'}.$$

Because V is finite-dimensional, so is V'; indeed $\dim V' = \dim V$ by 3.111. Hence 5.7 applies both to the operator T on V and to the operator T' on V'.
Now chain the equivalences:

$$\begin{aligned} \lambda \text{ is an eigenvalue of } T &\iff T - \lambda I \text{ is not surjective} \\ &\iff (T - \lambda I)' \text{ is not injective} \\ &\iff T' - \lambda I_{V'} \text{ is not injective} \\ &\iff \lambda \text{ is an eigenvalue of } T'. \end{aligned}$$

Here the first equivalence is the equivalence of (a) and (c) in 5.7 applied to $T \in \mathcal{L}(V)$; the second is 3.129 (S is surjective if and only if S' is injective, valid because V is finite-dimensional), applied to the operator $S = T - \lambda I$ and then negated; the third is the displayed identity $(T - \lambda I)' = T' - \lambda I_{V'}$; and the fourth is the equivalence of (a) and (b) in 5.7 applied to $T' \in \mathcal{L}(V')$, which is legitimate since V' is finite-dimensional.
This proves that λ is an eigenvalue of T if and only if λ is an eigenvalue of T'.
Two remarks. First, finite-dimensionality is genuinely used: 5.7 and 3.129 both require it. Second, although T and T' have the same eigenvalues, there is in general no natural identification of their eigenvectors, since T' acts on the dual space rather than on V.

Problem 5A.16. Suppose $v_1, \ldots, v_n$ is a basis of V and $T \in \mathcal{L}(V)$. Prove that if λ is an eigenvalue of T, then

$$|\lambda| \leq n \max\big\{ \big|\mathcal{M}(T)_{j,k}\big| : 1 \leq j, k \leq n \big\},$$

where $\mathcal{M}(T)_{j,k}$ denotes the entry in row j, column k of the matrix of T with respect to the basis $v_1, \dots, v_n$.
See Exercise 19 in Section 6A for a different bound on $|\lambda|$.

Solution. Write $A = \mathcal{M}(T)$ for the matrix of T with respect to the basis $v_1, \dots, v_n$, so that by the definition of the matrix of a linear map (3.31), applied with the basis $v_1, \dots, v_n$ used for both the domain and the codomain,

$$Tv_k = \sum_{j=1}^{n} A_{j,k}\, v_j \qquad \text{for each } k = 1, \dots, n,$$

and set

$$C = \max\big\{|A_{j,k}| : 1 \le j, k \le n\big\}.$$

Our goal is to show $|\lambda| \le nC$.
Suppose λ is an eigenvalue of T, and let $v \in V$ be an eigenvector corresponding to λ; thus $v \neq 0$ and $Tv = \lambda v$. Because $v_1, \dots, v_n$ is a basis of V, there are unique scalars $c_1, \dots, c_n \in \mathbf{F}$ with

$$v = \sum_{k=1}^{n} c_k v_k.$$

Since $v \neq 0$, not all the c_k are 0. Choose an index $m \in \{1, \dots, n\}$ with

$$|c_m| = \max\big\{|c_1|, \dots, |c_n|\big\},$$

so in particular $|c_m| > 0$.
Now compute Tv two ways. Using linearity of T and the formula for Tv_k,

$$Tv = \sum_{k=1}^{n} c_k\, Tv_k = \sum_{k=1}^{n} c_k \sum_{j=1}^{n} A_{j,k} v_j = \sum_{j=1}^{n} \Big(\sum_{k=1}^{n} A_{j,k} c_k \Big) v_j.$$

On the other hand $Tv = \lambda v = \sum_{j=1}^{n} \lambda c_j\, v_j$. Because the representation of a vector in terms of a basis is unique, comparing coefficients of v_j gives

$$\lambda c_j = \sum_{k=1}^{n} A_{j,k} c_k \qquad \text{for each } j = 1, \dots, n.$$

Apply this with $j = m$ and take absolute values. Using the triangle inequality and then the bounds $|A_{m,k}| \le C$ and $|c_k| \le |c_m|$, we get

$$|\lambda|\, |c_m| = \Big| \sum_{k=1}^{n} A_{m,k} c_k \Big| \le \sum_{k=1}^{n} |A_{m,k}|\, |c_k| \le \sum_{k=1}^{n} C\, |c_m| = nC|c_m|.$$

Since $|c_m| > 0$, we may divide both sides by $|c_m|$ to conclude

$$|\lambda| \le nC = n \max\big\{\, \big|\mathcal{M}(T)_{j,k}\big| : 1 \le j, k \le n \,\big\},$$

as desired.

Problem 5A.17. Suppose $\mathbf{F} = \mathbf{R}$, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{R}$. Prove that λ is an eigenvalue of T if and only if λ is an eigenvalue of the complexification $T_{\mathbf{C}}$.
See Exercise 33 in Section 3B for the definition of $T_{\mathbf{C}}$.

Solution. Recall the definitions. The complexification $V_{\mathbf{C}}$ of the real vector space V is the set $V \times V$, whose elements are written $u + iv$ with $u, v \in V$, with addition computed componentwise and with complex scalar multiplication given by

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

for all $a, b \in \mathbf{R}$ and all $u, v \in V$ (Exercise 8 in Section 1B). The complexification of $T \in \mathcal{L}(V)$ is the complex-linear map $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$ defined by

$$T_{\mathbf{C}}(u + iv) = Tu + iTv$$

for all $u, v \in V$ (Exercise 33 in Section 3B). Note that $u + iv$ is the zero vector of $V_{\mathbf{C}}$ exactly when $u = 0$ and $v = 0$, since $V_{\mathbf{C}}$ is $V \times V$ as a set.

A preliminary computation will be used in both directions. Since λ is real, $\lambda = \lambda + 0i$, so for all $u, v \in V$,

$$\lambda(u + iv) = (\lambda + 0i)(u + iv) = (\lambda u - 0 \cdot v) + i(\lambda v + 0 \cdot u) = \lambda u + i\lambda v.$$

First suppose λ is an eigenvalue of T. Then there exists $v \in V$ with $v \neq 0$ and $Tv = \lambda v$. Consider the vector $v + i0 \in V_{\mathbf{C}}$. It is nonzero, because its first component v is nonzero. Furthermore

$$T_{\mathbf{C}}(v + i0) = Tv + iT0 = \lambda v + i0 = \lambda(v + i0),$$

where the last equality uses the preliminary computation with $u = v$ and the second vector equal to 0. Thus $v + i0$ is an eigenvector of $T_{\mathbf{C}}$ corresponding to λ, so λ is an eigenvalue of $T_{\mathbf{C}}$.

Conversely, suppose λ is an eigenvalue of $T_{\mathbf{C}}$. Then there exist $u, v \in V$, not both 0, such that

$$T_{\mathbf{C}}(u + iv) = \lambda(u + iv).$$

The left side equals $Tu + iTv$ and, by the preliminary computation, the right side equals $\lambda u + i\lambda v$. Two elements of $V_{\mathbf{C}} = V \times V$ are equal exactly when their two components agree, so

$$Tu = \lambda u \qquad \text{and} \qquad Tv = \lambda v.$$

Since u and v are not both 0, at least one of them is a nonzero vector in V satisfying the eigenvector equation for T with the scalar λ. Hence λ is an eigenvalue of T.

This completes the proof.

Remark. A slicker route uses Exercise 33(b) in Section 3B, which says that $S_{\mathbf{C}}$ is injective if and only if S is injective. Since

$$(T - \lambda I)_{\mathbf{C}}(u + iv) = (T - \lambda I)u + i(T - \lambda I)v = (Tu + iTv) - \lambda(u + iv) = (T_{\mathbf{C}} - \lambda I)(u + iv)$$

for all $u, v \in V$ (using again that λ is real), we have $(T - \lambda I)_{\mathbf{C}} = T_{\mathbf{C}} - \lambda I$. Because a scalar μ is an eigenvalue of an operator S precisely when $S - \mu I$ fails to be injective (a restatement of the definition, valid without any finite-dimensionality hypothesis), we get

$$\begin{aligned} \lambda \text{ is an eigenvalue of } T &\iff T - \lambda I \text{ is not injective} \\ &\iff T_{\mathbf{C}} - \lambda I \text{ is not injective} \\ &\iff \lambda \text{ is an eigenvalue of } T_{\mathbf{C}}. \end{aligned}$$

Note that no finite-dimensionality assumption on V is needed anywhere.

Problem 5A.18. Suppose $\mathbf{F} = \mathbf{R}$, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{C}$. Prove that λ is an eigenvalue of the complexification $T_{\mathbf{C}}$ if and only if $\overline{\lambda}$ is an eigenvalue of $T_{\mathbf{C}}$.

Solution. As in Exercise 17, $V_{\mathbf{C}} = V \times V$ with elements written $u + iv$ for $u, v \in V$, complex scalar

multiplication given by

$$(a+bi)(u+iv) = (au-bv) + i(av+bu) \qquad (a,b \in \mathbf{R},\ u,v \in V),$$

and $T_{\mathbf{C}}(u+iv) = Tu + iTv$.
Define the conjugation map $\sigma\colon V_{\mathbf{C}} \to V_{\mathbf{C}}$ by

$$\sigma(u+iv) = u - iv = u + i(-v) \qquad \text{for all } u,v \in V.$$

This is a well-defined function on $V_{\mathbf{C}} = V \times V$, namely $(u,v) \mapsto (u,-v)$. We record four properties of σ.
First, σ is additive: for $u_1, v_1, u_2, v_2 \in V$,

$$\sigma\big((u_1+iv_1)+(u_2+iv_2)\big) = \sigma\big((u_1+u_2)+i(v_1+v_2)\big) = (u_1+u_2) + i\big(-(v_1+v_2)\big),$$

which equals $\sigma(u_1+iv_1) + \sigma(u_2+iv_2)$.
Second, σ is conjugate-homogeneous: if $\mu = a+bi$ with $a,b \in \mathbf{R}$ and $u,v \in V$, then

$$\sigma\big(\mu(u+iv)\big) = \sigma\big((au-bv)+i(av+bu)\big) = (au-bv) + i\big(-av-bu\big),$$

while

$$\overline{\mu}\,\sigma(u+iv) = (a-bi)\big(u+i(-v)\big) = \big(au-(-b)(-v)\big) + i\big(a(-v)+(-b)u\big) = (au-bv)+i(-av-bu).$$

The two right-hand sides agree, so $\sigma(\mu w) = \overline{\mu}\,\sigma(w)$ for all $\mu \in \mathbf{C}$ and $w \in V_{\mathbf{C}}$.
Third, σ is a bijection of $V_{\mathbf{C}}$ onto itself with $\sigma\circ\sigma$ equal to the identity, since $\sigma(\sigma(u+iv)) = \sigma(u+i(-v)) = u+iv$. In particular $\sigma(w) = 0$ if and only if $w = 0$.
Fourth, σ commutes with $T_{\mathbf{C}}$: for all $u,v \in V$,

$$\sigma\big(T_{\mathbf{C}}(u+iv)\big) = \sigma(Tu+iTv) = Tu + i(-Tv) = Tu + i\,T(-v) = T_{\mathbf{C}}\big(u+i(-v)\big) = T_{\mathbf{C}}\big(\sigma(u+iv)\big),$$

where we used the linearity of T in the form $T(-v) = -Tv$. Thus $\sigma \circ T_{\mathbf{C}} = T_{\mathbf{C}} \circ \sigma$.
Now suppose $\lambda \in \mathbf{C}$ is an eigenvalue of $T_{\mathbf{C}}$. Choose $w \in V_{\mathbf{C}}$ with $w \neq 0$ and $T_{\mathbf{C}}w = \lambda w$. Then $\sigma(w) \neq 0$ by the third property, and

$$T_{\mathbf{C}}\big(\sigma(w)\big) = \sigma\big(T_{\mathbf{C}}w\big) = \sigma(\lambda w) = \overline{\lambda}\,\sigma(w),$$

using the fourth property and then the second. Hence $\sigma(w)$ is an eigenvector of $T_{\mathbf{C}}$ corresponding to $\overline{\lambda}$, so $\overline{\lambda}$ is an eigenvalue of $T_{\mathbf{C}}$.
For the other direction, suppose $\overline{\lambda}$ is an eigenvalue of $T_{\mathbf{C}}$. Applying what we just proved with $\overline{\lambda}$ in place of λ shows that $\overline{\overline{\lambda}}$ is an eigenvalue of $T_{\mathbf{C}}$; since $\overline{\overline{\lambda}} = \lambda$, this says λ is an eigenvalue of $T_{\mathbf{C}}$.
Therefore λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if $\overline{\lambda}$ is an eigenvalue of $T_{\mathbf{C}}$.
Concretely, unwinding the abstraction: if $\lambda = a+bi$ and $T_{\mathbf{C}}(u+iv) = \lambda(u+iv)$ with $(u,v) \neq (0,0)$, then comparing components gives $Tu = au - bv$ and $Tv = av + bu$, and one checks directly that these same two equations say exactly that $T_{\mathbf{C}}(u-iv) = \overline{\lambda}(u-iv)$, with $u - iv \neq 0$.
Note that the argument uses in an essential way that T is the complexification of a real operator: this is what makes $T_{\mathbf{C}}$ commute with σ. The statement is false for a general operator on a complex vector space, for example for the operator on $\mathbf{C}$ given by multiplication by i, whose only eigenvalue is i.

Problem 5A.19. Show that the forward shift operator $T \in \mathcal{L}(\mathbf{F}^\infty)$ defined by

$$T(z_1, z_2, \dots) = (0, z_1, z_2, \dots)$$

has no eigenvalues.

Solution. Suppose $\lambda \in \mathbf{F}$ and $z = (z_1, z_2, z_3, \dots) \in \mathbf{F}^\infty$ satisfy $Tz = \lambda z$. We show $z = 0$; since an eigenvector must be nonzero, this proves that no $\lambda \in \mathbf{F}$ is an eigenvalue of T.
The equation $Tz = \lambda z$ reads

$$(0, z_1, z_2, z_3, \dots) = (\lambda z_1, \lambda z_2, \lambda z_3, \dots).$$

Two elements of $\mathbf{F}^\infty$ are equal exactly when all their coordinates agree, so comparing the first coordinate and then the $(k+1)$-st coordinate for each $k \geq 1$ gives

$$\lambda z_1 = 0 \qquad \text{and} \qquad z_k = \lambda z_{k+1} \quad \text{for every positive integer } k.$$

Case 1: $\lambda = 0$. Then the second family of equations reads $z_k = 0 \cdot z_{k+1} = 0$ for every positive integer k. Hence $z = 0$.
Case 2: $\lambda \neq 0$. From $\lambda z_1 = 0$ and $\lambda \neq 0$ we get $z_1 = 0$. Now the equations $z_k = \lambda z_{k+1}$ can be rewritten as

$$z_{k+1} = \frac{z_k}{\lambda} \qquad \text{for every positive integer } k.$$

We prove by induction on k that $z_k = 0$ for every positive integer k. The base case $z_1 = 0$ has just been established. If $z_k = 0$ for some $k \geq 1$, then $z_{k+1} = z_k/\lambda = 0$. Hence $z_k = 0$ for all k, so $z = 0$.
In both cases $z = 0$. Thus for every $\lambda \in \mathbf{F}$ the only solution of $Tz = \lambda z$ is $z = 0$, that is, $\operatorname{null}(T - \lambda I) = \{0\}$. Therefore T has no eigenvalues.
This example shows that the equivalence of (b) and (c) in 5.7 genuinely requires finite-dimensionality. Here $T - 0I = T$ is injective, so 0 is not an eigenvalue of T, yet T is not surjective: no vector in the range of T has a nonzero first coordinate. Compare Exercise 20, where the backward shift, which is surjective but not injective, has every element of $\mathbf{F}$ as an eigenvalue. (The equivalence of (a) and (b) in 5.7 does hold in every vector space, since $Tv = \lambda v$ is the same equation as $(T - \lambda I)v = 0$; it is (c) and (d) that decouple in infinite dimensions.)

Problem 5A.20. Define the backward shift operator $S \in \mathcal{L}(\mathbf{F}^\infty)$ by

$$S(z_1, z_2, z_3, \dots) = (z_2, z_3, \dots).$$

(a) Show that every element of $\mathbf{F}$ is an eigenvalue of S.
(b) Find all eigenvectors of S.

Solution. (a) Let $\lambda \in \mathbf{F}$ be arbitrary. Define

$$w_\lambda = (1, \lambda, \lambda^2, \lambda^3, \dots) \in \mathbf{F}^\infty,$$

that is, the sequence whose k-th coordinate is λ^{k-1}, with the convention $\lambda^0 = 1$ (so $w_0 = (1, 0, 0, \dots)$). Then $w_\lambda \neq 0$, because its first coordinate is 1.
Applying S deletes the first coordinate:

$$S w_\lambda = (\lambda, \lambda^2, \lambda^3, \dots).$$

The k-th coordinate of $S w_\lambda$ is λ^k, which equals $\lambda \cdot \lambda^{k-1}$, the k-th coordinate of λw_λ. Hence

$$S w_\lambda = \lambda w_\lambda,$$

so w_λ is an eigenvector of S corresponding to λ. Thus every $\lambda \in \mathbf{F}$ is an eigenvalue of S. (For $\lambda = 0$ this reads $S(1, 0, 0, \dots) = (0, 0, \dots) = 0 \cdot w_0$, which is correct.)
(b) We claim that the eigenvectors of S are exactly the vectors

$$c\,(1, \lambda, \lambda^2, \lambda^3, \dots) = c\, w_\lambda \qquad \text{with } \lambda \in \mathbf{F} \text{ and } c \in \mathbf{F},\ c \neq 0,$$

and that $c\, w_\lambda$ is an eigenvector corresponding to the eigenvalue λ.

Each such vector is indeed an eigenvector: it is nonzero because its first coordinate is $c \neq 0$, and $S(cw_\lambda) = c\,Sw_\lambda = c\lambda w_\lambda = \lambda(cw_\lambda)$ by part (a) and the linearity of S.
Conversely, suppose $z = (z_1, z_2, z_3, \dots)$ is an eigenvector of S, say $Sz = \lambda z$ with $z \neq 0$ and $\lambda \in \mathbf{F}$. Written out, the equation $Sz = \lambda z$ is

$$(z_2, z_3, z_4, \dots) = (\lambda z_1, \lambda z_2, \lambda z_3, \dots),$$

and comparing the k-th coordinates gives

$$z_{k+1} = \lambda z_k \qquad \text{for every positive integer } k.$$

By induction on k this yields

$$z_k = \lambda^{k-1} z_1 \qquad \text{for every positive integer } k :$$

the case $k = 1$ is the trivial identity $z_1 = \lambda^0 z_1$, and if $z_k = \lambda^{k-1} z_1$ then $z_{k+1} = \lambda z_k = \lambda^k z_1$.
If $z_1 = 0$, the formula forces $z_k = 0$ for every k, so $z = 0$, contradicting the assumption that z is an eigenvector. Hence $z_1 \neq 0$, and setting $c = z_1$ we have

$$z = (c, c\lambda, c\lambda^2, \dots) = c\,w_\lambda \qquad \text{with } c \neq 0,$$

as claimed.
Two consequences are worth recording. First, for each $\lambda \in \mathbf{F}$ the eigenspace of S corresponding to λ is

$$\operatorname{null}(S - \lambda I) = \{c\,w_\lambda : c \in \mathbf{F}\} = \operatorname{span}(w_\lambda),$$

a one-dimensional subspace of $\mathbf{F}^\infty$. Second, distinct scalars give distinct eigenspaces: if $\lambda \neq \mu$ then w_λ and w_μ differ in their second coordinate, and since both have first coordinate 1 neither is a scalar multiple of the other. So the eigenvectors of S form exactly the union over $\lambda \in \mathbf{F}$ of the punctured lines $\operatorname{span}(w_\lambda) \setminus \{0\}$, and these lines are pairwise distinct.
Note that S has infinitely many distinct eigenvalues when $\mathbf{F}$ is infinite, which is consistent with 5.12 only because $\mathbf{F}^\infty$ is infinite-dimensional.

Problem 5A.21. Suppose $T \in \mathcal{L}(V)$ is invertible.
(a) Suppose $\lambda \in \mathbf{F}$ with $\lambda \neq 0$. Prove that λ is an eigenvalue of T if and only if $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}.
(b) Prove that T and T^{-1} have the same eigenvectors.

Solution. We first record a computation that does all the work. Suppose $\lambda \in \mathbf{F}$ with $\lambda \neq 0$ and suppose $v \in V$ satisfies $Tv = \lambda v$. Applying T^{-1} to both sides and using the linearity of T^{-1} together with $T^{-1}T = I$, we obtain

$$v = T^{-1}(Tv) = T^{-1}(\lambda v) = \lambda\,T^{-1}v.$$

Since $\lambda \neq 0$, we may divide by λ to conclude

$$T^{-1}v = \frac{1}{\lambda}\,v.$$

Thus every solution of $Tv = \lambda v$ is also a solution of $T^{-1}v = \frac{1}{\lambda}v$. Note that the vector v itself is carried over unchanged, so a nonzero such v is simultaneously an eigenvector of T corresponding to λ and an eigenvector of T^{-1} corresponding to $\frac{1}{\lambda}$.
(a) Suppose λ is an eigenvalue of T, with $\lambda \neq 0$. Choose an eigenvector v, so $v \neq 0$ and $Tv = \lambda v$. By the computation above, $T^{-1}v = \frac{1}{\lambda}v$ with $v \neq 0$, so $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}.
Conversely, suppose $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}. The operator T^{-1} is itself invertible, with $(T^{-1})^{-1} = T$, and the scalar $\frac{1}{\lambda}$ is nonzero. So the computation above, applied with T^{-1} in place of T and $\frac{1}{\lambda}$ in place

of λ, shows that any eigenvector v of T^{-1} corresponding to $\frac{1}{\lambda}$ satisfies

$$(T^{-1})^{-1}v = \frac{1}{1/\lambda}\,v, \qquad \text{that is,} \qquad Tv = \lambda v,$$

with $v \neq 0$. Hence λ is an eigenvalue of T.
This proves that for $\lambda \neq 0$, the scalar λ is an eigenvalue of T if and only if $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}. The hypothesis $\lambda \neq 0$ is needed only so that $\frac{1}{\lambda}$ makes sense; it excludes nothing, because 0 is never an eigenvalue of an invertible operator. Indeed, if T is invertible then T is injective, so $\operatorname{null} T = \{0\}$, so there is no nonzero v with $Tv = 0 \cdot v$.
(b) Recall that a vector $v \in V$ is called an eigenvector of an operator S when $v \neq 0$ and $Sv = \mu v$ for some eigenvalue μ of S; equivalently, $v \neq 0$ and Sv is a scalar multiple of v.
Suppose v is an eigenvector of T, say $v \neq 0$ and $Tv = \lambda v$ for some $\lambda \in \mathbf{F}$. Then $\lambda \neq 0$: otherwise $Tv = 0$ with $v \neq 0$, contradicting the injectivity of the invertible operator T. So the computation at the start applies and gives

$$T^{-1}v = \frac{1}{\lambda}\,v$$

with $v \neq 0$. Hence v is an eigenvector of T^{-1} (corresponding to the eigenvalue $\frac{1}{\lambda}$).
Conversely, suppose v is an eigenvector of T^{-1}. Applying the paragraph above to the invertible operator T^{-1}, whose inverse is T, shows that v is an eigenvector of $(T^{-1})^{-1} = T$.
Therefore the set of eigenvectors of T equals the set of eigenvectors of T^{-1}. More precisely, v is an eigenvector of T corresponding to λ if and only if v is an eigenvector of T^{-1} corresponding to $\frac{1}{\lambda}$; equivalently, for every nonzero $\lambda \in \mathbf{F}$,

$$\operatorname{null}(T - \lambda I) = \operatorname{null}\Big(T^{-1} - \frac{1}{\lambda}I\Big).$$

So T and T^{-1} have the same eigenvectors, though the corresponding eigenvalues are inverted. Note that no finite-dimensionality assumption on V was used.

Problem 5A.22. Suppose $T \in \mathcal{L}(V)$ and there exist nonzero vectors u and w in V such that

$$Tu = 3w \quad \text{and} \quad Tw = 3u.$$

Prove that 3 or -3 is an eigenvalue of T.

Solution. Consider the two vectors $u + w$ and $u - w$. Applying T and using the two hypotheses gives

$$T(u + w) = Tu + Tw = 3w + 3u = 3(u + w)$$

and

$$T(u - w) = Tu - Tw = 3w - 3u = -3(u - w).$$

Now observe that $u+w$ and $u-w$ cannot both be the zero vector: their sum is $(u+w)+(u-w) = 2u$, which is nonzero because $u \neq 0$ and $2 \neq 0$ in $\mathbf{F}$ (here $\mathbf{F}$ is $\mathbb{R}$ or $\mathbb{C}$).
If $u + w \neq 0$, then the first displayed equation exhibits a nonzero vector sent by T to 3 times itself, so by the definition of eigenvalue (5.5) the scalar 3 is an eigenvalue of T, with $u + w$ an eigenvector corresponding to it.
Otherwise $u + w = 0$, and then $w = -u$, so $u - w = 2u \neq 0$. The second displayed equation then exhibits the nonzero vector $u - w$ with $T(u - w) = -3(u - w)$, so by 5.5 the scalar -3 is an eigenvalue of T.
In either case, 3 or -3 is an eigenvalue of T, as desired.

Problem 5A.23. Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that ST and TS have the same eigenvalues.

Solution. By symmetry (interchanging the roles of S and T), it suffices to prove that every eigenvalue of ST is an eigenvalue of TS. So suppose $\lambda \in \mathbf{F}$ is an eigenvalue of ST. We consider two cases.
Case 1: $\lambda \neq 0$. Because λ is an eigenvalue of ST, there exists $v \in V$ with $v \neq 0$ and $(ST)v = \lambda v$. Let $w = Tv$.
First we check $w \neq 0$. If $Tv = 0$, then $\lambda v = S(Tv) = S0 = 0$; since $\lambda \neq 0$ this forces $v = 0$, contradicting $v \neq 0$. Hence $w = Tv \neq 0$.
Now compute, using the linearity of T:

$$(TS)w = (TS)(Tv) = T\big(S(Tv)\big) = T\big((ST)v\big) = T(\lambda v) = \lambda\, Tv = \lambda w.$$

Thus w is a nonzero vector with $(TS)w = \lambda w$, so w is an eigenvector of TS corresponding to λ, and therefore λ is an eigenvalue of TS. (Note that finite-dimensionality was not needed in this case.)
Case 2: $\lambda = 0$. Because 0 is an eigenvalue of ST and V is finite-dimensional, the operator $ST - 0 \cdot I = ST$ is not invertible, by 5.7.
We claim that consequently S is not invertible or T is not invertible. Indeed, if both S and T were invertible, then ST would be invertible (with inverse $T^{-1}S^{-1}$), contradicting the previous paragraph.
Now suppose, for contradiction, that TS is invertible. Then S is injective: if $Sv = 0$, then $(TS)v = T0 = 0$, and injectivity of the invertible operator TS gives $v = 0$. Because V is finite-dimensional, an injective operator on V is invertible (by 3.65), so S is invertible. Then $T = (TS)S^{-1}$ is a composition of invertible operators, hence invertible. So both S and T are invertible, contradicting the claim of the previous paragraph.
Therefore TS is not invertible, and so $TS - 0 \cdot I$ is not invertible; by 5.7 this means 0 is an eigenvalue of TS.
In both cases λ is an eigenvalue of TS. By the symmetry noted at the start, ST and TS have exactly the same eigenvalues.

Problem 5A.24. Suppose A is an n-by-n matrix with entries in $\mathbf{F}$. Define $T \in \mathcal{L}(\mathbf{F}^n)$ by $Tx = Ax$, where elements of $\mathbf{F}^n$ are thought of as n-by-1 column vectors.
(a) Suppose the sum of the entries in each row of A equals 1. Prove that 1 is an eigenvalue of T.
(b) Suppose the sum of the entries in each column of A equals 1. Prove that 1 is an eigenvalue of T.

Solution. Write $A_{j,k}$ for the entry of A in row j, column k. Recall that for a column vector $x \in \mathbf{F}^n$, the j^{th} entry of Ax is $\sum_{k=1}^{n} A_{j,k} x_k$.
(a) Let x be the column vector all of whose entries equal 1; thus $x \neq 0$. For each $j \in \{1, \dots, n\}$, the j^{th} entry of Ax equals

$$\sum_{k=1}^{n} A_{j,k} \cdot 1 = \sum_{k=1}^{n} A_{j,k} = 1,$$

where the last equality is the hypothesis that the entries in row j of A sum to 1. Hence every entry of Ax equals 1, which means $Tx = Ax = x = 1 \cdot x$. Because $x \neq 0$, the definition of eigenvalue (5.5) gives that 1 is an eigenvalue of T, with x an eigenvector corresponding to it.
(b) Now assume $\sum_{j=1}^{n} A_{j,k} = 1$ for each $k \in \{1, \dots, n\}$. Let $B = A - I_n$, where I_n is the n-by-n identity matrix, so that $Bx = Tx - x = (T - I)x$ for all $x \in \mathbf{F}^n$. Each column of B sums to 0: for each k,

$$\sum_{j=1}^{n} B_{j,k} = \Big(\sum_{j=1}^{n} A_{j,k}\Big) - 1 = 1 - 1 = 0,$$

since subtracting I_n subtracts exactly 1 from the single diagonal entry of column k.
Let

$$U = \Big\{ y \in \mathbf{F}^n : \sum_{j=1}^{n} y_j = 0 \Big\},$$

which is a subspace of $\mathbf{F}^n$, being the null space of the linear functional $y \mapsto y_1 + \cdots + y_n$ on $\mathbf{F}^n$. This functional is surjective onto $\mathbf{F}$ (it sends $(c, 0, \ldots, 0)$ to c), so by the fundamental theorem of linear maps (3.21) we get $\dim U = n - 1$. In particular $U \neq \mathbf{F}^n$.
We claim $\operatorname{range}(T - I) \subseteq U$. Indeed, for any $x \in \mathbf{F}^n$, interchanging the order of summation gives

$$\sum_{j=1}^{n} \big(Bx\big)_j = \sum_{j=1}^{n} \sum_{k=1}^{n} B_{j,k} x_k = \sum_{k=1}^{n} x_k \Big(\sum_{j=1}^{n} B_{j,k} \Big) = \sum_{k=1}^{n} x_k \cdot 0 = 0,$$

so $(T - I)x = Bx \in U$, proving the claim.
Hence $\operatorname{range}(T - I) \subseteq U \subsetneq \mathbf{F}^n$, so $T - I$ is not surjective. Because $\mathbf{F}^n$ is finite-dimensional, 5.7 (the equivalence of conditions (a) and (c) there, with $\lambda = 1$) implies that 1 is an eigenvalue of T.

Problem 5A.25. Suppose $T \in \mathcal{L}(V)$ and u, w are eigenvectors of T such that $u + w$ is also an eigenvector of T. Prove that u and w are eigenvectors of T corresponding to the same eigenvalue.

Solution. Let $\alpha, \beta, \lambda \in \mathbf{F}$ be the eigenvalues to which u, w, and $u + w$ correspond; thus $u, w, u + w$ are all nonzero and

$$Tu = \alpha u, \qquad Tw = \beta w, \qquad T(u + w) = \lambda(u + w).$$

Our goal is to show $\alpha = \beta$.
By linearity of T,

$$\lambda u + \lambda w = T(u + w) = Tu + Tw = \alpha u + \beta w,$$

and therefore

$$(\alpha - \lambda)u + (\beta - \lambda)w = 0.$$

Suppose, for contradiction, that $\alpha \neq \beta$. Then u and w are eigenvectors of T corresponding to the distinct eigenvalues α and β, so the list u, w is linearly independent by 5.11. Linear independence applied to the displayed equation above forces

$$\alpha - \lambda = 0 \quad \text{and} \quad \beta - \lambda = 0,$$

so $\alpha = \lambda = \beta$. This contradicts the assumption $\alpha \neq \beta$.
Hence $\alpha = \beta$, which says exactly that u and w are eigenvectors of T corresponding to the same eigenvalue. (Moreover, feeding $\alpha = \beta$ back into the equation $(\alpha - \lambda)u + (\beta - \lambda)w = 0$ gives $(\alpha - \lambda)(u + w) = 0$; since $u + w \neq 0$ we get $\lambda = \alpha = \beta$, so all three eigenvectors correspond to the same eigenvalue.)

Problem 5A.26. Suppose $T \in \mathcal{L}(V)$ is such that every nonzero vector in V is an eigenvector of T. Prove that T is a scalar multiple of the identity operator.

Solution. If $V = \{0\}$, then $\mathcal{L}(V)$ contains only the zero operator, which equals $0 \cdot I$, and we are done. So assume $V \neq \{0\}$.
By hypothesis, for each nonzero $v \in V$ there exists a scalar $a_v \in \mathbf{F}$ such that

$$Tv = a_v v.$$

Note that a_v is uniquely determined by v: if $av = a'v$ with $v \neq 0$, then $(a - a')v = 0$, forcing $a = a'$.
We claim that $a_u = a_w$ for all nonzero $u, w \in V$. Fix nonzero $u, w \in V$ and consider two cases.
Case 1: the list u, w is linearly dependent. Then there exist scalars a, b, not both 0, with $au + bw = 0$. If $b = 0$ then $au = 0$ with $a \neq 0$, forcing $u = 0$, a contradiction; so $b \neq 0$ and $w = -(a/b)u$. Setting $c = -a/b$ we have $w = cu$, and $c \neq 0$ because $w \neq 0$. Then

$$a_w w = Tw = T(cu) = cTu = c\, a_u u = a_u(cu) = a_u w,$$

and since $w \neq 0$ the uniqueness noted above gives $a_w = a_u$.

Case 2: the list u, w is linearly independent. Then $u + w \neq 0$, so $u + w$ is also an eigenvector of T, with $T(u + w) = a_{u+w}(u + w)$. On the other hand, by linearity

$$T(u + w) = Tu + Tw = a_u u + a_w w.$$

Subtracting the two expressions for $T(u + w)$ yields

$$(a_{u+w} - a_u)u + (a_{u+w} - a_w)w = 0.$$

Because u, w is linearly independent, both coefficients vanish, so $a_u = a_{u+w} = a_w$.
In both cases $a_u = a_w$, proving the claim. Hence there is a single scalar $\lambda \in \mathbf{F}$ with $a_v = \lambda$ for every nonzero $v \in V$ (take $\lambda = a_{v_0}$ for any fixed nonzero $v_0 \in V$, which exists because $V \neq \{0\}$).
Therefore $Tv = \lambda v$ for every nonzero $v \in V$, and also $T0 = 0 = \lambda \cdot 0$. Thus $Tv = \lambda v = (\lambda I)v$ for all $v \in V$, which means $T = \lambda I$. So T is a scalar multiple of the identity operator.

Problem 5A.27. Suppose that V is finite-dimensional and $k \in \{1, \dots, \dim V - 1\}$. Suppose $T \in \mathcal{L}(V)$ is such that every subspace of V of dimension k is invariant under T. Prove that T is a scalar multiple of the identity operator.

Solution. Let $n = \dim V$. The hypothesis that the set $\{1, \dots, n - 1\}$ contains k forces $n \geq 2$.
The strategy is to show that every nonzero vector of V is an eigenvector of T, and then to apply Exercise 26 in this section.
So let $v \in V$ with $v \neq 0$, and suppose for contradiction that $Tv \notin \operatorname{span}(v)$. Then the list v, Tv is linearly independent: if $av + bTv = 0$ with $b \neq 0$ we would get $Tv = -(a/b)v \in \operatorname{span}(v)$, contrary to assumption; hence $b = 0$, and then $av = 0$ with $v \neq 0$ gives $a = 0$.
By 2.32 we may extend v, Tv to a basis

$$v,\ Tv,\ w_3,\ \dots,\ w_n$$

of V. Now set

$$U = \operatorname{span}(v, w_3, \dots, w_{k+1}),$$

where the list $w_3, \dots, w_{k+1}$ is empty if $k = 1$. This makes sense because $k + 1 \leq n$ (as $k \leq n - 1$), so the vectors $w_3, \dots, w_{k+1}$ are among the basis vectors listed above. The list $v, w_3, \dots, w_{k+1}$ has $1 + (k - 1) = k$ vectors and is a sublist of a basis, hence is linearly independent; therefore it is a basis of U and $\dim U = k$.
By hypothesis U is invariant under T. Since $v \in U$, invariance gives $Tv \in U$. So Tv can be written as a linear combination

$$Tv = c_1 v + c_3 w_3 + \cdots + c_{k+1} w_{k+1}$$

for some scalars $c_1, c_3, \dots, c_{k+1}$. But Tv is itself one of the vectors in the basis $v, Tv, w_3, \dots, w_n$, so it also has the representation $Tv = 0 \cdot v + 1 \cdot Tv + 0 \cdot w_3 + \cdots + 0 \cdot w_n$ in that basis. The displayed equation gives a second representation of Tv in that same basis, one in which the coefficient of Tv is 0. Because representations in a basis are unique (see 2.28), we obtain $1 = 0$, a contradiction.
Hence $Tv \in \operatorname{span}(v)$ for every nonzero $v \in V$; that is, for each nonzero v there is a scalar a_v with $Tv = a_v v$. By the remark following 5.8, a nonzero vector v is an eigenvector of T if and only if Tv is a scalar multiple of v, so every nonzero vector of V is an eigenvector of T.
By Exercise 26 in this section (solved above), T is a scalar multiple of the identity operator. For completeness, here is the argument again in the present setting. Because $n \geq 2$, the space V has two linearly independent vectors, and for any linearly independent u, w we have

$$a_{u+w} u + a_{u+w} w = T(u + w) = Tu + Tw = a_u u + a_w w,$$

so $(a_{u+w} - a_u)u + (a_{u+w} - a_w)w = 0$ and linear independence gives $a_u = a_{u+w} = a_w$. If instead u, w are both nonzero and linearly dependent, then $w = cu$ for some $c \neq 0$, and $a_w w = Tw = cTu = a_u(cu) = a_u w$, so $a_u = a_w$ because $w \neq 0$. Thus a_v takes a single value λ over all nonzero $v \in V$, and since $T0 = 0 = \lambda \cdot 0$ we conclude $Tv = \lambda v$ for all $v \in V$, that is, $T = \lambda I$.

Problem 5A.28. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T has at most $1 + \dim \operatorname{range} T$ distinct eigenvalues.

Solution. If T has no eigenvalues, the result is immediate because $1 + \dim \operatorname{range} T \geq 1 > 0$. So suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T, with $m \geq 1$; we must show $m \leq 1 + \dim \operatorname{range} T$.
For each $j \in \{1, \ldots, m\}$ choose an eigenvector v_j corresponding to λ_j, so $v_j \neq 0$ and $Tv_j = \lambda_j v_j$. By 5.11, the list $v_1, \ldots, v_m$ is linearly independent.
At most one of the distinct scalars $\lambda_1, \ldots, \lambda_m$ equals 0. After relabeling, we may assume that $\lambda_1, \ldots, \lambda_{m-1}$ are all nonzero (if none of the eigenvalues is 0, this holds after any labeling; if one of them is 0, put it last).
For each $j \in \{1, \ldots, m-1\}$ we have $\lambda_j \neq 0$, so

$$v_j = \frac{1}{\lambda_j}\lambda_j v_j = \frac{1}{\lambda_j} Tv_j = T\Big(\frac{v_j}{\lambda_j}\Big) \in \operatorname{range} T.$$

Thus $v_1, \ldots, v_{m-1}$ is a linearly independent list (being a sublist of the linearly independent list $v_1, \ldots, v_m$) of vectors in the subspace $\operatorname{range} T$. Now $\operatorname{range} T$ is finite-dimensional, being a subspace of the finite-dimensional space V, so it has a basis, and that basis is a spanning list of $\operatorname{range} T$ of length $\dim \operatorname{range} T$. By 2.22 the length of our linearly independent list is at most the length of that spanning list, so

$$m - 1 \leq \dim \operatorname{range} T,$$

that is,

$$m \leq 1 + \dim \operatorname{range} T.$$

Since $\lambda_1, \ldots, \lambda_m$ was an arbitrary list of distinct eigenvalues of T, the operator T has at most $1 + \dim \operatorname{range} T$ distinct eigenvalues.

Problem 5A.29. Suppose $T \in \mathcal{L}(\mathbf{R}^3)$ and -4, 5, and $\sqrt{7}$ are eigenvalues of T. Prove that there exists $x \in \mathbf{R}^3$ such that $Tx - 9x = (-4, 5, \sqrt{7})$.

Solution. The vector $(-4, 5, \sqrt{7})$ plays no special role here; what must be shown is that $T - 9I$ is surjective.
The numbers -4, 5, and $\sqrt{7}$ are three distinct eigenvalues of T. Since $\dim \mathbf{R}^3 = 3$, the result 5.12 says that T has at most 3 distinct eigenvalues. Hence -4, 5, $\sqrt{7}$ are all the eigenvalues that T has.
(For completeness, here is the argument behind 5.12 in this case. Let v_1, v_2, v_3 be eigenvectors of T corresponding to -4, 5, $\sqrt{7}$ respectively. By 5.11, a list of eigenvectors corresponding to distinct eigenvalues is linearly independent, so v_1, v_2, v_3 is a linearly independent list of length $3 = \dim \mathbf{R}^3$ and therefore a basis of $\mathbf{R}^3$. If λ were a fourth, different, eigenvalue with eigenvector v_4, then v_1, v_2, v_3, v_4 would be a linearly independent list of length 4 in a 3-dimensional space, which is impossible.)
Because $9 \neq -4$, $9 \neq 5$, and $9 \neq \sqrt{7}$ (indeed $\sqrt{7} < 3 < 9$), the number 9 is not an eigenvalue of T. By 5.7, applied with $\lambda = 9$, the operator $T - 9I$ is therefore injective. Since $\mathbf{R}^3$ is finite-dimensional, 3.65 shows that the injective operator $T - 9I$ is also surjective.
Surjectivity of $T - 9I$ means precisely that every vector of $\mathbf{R}^3$ is in the range of $T - 9I$. In particular there exists $x \in \mathbf{R}^3$ with

$$(T - 9I)x = (-4, 5, \sqrt{7}),$$

that is, $Tx - 9x = (-4, 5, \sqrt{7})$, as desired.

Problem 5A.30. Suppose $T \in \mathcal{L}(V)$ and $(T - 2I)(T - 3I)(T - 4I) = 0$. Suppose λ is an eigenvalue of T. Prove that $\lambda = 2$ or $\lambda = 3$ or $\lambda = 4$.

Solution. Let λ be an eigenvalue of T. By the definition 5.5 of eigenvalue, there exists $v \in V$ with $v \neq 0$ and $Tv = \lambda v$.

The key observation is that v is an eigenvector for every operator of the form $T - \mu I$: for each $\mu \in \mathbf{F}$,

$$(T - \mu I)v = Tv - \mu v = \lambda v - \mu v = (\lambda - \mu)v.$$

Now apply the three factors to v, working from the right (the order in which they act on v), using the displayed identity three times:

$$\begin{aligned}(T - 4I)v &= (\lambda - 4)v,\\ (T - 3I)(T - 4I)v &= (\lambda - 4)(T - 3I)v = (\lambda - 4)(\lambda - 3)v,\\ (T - 2I)(T - 3I)(T - 4I)v &= (\lambda - 4)(\lambda - 3)(T - 2I)v = (\lambda - 4)(\lambda - 3)(\lambda - 2)v.\end{aligned}$$

(Each step uses linearity of the operator being applied, to pull the scalar out front.)
By hypothesis $(T - 2I)(T - 3I)(T - 4I)$ is the zero operator, so the left side of the last line equals 0. Hence

$$(\lambda - 2)(\lambda - 3)(\lambda - 4)\, v = 0.$$

Because $v \neq 0$, the scalar multiplying it must be 0; indeed if a scalar c satisfies $cv = 0$ with $v \neq 0$ and $c \neq 0$, then $v = c^{-1}(cv) = 0$, a contradiction. Thus

$$(\lambda - 2)(\lambda - 3)(\lambda - 4) = 0.$$

Since $\mathbf{F}$ (which is $\mathbf{R}$ or $\mathbf{C}$) has no zero divisors, one of the three factors is 0. Therefore $\lambda = 2$ or $\lambda = 3$ or $\lambda = 4$.
Note that no finite-dimensionality of V was needed. Note also that the converse fails: the hypothesis does not force each of 2, 3, 4 to be an eigenvalue (for instance $T = 2I$ satisfies $(T-2I)(T-3I)(T-4I) = 0$ and has only the eigenvalue 2).

Problem 5A.31. Give an example of $T \in \mathcal{L}(\mathbf{R}^2)$ such that $T^4 = -I$.

Solution. Take T to be counterclockwise rotation of the plane by 45 degrees, which is the operator on $\mathbf{R}^2$ defined by

$$T(x, y) = \left(\frac{x - y}{\sqrt{2}},\ \frac{x + y}{\sqrt{2}}\right).$$

This map is linear, being given by the matrix

$$\mathcal{M}(T) = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}$$

with respect to the standard basis.
The idea is geometric: rotating by 45 degrees four times rotates by 180 degrees, and rotation by 180 degrees is exactly the operator $-I$. Here is the verification by direct computation.
First compute T^2. For $(x, y) \in \mathbf{R}^2$,

$$\begin{aligned}T^2(x, y) &= T\left(\frac{x - y}{\sqrt{2}},\ \frac{x + y}{\sqrt{2}}\right)\\ &= \left(\frac{1}{\sqrt{2}}\left(\frac{x - y}{\sqrt{2}} - \frac{x + y}{\sqrt{2}}\right),\ \frac{1}{\sqrt{2}}\left(\frac{x - y}{\sqrt{2}} + \frac{x + y}{\sqrt{2}}\right)\right)\\ &= \left(\frac{-2y}{2},\ \frac{2x}{2}\right)\\ &= (-y,\, x).\end{aligned}$$

So T^2 is rotation by 90 degrees. Applying this twice,

$$T^4(x, y) = T^2\big(T^2(x, y)\big) = T^2(-y, x) = (-x, -y) = -(x, y).$$

Hence $T^4 = -I$, as required.

Remark. Any operator T on a real vector space with $T^4 = -I$ must have no eigenvalues at all: if $Tv = \lambda v$ with $v \neq 0$, then applying T four times gives $-v = T^4 v = \lambda^4 v$, so $(\lambda^4 + 1)v = 0$ and hence $\lambda^4 = -1$, which is impossible for $\lambda \in \mathbf{R}$. In particular the operator T above has no eigenvalues, in the same spirit as the rotation by 90 degrees in 5.9.

Problem 5A.32. Suppose $T \in \mathcal{L}(V)$ has no eigenvalues and $T^4 = I$. Prove that $T^2 = -I$.

Solution. Because T has no eigenvalues, in particular neither 1 nor -1 is an eigenvalue of T. By the definition 5.5 of eigenvalue, this says:

$$\text{if } u \in V \text{ and } Tu = u, \text{ then } u = 0; \qquad \text{if } w \in V \text{ and } Tw = -w, \text{ then } w = 0.$$

Call these two statements the no-eigenvalue conditions.
Since polynomials in T multiply as polynomials do (5.17), and $z^4 - 1 = (z^2 - 1)(z^2 + 1) = (z - 1)(z + 1)(z^2 + 1)$, the hypothesis $T^4 = I$ gives

$$(T - I)(T + I)(T^2 + I) = T^4 - I = 0.$$

(One can also check this directly by expanding: $(T-I)(T+I) = T^2-I$, and $(T^2-I)(T^2+I) = T^4-I$.)
We must show $T^2 + I = 0$. Suppose not. Then there exists $v \in V$ with

$$w := (T^2 + I)v \neq 0.$$

Applying the displayed identity to v gives

$$(T - I)(T + I)w = (T - I)(T + I)(T^2 + I)v = 0.$$

Now set $u := (T + I)w$. Two cases:
If $u = 0$, then $(T + I)w = 0$, i.e. $Tw = -w$ with $w \neq 0$. Thus -1 is an eigenvalue of T, contradicting the no-eigenvalue conditions.
If $u \neq 0$, then from $(T-I)u = (T-I)(T+I)w = 0$ we get $Tu = u$ with $u \neq 0$. Thus 1 is an eigenvalue of T, again contradicting the no-eigenvalue conditions.
Both cases are impossible, so our supposition was false. Hence $(T^2 + I)v = 0$ for every $v \in V$, that is, $T^2 + I = 0$, which says

$$T^2 = -I.$$

Remark. The hypothesis is not vacuous: the rotation by 45 degrees T of Exercise 5A.31 satisfies $T^4 = -I$, while its square, rotation by 90 degrees, satisfies $S(x, y) = (-y, x)$, $S^4 = I$, $S^2 = -I$, and S has no eigenvalues on $\mathbf{R}^2$. Note also that no finite-dimensionality is assumed, and indeed over $\mathbf{C}$ with V finite-dimensional and nonzero the hypothesis could not hold, since every operator on such a space has an eigenvalue.

Problem 5A.33. Suppose $T \in \mathcal{L}(V)$ and m is a positive integer.
(a) Prove that T is injective if and only if T^m is injective.
(b) Prove that T is surjective if and only if T^m is surjective.

Solution. Note that V is not assumed to be finite-dimensional, so the arguments below must not use 3.65; they are direct. Throughout, T^m means the m-fold composition $T \cdots T$ (5.13), and we use $T^m = T^{m-1}T = TT^{m-1}$ for $m \geq 2$.
(a) We use 3.15: a linear map is injective if and only if its null space equals $\{0\}$.
First suppose T is injective. We show by induction on m that T^m is injective. The case $m = 1$ is the hypothesis. Suppose $m \geq 2$ and T^{m-1} is injective. If $v \in V$ satisfies $T^m v = 0$, then

$$T^{m-1}(Tv) = T^m v = 0,$$

so $Tv = 0$ by injectivity of T^{m-1}, and hence $v = 0$ by injectivity of T. Thus $\operatorname{null} T^m = \{0\}$ and T^m is

injective, completing the induction.
Conversely suppose T^m is injective. Let $v \in V$ with $Tv = 0$. Then

$$T^m v = T^{m-1}(Tv) = T^{m-1}0 = 0$$

(when $m = 1$ this reads $T^1 v = Tv = 0$ directly). Injectivity of T^m gives $v = 0$. Hence $\operatorname{null} T = \{0\}$ and T is injective.
(b) Recall that a linear map $S \in \mathcal{L}(V)$ is surjective exactly when $\operatorname{range} S = V$, where $\operatorname{range} S = S(V) = \{Sv : v \in V\}$.
First suppose T is surjective, i.e. $T(V) = V$. We show by induction on m that $T^m(V) = V$. The case $m = 1$ is the hypothesis. Suppose $m \geq 2$ and $T^{m-1}(V) = V$. Then

$$T^m(V) = T^{m-1}\big(T(V)\big) = T^{m-1}(V) = V.$$

So T^m is surjective.
Conversely suppose T^m is surjective, i.e. $T^m(V) = V$. Since $T^m = TT^{m-1}$,

$$V = T^m(V) = T\big(T^{m-1}(V)\big) \subseteq T(V) \subseteq V,$$

where the first inclusion holds because $T^{m-1}(V) \subseteq V$ and T preserves inclusions of sets. (For $m = 1$ the chain is trivial.) Hence $T(V) = V$, so T is surjective.
Remark. Both statements are genuinely about operators, i.e. maps from V to itself, since only then are the powers T^m defined. If V happens to be finite-dimensional, then (a) and (b) are equivalent to each other by 3.65 and each is equivalent to invertibility of T; but the proofs above need none of that.

Problem 5A.34. Suppose V is finite-dimensional and $v_1, \ldots, v_m \in V$. Prove that the list $v_1, \ldots, v_m$ is linearly independent if and only if there exists $T \in \mathcal{L}(V)$ such that $v_1, \ldots, v_m$ are eigenvectors of T corresponding to distinct eigenvalues.

Solution. One direction is 5.11; the other is a construction using the linear map lemma.
Suppose first that there exists $T \in \mathcal{L}(V)$ and distinct scalars $\lambda_1, \ldots, \lambda_m \in \mathbf{F}$ such that v_k is an eigenvector of T corresponding to λ_k for each k. By 5.11, every list of eigenvectors of T corresponding to distinct eigenvalues is linearly independent. Hence $v_1, \ldots, v_m$ is linearly independent.
Now suppose $v_1, \ldots, v_m$ is linearly independent. Since V is finite-dimensional, 2.32 lets us extend this list to a basis

$$v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$$

of V, where $n = \dim V$. By the linear map lemma 3.4, there is a (unique) linear map $T : V \to V$ with

$$Tv_k = k\, v_k \qquad \text{for } k = 1, \ldots, n.$$

(The lemma applies because we are prescribing the values of T on a basis; the prescribed values kv_k are legitimate vectors of V.) Thus $T \in \mathcal{L}(V)$.
For each $k \in \{1, \ldots, m\}$ we have $v_k \neq 0$, because $v_1, \ldots, v_m$ is linearly independent and a list containing 0 is linearly dependent. Together with $Tv_k = kv_k$, the definition 5.8 shows that v_k is an eigenvector of T corresponding to the eigenvalue k. The numbers $1, 2, \ldots, m$ are distinct elements of $\mathbf{F}$ (recall $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$, so distinct positive integers remain distinct in $\mathbf{F}$). Hence $v_1, \ldots, v_m$ are eigenvectors of T corresponding to distinct eigenvalues, as required.
Remark. Any n distinct scalars $\mu_1, \ldots, \mu_n \in \mathbf{F}$ would do in place of $1, \ldots, n$; the operator so constructed is the diagonal operator whose matrix with respect to the basis $v_1, \ldots, v_n$ has $\mu_1, \ldots, \mu_n$ down the diagonal. The finite-dimensionality of V is used only to extend the list to a basis, and indirectly forces $m \leq \dim V$.

Problem 5A.35. Suppose that $\lambda_1, \dots, \lambda_n$ is a list of distinct real numbers. Prove that the list $e^{\lambda_1 x}, \dots, e^{\lambda_n x}$ is linearly independent in the vector space of real-valued functions on $\mathbf{R}$.
Hint: Let $V = \operatorname{span}(e^{\lambda_1 x}, \dots, e^{\lambda_n x})$, and define an operator $D \in \mathcal{L}(V)$ by $Df = f'$. Find eigenvalues and eigenvectors of D.

Solution. Here $e^{\lambda_k x}$ denotes the function $\mathbf{R} \to \mathbf{R}$ given by $x \mapsto e^{\lambda_k x}$; write f_k for this function. Let W denote the real vector space of all real-valued functions on $\mathbf{R}$, and set

$$V = \operatorname{span}(f_1, \dots, f_n) \subseteq W.$$

Then V is a subspace of W, hence itself a vector space over $\mathbf{R}$, and it is finite-dimensional (spanned by a list of length n).
Step 1: differentiation is an operator on V. Each f_k is differentiable on $\mathbf{R}$, with

$$f_k' = \lambda_k f_k.$$

Consequently every element of V, being a finite linear combination $a_1 f_1 + \cdots + a_n f_n$, is differentiable, and

$$(a_1 f_1 + \cdots + a_n f_n)' = a_1 \lambda_1 f_1 + \cdots + a_n \lambda_n f_n \in V.$$

So the assignment $Df = f'$ maps V into V. It is linear, because differentiation satisfies $(f+g)' = f' + g'$ and $(af)' = af'$ for differentiable f, g and $a \in \mathbf{R}$. Hence $D \in \mathcal{L}(V)$.
Step 2: eigenvalues and eigenvectors of D. For each k the function f_k is not the zero element of V, since $f_k(x) = e^{\lambda_k x} > 0$ for all $x \in \mathbf{R}$; in particular $f_k(0) = 1 \neq 0$. And

$$Df_k = f_k' = \lambda_k f_k.$$

By the definitions 5.5 and 5.8, each λ_k is an eigenvalue of D and f_k is an eigenvector of D corresponding to λ_k.
Step 3: conclusion. The numbers $\lambda_1, \dots, \lambda_n$ are distinct by hypothesis. Thus $f_1, \dots, f_n$ is a list of eigenvectors of D corresponding to distinct eigenvalues, so by 5.11 the list $f_1, \dots, f_n$ is linearly independent in V.
Finally, linear independence does not depend on the ambient space: if $a_1, \dots, a_n \in \mathbf{R}$ satisfy $a_1 f_1 + \cdots + a_n f_n = 0$ as an equation in W, then, since all the vectors involved lie in the subspace V, this is also an equation in V, whence $a_1 = \cdots = a_n = 0$. Therefore $e^{\lambda_1 x}, \dots, e^{\lambda_n x}$ is linearly independent in the vector space of real-valued functions on $\mathbf{R}$.
Remark. The same argument shows the list is linearly independent over $\mathbf{C}$ if the λ_k are distinct complex numbers and one works with complex-valued functions on $\mathbf{R}$. Exercise 5A.36 (about $\cos(\lambda_1 x), \dots, \cos(\lambda_n x)$ for distinct positive λ_k) can be handled by the same device applied to D^2, whose eigenvalues on the relevant span are the distinct numbers $-\lambda_k^2$.

Problem 5A.36. Suppose that $\lambda_1, \dots, \lambda_n$ is a list of distinct positive numbers. Prove that the list $\cos(\lambda_1 x), \dots, \cos(\lambda_n x)$ is linearly independent in the vector space of real-valued functions on $\mathbb{R}$.

Solution. For each k write c_k for the function $x \mapsto \cos(\lambda_k x)$, which is an infinitely differentiable real-valued function on $\mathbb{R}$. Let

$$V = \operatorname{span}(c_1, \dots, c_n),$$

a subspace of the vector space of real-valued functions on $\mathbb{R}$.
Define $T\colon V \to V$ by $Tf = f''$. This makes sense: differentiation twice is a linear map on the space of twice-differentiable functions, and

$$c_k''(x) = \frac{d^2}{dx^2} \cos(\lambda_k x) = -\lambda_k^2 \cos(\lambda_k x),$$

so $Tc_k = -\lambda_k^2 c_k \in V$ for each k. Because T is linear and maps each element of a spanning list of V

into V, it maps all of V into V. Thus $T \in \mathcal{L}(V)$.
Each c_k is a nonzero element of V, since $c_k(0) = \cos 0 = 1 \neq 0$. Hence the displayed equation $Tc_k = -\lambda_k^2 c_k$ says exactly that c_k is an eigenvector of T corresponding to the eigenvalue $-\lambda_k^2$.
These n eigenvalues are distinct. Indeed, if $-\lambda_j^2 = -\lambda_k^2$ then $\lambda_j^2 = \lambda_k^2$, so $(\lambda_j - \lambda_k)(\lambda_j + \lambda_k) = 0$; since λ_j and λ_k are positive we have $\lambda_j + \lambda_k > 0$, forcing $\lambda_j = \lambda_k$ and therefore $j = k$ (the λ's being distinct). This is the only place where positivity is used, and it is genuinely needed: without it, $\cos(x)$ and $\cos(-x)$ would be equal functions.
So $c_1, \ldots, c_n$ is a list of eigenvectors of T corresponding to n distinct eigenvalues of T. By 5.11, this list is linearly independent in V.
Finally, linear independence of a list is a statement about the vanishing of linear combinations of its entries, and does not depend on which ambient vector space we regard the entries as living in. Since V is a subspace of the space of real-valued functions on $\mathbb{R}$, the list $\cos(\lambda_1 x), \ldots, \cos(\lambda_n x)$ is linearly independent in the vector space of real-valued functions on $\mathbb{R}$.

Problem 5A.37. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Define $\mathcal{A} \in \mathcal{L}(\mathcal{L}(V))$ by

$$\mathcal{A}(S) = TS$$

for each $S \in \mathcal{L}(V)$. Prove that the set of eigenvalues of T equals the set of eigenvalues of $\mathcal{A}$.

Solution. If $V = \{0\}$ then $\mathcal{L}(V) = \{0\}$ as well, so neither V nor $\mathcal{L}(V)$ contains a nonzero vector and both T and $\mathcal{A}$ have empty sets of eigenvalues. So assume from now on that $V \neq \{0\}$, and let $n = \dim V \geq 1$.
First note that $\mathcal{A}$ really is an operator on $\mathcal{L}(V)$: it maps $\mathcal{L}(V)$ into $\mathcal{L}(V)$ because a composition of operators on V is an operator on V, and it is linear because $T(S_1+S_2) = TS_1+TS_2$ and $T(\alpha S) = \alpha(TS)$ for all $S_1, S_2, S \in \mathcal{L}(V)$ and $\alpha \in \mathbf{F}$.
One inclusion: every eigenvalue of $\mathcal{A}$ is an eigenvalue of T. Suppose $\lambda \in \mathbf{F}$ is an eigenvalue of $\mathcal{A}$. Then there exists $S \in \mathcal{L}(V)$ with $S \neq 0$ and $\mathcal{A}(S) = \lambda S$, which says

$$TS = \lambda S.$$

Because $S \neq 0$, there exists $v \in V$ with $Sv \neq 0$. Applying both sides of the displayed equation to v gives

$$T(Sv) = \lambda(Sv).$$

Since Sv is a nonzero vector in V, it is an eigenvector of T corresponding to λ. Hence λ is an eigenvalue of T.
The other inclusion: every eigenvalue of T is an eigenvalue of $\mathcal{A}$. Suppose $\lambda \in \mathbf{F}$ is an eigenvalue of T, and let $v \in V$ be a corresponding eigenvector, so $v \neq 0$ and $Tv = \lambda v$. Let $v_1, \ldots, v_n$ be a basis of V. By the linear map lemma 3.4 there is a unique $S \in \mathcal{L}(V)$ with

$$Sv_1 = v \quad \text{and} \quad Sv_j = 0 \text{ for } j = 2, \ldots, n.$$

Then $S \neq 0$, because $Sv_1 = v \neq 0$.
Now take any $w \in V$ and write $w = a_1 v_1 + \cdots + a_n v_n$ with $a_1, \ldots, a_n \in \mathbf{F}$. Then $Sw = a_1 v$, and therefore

$$(TS)w = T(a_1 v) = a_1 Tv = a_1 \lambda v = \lambda(a_1 v) = \lambda(Sw) = (\lambda S)w.$$

Since $w \in V$ was arbitrary, $TS = \lambda S$, which is to say $\mathcal{A}(S) = \lambda S$ with $S \neq 0$. Hence λ is an eigenvalue of $\mathcal{A}$.
The two inclusions together show that the set of eigenvalues of T equals the set of eigenvalues of $\mathcal{A}$.
(The idea behind the second half: an operator S whose range is contained in the eigenspace $E(\lambda, T)$ automatically satisfies $TS = \lambda S$, and such nonzero S exist as soon as $E(\lambda, T) \neq \{0\}$.)

Problem 5A.38. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is a subspace of V invariant under T. The quotient operator $T/U \in \mathcal{L}(V/U)$ is defined by

$$(T/U)(v+U) = Tv + U$$

for each $v \in V$.
(a) Show that the definition of T/U makes sense (which requires using the condition that U is invariant under T) and show that T/U is an operator on V/U.
(b) Show that each eigenvalue of T/U is an eigenvalue of T.

Solution. (a) The formula $(T/U)(v+U) = Tv + U$ prescribes the value of T/U at an element of V/U by choosing a representative v of that coset. For this to define a function on V/U we must check that the answer does not depend on the representative chosen.
So suppose $v, w \in V$ satisfy $v + U = w + U$. Then $v - w \in U$. Because U is invariant under T, we get $T(v-w) \in U$, and by linearity of T this says $Tv - Tw \in U$, which is equivalent to

$$Tv + U = Tw + U.$$

Thus the two candidate values agree, and T/U is a well-defined function from V/U to V/U. This is exactly where invariance of U was used, and it cannot be dispensed with: if $T(v-w) \notin U$ for some $v - w \in U$, the two representatives would give different answers.
Now we check that T/U is linear. For $v, w \in V$, using the definition of addition in V/U and the linearity of T,

$$\begin{aligned}(T/U)\big((v+U)+(w+U)\big) &= (T/U)\big((v+w)+U\big)\\ &= T(v+w)+U\\ &= (Tv+Tw)+U\\ &= (Tv+U)+(Tw+U)\\ &= (T/U)(v+U)+(T/U)(w+U).\end{aligned}$$

Similarly, for $\alpha \in \mathbf{F}$ and $v \in V$,

$$(T/U)\big(\alpha(v+U)\big) = (T/U)(\alpha v + U) = T(\alpha v) + U = \alpha(Tv) + U = \alpha\big((T/U)(v+U)\big).$$

Hence T/U is a linear map from V/U to V/U; since its domain and codomain are the same vector space V/U, it is an operator on V/U, that is, $T/U \in \mathcal{L}(V/U)$.
(b) Suppose $\lambda \in \mathbf{F}$ is an eigenvalue of T/U. Then there exists $v \in V$ with $v + U \neq 0 + U$, equivalently $v \notin U$, such that

$$(T/U)(v+U) = \lambda(v+U),$$

that is, $Tv + U = \lambda v + U$, that is,

$$(T - \lambda I)v \in U.$$

Write $u = (T - \lambda I)v \in U$.
Because U is invariant under T and obviously invariant under I, the subspace U is invariant under $T - \lambda I$. So we may consider the restriction operator

$$(T - \lambda I)|_U \in \mathcal{L}(U),$$

and note that U is finite-dimensional, being a subspace of the finite-dimensional space V. There are two cases.
Case 1: $(T - \lambda I)|_U$ is not injective. Then there exists a nonzero $u' \in U$ with $(T - \lambda I)u' = 0$, that is, $Tu' = \lambda u'$. Thus u' is an eigenvector of T corresponding to λ, so λ is an eigenvalue of T.
Case 2: $(T - \lambda I)|_U$ is injective. Since U is finite-dimensional, injectivity of an operator on U implies surjectivity onto U (3.65). Hence there exists $u' \in U$ with

$$(T - \lambda I)u' = u.$$

Then

$$(T - \lambda I)(v - u') = (T - \lambda I)v - (T - \lambda I)u' = u - u = 0.$$

Moreover $v - u' \neq 0$: if $v - u'$ were 0 then $v = u' \in U$, contradicting $v \notin U$. So $v - u'$ is a nonzero vector with $T(v - u') = \lambda(v - u')$, making it an eigenvector of T corresponding to λ. Again λ is an eigenvalue of T.
In both cases λ is an eigenvalue of T, which completes the proof.

Problem 5A.39. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T has an eigenvalue if and only if there exists a subspace of V of dimension $\dim V - 1$ that is invariant under T.

Solution. Note first that if $V = \{0\}$ then both sides of the asserted equivalence are false: T has no eigenvalue, because V contains no nonzero vector, and no subspace has dimension $\dim V - 1 = -1$. So the equivalence holds in that case. In each direction below, the hypothesis being assumed forces $\dim V \geq 1$ (an eigenvector is a nonzero vector of V; a subspace of dimension $\dim V - 1$ exists only if $\dim V - 1 \geq 0$), so no separate treatment of $V = \{0\}$ is needed there.
Suppose first that T has an eigenvalue $\lambda \in \mathbf{F}$. Set

$$W = \operatorname{range}(T - \lambda I).$$

Because λ is an eigenvalue, $T - \lambda I$ is not injective (5.7), hence not surjective (5.7 again, or 3.65). Therefore $W \neq V$, and since W is a subspace of the finite-dimensional space V we get

$$\dim W \leq \dim V - 1.$$

Choose a basis $w_1, \ldots, w_m$ of W and extend it to a basis $w_1, \ldots, w_m, u_{m+1}, \ldots, u_n$ of V, where $n = \dim V$; this is possible by 2.32. Since $m \leq n - 1$, we may define

$$U = \operatorname{span}(w_1, \ldots, w_m, u_{m+1}, \ldots, u_{n-1}),$$

the span of the first $n - 1$ vectors of that basis. Then $\dim U = n - 1 = \dim V - 1$, because a sublist of a linearly independent list is linearly independent. Also $W \subseteq U$, since $w_1, \ldots, w_m$ are among the spanning vectors of U.
Now U is invariant under T: if $u \in U$ then

$$Tu = (T - \lambda I)u + \lambda u \in W + U \subseteq U,$$

where the first term lies in $W \subseteq U$ by definition of W, the second lies in U because U is a subspace, and U is closed under addition. So U is a subspace of V of dimension $\dim V - 1$ invariant under T.
Conversely, suppose U is a subspace of V with $\dim U = \dim V - 1$ that is invariant under T. By 3.105,

$$\dim V/U = \dim V - \dim U = 1.$$

By 5A.38(a), the quotient operator T/U is a well-defined operator on V/U.
Since $\dim V/U = 1$, pick any nonzero vector $w \in V/U$; then w is a basis of V/U, so $\operatorname{span}(w) = V/U$. In particular $(T/U)w \in V/U = \operatorname{span}(w)$, so there exists $\lambda \in \mathbf{F}$ with

$$(T/U)w = \lambda w.$$

As $w \neq 0$, this exhibits λ as an eigenvalue of T/U. By 5A.38(b), every eigenvalue of T/U is an eigenvalue of T. Hence λ is an eigenvalue of T, and in particular T has an eigenvalue.
(Note that the argument shows more: for every hyperplane U invariant under T, the scalar by which T acts on the one-dimensional quotient V/U is an eigenvalue of T.)

Problem 5A.40. Suppose $S, T \in \mathcal{L}(V)$ and S is invertible. Suppose $p \in \mathcal{P}(\mathbf{F})$ is a polynomial. Prove that

$$p(STS^{-1}) = S\,p(T)\,S^{-1}.$$

Solution. The key step is the corresponding statement for powers.
Claim: $(STS^{-1})^k = ST^kS^{-1}$ for every nonnegative integer k.
We prove this by induction on k. For $k = 0$ both sides equal the identity operator I: by the convention 5.13 that $R^0 = I$ for $R \in \mathcal{L}(V)$, the left side is I, and the right side is $SIS^{-1} = SS^{-1} = I$.
Now suppose $k \geq 0$ and $(STS^{-1})^k = ST^kS^{-1}$. Then

$$\begin{aligned}(STS^{-1})^{k+1} &= (STS^{-1})^k(STS^{-1}) \\ &= (ST^kS^{-1})(STS^{-1}) \\ &= ST^k(S^{-1}S)TS^{-1} \\ &= ST^kI\,TS^{-1} \\ &= ST^{k+1}S^{-1},\end{aligned}$$

where we used associativity of composition of linear maps and $S^{-1}S = I$. This completes the induction and proves the claim.
Now write $p(z) = a_0 + a_1z + \cdots + a_mz^m$ with $a_0, \ldots, a_m \in \mathbf{F}$. By the definition 5.14 of a polynomial applied to an operator,

$$p(STS^{-1}) = a_0I + a_1(STS^{-1}) + \cdots + a_m(STS^{-1})^m.$$

Applying the claim to each term gives

$$p(STS^{-1}) = a_0SIS^{-1} + a_1STS^{-1} + \cdots + a_mST^mS^{-1}.$$

The map $R \mapsto SRS^{-1}$ from $\mathcal{L}(V)$ to $\mathcal{L}(V)$ is linear, since composition of linear maps is distributive over addition and commutes with scalar multiplication: $S(R_1 + R_2)S^{-1} = SR_1S^{-1} + SR_2S^{-1}$ and $S(\alpha R)S^{-1} = \alpha\,SRS^{-1}$. Hence we may pull the sum and the scalars inside:

$$p(STS^{-1}) = S\big(a_0I + a_1T + \cdots + a_mT^m\big)S^{-1} = S\,p(T)\,S^{-1},$$

as desired.

Problem 5A.41. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T. Prove that U is invariant under $p(T)$ for every polynomial $p \in \mathcal{P}(\mathbf{F})$.

Solution. We first show that U is invariant under T^k for every nonnegative integer k, by induction on k.
For $k = 0$ we have $T^0 = I$ (see 5.13), and $Iu = u \in U$ for every $u \in U$, so U is invariant under T^0.
Now suppose $k \geq 0$ and U is invariant under T^k. Let $u \in U$. Then $T^ku \in U$ by the induction hypothesis, and hence

$$T^{k+1}u = T(T^ku) \in U$$

because U is invariant under T. Thus U is invariant under T^{k+1}, completing the induction.
Now let $p \in \mathcal{P}(\mathbf{F})$ and write $p(z) = a_0 + a_1z + \cdots + a_mz^m$ with $a_0, \ldots, a_m \in \mathbf{F}$. By 5.14,

$$p(T) = a_0I + a_1T + \cdots + a_mT^m.$$

Let $u \in U$. Then

$$p(T)u = a_0u + a_1(Tu) + \cdots + a_m(T^mu).$$

Each vector T^ku on the right side lies in U by the paragraph above. Since U is a subspace, it is closed under scalar multiplication and under addition, so the entire linear combination $p(T)u$ lies in U.

As $u \in U$ was arbitrary, U is invariant under $p(T)$. Note that no finite-dimensionality hypothesis was needed anywhere in this argument.

Problem 5A.42. Define $T \in \mathcal{L}(\mathbf{F}^n)$ by $T(x_1, x_2, x_3, \ldots, x_n) = (x_1, 2x_2, 3x_3, \ldots, nx_n)$.
(a) Find all eigenvalues and eigenvectors of T.
(b) Find all subspaces of $\mathbf{F}^n$ that are invariant under T.

Solution. Let $e_1, \ldots, e_n$ denote the standard basis of $\mathbf{F}^n$, so e_j has 1 in slot j and 0 elsewhere. The definition of T says precisely that

$$Te_j = j\, e_j \qquad \text{for } j = 1, \ldots, n.$$

Throughout, recall that $\mathbf{F}$ is $\mathbb{R}$ or $\mathbb{C}$, so the scalars $1, 2, \ldots, n$ are distinct elements of $\mathbf{F}$ and $j - k \neq 0$ in $\mathbf{F}$ whenever $j \neq k$.
(a) The eigenvalues of T are exactly $1, 2, \ldots, n$, and for each j the eigenvectors corresponding to j are exactly the nonzero scalar multiples of e_j; equivalently $E(j, T) = \operatorname{span}(e_j)$.
To see this, let $\lambda \in \mathbf{F}$ and let $x = (x_1, \ldots, x_n) \in \mathbf{F}^n$. Comparing coordinates, the equation $Tx = \lambda x$ says

$$k\, x_k = \lambda x_k \quad \text{for } k = 1, \ldots, n, \qquad \text{i.e.} \qquad (k - \lambda)x_k = 0 \quad \text{for } k = 1, \ldots, n.$$

If $\lambda \notin \{1, \ldots, n\}$, then $k - \lambda \neq 0$ for every k, so each of these equations forces $x_k = 0$. Hence $x = 0$, so there is no eigenvector corresponding to λ and λ is not an eigenvalue.
If $\lambda = j$ for some $j \in \{1, \ldots, n\}$, then the equation with $k = j$ is automatic, while for $k \neq j$ we have $k - j \neq 0$ and so $x_k = 0$. Thus the solutions of $Tx = jx$ are exactly the vectors $x = x_j e_j$ with $x_j \in \mathbf{F}$, that is, $E(j, T) = \operatorname{span}(e_j)$. Since $e_j \neq 0$, this space contains nonzero vectors, so j is indeed an eigenvalue, and its eigenvectors are the vectors $c\, e_j$ with $c \in \mathbf{F}$ and $c \neq 0$.
(b) The subspaces of $\mathbf{F}^n$ invariant under T are exactly the 2^n coordinate subspaces

$$U_\Omega = \operatorname{span}(e_j : j \in \Omega), \qquad \Omega \subseteq \{1, \ldots, n\},$$

where $U_\emptyset = \{0\}$ and $U_{\{1,\ldots,n\}} = \mathbf{F}^n$.
Each U_Ω is invariant under T: the spanning vectors satisfy $Te_j = je_j \in U_\Omega$ for $j \in \Omega$, and since T is linear it therefore maps every linear combination of those vectors back into U_Ω.
Conversely, suppose U is a subspace of $\mathbf{F}^n$ invariant under T. For each $j \in \{1, \ldots, n\}$ define the Lagrange-type polynomial

$$p_j(z) = \prod_{k \neq j,\, 1 \leq k \leq n} \frac{z - k}{j - k} \in \mathcal{P}(\mathbf{F}),$$

which is legitimate because $j - k \neq 0$ for $k \neq j$. (If $n = 1$ the product is empty, so p_1 is the constant polynomial 1.) By construction $p_j(j) = 1$ and $p_j(k) = 0$ for every $k \in \{1, \ldots, n\}$ with $k \neq j$.
Since $Te_k = ke_k$, induction on the degree gives $T^m e_k = k^m e_k$ for every $m \geq 0$, and hence by 5.14, for any polynomial p,

$$p(T)e_k = p(k)\, e_k.$$

In particular $p_j(T)e_k = 0$ for $k \neq j$ and $p_j(T)e_j = e_j$.
Now let $x = (x_1, \ldots, x_n) = x_1 e_1 + \cdots + x_n e_n$ be any element of U. By 5A.41, U is invariant under $p_j(T)$, so

$$p_j(T)x = \sum_{k=1}^{n} x_k\, p_j(T)e_k = x_j\, e_j \in U.$$

Therefore, whenever $x_j \neq 0$, dividing by the scalar x_j gives $e_j \in U$.
Let $\Omega = \{j : e_j \in U\}$. The previous paragraph shows that every $x \in U$ has $x_j = 0$ for all $j \notin \Omega$, so $x \in U_\Omega$; hence $U \subseteq U_\Omega$. The reverse inclusion $U_\Omega \subseteq U$ is immediate from the definition of Ω, since U is a subspace containing each e_j with $j \in \Omega$. Thus $U = U_\Omega$.
Finally, distinct subsets give distinct subspaces. Fix $j \in \{1, \ldots, n\}$ and $\Omega \subseteq \{1, \ldots, n\}$. If $j \in \Omega$ then certainly $e_j \in U_\Omega$. Conversely, if $j \notin \Omega$ and we had $e_j \in U_\Omega$, then e_j would be a linear combination

of the vectors e_k with $k \in \Omega$, contradicting the linear independence of $e_1, \dots, e_n$. So $e_j \in U_\Omega$ if and only if $j \in \Omega$, which recovers Ω from U_Ω. Hence $\Omega \mapsto U_\Omega$ is injective, and T has exactly 2^n invariant subspaces.

Problem 5A.43. Suppose that V is finite-dimensional, $\dim V > 1$, and $T \in \mathcal{L}(V)$. Prove that

$$\{p(T) : p \in \mathcal{P}(\mathbf{F})\} \neq \mathcal{L}(V).$$

Solution. Write

$$\mathcal{E} = \{p(T) : p \in \mathcal{P}(\mathbf{F})\},$$

and let $n = \dim V$, so $n \geq 2$ by hypothesis. The key point is that $\mathcal{E}$ is commutative while $\mathcal{L}(V)$ is not. First, any two elements of $\mathcal{E}$ commute with each other: if $p, q \in \mathcal{P}(\mathbf{F})$, then by 5.17(b)

$$p(T)\, q(T) = q(T)\, p(T).$$

Now we produce two operators on V that do not commute. Because V is finite-dimensional with $\dim V = n \geq 2$, there is a basis $v_1, \dots, v_n$ of V by 2.31; in particular v_1 and v_2 both exist, and $v_2 \neq 0$ because a basis is linearly independent. By the linear map lemma 3.4 there are unique operators $R, S \in \mathcal{L}(V)$ satisfying

$$Rv_1 = v_2, \qquad Rv_j = 0 \ \text{ for } j = 2, \dots, n,$$
$$Sv_1 = v_1, \qquad Sv_j = 0 \ \text{ for } j = 2, \dots, n.$$

Then

$$(RS)v_1 = R(Sv_1) = Rv_1 = v_2, \qquad (SR)v_1 = S(Rv_1) = Sv_2 = 0.$$

Since $v_2 \neq 0$, the operators RS and SR disagree at v_1, so $RS \neq SR$.
Suppose now, seeking a contradiction, that $\mathcal{E} = \mathcal{L}(V)$. Then $R \in \mathcal{E}$ and $S \in \mathcal{E}$, so by the first paragraph $RS = SR$, contradicting the computation above. Hence $\mathcal{E} \neq \mathcal{L}(V)$, as desired.
Remark: here is a second proof, which extracts more from the assumption. It shows that $\mathcal{E} = \mathcal{L}(V)$ would force $\mathcal{E}$ to have dimension 1, whereas $\mathcal{L}(V)$ has dimension at least 4. Suppose again $\mathcal{E} = \mathcal{L}(V)$. Taking $q(z) = z$ in 5.17(b) shows that every element of $\mathcal{E}$ commutes with T; hence under our assumption every operator on V commutes with T. We claim every nonzero vector of V is then an eigenvector of T. Indeed, suppose $v \in V$ is nonzero and $Tv \notin \operatorname{span}(v)$. Then v, Tv is linearly independent, so it extends to a basis $v, Tv, u_3, \dots, u_n$ of V (see 2.32). By 3.4 there is $S \in \mathcal{L}(V)$ with

$$Sv = v, \qquad S(Tv) = 0, \qquad Su_j = 0 \ \text{ for } j = 3, \dots, n.$$

Then $(ST)v = S(Tv) = 0$, while $(TS)v = T(Sv) = Tv \neq 0$ (note $Tv \neq 0$ because v, Tv is linearly independent). Thus $ST \neq TS$, contradicting that T commutes with every operator. Hence for each nonzero $v \in V$ there exists $\lambda_v \in \mathbf{F}$ with $Tv = \lambda_v v$.
These scalars are all equal. Let $v, w \in V$ be nonzero. If $w = cv$ for some $c \in \mathbf{F} \setminus \{0\}$, then $\lambda_w cv = Tw = cTv = c\lambda_v v$, and $cv \neq 0$ forces $\lambda_w = \lambda_v$. If instead v, w is linearly independent, then $v + w \neq 0$ and

$$\lambda_{v+w} v + \lambda_{v+w} w = T(v + w) = Tv + Tw = \lambda_v v + \lambda_w w,$$

so comparing coefficients with respect to the linearly independent list v, w gives $\lambda_v = \lambda_{v+w} = \lambda_w$. Writing λ for this common value, we get $Tv = \lambda v$ for all $v \in V$ (including $v = 0$), that is, $T = \lambda I$.
But then for every $p \in \mathcal{P}(\mathbf{F})$, say $p(z) = a_0 + a_1 z + \cdots + a_m z^m$, we have $T^k = \lambda^k I$ and hence

$$p(T) = (a_0 + a_1\lambda + \cdots + a_m\lambda^m) I = p(\lambda) I.$$

Therefore $\mathcal{E} = \{cI : c \in \mathbf{F}\} = \operatorname{span}(I)$, which is a subspace of dimension 1. On the other hand $\dim \mathcal{L}(V) = (\dim V)^2 = n^2 \geq 4$ by 3.72. Since $1 \neq n^2$, the subspaces $\mathcal{E}$ and $\mathcal{L}(V)$ cannot be equal, again a contradiction. Either argument establishes the desired conclusion.
Finally, note where the hypothesis $\dim V > 1$ is used: it is essential. If $\dim V = 1$, then every operator on V is a scalar multiple of I (as $Tv \in V = \operatorname{span}(v)$ for a basis vector v), so $\mathcal{L}(V) = \{cI : c \in \mathbf{F}\}$, and taking constant polynomials $p(z) = c$ gives $\mathcal{E} = \mathcal{L}(V)$.

5.2 Exercises 5B

Problem 5B.1. Suppose $T \in \mathcal{L}(V)$. Prove that 9 is an eigenvalue of T^2 if and only if 3 or -3 is an eigenvalue of T.

Solution. Note that V is not assumed to be finite-dimensional, so the argument below is arranged to avoid any use of finite-dimensionality.
First suppose 3 or -3 is an eigenvalue of T. Thus there exists $v \in V$ with $v \neq 0$ such that $Tv = 3v$ or $Tv = -3v$. In either case,

$$T^2v = T(Tv) = T(\pm 3v) = \pm 3\,Tv = (\pm 3)^2 v = 9v.$$

Because $v \neq 0$, this shows that 9 is an eigenvalue of T^2.
Conversely, suppose 9 is an eigenvalue of T^2. Thus there exists $v \in V$ with $v \neq 0$ such that $T^2v = 9v$, which we can rewrite as

$$(T^2 - 9I)v = 0.$$

Because polynomials of T can be factored just as ordinary polynomials can (see 5.17), we have $T^2 - 9I = (T - 3I)(T + 3I)$. Hence

$$(T - 3I)\big((T + 3I)v\big) = 0.$$

Let $w = (T + 3I)v$. There are two cases.
Case 1: $w = 0$. Then $Tv = -3v$ with $v \neq 0$, so -3 is an eigenvalue of T.
Case 2: $w \neq 0$. Then the displayed equation says $(T - 3I)w = 0$, so $Tw = 3w$ with $w \neq 0$, and hence 3 is an eigenvalue of T.
In both cases 3 or -3 is an eigenvalue of T, completing the proof.

Problem 5B.2. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ has no eigenvalues. Prove that every subspace of V invariant under T is either $\{0\}$ or infinite-dimensional.

Solution. Let U be a subspace of V that is invariant under T. We must show that if $U \neq \{0\}$ then U is infinite-dimensional. So suppose, seeking a contradiction, that $U \neq \{0\}$ and U is finite-dimensional. Because U is invariant under T, the restriction $T|_U$ is a well-defined operator on U, that is, $T|_U \in \mathcal{L}(U)$. Now U is a finite-dimensional nonzero complex vector space (it is a vector space over $\mathbf{C}$ because it is a subspace of the complex vector space V). Hence 5.19, which states that every operator on a finite-dimensional nonzero complex vector space has an eigenvalue, implies that $T|_U$ has an eigenvalue $\lambda \in \mathbf{C}$.
Thus there exists $u \in U$ with $u \neq 0$ such that

$$(T|_U)u = \lambda u.$$

But $(T|_U)u = Tu$ by definition of the restriction. Hence $Tu = \lambda u$ with $u \in V$ and $u \neq 0$, which means that λ is an eigenvalue of T. This contradicts the hypothesis that T has no eigenvalues.
Therefore no invariant subspace of V can be both nonzero and finite-dimensional. In other words, every subspace of V invariant under T is either $\{0\}$ or infinite-dimensional, as desired.

Problem 5B.3. Suppose n is an integer with $n > 1$ and $T \in \mathcal{L}(\mathbf{F}^n)$ is defined by

$$T(x_1, \ldots, x_n) = (x_1 + \cdots + x_n, \ldots, x_1 + \cdots + x_n).$$

(a) Find all eigenvalues and eigenvectors of T.
(b) Find the minimal polynomial of T.
[The matrix of T with respect to the standard basis of $\mathbf{F}^n$ consists of all 1's.]

Solution. Write $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbf{F}^n$ and, for $x = (x_1, \ldots, x_n) \in \mathbf{F}^n$, put

$$s(x) = x_1 + \cdots + x_n.$$

Then s is a linear functional on $\mathbf{F}^n$ and the definition of T says exactly that

$$Tx = s(x)\,\mathbf{1} \qquad \text{for all } x \in \mathbf{F}^n.$$

In particular $\operatorname{range} T \subseteq \operatorname{span}(\mathbf{1})$, and since $T\mathbf{1} = n\mathbf{1} \neq 0$ (note $n \neq 0$ in $\mathbf{F}$ because $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$), we have $\operatorname{range} T = \operatorname{span}(\mathbf{1})$.
(a) We claim the eigenvalues of T are exactly 0 and n.
First, 0 is an eigenvalue. Indeed $Tx = 0$ if and only if $s(x)\mathbf{1} = 0$, which (as $\mathbf{1} \neq 0$) holds if and only if $s(x) = 0$. Thus

$$\operatorname{null} T = \{(x_1, \ldots, x_n) \in \mathbf{F}^n : x_1 + \cdots + x_n = 0\},$$

which has dimension $n - 1 \geq 1$ because $n > 1$ (it is the null space of the surjective linear functional s, so 3.21 gives $\dim \operatorname{null} s = n - 1$). Hence $\operatorname{null} T \neq \{0\}$ and 0 is an eigenvalue of T, with eigenvectors precisely the nonzero vectors whose coordinates sum to 0.
Second, n is an eigenvalue: $T\mathbf{1} = s(\mathbf{1})\mathbf{1} = n\mathbf{1}$ and $\mathbf{1} \neq 0$.
Now suppose λ is an eigenvalue of T with $\lambda \neq 0$, say $Tx = \lambda x$ with $x \neq 0$. Then

$$\lambda x = Tx = s(x)\,\mathbf{1},$$

so $x = \frac{s(x)}{\lambda}\mathbf{1}$. Thus $x = c\mathbf{1}$ for some $c \in \mathbf{F}$, and $c \neq 0$ because $x \neq 0$. Applying T gives

$$\lambda c\mathbf{1} = Tx = c\,T\mathbf{1} = cn\mathbf{1},$$

and cancelling the nonzero scalar c and the nonzero vector $\mathbf{1}$ yields $\lambda = n$. This same computation shows that the eigenvectors corresponding to the eigenvalue n are precisely the nonzero scalar multiples of $\mathbf{1}$; in other words the eigenspace of T corresponding to n equals $\operatorname{span}(\mathbf{1})$.
Summarizing: the eigenvalues of T are 0 and n (these are distinct because $n > 1$). The eigenvectors corresponding to 0 are the nonzero $(x_1, \ldots, x_n)$ with $x_1 + \cdots + x_n = 0$; the eigenvectors corresponding to n are the nonzero scalar multiples of $(1, 1, \ldots, 1)$.
(b) We claim the minimal polynomial of T is

$$z^2 - nz = z(z - n).$$

First, this polynomial applied to T gives 0: for every $x \in \mathbf{F}^n$,

$$T^2x = T\big(s(x)\mathbf{1}\big) = s(x)\,T\mathbf{1} = s(x)\,n\mathbf{1} = n\,Tx,$$

so $T^2 - nT = 0$. Thus the monic polynomial $q(z) = z^2 - nz$ satisfies $q(T) = 0$, and by 5.29 the minimal polynomial of T divides q.
It remains to rule out a minimal polynomial of degree 0 or 1. The only monic polynomial of degree 0 is the constant polynomial 1, and 1 applied to T is $I \neq 0$ (here $\mathbf{F}^n \neq \{0\}$). If the minimal polynomial had degree 1, say $z - c$, then $T - cI = 0$, that is, $T = cI$. Applying this to $\mathbf{1}$ gives $n\mathbf{1} = c\mathbf{1}$, so $c = n$; applying it to a nonzero $x \in \operatorname{null} T$ (such x exists because $n > 1$) gives $0 = cx$, so $c = 0$. These force $n = 0$, contradicting $n > 1$. (Equivalently, by 5.27(a) the minimal polynomial has both 0 and n among its zeros, so its degree is at least 2.)
Hence the minimal polynomial of T has degree 2, divides the monic degree-2 polynomial $z(z - n)$, and is monic; therefore it equals

$$z^2 - nz.$$

Problem 5B.4. Suppose $\mathbf{F} = \mathbf{C}$, $T \in \mathcal{L}(V)$, $p \in \mathcal{P}(\mathbf{C})$ is a nonconstant polynomial, and $\alpha \in \mathbf{C}$. Prove that α is an eigenvalue of $p(T)$ if and only if $\alpha = p(\lambda)$ for some eigenvalue λ of T.

Solution. First suppose $\alpha = p(\lambda)$ for some eigenvalue λ of T. Choose $v \in V$ with $v \neq 0$ and $Tv = \lambda v$. Repeated application of T gives $T^k v = \lambda^k v$ for every nonnegative integer k, and hence, writing $p(z) = a_0 + a_1 z + \cdots + a_m z^m$,

$$p(T)v = a_0 v + a_1 Tv + \cdots + a_m T^m v = (a_0 + a_1\lambda + \cdots + a_m\lambda^m)v = p(\lambda)v = \alpha v.$$

Because $v \neq 0$, this shows that α is an eigenvalue of $p(T)$. Note that this direction did not use $\mathbf{F} = \mathbf{C}$. Conversely, suppose α is an eigenvalue of $p(T)$. Define $q \in \mathcal{P}(\mathbf{C})$ by

$$q(z) = p(z) - \alpha.$$

Then q is nonconstant (it differs from the nonconstant polynomial p by a constant), so $m = \deg q = \deg p \geq 1$. By the second version of the fundamental theorem of algebra (4.13), there exist $c, \lambda_1, \ldots, \lambda_m \in \mathbf{C}$ with $c \neq 0$ such that

$$q(z) = c(z - \lambda_1) \cdots (z - \lambda_m)$$

for all $z \in \mathbf{C}$. Because the map $r \mapsto r(T)$ from $\mathcal{P}(\mathbf{C})$ to $\mathcal{L}(V)$ preserves products (5.17), we obtain

$$p(T) - \alpha I = q(T) = c\,(T - \lambda_1 I) \cdots (T - \lambda_m I).$$

Since α is an eigenvalue of $p(T)$, the operator $p(T) - \alpha I$ is not injective: there exists $v \neq 0$ with $(p(T) - \alpha I)v = 0$. Now if every one of the operators $T - \lambda_1 I, \ldots, T - \lambda_m I$ were injective, then their composition would be injective, and multiplying by the nonzero scalar c preserves injectivity, so $p(T) - \alpha I$ would be injective, a contradiction. Hence $T - \lambda_j I$ is not injective for some $j \in \{1, \ldots, m\}$, which means precisely that λ_j is an eigenvalue of T.
Finally, λ_j is a zero of q by the factorization above, so

$$0 = q(\lambda_j) = p(\lambda_j) - \alpha,$$

giving $\alpha = p(\lambda_j)$ with λ_j an eigenvalue of T, as desired.
(The hypothesis $\mathbf{F} = \mathbf{C}$ was used only in this second direction, to factor q into linear factors. Note also that neither direction required V to be finite-dimensional.)

Problem 5B.5. Give an example of an operator on $\mathbf{R}^2$ that shows the result in Exercise 4 does not hold if $\mathbf{C}$ is replaced with $\mathbf{R}$.

Solution. Take $\mathbf{F} = \mathbf{R}$, let $V = \mathbf{R}^2$, and define $T \in \mathcal{L}(\mathbf{R}^2)$ by

$$T(x, y) = (-y, x)$$

(counterclockwise rotation by 90 degrees). Let $p \in \mathcal{P}(\mathbf{R})$ be the polynomial $p(z) = z^2$, which is nonconstant, and let $\alpha = -1 \in \mathbf{R}$.
First, T has no eigenvalues. Suppose $\lambda \in \mathbf{R}$ and $(x, y) \in \mathbf{R}^2$ satisfy $T(x, y) = \lambda(x, y)$. Then

$$-y = \lambda x \quad \text{and} \quad x = \lambda y.$$

Substituting the second equation into the first gives $-y = \lambda^2 y$, that is, $(\lambda^2 + 1)y = 0$. Since $\lambda^2 + 1 > 0$ for every real λ, we get $y = 0$, and then $x = \lambda y = 0$. Thus the only solution is $(x, y) = (0, 0)$, so no $\lambda \in \mathbf{R}$ is an eigenvalue of T.
Second, $\alpha = -1$ is an eigenvalue of $p(T) = T^2$. Indeed

$$T^2(x, y) = T(-y, x) = (-x, -y) = -(x, y),$$

so $T^2 = -I$. Hence every nonzero vector of $\mathbf{R}^2$ is an eigenvector of T^2 corresponding to the eigenvalue -1.

Thus $\alpha = -1$ is an eigenvalue of $p(T)$, but there is no eigenvalue λ of T with $\alpha = p(\lambda)$, for the trivial reason that T has no eigenvalues at all. Hence the implication
if α is an eigenvalue of $p(T)$, then $\alpha = p(\lambda)$ for some eigenvalue λ of T
fails here, so the result of Exercise 4 does not hold with $\mathbf{C}$ replaced by $\mathbf{R}$. (Only this implication can fail: the reverse implication, proved in Exercise 4 without using $\mathbf{F} = \mathbf{C}$, is valid for $\mathbf{F} = \mathbf{R}$ as well. The failure traces to the fact that $p(z) - \alpha = z^2 + 1$ has no zero in $\mathbf{R}$, so the factorization into linear factors used in Exercise 4 is unavailable.)

Problem 5B.6. Suppose $T \in \mathcal{L}(\mathbf{F}^2)$ is defined by $T(w, z) = (-z, w)$. Find the minimal polynomial of T.

Solution. The minimal polynomial of T is

$$z^2 + 1,$$

for both $\mathbf{F} = \mathbf{R}$ and $\mathbf{F} = \mathbf{C}$.
To verify this, first compute T^2: for all $(w, z) \in \mathbf{F}^2$,

$$T^2(w, z) = T(-z, w) = (-w, -z) = -(w, z).$$

Thus $T^2 = -I$, so $T^2 + I = 0$. In other words, the monic polynomial $q(z) = z^2 + 1$ satisfies $q(T) = 0$. By 5.29, the minimal polynomial of T divides q.
Now we show the minimal polynomial has degree 2, which forces it to equal the monic degree-2 polynomial q. The only monic polynomial of degree 0 is the constant polynomial 1, and applying it to T gives I, which is not the zero operator since $\mathbf{F}^2 \neq \{0\}$. If the minimal polynomial had degree 1, say $z - c$ with $c \in \mathbf{F}$, then $T = cI$; but

$$T(1, 0) = (0, 1),$$

which is not a scalar multiple of $(1, 0)$, so T is not a scalar multiple of the identity. Hence the minimal polynomial has degree at least 2.
Therefore the minimal polynomial of T is $z^2 + 1$.
Two remarks. If $\mathbf{F} = \mathbf{C}$, then $z^2 + 1 = (z - i)(z + i)$, so by 5.27(a) the eigenvalues of T are i and $-i$, in agreement with Example 5.9 in the text. If $\mathbf{F} = \mathbf{R}$, then $z^2 + 1$ has no real zeros, so by 5.27(a) the operator T has no eigenvalues, in agreement with Exercise 5 above.

Problem 5B.7. (a) Give an example of $S, T \in \mathcal{L}(\mathbf{F}^2)$ such that the minimal polynomial of ST does not equal the minimal polynomial of TS.
(b) Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that if at least one of S, T is invertible, then the minimal polynomial of ST equals the minimal polynomial of TS.
[Hint: Show that if S is invertible and $p \in \mathcal{P}(\mathbf{F})$, then $p(TS) = S^{-1}p(ST)S$.]

Solution. (a) Define $S, T \in \mathcal{L}(\mathbf{F}^2)$ by

$$S(w, z) = (z, 0) \qquad \text{and} \qquad T(w, z) = (w, 0).$$

With respect to the standard basis e_1, e_2 of $\mathbf{F}^2$, the matrix of S has first row $0, 1$ and second row $0, 0$, while the matrix of T has first row $1, 0$ and second row $0, 0$.
Compute ST: for all $(w, z) \in \mathbf{F}^2$,

$$(ST)(w, z) = S(T(w, z)) = S(w, 0) = (0, 0).$$

Thus $ST = 0$, and the minimal polynomial of the zero operator is z (it is monic, it annihilates ST, and the only monic polynomial of degree 0 is 1, which applied to ST gives $I \neq 0$).
Compute TS: for all $(w, z) \in \mathbf{F}^2$,

$$(TS)(w, z) = T(S(w, z)) = T(z, 0) = (z, 0).$$

Thus $TS = S$. Now $TS \neq 0$ because $(TS)(0,1) = (1,0) \neq 0$, and

$$(TS)^2(w,z) = (TS)(z,0) = (0,0),$$

so $(TS)^2 = 0$. Hence the monic polynomial z^2 annihilates TS, while no monic polynomial of degree 1 does: if $TS = cI$ then applying both sides to $(1,0)$ gives $(0,0) = (c,0)$, so $c = 0$, contradicting $TS \neq 0$. Therefore the minimal polynomial of TS is z^2.
So the minimal polynomial of ST is z and the minimal polynomial of TS is z^2; these are not equal.
(b) Because the statement to be proved is symmetric in the letters S and T (interchanging the names S and T interchanges ST and TS, and asserts the same equality of minimal polynomials), we may assume without loss of generality that S is invertible.
We first prove the hint: if S is invertible and $p \in \mathcal{P}(\mathbf{F})$, then

$$p(TS) = S^{-1}p(ST)S.$$

Begin with powers. For $k = 0$ we have $S^{-1}(ST)^0S = S^{-1}IS = I = (TS)^0$. For an integer $k \geq 1$,

$$S^{-1}(ST)^kS = S^{-1}\underbrace{(ST)(ST)\cdots(ST)}_{k \text{ times}}S = (S^{-1}S)\underbrace{(TS)(TS)\cdots(TS)}_{k \text{ times}} = (TS)^k,$$

where the middle equality is just a regrouping of the $2k$ factors alternating between S and T, together with $S^{-1}S = I$. Thus

$$(TS)^k = S^{-1}(ST)^kS \qquad \text{for every integer } k \geq 0.$$

Now write $p(z) = a_0 + a_1z + \cdots + a_mz^m$. Using the equation above and the linearity of the maps $A \mapsto S^{-1}A$ and $A \mapsto AS$,

$$\begin{aligned} p(TS) &= \sum_{k=0}^{m} a_k(TS)^k = \sum_{k=0}^{m} a_kS^{-1}(ST)^kS \\ &= S^{-1}\Big(\sum_{k=0}^{m} a_k(ST)^k\Big)S = S^{-1}p(ST)S, \end{aligned}$$

which proves the hint.
Consequently, for every $p \in \mathcal{P}(\mathbf{F})$,

$$p(TS) = 0 \iff S^{-1}p(ST)S = 0 \iff p(ST) = 0,$$

where the second equivalence holds because S and S^{-1} are invertible: if $S^{-1}p(ST)S = 0$ then $p(ST) = S\big(S^{-1}p(ST)S\big)S^{-1} = 0$, and conversely if $p(ST) = 0$ then clearly $S^{-1}p(ST)S = 0$.
Thus the set of polynomials in $\mathcal{P}(\mathbf{F})$ that annihilate TS is exactly the same as the set of polynomials that annihilate ST. Because V is finite-dimensional, each of ST and TS has a minimal polynomial, namely the unique monic polynomial of smallest degree in its annihilating set (5.22). Since the two annihilating sets coincide, the unique monic polynomial of smallest degree in each is the same polynomial. Hence the minimal polynomial of ST equals the minimal polynomial of TS.
(Alternatively: if p is the minimal polynomial of ST, then $p(TS) = S^{-1}p(ST)S = 0$, so by 5.29 the minimal polynomial of TS divides p; the symmetric argument, using $(ST) = S(TS)S^{-1}$, shows that the minimal polynomial of ST divides the minimal polynomial of TS. Two monic polynomials each dividing the other are equal.)

Problem 5B.8. Suppose $T \in \mathcal{L}(\mathbb{R}^2)$ is the operator of counterclockwise rotation by 1°. Find the minimal polynomial of T.
[Because $\dim \mathbb{R}^2 = 2$, the degree of the minimal polynomial of T is at most 2. Thus the minimal polynomial of T is not the tempting polynomial $x^{180} + 1$, even though $T^{180} = -I$.]

Solution. Write $\theta = \pi/180$, so that T is counterclockwise rotation by the angle θ. With respect to the standard basis of $\mathbb{R}^2$, the matrix of T is

$$A = \mathcal{M}(T) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

We claim that the minimal polynomial of T is

$$z^2 - (2\cos\theta)\, z + 1 = z^2 - \left(2\cos\tfrac{\pi}{180}\right) z + 1.$$

First we check that this polynomial applied to T gives 0. Using the double-angle formulas,

$$A^2 = \begin{pmatrix} \cos^2\theta - \sin^2\theta & -2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{pmatrix} = \begin{pmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix},$$

which of course just says that rotating twice by θ is rotation by 2θ. Hence

$$A^2 - (2\cos\theta)A + I = \begin{pmatrix} \cos 2\theta - 2\cos^2\theta + 1 & -\sin 2\theta + 2\sin\theta\cos\theta \\ \sin 2\theta - 2\sin\theta\cos\theta & \cos 2\theta - 2\cos^2\theta + 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

because $\cos 2\theta = 2\cos^2\theta - 1$ and $\sin 2\theta = 2\sin\theta\cos\theta$. The map $S \mapsto \mathcal{M}(S)$ is injective and satisfies $\mathcal{M}(SU) = \mathcal{M}(S)\mathcal{M}(U)$ (see 3.43), so the matrix identity above gives

$$T^2 - (2\cos\theta)T + I = 0.$$

Now we check minimality. The minimal polynomial of T cannot have degree 0: the only monic polynomial of degree 0 is the constant polynomial 1, and applying it to T gives $I \neq 0$ because $\dim \mathbb{R}^2 = 2 > 0$. The minimal polynomial of T cannot have degree 1 either: a monic polynomial $z - \lambda$ satisfies $(T - \lambda I) = 0$ only if $T = \lambda I$, and T is not a scalar multiple of the identity because $\sin\theta = \sin\frac{\pi}{180} \neq 0$, so the entry in row 2, column 1 of A is nonzero while the corresponding entry of $\mathcal{M}(\lambda I)$ is 0. (Equivalently: T has no eigenvalues, since rotation by 1° maps no nonzero vector of $\mathbb{R}^2$ to a scalar multiple of itself, so by 5.27(a) the minimal polynomial of T has no real zeros and hence is not of the form $z - \lambda$.)
Therefore the minimal polynomial of T has degree exactly 2. Since $z^2 - (2\cos\theta)z + 1$ is a monic polynomial of degree 2 that when applied to T gives 0, the uniqueness statement in 5.22 shows that it is the minimal polynomial. Thus the minimal polynomial of T is

$$z^2 - \left(2\cos\tfrac{\pi}{180}\right) z + 1.$$

As a check on the remark in the statement: this polynomial has no real zeros, since its discriminant is $4\cos^2\theta - 4 < 0$; its complex zeros are $\cos\theta \pm i\sin\theta = e^{\pm i\pi/180}$. Also, $x^{180} + 1$ is indeed a polynomial multiple of the minimal polynomial (as it must be, by 5.29, since $T^{180} = -I$), because $e^{\pm i\pi/180}$ are zeros of $x^{180} + 1$.

Problem 5B.9. Suppose $T \in \mathcal{L}(V)$ is such that with respect to some basis of V, all entries of the matrix of T are rational numbers. Explain why all coefficients of the minimal polynomial of T are rational numbers.

Solution. Here V is finite-dimensional (otherwise T has no matrix and no minimal polynomial); let $n = \dim V$. If $n = 0$, then I is the zero operator on V and the minimal polynomial of T is the constant polynomial 1 (see the proof of 5.22), whose coefficients are rational, so the conclusion holds; assume from now on that $n \geq 1$. Fix the basis in the statement and let $A = \mathcal{M}(T)$ with respect to it, so every entry of A is rational. Let p be the minimal polynomial of T and let $m = \deg p$. Because $\mathcal{M}$ is an injective linear map satisfying $\mathcal{M}(SU) = \mathcal{M}(S)\mathcal{M}(U)$ (see 3.43), for a polynomial $q \in \mathcal{P}(\mathbf{F})$ we have $q(T) = 0$ if and only if $q(A) = 0$, where $q(A)$ is computed with $A^0 = I$. Note that A^k has rational entries for every $k \geq 0$, since products and sums of rational matrices have rational entries.
The one point requiring care is that the field $\mathbf{F}$ is $\mathbb{R}$ or $\mathbb{C}$, not $\mathbb{Q}$, so a priori the coefficients of p are

only known to lie in $\mathbf{F}$. The following standard fact bridges the gap.

Lemma. Let B be an N-by-k matrix with rational entries. If the equation $Bx = 0$ has a solution $x \in \mathbf{F}^k$ with $x \neq 0$, then it has a solution $x \in \mathbb{Q}^k$ with $x \neq 0$.

Proof of Lemma. Row reduce B to a matrix R in reduced row echelon form. All the row operations used (swapping two rows, multiplying a row by a nonzero scalar, adding a multiple of one row to another) can be carried out using only rational scalars, because at each stage the entries of the current matrix are rational; hence R has rational entries. Row operations do not change the solution set of the associated homogeneous system, over $\mathbb{Q}$ or over $\mathbf{F}$: each operation is invertible by another operation of the same kind. So $Bx = 0$ and $Rx = 0$ have the same solutions in $\mathbf{F}^k$ and also the same solutions in $\mathbb{Q}^k$. If every one of the k columns of R were a pivot column, then $Rx = 0$ would force $x = 0$ over any field containing $\mathbb{Q}$, contradicting the existence of a nonzero solution in $\mathbf{F}^k$. Hence some column of R is not a pivot column; setting that free variable equal to 1, the other free variables equal to 0, and solving for the pivot variables (which involves only the rational entries of R) produces a nonzero solution in $\mathbb{Q}^k$. This proves the Lemma.

Now regard the n-by-n matrices with rational entries as a vector space over $\mathbb{Q}$; it has dimension n^2 over $\mathbb{Q}$, since the matrices with a single entry equal to 1 and all other entries 0 form a basis. Thus the $n^2 + 1$ matrices $I, A, A^2, \ldots, A^{n^2}$ are linearly dependent over $\mathbb{Q}$. Let m' be the smallest nonnegative integer such that $I, A, \ldots, A^{m'}$ is linearly dependent over $\mathbb{Q}$; so $m' \leq n^2$. Choose rationals $b_0, \ldots, b_{m'}$, not all 0, with

$$b_0 I + b_1 A + \cdots + b_{m'} A^{m'} = 0.$$

If $b_{m'}$ were 0, this would exhibit a nontrivial rational dependence among $I, A, \ldots, A^{m'-1}$, contradicting the minimality of m' (and when $m' = 0$ it would force all coefficients to be 0). So $b_{m'} \neq 0$, and dividing by $b_{m'}$ we obtain a monic polynomial

$$q(z) = \frac{b_0}{b_{m'}} + \frac{b_1}{b_{m'}} z + \cdots + z^{m'}$$

with rational coefficients, of degree m', such that $q(A) = 0$ and hence $q(T) = 0$.

Because $q(T) = 0$ and $q \neq 0$, the polynomial q is a polynomial multiple of p by 5.29; a nonzero polynomial multiple of p has degree at least $\deg p$, so $m' \geq m$.

For the reverse inequality, apply the Lemma. Since $p(A) = 0$ and p is monic of degree m, the matrices $I, A, \ldots, A^m$ are linearly dependent over $\mathbf{F}$. Listing the entries of these $m+1$ matrices as the columns of an n^2-by-$(m+1)$ matrix B with rational entries, a nontrivial $\mathbf{F}$-linear dependence among $I, A, \ldots, A^m$ is exactly a nonzero solution $x \in \mathbf{F}^{m+1}$ of $Bx = 0$. By the Lemma there is a nonzero solution in $\mathbb{Q}^{m+1}$, which is a nontrivial rational linear dependence among $I, A, \ldots, A^m$. By the minimality of m' this gives $m' \leq m$.

Therefore $m' = m$, and q is a monic polynomial of degree m with rational coefficients satisfying $q(T) = 0$. By the uniqueness part of 5.22 (the minimal polynomial is the unique monic polynomial of smallest degree annihilating T), we conclude $q = p$. Hence all coefficients of the minimal polynomial of T are rational.

Problem 5B.10. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $v \in V$. Prove that

$$\operatorname{span}(v, Tv, \ldots, T^m v) = \operatorname{span}(v, Tv, \ldots, T^{\dim V - 1} v)$$

for all integers $m \geq \dim V - 1$.

Solution. Let $n = \dim V$. If $n = 0$, then $V = \{0\}$, so $v = 0$ and both sides of the asserted equality equal $\{0\}$ (the right side being the span of the empty list); thus the result is trivial in that case. Assume from now on that $n \geq 1$, so that the list $v, Tv, \ldots, T^{n-1} v$ is a genuine list. Fix an integer $m \geq n - 1$. Let k be the smallest nonnegative integer such that the list

$$v, Tv, \ldots, T^k v$$

is linearly dependent. Such a k exists: the list $v, Tv, \ldots, T^n v$ has length $n + 1 > \dim V$, so it is linearly

dependent by 2.22. Note also that $k \leq n$.
Set

$$U = \operatorname{span}(v, Tv, \ldots, T^{k-1}v),$$

where $U = \{0\}$ if $k = 0$ (the span of the empty list). By the minimality of k, the list $v, Tv, \ldots, T^{k-1}v$ is linearly independent.
Claim: $T^k v \in U$. Indeed, since $v, Tv, \ldots, T^k v$ is linearly dependent, there exist scalars $a_0, \ldots, a_k \in \mathbf{F}$, not all 0, with

$$a_0 v + a_1 Tv + \cdots + a_k T^k v = 0.$$

If $a_k = 0$, this would be a nontrivial linear dependence of $v, Tv, \ldots, T^{k-1}v$, contradicting the linear independence of that list (when $k = 0$ the displayed equation reads $a_0 v = 0$ with $a_0 \neq 0$, forcing $v = 0 \in U$, so the claim holds in that case too). Hence $a_k \neq 0$, and

$$T^k v = -\frac{1}{a_k}\left(a_0 v + a_1 Tv + \cdots + a_{k-1} T^{k-1} v\right) \in U,$$

proving the claim.
Next, U is invariant under T. To see this, it suffices to check that T maps each vector in the spanning list $v, Tv, \ldots, T^{k-1}v$ into U. For $0 \leq j \leq k-2$ we have $T(T^j v) = T^{j+1}v \in U$ because $j + 1 \leq k - 1$. For $j = k - 1$ we have $T(T^{k-1}v) = T^k v \in U$ by the claim. Since T is linear and U is the span of these vectors, $Tu \in U$ for every $u \in U$.
Now we show by induction that $T^i v \in U$ for every integer $i \geq 0$. For $0 \leq i \leq k - 1$ this holds by the definition of U (and for $k = 0$ the base case is $v = 0 \in U$, established above). If $T^i v \in U$ for some i, then $T^{i+1}v = T(T^i v) \in U$ because U is invariant under T. This completes the induction.
Finally we chain three inclusions. First,

$$\operatorname{span}(v, Tv, \ldots, T^{n-1}v) \subseteq \operatorname{span}(v, Tv, \ldots, T^m v),$$

because $m \geq n - 1$, so the second list contains every vector of the first list. Second,

$$\operatorname{span}(v, Tv, \ldots, T^m v) \subseteq U,$$

because $T^i v \in U$ for every i and U is a subspace. Third,

$$U = \operatorname{span}(v, Tv, \ldots, T^{k-1}v) \subseteq \operatorname{span}(v, Tv, \ldots, T^{n-1}v),$$

because $k \leq n$, so $k - 1 \leq n - 1$ and every vector spanning U appears in the list $v, Tv, \ldots, T^{n-1}v$ (this is trivially true when $k = 0$, since then $U = \{0\}$).
Combining the three inclusions, all four subspaces are equal; in particular

$$\operatorname{span}(v, Tv, \ldots, T^m v) = \operatorname{span}(v, Tv, \ldots, T^{\dim V - 1}v)$$

for every integer $m \geq \dim V - 1$, as desired.

Problem 5B.11. Suppose V is a two-dimensional vector space, $T \in \mathcal{L}(V)$, and the matrix of T with respect to some basis of V is $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$.
(a) Show that $T^2 - (a + d)T + (ad - bc)I = 0$.
(b) Show that the minimal polynomial of T equals

$$\begin{cases} z - a & \text{if } b = c = 0 \text{ and } a = d, \\ z^2 - (a + d)z + (ad - bc) & \text{otherwise.} \end{cases}$$

Solution. Let v_1, v_2 be the basis of V in question and let

$$A = \mathcal{M}(T) = \begin{pmatrix} a & c \\ b & d \end{pmatrix},$$

so $Tv_1 = av_1 + bv_2$ and $Tv_2 = cv_1 + dv_2$.
(a) By 3.43, $\mathcal{M}(T^2) = A^2$, and $\mathcal{M}$ is linear with $\mathcal{M}(I)$ equal to the 2-by-2 identity matrix. Because $\mathcal{M}$ is injective (a linear map whose matrix is 0 sends each basis vector to 0 and hence is the 0 map), it suffices to verify the corresponding matrix identity

$$A^2 - (a+d)A + (ad-bc)I = 0.$$

Multiplying out,

$$A^2 = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2+cb & ac+cd \\ ba+db & bc+d^2 \end{pmatrix},$$

so

$$A^2 - (a+d)A + (ad-bc)I = \begin{pmatrix} a^2+bc-a^2-ad+ad-bc & ac+cd-(a+d)c \\ ab+bd-(a+d)b & bc+d^2-ad-d^2+ad-bc \end{pmatrix},$$

and each of the four entries is 0. Hence $T^2 - (a+d)T + (ad-bc)I = 0$.
(b) Let p denote the minimal polynomial of T and put

$$g(z) = z^2 - (a+d)z + (ad-bc),$$

a monic polynomial of degree 2 with $g(T) = 0$ by (a).
First note that $\deg p \geq 1$: the only monic polynomial of degree 0 is the constant polynomial 1, and applying it to T gives I, which is not 0 because $\dim V = 2 > 0$. Also $\deg p \leq 2$, either by 5.22 ($\deg p \leq \dim V = 2$) or because $g(T) = 0$ forces p to divide g by 5.29. So $\deg p \in \{1, 2\}$.
Now $\deg p = 1$ means $p(z) = z - \lambda$ for some $\lambda \in \mathbf{F}$ with $T - \lambda I = 0$, that is, $T = \lambda I$. Since $\mathcal{M}$ is injective, $T = \lambda I$ holds if and only if

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix},$$

that is, if and only if $b = c = 0$ and $a = d = \lambda$.
Case 1: $b = c = 0$ and $a = d$. Then $T = aI$, so $z - a$ is a monic polynomial of degree 1 that when applied to T gives 0. Since $\deg p \geq 1$, the polynomial $z - a$ has the smallest possible degree among monic annihilating polynomials, and by the uniqueness in 5.22 we get $p(z) = z - a$. (Consistently with (a), in this case $g(z) = z^2 - 2az + a^2 = (z-a)^2$, a polynomial multiple of p, as 5.29 requires.)
Case 2: not ($b = c = 0$ and $a = d$). By the paragraph above, T is not of the form λI, so no monic polynomial of degree 1 annihilates T; combined with $\deg p \geq 1$ this gives $\deg p = 2$. Since g is monic of degree 2 with $g(T) = 0$, the uniqueness in 5.22 gives

$$p(z) = z^2 - (a+d)z + (ad-bc).$$

This is exactly the claimed formula.

Problem 5B.12. Define $T \in \mathcal{L}(\mathbf{F}^n)$ by $T(x_1, x_2, x_3, \ldots, x_n) = (x_1, 2x_2, 3x_3, \ldots, nx_n)$. Find the minimal polynomial of T.

Solution. The minimal polynomial of T is

$$(z-1)(z-2)(z-3)\cdots(z-n) = \prod_{j=1}^{n}(z-j).$$

Let $e_1, \dots, e_n$ denote the standard basis of $\mathbf{F}^n$. The definition of T says precisely that

$$Te_j = j\, e_j \qquad \text{for } j = 1, \dots, n,$$

so each e_j is an eigenvector of T with eigenvalue j. (Here $1, 2, \dots, n$ are viewed as elements of $\mathbf{F}$, and they are distinct because $\mathbf{F}$ is $\mathbb{R}$ or $\mathbb{C}$.)
Repeated application of T gives $T^k e_j = j^k e_j$ for every nonnegative integer k. Hence for every polynomial $q(z) = c_0 + c_1 z + \cdots + c_k z^k \in \mathcal{P}(\mathbf{F})$ we have

$$q(T)e_j = c_0 e_j + c_1 j e_j + \cdots + c_k j^k e_j = q(j)\, e_j$$

for $j = 1, \dots, n$. Because $e_1, \dots, e_n$ is a basis of $\mathbf{F}^n$ and $q(T)$ is linear, we conclude

$$q(T) = 0 \iff q(j) = 0 \text{ for each } j = 1, \dots, n,$$

that is, $q(T) = 0$ if and only if each of the n distinct numbers $1, 2, \dots, n$ is a zero of q.
Now let p denote the minimal polynomial of T. By the equivalence above, p has $1, 2, \dots, n$ among its zeros. Since $p \neq 0$ and a nonzero polynomial has at most $\deg p$ distinct zeros (see 4.8), it follows that $\deg p \geq n$.
On the other hand, set $g(z) = (z-1)(z-2)\cdots(z-n)$. This is a monic polynomial of degree n, and $g(j) = 0$ for each $j = 1, \dots, n$, so $g(T) = 0$ by the equivalence above. Hence the smallest degree of a monic polynomial annihilating T is at most n, so $\deg p \leq n$.
Therefore $\deg p = n$, and since g is monic of degree n with $g(T) = 0$, the uniqueness statement in 5.22 gives $p = g$. Thus the minimal polynomial of T is

$$(z-1)(z-2)\cdots(z-n).$$

(As a sanity check, note that this is consistent with 5.22 and 5.27: the degree n equals $\dim \mathbf{F}^n$, which is the largest degree allowed, and the zeros $1, 2, \dots, n$ of the minimal polynomial are exactly the eigenvalues of T, since T is a diagonal operator with diagonal entries $1, 2, \dots, n$.)

Problem 5B.13. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $p \in \mathcal{P}(\mathbf{F})$. Prove that there exists a unique $r \in \mathcal{P}(\mathbf{F})$ such that $p(T) = r(T)$ and $\deg r$ is less than the degree of the minimal polynomial of T.

Solution. Let q denote the minimal polynomial of T and let $m = \deg q$. (The minimal polynomial exists by 5.22, since V is finite-dimensional.)
Existence. By the division algorithm for polynomials (4.9) applied to p and q, there exist $s, r \in \mathcal{P}(\mathbf{F})$ such that

$$p = qs + r \quad \text{and} \quad \deg r < \deg q = m.$$

Applying this identity to T and using $q(T) = 0$, we get

$$p(T) = q(T)s(T) + r(T) = 0\, s(T) + r(T) = r(T).$$

Thus r has both required properties: $p(T) = r(T)$ and $\deg r < m$.
Uniqueness. Suppose $r_1, r_2 \in \mathcal{P}(\mathbf{F})$ both satisfy the conditions, so

$$p(T) = r_1(T) = r_2(T), \qquad \deg r_1 < m, \qquad \deg r_2 < m.$$

Let $u = r_1 - r_2$. Then

$$u(T) = r_1(T) - r_2(T) = 0,$$

and $\deg u < m$ (the degree of a difference of two polynomials is at most the maximum of their degrees). By 5.29, u is a polynomial multiple of q: there exists $s \in \mathcal{P}(\mathbf{F})$ with $u = qs$. If $s \neq 0$, then

$$\deg u = \deg q + \deg s \geq m,$$

contradicting $\deg u < m$. Hence $s = 0$, so $u = 0$ and therefore $r_1 = r_2$.
(Here we use Axler's convention that $\deg 0 = -\infty$, so the inequality $\deg u < m$ is automatic when $u = 0$ and the argument above is not circular: we derived $s = 0$ from the degree count, which applies whenever $s \neq 0$.)
This proves that there is exactly one $r \in \mathcal{P}(\mathbf{F})$ with $p(T) = r(T)$ and $\deg r$ less than the degree of the minimal polynomial of T; it is the remainder obtained on dividing p by the minimal polynomial of T.

Problem 5B.14. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$ has minimal polynomial $4 + 5z - 6z^2 - 7z^3 + 2z^4 + z^5$. Find the minimal polynomial of T^{-1}.

Solution. Write $p(z) = 4 + 5z - 6z^2 - 7z^3 + 2z^4 + z^5$ for the minimal polynomial of T. Its constant term is $4 \neq 0$, so T is invertible by 5.32; thus T^{-1} makes sense. We claim that the minimal polynomial of T^{-1} is the reversal of p, normalized to be monic:

$$\frac{1}{4} + \frac{1}{2}z - \frac{7}{4}z^2 - \frac{3}{2}z^3 + \frac{5}{4}z^4 + z^5.$$

The key observation is the following general fact.
Reversal Lemma. Suppose $S \in \mathcal{L}(V)$ is invertible and $u(z) = b_0 + b_1 z + \cdots + b_k z^k$ is a polynomial with $u(S) = 0$. Define the reversed polynomial

$$\hat{u}(z) = b_k + b_{k-1}z + \cdots + b_1 z^{k-1} + b_0 z^k.$$

Then $\hat{u}(S^{-1}) = 0$.
Proof of Lemma. Because S is invertible, so is S^k, and powers of S and S^{-1} commute with each other. We compute

$$\hat{u}(S^{-1}) = \sum_{j=0}^{k} b_{k-j} S^{-j} = S^{-k} \sum_{j=0}^{k} b_{k-j} S^{k-j} = S^{-k} \sum_{i=0}^{k} b_i S^i = S^{-k} u(S) = S^{-k} 0 = 0,$$

where in the third equality we substituted $i = k - j$. This proves the Lemma.
Now apply the Lemma with $S = T$, $u = p$, and $k = 5$. Since $p(T) = 0$,

$$\hat{p}(z) = 1 + 2z - 7z^2 - 6z^3 + 5z^4 + 4z^5$$

satisfies $\hat{p}(T^{-1}) = 0$. Dividing by the leading coefficient 4 (which is nonzero precisely because the constant term of p is nonzero), the monic polynomial

$$g(z) = \tfrac{1}{4}\hat{p}(z) = \frac{1}{4} + \frac{1}{2}z - \frac{7}{4}z^2 - \frac{3}{2}z^3 + \frac{5}{4}z^4 + z^5$$

has degree 5 and satisfies $g(T^{-1}) = 0$. Hence the minimal polynomial of T^{-1} has degree at most 5.
For the reverse inequality, let $\tilde{q}$ be the minimal polynomial of T^{-1} and let $k = \deg \tilde{q}$; so $k \leq 5$ by the previous paragraph. The operator T^{-1} is invertible (its inverse is T), so by 5.32 the constant term of $\tilde{q}$ is nonzero. Applying the Reversal Lemma with $S = T^{-1}$ and $u = \tilde{q}$, the reversed polynomial $\hat{\tilde{q}}$ satisfies

$$\hat{\tilde{q}}\left((T^{-1})^{-1}\right) = \hat{\tilde{q}}(T) = 0.$$

Moreover $\hat{\tilde{q}}$ has degree exactly k, because its coefficient of z^k is the constant term of $\tilde{q}$, which is nonzero. In particular $\hat{\tilde{q}} \neq 0$, so by 5.29 it is a nonzero polynomial multiple of p, whence

$$k = \deg \hat{\tilde{q}} \geq \deg p = 5.$$

Combining the two inequalities gives $k = 5$. Since g is a monic polynomial of degree 5 with $g(T^{-1}) = 0$, the uniqueness statement in 5.22 shows that g is the minimal polynomial of T^{-1}. Thus the minimal polynomial of T^{-1} is

$$\frac{1}{4} + \frac{1}{2}z - \frac{7}{4}z^2 - \frac{3}{2}z^3 + \frac{5}{4}z^4 + z^5.$$

The same computation proves the general statement: if T is invertible with minimal polynomial $a_0 + a_1 z + \cdots + a_{m-1}z^{m-1} + z^m$ (so $a_0 \neq 0$ by 5.32), then the minimal polynomial of T^{-1} is

$$\frac{1}{a_0}\left(1 + a_{m-1}z + a_{m-2}z^2 + \cdots + a_1 z^{m-1}\right) + z^m.$$

Problem 5B.15. Suppose V is a finite-dimensional complex vector space with $\dim V > 0$ and $T \in \mathcal{L}(V)$. Define $f\colon \mathbf{C} \to \mathbf{R}$ by

$$f(\lambda) = \dim \operatorname{range}(T - \lambda I).$$

Prove that f is not continuous.

Solution. Write $n = \dim V$, so $n \geq 1$. Throughout we use the fundamental theorem of linear maps (3.21), which for each $\lambda \in \mathbf{C}$ gives

$$f(\lambda) = \dim \operatorname{range}(T - \lambda I) = n - \dim \operatorname{null}(T - \lambda I).$$

First we record the two values that f can take near an eigenvalue.
If $\lambda \in \mathbf{C}$ is not an eigenvalue of T, then $T - \lambda I$ is injective, so $\dim \operatorname{null}(T - \lambda I) = 0$ and hence $f(\lambda) = n$.
If $\lambda \in \mathbf{C}$ is an eigenvalue of T, then by definition there is a nonzero vector in $\operatorname{null}(T - \lambda I)$, so $\dim \operatorname{null}(T - \lambda I) \geq 1$ and hence $f(\lambda) \leq n - 1$.
Because V is a nonzero finite-dimensional complex vector space, the operator T has an eigenvalue (5.19); call it λ_0. By the two paragraphs above, $f(\lambda_0) \leq n - 1$.
Also, T has at most n distinct eigenvalues (5.12), so the set E of eigenvalues of T is finite. Hence for every $\varepsilon > 0$ the punctured disk $\{\lambda \in \mathbf{C} : 0 < |\lambda - \lambda_0| < \varepsilon\}$, which is an infinite set, is not contained in the finite set E; thus it contains some λ that is not an eigenvalue of T. Choosing $\varepsilon = 1/k$ for $k = 1, 2, \ldots$ produces a sequence $\lambda_1, \lambda_2, \ldots$ of complex numbers with $\lambda_k \to \lambda_0$ and with no λ_k an eigenvalue of T. For each k we have $f(\lambda_k) = n$, so

$$\lim_{k \to \infty} f(\lambda_k) = n \neq f(\lambda_0),$$

the inequality holding because $f(\lambda_0) \leq n - 1 < n$. Thus f fails to be continuous at λ_0, and in particular f is not a continuous function.
Remark. An alternative phrasing of the same idea: f takes only values in the set $\{0, 1, \ldots, n\}$ of integers, and $\mathbf{C}$ is connected, so if f were continuous its image would be a connected subset of $\mathbf{R}$ consisting of integers, forcing f to be constant. But the computation above shows $f(\lambda_0) < n = f(\lambda)$ for any non-eigenvalue λ, so f is not constant. Note that both arguments use the existence of an eigenvalue, which is exactly where the hypotheses that $\mathbf{F} = \mathbf{C}$ and $\dim V > 0$ enter; for T the operator on $\mathbf{R}^2$ given by $T(x, y) = (-y, x)$, the analogous function on $\mathbf{R}$ is constantly 2 and hence continuous.

Problem 5B.16. Suppose $a_0, \ldots, a_{n-1} \in \mathbf{F}$. Let T be the operator on $\mathbf{F}^n$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & -a_2 \\ & & \ddots & & \vdots \\ & & & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{pmatrix}.$$

Here all entries of the matrix are 0 except for all 1's on the line under the diagonal and the entries in the last column (some of which might also be 0). Show that the minimal polynomial of T is the polynomial

$$a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n.$$

The matrix above is called the companion matrix of the polynomial above. This exercise shows that every monic polynomial is the minimal polynomial of some operator. Hence a formula or an algorithm that could produce exact eigenvalues for each operator on each $\mathbf{F}^n$ could then produce

exact zeros for each polynomial [by 5.27(a)]. Thus there is no such formula or algorithm. However, efficient numerical methods exist for obtaining very good approximations for the eigenvalues of an operator.

Solution. Let $e_1, \dots, e_n$ denote the standard basis of $\mathbf{F}^n$, and let

$$p(z) = a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n,$$

which is monic of degree n. Recall that the j-th column of $\mathcal{M}(T)$ lists the coefficients needed to write Te_j as a linear combination of $e_1, \dots, e_n$. Reading off the displayed matrix, we therefore have

$$Te_j = e_{j+1} \quad \text{for } j = 1, \dots, n-1,$$

and

$$Te_n = -a_0 e_1 - a_1 e_2 - \cdots - a_{n-1} e_n.$$

Step 1: the powers of T applied to e_1. From the first display, an easy induction gives

$$T^k e_1 = e_{k+1} \quad \text{for } k = 0, 1, \dots, n-1.$$

Indeed this holds for $k = 0$, and if it holds for some $k \le n-2$ then $T^{k+1} e_1 = T(T^k e_1) = Te_{k+1} = e_{k+2}$.
Step 2: $p(T) = 0$. Using Step 1 with $k = n-1$ and then the formula for Te_n,

$$T^n e_1 = T(T^{n-1} e_1) = Te_n = -a_0 e_1 - a_1 e_2 - \cdots - a_{n-1} e_n = -\big(a_0 e_1 + a_1 T e_1 + \cdots + a_{n-1} T^{n-1} e_1\big),$$

where the last equality again uses Step 1 to rewrite e_{k+1} as $T^k e_1$. Rearranging,

$$p(T) e_1 = a_0 e_1 + a_1 T e_1 + \cdots + a_{n-1} T^{n-1} e_1 + T^n e_1 = 0.$$

Now let $j \in \{1, \dots, n\}$. By Step 1, $e_j = T^{j-1} e_1$, and because powers of T commute with $p(T)$,

$$p(T) e_j = p(T) T^{j-1} e_1 = T^{j-1}\big(p(T) e_1\big) = T^{j-1}(0) = 0.$$

Thus $p(T)$ sends every vector of a basis of $\mathbf{F}^n$ to 0, so $p(T) = 0$.
Step 3: no monic polynomial of degree less than n annihilates T. Suppose $q \in \mathcal{P}(\mathbf{F})$ is monic with $\deg q = m < n$, say

$$q(z) = b_0 + b_1 z + \cdots + b_{m-1} z^{m-1} + z^m.$$

Applying $q(T)$ to e_1 and using Step 1 (legitimate because $m \le n-1$, so all the exponents $0, 1, \dots, m$ occurring here are at most $n-1$),

$$q(T) e_1 = b_0 e_1 + b_1 e_2 + \cdots + b_{m-1} e_m + e_{m+1}.$$

The right side is a linear combination of the linearly independent list $e_1, \dots, e_{m+1}$ in which the coefficient of e_{m+1} equals $1 \neq 0$; hence $q(T) e_1 \neq 0$ and so $q(T) \neq 0$.
Conclusion. Let r be the minimal polynomial of T. By Step 2 and 5.22 (which states that the minimal polynomial is the monic polynomial of smallest degree annihilating T), we have $\deg r \le n$. By Step 3, $\deg r \ge n$, since r is monic and $r(T) = 0$. Hence $\deg r = n$. Now r and p are both monic polynomials of degree n with $r(T) = p(T) = 0$, so the uniqueness assertion of 5.22 gives $r = p$. That is, the minimal polynomial of T is

$$a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n,$$

as desired. (Example 5.26 is the case $n = 5$, $a_0 = 3$, $a_1 = -6$, $a_2 = a_3 = a_4 = 0$.)

Problem 5B.17. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and p is the minimal polynomial of T. Suppose $\lambda \in \mathbf{F}$. Show that the minimal polynomial of $T - \lambda I$ is the polynomial q defined by

$q(z) = p(z + \lambda)$.

Solution. We first isolate the computational fact that drives everything.
Lemma. Suppose $A \in \mathcal{L}(V)$, $\mu \in \mathbf{F}$, and $r \in \mathcal{P}(\mathbf{F})$. Define $r_\mu \in \mathcal{P}(\mathbf{F})$ by $r_\mu(z) = r(z + \mu)$. Then:

- (i) if r is monic of degree m, then r_μ is monic of degree m;
- (ii) $r_\mu(A) = r(A + \mu I)$.

Proof of the lemma. Write $r(z) = c_0 + c_1 z + \cdots + c_m z^m$ with $c_m = 1$.
(i) Then $r_\mu(z) = \sum_{k=0}^{m} c_k (z + \mu)^k$. Expanding each $(z + \mu)^k$ by the binomial theorem, the only contribution of degree m comes from $c_m(z + \mu)^m$ and equals $c_m z^m = z^m$; every other term has degree at most $m - 1$. Hence r_μ has degree m and its coefficient of z^m is 1, so r_μ is monic of degree m.
(ii) As polynomials, $r_\mu = \sum_{k=0}^{m} c_k\, s^k$, where $s \in \mathcal{P}(\mathbf{F})$ is defined by $s(z) = z + \mu$ and s^k denotes the k-fold product of s with itself (with $s^0 = 1$). The map $\mathcal{P}(\mathbf{F}) \to \mathcal{L}(V)$ given by $r \mapsto r(A)$ is linear, as is immediate from the definition of $r(A)$ (5.14), and it is multiplicative by 5.17(a); applying 5.17(a) repeatedly gives $(s^k)(A) = \big(s(A)\big)^k = (A + \mu I)^k$. Therefore

$$r_\mu(A) = \sum_{k=0}^{m} c_k (A + \mu I)^k = r(A + \mu I),$$

which proves (ii) and the lemma.
Now let $m = \deg p$ and put $S = T - \lambda I$, so that $S + \lambda I = T$.
Step 1: q is monic of degree m and $q(S) = 0$. Since $q = p_\lambda$ in the notation of the lemma, part (i) shows q is monic with $\deg q = m$. By part (ii) applied with $A = S$, $\mu = \lambda$, $r = p$,

$$q(S) = p(S + \lambda I) = p(T) = 0,$$

the last equality because p is the minimal polynomial of T.
Step 2: no monic polynomial of degree smaller than m annihilates S. Let $\tilde{q}$ be the minimal polynomial of S and set $m' = \deg \tilde{q}$. Define $\tilde{p} \in \mathcal{P}(\mathbf{F})$ by $\tilde{p}(z) = \tilde{q}(z - \lambda)$, i.e. $\tilde{p} = \tilde{q}_{-\lambda}$. By the lemma (i), $\tilde{p}$ is monic of degree m', and by the lemma (ii) applied with $A = T$, $\mu = -\lambda$, $r = \tilde{q}$,

$$\tilde{p}(T) = \tilde{q}(T - \lambda I) = \tilde{q}(S) = 0.$$

So $\tilde{p}$ is a monic polynomial of degree m' that annihilates T; since p is the minimal polynomial of T, minimality of its degree (5.22) forces $m \le m'$.
Step 3: conclusion. By Step 1, q is a monic polynomial of degree m with $q(S) = 0$, so the minimality of $\deg \tilde{q}$ gives $m' \le m$. Combined with Step 2, $m' = m$. Thus q and $\tilde{q}$ are both monic polynomials of degree m that annihilate S, so the uniqueness assertion in 5.22 gives $q = \tilde{q}$. That is, the minimal polynomial of $T - \lambda I$ is the polynomial q defined by $q(z) = p(z + \lambda)$.
Remark. This is consistent with 5.27(a): the zeros of q are exactly the numbers z with $p(z + \lambda) = 0$, i.e. the numbers $\alpha - \lambda$ where α ranges over the eigenvalues of T, and these are precisely the eigenvalues of $T - \lambda I$.

Problem 5B.18. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and p is the minimal polynomial of T. Suppose $\lambda \in \mathbf{F} \setminus \{0\}$. Show that the minimal polynomial of λT is the polynomial q defined by

$$q(z) = \lambda^{\deg p}\, p\Big(\frac{z}{\lambda}\Big).$$

Solution. As in Exercise 17, we begin with a lemma that does all the work.
Lemma. Suppose $A \in \mathcal{L}(V)$, $\mu \in \mathbf{F} \setminus \{0\}$, and $r \in \mathcal{P}(\mathbf{F})$ is monic with $\deg r = m$. Define $r^{[\mu]} \in \mathcal{P}(\mathbf{F})$ by

$$r^{[\mu]}(z) = \mu^m\, r\Big(\frac{z}{\mu}\Big).$$

Then:

- (i) $r^{[\mu]}$ is a monic polynomial of degree m;
- (ii) $r^{[\mu]}(\mu A) = \mu^m\, r(A)$; in particular $r^{[\mu]}(\mu A) = 0$ if and only if $r(A) = 0$.

Proof of the lemma. Write $r(z) = c_0 + c_1 z + \cdots + c_m z^m$ with $c_m = 1$.
(i) For every z,

$$r^{[\mu]}(z) = \mu^m \sum_{k=0}^{m} c_k \frac{z^k}{\mu^k} = \sum_{k=0}^{m} c_k \mu^{m-k} z^k,$$

which makes sense because $\mu \neq 0$. So $r^{[\mu]}$ is a polynomial of degree at most m whose coefficient of z^m is $c_m \mu^0 = 1$. Hence $r^{[\mu]}$ is monic of degree m.
(ii) Applying the displayed formula for the coefficients of $r^{[\mu]}$ to the operator μA, and using $(\mu A)^k = \mu^k A^k$,

$$r^{[\mu]}(\mu A) = \sum_{k=0}^{m} c_k \mu^{m-k} (\mu A)^k = \sum_{k=0}^{m} c_k \mu^{m-k} \mu^k A^k = \mu^m \sum_{k=0}^{m} c_k A^k = \mu^m r(A).$$

Since $\mu \neq 0$ we have $\mu^m \neq 0$, so $r^{[\mu]}(\mu A) = 0$ exactly when $r(A) = 0$. This proves the lemma.
Now let $m = \deg p$, so that the polynomial q of the exercise is exactly $p^{[\lambda]}$.
Step 1: q is monic of degree m and $q(\lambda T) = 0$. By the lemma (i) with $r = p$ and $\mu = \lambda$, the polynomial $q = p^{[\lambda]}$ is monic of degree m. By the lemma (ii) with $A = T$,

$$q(\lambda T) = \lambda^m p(T) = \lambda^m \cdot 0 = 0,$$

since p is the minimal polynomial of T.
Step 2: comparison of degrees. Let $\tilde{q}$ be the minimal polynomial of λT and set $m' = \deg \tilde{q}$. By Step 1 and the minimality of $\deg \tilde{q}$ (5.22), we get $m' \leq m$.
For the reverse inequality, apply the lemma to the monic polynomial $\tilde{q}$, the scalar $\mu = 1/\lambda$ (nonzero because $\lambda \neq 0$), and the operator $A = \lambda T$. Part (i) says that

$$\tilde{p} := \tilde{q}^{[1/\lambda]}, \qquad \tilde{p}(z) = \lambda^{-m'} \tilde{q}(\lambda z),$$

is monic of degree m'; part (ii) says

$$\tilde{p}\big(\tfrac{1}{\lambda}(\lambda T)\big) = \lambda^{-m'}\, \tilde{q}(\lambda T) = 0.$$

Since $\frac{1}{\lambda}(\lambda T) = T$, this reads $\tilde{p}(T) = 0$. So $\tilde{p}$ is a monic polynomial of degree m' annihilating T, and because p is the minimal polynomial of T we conclude $m \leq m'$.
Step 3: conclusion. Steps 1 and 2 give $m' = m$. Thus q and $\tilde{q}$ are both monic polynomials of degree m with $q(\lambda T) = \tilde{q}(\lambda T) = 0$, so the uniqueness assertion of 5.22 gives $q = \tilde{q}$. Hence the minimal polynomial of λT is

$$q(z) = \lambda^{\deg p}\, p\Big(\frac{z}{\lambda}\Big),$$

as desired.
Remark. The hypothesis $\lambda \neq 0$ is needed: if $\lambda = 0$ and $\dim V > 0$, then $\lambda T = 0$ has minimal polynomial z, regardless of p. Note also that the zeros of q are exactly the numbers $\lambda\alpha$ with $p(\alpha) = 0$, matching the fact that the eigenvalues of λT are λ times the eigenvalues of T [consistent with 5.27(a)].

Problem 5B.19. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\mathcal{E}$ be the subspace of $\mathcal{L}(V)$ defined by

$$\mathcal{E} = \{q(T) : q \in \mathcal{P}(\mathbf{F})\}.$$

Prove that $\dim \mathcal{E}$ equals the degree of the minimal polynomial of T.

Solution. Let p be the minimal polynomial of T and let $m = \deg p$. We prove that the list

$$I,\ T,\ T^2,\ \ldots,\ T^{m-1}$$

(of length m; the empty list if $m = 0$) is a basis of $\mathcal{E}$, which gives $\dim \mathcal{E} = m$ as required.

Each of these operators lies in $\mathcal{E}$: taking $q(z) = z^k$ gives $q(T) = T^k$, and $q(z) = 1$ gives $q(T) = I$.
The list spans $\mathcal{E}$. Let $q \in \mathcal{P}(\mathbf{F})$ be arbitrary. By the division algorithm for polynomials (4.9), there exist $s, r \in \mathcal{P}(\mathbf{F})$ with

$$q = ps + r \quad \text{and} \quad \deg r < \deg p = m.$$

Applying these polynomials to T and using 5.17(a) together with $p(T) = 0$,

$$q(T) = p(T)s(T) + r(T) = 0 \cdot s(T) + r(T) = r(T).$$

Writing $r(z) = b_0 + b_1 z + \cdots + b_{m-1} z^{m-1}$ (some of the b_k possibly 0, which is legitimate since $\deg r \leq m - 1$), we obtain

$$q(T) = r(T) = b_0 I + b_1 T + \cdots + b_{m-1} T^{m-1}.$$

Thus every element of $\mathcal{E}$ is a linear combination of $I, T, \ldots, T^{m-1}$, so this list spans $\mathcal{E}$. (If $m = 0$, then p is the constant polynomial 1, so $I = p(T) = 0$; the display above then shows $q(T) = 0$ for all q, so $\mathcal{E} = \{0\}$ is spanned by the empty list. This degenerate case occurs exactly when $\dim V = 0$.)
The list is linearly independent. Suppose $c_0, \ldots, c_{m-1} \in \mathbf{F}$ satisfy

$$c_0 I + c_1 T + \cdots + c_{m-1} T^{m-1} = 0,$$

and suppose, seeking a contradiction, that not all the c_k equal 0. Let j be the largest index with $c_j \neq 0$, and define

$$r(z) = \frac{c_0}{c_j} + \frac{c_1}{c_j} z + \cdots + \frac{c_{j-1}}{c_j} z^{j-1} + z^j.$$

Then r is a monic polynomial with $\deg r = j \leq m - 1 < m$, and dividing the displayed equation by c_j shows $r(T) = 0$. This contradicts 5.22, which states that the minimal polynomial p is the monic polynomial of smallest degree with $p(T) = 0$. Hence all c_k equal 0, and the list is linearly independent.
Since $I, T, \ldots, T^{m-1}$ is a linearly independent spanning list of $\mathcal{E}$, it is a basis of $\mathcal{E}$, and therefore

$$\dim \mathcal{E} = m = \deg p,$$

the degree of the minimal polynomial of T.

Problem 5B.20. Suppose $T \in \mathcal{L}(\mathbf{F}^4)$ is such that the eigenvalues of T are $3, 5, 8$. Prove that

$$(T - 3I)^2 (T - 5I)^2 (T - 8I)^2 = 0.$$

Solution. Let p be the minimal polynomial of T. Since $\dim \mathbf{F}^4 = 4$, we have $\deg p \leq 4$ by 5.22. By 5.27(a), the zeros of p are exactly the eigenvalues of T, which by hypothesis are precisely the three numbers $3, 5, 8$.
Step 1: $(z - 3)(z - 5)(z - 8)$ divides p. First note $\deg p \geq 1$: a monic polynomial of degree 0 is the constant polynomial 1, which has no zeros, whereas $p(3) = 0$. So 4.6 applies to p: because 3 is a zero of p, we may write $p(z) = (z - 3)q(z)$ for some $q \in \mathcal{P}(\mathbf{F})$ with $\deg q = \deg p - 1$. Evaluating at 5 gives $0 = p(5) = (5 - 3)q(5) = 2q(5)$, so $q(5) = 0$. Since q is not the zero polynomial (its leading coefficient equals that of p, namely 1) and q has a zero, we get $\deg q \geq 1$, so 4.6 applies again: $q(z) = (z - 5)s(z)$ for some $s \in \mathcal{P}(\mathbf{F})$ with $\deg s = \deg q - 1$, and thus $p(z) = (z - 3)(z - 5)s(z)$. Evaluating at 8 gives $0 = p(8) = (8-3)(8-5)s(8) = 15\, s(8)$, so $s(8) = 0$; as before s has leading coefficient 1, hence $\deg s \geq 1$ and 4.6 gives $s(z) = (z - 8)t(z)$ for some $t \in \mathcal{P}(\mathbf{F})$. Therefore

$$p(z) = (z - 3)(z - 5)(z - 8)\, t(z).$$

Comparing highest-degree coefficients and using that p is monic, t is monic; and $\deg t = \deg p - 3 \leq 4 - 3 = 1$.
Step 2: p divides $(z - 3)^2 (z - 5)^2 (z - 8)^2$. There are two cases.
If $\deg t = 0$, then t is the constant polynomial 1, so $p(z) = (z - 3)(z - 5)(z - 8)$ and

$$(z - 3)^2 (z - 5)^2 (z - 8)^2 = p(z) \big[(z - 3)(z - 5)(z - 8)\big],$$

a polynomial multiple of p.
If $\deg t = 1$, then, t being monic, $t(z) = z - c$ for some $c \in \mathbf{F}$. Then c is a zero of p, hence an eigenvalue of T by 5.27(a), so $c \in \{3, 5, 8\}$. Let d and e be the other two elements of $\{3, 5, 8\}$, so that

$$p(z) = (z - c)^2(z - d)(z - e).$$

Then

$$(z - 3)^2(z - 5)^2(z - 8)^2 = (z - c)^2(z - d)^2(z - e)^2 = p(z)\big[(z - d)(z - e)\big],$$

again a polynomial multiple of p.
Step 3: conclusion. In both cases the polynomial r defined by $r(z) = (z - 3)^2(z - 5)^2(z - 8)^2$ is a polynomial multiple of the minimal polynomial of T. Hence $r(T) = 0$ by 5.29. Since

$$r(T) = (T - 3I)^2(T - 5I)^2(T - 8I)^2$$

by 5.17(a), we conclude that $(T - 3I)^2(T - 5I)^2(T - 8I)^2 = 0$, as desired.
Remark. The squares cannot simply be dropped. The operator $(T - 3I)(T - 5I)(T - 8I)$ need not be 0: if $p(z) = (z - 3)^2(z - 5)(z - 8)$, then $(z - 3)(z - 5)(z - 8)$ is a nonzero polynomial of degree 3 that is not a polynomial multiple of p, so by 5.29 the corresponding operator is nonzero. The point of the exercise is that squaring all three factors covers whichever of the three eigenvalues is the repeated one in p, without our having to know which it is.

Problem 5B.21. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that the minimal polynomial of T has degree at most $1 + \dim \operatorname{range} T$.
If $\dim \operatorname{range} T < \dim V - 1$, then this exercise gives a better upper bound than 5.22 for the degree of the minimal polynomial of T.

Solution. Let $U = \operatorname{range} T$ and let $r = \dim U$. Note that U is invariant under T: if $u \in U$, then $Tu \in \operatorname{range} T = U$. Thus $T|_U \in \mathcal{L}(U)$ makes sense.
Step 1: powers of T agree with powers of $T|_U$ on U. We claim that

$$T^k u = (T|_U)^k u \quad \text{for all } u \in U \text{ and all integers } k \geq 0.$$

This holds for $k = 0$ (both sides equal u). If it holds for some $k \geq 0$, then for $u \in U$ we have $(T|_U)^k u \in U$ (because U is invariant under T, so $T|_U$ maps U into U), and hence

$$T^{k+1}u = T\big(T^k u\big) = T\big((T|_U)^k u\big) = (T|_U)\big((T|_U)^k u\big) = (T|_U)^{k+1}u,$$

where the third equality holds because T and $T|_U$ agree on the vector $(T|_U)^k u$ of U. This completes the induction. Consequently, for every polynomial $q \in \mathcal{P}(\mathbf{F})$,

$$q(T)u = q(T|_U)u \qquad \text{for all } u \in U,$$

since $q(T)u$ and $q(T|_U)u$ are the same linear combination of the vectors $T^k u = (T|_U)^k u$.
Step 2: construction of a monic annihilating polynomial of degree at most $1 + r$. Let q be the minimal polynomial of the operator $T|_U \in \mathcal{L}(U)$. By 5.22 applied to $T|_U$ on the finite-dimensional vector space U, the polynomial q is monic with

$$\deg q \leq \dim U = r.$$

Define $s \in \mathcal{P}(\mathbf{F})$ by $s(z) = z\, q(z)$. Then s is monic (the product of monic polynomials is monic) and

$$\deg s = 1 + \deg q \leq 1 + r.$$

We claim $s(T) = 0$. Indeed, by 5.17(a) we have $s(T) = q(T)\, T$, so for every $v \in V$,

$$s(T)v = q(T)(Tv) = q(T|_U)(Tv) = 0,$$

where the second equality uses the last display of Step 1 with $u = Tv \in \operatorname{range} T = U$, and the last equality holds because $q(T|_U) = 0$ by the definition of the minimal polynomial of $T|_U$. Hence $s(T) = 0$.

Step 3: conclusion. Let p be the minimal polynomial of T. By 5.22, p is the monic polynomial of smallest degree with $p(T) = 0$; since s is a monic polynomial with $s(T) = 0$, we conclude

$$\deg p \le \deg s \le 1 + r = 1 + \dim \operatorname{range} T,$$

as desired.
Remark on the two bounds. If T is not injective, then $\dim \operatorname{range} T \le \dim V - 1$ and this bound is at least as good as the bound $\deg p \le \dim V$ from 5.22; it is strictly better when $\dim \operatorname{range} T < \dim V - 1$. For example, if $T \neq 0$ has rank 1 on a space of dimension 100, then the minimal polynomial of T has degree at most 2, whereas 5.22 only gives 100. The bound is sharp: for $T = 0$ on a nonzero space the minimal polynomial is z, of degree $1 = 1 + \dim \operatorname{range} T$.

Problem 5B.22. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if $I \in \operatorname{span}(T, T^2, \dots, T^{\dim V})$.

Solution. Let $n = \dim V$. If $n = 0$ the assertion is trivial: the only operator on $V = \{0\}$ is 0, which equals I and is invertible, while the list $T, T^2, \dots, T^{\dim V}$ is empty and the span of the empty list is $\{0\}$, which contains $I = 0$. So both sides of the equivalence hold. Assume from now on that $n \ge 1$.
First suppose T is invertible. Let

$$p(z) = c_0 + c_1 z + \cdots + c_{m-1} z^{m-1} + z^m$$

be the minimal polynomial of T; by 5.22 it exists and $m = \deg p \le n$. First note $m \ge 1$: if $m = 0$ then p is the constant polynomial 1, so $p(T) = I$, and $p(T) = 0$ would force $I = 0$ and hence $V = \{0\}$, a case already excluded. Because T is invertible, 5.32 tells us that the constant term c_0 of p is nonzero. Now $p(T) = 0$ means

$$c_0 I + c_1 T + \cdots + c_{m-1} T^{m-1} + T^m = 0,$$

and dividing by c_0 and rearranging gives

$$I = -\frac{c_1}{c_0} T - \cdots - \frac{c_{m-1}}{c_0} T^{m-1} - \frac{1}{c_0} T^m.$$

The right side is a linear combination of $T, T^2, \dots, T^m$ (when $m = 1$ it reduces to $I = -\frac{1}{c_0} T$). Since $m \le n$, this exhibits I as an element of $\operatorname{span}(T, T^2, \dots, T^n)$, as desired.
Conversely, suppose $I \in \operatorname{span}(T, T^2, \dots, T^n)$. Then there exist $a_1, \dots, a_n \in \mathbf{F}$ with

$$I = a_1 T + a_2 T^2 + \cdots + a_n T^n.$$

Set $S = a_1 I + a_2 T + \cdots + a_n T^{n-1} \in \mathcal{L}(V)$. Because every power of T commutes with T, factoring T out of the right side above on either side gives

$$TS = a_1 T + a_2 T^2 + \cdots + a_n T^n = I \quad \text{and} \quad ST = a_1 T + a_2 T^2 + \cdots + a_n T^n = I.$$

Thus S is an inverse of T, so T is invertible.
(Alternatively, for the converse one may argue: if $Tv = 0$ then $T^k v = 0$ for every $k \ge 1$, so $v = Iv = a_1 Tv + \cdots + a_n T^n v = 0$; hence T is injective and therefore, V being finite-dimensional, invertible by 3.65.)

Problem 5B.23. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $n = \dim V$. Prove that if $v \in V$, then $\operatorname{span}(v, Tv, \dots, T^{n-1} v)$ is invariant under T.

Solution. Let $v \in V$ and put

$$U = \operatorname{span}(v, Tv, \dots, T^{n-1} v).$$

(If $n = 0$ then $V = \{0\}$, $U = \{0\}$, and there is nothing to prove; so assume $n \ge 1$.) Since $T(T^k v) =$

$T^{k+1}v$, which lies in U whenever $k \leq n-2$, the only thing that must be shown is

$$T^n v \in U.$$

Let m be the smallest nonnegative integer such that

$$T^m v \in \operatorname{span}(v, Tv, \ldots, T^{m-1}v)$$

(for $m = 0$ this condition reads $v \in \operatorname{span}() = \{0\}$, i.e. $v = 0$). Such an m exists with $m \leq n$: indeed the list $v, Tv, \ldots, T^n v$ has length $n+1 > \dim V$, so it is linearly dependent by 2.22, and hence by the linear dependence lemma 2.19 one of its terms lies in the span of the terms preceding it.
Set $W = \operatorname{span}(v, Tv, \ldots, T^{m-1}v)$. I claim W is invariant under T. Indeed, for $0 \leq j \leq m-2$ we have $T(T^j v) = T^{j+1}v \in W$ by inspection, and $T(T^{m-1}v) = T^m v \in W$ by the defining property of m. Since W is spanned by $v, Tv, \ldots, T^{m-1}v$ and T maps each of these into W, linearity gives $TW \subseteq W$. (When $m = 0$ we have $W = \{0\}$ and $v = 0$, so W is trivially invariant.)
Because $v \in W$ when $m \geq 1$, and $v = 0 \in W$ when $m = 0$, invariance of W gives by induction that $T^k v \in W$ for every integer $k \geq 0$.
Now two cases.
Case 1: $m \leq n-1$. Then the spanning list of W is a sublist of the spanning list of U, so $W \subseteq U$; and by the previous paragraph each of $v, Tv, \ldots, T^{n-1}v$ lies in W, so $U \subseteq W$. Hence $U = W$, which we showed is invariant under T. In particular $T^n v \in W = U$.
Case 2: $m = n$. Minimality of m says that for every k with $0 \leq k \leq n-1$ we have $T^k v \notin \operatorname{span}(v, \ldots, T^{k-1}v)$. By the linear dependence lemma 2.19, this means the list $v, Tv, \ldots, T^{n-1}v$ is linearly independent. It has length $n = \dim V$, so it is a basis of V (by 2.38), and therefore $U = V$. Of course V is invariant under T, and in particular $T^n v \in V = U$.
In both cases U is invariant under T, as desired.

Problem 5B.24. Suppose V is a finite-dimensional complex vector space. Suppose $T \in \mathcal{L}(V)$ is such that 5 and 6 are eigenvalues of T and that T has no other eigenvalues. Prove that $(T-5I)^{\dim V-1}(T-6I)^{\dim V-1} = 0$.

Solution. Let $n = \dim V$ and let p be the minimal polynomial of T; by 5.22, p is monic and $\deg p \leq n$. Because T has two distinct eigenvalues, 5.12 (an operator on V has at most $\dim V$ distinct eigenvalues) gives $n \geq 2$; thus $n - 1 \geq 1$.
Because $\mathbf{F} = \mathbf{C}$, part (b) of 5.27 says that p factors as a product of linear factors,

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m),$$

and by part (a) of 5.27 the zeros of p are exactly the eigenvalues of T, namely 5 and 6. Hence each $\lambda_i \in \{5, 6\}$, and both values actually occur. Therefore

$$p(z) = (z-5)^j (z-6)^k$$

for some integers $j, k \geq 1$ with $j + k = \deg p \leq n$.
From $j + k \leq n$ and $k \geq 1$ we get $j \leq n-1$; symmetrically $k \leq n-1$. Consequently the exponents $n-1-j$ and $n-1-k$ are nonnegative integers, and

$$(z-5)^{n-1}(z-6)^{n-1} = (z-5)^j (z-6)^k \cdot (z-5)^{n-1-j}(z-6)^{n-1-k} = p(z)\, q(z),$$

where $q(z) = (z-5)^{n-1-j}(z-6)^{n-1-k} \in \mathcal{P}(\mathbf{C})$.
Thus $(z-5)^{n-1}(z-6)^{n-1}$ is a polynomial multiple of the minimal polynomial of T, so by 5.29 evaluating it at T gives 0. Explicitly,

$$(T-5I)^{n-1}(T-6I)^{n-1} = p(T)\, q(T) = 0 \cdot q(T) = 0,$$

as desired.

Problem 5B.25. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is a subspace of V that is invariant under T.
(a) Prove that the minimal polynomial of T is a polynomial multiple of the minimal polynomial of the quotient operator T/U.
(b) Prove that

$$\text{(minimal polynomial of } T|_U) \times \text{(minimal polynomial of } T/U)$$

is a polynomial multiple of the minimal polynomial of T.
[The quotient operator T/U was defined in Exercise 38 in Section 5A.]

Solution. Recall from Exercise 38 in Section 5A that $T/U \in \mathcal{L}(V/U)$ is defined by

$$(T/U)(v+U) = Tv + U \qquad (v \in V),$$

which makes sense precisely because U is invariant under T. Note that V/U is finite-dimensional (indeed $\dim V/U = \dim V - \dim U$ by 3.105), and U is finite-dimensional, so $T|_U$ and T/U both have minimal polynomials, by 5.22.
We first record two computational facts.
Fact 1. For every nonnegative integer k and every $v \in V$ we have $(T/U)^k(v+U) = T^k v + U$. This holds for $k = 0$ trivially, and if it holds for k then

$$(T/U)^{k+1}(v+U) = (T/U)\big(T^k v + U\big) = T^{k+1}v + U,$$

completing the induction. Because the quotient map $\pi\colon v \mapsto v+U$ is linear (3.104), it follows that for every $q \in \mathcal{P}(\mathbf{F})$,

$$q(T/U)(v+U) = q(T)v + U \qquad \text{for all } v \in V.$$

Fact 2. For every $q \in \mathcal{P}(\mathbf{F})$ and every $u \in U$ we have $q(T)u = q(T|_U)u$. Indeed $T^k u = (T|_U)^k u$ for all $k \geq 0$ (by induction, using that U is invariant under T so all these vectors stay in U), and then take linear combinations.
(a) Let p be the minimal polynomial of T, so $p(T) = 0$. For each $v \in V$, Fact 1 gives

$$p(T/U)(v+U) = p(T)v + U = 0 + U,$$

which is the zero vector of V/U. Since every element of V/U has the form $v+U$, this says $p(T/U) = 0$. By 5.29 applied to the operator T/U, the polynomial p is a polynomial multiple of the minimal polynomial of T/U.
(b) Let r be the minimal polynomial of $T|_U$ and let s be the minimal polynomial of T/U. By 5.29 (applied to T) it suffices to prove that

$$(rs)(T) = 0.$$

Let $v \in V$. Since $s(T/U) = 0$, Fact 1 gives

$$s(T)v + U = s(T/U)(v+U) = 0 + U,$$

which means

$$s(T)v \in U.$$

Now apply $r(T)$ to this vector. Because $s(T)v \in U$, Fact 2 gives

$$r(T)\big(s(T)v\big) = r(T|_U)\big(s(T)v\big) = 0,$$

the last equality because $r(T|_U) = 0$ by definition of the minimal polynomial of $T|_U$.
Since $(rs)(T) = r(T)s(T)$ (multiplicativity of the polynomial functional calculus, 5.17), we have shown $(rs)(T)v = 0$ for every $v \in V$, i.e. $(rs)(T) = 0$. By 5.29, rs is a polynomial multiple of the minimal polynomial of T, as desired.

Problem 5B.26. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is a subspace of V that is invariant under T. Prove that the set of eigenvalues of T equals the union of the set of eigenvalues of $T|_U$ and the set of eigenvalues of T/U.

Solution. Write p for the minimal polynomial of T, r for the minimal polynomial of $T|_U$, and s for the minimal polynomial of T/U. All three exist by 5.22, because V, U, and V/U are finite-dimensional. All three are monic, hence nonzero.
For a polynomial q, let $Z(q) \subseteq \mathbf{F}$ denote the set of zeros of q in $\mathbf{F}$. By 5.27(a),

$$Z(p) = \{\text{eigenvalues of } T\}, \quad Z(r) = \{\text{eigenvalues of } T|_U\}, \quad Z(s) = \{\text{eigenvalues of } T/U\}.$$

So it suffices to prove $Z(p) = Z(r) \cup Z(s)$.
Step 1: $Z(r) \cup Z(s) \subseteq Z(p)$. By 5.31, p is a polynomial multiple of r: write $p = rg$ for some $g \in \mathcal{P}(\mathbf{F})$. If $\lambda \in Z(r)$ then $p(\lambda) = r(\lambda)g(\lambda) = 0$, so $\lambda \in Z(p)$. By part (a) of Exercise 25 above, p is likewise a polynomial multiple of s, say $p = sh$; if $\lambda \in Z(s)$ then $p(\lambda) = s(\lambda)h(\lambda) = 0$, so $\lambda \in Z(p)$. Hence $Z(r) \cup Z(s) \subseteq Z(p)$.
Step 2: $Z(p) \subseteq Z(r) \cup Z(s)$. By part (b) of Exercise 25 above, the product rs is a polynomial multiple of p: write $rs = pk$ for some $k \in \mathcal{P}(\mathbf{F})$. Suppose $\lambda \in Z(p)$. Then

$$r(\lambda)s(\lambda) = (rs)(\lambda) = p(\lambda)k(\lambda) = 0.$$

Because $\mathbf{F}$ ($= \mathbf{R}$ or $\mathbf{C}$) has no zero divisors, $r(\lambda) = 0$ or $s(\lambda) = 0$, i.e. $\lambda \in Z(r) \cup Z(s)$.
Combining the two steps, $Z(p) = Z(r) \cup Z(s)$, which by the identification above says exactly that the set of eigenvalues of T is the union of the set of eigenvalues of $T|_U$ and the set of eigenvalues of T/U.
(Degenerate cases are covered by the argument: if $U = \{0\}$ then $T|_U$ is the operator on the zero space, whose minimal polynomial is the constant polynomial 1, with no zeros and correspondingly no eigenvalues; similarly if $U = V$ then T/U has minimal polynomial 1.)

Problem 5B.27. Suppose $\mathbf{F} = \mathbf{R}$, V is finite-dimensional, and $T \in \mathcal{L}(V)$. Prove that the minimal polynomial of $T_{\mathbf{C}}$ equals the minimal polynomial of T.
[The complexification $T_{\mathbf{C}}$ was defined in Exercise 33 of Section 3B.]

Solution. Recall the definitions. The complexification $V_{\mathbf{C}}$ (Exercise 8 in Section 1B) is $V \times V$ with elements written $u + iv$ ($u, v \in V$), with

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2), \qquad (a + bi)(u + iv) = (au - bv) + i(av + bu)$$

for $a, b \in \mathbf{R}$; this is a complex vector space. The complexification of T is $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$ defined by $T_{\mathbf{C}}(u + iv) = Tu + iTv$ (Exercise 33 in Section 3B).
$V_{\mathbf{C}}$ is finite-dimensional. Indeed, if $v_1, \ldots, v_n$ is a basis of V over $\mathbf{R}$, then $v_1, \ldots, v_n$ (viewed as $v_j + i0$) is a basis of $V_{\mathbf{C}}$ over $\mathbf{C}$: it spans because $u + iv = \sum_j (a_j + ib_j)v_j$ when $u = \sum a_j v_j$ and $v = \sum b_j v_j$, and it is linearly independent because $\sum_j (a_j + ib_j)v_j = \big(\sum_j a_j v_j\big) + i\big(\sum_j b_j v_j\big) = 0$ forces $\sum a_j v_j = 0$ and $\sum b_j v_j = 0$, hence all $a_j = b_j = 0$. Thus $T_{\mathbf{C}}$ has a minimal polynomial $q \in \mathcal{P}(\mathbf{C})$, by 5.22. Let $p \in \mathcal{P}(\mathbf{R})$ be the minimal polynomial of T.
Key computation. If $g \in \mathcal{P}(\mathbf{R})$ is a polynomial with real coefficients, then

$$g(T_{\mathbf{C}})(u + iv) = g(T)u + i\, g(T)v \qquad \text{for all } u, v \in V. \quad (\dagger)$$

To see this, first note $(T_{\mathbf{C}})^k(u + iv) = T^k u + iT^k v$ for every $k \geq 0$, by induction on k (the case $k = 0$ is clear, and applying $T_{\mathbf{C}}$ to $T^k u + iT^k v$ gives $T^{k+1}u + iT^{k+1}v$). Since g has real coefficients $a_0, \ldots, a_d$, and real scalar multiplication on $V_{\mathbf{C}}$ satisfies $a(x + iy) = ax + iay$, summing gives ($\dagger$).
p is a polynomial multiple of q. Taking $g = p$ in ($\dagger$) and using $p(T) = 0$ gives $p(T_{\mathbf{C}})(u + iv) = 0 + i0 = 0$ for all $u, v \in V$, i.e. $p(T_{\mathbf{C}}) = 0$. By 5.29 applied to $T_{\mathbf{C}}$, the polynomial p (viewed in $\mathcal{P}(\mathbf{C})$) is a polynomial multiple of q.

q is a polynomial multiple of p. Write $q = q_1 + iq_2$ where $q_1, q_2 \in \mathcal{P}(\mathbf{R})$ are obtained by taking the real and imaginary parts of the coefficients of q. Fix $u \in V$ and consider the element $u + i0 \in V_{\mathbf{C}}$. Using (†) for the real polynomials q_1 and q_2, and then multiplying by the complex scalar i (which sends $x + iy$ to $-y + ix$),

$$q(T_{\mathbf{C}})(u+i0) = q_1(T_{\mathbf{C}})(u+i0) + i\, q_2(T_{\mathbf{C}})(u+i0) = \big(q_1(T)u + i0\big) + i\big(q_2(T)u + i0\big) = q_1(T)u + i\, q_2(T)u.$$

Since $q(T_{\mathbf{C}}) = 0$, the left side is $0 = 0 + i0$, and equality in $V_{\mathbf{C}} = V \times V$ is coordinatewise, so

$$q_1(T)u = 0 \quad \text{and} \quad q_2(T)u = 0.$$

As $u \in V$ was arbitrary, $q_1(T) = 0$ and $q_2(T) = 0$. By 5.29 applied to T, both q_1 and q_2 are polynomial multiples of p: say $q_1 = pg_1$ and $q_2 = pg_2$ with $g_1, g_2 \in \mathcal{P}(\mathbf{R})$. Hence

$$q = q_1 + iq_2 = p\,(g_1 + ig_2),$$

so q is a polynomial multiple of p in $\mathcal{P}(\mathbf{C})$.
Conclusion. We have $q = ph$ and $p = qg$ for some $g, h \in \mathcal{P}(\mathbf{C})$. Then $p = phg$, and since $p \neq 0$ (it is monic) this forces $hg = 1$, so h is a nonzero constant. Comparing leading coefficients in $q = ph$ and using that both p and q are monic gives $h = 1$. Therefore

$$q = p,$$

i.e. the minimal polynomial of $T_{\mathbf{C}}$ equals the minimal polynomial of T.

Problem 5B.28. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that the minimal polynomial of $T' \in \mathcal{L}(V')$ equals the minimal polynomial of T.
[The dual map T' was defined in Section 3F.]

Solution. Recall from 3.118 that for $T \in \mathcal{L}(V)$ the dual map is $T' \in \mathcal{L}(V')$ given by $T'(\varphi) = \varphi \circ T$ for $\varphi \in V'$. Since V is finite-dimensional, so is V', with $\dim V' = \dim V$ (by 3.111); hence T' has a minimal polynomial, by 5.22.
Step 1: $g(T') = \big(g(T)\big)'$ for every $g \in \mathcal{P}(\mathbf{F})$.
First, $(T^k)' = (T')^k$ for every nonnegative integer k. For $k = 0$ this says $I' = I$, where the left I is the identity on V and the right I is the identity on V'; that is true because $I'(\varphi) = \varphi \circ I = \varphi$. For the inductive step, 3.120(c) with $S = T^k$ and the map T gives

$$(T^{k+1})' = (T^k T)' = T'\,(T^k)' = T'(T')^k = (T')^{k+1}.$$

Next, the map $S \mapsto S'$ from $\mathcal{L}(V)$ to $\mathcal{L}(V')$ is linear, by parts (a) and (b) of 3.120. So if $g(z) = a_0 + a_1 z + \cdots + a_d z^d$ with $a_j \in \mathbf{F}$, then

$$\big(g(T)\big)' = \big(a_0 I + a_1 T + \cdots + a_d T^d\big)' = a_0 I' + a_1 T' + \cdots + a_d (T^d)' = a_0 I + a_1 T' + \cdots + a_d (T')^d = g(T').$$

Step 2: for $S \in \mathcal{L}(V)$ we have $S = 0$ if and only if $S' = 0$.
By 3.130(a), $\dim \operatorname{range} S' = \dim \operatorname{range} S$. Now $S = 0$ if and only if $\dim \operatorname{range} S = 0$, and $S' = 0$ if and only if $\dim \operatorname{range} S' = 0$; the displayed equality of dimensions makes these two conditions equivalent.
(Directly: if $S \neq 0$, pick $v \in V$ with $Sv \neq 0$; extending Sv to a basis of V and taking the first element of the dual basis produces $\varphi \in V'$ with $\varphi(Sv) = 1$, so $(S'\varphi)(v) = \varphi(Sv) \neq 0$ and $S' \neq 0$. Conversely if $S = 0$ then $S'\varphi = \varphi \circ 0 = 0$ for all φ.)
Step 3: conclusion.
Combining Steps 1 and 2, for every $g \in \mathcal{P}(\mathbf{F})$,

$$g(T) = 0 \iff \big(g(T)\big)' = 0 \iff g(T') = 0.$$

Thus T and T' are annihilated by exactly the same polynomials. The minimal polynomial of an operator is the unique monic polynomial of smallest degree annihilating it (5.22), and that description

depends only on the set of annihilating polynomials. Hence the minimal polynomial of T' equals the minimal polynomial of T.
(Spelled out: let p be the minimal polynomial of T and $\tilde{p}$ that of T'. Since $p(T) = 0$ we get $p(T') = 0$, so $\deg \tilde{p} \le \deg p$; since $\tilde{p}(T') = 0$ we get $\tilde{p}(T) = 0$, so $\deg p \le \deg \tilde{p}$. Thus $\deg p = \deg \tilde{p}$, and both are monic polynomials of that degree annihilating T; uniqueness in 5.22 gives $\tilde{p} = p$.)

Problem 5B.29. Show that every operator on a finite-dimensional vector space of dimension at least two has an invariant subspace of dimension two.
[Exercise 6 in Section 5C will give an improvement of this result when $\mathbf{F} = \mathbf{C}$.]

Solution. Throughout, V is a finite-dimensional vector space over $\mathbf{F}$ (which is $\mathbf{R}$ or $\mathbf{C}$) with $\dim V \ge 2$, and $T \in \mathcal{L}(V)$.
Step 1: every monic polynomial of degree at least two has a monic divisor of degree exactly two.
Let $q \in \mathcal{P}(\mathbf{F})$ be monic with $\deg q = m \ge 2$.
If $\mathbf{F} = \mathbf{C}$, then q is nonconstant, so 4.13 provides $c, \lambda_1, \dots, \lambda_m \in \mathbf{C}$ with $q(z) = c(z - \lambda_1) \cdots (z - \lambda_m)$. Comparing coefficients of z^m on the two sides shows $c = 1$, because q is monic. Thus

$$q(z) = (z - \lambda_1) \cdots (z - \lambda_m).$$

Because $m \ge 2$, the monic polynomial $g(z) = (z - \lambda_1)(z - \lambda_2)$ has degree 2 and $q = gh$, where $h(z) = (z - \lambda_3) \cdots (z - \lambda_m)$ is monic of degree $m - 2$ (interpreting an empty product as the constant polynomial 1).
If $\mathbf{F} = \mathbf{R}$, then again q is nonconstant, so 4.16 provides a factorization

$$q(x) = (x - \lambda_1) \cdots (x - \lambda_j)(x^2 + b_1 x + c_1) \cdots (x^2 + b_k x + c_k),$$

where $\lambda_1, \dots, \lambda_j, b_1, c_1, \dots, b_k, c_k \in \mathbf{R}$ and $b_i^2 < 4c_i$ for each i; the leading constant supplied by 4.16 equals 1 because q is monic, and comparing degrees gives $j + 2k = m$. If $k \ge 1$, take $g(x) = x^2 + b_1 x + c_1$; if $k = 0$, then $j = m \ge 2$ and we take $g(x) = (x - \lambda_1)(x - \lambda_2)$. In either case g is monic of degree 2 and $q = gh$ for a monic h of degree $m - 2$ (the product of the remaining factors).
So in both cases there are monic $g, h \in \mathcal{P}(\mathbf{F})$ with $\deg g = 2$, $\deg h = m - 2$, and $q = gh$.
Step 2: reduction to the case where T is not a scalar multiple of the identity.
Suppose first that $T = \lambda I$ for some $\lambda \in \mathbf{F}$. Then every subspace of V is invariant under T. Because $\dim V \ge 2$, there exist linearly independent $u_1, u_2 \in V$, and $\operatorname{span}(u_1, u_2)$ is a two-dimensional invariant subspace. So the result holds in this case.
Now suppose T is not a scalar multiple of the identity. We claim there exists $v \in V$ such that the list v, Tv is linearly independent. To see this, suppose no such v exists. Then for each nonzero $v \in V$ the list v, Tv is linearly dependent, so there exist $a, b \in \mathbf{F}$, not both 0, with $av + bTv = 0$. If $b = 0$, then $a \ne 0$ and $av = 0$, forcing $v = 0$, a contradiction; hence $b \ne 0$ and $Tv = \lambda_v v$ with $\lambda_v = -a/b \in \mathbf{F}$.
We show that all these scalars are equal. Suppose $u, w \in V$ are nonzero. If u, w is linearly dependent, then (by the same two-line argument, applied to the dependent list u, w of nonzero vectors) $w = tu$ for some $t \in \mathbf{F} \setminus \{0\}$, and

$$\lambda_w w = Tw = tTu = t\lambda_u u = \lambda_u w,$$

so $\lambda_w = \lambda_u$ (because $w \ne 0$). If u, w is linearly independent, then $u + w \ne 0$ and

$$\lambda_{u+w} u + \lambda_{u+w} w = T(u + w) = Tu + Tw = \lambda_u u + \lambda_w w;$$

comparing coefficients with respect to the linearly independent list u, w gives $\lambda_u = \lambda_{u+w} = \lambda_w$. Hence there is a single $\lambda \in \mathbf{F}$ with $Tv = \lambda v$ for all nonzero $v \in V$, and thus $T = \lambda I$ (the equation $Tv = \lambda v$ also holds trivially for $v = 0$). This contradicts our assumption, proving the claim.
Step 3: the minimal polynomial of a vector.
Fix $v \in V$ such that v, Tv is linearly independent; in particular $v \ne 0$. Because V is finite-dimensional, T has a minimal polynomial p (5.22), and $p(T) = 0$, so $p(T)v = 0$. Thus the set

$$\{q \in \mathcal{P}(\mathbf{F}) : q \text{ is monic and } q(T)v = 0\}$$

is nonempty. Choose an element q of this set of smallest degree, and put $m = \deg q$.
Minimality has the following consequence, used twice below: if $r \in \mathcal{P}(\mathbf{F})$ satisfies $\deg r < m$ and $r(T)v = 0$, then $r = 0$. Indeed, if $r \neq 0$, then dividing r by its leading coefficient produces a monic polynomial of degree $\deg r < m$ that sends v to 0 under the operator substitution, contradicting the choice of q.
We have $m \geq 2$. Indeed, $m \neq 0$: the only monic polynomial of degree 0 is the constant polynomial 1, whose associated operator is I, and $Iv = v \neq 0$. And $m \neq 1$: a monic polynomial of degree 1 has the form $z - \lambda$, and $(T - \lambda I)v = 0$ would give $Tv = \lambda v$, contradicting the linear independence of v, Tv.
Step 4: construction of a two-dimensional invariant subspace.
By Step 1 there are monic $g, h \in \mathcal{P}(\mathbf{F})$ with $q = gh$, $\deg h = m - 2$, and

$$g(z) = z^2 + bz + c$$

for some $b, c \in \mathbf{F}$. Define

$$w = h(T)v.$$

Then $w \neq 0$. Indeed, h is monic and hence nonzero, and $\deg h = m - 2 < m$; if $w = h(T)v$ were 0, the minimality consequence in Step 3 would force $h = 0$, a contradiction.
By the multiplicative property 5.17,

$$g(T)w = g(T)\big(h(T)v\big) = (gh)(T)v = q(T)v = 0,$$

that is,

$$T^2 w + b\,Tw + c\,w = 0, \qquad \text{so} \qquad T^2 w = -b\,Tw - c\,w.$$

Let $W = \operatorname{span}(w, Tw)$. Then W is invariant under T: a typical element of W is $\alpha w + \beta\,Tw$ with $\alpha, \beta \in \mathbf{F}$, and

$$T(\alpha w + \beta\,Tw) = \alpha\,Tw + \beta\,T^2 w = \alpha\,Tw + \beta(-b\,Tw - c\,w) = (-\beta c)w + (\alpha - \beta b)Tw \in W.$$

Finally, $\dim W = 2$. Certainly $\dim W \leq 2$, because W is spanned by a list of length 2. If $\dim W$ were less than 2, then since $w \in W$ and $w \neq 0$ we would have $\dim W = 1$ and $W = \operatorname{span}(w)$; as $Tw \in W$, this gives $Tw = \lambda w$ for some $\lambda \in \mathbf{F}$. Then the polynomial r defined by $r(z) = (z - \lambda)h(z)$ is monic of degree $m - 1 < m$ and satisfies, again by 5.17,

$$r(T)v = (T - \lambda I)\big(h(T)v\big) = (T - \lambda I)w = Tw - \lambda w = 0,$$

contradicting the minimality of m established in Step 3.
Hence W is a two-dimensional subspace of V invariant under T, completing the proof.
Two remarks. First, when $m = 2$ the polynomial h is the constant 1, so $w = v$ and $W = \operatorname{span}(v, Tv)$; the general construction just first pushes v forward by $h(T)$ so that the remaining annihilator of the new vector has degree exactly 2. Second, the hypothesis $\dim V \geq 2$ is needed only to guarantee that V has a two-dimensional subspace at all; it is used in Step 2, while in Steps 3 and 4 the existence of v with v, Tv linearly independent already forces $\dim V \geq 2$.

5.3 Exercises 5C

Problem 5C.1. Prove or give a counterexample: If $T \in \mathcal{L}(V)$ and T^2 has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some basis of V.

Solution. The statement is false. Here is a counterexample.
Take $\mathbf{F} = \mathbf{R}$, take $V = \mathbf{R}^2$, and define $T \in \mathcal{L}(\mathbf{R}^2)$ by

$$T(x, y) = (-y, x).$$

(This is counterclockwise rotation by 90 degrees.)

First we compute T^2. For all $(x, y) \in \mathbf{R}^2$,

$$T^2(x, y) = T(-y, x) = (-x, -y),$$

so $T^2 = -I$. The matrix of $-I$ with respect to any basis of $\mathbf{R}^2$ is

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

which is upper triangular. Thus T^2 has an upper-triangular matrix with respect to some basis of V (indeed, with respect to every basis of V).
Now we show that T has no upper-triangular matrix with respect to any basis of $\mathbf{R}^2$. We first check that T has no eigenvalues. Suppose $\lambda \in \mathbf{R}$ and $(x, y) \in \mathbf{R}^2$ satisfy $T(x, y) = \lambda(x, y)$. Then

$$-y = \lambda x \quad \text{and} \quad x = \lambda y.$$

Substituting the first equation into the second gives $x = \lambda(-\lambda x) = -\lambda^2 x$, hence $(1+\lambda^2)x = 0$. Because $1 + \lambda^2 > 0$ for every real λ, we get $x = 0$; then $-y = \lambda x = 0$, so $y = 0$ as well. Hence the only vector satisfying $T(x, y) = \lambda(x, y)$ is $(0, 0)$, and therefore T has no eigenvalues.
Suppose now, seeking a contradiction, that T had an upper-triangular matrix with respect to some basis v_1, v_2 of $\mathbf{R}^2$, with diagonal entries λ_1, λ_2. By 5.41, the eigenvalues of T are exactly the entries on the diagonal of that matrix; in particular λ_1 would be an eigenvalue of T (this is immediate anyway, since upper-triangularity forces $Tv_1 = \lambda_1 v_1$ with $v_1 \neq 0$). This contradicts the previous paragraph.
Hence T^2 has an upper-triangular matrix with respect to some basis of V, but T does not.
Note that any counterexample must have $\mathbf{F} = \mathbf{R}$: if $\mathbf{F} = \mathbf{C}$ and V is finite-dimensional, then by 5.47 every operator on V has an upper-triangular matrix with respect to some basis, so the implication holds vacuously in the complex finite-dimensional case.

Problem 5C.2. Suppose A and B are upper-triangular matrices of the same size, with $\alpha_1, \dots, \alpha_n$ on the diagonal of A and $\beta_1, \dots, \beta_n$ on the diagonal of B.

- (a) Show that $A + B$ is an upper-triangular matrix with $\alpha_1 + \beta_1, \dots, \alpha_n + \beta_n$ on the diagonal.
- (b) Show that AB is an upper-triangular matrix with $\alpha_1\beta_1, \dots, \alpha_n\beta_n$ on the diagonal.

[The results in this exercise are used in the proof of 5.81.]

Solution. Throughout, A and B are n-by-n matrices, and $A_{j,k}$ denotes the entry of A in row j, column k. The hypothesis that A and B are upper triangular says precisely (by 5.38) that

$$A_{j,k} = 0 \quad \text{and} \quad B_{j,k} = 0 \qquad \text{whenever } j > k,$$

and the hypothesis about the diagonals says that $A_{k,k} = \alpha_k$ and $B_{k,k} = \beta_k$ for each $k \in \{1, \dots, n\}$.
(a) By the definition of matrix addition, $(A + B)_{j,k} = A_{j,k} + B_{j,k}$ for all j, k. If $j > k$, then $A_{j,k} = 0$ and $B_{j,k} = 0$, so $(A + B)_{j,k} = 0$. Thus every entry of $A + B$ below the diagonal equals 0, which means that $A + B$ is upper triangular. Furthermore, for each $k \in \{1, \dots, n\}$ the k-th diagonal entry of $A + B$ is

$$(A + B)_{k,k} = A_{k,k} + B_{k,k} = \alpha_k + \beta_k.$$

Hence $A + B$ is upper triangular with $\alpha_1 + \beta_1, \dots, \alpha_n + \beta_n$ on its diagonal.
(b) By the definition of matrix multiplication (3.41),

$$(AB)_{j,k} = \sum_{r=1}^{n} A_{j,r} B_{r,k}$$

for all $j, k \in \{1, \dots, n\}$.
Fix j and k, and consider a term $A_{j,r}B_{r,k}$ of the sum above that is nonzero. Then $A_{j,r} \neq 0$, which (because A is upper triangular) forces $j \leq r$; and $B_{r,k} \neq 0$, which (because B is upper triangular)

forces $r \le k$. Hence every nonzero term of the sum has

$$j \le r \le k.$$

First suppose $j > k$. Then no index r satisfies $j \le r \le k$, so every term of the sum is 0, giving $(AB)_{j,k} = 0$. Thus all entries of AB below the diagonal equal 0, which means that AB is upper triangular.
Now suppose $j = k$. Then the only index r satisfying $j \le r \le k$ is $r = k$, so all terms of the sum other than the one with $r = k$ equal 0. Hence

$$(AB)_{k,k} = A_{k,k}B_{k,k} = \alpha_k\beta_k.$$

Thus AB is upper triangular with $\alpha_1\beta_1, \dots, \alpha_n\beta_n$ on its diagonal.

Problem 5C.3. Suppose $T \in \mathcal{L}(V)$ is invertible and $v_1, \dots, v_n$ is a basis of V with respect to which the matrix of T is upper triangular, with $\lambda_1, \dots, \lambda_n$ on the diagonal. Show that the matrix of T^{-1} is also upper triangular with respect to the basis $v_1, \dots, v_n$, with

$$\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}$$

on the diagonal.

Solution. Because V has the basis $v_1, \dots, v_n$, the vector space V is finite-dimensional with $\dim V = n$. For each $k \in \{1, \dots, n\}$ set

$$U_k = \operatorname{span}(v_1, \dots, v_k).$$

Since $v_1, \dots, v_k$ is a sublist of a basis, it is linearly independent, and hence $\dim U_k = k$.
Step 1: each λ_k is nonzero. Because T is invertible, T is injective, so 0 is not an eigenvalue of T. By 5.41 the eigenvalues of T are exactly the diagonal entries $\lambda_1, \dots, \lambda_n$ of the given upper-triangular matrix. Hence $\lambda_k \neq 0$ for each k, and the numbers $1/\lambda_k$ make sense.
Step 2: the matrix of T^{-1} with respect to $v_1, \dots, v_n$ is upper triangular. Fix $k \in \{1, \dots, n\}$. Because the matrix of T with respect to $v_1, \dots, v_n$ is upper triangular, 5.39 tells us that U_k is invariant under T. Thus $T|_{U_k}$ is an operator on U_k. Since T is injective on V, its restriction $T|_{U_k}$ is injective; because U_k is finite-dimensional, injectivity of the operator $T|_{U_k}$ implies its surjectivity (by 3.65). Hence

$$T(U_k) = U_k.$$

In particular $v_k \in U_k = T(U_k)$, so there exists $u \in U_k$ with $Tu = v_k$, which gives $T^{-1}v_k = u \in U_k$. Thus

$$T^{-1}v_k \in \operatorname{span}(v_1, \dots, v_k) \qquad \text{for each } k \in \{1, \dots, n\}.$$

By the implication (c) $\Longrightarrow$ (a) of 5.39, the matrix of T^{-1} with respect to $v_1, \dots, v_n$ is upper triangular.
Step 3: identification of the diagonal entries. Let $\mu_1, \dots, \mu_n$ denote the diagonal entries of the matrix of T^{-1} with respect to $v_1, \dots, v_n$. By 3.43 (applied with all three bases equal to $v_1, \dots, v_n$),

$$\mathcal{M}(T)\,\mathcal{M}(T^{-1}) = \mathcal{M}(TT^{-1}) = \mathcal{M}(I),$$

and $\mathcal{M}(I)$ is the n-by-n identity matrix, whose diagonal entries all equal 1. On the other hand, $\mathcal{M}(T)$ and $\mathcal{M}(T^{-1})$ are upper-triangular matrices of the same size with diagonals $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$ respectively, so by part (b) of Exercise 2 the diagonal entries of $\mathcal{M}(T)\mathcal{M}(T^{-1})$ are $\lambda_1\mu_1, \dots, \lambda_n\mu_n$. Comparing the two descriptions of the diagonal gives

$$\lambda_k\mu_k = 1 \qquad \text{for each } k \in \{1, \dots, n\},$$

and therefore $\mu_k = 1/\lambda_k$.
Hence the matrix of T^{-1} with respect to $v_1, \dots, v_n$ is upper triangular with $1/\lambda_1, \dots, 1/\lambda_n$ on its diagonal, as desired.

Problem 5C.4. Give an example of an operator whose matrix with respect to some basis contains only 0's on the diagonal, but the operator is invertible.
[This exercise and the exercise below show that 5.41 fails without the hypothesis that an upper-triangular matrix is under consideration.]

Solution. Define $T \in \mathcal{L}(\mathbf{F}^2)$ by

$$T(x, y) = (y, x).$$

Let $e_1 = (1, 0)$, $e_2 = (0, 1)$ be the standard basis of $\mathbf{F}^2$. Then $Te_1 = e_2 = 0 \cdot e_1 + 1 \cdot e_2$ and $Te_2 = e_1 = 1 \cdot e_1 + 0 \cdot e_2$, so

$$\mathcal{M}(T) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Both diagonal entries of this matrix equal 0.
Nevertheless T is invertible: for all $(x, y) \in \mathbf{F}^2$ we have

$$T\big(T(x, y)\big) = T(y, x) = (x, y),$$

so $T^2 = I$. Thus T is invertible, with $T^{-1} = T$.
Note that this does not contradict 5.41, because the matrix above is not upper triangular (its entry in row 2, column 1 is $1 \neq 0$). Indeed, the eigenvalues of T are 1 and -1 (with eigenvectors $(1, 1)$ and $(1, -1)$), and neither of these appears on the diagonal of $\mathcal{M}(T)$.

Problem 5C.5. Give an example of an operator whose matrix with respect to some basis contains only nonzero numbers on the diagonal, but the operator is not invertible.

Solution. Define $T \in \mathcal{L}(\mathbf{F}^2)$ by

$$T(x, y) = (x + y,\ x + y).$$

Let $e_1 = (1, 0)$, $e_2 = (0, 1)$ be the standard basis of $\mathbf{F}^2$. Then $Te_1 = (1, 1) = e_1 + e_2$ and $Te_2 = (1, 1) = e_1 + e_2$, so

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Both diagonal entries of this matrix equal 1, which is nonzero.
However T is not invertible: the vector $(1, -1)$ is nonzero and

$$T(1, -1) = (1 + (-1),\ 1 + (-1)) = (0, 0),$$

so T is not injective, and hence not invertible.
Again this is consistent with 5.41, because the matrix above is not upper triangular. (Here 0 is an eigenvalue of T, yet 0 does not appear on the diagonal of $\mathcal{M}(T)$.)

Problem 5C.6. Suppose $\mathbf{F} = \mathbf{C}$, V is finite-dimensional, and $T \in \mathcal{L}(V)$. Prove that if $k \in \{1, \dots, \dim V\}$, then V has a k-dimensional subspace invariant under T.

Solution. Let $n = \dim V$ and suppose $k \in \{1, \dots, n\}$ (in particular $n \geq 1$, so $V \neq \{0\}$).
Because $\mathbf{F} = \mathbf{C}$ and V is a finite-dimensional complex vector space, 5.47 tells us that there is a basis $v_1, \dots, v_n$ of V with respect to which T has an upper-triangular matrix.
Let

$$U = \operatorname{span}(v_1, \dots, v_k).$$

By the implication (a) $\Longrightarrow$ (b) of 5.39, the subspace $\operatorname{span}(v_1, \dots, v_k)$ is invariant under T for each $k = 1, \dots, n$. In particular U is invariant under T.
Finally, $v_1, \dots, v_k$ is a sublist of the basis $v_1, \dots, v_n$ of V, hence is linearly independent, and by construction it spans U. Therefore $v_1, \dots, v_k$ is a basis of U and $\dim U = k$.

Thus U is a k-dimensional subspace of V invariant under T, as desired.

Problem 5C.7. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $v \in V$.
- (a) Prove that there exists a unique monic polynomial p_v of smallest degree such that $p_v(T)v = 0$.
- (b) Prove that the minimal polynomial of T is a polynomial multiple of p_v.

Solution. Recall (5.21) that a monic polynomial is a polynomial whose highest-degree coefficient equals 1; in particular the constant polynomial 1 is monic, of degree 0.
(a) Let

$$S = \{\, q \in \mathcal{P}(\mathbf{F}) : q \text{ is monic and } q(T)v = 0 \,\}.$$

First we show that $S \neq \varnothing$. Because V is finite-dimensional and $T \in \mathcal{L}(V)$, the operator T has a minimal polynomial p, which is monic and satisfies $p(T) = 0$ (by 5.22). Hence $p(T)v = 0$, so $p \in S$.
Thus $\{\deg q : q \in S\}$ is a nonempty set of nonnegative integers, and so it has a smallest element m. Choose $p_v \in S$ with $\deg p_v = m$. Then p_v is a monic polynomial of smallest degree among monic polynomials annihilating v, and $p_v(T)v = 0$. This proves existence.
For uniqueness, suppose $q \in S$ also satisfies $\deg q = m$. Consider the polynomial $p_v - q$. Both p_v and q are monic of degree m, so their degree-m coefficients are both 1 and cancel in the difference; hence either $p_v - q = 0$ or $\deg(p_v - q) < m$. Moreover, by the linearity of the map $r \mapsto r(T)$ (see 5.14),

$$(p_v - q)(T)v = p_v(T)v - q(T)v = 0 - 0 = 0.$$

Suppose $p_v - q \neq 0$, and let $c \neq 0$ be its highest-degree coefficient. Then $c^{-1}(p_v - q)$ is a monic polynomial of degree $\deg(p_v - q) < m$, and

$$\big(c^{-1}(p_v - q)\big)(T)v = c^{-1}\big((p_v - q)(T)v\big) = 0,$$

so $c^{-1}(p_v - q) \in S$ has degree less than m. This contradicts the minimality of m. Hence $p_v - q = 0$, that is, $q = p_v$. This proves uniqueness.
Notice also that $m \leq \deg p \leq \dim V$ by 5.22, so p_v has degree at most $\dim V$.
(b) Let p be the minimal polynomial of T. Because $p_v \neq 0$ (it is monic), the division algorithm for polynomials (4.9) gives polynomials $s, r \in \mathcal{P}(\mathbf{F})$ with

$$p = s\,p_v + r \quad \text{and} \quad \deg r < \deg p_v.$$

Applying this identity at T, using the additivity of $r \mapsto r(T)$ (immediate from 5.14) and the multiplicative property 5.17(a), we get

$$p(T) = s(T)\,p_v(T) + r(T).$$

Apply both sides to v. Since $p(T) = 0$ (by 5.22) and $p_v(T)v = 0$,

$$0 = p(T)v = s(T)\big(p_v(T)v\big) + r(T)v = s(T)0 + r(T)v = r(T)v.$$

Suppose $r \neq 0$, and let $c \neq 0$ be its highest-degree coefficient. Then $c^{-1}r$ is monic, has degree $\deg r < \deg p_v$, and satisfies $(c^{-1}r)(T)v = c^{-1}\big(r(T)v\big) = 0$. This contradicts the fact, established in part (a), that p_v has smallest degree among the monic polynomials q with $q(T)v = 0$.
Hence $r = 0$, and therefore

$$p = s\,p_v,$$

which shows that the minimal polynomial of T is a polynomial multiple of p_v.

Problem 5C.8. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and there exists a nonzero vector $v \in V$ such that $T^2v + 2Tv = -2v$.
(a) Prove that if $\mathbb{F} = \mathbb{R}$, then there does not exist a basis of V with respect to which T has an upper-triangular matrix.

(b) Prove that if $\mathbb{F} = \mathbb{C}$ and A is an upper-triangular matrix that equals the matrix of T with respect to some basis of V, then $-1 + i$ or $-1 - i$ appears on the diagonal of A.

Solution. The hypothesis says exactly that

$$(T^2 + 2T + 2I)v = 0, \qquad v \neq 0.$$

We begin with a construction that serves both parts.
Step 1: a nonzero invariant subspace on which T satisfies a quadratic equation.
Let

$$U = \operatorname{span}(v, Tv).$$

Then $U \neq \{0\}$, because $v \in U$ and $v \neq 0$. Furthermore U is invariant under T: indeed $Tv \in U$, and

$$T(Tv) = T^2v = -2Tv - 2v \in U,$$

so T maps each vector of the spanning list v, Tv into U, and hence maps all of U into U (because T is linear).
Let $S = T|_U \in \mathcal{L}(U)$. I claim that

$$S^2 + 2S + 2I_U = 0,$$

where I_U denotes the identity operator on U. To see this, note that $(T^2+2T+2I)v = 0$ by hypothesis, and applying T to that equation gives

$$(T^2 + 2T + 2I)(Tv) = T\big((T^2 + 2T + 2I)v\big) = T0 = 0,$$

where the first equality holds because T commutes with every polynomial in T. Thus the operator $T^2 + 2T + 2I$ vanishes on both vectors of the spanning list v, Tv of U, and hence vanishes on all of U. Since U is invariant under T, the restriction of $T^2 + 2T + 2I$ to U equals $S^2 + 2S + 2I_U$. This proves the claim.
Step 2: the possible eigenvalues of S.
Suppose $\mu \in \mathbb{F}$ is an eigenvalue of S, with eigenvector $u \in U$, $u \neq 0$. Then $S^2u = \mu^2u$, so

$$0 = (S^2 + 2S + 2I_U)u = (\mu^2 + 2\mu + 2)u.$$

Since $u \neq 0$, we get $\mu^2 + 2\mu + 2 = 0$, i.e. $(\mu + 1)^2 = -1$.
If $\mathbb{F} = \mathbb{R}$ this is impossible, since $(\mu+1)^2 \geq 0$ for every real μ. Hence S has no eigenvalue when $\mathbb{F} = \mathbb{R}$.
If $\mathbb{F} = \mathbb{C}$, then $(\mu + 1)^2 = -1$ forces $\mu + 1 = i$ or $\mu + 1 = -i$, so

$$\mu = -1 + i \quad \text{or} \quad \mu = -1 - i.$$

Step 3: the two parts.
(a) Suppose $\mathbb{F} = \mathbb{R}$ and, seeking a contradiction, suppose T has an upper-triangular matrix with respect to some basis of V, with diagonal entries $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. By 5.40,

$$(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0.$$

The subspace U is invariant under T, hence invariant under each $T - \lambda_k I$; therefore restricting the displayed equation to U gives

$$(S - \lambda_1 I_U) \cdots (S - \lambda_n I_U) = 0,$$

an equation between operators on U.
By Step 2 no real number is an eigenvalue of S. In particular, for each k the operator $S - \lambda_k I_U$ is injective (5.7). A composition of injective maps is injective, so $(S - \lambda_1 I_U) \cdots (S - \lambda_n I_U)$ is injective. But we have just shown that this operator is 0, and the zero operator on U is injective only if $U = \{0\}$. This contradicts $U \neq \{0\}$ from Step 1.
Hence no basis of V gives T an upper-triangular matrix when $\mathbb{F} = \mathbb{R}$.

(b) Suppose $\mathbb{F} = \mathbb{C}$ and A is an upper-triangular matrix equal to the matrix of T with respect to some basis of V; let $\lambda_1, \dots, \lambda_n$ be the diagonal entries of A. Exactly as in part (a), 5.40 gives $(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0$, and restricting to the invariant subspace U yields

$$(S - \lambda_1 I_U) \cdots (S - \lambda_n I_U) = 0.$$

Since $U \neq \{0\}$, this composition is not injective, so at least one factor $S - \lambda_k I_U$ fails to be injective (a composition of injective operators is injective). By 5.7 that λ_k is an eigenvalue of S, so by Step 2

$$\lambda_k = -1 + i \quad \text{or} \quad \lambda_k = -1 - i.$$

Thus $-1 + i$ or $-1 - i$ appears on the diagonal of A, as desired.
Two remarks. First, the argument in (b) never needed 5.19; but one can also phrase it that way: U is a nonzero finite-dimensional complex vector space, so S has an eigenvalue by 5.19, which by Step 2 is $-1 + i$ or $-1 - i$; an eigenvector of S in U is an eigenvector of T in V, so this number is an eigenvalue of T, and by 5.41 the eigenvalues of T are precisely the diagonal entries of A.
Second, part (a) shows in particular that the hypothesis of part (b) cannot be met over $\mathbb{R}$: over $\mathbb{R}$ the operator T has no upper-triangular matrix at all. Concretely, $T \in \mathcal{L}(\mathbb{R}^2)$ defined by $T(x, y) = (-x - y, x - y)$ satisfies $T^2 + 2T + 2I = 0$ and has no eigenvalue.

Problem 5C.9. Suppose B is a square matrix with complex entries. Prove that there exists an invertible square matrix A with complex entries such that $A^{-1}BA$ is an upper-triangular matrix.

Solution. Say B is an n-by-n matrix with complex entries. Let $e_1, \dots, e_n$ denote the standard basis of $\mathbb{C}^n$ and define $T \in \mathcal{L}(\mathbb{C}^n)$ to be the unique operator whose matrix with respect to $e_1, \dots, e_n$ is B; explicitly, if $B = (B_{j,k})$, then

$$Te_k = \sum_{j=1}^{n} B_{j,k}\, e_j \qquad \text{for each } k = 1, \dots, n,$$

and T is extended linearly. Thus

$$\mathcal{M}\big(T, (e_1, \dots, e_n)\big) = B.$$

Because $\mathbb{C}^n$ is a finite-dimensional complex vector space, 5.47 tells us that there exists a basis $u_1, \dots, u_n$ of $\mathbb{C}^n$ with respect to which T has an upper-triangular matrix. Put

$$C = \mathcal{M}\big(T, (u_1, \dots, u_n)\big),$$

so C is upper triangular.
Now let

$$A = \mathcal{M}\big(I, (u_1, \dots, u_n), (e_1, \dots, e_n)\big),$$

the matrix of the identity operator on $\mathbb{C}^n$ with respect to the basis $u_1, \dots, u_n$ in the domain and the basis $e_1, \dots, e_n$ in the target. Concretely, the k^{th} column of A is the list of coordinates of u_k with respect to the standard basis, which is just u_k itself written as a column vector.
The matrix A is invertible. Indeed, if

$$A' = \mathcal{M}\big(I, (e_1, \dots, e_n), (u_1, \dots, u_n)\big),$$

then by the formula for the matrix of a product of linear maps (3.81), $AA' = \mathcal{M}(I, (e), (e))$ is the identity matrix and $A'A = \mathcal{M}(I, (u), (u))$ is the identity matrix. Hence A is invertible with $A^{-1} = A'$.
The change-of-basis formula 3.84, applied with the bases $u_1, \dots, u_n$ and $e_1, \dots, e_n$, gives

$$C = A^{-1}BA.$$

Since C is upper triangular, this is exactly the desired conclusion: A is an invertible square matrix with complex entries such that $A^{-1}BA$ is upper triangular.

Note that this argument genuinely uses $\mathbb{F} = \mathbb{C}$. The analogous statement over $\mathbb{R}$ is false: taking

$$B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and any invertible real A, the matrix $A^{-1}BA$ is the matrix of the corresponding operator on $\mathbb{R}^2$ with respect to some basis, and if it were upper triangular then by 5.41 its diagonal entries would be eigenvalues of that operator, which is rotation by 90 degrees and has no real eigenvalue.

Problem 5C.10. Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of V. Show that the following are equivalent.
(a) The matrix of T with respect to $v_1, \dots, v_n$ is lower triangular.
(b) $\operatorname{span}(v_k, \dots, v_n)$ is invariant under T for each $k = 1, \dots, n$.
(c) $Tv_k \in \operatorname{span}(v_k, \dots, v_n)$ for each $k = 1, \dots, n$.
[A square matrix is called lower triangular if all entries above the diagonal are 0.]

Solution. The cleanest route is to reverse the order of the basis and quote 5.39, which is the corresponding statement for upper-triangular matrices. So define

$$w_j = v_{n+1-j} \qquad \text{for } j = 1, \dots, n,$$

so that $w_1, \dots, w_n$ is the list $v_n, v_{n-1}, \dots, v_1$. This is again a basis of V, since it is the same finite set of vectors listed in a different order.
Reversal turns lower triangular into upper triangular. Let

$$A = \mathcal{M}\big(T, (v_1, \dots, v_n)\big), \qquad B = \mathcal{M}\big(T, (w_1, \dots, w_n)\big).$$

Fix $k \in \{1, \dots, n\}$ and write $Tv_{n+1-k} = \sum_{i=1}^{n} A_{i,n+1-k} v_i$, which is the definition of the column $n+1-k$ of A. Re-indexing the sum by $i = n + 1 - j$ and using $v_{n+1-j} = w_j$ gives

$$Tw_k = Tv_{n+1-k} = \sum_{j=1}^{n} A_{n+1-j,\, n+1-k}\, w_j.$$

Hence

$$B_{j,k} = A_{n+1-j,\, n+1-k} \qquad \text{for all } j, k,$$

so B is obtained from A by rotating it 180 degrees. Now

$$j > k \iff n + 1 - j < n + 1 - k,$$

so the entries of B strictly below its diagonal are exactly the entries of A strictly above its diagonal. Therefore

$$B \text{ is upper triangular} \iff A \text{ is lower triangular.}$$

Reversal matches up the spans. For each $m \in \{1, \dots, n\}$,

$$\operatorname{span}(w_1, \dots, w_m) = \operatorname{span}(v_n, v_{n-1}, \dots, v_{n+1-m}) = \operatorname{span}(v_{n+1-m}, \dots, v_n),$$

because span does not depend on the order in which the vectors are listed. Writing $k = n + 1 - m$, the correspondence $m \leftrightarrow k$ is a bijection of $\{1, \dots, n\}$ with itself, and under it

$$\operatorname{span}(w_1, \dots, w_m) = \operatorname{span}(v_k, \dots, v_n), \qquad w_m = v_k.$$

Now compare the three conditions of this exercise, stated for T and $v_1, \dots, v_n$, with the three conditions of 5.39, stated for T and $w_1, \dots, w_n$.
Condition (a) here says A is lower triangular, which by the first paragraph above holds if and only if B is upper triangular, which is condition (a) of 5.39 for the basis $w_1, \dots, w_n$.

Condition (b) here says that $\operatorname{span}(v_k, \dots, v_n)$ is invariant under T for each $k = 1, \dots, n$. By the span identity above, this says exactly that $\operatorname{span}(w_1, \dots, w_m)$ is invariant under T for each $m = 1, \dots, n$, which is condition (b) of 5.39 for the basis $w_1, \dots, w_n$.
Condition (c) here says that $Tv_k \in \operatorname{span}(v_k, \dots, v_n)$ for each k. By the span identity and $w_m = v_k$, this says exactly that $Tw_m \in \operatorname{span}(w_1, \dots, w_m)$ for each $m = 1, \dots, n$, which is condition (c) of 5.39 for the basis $w_1, \dots, w_n$.
Thus (a), (b), (c) of this exercise are respectively equivalent to (a), (b), (c) of 5.39 applied to the basis $w_1, \dots, w_n$. Since those three are equivalent to each other by 5.39, so are these three. This completes the proof.
For the reader who prefers a direct argument, here it is in outline, mirroring the proof of 5.39.
Assume (a). Then $A_{i,j} = 0$ whenever $i < j$, so $Tv_j = \sum_{i \geq j} A_{i,j} v_i \in \operatorname{span}(v_j, \dots, v_n)$ for every j. If $j \geq k$ then $\operatorname{span}(v_j, \dots, v_n) \subseteq \operatorname{span}(v_k, \dots, v_n)$, so $Tv_j \in \operatorname{span}(v_k, \dots, v_n)$ for every $j \geq k$; hence $\operatorname{span}(v_k, \dots, v_n)$ is invariant under T, giving (b).
Assume (b). Fix k. Then $v_k \in \operatorname{span}(v_k, \dots, v_n)$, and that subspace is invariant under T, so $Tv_k \in \operatorname{span}(v_k, \dots, v_n)$, giving (c).
Assume (c). Then for each k we may write Tv_k as a linear combination of $v_k, \dots, v_n$ only, which means $A_{i,k} = 0$ for all $i < k$. Thus every entry of A above the diagonal is 0, giving (a).

Problem 5C.11. Suppose $\mathbb{F} = \mathbb{C}$ and V is finite-dimensional. Prove that if $T \in \mathcal{L}(V)$, then there exists a basis of V with respect to which T has a lower-triangular matrix.

Solution. If $V = \{0\}$ the statement is vacuous (the empty basis gives the empty matrix), so assume $n = \dim V \geq 1$.
Because V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$, 5.47 provides a basis $v_1, \dots, v_n$ of V with respect to which T has an upper-triangular matrix. By 5.39 this means

$$Tv_k \in \operatorname{span}(v_1, \dots, v_k) \qquad \text{for each } k = 1, \dots, n.$$

Now reverse the list. Define

$$w_j = v_{n+1-j} \qquad \text{for } j = 1, \dots, n.$$

Then $w_1, \dots, w_n$ is the list $v_n, \dots, v_1$, which is again a basis of V (the same vectors, listed in a different order).
I claim that T has a lower-triangular matrix with respect to $w_1, \dots, w_n$. By the equivalence proved in Exercise 10 of this section, it suffices to verify condition (c) there, namely

$$Tw_k \in \operatorname{span}(w_k, \dots, w_n) \qquad \text{for each } k = 1, \dots, n.$$

Fix k and set $m = n + 1 - k$, so $w_k = v_m$. Since span does not depend on the order in which vectors are listed,

$$\operatorname{span}(w_k, w_{k+1}, \dots, w_n) = \operatorname{span}(v_{n+1-k}, v_{n-k}, \dots, v_1) = \operatorname{span}(v_1, \dots, v_m).$$

By the displayed consequence of 5.39 above, $Tv_m \in \operatorname{span}(v_1, \dots, v_m)$. Therefore

$$Tw_k = Tv_m \in \operatorname{span}(v_1, \dots, v_m) = \operatorname{span}(w_k, \dots, w_n),$$

which is exactly what was needed.
Hence, by Exercise 10, the matrix of T with respect to the basis $w_1, \dots, w_n$ is lower triangular.
Alternatively, one can argue purely matricially: if A is the upper-triangular matrix of T with respect to $v_1, \dots, v_n$, then, as computed in the solution to Exercise 10, the matrix B of T with respect to the reversed basis satisfies $B_{j,k} = A_{n+1-j,\, n+1-k}$, so B is the 180-degree rotation of A and is therefore lower triangular.

Problem 5C.12. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V, and U is a subspace of V that is invariant under T.
(a) Prove that $T|_U$ has an upper-triangular matrix with respect to some basis of U.
(b) Prove that the quotient operator T/U has an upper-triangular matrix with respect to some basis of V/U.
[The quotient operator T/U was defined in Exercise 38 in Section 5A.]

Solution. Let $n = \dim V$ and let $v_1, \dots, v_n$ be a basis of V with respect to which T has an upper-triangular matrix. Put

$$V_0 = \{0\}, \qquad V_k = \operatorname{span}(v_1, \dots, v_k) \quad (k = 1, \dots, n).$$

By 5.39 each V_k is invariant under T, and $\dim V_k = k$ because $v_1, \dots, v_k$ is linearly independent. Thus we have a chain

$$\{0\} = V_0 \subseteq V_1 \subseteq \cdots \subseteq V_n = V$$

of T-invariant subspaces with $\dim V_k = k$. Both parts are proved by intersecting this chain with U, respectively pushing it forward to V/U.
Part (a).
For $k = 0, 1, \dots, n$ set

$$U_k = U \cap V_k.$$

Each U_k is invariant under T: if $u \in U_k$, then $Tu \in U$ because U is invariant under T, and $Tu \in V_k$ because V_k is invariant under T, so $Tu \in U_k$. Also $U_0 = \{0\}$ and $U_n = U$, and $U_{k-1} \subseteq U_k$.
The dimensions increase by at most 1 at each step. Indeed, since $V_{k-1} \subseteq V_k$ we have $U_k \cap V_{k-1} = U \cap V_{k-1} = U_{k-1}$, so by 2.43,

$$\dim(U_k + V_{k-1}) = \dim U_k + \dim V_{k-1} - \dim U_{k-1} = \dim U_k - \dim U_{k-1} + (k-1).$$

On the other hand $U_k + V_{k-1} \subseteq V_k$, so $\dim(U_k + V_{k-1}) \le k$. Combining,

$$0 \le \dim U_k - \dim U_{k-1} \le 1 \qquad \text{for each } k = 1, \dots, n.$$

Let $m = \dim U$. If $m = 0$ then $U = \{0\}$ and the empty list is a basis of U, with respect to which the (empty) matrix of $T|_U$ is upper triangular; so assume $m \ge 1$. Since $\dim U_0 = 0$, $\dim U_n = m$, and each step increases the dimension by 0 or 1, there are exactly m indices k at which the dimension increases; call them

$$k_1 < k_2 < \cdots < k_m.$$

Moreover, for each j the number of jumps among $1, \dots, k_j$ is j, so

$$\dim U_{k_j} = j.$$

For each j choose $u_j \in U_{k_j}$ with $u_j \notin U_{k_j - 1}$; such a vector exists because $\dim U_{k_j} > \dim U_{k_j - 1}$, so the inclusion $U_{k_j - 1} \subseteq U_{k_j}$ is strict.
I claim that

$$\operatorname{span}(u_1, \dots, u_j) = U_{k_j} \qquad \text{for each } j = 1, \dots, m.$$

We induct on j. For $j = 1$: $u_1 \notin U_{k_1 - 1} \supseteq \{0\}$, so $u_1 \neq 0$, and $\dim U_{k_1} = 1$, so $\operatorname{span}(u_1) = U_{k_1}$. Now let $j \ge 2$ and assume $\operatorname{span}(u_1, \dots, u_{j-1}) = U_{k_{j-1}}$. Since $k_{j-1} \le k_j - 1$ we have $U_{k_{j-1}} \subseteq U_{k_j - 1}$, and $u_j \notin U_{k_j - 1}$, so

$$u_j \notin \operatorname{span}(u_1, \dots, u_{j-1}).$$

Hence $u_1, \dots, u_j$ is linearly independent. All of these vectors lie in U_{k_j}, which has dimension j; a linearly independent list of length j in a j-dimensional space is a basis of it (2.38), so $\operatorname{span}(u_1, \dots, u_j) = U_{k_j}$. This proves the claim.
Taking $j = m$ gives $\operatorname{span}(u_1, \dots, u_m) = U_{k_m}$, a subspace of U of dimension $m = \dim U$, hence equal to U. Thus $u_1, \dots, u_m$ is a basis of U.

Finally, for each j the subspace $\operatorname{span}(u_1,\dots,u_j) = U_{k_j}$ is invariant under T and is contained in U, hence is invariant under $T|_U$. By 5.39, applied to the operator $T|_U \in \mathcal{L}(U)$ and the basis $u_1,\dots,u_m$ of U (implication (b) implies (a) there), the matrix of $T|_U$ with respect to $u_1,\dots,u_m$ is upper triangular.
Part (b).
Let $\pi\colon V \to V/U$ be the quotient map, $\pi v = v + U$. Recall (Exercise 38 in Section 5A) that because U is invariant under T, the quotient operator $T/U \in \mathcal{L}(V/U)$ is well defined by

$$(T/U)(v+U) = Tv + U, \qquad \text{i.e.} \qquad (T/U)\circ\pi = \pi\circ T.$$

For $k = 0, 1, \dots, n$ set

$$W_k = \pi(V_k) = \operatorname{span}\big(\pi(v_1),\dots,\pi(v_k)\big),$$

the second equality holding because π is linear and $v_1,\dots,v_k$ spans V_k. Each W_k is invariant under T/U: for $i \le k$ we have $Tv_i \in V_k$ (as V_k is invariant under T), hence

$$(T/U)\big(\pi(v_i)\big) = \pi(Tv_i) \in \pi(V_k) = W_k,$$

and since T/U is linear and $\pi(v_1),\dots,\pi(v_k)$ spans W_k, all of W_k is mapped into W_k.
Because π is surjective and $v_1,\dots,v_n$ spans V, the list $\pi(v_1),\dots,\pi(v_n)$ spans V/U. Reduce it to a basis by the procedure in the proof of 2.30: run through the list from left to right and delete $\pi(v_k)$ whenever $\pi(v_k) \in \operatorname{span}(\pi(v_1),\dots,\pi(v_{k-1}))$. Let

$$k_1 < k_2 < \cdots < k_p$$

be the indices that survive, and set $w_j = \pi(v_{k_j})$; then $w_1,\dots,w_p$ is a basis of V/U.
The key feature of this reduction is that it never changes the span of an initial segment:

$$\operatorname{span}\big(w_j : k_j \le k\big) = \operatorname{span}\big(\pi(v_1),\dots,\pi(v_k)\big) = W_k \qquad \text{for each } k = 0,1,\dots,n.$$

This is proved by induction on k. For $k = 0$ both sides are $\{0\}$. For the step from $k-1$ to k: if k is one of the retained indices, both sides gain the single vector $\pi(v_k)$, so the equality persists; if k is deleted, then $\pi(v_k) \in \operatorname{span}(\pi(v_1),\dots,\pi(v_{k-1}))$, so the right side is unchanged, and the left side is unchanged as well.
Taking $k = k_j$ in the display above gives

$$\operatorname{span}(w_1,\dots,w_j) = W_{k_j} \qquad \text{for each } j = 1,\dots,p,$$

and we showed above that each W_{k_j} is invariant under T/U. By 5.39, applied to the operator $T/U \in \mathcal{L}(V/U)$ and the basis $w_1,\dots,w_p$ of V/U, the matrix of T/U with respect to $w_1,\dots,w_p$ is upper triangular.
(Incidentally $p = \dim V/U = n - \dim U$ by 3.105, so the bases produced in parts (a) and (b) have lengths adding up to n.)

Problem 5C.13. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Suppose there exists a subspace U of V that is invariant under T such that $T|_U$ has an upper-triangular matrix with respect to some basis of U and also T/U has an upper-triangular matrix with respect to some basis of V/U. Prove that T has an upper-triangular matrix with respect to some basis of V.

Solution. This is the converse of Exercise 12: it says that upper-triangularizability can be assembled from a piece and its quotient.
Let $u_1,\dots,u_m$ be a basis of U with respect to which $T|_U$ has an upper-triangular matrix, and let $\widetilde{w}_1,\dots,\widetilde{w}_p$ be a basis of V/U with respect to which T/U has an upper-triangular matrix. Let $\pi\colon V \to V/U$ be the quotient map, $\pi v = v + U$; recall that π is linear and surjective with $\operatorname{null}\pi = U$, and that the quotient operator satisfies

$$(T/U)\circ\pi = \pi\circ T.$$

Because π is surjective, for each j we may choose $w_j \in V$ with $\pi(w_j) = \widetilde{w}_j$.
Step 1: the combined list is a basis of V.
I claim that

$$u_1, \dots, u_m, w_1, \dots, w_p$$

is a basis of V.
It spans V: given $v \in V$, write $\pi(v) = c_1\widetilde{w}_1 + \cdots + c_p\widetilde{w}_p$ with $c_1, \dots, c_p \in \mathbb{F}$, which is possible because $\widetilde{w}_1, \dots, \widetilde{w}_p$ is a basis of V/U. Then

$$\pi\big(v - c_1w_1 - \cdots - c_pw_p\big) = \pi(v) - c_1\widetilde{w}_1 - \cdots - c_p\widetilde{w}_p = 0,$$

so $v - c_1w_1 - \cdots - c_pw_p \in \operatorname{null}\pi = U = \operatorname{span}(u_1, \dots, u_m)$. Hence $v \in \operatorname{span}(u_1, \dots, u_m, w_1, \dots, w_p)$.
It is linearly independent: suppose $a_1, \dots, a_m, b_1, \dots, b_p \in \mathbb{F}$ satisfy

$$a_1u_1 + \cdots + a_mu_m + b_1w_1 + \cdots + b_pw_p = 0.$$

Applying π and using $\pi(u_i) = 0$ (because $u_i \in U = \operatorname{null}\pi$) gives

$$b_1\widetilde{w}_1 + \cdots + b_p\widetilde{w}_p = 0,$$

so $b_1 = \cdots = b_p = 0$ by the linear independence of $\widetilde{w}_1, \dots, \widetilde{w}_p$. The displayed equation then reduces to $a_1u_1 + \cdots + a_mu_m = 0$, whence $a_1 = \cdots = a_m = 0$ by the linear independence of $u_1, \dots, u_m$.
So the combined list is a basis of V; call it $x_1, \dots, x_{m+p}$, where $x_i = u_i$ for $1 \le i \le m$ and $x_{m+j} = w_j$ for $1 \le j \le p$.
Step 2: verifying condition (c) of 5.39.
By 5.39 it suffices to show

$$Tx_k \in \operatorname{span}(x_1, \dots, x_k) \qquad \text{for each } k = 1, \dots, m+p.$$

First let $k = i \le m$. Because the matrix of $T|_U$ with respect to $u_1, \dots, u_m$ is upper triangular, 5.39 (applied to $T|_U$) gives

$$Tu_i = (T|_U)(u_i) \in \operatorname{span}(u_1, \dots, u_i) = \operatorname{span}(x_1, \dots, x_i),$$

as required.
Now let $k = m + j$ with $1 \le j \le p$. Because the matrix of T/U with respect to $\widetilde{w}_1, \dots, \widetilde{w}_p$ is upper triangular, 5.39 (applied to T/U) gives

$$(T/U)(\widetilde{w}_j) \in \operatorname{span}(\widetilde{w}_1, \dots, \widetilde{w}_j),$$

so there exist $c_1, \dots, c_j \in \mathbb{F}$ with

$$\pi(Tw_j) = (T/U)\big(\pi(w_j)\big) = (T/U)(\widetilde{w}_j) = c_1\widetilde{w}_1 + \cdots + c_j\widetilde{w}_j = \pi(c_1w_1 + \cdots + c_jw_j).$$

Therefore

$$Tw_j - (c_1w_1 + \cdots + c_jw_j) \in \operatorname{null}\pi = U = \operatorname{span}(u_1, \dots, u_m),$$

and hence

$$Tx_{m+j} = Tw_j \in \operatorname{span}(u_1, \dots, u_m, w_1, \dots, w_j) = \operatorname{span}(x_1, \dots, x_{m+j}),$$

as required.
Thus condition (c) of 5.39 holds for the basis $x_1, \dots, x_{m+p}$ of V, so by 5.39 the matrix of T with respect to $x_1, \dots, x_{m+p}$ is upper triangular. In block form, that matrix is

$$\begin{pmatrix} \mathcal{M}(T|_U) & * \\ 0 & \mathcal{M}(T/U) \end{pmatrix},$$

with the two upper-triangular blocks coming from the hypotheses.

Problem 5C.14. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T has an upper-triangular matrix with respect to some basis of V if and only if the dual operator T' has an upper-triangular matrix with respect to some basis of the dual space V'.

Solution. Recall (3.118) that the dual operator $T' \in \mathcal{L}(V')$ is defined by

$$T'(\varphi) = \varphi \circ T \qquad \text{for } \varphi \in V'.$$

Also V' is finite-dimensional with $\dim V' = \dim V$ (3.111), so both T and T' have minimal polynomials (5.24).
If $V = \{0\}$ then $V' = \{0\}$ and both statements hold vacuously, so assume $V \neq \{0\}$.
The proof runs through the minimal polynomial. We show that T and T' have the **same** minimal polynomial, after which 5.44 gives both implications at once.
Step 1: if $S \in \mathcal{L}(V)$ and $S' = 0$, then $S = 0$.
Suppose $S \neq 0$. Choose $v \in V$ with $Sv \neq 0$. Since $Sv \neq 0$, the length-one list Sv is linearly independent, so it extends to a basis $Sv, y_2, \ldots, y_n$ of V (2.32). Let $\varphi \in V'$ be the first element of the corresponding dual basis (3.112), so that $\varphi(Sv) = 1$. Then

$$\big(S'(\varphi)\big)(v) = (\varphi \circ S)(v) = \varphi(Sv) = 1 \neq 0,$$

so $S'(\varphi) \neq 0$ and hence $S' \neq 0$. This proves the contrapositive.
Step 2: $\big(p(T)\big)' = p(T')$ for every $p \in \mathcal{P}(\mathbb{F})$.
First, $I' = I$, where on the left I is the identity on V and on the right the identity on V': indeed $I'(\varphi) = \varphi \circ I = \varphi$ for every $\varphi \in V'$.
Next, $(T^k)' = (T')^k$ for every nonnegative integer k. This holds for $k = 0$ by the previous paragraph, and if it holds for k then by 3.120(c), applied with the two maps T^k and T,

$$(T^{k+1})' = (T^k\, T)' = T'\,(T^k)' = T'(T')^k = (T')^{k+1},$$

completing the induction.
Finally, if $p(z) = a_0 + a_1 z + \cdots + a_d z^d$ with $a_0, \ldots, a_d \in \mathbb{F}$, then by 3.120(a) and 3.120(b) the map $S \mapsto S'$ is linear, so

$$\big(p(T)\big)' = \Big(\sum_{k=0}^{d} a_k T^k\Big)' = \sum_{k=0}^{d} a_k (T^k)' = \sum_{k=0}^{d} a_k (T')^k = p(T').$$

Step 3: T and T' have the same minimal polynomial.
Let p be the minimal polynomial of T and q the minimal polynomial of T'. By Step 2,

$$p(T') = \big(p(T)\big)' = 0' = 0,$$

so $\deg q \leq \deg p$, since q is a monic polynomial of smallest degree annihilating T'. In the other direction, again by Step 2,

$$\big(q(T)\big)' = q(T') = 0,$$

so $q(T) = 0$ by Step 1, whence $\deg p \leq \deg q$. Thus $\deg p = \deg q$, and q is a monic polynomial of degree $\deg p$ with $q(T) = 0$. Because the minimal polynomial of T is the unique monic polynomial of smallest degree annihilating T (5.24), we conclude $q = p$.
Step 4: conclusion.
By 5.44 applied to T on V, the operator T has an upper-triangular matrix with respect to some basis of V if and only if

$$p = (z - \lambda_1) \cdots (z - \lambda_m) \quad \text{for some } \lambda_1, \ldots, \lambda_m \in \mathbb{F}.$$

By 5.44 applied to T' on the finite-dimensional space V', the operator T' has an upper-triangular matrix with respect to some basis of V' if and only if its minimal polynomial q has that same form. Since $p = q$ by Step 3, the two conditions are equivalent, which is exactly the assertion to be proved.

A concrete construction for one direction. It is worth seeing how a triangularizing basis of V' is built explicitly from one of V. Suppose $v_1, \dots, v_n$ is a basis of V with respect to which $A = \mathcal{M}(T)$ is upper triangular, and let $\varphi_1, \dots, \varphi_n$ be the dual basis, which is a basis of V' by 3.116. By 3.132,

$$\mathcal{M}\big(T', (\varphi_1, \dots, \varphi_n)\big) = A^{\mathrm{t}},$$

which is lower triangular because A is upper triangular. Now reverse the order of that basis: by the computation in the solution to Exercise 10, the matrix of T' with respect to $\varphi_n, \varphi_{n-1}, \dots, \varphi_1$ is the 180-degree rotation of A^{t}, hence upper triangular. The same construction applied to T' produces a triangularizing basis of V'' for T'', and one can then transport it back to V through the canonical isomorphism $V \to V''$; the minimal-polynomial argument above avoids that last step.

5.4 Exercises 5D

Problem 5D.1. Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$.
(a) Prove that if $T^4 = I$, then T is diagonalizable.
(b) Prove that if $T^4 = T$, then T is diagonalizable.
(c) Give an example of an operator $T \in \mathcal{L}(\mathbf{C}^2)$ such that $T^4 = T^2$ and T is not diagonalizable.

Solution. Throughout we may assume $V \neq \{0\}$, since the operator on the zero vector space is diagonalizable (its matrix with respect to the empty basis is the empty diagonal matrix) and all three hypotheses hold vacuously there.
We begin with a lemma that will settle both (a) and (b).
Lemma. Suppose V is a finite-dimensional complex vector space, $T \in \mathcal{L}(V)$, and q is a nonconstant polynomial in $\mathcal{P}(\mathbf{C})$ whose zeros are all simple (that is, $q(z) = c(z - \beta_1) \cdots (z - \beta_n)$ with $c \neq 0$ and $\beta_1, \dots, \beta_n$ distinct) such that $q(T) = 0$. Then T is diagonalizable.
To prove the lemma, let p denote the minimal polynomial of T. Because $q(T) = 0$, the result 5.29 tells us that q is a polynomial multiple of p; write $q = ps$ with $s \in \mathcal{P}(\mathbf{C})$. Because $\mathbf{C}$ is algebraically closed, the factorization 4.13 lets us write

$$p(z) = (z - \alpha_1)^{m_1} \cdots (z - \alpha_k)^{m_k},$$

where $\alpha_1, \dots, \alpha_k \in \mathbf{C}$ are distinct and each m_j is a positive integer (here p is monic, and $k \geq 1$ because $\deg p \geq 1$ as $V \neq \{0\}$). Suppose some $m_j \geq 2$. Then $(z - \alpha_j)^2$ divides p, and hence $(z - \alpha_j)^2$ divides $q = ps$; writing $q = (z - \alpha_j)^2 h$ and factoring h into linear factors produces a factorization of q into linear factors in which $z - \alpha_j$ occurs at least twice. By the uniqueness assertion in 4.13 this contradicts the assumption that the zeros $\beta_1, \dots, \beta_n$ of q are distinct. Hence $m_1 = \cdots = m_k = 1$, so

$$p(z) = (z - \alpha_1) \cdots (z - \alpha_k)$$

for the distinct numbers $\alpha_1, \dots, \alpha_k$. Now 5.62 shows that T is diagonalizable, completing the proof of the lemma.
(a) Suppose $T^4 = I$. Let $q(z) = z^4 - 1$. Then $q(T) = T^4 - I = 0$. Over $\mathbf{C}$ we have

$$q(z) = (z - 1)(z + 1)(z - i)(z + i),$$

so the zeros of q are $1, -1, i, -i$, which are four distinct numbers; thus q has no repeated zeros. The lemma now shows that T is diagonalizable.
(b) Suppose $T^4 = T$. Let $q(z) = z^4 - z$. Then $q(T) = T^4 - T = 0$. Factoring,

$$q(z) = z(z^3 - 1) = z(z - 1)(z - \omega)(z - \omega^2), \qquad \text{where } \omega = \tfrac{-1 + i\sqrt{3}}{2} = e^{2\pi i/3}.$$

The four zeros $0, 1, \omega, \omega^2$ of q are distinct: the three cube roots of 1 are distinct because ω is not real and $\omega^2 = \overline{\omega} \neq \omega$, and none of them equals 0. Hence q has no repeated zeros, and the lemma shows that T is diagonalizable.

(c) Define $T \in \mathcal{L}(\mathbf{C}^2)$ by

$$T(w, z) = (z, 0).$$

Then $T^2(w, z) = T(z, 0) = (0, 0)$, so $T^2 = 0$ and therefore $T^4 = (T^2)^2 = 0 = T^2$. Thus $T^4 = T^2$.
However T is not diagonalizable. Indeed, suppose T had a diagonal matrix with respect to some basis v_1, v_2 of $\mathbf{C}^2$, say $Tv_k = \lambda_k v_k$. Then $0 = T^2 v_k = \lambda_k^2 v_k$ and $v_k \neq 0$, so $\lambda_k = 0$ for each k. Hence $Tv_k = 0$ for each k, which forces $T = 0$. But $T(0, 1) = (1, 0) \neq (0, 0)$, so $T \neq 0$, a contradiction. Thus T is not diagonalizable.
(Equivalently: $T^2 = 0$ and $T \neq 0$ show that the minimal polynomial of T is z^2, which does not have the form required by 5.62.)
This also shows why the hypothesis in the lemma cannot be relaxed: $z^4 - z^2 = z^2(z - 1)(z + 1)$ has 0 as a repeated zero.

Problem 5D.2. Suppose $T \in \mathcal{L}(V)$ has a diagonal matrix A with respect to some basis of V. Prove that if $\lambda \in \mathbf{F}$, then λ appears on the diagonal of A precisely $\dim E(\lambda, T)$ times.

Solution. Let $v_1, \ldots, v_n$ be a basis of V with respect to which the matrix of T is the diagonal matrix A, and let $\lambda_1, \ldots, \lambda_n \in \mathbf{F}$ be the diagonal entries of A. By the definition of the matrix of an operator, the k^{th} column of A gives the coefficients of Tv_k in the basis $v_1, \ldots, v_n$; because A is diagonal this says

$$Tv_k = \lambda_k v_k \qquad \text{for each } k = 1, \ldots, n.$$

Fix $\lambda \in \mathbf{F}$ and set

$$S = \{k \in \{1, \ldots, n\} : \lambda_k = \lambda\},$$

so that the number of times λ appears on the diagonal of A is exactly the number of elements of S. We claim that

$$E(\lambda, T) = \operatorname{span}(v_k : k \in S).$$

First we prove the inclusion $\supseteq$. If $k \in S$, then $Tv_k = \lambda_k v_k = \lambda v_k$, so $v_k \in E(\lambda, T)$. Because $E(\lambda, T) = \operatorname{null}(T - \lambda I)$ is a subspace of V (see 5.52), it contains the span of the v_k with $k \in S$.
Now we prove the inclusion $\subseteq$. Suppose $v \in E(\lambda, T)$. Because $v_1, \ldots, v_n$ is a basis of V, there exist unique $a_1, \ldots, a_n \in \mathbf{F}$ with

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Applying T and using $Tv_k = \lambda_k v_k$ gives

$$Tv = a_1 \lambda_1 v_1 + \cdots + a_n \lambda_n v_n,$$

while $Tv = \lambda v = a_1 \lambda v_1 + \cdots + a_n \lambda v_n$. Because the representation of the vector Tv as a linear combination of the basis $v_1, \ldots, v_n$ is unique, we may equate coefficients, obtaining

$$a_k \lambda_k = a_k \lambda, \qquad \text{that is,} \qquad a_k(\lambda_k - \lambda) = 0$$

for each k. Hence $a_k = 0$ whenever $\lambda_k \neq \lambda$, that is, whenever $k \notin S$. Thus $v = \sum_{k \in S} a_k v_k \in \operatorname{span}(v_k : k \in S)$, completing the proof of the claim.
Finally, the list $(v_k)_{k \in S}$ is a sublist of the linearly independent list $v_1, \ldots, v_n$ and hence is linearly independent. Being also a spanning list of $E(\lambda, T)$ by the claim, it is a basis of $E(\lambda, T)$. Therefore

$$\dim E(\lambda, T) = \#S = \text{the number of times } \lambda \text{ appears on the diagonal of } A,$$

as desired.
Note that the statement covers every $\lambda \in \mathbf{F}$, including those that are not eigenvalues of T: such a λ satisfies $\dim E(\lambda, T) = 0$ and indeed does not appear on the diagonal of A at all.

Problem 5D.3. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that if the operator T is diagonalizable, then $V = \operatorname{null} T \oplus \operatorname{range} T$.

Solution. Suppose T is diagonalizable. By 5.55(b) there is a basis $v_1, \dots, v_n$ of V consisting of eigenvectors of T; say

$$Tv_k = \lambda_k v_k \qquad \text{for each } k = 1, \dots, n,$$

where $\lambda_1, \dots, \lambda_n \in \mathbf{F}$ (these need not be distinct). Split the index set according to whether the corresponding scalar is zero:

$$S = \{k : \lambda_k = 0\}, \qquad S' = \{k : \lambda_k \neq 0\}.$$

Set $U = \operatorname{span}(v_k : k \in S)$ and $W = \operatorname{span}(v_k : k \in S')$.
First, $V = U \oplus W$. Indeed $U + W$ contains every v_k and hence equals V. Moreover, if $u + w = 0$ with $u \in U$ and $w \in W$, then writing $u = \sum_{k \in S} a_k v_k$ and $w = \sum_{k \in S'} a_k v_k$ we get $\sum_{k=1}^{n} a_k v_k = 0$, so all $a_k = 0$ by the linear independence of $v_1, \dots, v_n$; hence $u = w = 0$. By 1.45 the sum $U + W$ is a direct sum, so $V = U \oplus W$.
Next, $U = \operatorname{null} T$. This is the case $\lambda = 0$ of the claim proved in Exercise 2 of this section, since $\operatorname{null} T = \operatorname{null}(T - 0I) = E(0, T)$. For completeness, here is the argument again. If $k \in S$ then $Tv_k = 0 \cdot v_k = 0$, so $U \subseteq \operatorname{null} T$. Conversely, if $v = \sum_{k=1}^{n} a_k v_k$ satisfies $Tv = 0$, then $\sum_{k=1}^{n} a_k \lambda_k v_k = 0$, so $a_k \lambda_k = 0$ for every k by linear independence; hence $a_k = 0$ for every $k \in S'$, and thus $v \in U$.
Finally, $W = \operatorname{range} T$. Because $v_1, \dots, v_n$ spans V, we have

$$\operatorname{range} T = \operatorname{span}(Tv_1, \dots, Tv_n) = \operatorname{span}(\lambda_1 v_1, \dots, \lambda_n v_n).$$

The terms with $k \in S$ contribute the zero vector, and for $k \in S'$ we have $\operatorname{span}(\lambda_k v_k) = \operatorname{span}(v_k)$ because $\lambda_k \neq 0$. Hence $\operatorname{range} T = \operatorname{span}(v_k : k \in S') = W$.
Combining the three displayed facts gives

$$V = U \oplus W = \operatorname{null} T \oplus \operatorname{range} T,$$

as desired.
(Both extreme cases are allowed: if $S = \varnothing$ then $\operatorname{null} T = \{0\}$ and T is invertible, while if $S' = \varnothing$ then $T = 0$.)

Problem 5D.4. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that the following are equivalent.
(a) $V = \operatorname{null} T \oplus \operatorname{range} T$.
(b) $V = \operatorname{null} T + \operatorname{range} T$.
(c) $\operatorname{null} T \cap \operatorname{range} T = \{0\}$.

Solution. Throughout, write $N = \operatorname{null} T$ and $R = \operatorname{range} T$; these are subspaces of V, and by the fundamental theorem of linear maps (3.21),

$$\dim N + \dim R = \dim V.$$

We prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a).
(a) $\Rightarrow$ (b). The notation $N \oplus R$ means the sum $N + R$ together with the assertion that this sum is a direct sum. So if $V = N \oplus R$, then in particular $V = N + R$.
(b) $\Rightarrow$ (c). Suppose $V = N + R$. By the formula for the dimension of a sum (2.43),

$$\dim V = \dim(N + R) = \dim N + \dim R - \dim(N \cap R) = \dim V - \dim(N \cap R),$$

where the last equality uses the fundamental theorem of linear maps displayed above. Hence $\dim(N \cap R) = 0$. A vector space of dimension 0 is the zero space (the empty list is a basis of it, so it has no nonzero vectors). Thus $N \cap R = \{0\}$.

(c) $\Rightarrow$ (a). Suppose $N \cap R = \{0\}$. By 1.46, the sum $N + R$ is a direct sum, so $N \oplus R$ makes sense and is a subspace of V (a sum of two subspaces of V is a subspace of V). By 2.43,

$$\dim(N \oplus R) = \dim N + \dim R - \dim(N \cap R) = \dim N + \dim R = \dim V,$$

again using 3.21 for the last equality. Thus $N \oplus R$ is a subspace of the finite-dimensional space V having the same dimension as V, so $N \oplus R = V$ by 2.39. Hence $V = \operatorname{null} T \oplus \operatorname{range} T$, which is (a). This completes the cycle of implications, so (a), (b), and (c) are equivalent.
Note that finite-dimensionality is essential. For example, let $V = \mathcal{P}(\mathbf{R})$ and let T be the differentiation operator $Tp = p'$. Then $\operatorname{range} T = \mathcal{P}(\mathbf{R}) = V$, so (b) holds, but $\operatorname{null} T$ is the set of constant polynomials, which is not $\{0\}$ and is contained in $\operatorname{range} T$; hence (c) fails.

Problem 5D.5. Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Prove that T is diagonalizable if and only if

$$V = \operatorname{null}(T - \lambda I) \oplus \operatorname{range}(T - \lambda I)$$

for every $\lambda \in \mathbf{C}$.

Solution. If $V = \{0\}$ then both conditions hold trivially, so assume $V \neq \{0\}$.
First suppose T is diagonalizable, and let $\lambda \in \mathbf{C}$. By 5.55(b) there is a basis $v_1, \ldots, v_n$ of V consisting of eigenvectors of T, say $Tv_k = \lambda_k v_k$. Then

$$(T - \lambda I)v_k = (\lambda_k - \lambda)v_k$$

for each k, so the very same basis consists of eigenvectors of the operator $T - \lambda I$. Hence $T - \lambda I$ is diagonalizable, again by 5.55. Applying Exercise 3 of this section to the operator $T - \lambda I$ now gives

$$V = \operatorname{null}(T - \lambda I) \oplus \operatorname{range}(T - \lambda I),$$

as desired. (This direction did not use $\mathbf{F} = \mathbf{C}$.)
To prove the other direction, suppose

$$V = \operatorname{null}(T - \lambda I) \oplus \operatorname{range}(T - \lambda I)$$

for every $\lambda \in \mathbf{C}$. By 1.46, this hypothesis gives

$$\operatorname{null}(T - \lambda I) \cap \operatorname{range}(T - \lambda I) = \{0\} \qquad \text{for every } \lambda \in \mathbf{C}.$$

Let p be the minimal polynomial of T. Because $V \neq \{0\}$, p is nonconstant, so by the fundamental theorem of algebra in the factored form 4.13 we may write

$$p(z) = (z - \lambda_1)^{m_1} \cdots (z - \lambda_k)^{m_k},$$

where $\lambda_1, \ldots, \lambda_k \in \mathbf{C}$ are distinct and each m_j is a positive integer. (Here we use that $\mathbf{F} = \mathbf{C}$.)
We claim each $m_j = 1$. Suppose instead that $m_j \geq 2$ for some j; relabel so that $j = 1$. Define polynomials

$$r(z) = \frac{p(z)}{(z - \lambda_1)^2}, \qquad q(z) = \frac{p(z)}{z - \lambda_1} = (z - \lambda_1)r(z);$$

both are genuine polynomials because $m_1 \geq 2$. Now q is a nonzero polynomial with $\deg q = \deg p - 1 < \deg p$, so q is not a polynomial multiple of p (a nonzero polynomial multiple of p has degree at least $\deg p$). Hence $q(T) \neq 0$ by 5.29. Choose $v \in V$ with

$$q(T)v \neq 0,$$

and set $w = r(T)v$ and $u = (T - \lambda_1 I)w$. By 5.17,

$$u = (T - \lambda_1 I)r(T)v = q(T)v \neq 0.$$

By construction $u \in \operatorname{range}(T - \lambda_1 I)$. Moreover, using 5.17 again,

$$(T - \lambda_1 I)u = (T - \lambda_1 I)^2 r(T)v = p(T)v = 0,$$

because $p(T) = 0$. Hence $u \in \operatorname{null}(T - \lambda_1 I)$. Thus u is a nonzero vector lying in

$$\operatorname{null}(T - \lambda_1 I) \cap \operatorname{range}(T - \lambda_1 I),$$

contradicting the displayed consequence of our hypothesis (taken with $\lambda = \lambda_1$).
Therefore $m_1 = \cdots = m_k = 1$, so the minimal polynomial of T is

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_k)$$

with $\lambda_1, \ldots, \lambda_k$ distinct. By 5.62, T is diagonalizable.

Problem 5D.6. Suppose $T \in \mathcal{L}(\mathbf{F}^5)$ and $\dim E(8, T) = 4$. Prove that $T - 2I$ or $T - 6I$ is invertible.

Solution. Suppose, seeking a contradiction, that neither $T - 2I$ nor $T - 6I$ is invertible.
Because $\mathbf{F}^5$ is finite-dimensional, an operator on it is invertible if and only if it is injective (by 3.65).
Hence $T - 2I$ is not injective, so $\operatorname{null}(T - 2I) \neq \{0\}$; that is,

$$E(2, T) \neq \{0\}, \qquad \text{so} \qquad \dim E(2, T) \geq 1.$$

Likewise $T - 6I$ is not injective, so

$$E(6, T) \neq \{0\}, \qquad \text{so} \qquad \dim E(6, T) \geq 1.$$

Thus 2 and 6 are eigenvalues of T. Also $\dim E(8, T) = 4 \neq 0$, so 8 is an eigenvalue of T.
Now 2, 6, 8 are three distinct eigenvalues of T, so 5.54 applies and gives

$$\dim E(2, T) + \dim E(6, T) + \dim E(8, T) \leq \dim \mathbf{F}^5 = 5.$$

But the left side is at least $1 + 1 + 4 = 6$. This contradiction shows that our supposition was false.
Hence $T - 2I$ is invertible or $T - 6I$ is invertible.
(The same argument shows more: with $\dim E(8, T) = 4$, the operator $T - \lambda I$ can fail to be invertible for at most one $\lambda \in \mathbf{F} \setminus \{8\}$, since two such λ would force the eigenspace dimensions to sum to at least $6 > 5$.)

Problem 5D.7. Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that

$$E(\lambda, T) = E\big(\tfrac{1}{\lambda}, T^{-1}\big)$$

for every $\lambda \in \mathbf{F}$ with $\lambda \neq 0$.

Solution. Fix $\lambda \in \mathbf{F}$ with $\lambda \neq 0$. Recall from 5.52 that

$$E(\lambda, T) = \{v \in V : Tv = \lambda v\}, \qquad E\big(\tfrac{1}{\lambda}, T^{-1}\big) = \big\{v \in V : T^{-1}v = \tfrac{1}{\lambda}v\big\},$$

so it suffices to show that for $v \in V$,

$$Tv = \lambda v \iff T^{-1}v = \tfrac{1}{\lambda}v.$$

First suppose $v \in E(\lambda, T)$, so that $Tv = \lambda v$. Apply the linear map T^{-1} to both sides:

$$v = T^{-1}(Tv) = T^{-1}(\lambda v) = \lambda\, T^{-1}v.$$

Because $\lambda \neq 0$, we may divide by λ, obtaining $T^{-1}v = \frac{1}{\lambda}v$. Hence $v \in E\big(\frac{1}{\lambda}, T^{-1}\big)$.

Conversely, suppose $v \in E\left(\frac{1}{\lambda}, T^{-1}\right)$, so that $T^{-1}v = \frac{1}{\lambda}v$. Apply T to both sides:

$$v = T(T^{-1}v) = T\left(\tfrac{1}{\lambda}v\right) = \tfrac{1}{\lambda}\,Tv.$$

Multiplying both sides by λ gives $Tv = \lambda v$, so $v \in E(\lambda, T)$.
The two inclusions give $E(\lambda, T) = E\left(\frac{1}{\lambda}, T^{-1}\right)$, as desired.
Two remarks. First, the hypothesis $\lambda \neq 0$ is needed only to make $\frac{1}{\lambda}$ meaningful; and in fact 0 is never an eigenvalue of an invertible operator, because $Tv = 0$ with T injective forces $v = 0$, so $E(0, T) = \{0\}$.
Second, the result says in particular that λ is an eigenvalue of T if and only if $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}, and that the corresponding eigenspaces are literally the same subspace of V (in particular they have the same dimension). No finite-dimensionality was used.

Problem 5D.8. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct nonzero eigenvalues of T. Prove that

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim \operatorname{range} T.$$

Solution. If T has no nonzero eigenvalues then the left side is an empty sum, equal to 0, and the inequality is trivial. So assume $m \geq 1$.
Step 1: each eigenspace of a nonzero eigenvalue sits inside the range. Fix $k \in \{1, \ldots, m\}$ and let $v \in E(\lambda_k, T)$, so $Tv = \lambda_k v$. Because $\lambda_k \neq 0$ we may divide by it:

$$v = T\left(\tfrac{1}{\lambda_k}v\right),$$

which exhibits v as an element of $\operatorname{range} T$. Hence

$$E(\lambda_k, T) \subseteq \operatorname{range} T \quad \text{for each } k \in \{1, \ldots, m\}.$$

Consequently the sum $U := E(\lambda_1, T) + \cdots + E(\lambda_m, T)$, being the set of all sums of elements of these subspaces, is also contained in the subspace $\operatorname{range} T$.
Step 2: the sum is direct, so its dimension is the sum of the dimensions. The numbers $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T, so by 5.54 the sum $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum. Now 3.94 (a sum is a direct sum if and only if dimensions add up), applied to the subspaces $E(\lambda_1, T), \ldots, E(\lambda_m, T)$ of the finite-dimensional vector space V, gives

$$\dim U = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T).$$

Step 3: combine. Since U is a subspace of $\operatorname{range} T$ and $\operatorname{range} T$ is finite-dimensional (it is a subspace of the finite-dimensional space V), 2.37 gives $\dim U \leq \dim \operatorname{range} T$. Therefore

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) = \dim U \leq \dim \operatorname{range} T,$$

as desired.
Remark: the hypothesis that the eigenvalues are nonzero is essential in Step 1. For instance, if T is the zero operator on a nonzero space V, then $E(0, T) = V$ has dimension $\dim V > 0 = \dim \operatorname{range} T$.

Problem 5D.9. Suppose $R, T \in \mathcal{L}(\mathbf{F}^3)$ each have $2, 6, 7$ as eigenvalues. Prove that there exists an invertible operator $S \in \mathcal{L}(\mathbf{F}^3)$ such that $R = S^{-1}TS$.

Solution. Write $\lambda_1 = 2$, $\lambda_2 = 6$, $\lambda_3 = 7$.
Step 1: both operators are diagonalizable. The operator R has the three distinct eigenvalues $2, 6, 7$, and $\dim \mathbf{F}^3 = 3$. Thus R has $\dim \mathbf{F}^3$ distinct eigenvalues, so R is diagonalizable by 5.58. The same argument applies to T.
Therefore there is a basis u_1, u_2, u_3 of $\mathbf{F}^3$ with $Ru_k = \lambda_k u_k$ for each k, and a basis v_1, v_2, v_3 of

$\mathbf{F}^3$ with $Tv_k = \lambda_k v_k$ for each k. (Concretely: pick an eigenvector for each of $2, 6, 7$; eigenvectors corresponding to distinct eigenvalues are linearly independent by 5.11, and a linearly independent list of length $3 = \dim \mathbf{F}^3$ is a basis by 2.38. The labelling can be chosen so that the k-th vector of each list corresponds to the same eigenvalue λ_k.)
Step 2: build S. Since u_1, u_2, u_3 is a basis of $\mathbf{F}^3$, there is a unique linear map $S \in \mathcal{L}(\mathbf{F}^3)$ with

$$Su_k = v_k \quad \text{for } k = 1, 2, 3$$

(by 3.4, the existence and uniqueness of a linear map with prescribed values on a basis). The range of S contains v_1, v_2, v_3, which span $\mathbf{F}^3$; hence S is surjective, and therefore invertible by 3.65.
Step 3: verify the similarity. For each k,

$$(S^{-1}TS)u_k = S^{-1}Tv_k = S^{-1}(\lambda_k v_k) = \lambda_k S^{-1}v_k = \lambda_k u_k = Ru_k.$$

Thus the linear maps $S^{-1}TS$ and R agree on the basis u_1, u_2, u_3 of $\mathbf{F}^3$, and hence they are equal. Therefore $R = S^{-1}TS$ with S invertible, as required.

Problem 5D.10. Find $R, T \in \mathcal{L}(\mathbf{F}^4)$ such that R and T each have $2, 6, 7$ as eigenvalues, R and T have no other eigenvalues, and there does not exist an invertible operator $S \in \mathcal{L}(\mathbf{F}^4)$ such that $R = S^{-1}TS$.

Solution. Define $R, T \in \mathcal{L}(\mathbf{F}^4)$ by

$$R(z_1, z_2, z_3, z_4) = (2z_1,\ 6z_2,\ 7z_3,\ 7z_4),$$

$$T(z_1, z_2, z_3, z_4) = (2z_1,\ 6z_2,\ 7z_3 + z_4,\ 7z_4).$$

We show these work.
Step 1: eigenvalues. With respect to the standard basis e_1, e_2, e_3, e_4 of $\mathbf{F}^4$, the matrices of R and T are

$$\mathcal{M}(R) = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix}, \qquad \mathcal{M}(T) = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 7 & 1 \\ 0 & 0 & 0 & 7 \end{pmatrix}.$$

Both are upper triangular with diagonal entries $2, 6, 7, 7$. By 5.41, the eigenvalues of an operator having an upper-triangular matrix are exactly the diagonal entries of that matrix. Hence R and T each have eigenvalues $2, 6, 7$ and no others.
Step 2: the 7-eigenspaces have different dimensions. Clearly $E(7, R) = \operatorname{span}(e_3, e_4)$, so $\dim E(7, R) = 2$. For T: the equation $T(z_1, z_2, z_3, z_4) = 7(z_1, z_2, z_3, z_4)$ says

$$2z_1 = 7z_1, \quad 6z_2 = 7z_2, \quad 7z_3 + z_4 = 7z_3, \quad 7z_4 = 7z_4.$$

The first two equations force $z_1 = z_2 = 0$, and the third forces $z_4 = 0$. Thus $E(7, T) = \operatorname{span}(e_3)$ and $\dim E(7, T) = 1$.
Step 3: similar operators have eigenspaces of equal dimension. Suppose, for contradiction, that $S \in \mathcal{L}(\mathbf{F}^4)$ is invertible and $R = S^{-1}TS$; equivalently, $SR = TS$. Let $\lambda \in \mathbf{F}$ and let $v \in E(\lambda, R)$, so $Rv = \lambda v$. Then

$$T(Sv) = (TS)v = (SR)v = S(Rv) = S(\lambda v) = \lambda(Sv),$$

so $Sv \in E(\lambda, T)$. Hence S maps $E(\lambda, R)$ into $E(\lambda, T)$, and the restriction of S to $E(\lambda, R)$ is injective because S is injective. An injective linear map from one finite-dimensional space to another forces

$$\dim E(\lambda, R) \le \dim E(\lambda, T)$$

(by the fundamental theorem of linear maps 3.21, since the null space of the restriction is $\{0\}$).
Applying this with $\lambda = 7$ gives $2 = \dim E(7, R) \le \dim E(7, T) = 1$, a contradiction. Therefore no invertible $S \in \mathcal{L}(\mathbf{F}^4)$ satisfies $R = S^{-1}TS$.

Remark: equivalently, R is diagonalizable (its matrix above is already diagonal) while T is not, since

$$\dim E(2,T) + \dim E(6,T) + \dim E(7,T) = 1 + 1 + 1 = 3 < 4 = \dim \mathbf{F}^4,$$

so condition (d) of 5.55 fails for T. Diagonalizability is preserved by similarity (if $u_1, \ldots, u_4$ is a basis of eigenvectors of R and $SR = TS$ with S invertible, then $Su_1, \ldots, Su_4$ is a basis of eigenvectors of T), so again R and T cannot be similar. This works over $\mathbf{F} = \mathbf{R}$ and over $\mathbf{F} = \mathbf{C}$ alike.

Problem 5D.11. Find $T \in \mathcal{L}(\mathbf{C}^3)$ such that 6 and 7 are eigenvalues of T and such that T does not have a diagonal matrix with respect to any basis of $\mathbf{C}^3$.

Solution. Define $T \in \mathcal{L}(\mathbf{C}^3)$ by

$$T(z_1, z_2, z_3) = (6z_1 + z_2,\ 6z_2,\ 7z_3).$$

Step 1: 6 and 7 are eigenvalues. With respect to the standard basis e_1, e_2, e_3 of $\mathbf{C}^3$, the matrix of T is

$$\mathcal{M}(T) = \begin{pmatrix} 6 & 1 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 7 \end{pmatrix},$$

which is upper triangular with diagonal entries $6, 6, 7$. By 5.41 the eigenvalues of T are exactly 6 and 7. (Directly: $Te_1 = 6e_1$ and $Te_3 = 7e_3$, so 6 and 7 are indeed eigenvalues.)
Step 2: T is not diagonalizable. Compute the eigenspaces. The equation $T(z_1, z_2, z_3) = 6(z_1, z_2, z_3)$ reads

$$6z_1 + z_2 = 6z_1, \quad 6z_2 = 6z_2, \quad 7z_3 = 6z_3,$$

which forces $z_2 = 0$ and $z_3 = 0$. Hence $E(6,T) = \operatorname{span}(e_1)$, of dimension 1.
The equation $T(z_1, z_2, z_3) = 7(z_1, z_2, z_3)$ reads

$$6z_1 + z_2 = 7z_1, \quad 6z_2 = 7z_2, \quad 7z_3 = 7z_3,$$

so $z_2 = 0$ from the second equation and then $z_1 = 0$ from the first. Hence $E(7,T) = \operatorname{span}(e_3)$, of dimension 1.
Since 6 and 7 are the only eigenvalues of T,

$$\dim E(6,T) + \dim E(7,T) = 1 + 1 = 2 < 3 = \dim \mathbf{C}^3,$$

so condition (d) of 5.55 fails. Because the four conditions of 5.55 are equivalent, condition (a) fails too: T is not diagonalizable, i.e. T has a diagonal matrix with respect to no basis of $\mathbf{C}^3$.
Alternative verification via the minimal polynomial: suppose T were diagonalizable. Then by 5.62 the minimal polynomial of T equals $(z - \mu_1) \cdots (z - \mu_r)$ for some distinct $\mu_1, \ldots, \mu_r \in \mathbf{C}$. The zeros of the minimal polynomial are exactly the eigenvalues of T (by 5.27), and by Step 1 those eigenvalues are 6 and 7; hence $\{\mu_1, \ldots, \mu_r\} = \{6, 7\}$ and the minimal polynomial of T would be $(z - 6)(z - 7)$, forcing $(T - 6I)(T - 7I) = 0$. But $Te_2 = e_1 + 6e_2$, so $(T - 7I)e_2 = e_1 - e_2$, while $(T - 6I)e_1 = 0$ and $(T - 6I)e_2 = e_1$; therefore

$$(T - 6I)(T - 7I)e_2 = (T - 6I)(e_1 - e_2) = 0 - e_1 = -e_1 \neq 0.$$

Thus $(T - 6I)(T - 7I) \neq 0$, a contradiction. Hence T is not diagonalizable.

Problem 5D.12. Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is such that 6 and 7 are eigenvalues of T. Furthermore, suppose T does not have a diagonal matrix with respect to any basis of $\mathbf{C}^3$. Prove that there exists

$(z_1, z_2, z_3) \in \mathbf{C}^3$ such that

$$T(z_1, z_2, z_3) = (6 + 8z_1,\ 7 + 8z_2,\ 13 + 8z_3).$$

Solution. Reformulate the desired conclusion. For $z = (z_1, z_2, z_3) \in \mathbf{C}^3$,

$$Tz = (6 + 8z_1,\ 7 + 8z_2,\ 13 + 8z_3) \iff Tz - 8z = (6, 7, 13) \iff (T - 8I)z = (6, 7, 13).$$

So we must show that $(6, 7, 13) \in \text{range}(T - 8I)$. We prove the stronger statement that $T - 8I$ is surjective.
Step 1: 8 is not an eigenvalue of T. Suppose it were. Then 6, 7, 8 would be three distinct eigenvalues of T, and $\dim \mathbf{C}^3 = 3$, so T would have $\dim \mathbf{C}^3$ distinct eigenvalues. By 5.58, T would then be diagonalizable, i.e. T would have a diagonal matrix with respect to some basis of $\mathbf{C}^3$. This contradicts the hypothesis. Hence 8 is not an eigenvalue of T.
Step 2: $T - 8I$ is surjective. Because 8 is not an eigenvalue of T, there is no nonzero $v \in \mathbf{C}^3$ with $Tv = 8v$; that is, $\text{null}(T - 8I) = \{0\}$, so $T - 8I$ is injective (by 3.15, a linear map is injective if and only if its null space is $\{0\}$). Since $T - 8I$ is an operator on the finite-dimensional vector space $\mathbf{C}^3$, injectivity is equivalent to surjectivity by 3.65. Hence $T - 8I$ is surjective.
Step 3: conclude. By Step 2 there exists $z = (z_1, z_2, z_3) \in \mathbf{C}^3$ with

$$(T - 8I)(z_1, z_2, z_3) = (6, 7, 13),$$

and by the reformulation at the start this means

$$T(z_1, z_2, z_3) = (6 + 8z_1,\ 7 + 8z_2,\ 13 + 8z_3),$$

as required.
Remark: nothing is special about the vector $(6, 7, 13)$; the argument shows that for every $w \in \mathbf{C}^3$ there is a $z \in \mathbf{C}^3$ with $Tz = w + 8z$. What matters is only that $8 \notin \{6, 7\}$, since a third distinct eigenvalue would force diagonalizability on a 3-dimensional space.

Problem 5D.13. Suppose A is a diagonal matrix with distinct entries on the diagonal and B is a matrix of the same size as A. Show that $AB = BA$ if and only if B is a diagonal matrix.

Solution. Say A and B are n-by-n matrices, and write $A_{j,k}$ and $B_{j,k}$ for their entries. That A is diagonal with distinct diagonal entries means: there exist $\lambda_1, \ldots, \lambda_n \in \mathbf{F}$, distinct from one another, with

$$A_{j,k} = \begin{cases} \lambda_j & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Compute both products entrywise using the entry formula for matrix multiplication (3.47). For $j, k \in \{1, \ldots, n\}$,

$$(AB)_{j,k} = \sum_{r=1}^{n} A_{j,r} B_{r,k} = A_{j,j} B_{j,k} = \lambda_j B_{j,k},$$

where the middle equality holds because $A_{j,r} = 0$ for every $r \neq j$. Similarly,

$$(BA)_{j,k} = \sum_{r=1}^{n} B_{j,r} A_{r,k} = B_{j,k} A_{k,k} = \lambda_k B_{j,k},$$

because $A_{r,k} = 0$ for every $r \neq k$.
Two matrices of the same size are equal exactly when all their corresponding entries are equal. Hence

$$AB = BA \iff \lambda_j B_{j,k} = \lambda_k B_{j,k} \text{ for all } j, k \iff (\lambda_j - \lambda_k) B_{j,k} = 0 \text{ for all } j, k.$$

Now suppose $AB = BA$, and let $j \neq k$. Since the diagonal entries of A are distinct, $\lambda_j - \lambda_k \neq 0$, and the displayed condition forces $B_{j,k} = 0$. Thus every off-diagonal entry of B vanishes, which says exactly that B is a diagonal matrix.
Conversely, suppose B is diagonal. If $j \neq k$ then $B_{j,k} = 0$, so $(\lambda_j - \lambda_k)B_{j,k} = 0$; if $j = k$ then $\lambda_j - \lambda_k = 0$, so again $(\lambda_j - \lambda_k)B_{j,k} = 0$. Hence the condition above holds for all j, k, and therefore $AB = BA$. (Explicitly, both products are the diagonal matrix whose j-th diagonal entry is $\lambda_j B_{j,j}$.)
Note that the distinctness of the diagonal entries of A is needed only for the forward implication; the converse holds for any diagonal A and B. Distinctness is genuinely necessary: if $A = I$, then $AB = BA$ for every B, diagonal or not.

Problem 5D.14. (a) Give an example of a finite-dimensional complex vector space and an operator T on that vector space such that T^2 is diagonalizable but T is not diagonalizable.
(b) Suppose $\mathbf{F} = \mathbf{C}$, k is a positive integer, and $T \in \mathcal{L}(V)$ is invertible. Prove that T is diagonalizable if and only if T^k is diagonalizable.

Solution. (a) Take $V = \mathbf{C}^2$ and define $T \in \mathcal{L}(\mathbf{C}^2)$ by

$$T(z_1, z_2) = (z_2, 0).$$

Then $T^2(z_1, z_2) = T(z_2, 0) = (0, 0)$, so $T^2 = 0$. The matrix of the zero operator with respect to any basis is the zero matrix, which is diagonal; hence T^2 is diagonalizable.
But T is not diagonalizable. Indeed, the matrix of T with respect to the standard basis is upper triangular with both diagonal entries equal to 0, so by 5.41 the only eigenvalue of T is 0. Moreover

$$E(0, T) = \operatorname{null} T = \{(z_1, z_2) \in \mathbf{C}^2 : z_2 = 0\} = \operatorname{span}\big((1, 0)\big),$$

which has dimension 1. Thus $\dim E(0, T) = 1 < 2 = \dim \mathbf{C}^2$, so condition (d) of 5.55 fails and T is not diagonalizable.
(This also shows the invertibility hypothesis in (b) cannot be dropped: here T is not invertible.)
(b) Throughout, V is finite-dimensional (as is implicit in speaking of diagonalizability and of the minimal polynomial), $\mathbf{F} = \mathbf{C}$, $k \geq 1$, and $T \in \mathcal{L}(V)$ is invertible.
First suppose T is diagonalizable. Then V has a basis $v_1, \ldots, v_n$ consisting of eigenvectors of T, say $Tv_j = \lambda_j v_j$. Then

$$T^k v_j = \lambda_j^k v_j$$

for each j, so the same basis consists of eigenvectors of T^k. Hence T^k is diagonalizable. (This direction uses neither $\mathbf{F} = \mathbf{C}$ nor invertibility.)
Now suppose T^k is diagonalizable; we show T is diagonalizable.
Step 1: the minimal polynomial of T^k splits into distinct linear factors with nonzero roots. By 5.62, since T^k is diagonalizable its minimal polynomial equals

$$(z - \mu_1) \cdots (z - \mu_m)$$

for some distinct $\mu_1, \ldots, \mu_m \in \mathbf{C}$. Because T is invertible, so is T^k, and hence 0 is not an eigenvalue of T^k. By 5.27, the zeros of the minimal polynomial of T^k are precisely the eigenvalues of T^k; thus $\mu_j \neq 0$ for each j.
Step 2: produce a polynomial annihilating T with no repeated zeros. Define

$$p(z) = (z^k - \mu_1)(z^k - \mu_2) \cdots (z^k - \mu_m).$$

Because the map $q \mapsto q(T)$ from $\mathcal{P}(\mathbf{C})$ to $\mathcal{L}(V)$ is linear and multiplicative (5.17: $(q_1 q_2)(T) = q_1(T) q_2(T)$), and because z^k evaluated at T is T^k, evaluating the product p at T gives

$$p(T) = (T^k - \mu_1 I) \cdots (T^k - \mu_m I),$$

which is exactly the minimal polynomial of T^k evaluated at T^k. Hence $p(T) = 0$.

We claim p has km distinct zeros. Fix j. Since $\mu_j \neq 0$, write $\mu_j = r_j e^{i\theta_j}$ with $r_j > 0$; the equation $z^k = \mu_j$ has exactly the k distinct solutions

$$r_j^{1/k} e^{i(\theta_j + 2\pi t)/k}, \qquad t = 0, 1, \ldots, k-1,$$

so $z^k - \mu_j$ has k distinct zeros. Moreover, if α is a zero of $z^k - \mu_j$ and also of $z^k - \mu_{j'}$, then $\mu_j = \alpha^k = \mu_{j'}$, forcing $j = j'$ since $\mu_1, \ldots, \mu_m$ are distinct. So the zero sets of the m factors are pairwise disjoint, and p has exactly km distinct zeros. As $\deg p = km$ and p is monic, the factorization of p guaranteed by 4.13 must be

$$p(z) = (z - \alpha_1)(z - \alpha_2) \cdots (z - \alpha_{km})$$

with $\alpha_1, \ldots, \alpha_{km}$ distinct. In particular p has no repeated zeros: there is no $\alpha \in \mathbf{C}$ with $(z - \alpha)^2$ dividing p (such an α would have to occur at least twice among $\alpha_1, \ldots, \alpha_{km}$, by the uniqueness of the factorization in 4.13).

Step 3: transfer this to the minimal polynomial of T. Let q denote the minimal polynomial of T. Since $p(T) = 0$, 5.29 shows that p is a polynomial multiple of q: there is $s \in \mathcal{P}(\mathbf{C})$ with $p = qs$.

Because $\mathbf{F} = \mathbf{C}$ and q is monic, 4.13 gives

$$q(z) = (z - \lambda_1)^{d_1} \cdots (z - \lambda_r)^{d_r}$$

with $\lambda_1, \ldots, \lambda_r \in \mathbf{C}$ distinct and each $d_i \geq 1$. If some $d_i \geq 2$, then $(z - \lambda_i)^2$ divides q, hence divides $p = qs$, contradicting the conclusion of Step 2 that p has no repeated zeros. Therefore $d_i = 1$ for every i, so

$$q(z) = (z - \lambda_1) \cdots (z - \lambda_r)$$

with $\lambda_1, \ldots, \lambda_r$ distinct. By 5.62, T is diagonalizable.

This completes both implications, so T is diagonalizable if and only if T^k is.

Remark: the hypothesis $\mathbf{F} = \mathbf{C}$ is used twice in the harder direction: to split $z^k - \mu_j$ into k distinct linear factors, and to factor q into linear factors. Over $\mathbf{R}$ the statement fails: rotation by 90 degrees on $\mathbf{R}^2$ is invertible and not diagonalizable (it has no eigenvalues), yet its square is $-I$, which is diagonalizable. Invertibility is used only to guarantee $\mu_j \neq 0$, and part (a) shows it cannot be omitted.

Problem 5D.15. Suppose V is a finite-dimensional complex vector space, $T \in \mathcal{L}(V)$, and p is the minimal polynomial of T. Prove that the following are equivalent.
(a) T is diagonalizable.
(b) There does not exist $\lambda \in \mathbf{C}$ such that p is a polynomial multiple of $(z - \lambda)^2$.
(c) p and its derivative p' have no zeros in common.
(d) The greatest common divisor of p and p' is the constant polynomial 1.
The greatest common divisor of p and p' is the monic polynomial q of largest degree such that p and p' are both polynomial multiples of q. The Euclidean algorithm for polynomials (look it up) can quickly determine the greatest common divisor of two polynomials, without requiring any information about the zeros of the polynomials. Thus the equivalence of (a) and (d) above shows that we can determine whether T is diagonalizable without knowing anything about the zeros of p.

Solution. Throughout, p denotes the minimal polynomial of T, which exists and is monic because V is finite-dimensional (by 5.22 and 5.24).

We first dispose of the degenerate case $V = \{0\}$, which the standing assumptions of Chapter 5 do not exclude. If $V = \{0\}$, then $I = 0$, so the constant polynomial 1 satisfies $p(T) = 0$ and hence $p = 1$. Then: T has a diagonal matrix with respect to the empty basis of V, so (a) holds; $p = 1$ is not a polynomial multiple of any $(z - \lambda)^2$, since a nonzero multiple of $(z - \lambda)^2$ has degree at least 2 and the zero polynomial is not p, so (b) holds; p has no zeros at all, so (c) holds; and the only monic polynomial of which the constant polynomial 1 is a multiple is 1 itself, so the greatest common divisor of p and p' is 1 and (d) holds. Thus all four statements are true and the equivalence is trivial.

So assume from now on that $V \neq \{0\}$. Then $I \neq 0$, so $\deg p \geq 1$. We prove (a) $\Leftrightarrow$ (b), then (b) $\Leftrightarrow$ (c),

then (c) $\Leftrightarrow$ (d).
Step 1: (a) $\Leftrightarrow$ (b).
First suppose (b) fails, so that $p = (z-\lambda)^2 r$ for some $\lambda \in \mathbf{C}$ and some $r \in \mathcal{P}(\mathbf{C})$. We show T is not diagonalizable. If T were diagonalizable, then by 5.62 we could write

$$p = (z-\mu_1)\cdots(z-\mu_m)$$

for some distinct $\mu_1, \ldots, \mu_m \in \mathbf{C}$. Since $p(\lambda) = 0$, we have $(\lambda-\mu_1)\cdots(\lambda-\mu_m) = 0$, so $\lambda = \mu_j$ for exactly one j (the μ_k are distinct). Put $s = \prod_{k\neq j}(z-\mu_k)$, so that $p = (z-\lambda)s$ and

$$s(\lambda) = \prod_{k\neq j}(\lambda-\mu_k) \neq 0.$$

On the other hand $p = (z-\lambda)\big((z-\lambda)r\big)$. Hence $(z-\lambda)\big(s-(z-\lambda)r\big) = 0$ in $\mathcal{P}(\mathbf{C})$. Because $\mathcal{P}(\mathbf{C})$ has no zero divisors (the degree of a product of nonzero polynomials is the sum of the degrees), we get $s = (z-\lambda)r$ and therefore $s(\lambda) = 0$, a contradiction. Thus T is not diagonalizable, and (a) fails.
Conversely, suppose (b) holds. Because V is a complex vector space, 5.27(b) tells us that

$$p = (z-\lambda_1)\cdots(z-\lambda_n)$$

for some $\lambda_1, \ldots, \lambda_n \in \mathbf{C}$, not necessarily distinct, with $n = \deg p \geq 1$. If $\lambda_j = \lambda_k$ for some $j \neq k$, then p is a polynomial multiple of $(z-\lambda_j)^2$, contradicting (b). Hence $\lambda_1, \ldots, \lambda_n$ are distinct, and 5.62 shows that T is diagonalizable. This proves (b) $\Rightarrow$ (a), so (a) and (b) are equivalent.
Step 2: (b) $\Leftrightarrow$ (c).
Suppose (b) fails: $p = (z-\lambda)^2 r$ for some $\lambda \in \mathbf{C}$ and $r \in \mathcal{P}(\mathbf{C})$. Differentiating,

$$p' = 2(z-\lambda)r + (z-\lambda)^2 r'.$$

Hence $p(\lambda) = 0$ and $p'(\lambda) = 0$, so λ is a common zero of p and p'; thus (c) fails.
Conversely, suppose (c) fails, so there is $\lambda \in \mathbf{C}$ with $p(\lambda) = p'(\lambda) = 0$. Because λ is a zero of p, the factor theorem (4.6) gives $q \in \mathcal{P}(\mathbf{C})$ with $p = (z-\lambda)q$. Differentiating,

$$p' = q + (z-\lambda)q',$$

so $0 = p'(\lambda) = q(\lambda)$. Now $q \neq 0$, because $p = (z-\lambda)q$ and $p \neq 0$; a nonzero polynomial with a zero cannot be constant, so $\deg q \geq 1$ and 4.6 applies to q. It gives $q = (z-\lambda)r$ for some $r \in \mathcal{P}(\mathbf{C})$, and therefore $p = (z-\lambda)^2 r$, so (b) fails. Hence (b) and (c) are equivalent.
Step 3: (c) $\Leftrightarrow$ (d).
First note that the greatest common divisor is well defined here. We have $\deg p \geq 1$, and writing $p = z^d + a_{d-1}z^{d-1} + \cdots + a_0$ with $d = \deg p$ we get $p' = dz^{d-1} + \cdots$, whose leading coefficient d is a nonzero complex number; hence p' is a nonzero polynomial with $\deg p' = \deg p - 1$. The constant polynomial 1 is a monic common divisor of p and p', and any monic q of which the nonzero polynomial p' is a multiple satisfies $\deg q \leq \deg p'$. So among the monic common divisors of p and p' there is one of largest degree, and the greatest common divisor exists.
Suppose (c) fails, so $p(\lambda) = p'(\lambda) = 0$ for some $\lambda \in \mathbf{C}$. By 4.6 both p and p' are polynomial multiples of $z-\lambda$, which is monic of degree 1. Hence the greatest common divisor of p and p' has degree at least 1 and so is not the constant polynomial 1; thus (d) fails.
Suppose (d) fails, so the greatest common divisor q of p and p' satisfies $\deg q \geq 1$. By the fundamental theorem of algebra (4.12), q has a zero $\lambda \in \mathbf{C}$. Since p and p' are both polynomial multiples of q, we get $p(\lambda) = p'(\lambda) = 0$, so p and p' have a common zero and (c) fails. Hence (c) and (d) are equivalent.
Combining the three steps, (a), (b), (c), (d) are all equivalent.

Problem 5D.16. Suppose that $T \in \mathcal{L}(V)$ is diagonalizable. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Prove that a subspace U of V is invariant under T if and only if there exist

subspaces $U_1, \dots, U_m$ of V such that $U_k \subseteq E(\lambda_k, T)$ for each k and $U = U_1 \oplus \cdots \oplus U_m$.

Solution. Because T is diagonalizable, T has a diagonal matrix with respect to some basis of V; in particular V is finite-dimensional. By 5.55(c),

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$

First suppose $U_1, \dots, U_m$ are subspaces of V with $U_k \subseteq E(\lambda_k, T)$ for each k and $U = U_1 \oplus \cdots \oplus U_m$. Let $u \in U$. Then $u = u_1 + \cdots + u_m$ for some $u_k \in U_k$. For each k we have $u_k \in E(\lambda_k, T)$, so $Tu_k = \lambda_k u_k$, and $\lambda_k u_k \in U_k$ because U_k is a subspace. Hence

$$Tu = \lambda_1 u_1 + \cdots + \lambda_m u_m \in U_1 + \cdots + U_m = U.$$

Thus U is invariant under T.
Conversely, suppose U is a subspace of V invariant under T. For each $k = 1, \dots, m$ define

$$U_k = U \cap E(\lambda_k, T),$$

which is a subspace of V (an intersection of two subspaces) satisfying $U_k \subseteq E(\lambda_k, T)$. We show $U = U_1 \oplus \cdots \oplus U_m$.
The sum $U_1 + \cdots + U_m$ is a direct sum: if $u_1 + \cdots + u_m = 0$ with $u_k \in U_k$, then $u_k \in E(\lambda_k, T)$ for each k, and because $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum (by 5.54), each u_k equals 0. By 1.45 this shows $U_1 + \cdots + U_m$ is a direct sum.
Also $U_1 + \cdots + U_m \subseteq U$, because each $U_k \subseteq U$ and U is a subspace.
For the reverse inclusion, note that U is invariant under T and T is diagonalizable, so $T|_U$ is a diagonalizable operator on U (by 5.65). Because U is a subspace of the finite-dimensional space V, it is finite-dimensional (by 2.25), so 5.55(b) applied to $T|_U \in \mathcal{L}(U)$ shows that U has a basis $u^{(1)}, \dots, u^{(r)}$ consisting of eigenvectors of $T|_U$. If $u^{(j)}$ is an eigenvector of $T|_U$ with eigenvalue μ, then $Tu^{(j)} = \mu u^{(j)}$ with $u^{(j)} \neq 0$, so μ is an eigenvalue of T; thus $\mu = \lambda_k$ for some k, and therefore

$$u^{(j)} \in U \cap E(\lambda_k, T) = U_k.$$

So every vector in the basis $u^{(1)}, \dots, u^{(r)}$ of U lies in $U_1 \cup \cdots \cup U_m$, and consequently

$$U = \operatorname{span}\big(u^{(1)}, \dots, u^{(r)}\big) \subseteq U_1 + \cdots + U_m.$$

(If $U = \{0\}$ the list of basis vectors is empty and the inclusion is trivial.)
Combining the two inclusions gives $U = U_1 + \cdots + U_m$, and we showed this sum is direct. Hence $U = U_1 \oplus \cdots \oplus U_m$ with $U_k \subseteq E(\lambda_k, T)$ for each k, as required.

Problem 5D.17. Suppose V is finite-dimensional. Prove that $\mathcal{L}(V)$ has a basis consisting of diagonalizable operators.

Solution. Let $n = \dim V$. If $n = 0$, then $\mathcal{L}(V) = \{0\}$ and the empty list is a basis of $\mathcal{L}(V)$; it vacuously consists of diagonalizable operators. So assume $n \geq 1$ and let $v_1, \dots, v_n$ be a basis of V.
For $j, k \in \{1, \dots, n\}$, the linear map lemma (3.4) gives a unique operator $E_{j,k} \in \mathcal{L}(V)$ with

$$E_{j,k} v_k = v_j \quad \text{and} \quad E_{j,k} v_l = 0 \ \text{ for } l \neq k.$$

The matrix of $E_{j,k}$ with respect to $v_1, \dots, v_n$ is the matrix with 1 in row j, column k, and 0 everywhere else. These n^2 matrices form a basis of $\mathbf{F}^{n,n}$, and $\mathcal{M}$ is an isomorphism from $\mathcal{L}(V)$ onto $\mathbf{F}^{n,n}$ (by 3.71); hence the list of the n^2 operators $E_{j,k}$ is a basis of $\mathcal{L}(V)$. In particular $\dim \mathcal{L}(V) = n^2$, as also asserted by 3.72.
Now consider the list of n^2 operators consisting of

$$E_{j,j} \quad (j = 1, \dots, n) \qquad \text{and} \qquad E_{j,j} + E_{j,k} \quad (j \neq k).$$

There are n operators of the first kind and $n(n-1)$ of the second kind, for a total of n^2.
This list spans $\mathcal{L}(V)$: for $j \neq k$ we have $E_{j,k} = (E_{j,j} + E_{j,k}) - E_{j,j}$, so the span of the list contains every $E_{j,k}$ with $j \neq k$, and it obviously contains every $E_{j,j}$. Hence the span contains all n^2 operators $E_{j,k}$, which form a basis of $\mathcal{L}(V)$; therefore the span is all of $\mathcal{L}(V)$. A spanning list of length $\dim \mathcal{L}(V) = n^2$ is a basis (by 2.42). Thus our list is a basis of $\mathcal{L}(V)$.
It remains to check that every operator in this list is diagonalizable.
The operator $E_{j,j}$ satisfies $E_{j,j}v_j = v_j$ and $E_{j,j}v_l = 0$ for $l \neq j$. Thus $v_1, \ldots, v_n$ is a basis of V consisting of eigenvectors of $E_{j,j}$ (with eigenvalue 1 for v_j and eigenvalue 0 for the others), so $E_{j,j}$ is diagonalizable by 5.55(b).
Now fix $j \neq k$ and set $S = E_{j,j} + E_{j,k}$. Then

$$Sv_j = v_j, \qquad Sv_k = v_j, \qquad Sv_l = 0 \ \text{ for } l \notin \{j, k\}.$$

Consider the list of n vectors obtained from $v_1, \ldots, v_n$ by replacing v_k with $v_k - v_j$ and leaving all other entries unchanged. This list spans V, since $v_k = (v_k - v_j) + v_j$ shows that its span contains $v_1, \ldots, v_n$; being a spanning list of length $n = \dim V$, it is a basis of V (by 2.42). Each of its vectors is an eigenvector of S:

$$Sv_j = 1 \cdot v_j, \qquad S(v_k - v_j) = v_j - v_j = 0 = 0 \cdot (v_k - v_j), \qquad Sv_l = 0 \cdot v_l \ \text{ for } l \notin \{j, k\}.$$

So V has a basis consisting of eigenvectors of S, and 5.55(b) shows that S is diagonalizable. (Concretely, S has the diagonal matrix with a single 1 in the position corresponding to v_j and 0 elsewhere, with respect to this new basis.)
Hence the list displayed above is a basis of $\mathcal{L}(V)$ consisting of diagonalizable operators.

Problem 5D.18. Suppose that $T \in \mathcal{L}(V)$ is diagonalizable and U is a subspace of V that is invariant under T. Prove that the quotient operator T/U is a diagonalizable operator on V/U.
The quotient operator T/U was defined in Exercise 38 in Section 5A.

Solution. Because T is diagonalizable, V is finite-dimensional; hence so is V/U, with $\dim V/U = \dim V - \dim U$ (by 3.105). Let $\pi \colon V \to V/U$ denote the quotient map, defined by

$$\pi v = v + U.$$

The map π is linear and surjective, and by the definition of the quotient operator (Exercise 38 in Section 5A),

$$(T/U)(\pi v) = (T/U)(v + U) = Tv + U = \pi(Tv) \qquad \text{for every } v \in V.$$

Because T is diagonalizable, 5.55(b) gives a basis

$$v_1, \ldots, v_n$$

of V consisting of eigenvectors of T. Thus for each i there is $\lambda^{(i)} \in \mathbf{F}$ with $Tv_i = \lambda^{(i)} v_i$.
Now apply π. Because π is surjective and $v_1, \ldots, v_n$ spans V, the list

$$\pi v_1, \ldots, \pi v_n$$

spans V/U: indeed, every element of V/U has the form πv with $v \in V$, and writing $v = a_1 v_1 + \cdots + a_n v_n$ gives $\pi v = a_1 \pi v_1 + \cdots + a_n \pi v_n$ by linearity of π.
Moreover each πv_i is either 0 or an eigenvector of T/U. Indeed, writing $\lambda = \lambda^{(i)}$,

$$(T/U)(\pi v_i) = \pi(Tv_i) = \pi(\lambda v_i) = \lambda\, \pi v_i.$$

Thus if $\pi v_i \neq 0$, then πv_i is an eigenvector of T/U corresponding to the eigenvalue λ.
Discard from the list $\pi v_1, \ldots, \pi v_n$ every entry that equals 0; the remaining list still spans V/U and consists of eigenvectors of T/U. By 2.30 (every spanning list contains a basis), some sublist of it is a basis of V/U, and every vector in that basis is one of the retained πv_i, hence an eigenvector of T/U.

Therefore V/U has a basis consisting of eigenvectors of T/U, and 5.55(b) shows that T/U is diagonalizable. (If $U = V$, then $V/U = \{0\}$, all the πv_i are 0, the empty list is a basis of V/U, and T/U is trivially diagonalizable.)

Problem 5D.19. Prove or give a counterexample: If $T \in \mathcal{L}(V)$ and there exists a subspace U of V that is invariant under T such that $T|_U$ and T/U are both diagonalizable, then T is diagonalizable. See Exercise 13 in Section 5C for an analogous statement about upper-triangular matrices.

Solution. The statement is false. Here is a counterexample.
Take $V = \mathbf{F}^2$ and define $T \in \mathcal{L}(\mathbf{F}^2)$ by

$$T(x, y) = (y, 0).$$

Let

$$U = \{(x, 0) : x \in \mathbf{F}\} = \operatorname{span}\big((1, 0)\big).$$

First, U is invariant under T: if $(x, 0) \in U$, then $T(x, 0) = (0, 0) \in U$.
Second, $T|_U$ is diagonalizable. Indeed $T|_U = 0$, the zero operator on the one-dimensional space U, and its matrix with respect to the basis $(1, 0)$ of U is the 1-by-1 diagonal matrix $\big(0\big)$.
Third, T/U is diagonalizable. We have $\dim V/U = \dim V - \dim U = 2 - 1 = 1$ (by 3.105). For every $(x, y) \in \mathbf{F}^2$,

$$(T/U)\big((x, y) + U\big) = T(x, y) + U = (y, 0) + U = 0 + U,$$

since $(y, 0) \in U$. Thus T/U is the zero operator on the one-dimensional space V/U, and its matrix with respect to any basis of V/U is the 1-by-1 diagonal matrix $\big(0\big)$. (More generally, every operator on a vector space of dimension at most 1 is diagonalizable, since it is a scalar multiple of the identity.)
Finally, T is not diagonalizable. Suppose $\lambda \in \mathbf{F}$ and $(x, y) \in \mathbf{F}^2$ satisfy $T(x, y) = \lambda(x, y)$, that is, $(y, 0) = (\lambda x, \lambda y)$. If $\lambda \neq 0$, then the second coordinate gives $\lambda y = 0$, so $y = 0$, and then the first coordinate gives $\lambda x = y = 0$, so $x = 0$. Hence the only eigenvalue of T is 0, and

$$E(0, T) = \operatorname{null} T = \{(x, y) \in \mathbf{F}^2 : y = 0\} = U,$$

which has dimension 1. Since $\dim E(0, T) = 1 \neq 2 = \dim \mathbf{F}^2$, condition (d) of 5.55 fails, so T is not diagonalizable.
(Equivalently: $T \neq 0$ and $T^2 = 0$, so the minimal polynomial of T is z^2, which is not of the form $(z - \lambda_1) \cdots (z - \lambda_m)$ with distinct $\lambda_1, \ldots, \lambda_m$; now 5.62 shows T is not diagonalizable.)
Thus $T|_U$ and T/U can both be diagonalizable while T is not, so the statement is false.
Note the contrast with Exercise 13 in Section 5C, where the analogous statement for upper-triangular matrices is true. The reason for the difference: upper-triangularity with respect to a basis $v_1, \ldots, v_n$ asks only that $Tv_k \in \operatorname{span}(v_1, \ldots, v_k)$ for each k, a condition stated in terms of subspaces, so a basis of U can be concatenated with lifts of a basis of V/U. Diagonalizability instead demands the exact equations $Tv_k = \lambda_k v_k$, and these do not survive lifting: in the example above, $(0, 1)+U$ is an eigenvector of T/U with eigenvalue 0, but no vector (x, y) with $y \neq 0$ satisfies $T(x, y) = 0$.

Problem 5D.20. Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T is diagonalizable if and only if the dual operator T' is diagonalizable.

Solution. Recall that $T' \in \mathcal{L}(V')$ is defined by $T'\varphi = \varphi \circ T$ for $\varphi \in V'$ (see 3.118). Because V is finite-dimensional, so is V', with $\dim V' = \dim V$ (by 3.111); hence T' has a minimal polynomial. (If $V = \{0\}$, then $V' = \{0\}$ as well, and both T and T' are diagonalizable because the empty list is a basis; so assume $V \neq \{0\}$, which forces $V' \neq \{0\}$ since $\dim V' = \dim V$.) We show that T and T' have the same minimal polynomial, after which 5.62 finishes the proof at once.
Step 1: $q(T') = \big(q(T)\big)'$ for every $q \in \mathcal{P}(\mathbf{F})$.
First note $I' = I$, where on the left I is the identity on V and on the right the identity on V': indeed

$I'\varphi = \varphi \circ I = \varphi$ for every $\varphi \in V'$.
Next, $(T^k)' = (T')^k$ for every nonnegative integer k. This holds for $k = 0$ by the previous paragraph. If it holds for $k-1$ with $k \geq 1$, then by 3.120(c),

$$(T^k)' = (T\,T^{k-1})' = (T^{k-1})'\,T' = (T')^{k-1}(T') = (T')^k,$$

completing the induction.
Now let $q(z) = a_0 + a_1 z + \cdots + a_d z^d$ with $a_0, \ldots, a_d \in \mathbf{F}$. Using 3.120(a) and 3.120(b) and the previous paragraph,

$$\big(q(T)\big)' = \big(a_0 I + a_1 T + \cdots + a_d T^d\big)' = a_0 I' + a_1 T' + \cdots + a_d (T^d)' = a_0 I + a_1 T' + \cdots + a_d (T')^d,$$

which is exactly $q(T')$.
Step 2: For $S \in \mathcal{L}(V)$, we have $S' = 0$ if and only if $S = 0$.
If $S = 0$ then clearly $S' = 0$. Conversely, suppose $S \neq 0$; choose $v \in V$ with $Sv \neq 0$. Extend the linearly independent list Sv to a basis $Sv, u_2, \ldots, u_n$ of V (by 2.32), and use the linear map lemma (3.4) to obtain $\varphi \in V'$ with $\varphi(Sv) = 1$ and $\varphi(u_j) = 0$ for $j = 2, \ldots, n$. Then

$$(S'\varphi)(v) = (\varphi \circ S)(v) = \varphi(Sv) = 1 \neq 0,$$

so $S'\varphi \neq 0$ and hence $S' \neq 0$.
Step 3: T and T' have the same minimal polynomial.
Let $q \in \mathcal{P}(\mathbf{F})$. By Steps 1 and 2,

$$q(T) = 0 \iff \big(q(T)\big)' = 0 \iff q(T') = 0.$$

Thus a monic polynomial annihilates T if and only if it annihilates T'. The minimal polynomial of an operator is the monic polynomial of smallest degree annihilating it (5.24), so T and T' have the same minimal polynomial; call it p.
Step 4: Conclusion.
By 5.62, T is diagonalizable if and only if p equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some list of distinct $\lambda_1, \ldots, \lambda_m \in \mathbf{F}$. By 5.62 applied to the operator T' on the finite-dimensional space V', the operator T' is diagonalizable if and only if its minimal polynomial, which is also p, has that same form. Hence T is diagonalizable if and only if T' is diagonalizable.

Problem 5D.21. The Fibonacci sequence $F_0, F_1, F_2, \ldots$ is defined by

$$F_0 = 0, \quad F_1 = 1, \quad \text{and} \quad F_n = F_{n-2} + F_{n-1} \ \text{ for } n \geq 2.$$

Define $T \in \mathcal{L}(\mathbf{R}^2)$ by $T(x, y) = (y, x + y)$.
(a) Show that $T^n(0, 1) = (F_n, F_{n+1})$ for each nonnegative integer n.
(b) Find the eigenvalues of T.
(c) Find a basis of $\mathbf{R}^2$ consisting of eigenvectors of T.
(d) Use the solution to (c) to compute $T^n(0, 1)$. Conclude that

$$F_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right]$$

for each nonnegative integer n.
(e) Use (d) to conclude that if n is a nonnegative integer, then the Fibonacci number F_n is the integer that is closest to

$$\frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^n.$$

Each F_n is a nonnegative integer, even though the right side of the formula in (d) does not look like an integer. The number

$$\frac{1+\sqrt{5}}{2}$$

is called the golden ratio.

Solution. Throughout, write

$$\varphi = \frac{1+\sqrt{5}}{2}, \qquad \psi = \frac{1-\sqrt{5}}{2}.$$

Note $\varphi + \psi = 1$, $\varphi\psi = \frac{1-5}{4} = -1$, and $\varphi - \psi = \sqrt{5}$.
(a) We induct on n. For $n = 0$ we have $T^0 = I$, so $T^0(0,1) = (0,1) = (F_0, F_1)$ because $F_0 = 0$ and $F_1 = 1$.
Now suppose $n \geq 0$ and $T^n(0,1) = (F_n, F_{n+1})$. Then

$$T^{n+1}(0,1) = T\big(T^n(0,1)\big) = T(F_n, F_{n+1}) = (F_{n+1},\, F_n + F_{n+1}) = (F_{n+1}, F_{n+2}),$$

where the last equality uses the recurrence $F_{n+2} = F_n + F_{n+1}$ (valid since $n + 2 \geq 2$). This completes the induction, so $T^n(0,1) = (F_n, F_{n+1})$ for every nonnegative integer n.
(b) Suppose $\lambda \in \mathbf{R}$ and $(x,y) \in \mathbf{R}^2$ is nonzero with $T(x,y) = \lambda(x,y)$, that is,

$$y = \lambda x \qquad \text{and} \qquad x + y = \lambda y.$$

Substituting the first equation into the second gives $x + \lambda x = \lambda^2 x$, so

$$(\lambda^2 - \lambda - 1)x = 0.$$

If $x = 0$, then $y = \lambda x = 0$, contradicting $(x,y) \neq (0,0)$. Hence $x \neq 0$ and $\lambda^2 - \lambda - 1 = 0$, so

$$\lambda = \frac{1 \pm \sqrt{5}}{2}.$$

Conversely, each of these two numbers is an eigenvalue: if $\lambda^2 = \lambda + 1$, then

$$T(1,\lambda) = (\lambda, 1+\lambda) = (\lambda, \lambda^2) = \lambda(1,\lambda),$$

and $(1,\lambda) \neq (0,0)$. Thus the eigenvalues of T are exactly $\varphi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$.
(c) By the computation in (b), $(1,\varphi)$ is an eigenvector corresponding to φ and $(1,\psi)$ is an eigenvector corresponding to ψ. Since $\varphi \neq \psi$, these eigenvectors correspond to distinct eigenvalues and hence are linearly independent (by 5.11). A linearly independent list of length $2 = \dim \mathbf{R}^2$ is a basis (by 2.38). Thus

$$(1,\varphi), \quad (1,\psi)$$

is a basis of $\mathbf{R}^2$ consisting of eigenvectors of T. (In particular T is diagonalizable, with diagonal matrix $\begin{pmatrix} \varphi & 0 \\ 0 & \psi \end{pmatrix}$ with respect to this basis.)
(d) Write $(0,1) = a(1,\varphi) + b(1,\psi)$ with $a, b \in \mathbf{R}$. Comparing coordinates gives

$$a + b = 0 \qquad \text{and} \qquad a\varphi + b\psi = 1.$$

From the first equation $b = -a$; substituting into the second gives $a(\varphi - \psi) = 1$, and since $\varphi - \psi = \sqrt{5}$ we get

$$a = \frac{1}{\sqrt{5}}, \qquad b = -\frac{1}{\sqrt{5}}.$$

Applying T^n and using $T^n(1,\varphi) = \varphi^n(1,\varphi)$ and $T^n(1,\psi) = \psi^n(1,\psi)$,

$$T^n(0,1) = \frac{1}{\sqrt{5}}\varphi^n(1,\varphi) - \frac{1}{\sqrt{5}}\psi^n(1,\psi) = \left(\frac{\varphi^n - \psi^n}{\sqrt{5}},\ \frac{\varphi^{n+1} - \psi^{n+1}}{\sqrt{5}}\right).$$

By (a), the first coordinate of $T^n(0,1)$ equals F_n. Hence

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

for each nonnegative integer n.
(e) Since $2 < \sqrt{5} < 3$, we have $-1 < \psi = \frac{1-\sqrt{5}}{2} < 0$, so $|\psi| < 1$ and therefore $|\psi|^n \le 1$ for every nonnegative integer n. By (d),

$$\left|F_n - \frac{1}{\sqrt{5}}\varphi^n\right| = \frac{|\psi|^n}{\sqrt{5}} \le \frac{1}{\sqrt{5}} < \frac{1}{2},$$

where the last inequality holds because $\sqrt{5} > 2$.
Finally, an integer within distance less than $\frac{1}{2}$ of a real number t is the unique integer closest to t: if $m \in \mathbf{Z}$ satisfies $|t-m| < \frac{1}{2}$ and $m' \in \mathbf{Z}$ with $m' \ne m$, then

$$|t-m'| \ge |m-m'| - |t-m| > 1 - \tfrac{1}{2} = \tfrac{1}{2} > |t-m|.$$

Applying this with $t = \frac{1}{\sqrt{5}}\varphi^n$ and $m = F_n$ shows that F_n is the integer closest to

$$\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n,$$

as desired.

Problem 5D.22. Suppose $T \in \mathcal{L}(V)$ and A is an n-by-n matrix that is the matrix of T with respect to some basis of V. Prove that if

$$|A_{j,j}| > \sum_{\substack{k=1\\k\ne j}}^{n} |A_{j,k}|$$

for each $j \in \{1,\dots,n\}$, then T is invertible.
This exercise states that if the diagonal entries of the matrix of T are large compared to the nondiagonal entries, then T is invertible.

Solution. Let $v_1,\dots,v_n$ be the basis of V with respect to which $A = \mathcal{M}(T)$; since this basis has length n, we have $\dim V = n < \infty$.
Because $\dim V < \infty$, the operator T is invertible if and only if it is injective: invertibility is equivalent to injectivity together with surjectivity (3.63), and for an operator on a finite-dimensional vector space injectivity is equivalent to surjectivity (3.65). Thus T fails to be invertible exactly when $\operatorname{null} T \ne \{0\}$, which happens exactly when 0 is an eigenvalue of T.
So suppose, seeking a contradiction, that T is not invertible. Then 0 is an eigenvalue of T. By the Gershgorin disk theorem (5.67), 0 is contained in some Gershgorin disk of T with respect to $v_1,\dots,v_n$: there exists $j \in \{1,\dots,n\}$ such that

$$|0 - A_{j,j}| \le \sum_{\substack{k=1\\k\ne j}}^{n} |A_{j,k}|,$$

which is to say

$$|A_{j,j}| \le \sum_{\substack{k=1\\k\ne j}}^{n} |A_{j,k}|.$$

This contradicts the hypothesis, which asserts strict inequality in the opposite direction for every j. Hence T is invertible.

For completeness, here is the same argument written out without invoking 5.67, which makes the solution self-contained. Suppose T is not invertible. By 3.63 and 3.65, T is not injective, so there exists $w \in V$ with $w \neq 0$ and $Tw = 0$. Write

$$w = c_1 v_1 + \cdots + c_n v_n$$

with $c_1, \ldots, c_n \in \mathbb{F}$ not all 0. Applying T and using the definition of the matrix A (so that $Tv_k = \sum_{j=1}^{n} A_{j,k} v_j$) gives

$$0 = Tw = \sum_{k=1}^{n} c_k T v_k = \sum_{k=1}^{n} c_k \sum_{j=1}^{n} A_{j,k} v_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} A_{j,k} c_k \right) v_j.$$

Because $v_1, \ldots, v_n$ is a basis, each coefficient vanishes:

$$\sum_{k=1}^{n} A_{j,k} c_k = 0 \quad \text{for each } j \in \{1, \ldots, n\}.$$

Choose j so that $|c_j| = \max\{|c_1|, \ldots, |c_n|\}$; then $|c_j| > 0$ because not all c_k are 0. Isolating the $k = j$ term in the equation above gives

$$A_{j,j} c_j = -\sum_{\substack{k=1 \\ k \neq j}}^{n} A_{j,k} c_k,$$

and hence, by the triangle inequality and the maximality of $|c_j|$,

$$|A_{j,j}|\,|c_j| \leq \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}|\,|c_k| \leq |c_j| \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}|.$$

Dividing both sides by $|c_j| > 0$ yields $|A_{j,j}| \leq \sum_{k \neq j} |A_{j,k}|$, again contradicting the hypothesis. Therefore T is invertible.

Problem 5D.23. Suppose the definition of the Gershgorin disks is changed so that the radius of the k^{th} disk is the sum of the absolute values of the entries in column (instead of row) k of A, excluding the diagonal entry. Show that the Gershgorin disk theorem (5.67) still holds with this changed definition.

Solution. Fix $T \in \mathcal{L}(V)$ and a basis $v_1, \ldots, v_n$ of V, and let A denote the matrix of T with respect to this basis. In the definition 5.66 the j^{th} disk has radius $\sum_{k \neq j} |A_{j,k}|$, the sum of the absolute values of the off-diagonal entries in row j. Under the changed definition, the k^{th} disk is instead

$$D_k = \{z \in \mathbb{F} : |z - A_{k,k}| \leq r_k\}, \qquad \text{where } r_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} |A_{j,k}|.$$

We must show that every eigenvalue of T lies in D_k for at least one $k \in \{1, \ldots, n\}$.
Let $\lambda \in \mathbb{F}$ be an eigenvalue of T and let $w \in V$ be a corresponding eigenvector, so $w \neq 0$ and $Tw = \lambda w$. There exist $c_1, \ldots, c_n \in \mathbb{F}$, not all 0, such that

$$w = c_1 v_1 + \cdots + c_n v_n.$$

Exactly as in the proof of 5.67, applying T to both sides and using $Tv_k = \sum_{j=1}^{n} A_{j,k} v_j$ gives

$$\lambda w = \sum_{k=1}^{n} c_k T v_k = \sum_{k=1}^{n} c_k \sum_{j=1}^{n} A_{j,k} v_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} A_{j,k} c_k \right) v_j.$$

Because $v_1, \dots, v_n$ is a basis, the representation of λw in this basis is unique; comparing the coefficient of v_j on the two sides gives

$$\lambda c_j = \sum_{k=1}^{n} A_{j,k} c_k \quad \text{for each } j \in \{1, \dots, n\}.$$

Moving the $k = j$ term to the left side, we obtain

$$(\lambda - A_{j,j}) c_j = \sum_{\substack{k=1 \\ k \neq j}}^{n} A_{j,k} c_k,$$

and therefore, by the triangle inequality,

$$|\lambda - A_{j,j}|\, |c_j| \leq \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}|\, |c_k| \quad \text{for each } j \in \{1, \dots, n\}.$$

Note that unlike in the proof of 5.67 we do not single out a largest $|c_j|$; instead we sum these n inequalities over j and interchange the order of summation on the right:

$$\begin{aligned} \sum_{j=1}^{n} |\lambda - A_{j,j}|\, |c_j| &\leq \sum_{j=1}^{n} \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}|\, |c_k| \\ &= \sum_{k=1}^{n} |c_k| \sum_{\substack{j=1 \\ j \neq k}}^{n} |A_{j,k}| \\ &= \sum_{k=1}^{n} |c_k|\, r_k. \end{aligned}$$

The interchange is legitimate because the double sum is a finite sum of nonnegative numbers over the same index set $\{(j,k) : j \neq k\}$.

Now suppose, seeking a contradiction, that λ belongs to none of the disks $D_1, \dots, D_n$. Then

$$|\lambda - A_{k,k}| > r_k \quad \text{for each } k \in \{1, \dots, n\}.$$

Multiplying by $|c_k| \geq 0$ gives $|c_k|\, r_k \leq |c_k|\, |\lambda - A_{k,k}|$ for every k, with strict inequality for every k such that $c_k \neq 0$. Since $w \neq 0$, at least one c_k is nonzero, so summing over k produces a strict inequality:

$$\sum_{k=1}^{n} |c_k|\, r_k < \sum_{k=1}^{n} |c_k|\, |\lambda - A_{k,k}|.$$

Combining this with the displayed estimate above yields

$$\sum_{k=1}^{n} |c_k|\, |\lambda - A_{k,k}| \leq \sum_{k=1}^{n} |c_k|\, r_k < \sum_{k=1}^{n} |c_k|\, |\lambda - A_{k,k}|,$$

so this (finite, nonnegative) number is strictly less than itself, which is impossible. Hence $\lambda \in D_k$ for some k, which is exactly the Gershgorin disk theorem for the column version of the disks.

Remark. There is a second route to the same conclusion, for readers who have covered Section 3F. The matrix of the dual operator T' with respect to the dual basis $\varphi_1, \dots, \varphi_n$ is the transpose A^{t} of A (by 3.132), so the diagonal entries are unchanged while the row sums that build the ordinary Gershgorin disks of T' are precisely the column sums used here for T. Moreover T and T' have the same eigenvalues: for $\lambda \in \mathbb{F}$ we have $(T - \lambda I)' = T' - \lambda I$ by 3.120, and by 3.129 and 3.131 the map $T - \lambda I$ is surjective if and only if $(T - \lambda I)'$ is injective and injective if and only if $(T - \lambda I)'$ is surjective, so by 3.63 one of these operators is invertible exactly when the other is. Applying 5.67 to T' therefore gives the column-radius statement for T. The self-contained argument above avoids needing these facts about duality.

5.5 Exercises 5E

Problem 5E.1. Give an example of two commuting operators S, T on $\mathbf{F}^4$ such that there is a subspace of $\mathbf{F}^4$ that is invariant under S but not under T and there is a subspace of $\mathbf{F}^4$ that is invariant under T but not under S.

Solution. Let e_1, e_2, e_3, e_4 be the standard basis of $\mathbf{F}^4$ and define $S, T \in \mathcal{L}(\mathbf{F}^4)$ by

$$S(x_1, x_2, x_3, x_4) = (x_2, 0, 0, 0), \qquad T(x_1, x_2, x_3, x_4) = (0, 0, x_4, 0).$$

Equivalently, $Se_2 = e_1$ and $Se_1 = Se_3 = Se_4 = 0$, while $Te_4 = e_3$ and $Te_1 = Te_2 = Te_3 = 0$.
These operators commute. For every $(x_1, x_2, x_3, x_4) \in \mathbf{F}^4$,

$$(ST)(x_1, x_2, x_3, x_4) = S(0, 0, x_4, 0) = (0, 0, 0, 0)$$

because the second coordinate of $(0, 0, x_4, 0)$ is 0, and

$$(TS)(x_1, x_2, x_3, x_4) = T(x_2, 0, 0, 0) = (0, 0, 0, 0)$$

because the fourth coordinate of $(x_2, 0, 0, 0)$ is 0. Thus $ST = 0 = TS$, so S and T commute.
A subspace invariant under S but not under T. Let $U = \operatorname{span}(e_1, e_4)$. Then $Se_1 = 0 \in U$ and $Se_4 = 0 \in U$, so $SU \subseteq U$ and hence U is invariant under S. However $Te_4 = e_3 \notin U$ (since e_3 is not a linear combination of e_1 and e_4), so U is not invariant under T.
A subspace invariant under T but not under S. Let $W = \operatorname{span}(e_2, e_3)$. Then $Te_2 = 0 \in W$ and $Te_3 = 0 \in W$, so W is invariant under T. However $Se_2 = e_1 \notin W$, so W is not invariant under S.
Thus S and T have all the required properties. (The example works over $\mathbf{F} = \mathbf{R}$ and over $\mathbf{F} = \mathbf{C}$ alike.)
With respect to the standard basis, the matrices of S and T are

$$\mathcal{M}(S) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathcal{M}(T) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

whose product in either order is the zero matrix, confirming via 5.74 that S and T commute.

Problem 5E.2. Suppose $\mathcal{E}$ is a subset of $\mathcal{L}(V)$ and every element of $\mathcal{E}$ is diagonalizable. Prove that there exists a basis of V with respect to which every element of $\mathcal{E}$ has a diagonal matrix if and only if every pair of elements of $\mathcal{E}$ commutes.
[This exercise extends 5.76, which considers the case in which $\mathcal{E}$ contains only two elements. For this exercise, $\mathcal{E}$ may contain any number of elements, and $\mathcal{E}$ may even be an infinite set.]

Solution. Throughout, V is finite-dimensional: for Axler a basis is a finite list, so as soon as $\mathcal{E}$ contains one diagonalizable operator (5.50), V has a basis and hence is finite-dimensional.
First direction. Suppose $v_1, \ldots, v_n$ is a basis of V with respect to which every element of $\mathcal{E}$ has a diagonal matrix. Let $S, T \in \mathcal{E}$. Then $\mathcal{M}(S, (v_1, \ldots, v_n))$ and $\mathcal{M}(T, (v_1, \ldots, v_n))$ are diagonal matrices of the same size. The product of two diagonal matrices of the same size is the diagonal matrix whose diagonal is obtained by multiplying corresponding diagonal entries; since multiplication in $\mathbf{F}$ is commutative, any two diagonal matrices of the same size commute. Hence $\mathcal{M}(S)$ and $\mathcal{M}(T)$ commute, and therefore S and T commute by 5.74. Thus every pair of elements of $\mathcal{E}$ commutes.
Second direction. Now suppose every pair of elements of $\mathcal{E}$ commutes. We prove, by strong induction on $\dim V$, the following statement: if V is a finite-dimensional vector space and $\mathcal{E} \subseteq \mathcal{L}(V)$ is a set of diagonalizable operators any two of which commute, then V has a basis consisting of vectors that are eigenvectors of every element of $\mathcal{E}$; with respect to such a basis every element of $\mathcal{E}$ has a diagonal matrix.

Note first that the last clause is automatic: if $v_1, \dots, v_n$ is a basis of V and each v_k is an eigenvector of T, say $Tv_k = \mu_k v_k$, then the k^{th} column of $\mathcal{M}(T, (v_1, \dots, v_n))$ is μ_k times the k^{th} standard column, so $\mathcal{M}(T)$ is diagonal. So it suffices to produce a basis of common eigenvectors.
Base case. If $\dim V \le 1$, then every operator on V is a scalar multiple of the identity, so every nonzero vector of V is an eigenvector of every element of $\mathcal{E}$, and any basis of V works. (If $\dim V = 0$ the empty list is the required basis and the assertion is vacuous.)
Induction step. Let $n = \dim V \ge 2$ and suppose the statement holds for all vector spaces of dimension less than n.
Case 1: every element of $\mathcal{E}$ is a scalar multiple of the identity operator on V. Then every nonzero vector of V is an eigenvector of every element of $\mathcal{E}$, so any basis of V consists of common eigenvectors, and we are done.
Case 2: some $S \in \mathcal{E}$ is not a scalar multiple of the identity. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of S. Because S is diagonalizable, 5.55(c) gives

$$V = E(\lambda_1, S) \oplus \cdots \oplus E(\lambda_m, S).$$

If we had $m = 1$, this would say $V = E(\lambda_1, S)$, i.e. $S = \lambda_1 I$, contradicting the assumption of this case; hence $m \ge 2$. Since the sum above is direct and each summand is nonzero, $\dim E(\lambda_k, S) < \dim V = n$ for each $k \in \{1, \dots, m\}$.
Fix $k \in \{1, \dots, m\}$ and write $U = E(\lambda_k, S)$. Every $T \in \mathcal{E}$ commutes with S, so U is invariant under T by 5.75. Thus the restriction $T|_U$ is an operator on U, and we may define

$$\mathcal{E}_k = \{T|_U : T \in \mathcal{E}\} \subseteq \mathcal{L}(U).$$

Each element of $\mathcal{E}_k$ is diagonalizable: T is diagonalizable and U is invariant under T, so $T|_U$ is diagonalizable by 5.65. Moreover any two elements of $\mathcal{E}_k$ commute: if $T, R \in \mathcal{E}$, then U is invariant under both, and for every $u \in U$,

$$(T|_U)(R|_U)u = T(Ru) = (TR)u = (RT)u = R(Tu) = (R|_U)(T|_U)u,$$

so $(T|_U)(R|_U) = (R|_U)(T|_U)$.
Since $\dim U < n$, the induction hypothesis applied to U and $\mathcal{E}_k$ produces a basis $u_1^{(k)}, \dots, u_{d_k}^{(k)}$ of $U = E(\lambda_k, S)$ consisting of vectors that are eigenvectors of $T|_U$ for every $T \in \mathcal{E}$. If $T \in \mathcal{E}$ and u is an eigenvector of $T|_U$, then $Tu = (T|_U)u$ is a scalar multiple of u and $u \ne 0$, so u is an eigenvector of T itself. Hence each $u_j^{(k)}$ is an eigenvector of every element of $\mathcal{E}$.
Finally, concatenate these lists over $k = 1, \dots, m$:

$$u_1^{(1)}, \dots, u_{d_1}^{(1)},\ u_1^{(2)}, \dots, u_{d_2}^{(2)},\ \dots,\ u_1^{(m)}, \dots, u_{d_m}^{(m)}.$$

Because $V = E(\lambda_1, S) \oplus \cdots \oplus E(\lambda_m, S)$ and each sublist is a basis of the corresponding summand, this concatenated list is a basis of V (a list obtained by concatenating bases of the summands of a direct sum decomposition spans the sum and has length $\sum_k \dim E(\lambda_k, S) = \dim V$, hence is a basis). Every vector in this basis is an eigenvector of every element of $\mathcal{E}$, so by the observation at the start of this direction, every element of $\mathcal{E}$ has a diagonal matrix with respect to it. This completes the induction and the proof.
Note that no assumption whatsoever was needed on the cardinality of $\mathcal{E}$: the induction is on $\dim V$, not on the number of operators, so $\mathcal{E}$ may be infinite. (If $\mathcal{E} = \varnothing$ both sides of the asserted equivalence hold vacuously, and any basis of V works.)

Problem 5E.3. Suppose $S, T \in \mathcal{L}(V)$ are such that $ST = TS$. Suppose $p \in \mathcal{P}(\mathbf{F})$.
(a) Prove that $\operatorname{null} p(S)$ is invariant under T.
(b) Prove that $\operatorname{range} p(S)$ is invariant under T.
[See 5.18 for the special case $S = T$.]

Solution. The whole exercise rests on the following lemma.
Lemma. If $S, T \in \mathcal{L}(V)$ satisfy $ST = TS$ and $p \in \mathcal{P}(\mathbf{F})$, then $p(S)T = Tp(S)$.
Proof of lemma. First we show by induction on k that $S^k T = TS^k$ for every nonnegative integer k. For $k = 0$ this reads $IT = TI$, which is true. Suppose $k \geq 0$ and $S^k T = TS^k$. Then

$$S^{k+1}T = S(S^kT) = S(TS^k) = (ST)S^k = (TS)S^k = TS^{k+1},$$

completing the induction.
Now write $p(z) = a_0 + a_1 z + \cdots + a_N z^N$ with $a_0, \ldots, a_N \in \mathbf{F}$, so that $p(S) = a_0 I + a_1 S + \cdots + a_N S^N$. Then

$$p(S)T = \sum_{k=0}^{N} a_k S^k T = \sum_{k=0}^{N} a_k T S^k = T \sum_{k=0}^{N} a_k S^k = Tp(S),$$

where the middle equality uses the previous paragraph and the outer equalities use the distributivity and homogeneity of composition of linear maps. This proves the lemma.
(a) Suppose $v \in \operatorname{null} p(S)$, so $p(S)v = 0$. Then, using the lemma,

$$p(S)(Tv) = \big(p(S)T\big)v = \big(Tp(S)\big)v = T\big(p(S)v\big) = T0 = 0.$$

Thus $Tv \in \operatorname{null} p(S)$. Since $v \in \operatorname{null} p(S)$ was arbitrary, $\operatorname{null} p(S)$ is invariant under T.
(b) Suppose $v \in \operatorname{range} p(S)$. Then $v = p(S)u$ for some $u \in V$, and using the lemma again,

$$Tv = T\big(p(S)u\big) = \big(Tp(S)\big)u = \big(p(S)T\big)u = p(S)(Tu) \in \operatorname{range} p(S).$$

Thus $\operatorname{range} p(S)$ is invariant under T.
Taking $S = T$ (which trivially satisfies $ST = TS$) recovers 5.18: $\operatorname{null} p(T)$ and $\operatorname{range} p(T)$ are invariant under T.

Problem 5E.4. Prove or give a counterexample: If A is a diagonal matrix and B is an upper-triangular matrix of the same size as A, then A and B commute.

Solution. The statement is *false*. Here is a counterexample of size 2-by-2:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then A is a diagonal matrix and B is upper triangular (all entries below the diagonal are 0), and they have the same size. However

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad BA = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}.$$

Since $AB \neq BA$, the matrices A and B do not commute.
It is worth recording what does go wrong and when. If A is the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$ and B is any n-by-n matrix, then

$$(AB)_{j,k} = \lambda_j B_{j,k} \qquad \text{and} \qquad (BA)_{j,k} = B_{j,k}\lambda_k.$$

Hence $AB = BA$ if and only if $(\lambda_j - \lambda_k)B_{j,k} = 0$ for all j, k, that is, if and only if $B_{j,k} = 0$ whenever $\lambda_j \neq \lambda_k$. So a diagonal matrix commutes with every matrix of its size exactly when all its diagonal entries are equal (i.e. it is a scalar multiple of the identity); as soon as A has two distinct diagonal entries $\lambda_j \neq \lambda_k$ with $j < k$, the upper-triangular matrix B with a 1 in position (j, k) and 0 elsewhere fails to commute with A. The counterexample above is the case $n = 2$, $j = 1$, $k = 2$.

Problem 5E.5. Prove that a pair of operators on a finite-dimensional vector space commute if and only if their dual operators commute.
[See 3.118 for the definition of the dual of an operator.]

Solution. Suppose V is a finite-dimensional vector space and $S, T \in \mathcal{L}(V)$. Recall (3.118) that the dual operator $T' \in \mathcal{L}(V')$ is defined by $T'(\varphi) = \varphi \circ T$ for $\varphi \in V'$, and that by 3.120(c) we have $(RT)' = T'R'$ for all $R, T \in \mathcal{L}(V)$; note the reversal of order.
First direction. Suppose S and T commute, i.e. $ST = TS$. Then by 3.120(c),

$$T'S' = (ST)' = (TS)' = S'T',$$

so S' and T' commute.
Second direction. Suppose S' and T' commute, i.e. $S'T' = T'S'$. By 3.120(c) again, $S'T' = (TS)'$ and $T'S' = (ST)'$. Hence

$$(TS)' = (ST)'.$$

Set $R = ST - TS \in \mathcal{L}(V)$. By 3.120(a) and 3.120(b) (with $\lambda = -1$), the map $R \mapsto R'$ is linear, so

$$R' = (ST)' - (TS)' = 0.$$

It remains to see that $R' = 0$ forces $R = 0$. By 3.128(a), $\operatorname{null} R' = (\operatorname{range} R)^0$. Since $R' = 0$ we have $\operatorname{null} R' = V'$, so $(\operatorname{range} R)^0 = V'$. By 3.127(b), the annihilator of a subspace U of the finite-dimensional space V equals V' if and only if $U = \{0\}$; hence $\operatorname{range} R = \{0\}$, that is, $R = 0$. Therefore $ST = TS$, so S and T commute.
Alternative proof of the second direction. Let $v_1, \ldots, v_n$ be a basis of V and let $\varphi_1, \ldots, \varphi_n$ be the dual basis of V'. By 3.132, the matrix of a dual operator is the transpose of the matrix of the operator:

$$\mathcal{M}(S', (\varphi_1, \ldots, \varphi_n)) = \big(\mathcal{M}(S, (v_1, \ldots, v_n))\big)^{\mathrm{t}},$$

and likewise for T. Write $A = \mathcal{M}(S)$ and $B = \mathcal{M}(T)$ with respect to $v_1, \ldots, v_n$. Using $(XY)^{\mathrm{t}} = Y^{\mathrm{t}}X^{\mathrm{t}}$ and 5.74 twice,

$$S'T' = T'S' \iff A^{\mathrm{t}}B^{\mathrm{t}} = B^{\mathrm{t}}A^{\mathrm{t}} \iff (BA)^{\mathrm{t}} = (AB)^{\mathrm{t}} \iff BA = AB \iff ST = TS,$$

where the third equivalence holds because transposition is injective (indeed $(X^{\mathrm{t}})^{\mathrm{t}} = X$). This chain of equivalences in fact proves both directions at once.

Problem 5E.6. Suppose that V is a nonzero finite-dimensional complex vector space and $S, T \in \mathcal{L}(V)$ commute. Prove that there exist $\alpha, \lambda \in \mathbf{C}$ such that

$$\operatorname{range}(S - \alpha I) + \operatorname{range}(T - \lambda I) \neq V.$$

Solution. Let $n = \dim V$; since $V \neq \{0\}$ we have $n \geq 1$.
Because V is a finite-dimensional complex vector space and S, T commute, 5.80 gives a basis $v_1, \ldots, v_n$ of V with respect to which both S and T have upper-triangular matrices. Write

$$A = \mathcal{M}(S, (v_1, \ldots, v_n)), \qquad B = \mathcal{M}(T, (v_1, \ldots, v_n)),$$

both upper triangular. Define

$$\alpha = A_{n,n}, \qquad \lambda = B_{n,n},$$

the bottom-right entries; these are complex numbers. (By 5.41 they are in fact eigenvalues of S and of T respectively, though we do not need this.)
Let $U = \operatorname{span}(v_1, \ldots, v_{n-1})$, which is a subspace of V with $\dim U = n - 1 < n$, so $U \neq V$. (When $n = 1$ this list is empty and $U = \{0\}$, which is still a proper subspace since $V \neq \{0\}$.)
Claim: $\operatorname{range}(S - \alpha I) \subseteq U$. It suffices to show $(S - \alpha I)v_k \in U$ for each $k \in \{1, \ldots, n\}$, because

$v_1, \dots, v_n$ spans V and hence the vectors $(S-\alpha I)v_1, \dots, (S-\alpha I)v_n$ span $\operatorname{range}(S-\alpha I)$ (a linear map sends a spanning list to a spanning list of its range).
For $k \in \{1, \dots, n-1\}$: since A is upper triangular, 5.39 gives $Sv_k \in \operatorname{span}(v_1, \dots, v_k) \subseteq U$; also $\alpha v_k \in U$. Hence $(S-\alpha I)v_k \in U$.
For $k = n$: upper-triangularity of A says

$$Sv_n = \sum_{j=1}^{n} A_{j,n} v_j = \Big(\sum_{j=1}^{n-1} A_{j,n} v_j\Big) + A_{n,n} v_n = \Big(\sum_{j=1}^{n-1} A_{j,n} v_j\Big) + \alpha v_n.$$

Therefore

$$(S-\alpha I)v_n = Sv_n - \alpha v_n = \sum_{j=1}^{n-1} A_{j,n} v_j \in U.$$

This proves the claim. (In matrix language: $A - \alpha I$ is upper triangular with bottom-right entry 0, so its entire last row is zero, and hence every column of $A - \alpha I$ lies in the span of the first $n-1$ standard basis vectors.)
The identical argument with B and λ in place of A and α gives

$$\operatorname{range}(T-\lambda I) \subseteq U.$$

Since U is a subspace, it is closed under addition, so

$$\operatorname{range}(S-\alpha I) + \operatorname{range}(T-\lambda I) \subseteq U \neq V.$$

In particular $\operatorname{range}(S-\alpha I) + \operatorname{range}(T-\lambda I) \neq V$, as desired.
Remark. There is a dual formulation of the same fact. Writing $W = \operatorname{range}(S-\alpha I) + \operatorname{range}(T-\lambda I)$, which is a subspace of V, 3.127(a) says $W \neq V$ if and only if $W^0 \neq \{0\}$, that is, if and only if some nonzero $\varphi \in V'$ annihilates both ranges. By 3.128(a), φ annihilates $\operatorname{range}(S-\alpha I)$ exactly when $\varphi \in \operatorname{null}(S-\alpha I)' = \operatorname{null}(S' - \alpha I)$, i.e. exactly when $S'\varphi = \alpha\varphi$; similarly for T. So the assertion of the exercise is equivalent to S' and T' having a common eigenvector. Since S and T commute, so do S' and T' (Exercise 5 in this section), and V' is a nonzero finite-dimensional complex vector space (its dimension equals $\dim V$ by 3.111), so 5.78 supplies such a common eigenvector. This gives a second proof.

Problem 5E.7. Suppose V is a complex vector space, $S \in \mathcal{L}(V)$ is diagonalizable, and $T \in \mathcal{L}(V)$ commutes with S. Prove that there is a basis of V such that S has a diagonal matrix with respect to this basis and T has an upper-triangular matrix with respect to this basis.

Solution. Note first that V is finite-dimensional: by 5.50, saying S is diagonalizable means S has a diagonal matrix with respect to some basis of V, and matrices of operators are defined with respect to (finite) bases.
Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of S. Since S is diagonalizable, 5.55(c) gives

$$V = E(\lambda_1, S) \oplus \cdots \oplus E(\lambda_m, S).$$

(If $V = \{0\}$ the empty basis works trivially, so assume $V \neq \{0\}$; then $m \geq 1$.)
Fix $k \in \{1, \dots, m\}$. Since S and T commute, 5.75 says that $E(\lambda_k, S)$ is invariant under T, so $T|_{E(\lambda_k,S)}$ is an operator on the finite-dimensional complex vector space $E(\lambda_k, S)$. By 5.47 there is a basis

$$u_1^{(k)}, \dots, u_{d_k}^{(k)}$$

of $E(\lambda_k, S)$, where $d_k = \dim E(\lambda_k, S)$, with respect to which $T|_{E(\lambda_k,S)}$ has an upper-triangular matrix. By 5.39(c), this means

$$Tu_j^{(k)} = \big(T|_{E(\lambda_k,S)}\big)u_j^{(k)} \in \operatorname{span}\big(u_1^{(k)}, \dots, u_j^{(k)}\big) \quad \text{for each } j \in \{1, \dots, d_k\}.$$

Now form the concatenated list

$$v_1, \dots, v_n \;=\; u_1^{(1)}, \dots, u_{d_1}^{(1)},\; u_1^{(2)}, \dots, u_{d_2}^{(2)},\; \dots,\; u_1^{(m)}, \dots, u_{d_m}^{(m)},$$

where $n = d_1 + \cdots + d_m$. Because V is the direct sum of the eigenspaces $E(\lambda_k, S)$ and each sublist is a basis of the corresponding eigenspace, this list spans V and has length $\sum_{k=1}^{m} \dim E(\lambda_k, S) = \dim V$; hence $v_1, \dots, v_n$ is a basis of V.

The matrix of S is diagonal. Each v_i lies in some $E(\lambda_k, S)$ and is nonzero, so $Sv_i = \lambda_k v_i$ is a scalar multiple of v_i. Hence the i^{th} column of $\mathcal{M}(S, (v_1, \dots, v_n))$ has its only possibly nonzero entry in row i, which is exactly the statement that $\mathcal{M}(S, (v_1, \dots, v_n))$ is diagonal. Explicitly, the diagonal of this matrix consists of λ_1 repeated d_1 times, then λ_2 repeated d_2 times, and so on.

The matrix of T is upper triangular. By 5.39 it suffices to check that $Tv_i \in \operatorname{span}(v_1, \dots, v_i)$ for each $i \in \{1, \dots, n\}$. Given i, write $v_i = u_j^{(k)}$, so that the vectors $u_1^{(k)}, \dots, u_j^{(k)}$ occupy positions $d_1 + \cdots + d_{k-1} + 1,\; \dots,\; d_1 + \cdots + d_{k-1} + j = i$ in the concatenated list; in particular

$$\operatorname{span}\big(u_1^{(k)}, \dots, u_j^{(k)}\big) \subseteq \operatorname{span}(v_1, \dots, v_i).$$

By the displayed containment from the choice of the block basis, $Tv_i = Tu_j^{(k)} \in \operatorname{span}(u_1^{(k)}, \dots, u_j^{(k)}) \subseteq \operatorname{span}(v_1, \dots, v_i)$, as required.

Hence with respect to the basis $v_1, \dots, v_n$ the operator S has a diagonal matrix and the operator T has an upper-triangular matrix. Indeed the matrix of T is block diagonal with m upper-triangular diagonal blocks of sizes $d_1, \dots, d_m$, which is in particular upper triangular.

Remark. The hypothesis that $\mathbf{F} = \mathbf{C}$ is used only through 5.47, to upper-triangularize T on each eigenspace of S; the rest of the argument is valid over any field. The conclusion can fail over $\mathbf{R}$: take $S = I$ on $\mathbf{R}^2$ (diagonalizable, and it commutes with everything) and $T(x, y) = (-y, x)$, which has no eigenvalue and therefore no upper-triangular matrix with respect to any basis (by 5.41 a diagonal entry of an upper-triangular matrix would be an eigenvalue).

Problem 5E.8. Suppose $m = 3$ in Example 5.72 and D_w, D_z are the commuting partial differentiation operators on $\mathcal{P}_3(\mathbf{C}^2, \mathbf{C})$ from that example. Find a basis of $\mathcal{P}_3(\mathbf{C}^2, \mathbf{C})$ with respect to which D_w and D_z each have an upper-triangular matrix.

Solution. Take the monomial basis of $\mathcal{P}_3(\mathbf{C}^2, \mathbf{C})$ listed in order of increasing total degree:

$$v_1 = 1; \quad v_2 = w,\; v_3 = z; \quad v_4 = w^2,\; v_5 = wz,\; v_6 = z^2; \quad v_7 = w^3,\; v_8 = w^2 z,\; v_9 = wz^2,\; v_{10} = z^3.$$

These ten functions are exactly the monomials $w^j z^k$ with $j + k \le 3$, and by the description 5.73 of $\mathcal{P}_3(\mathbf{C}^2, \mathbf{C})$ they span that space; they are linearly independent because a polynomial in two variables that is the zero function on $\mathbf{C}^2$ has all coefficients equal to 0. (Indeed, if $\sum_{j+k\le 3} a_{j,k} w^j z^k = 0$ for all $(w, z) \in \mathbf{C}^2$, then fixing z gives a polynomial in the single variable w vanishing on all of $\mathbf{C}$, so each coefficient $\sum_k a_{j,k} z^k$ vanishes for every $z \in \mathbf{C}$, and then each $a_{j,k} = 0$ by the same one-variable reasoning.) Hence $v_1, \dots, v_{10}$ is a basis of $\mathcal{P}_3(\mathbf{C}^2, \mathbf{C})$, and $\dim \mathcal{P}_3(\mathbf{C}^2, \mathbf{C}) = 10$.

The point of this ordering is that both partial differentiation operators strictly decrease total degree. Precisely, for $j + k \le 3$,

$$D_w(w^j z^k) = j\, w^{j-1} z^k, \qquad D_z(w^j z^k) = k\, w^j z^{k-1},$$

where a term with a negative exponent is interpreted as 0 (that is, $D_w(w^j z^k) = 0$ when $j = 0$, and $D_z(w^j z^k) = 0$ when $k = 0$). In every case the output is a scalar multiple of a monomial of total degree $j + k - 1$, or is 0. Since our list places all monomials of total degree $d - 1$ strictly before every monomial of total degree d, we conclude that for each $i \in \{1, \dots, 10\}$,

$$D_w v_i \in \operatorname{span}(v_1, \dots, v_{i-1}) \subseteq \operatorname{span}(v_1, \dots, v_i), \qquad D_z v_i \in \operatorname{span}(v_1, \dots, v_{i-1}) \subseteq \operatorname{span}(v_1, \dots, v_i).$$

By 5.39, the condition $Dv_i \in \operatorname{span}(v_1, \dots, v_i)$ for each i is equivalent to $\mathcal{M}(D, (v_1, \dots, v_{10}))$ being upper triangular. Thus D_w and D_z each have an upper-triangular matrix with respect to this basis.

Explicitly, the actions on the basis vectors are

$$D_w:\ v_1\mapsto 0,\ v_2\mapsto v_1,\ v_3\mapsto 0,\ v_4\mapsto 2v_2,\ v_5\mapsto v_3,\ v_6\mapsto 0,\ v_7\mapsto 3v_4,\ v_8\mapsto 2v_5,\ v_9\mapsto v_6,\ v_{10}\mapsto 0,$$
$$D_z:\ v_1\mapsto 0,\ v_2\mapsto 0,\ v_3\mapsto v_1,\ v_4\mapsto 0,\ v_5\mapsto v_2,\ v_6\mapsto 2v_3,\ v_7\mapsto 0,\ v_8\mapsto v_4,\ v_9\mapsto 2v_5,\ v_{10}\mapsto 3v_6,$$

so with respect to $v_1,\dots,v_{10}$ the two matrices are

$$\mathcal{M}(D_w)=\begin{pmatrix}0&1&0&0&0&0&0&0&0&0\\0&0&0&2&0&0&0&0&0&0\\0&0&0&0&1&0&0&0&0&0\\0&0&0&0&0&0&3&0&0&0\\0&0&0&0&0&0&0&2&0&0\\0&0&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\end{pmatrix},\qquad \mathcal{M}(D_z)=\begin{pmatrix}0&0&1&0&0&0&0&0&0&0\\0&0&0&0&1&0&0&0&0&0\\0&0&0&0&0&2&0&0&0&0\\0&0&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&0&2&0\\0&0&0&0&0&0&0&0&0&3\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\end{pmatrix}.$$

Both matrices are upper triangular, as required. In fact both have only zeros on the diagonal, which by 5.41 confirms the observation made in Example 5.79 that 0 is the only eigenvalue of D_w and the only eigenvalue of D_z.
Note that the existence of such a basis is guaranteed in advance by 5.80, since D_w and D_z commute (Example 5.72) and $\mathcal{P}_3(\mathbf{C}^2,\mathbf{C})$ is a finite-dimensional complex vector space; the work above simply exhibits one.

Problem 5E.9. Suppose V is a finite-dimensional nonzero complex vector space. Suppose that $\mathcal{E}\subseteq\mathcal{L}(V)$ is such that S and T commute for all $S,T\in\mathcal{E}$.

- (a) Prove that there is a vector in V that is an eigenvector for every element of $\mathcal{E}$.
- (b) Prove that there is a basis of V with respect to which every element of $\mathcal{E}$ has an upper-triangular matrix.

[This exercise extends 5.78 and 5.80, which consider the case in which $\mathcal{E}$ contains only two elements. For this exercise, $\mathcal{E}$ may contain any number of elements, and $\mathcal{E}$ may even be an infinite set.]

Solution. Throughout, call a subset $\mathcal{E}$ of $\mathcal{L}(V)$ a commuting family if $ST=TS$ for all $S,T\in\mathcal{E}$.
Part (a). We prove the following statement by strong induction on n: if V is a complex vector space with $\dim V=n\geq 1$ and $\mathcal{E}\subseteq\mathcal{L}(V)$ is a commuting family, then some nonzero $v\in V$ is an eigenvector of every element of $\mathcal{E}$.
Base case $n=1$. Choose any nonzero $v\in V$; then $V=\operatorname{span}(v)$. For each $T\in\mathcal{E}$ we have $Tv\in V=\operatorname{span}(v)$, so $Tv=\lambda v$ for some $\lambda\in\mathbf{C}$. As $v\neq 0$, this makes v an eigenvector of T. Thus v works.
Inductive step. Let $n>1$ and suppose the statement holds for every complex vector space of dimension in $\{1,\dots,n-1\}$ and every commuting family of operators on it. Suppose $\dim V=n$ and $\mathcal{E}\subseteq\mathcal{L}(V)$ is a commuting family. There are two cases.
Case 1: every element of $\mathcal{E}$ is a scalar multiple of the identity operator on V. Then any nonzero $v\in V$ is an eigenvector of every element of $\mathcal{E}$, and we are done. (This case also covers $\mathcal{E}=\emptyset$, for which the assertion is vacuous.)
Case 2: some $S\in\mathcal{E}$ is not a scalar multiple of the identity. Because V is a nonzero finite-dimensional complex vector space, S has an eigenvalue $\lambda\in\mathbf{C}$ (by 5.19). Let

$$U=E(\lambda,S)=\operatorname{null}(S-\lambda I).$$

Since λ is an eigenvalue of S, we have $U\neq\{0\}$, so $\dim U\geq 1$. Also $U\neq V$: if $U=V$ then $Sv=\lambda v$ for every $v\in V$, meaning $S=\lambda I$, contrary to the choice of S. Hence $1\leq\dim U\leq n-1$.
Now let $T\in\mathcal{E}$. Then S and T commute, so $U=E(\lambda,S)$ is invariant under T by 5.75. Thus every element of $\mathcal{E}$ maps U into U, and we may form

$$\mathcal{E}|_U=\{T|_U:T\in\mathcal{E}\}\subseteq\mathcal{L}(U).$$

This is a commuting family: if $T_1, T_2 \in \mathcal{E}$ and $u \in U$, then $T_2 u \in U$ and $T_1 u \in U$, so

$$(T_1|_U)(T_2|_U)u = T_1 T_2 u = T_2 T_1 u = (T_2|_U)(T_1|_U)u.$$

Since U is a complex vector space with $1 \le \dim U \le n-1$, the induction hypothesis applies to U and $\mathcal{E}|_U$: there is a nonzero $v \in U$ that is an eigenvector of $T|_U$ for every $T \in \mathcal{E}$. Because $T|_U v = Tv$ for $v \in U$, this same v is an eigenvector of every $T \in \mathcal{E}$ (with $Tv = \alpha_T v$ for suitable $\alpha_T \in \mathbf{C}$ depending on T). This completes the induction and proves (a).

Note that the eigenvalues α_T generally depend on T; the assertion is that one single vector serves as an eigenvector for all of $\mathcal{E}$ simultaneously.

Part (b). Recall that if U is a subspace of V invariant under $T \in \mathcal{L}(V)$, then the quotient operator $T/U \in \mathcal{L}(V/U)$ is defined by $(T/U)(v+U) = Tv + U$ (see Exercise 38 in Section 5A). This is well defined: if $v + U = v' + U$, then $v - v' \in U$, so $Tv - Tv' = T(v - v') \in U$ by invariance, whence $Tv + U = Tv' + U$; linearity of T/U is immediate from linearity of T and the definitions of the vector space operations on V/U.

We prove by induction on $n = \dim V \ge 1$: if V is a complex vector space of dimension n and $\mathcal{E} \subseteq \mathcal{L}(V)$ is a commuting family, then there is a basis of V with respect to which every element of $\mathcal{E}$ has an upper-triangular matrix.

Base case $n = 1$. Every 1-by-1 matrix is upper triangular, so any basis of V works.

Inductive step. Let $n > 1$ and suppose the result holds for all complex vector spaces of dimension $n-1$. Let $\dim V = n$ and let $\mathcal{E} \subseteq \mathcal{L}(V)$ be a commuting family. By part (a) there is a nonzero $v_1 \in V$ that is an eigenvector of every element of $\mathcal{E}$. Put

$$U = \operatorname{span}(v_1),$$

so $\dim U = 1$ and $Tv_1 \in U$ for every $T \in \mathcal{E}$; hence U is invariant under every element of $\mathcal{E}$.

By 3.105, $\dim V/U = n - 1 \ge 1$, so V/U is a nonzero complex vector space. Let

$$\mathcal{E}/U = \{T/U : T \in \mathcal{E}\} \subseteq \mathcal{L}(V/U).$$

This is a commuting family: for $T_1, T_2 \in \mathcal{E}$ and $v \in V$,

$$(T_1/U)(T_2/U)(v+U) = (T_1/U)(T_2 v + U) = T_1 T_2 v + U = T_2 T_1 v + U = (T_2/U)(T_1/U)(v+U).$$

By the induction hypothesis there is a basis of V/U, which we may write as

$$u_2 + U,\ u_3 + U,\ \ldots,\ u_n + U$$

for some $u_2, \ldots, u_n \in V$, with respect to which every element of $\mathcal{E}/U$ has an upper-triangular matrix. We claim $v_1, u_2, \ldots, u_n$ is a basis of V. This list has length $n = \dim V$, so by 2.38 it suffices to prove linear independence. Suppose $a_1, \ldots, a_n \in \mathbf{C}$ and $a_1 v_1 + a_2 u_2 + \cdots + a_n u_n = 0$. Applying the quotient map $\pi : V \to V/U$ (which is linear, and satisfies $\pi(v_1) = 0 + U$ because $v_1 \in U$) gives

$$a_2(u_2 + U) + \cdots + a_n(u_n + U) = 0 + U.$$

Linear independence of $u_2 + U, \ldots, u_n + U$ in V/U forces $a_2 = \cdots = a_n = 0$. The original equation then reads $a_1 v_1 = 0$, and $v_1 \neq 0$ gives $a_1 = 0$. Hence $v_1, u_2, \ldots, u_n$ is a basis of V.

Finally we check upper-triangularity with respect to this basis. Let $T \in \mathcal{E}$. First, $Tv_1 \in \operatorname{span}(v_1)$ because v_1 is an eigenvector of T. Next, fix $k \in \{2, \ldots, n\}$. By 5.39 applied to the operator T/U and the basis $u_2 + U, \ldots, u_n + U$ of V/U, upper-triangularity of $\mathcal{M}(T/U)$ gives

$$(T/U)(u_k + U) \in \operatorname{span}(u_2 + U, \ldots, u_k + U),$$

that is, there exist $a_2, \ldots, a_k \in \mathbf{C}$ with

$$Tu_k + U = \sum_{j=2}^{k} a_j(u_j + U) = \Big(\sum_{j=2}^{k} a_j u_j\Big) + U.$$

Hence $Tu_k - \sum_{j=2}^{k} a_j u_j \in U = \operatorname{span}(v_1)$, so

$$Tu_k \in \operatorname{span}(v_1, u_2, \ldots, u_k).$$

Thus for every $T \in \mathcal{E}$, each vector of the basis $v_1, u_2, \ldots, u_n$ is mapped by T into the span of itself and the preceding basis vectors. By 5.39, every $T \in \mathcal{E}$ has an upper-triangular matrix with respect to $v_1, u_2, \ldots, u_n$. This completes the induction and proves (b).
Taking $\mathcal{E} = \{S, T\}$ recovers 5.78 in part (a) and 5.80 in part (b). Note that nothing in the argument required $\mathcal{E}$ to be finite: the single vector produced in (a), and the single basis produced in (b), work simultaneously for all the (possibly infinitely many) operators in $\mathcal{E}$.

Problem 5E.10. Give an example of two commuting operators S, T on a finite-dimensional real vector space such that $S + T$ has an eigenvalue that does not equal an eigenvalue of S plus an eigenvalue of T and ST has an eigenvalue that does not equal an eigenvalue of S times an eigenvalue of T.
[This exercise shows that 5.81 does not hold on real vector spaces.]

Solution. Take $V = \mathbf{R}^2$ and define $S, T \in \mathcal{L}(\mathbf{R}^2)$ by

$$S(x, y) = (-y, x), \qquad T(x, y) = (y, -x).$$

Thus $T = -S$ (geometrically, S is counterclockwise rotation by $90°$ and T is clockwise rotation by $90°$).
First, S and T commute: since $T = -S$, we have $ST = S(-S) = -S^2 = (-S)S = TS$.
Next, neither S nor T has any eigenvalue. Suppose $\lambda \in \mathbf{R}$ and $(x, y) \in \mathbf{R}^2$ satisfy $S(x, y) = \lambda(x, y)$. Then

$$-y = \lambda x \quad \text{and} \quad x = \lambda y.$$

Substituting the second equation into the first gives $-y = \lambda^2 y$, so $(\lambda^2 + 1)y = 0$. Because λ is real, $\lambda^2 + 1 > 0$, forcing $y = 0$; then $x = \lambda y = 0$. Hence the only vector satisfying $S(x, y) = \lambda(x, y)$ is $(0, 0)$, so S has no eigenvector and therefore no eigenvalue. Since $T = -S$, the equation $T(x, y) = \lambda(x, y)$ is the same as $S(x, y) = (-\lambda)(x, y)$, so likewise T has no eigenvalue.
Now compute $S + T$ and ST. Because $T = -S$ we get $S + T = 0$, the zero operator on $\mathbf{R}^2$. Also, for all $(x, y) \in \mathbf{R}^2$,

$$(ST)(x, y) = S(y, -x) = (x, y),$$

so $ST = I$, the identity operator on $\mathbf{R}^2$.
We can now conclude. The operator $S + T = 0$ has eigenvalue 0 (every nonzero vector of $\mathbf{R}^2$ is an eigenvector with eigenvalue 0). But S has no eigenvalues and T has no eigenvalues, so there is no number at all of the form (an eigenvalue of S) plus (an eigenvalue of T); in particular 0 is not such a number. Similarly, $ST = I$ has eigenvalue 1, but there is no number of the form (an eigenvalue of S) times (an eigenvalue of T), so 1 is not such a number. Hence S and T are as required, and 5.81 fails on this real vector space.
A sharper example. One might object that in the example above the conclusion holds only because S and T have no eigenvalues whatsoever. The failure of 5.81 over $\mathbf{R}$ is genuine even when both operators do have eigenvalues. Let $V = \mathbf{R}^3$ and define

$$S(x, y, z) = (-y,\ x,\ z), \qquad T(x, y, z) = (y,\ -x,\ 2z).$$

These commute: for all (x, y, z),

$$(ST)(x, y, z) = S(y, -x, 2z) = (x,\ y,\ 2z) = T(-y, x, z) = (TS)(x, y, z).$$

The eigenvalues of S are found from $-y = \lambda x$, $x = \lambda y$, $z = \lambda z$. As before the first two equations force $x = y = 0$, and then an eigenvector must have $z \neq 0$, giving $\lambda = 1$. So the only eigenvalue of S is 1. Likewise, $T(x, y, z) = \lambda(x, y, z)$ means $y = \lambda x$, $-x = \lambda y$, $2z = \lambda z$; the first two give $(1 + \lambda^2)y = 0$, hence $x = y = 0$, and then $z \neq 0$ forces $\lambda = 2$. So the only eigenvalue of T is 2.

Now $(S+T)(x,y,z) = (0,0,3z)$, whose eigenvalues are 0 (with eigenspace $\{(x,y,0) : x,y \in \mathbf{R}\}$) and 3 (with eigenspace $\{(0,0,z) : z \in \mathbf{R}\}$). The only number that is an eigenvalue of S plus an eigenvalue of T is $1+2=3$, so the eigenvalue 0 of $S+T$ is not of that form.
Also $(ST)(x,y,z) = (x,y,2z)$, whose eigenvalues are 1 and 2. The only number that is an eigenvalue of S times an eigenvalue of T is $1 \cdot 2 = 2$, so the eigenvalue 1 of ST is not of that form.
The proof of 5.81 breaks down over $\mathbf{R}$ precisely at 5.80, which relies on 5.78 and hence on 5.19 (every operator on a nonzero finite-dimensional complex vector space has an eigenvalue). Indeed, in neither example does S have an upper-triangular matrix with respect to any basis of the real vector space in question, so the diagonal bookkeeping used in the proof of 5.81 is unavailable. For the first example this is clear, since S has no eigenvalue at all while by 5.41 the diagonal entries of an upper-triangular matrix of S would be eigenvalues of S. For the second example, suppose S had an upper-triangular matrix with respect to some basis of $\mathbf{R}^3$. By 5.41 every diagonal entry of that matrix is an eigenvalue of S, and 1 is the only eigenvalue of S; hence $S-I$ would have a strictly upper-triangular matrix with respect to that basis, which forces $(S-I)^3 = 0$. But the plane $P = \{(x,y,0) : x,y \in \mathbf{R}\}$ is invariant under both S and $S-I$, and $(S-I)(x,y,0) = (-x-y,\, x-y,\, 0)$, which vanishes only when $x = y = 0$; thus $(S-I)|_P$ is injective, so $(S-I)^3$ is nonzero on P, a contradiction.

6 Inner Product Spaces

Contents

6.1 Exercises 6A 288
6.2 Exercises 6B 318
6.3 Exercises 6C 342

6.1 Exercises 6A

Problem 6A.1. Prove or give a counterexample: If $v_1, \dots, v_m \in V$, then

$$\sum_{j=1}^{m} \sum_{k=1}^{m} \langle v_j, v_k \rangle \geq 0.$$

Solution. The statement is true.
Set $v = v_1 + \cdots + v_m$. Additivity in the first slot (part of the definition 6.2 of an inner product), applied $m - 1$ times, gives

$$\sum_{j=1}^{m} \langle v_j, w \rangle = \langle v_1 + \cdots + v_m, w \rangle = \langle v, w \rangle$$

for every $w \in V$. Similarly, additivity in the second slot, which is 6.6(d), applied $m - 1$ times gives

$$\sum_{k=1}^{m} \langle u, v_k \rangle = \langle u, v_1 + \cdots + v_m \rangle = \langle u, v \rangle$$

for every $u \in V$.
Now compute the double sum, summing first over k and then over j:

$$\begin{aligned} \sum_{j=1}^{m} \sum_{k=1}^{m} \langle v_j, v_k \rangle &= \sum_{j=1}^{m} \left(\sum_{k=1}^{m} \langle v_j, v_k \rangle \right) \\ &= \sum_{j=1}^{m} \langle v_j, v \rangle \\ &= \langle v, v \rangle \\ &= \|v\|^2, \end{aligned}$$

where the last equality is the definition 6.7 of the norm.
By the positivity property in 6.2 we have $\langle v, v \rangle \geq 0$; recall that for a complex scalar the notation $\lambda \geq 0$ means that λ is real and nonnegative, so in particular the double sum is a nonnegative real number even when $\mathbf{F} = \mathbf{C}$. Hence

$$\sum_{j=1}^{m} \sum_{k=1}^{m} \langle v_j, v_k \rangle = \|v_1 + \cdots + v_m\|^2 \geq 0,$$

as claimed.

Problem 6A.2. Suppose $S \in \mathcal{L}(V)$. Define $\langle \cdot, \cdot \rangle_1$ by

$$\langle u, v \rangle_1 = \langle Su, Sv \rangle$$

for all $u, v \in V$. Show that $\langle \cdot, \cdot \rangle_1$ is an inner product on V if and only if S is injective.

Solution. Here V is an inner product space (see 6.5) with inner product $\langle \cdot, \cdot \rangle$, and $\langle \cdot, \cdot \rangle_1$ is the function on $V \times V$ defined above. We first show that $\langle \cdot, \cdot \rangle_1$ satisfies four of the five properties in the definition 6.2 of an inner product no matter what S is, and that the fifth property, definiteness, is exactly the assertion that S is injective.
Positivity. If $v \in V$, then $\langle v, v \rangle_1 = \langle Sv, Sv \rangle \geq 0$ by the positivity of $\langle \cdot, \cdot \rangle$.
Additivity in the first slot. If $u, v, w \in V$, then the linearity of S and the additivity of $\langle \cdot, \cdot \rangle$ in its first slot give

$$\langle u + v, w \rangle_1 = \langle S(u + v), Sw \rangle = \langle Su + Sv, Sw \rangle = \langle Su, Sw \rangle + \langle Sv, Sw \rangle = \langle u, w \rangle_1 + \langle v, w \rangle_1.$$

Homogeneity in the first slot. If $\lambda \in \mathbf{F}$ and $u, v \in V$, then

$$\langle \lambda u, v\rangle_1 = \langle S(\lambda u), Sv\rangle = \langle \lambda Su, Sv\rangle = \lambda\langle Su, Sv\rangle = \lambda\langle u, v\rangle_1.$$

Conjugate symmetry. If $u, v \in V$, then

$$\langle u, v\rangle_1 = \langle Su, Sv\rangle = \overline{\langle Sv, Su\rangle} = \overline{\langle v, u\rangle_1}.$$

Thus $\langle \cdot, \cdot\rangle_1$ is an inner product on V if and only if it also satisfies definiteness, that is, if and only if

$$\langle v, v\rangle_1 = 0 \implies v = 0$$

for all $v \in V$. (The reverse implication in the definiteness condition always holds: $\langle 0, 0\rangle_1 = \langle S0, S0\rangle = \langle 0, 0\rangle = 0$.) Now observe that for $v \in V$,

$$\langle v, v\rangle_1 = 0 \iff \langle Sv, Sv\rangle = 0 \iff Sv = 0,$$

where the second equivalence is the definiteness of the inner product $\langle \cdot, \cdot\rangle$ on V. In other words, the set of $v \in V$ with $\langle v, v\rangle_1 = 0$ is exactly $\operatorname{null} S$.
Suppose now that S is injective. Then $\operatorname{null} S = \{0\}$ by 3.15. Hence if $\langle v, v\rangle_1 = 0$, the equivalence above gives $Sv = 0$ and thus $v = 0$. So definiteness holds and $\langle \cdot, \cdot\rangle_1$ is an inner product on V.
Conversely, suppose $\langle \cdot, \cdot\rangle_1$ is an inner product on V. If $v \in \operatorname{null} S$, then $\langle v, v\rangle_1 = \langle Sv, Sv\rangle = \langle 0, 0\rangle = 0$, and the definiteness of $\langle \cdot, \cdot\rangle_1$ then forces $v = 0$. Thus $\operatorname{null} S = \{0\}$, and 3.15 shows that S is injective.

Problem 6A.3. (a) Show that the function taking an ordered pair $((x_1, x_2), (y_1, y_2))$ of elements of $\mathbf{R}^2$ to $|x_1y_1| + |x_2y_2|$ is not an inner product on $\mathbf{R}^2$.
(b) Show that the function taking an ordered pair $((x_1, x_2, x_3), (y_1, y_2, y_3))$ of elements of $\mathbf{R}^3$ to $x_1y_1 + x_3y_3$ is not an inner product on $\mathbf{R}^3$.

Solution. (a) Write φ for the function in question, so

$$\varphi\big((x_1, x_2), (y_1, y_2)\big) = |x_1y_1| + |x_2y_2|$$

for all $(x_1, x_2), (y_1, y_2) \in \mathbf{R}^2$. Note that positivity and definiteness do hold for φ, because $\varphi(v, v) = x_1^2 + x_2^2$ when $v = (x_1, x_2)$; the failure is a failure of linearity in the first slot.
Take $\lambda = -1$ and $u = v = (1, 0)$. Then

$$\varphi(\lambda u, v) = \varphi\big((-1, 0), (1, 0)\big) = |(-1)(1)| + |0 \cdot 0| = 1,$$

whereas

$$\lambda\varphi(u, v) = -\varphi\big((1, 0), (1, 0)\big) = -\big(|1 \cdot 1| + |0 \cdot 0|\big) = -1.$$

Since $1 \neq -1$, the function φ does not satisfy homogeneity in the first slot, so φ is not an inner product on $\mathbf{R}^2$.
(One can also see the failure through additivity in the first slot: with $u = (1, 0)$, $v = (-1, 0)$, and $w = (1, 0)$ we get $\varphi(u + v, w) = \varphi((0, 0), (1, 0)) = 0$ but $\varphi(u, w) + \varphi(v, w) = 1 + 1 = 2$.)
(b) Write ψ for the function in question, so

$$\psi\big((x_1, x_2, x_3), (y_1, y_2, y_3)\big) = x_1y_1 + x_3y_3$$

for all $(x_1, x_2, x_3), (y_1, y_2, y_3) \in \mathbf{R}^3$. This function is symmetric, is linear in its first slot, and satisfies positivity, since $\psi(v, v) = x_1^2 + x_3^2 \geq 0$ when $v = (x_1, x_2, x_3)$. However, definiteness fails.
Indeed, let $v = (0, 1, 0)$. Then $v \neq 0$, but

$$\psi(v, v) = 0 \cdot 0 + 0 \cdot 0 = 0.$$

The definiteness property in 6.2 requires that $\psi(v, v) = 0$ only for $v = 0$. Hence ψ is not an inner product on $\mathbf{R}^3$.

(Alternatively, let $S \in \mathcal{L}(\mathbf{R}^3)$ be defined by $S(x_1, x_2, x_3) = (x_1, 0, x_3)$. Then $\psi(u, v) = \langle Su, Sv\rangle$ for all $u, v \in \mathbf{R}^3$, where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. Because S is not injective, Exercise 6A.2 shows that ψ is not an inner product on $\mathbf{R}^3$.)

Problem 6A.4. Suppose $T \in \mathcal{L}(V)$ is such that $\|Tv\| \le \|v\|$ for every $v \in V$. Prove that $T - \sqrt{2}\, I$ is injective.

Solution. Note first that $\sqrt{2}$ is a real number and hence lies in $\mathbf{F}$ whether $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$, so $T - \sqrt{2}\, I \in \mathcal{L}(V)$ makes sense.
By 3.15, it suffices to show that $\operatorname{null}(T - \sqrt{2}\, I) = \{0\}$.
So suppose $v \in V$ and $(T - \sqrt{2}\, I)v = 0$, which means

$$Tv = \sqrt{2}\, v.$$

Taking norms and using 6.9(b) (with $\lambda = \sqrt{2}$, so $|\lambda| = \sqrt{2}$) gives

$$\|Tv\| = \|\sqrt{2}\, v\| = \sqrt{2}\, \|v\|.$$

On the other hand, the hypothesis gives $\|Tv\| \le \|v\|$. Combining these,

$$\sqrt{2}\, \|v\| \le \|v\|,$$

and therefore

$$(\sqrt{2} - 1)\|v\| \le 0.$$

Since $\sqrt{2} - 1 > 0$, this forces $\|v\| \le 0$. But norms are nonnegative, so $\|v\| = 0$, and hence $v = 0$ by 6.9(a).
Thus $\operatorname{null}(T - \sqrt{2}\, I) = \{0\}$, and so $T - \sqrt{2}\, I$ is injective by 3.15.
(The same argument shows that $T - \lambda I$ is injective for every scalar $\lambda \in \mathbf{F}$ with $|\lambda| > 1$; in other words, the hypothesis $\|Tv\| \le \|v\|$ for all v forces every eigenvalue of T to have absolute value at most 1.)

Problem 6A.5. Suppose V is a real inner product space.
(a) Show that $\langle u + v, u - v\rangle = \|u\|^2 - \|v\|^2$ for every $u, v \in V$.
(b) Show that if $u, v \in V$ have the same norm, then $u + v$ is orthogonal to $u - v$.
(c) Use (b) to show that the diagonals of a rhombus are perpendicular to each other.

Solution. Throughout, V is a real inner product space, so $\mathbf{F} = \mathbf{R}$ and every scalar equals its own complex conjugate. In particular the conjugate symmetry property in 6.2 becomes plain symmetry:

$$\langle u, v\rangle = \langle v, u\rangle \quad \text{for all } u, v \in V,$$

and 6.6(e) becomes $\langle u, \lambda v\rangle = \lambda\langle u, v\rangle$ for all $\lambda \in \mathbf{R}$.
(a) Let $u, v \in V$. Using additivity in the first slot (6.2), then additivity in the second slot (6.6(d)) together with $\langle x, -v\rangle = -\langle x, v\rangle$ (which is 6.6(e) with $\lambda = -1$), we get

$$\begin{aligned}\langle u + v, u - v\rangle &= \langle u, u - v\rangle + \langle v, u - v\rangle \\ &= \langle u, u\rangle - \langle u, v\rangle + \langle v, u\rangle - \langle v, v\rangle.\end{aligned}$$

Because V is a real inner product space, $\langle u, v\rangle = \langle v, u\rangle$, so the two middle terms cancel. Hence, using the definition 6.7 of the norm,

$$\langle u + v, u - v\rangle = \langle u, u\rangle - \langle v, v\rangle = \|u\|^2 - \|v\|^2,$$

as desired.
(b) Suppose $u, v \in V$ and $\|u\| = \|v\|$. Then by (a),

$$\langle u + v, u - v\rangle = \|u\|^2 - \|v\|^2 = 0.$$

By the definition 6.10 of orthogonality, $u+v$ is orthogonal to $u-v$.
(c) A rhombus is a parallelogram all four of whose sides have equal length. Place it in $\mathbf{R}^2$ (with the Euclidean inner product) with one vertex at the origin, and let u and v be the two vectors along the sides emanating from that vertex. Since the figure is a genuine parallelogram, u and v are linearly independent; its four vertices are

$$0, \quad u, \quad u+v, \quad v,$$

and its four sides are the segments from 0 to u, from u to $u+v$, from $u+v$ to v, and from v to 0. These sides are represented by the vectors u, v, $-u$, $-v$, so their lengths are $\|u\|, \|v\|, \|u\|, \|v\|$ (using 6.9(b)). The rhombus condition that all four sides have the same length therefore says precisely that

$$\|u\| = \|v\|.$$

The two diagonals of the parallelogram are the segment from 0 to $u+v$ and the segment from v to u; as vectors these diagonals are $u+v$ and $u-v$. Since u and v are linearly independent, both $u+v$ and $u-v$ are nonzero, so each diagonal really is a segment with a well-defined direction.
By (b), since $\|u\| = \|v\|$ we have

$$\langle u+v, u-v\rangle = 0,$$

so the two diagonals are orthogonal. As discussed after 6.10 (and proved in Exercise 6A.15), two nonzero vectors in $\mathbf{R}^2$ are orthogonal with respect to the Euclidean inner product exactly when the angle between them is a right angle. Hence the diagonals of a rhombus are perpendicular to each other.

Problem 6A.6. Suppose $u, v \in V$. Prove that $\langle u, v\rangle = 0 \iff \|u\| \le \|u+av\|$ for all $a \in \mathbf{F}$.

Solution. First suppose $\langle u, v\rangle = 0$, and let $a \in \mathbf{F}$. Then by 6.6(e),

$$\langle u, av\rangle = \overline{a}\langle u, v\rangle = 0,$$

so u and av are orthogonal. The Pythagorean theorem 6.12 therefore gives

$$\|u+av\|^2 = \|u\|^2 + \|av\|^2 \ge \|u\|^2.$$

Since norms are nonnegative, taking square roots yields $\|u\| \le \|u+av\|$. As $a \in \mathbf{F}$ was arbitrary, this proves the forward implication.
Conversely, suppose $\|u\| \le \|u+av\|$ for all $a \in \mathbf{F}$. If $v = 0$, then $\langle u, v\rangle = \langle u, 0\rangle = 0$ by 6.6(c) and we are done. So assume $v \ne 0$.
Because $v \ne 0$, the orthogonal decomposition 6.13 applies: setting

$$c = \frac{\langle u, v\rangle}{\|v\|^2} \quad \text{and} \quad w = u - cv,$$

we have $u = cv + w$ and $\langle w, v\rangle = 0$. Conjugate symmetry gives $\langle v, w\rangle = \overline{\langle w, v\rangle} = 0$, and hence by 6.6(a) applied in the first slot,

$$\langle cv, w\rangle = c\langle v, w\rangle = 0,$$

so cv and w are orthogonal. The Pythagorean theorem 6.12 then gives

$$\|u\|^2 = \|cv+w\|^2 = \|cv\|^2 + \|w\|^2 = |c|^2\|v\|^2 + \|w\|^2,$$

where the last equality uses 6.9(b).
Now apply the hypothesis with the particular scalar $a = -c$. Since $u + (-c)v = u - cv = w$, the hypothesis says $\|u\| \le \|w\|$, and hence $\|u\|^2 \le \|w\|^2$. Substituting the displayed identity for $\|u\|^2$ gives

$$|c|^2\|v\|^2 + \|w\|^2 \le \|w\|^2,$$

so $|c|^2\|v\|^2 \le 0$. Since $|c|^2\|v\|^2 \ge 0$, we conclude $|c|^2\|v\|^2 = 0$. Because $v \ne 0$, we have $\|v\| \ne 0$ by 6.9(a), and therefore $c = 0$. By the definition of c this means

$$\frac{\langle u, v\rangle}{\|v\|^2} = 0,$$

and hence $\langle u, v\rangle = 0$, as desired.
Geometrically, this exercise says that u is orthogonal to v exactly when 0 is the point of the line $\{-av : a \in \mathbf{F}\}$ closest to u, equivalently when u is a vector of smallest norm in the coset $u + \operatorname{span}(v)$.

Problem 6A.7. Suppose $u, v \in V$. Prove that $\|au + bv\| = \|bu + av\|$ for all $a, b \in \mathbf{R}$ if and only if $\|u\| = \|v\|$.

Solution. Note that a and b are required to be real numbers, although V may be a complex inner product space. The key is a single computation, valid for all $a, b \in \mathbf{R}$.
Let $a, b \in \mathbf{R}$. Using additivity in the first slot (6.2) and then additivity in the second slot (6.6(d)),

$$\|au + bv\|^2 = \langle au + bv, au + bv\rangle = \langle au, au\rangle + \langle au, bv\rangle + \langle bv, au\rangle + \langle bv, bv\rangle.$$

By 6.9(b) and the fact that a, b are real, $\langle au, au\rangle = \|au\|^2 = a^2\|u\|^2$ and $\langle bv, bv\rangle = b^2\|v\|^2$. For the cross terms, homogeneity in the first slot together with 6.6(e) gives, again using that a and b are real (so $\overline{a} = a$ and $\overline{b} = b$),

$$\langle au, bv\rangle = a\bar{b}\langle u, v\rangle = ab\langle u, v\rangle, \qquad \langle bv, au\rangle = b\overline{a}\langle v, u\rangle = ab\overline{\langle u, v\rangle},$$

the last step by conjugate symmetry. Since $z + \overline{z} = 2\operatorname{Re} z$ for every $z \in \mathbf{C}$, the two cross terms sum to $2ab\operatorname{Re}\langle u, v\rangle$. Hence

$$\|au + bv\|^2 = a^2\|u\|^2 + b^2\|v\|^2 + 2ab\operatorname{Re}\langle u, v\rangle.$$

Interchanging the roles of a and b in this identity (which is legitimate because it holds for all real a, b) gives

$$\|bu + av\|^2 = b^2\|u\|^2 + a^2\|v\|^2 + 2ab\operatorname{Re}\langle u, v\rangle.$$

Subtracting, the cross terms cancel and we obtain the identity

$$\|au + bv\|^2 - \|bu + av\|^2 = (a^2 - b^2)\big(\|u\|^2 - \|v\|^2\big)$$

for all $a, b \in \mathbf{R}$.
Now suppose $\|u\| = \|v\|$. Then $\|u\|^2 - \|v\|^2 = 0$, so the right side of the identity above is 0 for every choice of $a, b \in \mathbf{R}$. Hence $\|au + bv\|^2 = \|bu + av\|^2$, and since norms are nonnegative, $\|au + bv\| = \|bu + av\|$ for all $a, b \in \mathbf{R}$.
Conversely, suppose $\|au + bv\| = \|bu + av\|$ for all $a, b \in \mathbf{R}$. Taking $a = 1$ and $b = 0$ gives

$$\|u\| = \|1 \cdot u + 0 \cdot v\| = \|0 \cdot u + 1 \cdot v\| = \|v\|,$$

as desired. (Equivalently, the identity above with $a = 1$, $b = 0$ gives $\|u\|^2 - \|v\|^2 = 0$.)

Problem 6A.8. Suppose $a, b, c, x, y \in \mathbb{R}$ and $a^2 + b^2 + c^2 + x^2 + y^2 \le 1$. Prove that

$$a + b + c + 4x + 9y \le 10.$$

Solution. Work in $\mathbb{R}^5$ with its Euclidean inner product, so that

$$\langle (s_1, \dots, s_5), (t_1, \dots, t_5)\rangle = s_1t_1 + \cdots + s_5t_5.$$

Set

$$u = (a, b, c, x, y), \qquad w = (1, 1, 1, 4, 9).$$

Then

$$\langle u, w\rangle = a + b + c + 4x + 9y,$$

and

$$\|w\| = \sqrt{1^2 + 1^2 + 1^2 + 4^2 + 9^2} = \sqrt{1 + 1 + 1 + 16 + 81} = \sqrt{100} = 10.$$

The hypothesis $a^2+b^2+c^2+x^2+y^2 \le 1$ says exactly that $\|u\|^2 \le 1$, hence $\|u\| \le 1$.
Now the Cauchy–Schwarz inequality (6.14) gives

$$a+b+c+4x+9y = \langle u, w\rangle \le |\langle u, w\rangle| \le \|u\|\,\|w\| \le 1 \cdot 10 = 10,$$

as desired.
Remark on sharpness: equality forces $\|u\| = 1$ and equality in Cauchy–Schwarz, so u must be a scalar multiple of w; since also $\langle u, w\rangle = |\langle u, w\rangle|$, that multiple is nonnegative. Thus equality holds precisely for

$$(a, b, c, x, y) = \tfrac{1}{10}(1, 1, 1, 4, 9) = \left(\tfrac{1}{10}, \tfrac{1}{10}, \tfrac{1}{10}, \tfrac{2}{5}, \tfrac{9}{10}\right),$$

which indeed satisfies $a^2+b^2+c^2+x^2+y^2 = 1$ and $a+b+c+4x+9y = 10$. So the bound 10 cannot be improved.

Problem 6A.9. Suppose $u, v \in V$ and $\|u\| = \|v\| = 1$ and $\langle u, v\rangle = 1$. Prove that $u = v$.

Solution. Here V is an inner product space over $\mathbb{F}$, where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.
First note that conjugate symmetry gives

$$\langle v, u\rangle = \overline{\langle u, v\rangle} = \overline{1} = 1.$$

Now expand $\|u-v\|^2$ using the definition of the norm (6.7) together with 6.6: linearity in the first slot is 6.6(a), while in the second slot 6.6(d) and 6.6(e) give $\langle x, y-z\rangle = \langle x, y\rangle - \langle x, z\rangle$, since the complex conjugate of -1 is -1. Thus

$$\begin{aligned}\|u-v\|^2 &= \langle u-v, u-v\rangle \\ &= \langle u, u\rangle - \langle u, v\rangle - \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 - \langle u, v\rangle - \langle v, u\rangle + \|v\|^2 \\ &= 1 - 1 - 1 + 1 \\ &= 0.\end{aligned}$$

By 6.9(a), $\|u-v\| = 0$ implies $u - v = 0$, that is, $u = v$.

Problem 6A.10. Suppose $u, v \in V$ and $\|u\| \le 1$ and $\|v\| \le 1$. Prove that

$$\sqrt{1-\|u\|^2}\,\sqrt{1-\|v\|^2} \le 1 - |\langle u, v\rangle|.$$

Solution. Write $a = \|u\|$ and $b = \|v\|$. By hypothesis $a, b \in [0, 1]$, so both square roots above are defined and the quantities $1-a^2$ and $1-b^2$ are nonnegative.
Step 1: reduce to a statement about a and b. The Cauchy–Schwarz inequality (6.14) gives

$$|\langle u, v\rangle| \le \|u\|\,\|v\| = ab,$$

hence

$$1 - ab \le 1 - |\langle u, v\rangle|.$$

So it suffices to prove the numerical inequality

$$\sqrt{1-a^2}\,\sqrt{1-b^2} \le 1 - ab,$$

equivalently

$$\sqrt{1-a^2}\,\sqrt{1-b^2} + ab \le 1.$$

Step 2: prove the numerical inequality, again by Cauchy–Schwarz. Work in $\mathbb{R}^2$ with the Euclidean inner product and set

$$s = \left(\sqrt{1-a^2},\, a\right), \qquad t = \left(\sqrt{1-b^2},\, b\right).$$

Then

$$\|s\| = \sqrt{(1-a^2)+a^2} = 1, \qquad \|t\| = \sqrt{(1-b^2)+b^2} = 1,$$

and

$$\langle s, t\rangle = \sqrt{1-a^2}\,\sqrt{1-b^2} + ab.$$

By 6.14 applied in $\mathbb{R}^2$,

$$\sqrt{1-a^2}\,\sqrt{1-b^2} + ab = \langle s, t\rangle \le |\langle s, t\rangle| \le \|s\|\,\|t\| = 1.$$

Combining the two steps,

$$\sqrt{1-\|u\|^2}\,\sqrt{1-\|v\|^2} \le 1 - ab \le 1 - |\langle u, v\rangle|,$$

as desired.
Step 2 can also be checked by hand. Both sides of $\sqrt{1-a^2}\sqrt{1-b^2} \le 1-ab$ are nonnegative, because $ab \le 1$, so the inequality is equivalent to the one obtained by squaring:

$$(1-a^2)(1-b^2) \le (1-ab)^2.$$

Expanding, this says $1 - a^2 - b^2 + a^2b^2 \le 1 - 2ab + a^2b^2$, i.e. $2ab \le a^2 + b^2$, i.e. $0 \le (a-b)^2$, which is true.

Problem 6A.11. Find vectors $u, v \in \mathbb{R}^2$ such that u is a scalar multiple of $(1,3)$, v is orthogonal to $(1,3)$, and $(1,2) = u + v$.

Solution. This is exactly the orthogonal decomposition 6.13, applied with the vector being decomposed equal to $(1,2)$ and the vector being projected onto equal to $(1,3)$.
Following 6.13, put

$$c = \frac{\langle (1,2), (1,3)\rangle}{\|(1,3)\|^2}.$$

In $\mathbb{R}^2$ with the Euclidean inner product,

$$\langle (1,2), (1,3)\rangle = 1\cdot 1 + 2\cdot 3 = 7, \qquad \|(1,3)\|^2 = 1^2 + 3^2 = 10,$$

so $c = \frac{7}{10}$. Then 6.13 tells us to take

$$u = c\,(1,3) = \left(\tfrac{7}{10}, \tfrac{21}{10}\right), \qquad v = (1,2) - u = \left(1 - \tfrac{7}{10},\, 2 - \tfrac{21}{10}\right) = \left(\tfrac{3}{10}, -\tfrac{1}{10}\right).$$

Verification, so that the answer does not depend on citing 6.13:

- $u = \frac{7}{10}(1,3)$ is a scalar multiple of $(1,3)$.
- $\langle v, (1,3)\rangle = \frac{3}{10}\cdot 1 + \left(-\frac{1}{10}\right)\cdot 3 = \frac{3}{10} - \frac{3}{10} = 0$, so v is orthogonal to $(1,3)$.
- $u + v = \left(\frac{7}{10} + \frac{3}{10},\, \frac{21}{10} - \frac{1}{10}\right) = (1,2)$.

These vectors are moreover the only such pair. Indeed, suppose $u' = t(1,3)$ with $t \in \mathbb{R}$, that v' is orthogonal to $(1,3)$, and that $(1,2) = u' + v'$. Taking the inner product of both sides of $(1,2) = t(1,3) + v'$ with $(1,3)$ and using $\langle v', (1,3)\rangle = 0$ gives

$$7 = t\,\|(1,3)\|^2 = 10t,$$

so $t = \frac{7}{10}$, whence $u' = u$ and then $v' = (1,2) - u' = v$.

Problem 6A.12. Suppose a, b, c, d are positive numbers.
(a) Prove that $(a+b+c+d)\left(\dfrac{1}{a} + \dfrac{1}{b} + \dfrac{1}{c} + \dfrac{1}{d}\right) \ge 16$.
(b) For which positive numbers a, b, c, d is the inequality above an equality?

Solution. Work in $\mathbb{R}^4$ with the Euclidean inner product. Since $a, b, c, d > 0$, the following vectors are defined:

$$s = \left(\sqrt{a}, \sqrt{b}, \sqrt{c}, \sqrt{d}\right), \qquad t = \left(\frac{1}{\sqrt{a}}, \frac{1}{\sqrt{b}}, \frac{1}{\sqrt{c}}, \frac{1}{\sqrt{d}}\right).$$

Then

$$\langle s, t\rangle = \sqrt{a} \cdot \frac{1}{\sqrt{a}} + \sqrt{b} \cdot \frac{1}{\sqrt{b}} + \sqrt{c} \cdot \frac{1}{\sqrt{c}} + \sqrt{d} \cdot \frac{1}{\sqrt{d}} = 4,$$

while

$$\|s\|^2 = a + b + c + d, \qquad \|t\|^2 = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}.$$

(a) The Cauchy–Schwarz inequality (6.14) gives $|\langle s, t\rangle| \le \|s\| \, \|t\|$. Squaring (both sides are nonnegative) yields

$$16 = \langle s, t\rangle^2 \le \|s\|^2 \, \|t\|^2 = (a + b + c + d)\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right),$$

which is the desired inequality.
(b) Equality holds if and only if $a = b = c = d$.
First suppose $a = b = c = d$. Then

$$(a + b + c + d)\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right) = (4a) \cdot \frac{4}{a} = 16,$$

so the inequality is an equality.
Conversely, suppose equality holds. Then the chain in part (a) is an equality, so $|\langle s, t\rangle| = \|s\| \, \|t\|$, i.e. the Cauchy–Schwarz inequality is an equality for s and t. By the equality condition stated in 6.14, one of s, t is a scalar multiple of the other. Both are nonzero (their entries are positive), so we may write $t = \lambda s$ for some $\lambda \in \mathbb{R}$: if instead $s = \mu t$, then $\mu \neq 0$ because $s \neq 0$, and $t = \mu^{-1} s$. Comparing first coordinates,

$$\frac{1}{\sqrt{a}} = \lambda \sqrt{a} \quad \Longrightarrow \quad \lambda = \frac{1}{a},$$

and comparing the other three coordinates in the same way gives

$$\lambda = \frac{1}{a} = \frac{1}{b} = \frac{1}{c} = \frac{1}{d}.$$

Since a, b, c, d are positive, this forces $a = b = c = d$.
Thus the inequality in (a) is an equality exactly when $a = b = c = d$ (with no restriction on the common value, which may be any positive number).

Problem 6A.13. Show that the square of an average is less than or equal to the average of the squares. More precisely, show that if $a_1, \ldots, a_n \in \mathbb{R}$, then the square of the average of $a_1, \ldots, a_n$ is less than or equal to the average of $a_1^2, \ldots, a_n^2$.

Solution. We must show

$$\left(\frac{a_1 + \cdots + a_n}{n}\right)^2 \le \frac{a_1^2 + \cdots + a_n^2}{n}.$$

Work in $\mathbb{R}^n$ with the Euclidean inner product and set

$$a = (a_1, \ldots, a_n), \qquad e = (1, 1, \ldots, 1).$$

Then

$$\langle a, e\rangle = a_1 + \cdots + a_n, \qquad \|a\|^2 = a_1^2 + \cdots + a_n^2, \qquad \|e\|^2 = n.$$

The Cauchy–Schwarz inequality (6.14) gives $|\langle a, e\rangle| \le \|a\| \, \|e\|$; squaring both sides (both are nonnegative) yields

$$(a_1 + \cdots + a_n)^2 \le n \, (a_1^2 + \cdots + a_n^2).$$

This is precisely the case $x_j = a_j$, $y_j = 1$ of 6.16(a). Dividing both sides by $n^2 > 0$ gives

$$\left(\frac{a_1 + \cdots + a_n}{n}\right)^2 = \frac{(a_1 + \cdots + a_n)^2}{n^2} \le \frac{n(a_1^2 + \cdots + a_n^2)}{n^2} = \frac{a_1^2 + \cdots + a_n^2}{n},$$

which is the assertion that the square of the average of $a_1, \ldots, a_n$ is at most the average of $a_1^2, \ldots, a_n^2$. Equality analysis (not required, but immediate): by 6.14, equality in Cauchy–Schwarz holds if and only if one of a, e is a scalar multiple of the other. As $e \neq 0$, this happens exactly when $a = \lambda e$ for some $\lambda \in \mathbb{R}$, i.e. exactly when $a_1 = \cdots = a_n$. So the square of the average equals the average of the squares precisely when all the numbers are equal.

(In probabilistic language, the inequality says that the variance $\frac{1}{n}\sum_j a_j^2 - \left(\frac{1}{n}\sum_j a_j\right)^2$ of the list $a_1, \ldots, a_n$ is nonnegative, and vanishes only for a constant list.)

Problem 6A.14. Suppose $v \in V$ and $v \neq 0$. Prove that $v/\|v\|$ is the unique closest element on the unit sphere of V to v. More precisely, prove that if $u \in V$ and $\|u\| = 1$, then

$$\left\| v - \frac{v}{\|v\|} \right\| \le \|v - u\|,$$

with equality only if $u = v/\|v\|$.

Solution. Since $v \neq 0$, we have $\|v\| > 0$ by 6.9(a), so $v/\|v\|$ is defined; and $\|v/\|v\|\| = \|v\|^{-1}\|v\| = 1$ by 6.9(b), so $v/\|v\|$ does lie on the unit sphere. Throughout, $u \in V$ satisfies $\|u\| = 1$.

Step 1: compute the left side. Using 6.9(b),

$$\left\| v - \frac{v}{\|v\|} \right\| = \left\| \left(1 - \frac{1}{\|v\|}\right) v \right\| = \left| 1 - \frac{1}{\|v\|} \right| \|v\| = \big|\, \|v\| - 1 \,\big|,$$

the last step because $\|v\| > 0$.

Step 2: bound the right side from below. Expanding $\|v - u\|^2 = \langle v - u, v - u\rangle$ by the definition of the norm (6.7) and using conjugate symmetry (so that $\langle v, u\rangle + \langle u, v\rangle = \langle v, u\rangle + \overline{\langle v, u\rangle} = 2\,\mathrm{Re}\langle v, u\rangle$),

$$\|v - u\|^2 = \|v\|^2 - 2\,\mathrm{Re}\langle v, u\rangle + \|u\|^2 = \|v\|^2 - 2\,\mathrm{Re}\langle v, u\rangle + 1.$$

Now the real part of a scalar is at most its absolute value, and the Cauchy–Schwarz inequality (6.14) bounds that absolute value, so

$$\mathrm{Re}\langle v, u\rangle \le |\langle v, u\rangle| \le \|v\|\,\|u\| = \|v\|,$$

and therefore

$$\|v - u\|^2 \ge \|v\|^2 - 2\|v\| + 1 = \big(\|v\| - 1\big)^2.$$

Taking nonnegative square roots and using Step 1,

$$\|v - u\| \ge \big|\, \|v\| - 1 \,\big| = \left\| v - \frac{v}{\|v\|} \right\|,$$

which is the asserted inequality. (If $\mathbb{F} = \mathbb{R}$, read $\mathrm{Re}\langle v, u\rangle$ as $\langle v, u\rangle$; the argument is unchanged.)

Step 3: the equality case. Suppose $\|v - u\| = \big|\, \|v\| - 1 \,\big|$. Squaring and comparing with the identity in Step 2,

$$\|v\|^2 - 2\,\mathrm{Re}\langle v, u\rangle + 1 = \|v\|^2 - 2\|v\| + 1,$$

so $\mathrm{Re}\langle v, u\rangle = \|v\| = \|v\|\,\|u\|$. Since always $\mathrm{Re}\langle v, u\rangle \le |\langle v, u\rangle| \le \|v\|\,\|u\|$, both of these inequalities must be equalities. In particular $|\langle v, u\rangle| = \|v\|\,\|u\|$, so the Cauchy–Schwarz inequality is an equality for v and u; by the equality condition stated in 6.14, one of v, u is a scalar multiple of the other. Both v and u are nonzero, so in either case we may write

$$v = \lambda u \quad \text{for some } \lambda \in \mathbb{F},\ \lambda \neq 0.$$

(If instead $u = \mu v$, then $\mu \neq 0$ because $u \neq 0$, and $v = \mu^{-1}u$.)
For this λ we have $\langle v, u\rangle = \lambda\langle u, u\rangle = \lambda\|u\|^2 = \lambda$ and $\|v\| = |\lambda|\,\|u\| = |\lambda|$. The equality $\operatorname{Re}\langle v, u\rangle = \|v\|$ therefore reads

$$\operatorname{Re}\lambda = |\lambda|,$$

which forces λ to be a nonnegative real number; as $\lambda \neq 0$, we get $\lambda > 0$ and hence $\|v\| = |\lambda| = \lambda$. Consequently

$$u = \frac{v}{\lambda} = \frac{v}{\|v\|}.$$

Thus equality holds only if $u = v/\|v\|$. Since equality obviously does hold for $u = v/\|v\|$, the vector $v/\|v\|$ is the unique point of the unit sphere $\{u \in V : \|u\| = 1\}$ at minimum distance from v, that distance being $\big|\,\|v\| - 1\,\big|$.

Problem 6A.15. Suppose u, v are nonzero vectors in $\mathbf{R}^2$. Prove that

$$\langle u, v\rangle = \|u\|\,\|v\|\cos\theta,$$

where θ is the angle between u and v (thinking of u and v as arrows with initial point at the origin). Hint: Use the law of cosines on the triangle formed by u, v, and $u - v$.

Solution. Throughout, $\mathbf{R}^2$ carries the Euclidean inner product, so that $\|w\|$ is exactly the ordinary Euclidean length of the arrow w, and $\|u - v\|$ is the ordinary distance between the points u and v.
First we record an algebraic identity valid in any real inner product space. Because the inner product on $\mathbf{R}^2$ is real, it is symmetric, so $\langle u, v\rangle = \langle v, u\rangle$. Hence

$$\begin{aligned}\|u - v\|^2 &= \langle u - v,\, u - v\rangle \\ &= \langle u, u\rangle - \langle u, v\rangle - \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + \|v\|^2 - 2\langle u, v\rangle.\end{aligned}$$

Now let $\theta \in [0, \pi]$ denote the angle between the arrows u and v based at the origin.
Case 1: u and v are not scalar multiples of each other. Then the three points 0, u, v are not collinear, so they form a genuine triangle. The two sides emanating from the vertex 0 have lengths $\|u\|$ and $\|v\|$, the angle between them at that vertex is θ, and the side opposite that vertex is the segment joining u to v, whose length is $\|u - v\|$. The law of cosines from Euclidean geometry therefore gives

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\,\|v\|\cos\theta.$$

Comparing this with the identity computed above, we get

$$\|u\|^2 + \|v\|^2 - 2\langle u, v\rangle = \|u\|^2 + \|v\|^2 - 2\|u\|\,\|v\|\cos\theta,$$

and cancelling $\|u\|^2 + \|v\|^2$ and dividing by -2 yields $\langle u, v\rangle = \|u\|\,\|v\|\cos\theta$, as desired.
Case 2: one of u, v is a scalar multiple of the other. Since both are nonzero, we may write $v = tu$ for some $t \in \mathbf{R}$ with $t \neq 0$ (if instead $u = sv$ with $s \neq 0$, take $t = 1/s$). Here the triangle degenerates, so we verify the formula directly. If $t > 0$, then v points in the same direction as u, so $\theta = 0$ and $\cos\theta = 1$; moreover $\|v\| = |t|\,\|u\| = t\|u\|$ by 6.9(b), so

$$\langle u, v\rangle = \langle u, tu\rangle = t\|u\|^2 = \|u\|\big(t\|u\|\big) = \|u\|\,\|v\| = \|u\|\,\|v\|\cos\theta.$$

If $t < 0$, then v points in the direction opposite to u, so $\theta = \pi$ and $\cos\theta = -1$; now $\|v\| = |t|\,\|u\| = -t\|u\|$, so

$$\langle u, v\rangle = t\|u\|^2 = -\|u\|\big(-t\|u\|\big) = -\|u\|\,\|v\| = \|u\|\,\|v\|\cos\theta.$$

In both cases $\langle u, v\rangle = \|u\|\,\|v\|\cos\theta$, completing the proof.

Problem 6A.16. The angle between two vectors (thought of as arrows with initial point at the origin) in $\mathbf{R}^2$ or $\mathbf{R}^3$ can be defined geometrically. However, geometry is not as clear in $\mathbf{R}^n$ for $n > 3$. Thus the angle between two nonzero vectors $x, y \in \mathbf{R}^n$ is defined to be

$$\arccos \frac{\langle x, y \rangle}{\|x\| \, \|y\|},$$

where the motivation for this definition comes from Exercise 15. Explain why the Cauchy–Schwarz inequality is needed to show that this definition makes sense.

Solution. The function arccos is the inverse of the restriction of cos to $[0, \pi]$. Since cos maps $[0, \pi]$ bijectively onto $[-1, 1]$, the domain of arccos is exactly the interval $[-1, 1]$, and arccos maps that interval bijectively onto $[0, \pi]$. Thus the expression

$$\arccos \frac{\langle x, y \rangle}{\|x\| \, \|y\|}$$

denotes a real number if and only if the quotient $\dfrac{\langle x, y \rangle}{\|x\| \, \|y\|}$ lies in $[-1, 1]$; for a quotient outside $[-1, 1]$ the symbol arccos has no meaning, and the proposed definition of angle would be vacuous.
Two things therefore have to be checked, and each is supplied by the material of this section.
First, the quotient must be defined at all, meaning the denominator must be nonzero. Because x and y are assumed nonzero, 6.9(a) gives $\|x\| \neq 0$ and $\|y\| \neq 0$, so $\|x\| \, \|y\| > 0$.
Second, and this is the essential point, the quotient must land in the domain $[-1, 1]$ of arccos. This is precisely what the Cauchy–Schwarz inequality (6.14) provides: it states that

$$|\langle x, y \rangle| \leq \|x\| \, \|y\|.$$

Dividing both sides by the positive number $\|x\| \, \|y\|$ gives

$$\left| \frac{\langle x, y \rangle}{\|x\| \, \|y\|} \right| \leq 1, \qquad \text{that is,} \qquad -1 \leq \frac{\langle x, y \rangle}{\|x\| \, \|y\|} \leq 1.$$

Hence the quotient does lie in $[-1, 1]$, and so the angle $\arccos \frac{\langle x, y \rangle}{\|x\| \, \|y\|}$ is a well-defined real number in $[0, \pi]$ for every pair of nonzero $x, y \in \mathbf{R}^n$.
Without the Cauchy–Schwarz inequality there would be no reason to believe that the quotient stays within $[-1, 1]$; nothing in the mere definitions of $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ makes this evident. In coordinates, the assertion being used is the concrete inequality 6.16(a),

$$(x_1 y_1 + \cdots + x_n y_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2),$$

which is exactly Cauchy–Schwarz for the Euclidean inner product on $\mathbf{R}^n$.
It is worth adding that the definition is also consistent with the geometric notion of angle in the low-dimensional cases, which is what makes it a reasonable definition rather than merely a meaningful one: by Exercise 15, for nonzero $u, v \in \mathbf{R}^2$ the geometric angle $\theta \in [0, \pi]$ between u and v satisfies $\cos\theta = \frac{\langle u, v \rangle}{\|u\| \, \|v\|}$, and applying arccos (legitimate since $\theta \in [0, \pi]$) gives $\theta = \arccos \frac{\langle u, v \rangle}{\|u\| \, \|v\|}$. Finally, the equality case in 6.14 shows that the angle equals 0 or π exactly when one of x, y is a scalar multiple of the other, and the angle equals $\pi/2$ exactly when $\langle x, y \rangle = 0$, that is, when x and y are orthogonal in the sense of 6.10 — so the new definition matches the older notions where both apply.

Problem 6A.17. Prove that

$$\left(\sum_{k=1}^{n} a_k b_k \right)^2 \leq \left(\sum_{k=1}^{n} k a_k^2 \right) \left(\sum_{k=1}^{n} \frac{b_k^2}{k} \right)$$

for all real numbers $a_1, \dots, a_n$ and $b_1, \dots, b_n$.

Solution. The point is to split each product $a_k b_k$ so that the two halves have precisely the sizes appearing on the right side. Since $k \geq 1 > 0$ for each $k \in \{1, \dots, n\}$, the real number $\sqrt{k}$ is defined and nonzero, and we may write

$$a_k b_k = \left(\sqrt{k}\, a_k\right)\left(\frac{b_k}{\sqrt{k}}\right) \qquad \text{for } k = 1, \dots, n.$$

Define $x, y \in \mathbf{R}^n$ by

$$x = \left(\sqrt{1}\, a_1,\ \sqrt{2}\, a_2,\ \dots,\ \sqrt{n}\, a_n\right), \qquad y = \left(\frac{b_1}{\sqrt{1}},\ \frac{b_2}{\sqrt{2}},\ \dots,\ \frac{b_n}{\sqrt{n}}\right).$$

Writing $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$, so that $x_k = \sqrt{k}\, a_k$ and $y_k = b_k/\sqrt{k}$, we have

$$x_k y_k = a_k b_k, \qquad x_k^2 = k a_k^2, \qquad y_k^2 = \frac{b_k^2}{k}$$

for each k.
Now apply the Cauchy–Schwarz inequality (6.14) to x and y in $\mathbf{R}^n$ with the Euclidean inner product; in the explicit coordinate form recorded in 6.16(a), this says

$$(x_1 y_1 + \cdots + x_n y_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2).$$

Substituting the three displayed identities gives

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 \leq \left(\sum_{k=1}^{n} k a_k^2\right)\left(\sum_{k=1}^{n} \frac{b_k^2}{k}\right),$$

which is the desired inequality.
Equality analysis (not required, but immediate from the equality case of 6.14): equality holds if and only if one of x, y is a scalar multiple of the other. Unwinding the substitutions, this says that either there is $c \in \mathbf{R}$ with $b_k = c\, k\, a_k$ for every k (the case $y = cx$, since $b_k/\sqrt{k} = c\sqrt{k}\, a_k$), or there is $d \in \mathbf{R}$ with $k\, a_k = d\, b_k$ for every k (the case $x = dy$). No nonvanishing hypothesis on the a_k is needed for this, and the two alternatives are both required only to cover the degenerate cases $x = 0$ and $y = 0$.

Problem 6A.18. (a) Suppose $f\colon [1, \infty) \to [0, \infty)$ is continuous. Show that

$$\left(\int_1^\infty f\right)^2 \leq \int_1^\infty x^2 \big(f(x)\big)^2\, dx.$$

(b) For which continuous functions $f\colon [1, \infty) \to [0, \infty)$ is the inequality in (a) an equality with both sides finite?

Solution. Notation and conventions. For $b > 1$ put

$$F(b) = \int_1^b f, \qquad G(b) = \int_1^b x^2 \big(f(x)\big)^2\, dx.$$

Both integrands are continuous and nonnegative on $[1, \infty)$ (note $f \geq 0$ by hypothesis, and $x^2 f(x)^2 \geq 0$ always), so F and G are nondecreasing functions of b with nonnegative values. Consequently the improper integrals

$$L = \int_1^\infty f = \lim_{b \to \infty} F(b), \qquad M = \int_1^\infty x^2 \big(f(x)\big)^2\, dx = \lim_{b \to \infty} G(b)$$

exist in $[0, \infty]$: each limit is either a finite number (if the corresponding function is bounded) or $+\infty$. The inequality in (a) is the assertion $L^2 \leq M$, understood in $[0, \infty]$.

We shall use the inner product space of continuous real-valued functions on a closed bounded interval $[1, b]$, with inner product

$$\langle g, h\rangle = \int_1^b gh,$$

which is an inner product exactly as in 6.3(c) (there stated on $[-1, 1]$; the same verification works on any $[1, b]$). Cauchy–Schwarz (6.14) in this space is the integral inequality recorded in 6.16(b):

$$\left(\int_1^b gh\right)^2 \le \left(\int_1^b g^2\right)\left(\int_1^b h^2\right).$$

Part (a). If $M = \infty$ there is nothing to prove, since the right side is then $+\infty$. So the substance of the claim is that L is finite and $L^2 \le M$ whenever $M < \infty$; in fact the argument below handles both cases at once.

Fix $b > 1$ and apply the displayed Cauchy–Schwarz inequality on $[1, b]$ to the continuous functions

$$g(x) = \frac{1}{x}, \qquad h(x) = xf(x),$$

which are indeed continuous on $[1, b]$ (the first because $x \neq 0$ there). Since $g(x)h(x) = f(x)$, we obtain

$$F(b)^2 = \left(\int_1^b f\right)^2 \le \left(\int_1^b \frac{dx}{x^2}\right)\left(\int_1^b x^2\big(f(x)\big)^2\,dx\right) = \left(1 - \frac{1}{b}\right) G(b),$$

using $\int_1^b x^{-2}\,dx = \left[-x^{-1}\right]_1^b = 1 - \frac{1}{b}$. Since $0 < 1 - \frac{1}{b} < 1$ and $G(b) \ge 0$, this gives

$$F(b)^2 \le G(b) \le M \qquad \text{for every } b > 1.$$

If $M = \infty$ the desired inequality is trivial. If $M < \infty$, then the nondecreasing function F is bounded above by $\sqrt{M}$, hence $L = \lim_{b\to\infty} F(b)$ is finite and $L \le \sqrt{M}$; alternatively, let $b \to \infty$ in $F(b)^2 \le G(b)$ to get $L^2 \le M$ directly. In either case

$$\left(\int_1^\infty f\right)^2 \le \int_1^\infty x^2\big(f(x)\big)^2\,dx,$$

as required.

Remark on sharpness: the constant is exactly right because $\int_1^\infty x^{-2}\,dx = 1$, so the factor $\int_1^b g^2$ increases to 1 and no smaller constant would do. Part (b) confirms this by exhibiting the functions achieving equality.

Part (b). Claim: equality holds with both sides finite if and only if there is a constant $c \in [0, \infty)$ with

$$f(x) = \frac{c}{x^2} \qquad \text{for all } x \ge 1.$$

First suppose $f(x) = c/x^2$ with $c \ge 0$ (this is continuous on $[1, \infty)$ and takes values in $[0, \infty)$, as required). Then

$$\int_1^\infty f = c\int_1^\infty \frac{dx}{x^2} = c, \qquad \int_1^\infty x^2\big(f(x)\big)^2\,dx = \int_1^\infty x^2 \cdot \frac{c^2}{x^4}\,dx = c^2\int_1^\infty \frac{dx}{x^2} = c^2.$$

Both quantities are finite and $\left(\int_1^\infty f\right)^2 = c^2 = \int_1^\infty x^2 f(x)^2\,dx$, so equality holds with both sides finite. Conversely, suppose $L^2 = M < \infty$. In particular L is finite, and $L \ge 0$ because $f \ge 0$. Put $c = L$ and, for $b > 1$, consider

$$\Phi(b) = \int_1^b \left(xf(x) - \frac{c}{x}\right)^2 dx.$$

Expanding the square, and using $xf(x) \cdot \frac{c}{x} = cf(x)$, gives

$$\Phi(b) = \int_1^b x^2\big(f(x)\big)^2 dx - 2c\int_1^b f + c^2\int_1^b \frac{dx}{x^2} = G(b) - 2cF(b) + c^2\left(1 - \frac{1}{b}\right).$$

Letting $b \to \infty$ and using $G(b) \to M$, $F(b) \to L = c$, and $1 - \frac{1}{b} \to 1$ (all limits finite), we obtain

$$\lim_{b\to\infty} \Phi(b) = M - 2cL + c^2 = M - 2L^2 + L^2 = M - L^2 = 0.$$

But the integrand defining Φ is nonnegative, so Φ is a nondecreasing function of b with $\Phi(b) \geq 0$ for all b. A nondecreasing nonnegative function whose limit at ∞ equals 0 must be identically 0: indeed, if $\Phi(b_0) > 0$ for some b_0, then $\Phi(b) \geq \Phi(b_0) > 0$ for all $b \geq b_0$, so the limit would be at least $\Phi(b_0) > 0$. Hence

$$\int_1^b \Big(xf(x) - \frac{c}{x}\Big)^2 \, dx = 0 \qquad \text{for every } b > 1.$$

Finally, a continuous nonnegative function with zero integral over an interval is identically zero on that interval. (If φ is continuous, $\varphi \geq 0$, and $\varphi(x_0) = \varepsilon > 0$ for some $x_0 \in [1, b]$, then by continuity $\varphi > \varepsilon/2$ on some subinterval $J \subseteq [1, b]$ of positive length δ, whence $\int_1^b \varphi \geq \varepsilon\delta/2 > 0$.) Applying this to $\varphi(x) = \big(xf(x) - c/x\big)^2$ on each $[1, b]$ gives

$$xf(x) - \frac{c}{x} = 0 \qquad \text{for all } x \geq 1,$$

that is, $f(x) = c/x^2$ for all $x \geq 1$, with $c = L \geq 0$.
Therefore equality with both sides finite holds precisely for the functions $f(x) = c/x^2$ with c a nonnegative constant. The case $c = 0$ is the zero function, for which both sides equal 0.

Problem 6A.19. Suppose $v_1, \ldots, v_n$ is a basis of V and $T \in \mathcal{L}(V)$. Prove that if λ is an eigenvalue of T, then

$$|\lambda|^2 \leq \sum_{j=1}^{n} \sum_{k=1}^{n} \big|\mathcal{M}(T)_{j,k}\big|^2,$$

where $\mathcal{M}(T)_{j,k}$ denotes the entry in row j, column k of the matrix of T with respect to the basis $v_1, \ldots, v_n$.

Solution. Write $A_{j,k} = \mathcal{M}(T)_{j,k}$ for the entries of the matrix of T with respect to the basis $v_1, \ldots, v_n$. By the definition of that matrix (3.31), the k^{th} column of $\mathcal{M}(T)$ records the coefficients of Tv_k in the basis $v_1, \ldots, v_n$:

$$Tv_k = \sum_{j=1}^{n} A_{j,k}\, v_j \qquad \text{for } k = 1, \ldots, n.$$

Let $\lambda \in \mathbf{F}$ be an eigenvalue of T and choose a corresponding eigenvector $v \in V$ with $v \neq 0$ and $Tv = \lambda v$. Because $v_1, \ldots, v_n$ is a basis of V, there are unique scalars $c_1, \ldots, c_n \in \mathbf{F}$ with

$$v = \sum_{k=1}^{n} c_k v_k,$$

and since $v \neq 0$ not all the c_k are 0; hence

$$\sum_{k=1}^{n} |c_k|^2 > 0.$$

Now compute Tv two ways. On one hand, by linearity of T,

$$Tv = \sum_{k=1}^{n} c_k\, Tv_k = \sum_{k=1}^{n} c_k \sum_{j=1}^{n} A_{j,k} v_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} A_{j,k} c_k \right) v_j.$$

On the other hand,

$$Tv = \lambda v = \sum_{j=1}^{n} \lambda c_j\, v_j.$$

Since the representation of a vector in the basis $v_1, \dots, v_n$ is unique (equivalently, by the linear independence of $v_1, \dots, v_n$), comparing coefficients gives

$$\sum_{k=1}^{n} A_{j,k} c_k = \lambda c_j \qquad \text{for } j = 1, \dots, n.$$

Fix j and estimate the left side using the Cauchy–Schwarz inequality (6.14) in $\mathbf{F}^n$ with the Euclidean inner product $\langle x, y \rangle = \sum_k x_k \overline{y_k}$. Take

$$x = (c_1, \dots, c_n), \qquad y = \left(\overline{A_{j,1}}, \dots, \overline{A_{j,n}}\right),$$

so that $\langle x, y \rangle = \sum_{k=1}^{n} c_k A_{j,k}$, while $\|x\|^2 = \sum_{k=1}^{n} |c_k|^2$ and $\|y\|^2 = \sum_{k=1}^{n} |A_{j,k}|^2$ (using $|\overline{A_{j,k}}| = |A_{j,k}|$). Cauchy–Schwarz gives $|\langle x, y \rangle|^2 \le \|x\|^2 \|y\|^2$, that is,

$$|\lambda|^2 |c_j|^2 = \left| \sum_{k=1}^{n} A_{j,k} c_k \right|^2 \le \left(\sum_{k=1}^{n} |A_{j,k}|^2 \right) \left(\sum_{k=1}^{n} |c_k|^2 \right).$$

Summing this over $j = 1, \dots, n$ yields

$$|\lambda|^2 \sum_{j=1}^{n} |c_j|^2 \le \left(\sum_{j=1}^{n} \sum_{k=1}^{n} |A_{j,k}|^2 \right) \left(\sum_{k=1}^{n} |c_k|^2 \right).$$

Since $\sum_{k=1}^{n} |c_k|^2 > 0$, we may divide both sides by this positive number, obtaining

$$|\lambda|^2 \le \sum_{j=1}^{n} \sum_{k=1}^{n} |A_{j,k}|^2 = \sum_{j=1}^{n} \sum_{k=1}^{n} \left| \mathcal{M}(T)_{j,k} \right|^2,$$

as desired.
Note that no assumption was made about the basis $v_1, \dots, v_n$ beyond its being a basis, and no use was made of any inner product on V: the Cauchy–Schwarz inequality was applied in $\mathbf{F}^n$ to the coordinate vectors, not to vectors in V. In particular the bound holds for every basis of V simultaneously, so $|\lambda|$ is at most the infimum of $\left(\sum_{j,k} |\mathcal{M}(T)_{j,k}|^2 \right)^{1/2}$ taken over all bases of V.

Problem 6A.20. Prove that if $u, v \in V$, then $\big| \|u\| - \|v\| \big| \le \|u - v\|$.
The inequality above is called the reverse triangle inequality. For the reverse triangle inequality when $V = \mathbf{C}$, see Exercise 2 in Chapter 4.

Solution. Let $u, v \in V$. The absolute value of a real number t satisfies $|t| \le s$ if and only if both $t \le s$ and $-t \le s$. So with $t = \|u\| - \|v\|$ and $s = \|u - v\|$, it suffices to prove the two inequalities

$$\|u\| - \|v\| \le \|u - v\| \qquad \text{and} \qquad \|v\| - \|u\| \le \|u - v\|.$$

For the first, write u as the sum of two vectors and apply the triangle inequality (6.17):

$$\|u\| = \big\| (u - v) + v \big\| \le \|u - v\| + \|v\|.$$

Subtracting $\|v\|$ from both sides gives $\|u\| - \|v\| \le \|u - v\|$.
For the second, interchange the roles of u and v in the argument just given:

$$\|v\| = \big\| (v - u) + u \big\| \le \|v - u\| + \|u\|,$$

so $\|v\| - \|u\| \le \|v - u\|$. Finally $v - u = (-1)(u - v)$, so by 6.9(b),

$$\|v - u\| = |-1| \, \|u - v\| = \|u - v\|.$$

Hence $\|v\| - \|u\| \le \|u - v\|$.
Combining the two displayed inequalities gives

$$\big|\|u\| - \|v\|\big| \le \|u - v\|,$$

as desired.
Two remarks. First, replacing v by $-v$ (and using $\|-v\| = \|v\|$ from 6.9(b) again) gives the companion form $\big|\|u\| - \|v\|\big| \le \|u + v\|$. Second, taking $V = \mathbf{C}$ with the inner product $\langle w, z\rangle = w\overline{z}$, the norm is the usual absolute value, and the inequality becomes $\big||u| - |v|\big| \le |u - v|$ for complex numbers u, v, which is Exercise 2 in Chapter 4.

Problem 6A.21. Suppose $u, v \in V$ are such that

$$\|u\| = 3, \qquad \|u + v\| = 4, \qquad \|u - v\| = 6.$$

What number does $\|v\|$ equal?

Solution. The parallelogram equality (6.21) states that for all $u, v \in V$,

$$\|u + v\|^2 + \|u - v\|^2 = 2\big(\|u\|^2 + \|v\|^2\big).$$

Substituting the three given values $\|u\| = 3$, $\|u + v\| = 4$, $\|u - v\| = 6$ gives

$$4^2 + 6^2 = 2\big(3^2 + \|v\|^2\big),$$

that is,

$$16 + 36 = 18 + 2\|v\|^2.$$

Hence $2\|v\|^2 = 52 - 18 = 34$, so $\|v\|^2 = 17$. Since norms are nonnegative,

$$\|v\| = \sqrt{17}.$$

A consistency check: the data are realizable, so this is a genuine answer rather than a vacuous deduction. Take $V = \mathbf{R}^2$ with the Euclidean inner product, $u = (3, 0)$, and $v = (t, s)$. Then $\|u+v\|^2 - \|u-v\|^2 = 4\langle u, v\rangle = 12t$, and the data force $16 - 36 = 12t$, so $t = -\frac{5}{3}$; and $\|v\|^2 = t^2 + s^2 = 17$ then gives $s^2 = 17 - \frac{25}{9} = \frac{128}{9}$, so $s = \frac{8\sqrt{2}}{3}$ works. One checks directly that $u = (3, 0)$ and $v = \big(-\frac{5}{3}, \frac{8\sqrt{2}}{3}\big)$ satisfy all three given equations.
Note also that the answer is forced by the parallelogram equality alone, so $\|v\| = \sqrt{17}$ no matter which inner product space V is and which vectors u, v satisfy the three conditions.

Problem 6A.22. Show that if $u, v \in V$, then

$$\|u + v\|\,\|u - v\| \le \|u\|^2 + \|v\|^2.$$

Solution. For all real numbers a, b we have $0 \le (a - b)^2 = a^2 - 2ab + b^2$, and hence

$$ab \le \frac{a^2 + b^2}{2}.$$

Apply this with $a = \|u + v\|$ and $b = \|u - v\|$, both of which are real numbers. We get

$$\|u + v\|\,\|u - v\| \le \frac{\|u + v\|^2 + \|u - v\|^2}{2}.$$

The parallelogram equality (6.21) states that

$$\|u + v\|^2 + \|u - v\|^2 = 2\big(\|u\|^2 + \|v\|^2\big).$$

Substituting this into the previous inequality gives

$$\|u+v\|\,\|u-v\| \le \frac{2\big(\|u\|^2+\|v\|^2\big)}{2} = \|u\|^2+\|v\|^2,$$

as desired.
Note that equality in the first inequality holds if and only if $a=b$, so the chain above is an equality precisely when $\|u+v\| = \|u-v\|$. Expanding as in the proof of the triangle inequality 6.17, for all $u, v \in V$ we have

$$\|u \pm v\|^2 = \|u\|^2 + \|v\|^2 \pm 2\,\mathrm{Re}\langle u, v\rangle,$$

so equality holds if and only if $\mathrm{Re}\langle u, v\rangle = 0$ (which in a real inner product space means exactly that u and v are orthogonal).

Problem 6A.23. Suppose $v_1, \dots, v_m \in V$ are such that $\|v_k\| \le 1$ for each $k = 1, \dots, m$. Show that there exist $a_1, \dots, a_m \in \{1, -1\}$ such that

$$\|a_1 v_1 + \cdots + a_m v_m\| \le \sqrt{m}.$$

Solution. Let $A = \{1,-1\}^m$ be the set of all 2^m sign vectors $a = (a_1, \dots, a_m)$, and for $a \in A$ put

$$N(a) = \|a_1 v_1 + \cdots + a_m v_m\|^2.$$

We compute the sum of $N(a)$ over all $a \in A$ and show it is at most $2^m m$; it then follows that $N(a) \le m$ for at least one $a \in A$, which is what we want.
Fix $a \in A$. Since each a_k is real, $\overline{a_k} = a_k$, so expanding by additivity and homogeneity in each slot (using 6.6) gives

$$N(a) = \Big\langle \sum_{j=1}^m a_j v_j, \sum_{k=1}^m a_k v_k \Big\rangle = \sum_{j=1}^m \sum_{k=1}^m a_j a_k \langle v_j, v_k\rangle.$$

Now sum over all $a \in A$ and interchange the order of summation:

$$\sum_{a\in A} N(a) = \sum_{j=1}^m \sum_{k=1}^m \Big(\sum_{a\in A} a_j a_k\Big)\langle v_j, v_k\rangle.$$

Consider the inner sum $\sum_{a\in A} a_j a_k$.
If $j = k$, then $a_j a_k = a_j^2 = 1$ for every $a \in A$, so the inner sum equals 2^m.
If $j \ne k$, let $\sigma\colon A \to A$ be the map that flips the j^{th} coordinate, leaving all other coordinates unchanged. Then σ is an involution with no fixed points, so it partitions A into 2^{m-1} two-element sets $\{a, \sigma(a)\}$. For each such pair, $\sigma(a)_j \sigma(a)_k = (-a_j)a_k = -a_j a_k$, so the two contributions cancel. Hence the inner sum equals 0 when $j \ne k$.
Therefore

$$\sum_{a\in A} N(a) = 2^m \sum_{k=1}^m \langle v_k, v_k\rangle = 2^m \sum_{k=1}^m \|v_k\|^2 \le 2^m m,$$

where the last inequality uses $\|v_k\| \le 1$ for each k.
The sum of the 2^m numbers $N(a)$ is at most $2^m m$, so the smallest of them is at most m: if every $N(a)$ exceeded m, the sum would exceed $2^m m$. Thus there exists $a = (a_1, \dots, a_m) \in \{1,-1\}^m$ with

$$\|a_1 v_1 + \cdots + a_m v_m\|^2 = N(a) \le m,$$

and taking square roots gives $\|a_1 v_1 + \cdots + a_m v_m\| \le \sqrt{m}$, as desired.

Problem 6A.24. Prove or give a counterexample: If $\|\cdot\|$ is the norm associated with an inner product on $\mathbb{R}^2$, then there exists $(x, y) \in \mathbb{R}^2$ such that $\|(x, y)\| \neq \max\{|x|, |y|\}$.

Solution. The statement is true. In other words, no inner product on $\mathbb{R}^2$ has $\max\{|x|, |y|\}$ as its associated norm.
Suppose, seeking a contradiction, that $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{R}^2$ whose associated norm satisfies

$$\|(x, y)\| = \max\{|x|, |y|\} \quad \text{for every } (x, y) \in \mathbb{R}^2,$$

so that no vector as in the exercise exists.
Take $u = (1, 0)$ and $v = (0, 1)$. Then

$$\|u\| = \max\{1, 0\} = 1, \qquad \|v\| = \max\{0, 1\} = 1,$$

and

$$u + v = (1, 1), \qquad u - v = (1, -1),$$

so

$$\|u + v\| = \max\{1, 1\} = 1, \qquad \|u - v\| = \max\{1, |-1|\} = 1.$$

Since the norm comes from an inner product, the parallelogram equality (6.21) must hold for u and v. But

$$\|u + v\|^2 + \|u - v\|^2 = 1^2 + 1^2 = 2,$$

while

$$2\big(\|u\|^2 + \|v\|^2\big) = 2(1 + 1) = 4.$$

Thus $2 = 4$, a contradiction.
Hence no inner product on $\mathbb{R}^2$ induces the norm $(x, y) \mapsto \max\{|x|, |y|\}$. Equivalently, if $\|\cdot\|$ is the norm associated with an inner product on $\mathbb{R}^2$, then there must exist some $(x, y) \in \mathbb{R}^2$ with $\|(x, y)\| \neq \max\{|x|, |y|\}$, which is exactly the assertion of the exercise.
Concretely, the contradiction above shows that for any inner product norm on $\mathbb{R}^2$ at least one of the four vectors $(1, 0)$, $(0, 1)$, $(1, 1)$, $(1, -1)$ already has norm different from the maximum of the absolute values of its coordinates.

Problem 6A.25. Suppose $p > 0$. Prove that there is an inner product on $\mathbb{R}^2$ such that the associated norm is given by

$$\|(x, y)\| = \big(|x|^p + |y|^p\big)^{1/p}$$

for all $(x, y) \in \mathbb{R}^2$ if and only if $p = 2$.

Solution. Write $N_p(x, y) = (|x|^p + |y|^p)^{1/p}$ for $(x, y) \in \mathbb{R}^2$; this is a well-defined nonnegative real number for every $p > 0$. We must show that N_p is the norm associated with some inner product on $\mathbb{R}^2$ exactly when $p = 2$.
First suppose $p = 2$. Let $\langle \cdot, \cdot \rangle$ be the Euclidean inner product on $\mathbb{R}^2$, that is,

$$\big\langle (x_1, y_1), (x_2, y_2) \big\rangle = x_1 x_2 + y_1 y_2,$$

which is an inner product by 6.3(a). Its associated norm (6.7) is

$$\|(x, y)\| = \sqrt{\big\langle (x, y), (x, y) \big\rangle} = \sqrt{x^2 + y^2} = \big(|x|^2 + |y|^2\big)^{1/2} = N_2(x, y),$$

where we used $x^2 = |x|^2$ and $y^2 = |y|^2$ for real x, y. Thus for $p = 2$ an inner product with the required associated norm does exist.
Conversely, suppose $p > 0$ and there is an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^2$ whose associated norm $\|\cdot\|$ satisfies $\|(x, y)\| = N_p(x, y)$ for all $(x, y) \in \mathbb{R}^2$. Because this norm comes from an inner product, the parallelogram equality (6.21) holds for every pair of vectors.

Apply it to $u = (1, 0)$ and $v = (0, 1)$. We have

$$\|u\| = (1^p + 0^p)^{1/p} = 1, \qquad \|v\| = (0^p + 1^p)^{1/p} = 1,$$

and, since $u + v = (1, 1)$ and $u - v = (1, -1)$,

$$\|u + v\| = (1 + 1)^{1/p} = 2^{1/p}, \qquad \|u - v\| = \big(1^p + |-1|^p\big)^{1/p} = 2^{1/p}.$$

The parallelogram equality now reads

$$2^{2/p} + 2^{2/p} = 2\big(1^2 + 1^2\big),$$

that is,

$$2^{1+2/p} = 4 = 2^2.$$

Since the function $t \mapsto 2^t$ is injective on $\mathbb{R}$, this forces

$$1 + \frac{2}{p} = 2, \qquad \text{so} \qquad \frac{2}{p} = 1, \qquad \text{so} \qquad p = 2.$$

Combining the two directions: an inner product on $\mathbb{R}^2$ with associated norm N_p exists if and only if $p = 2$.

Problem 6A.26. Suppose V is a real inner product space. Prove that

$$\langle u, v\rangle = \frac{\|u + v\|^2 - \|u - v\|^2}{4}$$

for all $u, v \in V$.

Solution. Suppose $u, v \in V$. Because V is a real inner product space, the scalars are real and conjugation is the identity, so conjugate symmetry (6.2) says simply $\langle v, u\rangle = \langle u, v\rangle$.
Expanding using additivity and homogeneity in the first slot together with 6.6(d) and 6.6(e), we get

$$\begin{aligned}\|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + \|v\|^2 + 2\langle u, v\rangle.\end{aligned}$$

Replacing v by $-v$ (and using $\langle u, -v\rangle = -\langle u, v\rangle$, $\|-v\| = \|v\|$) gives

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\langle u, v\rangle.$$

Subtracting the second identity from the first, the terms $\|u\|^2 + \|v\|^2$ cancel and we obtain

$$\|u + v\|^2 - \|u - v\|^2 = 4\langle u, v\rangle.$$

Dividing by 4 yields

$$\langle u, v\rangle = \frac{\|u + v\|^2 - \|u - v\|^2}{4},$$

as desired.
This identity is called the polarization identity for real inner product spaces. It shows that a real inner product is completely determined by its associated norm.

Problem 6A.27. Suppose V is a complex inner product space. Prove that

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4}$$

for all $u, v \in V$.

Solution. Suppose $u, v \in V$. As in the proof of the triangle inequality (6.17), for all $u, w \in V$ we have

$$\begin{aligned}\|u+w\|^2 &= \langle u+w, u+w\rangle \\ &= \langle u,u\rangle + \langle w,w\rangle + \langle u,w\rangle + \langle w,u\rangle \\ &= \|u\|^2 + \|w\|^2 + \langle u,w\rangle + \overline{\langle u,w\rangle} \\ &= \|u\|^2 + \|w\|^2 + 2\operatorname{Re}\langle u,w\rangle,\end{aligned}$$

using conjugate symmetry and the fact that $z + \overline{z} = 2\operatorname{Re} z$ for $z \in \mathbb{C}$.
Taking $w = v$ and $w = -v$ (note $\|-v\| = \|v\|$ and $\langle u, -v\rangle = -\langle u, v\rangle$ by 6.6(e)) gives

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 + 2\operatorname{Re}\langle u,v\rangle, \qquad \|u-v\|^2 = \|u\|^2 + \|v\|^2 - 2\operatorname{Re}\langle u,v\rangle,$$

and therefore

$$\|u+v\|^2 - \|u-v\|^2 = 4\operatorname{Re}\langle u,v\rangle.$$

Next take $w = iv$. By 6.6(e), $\langle u, iv\rangle = \overline{i}\,\langle u,v\rangle = -i\langle u,v\rangle$. Writing $\langle u,v\rangle = a+bi$ with $a, b \in \mathbb{R}$, we get

$$\langle u, iv\rangle = -i(a+bi) = b - ai, \qquad \text{so} \qquad \operatorname{Re}\langle u, iv\rangle = b = \operatorname{Im}\langle u,v\rangle.$$

Also $\|iv\| = \sqrt{\langle iv, iv\rangle} = \sqrt{i\overline{i}\,\langle v,v\rangle} = \|v\|$, and likewise $\|-iv\| = \|v\|$. Hence, applying the displayed expansion with $w = iv$ and with $w = -iv$,

$$\|u+iv\|^2 = \|u\|^2 + \|v\|^2 + 2\operatorname{Im}\langle u,v\rangle, \qquad \|u-iv\|^2 = \|u\|^2 + \|v\|^2 - 2\operatorname{Im}\langle u,v\rangle,$$

and therefore

$$\|u+iv\|^2 - \|u-iv\|^2 = 4\operatorname{Im}\langle u,v\rangle.$$

Putting the two computations together, the numerator on the right side of the asserted formula equals

$$\big(\|u+v\|^2 - \|u-v\|^2\big) + \big(\|u+iv\|^2 - \|u-iv\|^2\big)i = 4\operatorname{Re}\langle u,v\rangle + 4\big(\operatorname{Im}\langle u,v\rangle\big)i = 4\langle u,v\rangle.$$

Dividing by 4 gives

$$\langle u,v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4},$$

as desired.
This is the polarization identity for complex inner product spaces; as in the real case (Exercise 26) it shows that the inner product is determined by the norm it induces.

Problem 6A.28. A norm on a vector space U is a function

$$\|\cdot\| : U \to [0, \infty)$$

such that $\|u\| = 0$ if and only if $u = 0$, $\|\alpha u\| = |\alpha|\|u\|$ for all $\alpha \in \mathbb{F}$ and all $u \in U$, and $\|u+v\| \le \|u\| + \|v\|$ for all $u, v \in U$. Prove that a norm satisfying the parallelogram equality comes from an inner product (in other words, show that if $\|\cdot\|$ is a norm on U satisfying the parallelogram equality, then there is an inner product $\langle\cdot,\cdot\rangle$ on U such that $\|u\| = \langle u,u\rangle^{1/2}$ for all $u \in U$).

Solution. Throughout, $\|\cdot\|$ is a norm on the vector space U over $\mathbb{F}$ satisfying

$$\|u+v\|^2+\|u-v\|^2=2\big(\|u\|^2+\|v\|^2\big)\qquad\text{for all } u,v\in U. \tag{P}$$

Note first two consequences of the norm axioms that will be used repeatedly. Taking $\alpha=-1$ in the homogeneity axiom gives $\|-u\|=\|u\|$. And from $\|a\|\le\|a-b\|+\|b\|$ and the same inequality with a,b interchanged we get the reverse triangle inequality

$$\big|\,\|a\|-\|b\|\,\big|\le\|a-b\|\qquad\text{for all } a,b\in U. \tag{R}$$

Guided by Exercises 26 and 27, define $\varphi\colon U\times U\to\mathbb{R}$ by

$$\varphi(u,v)=\frac{\|u+v\|^2-\|u-v\|^2}{4}.$$

We first prove that if $\mathbb{F}=\mathbb{R}$ then φ is an inner product on U with $\varphi(u,u)=\|u\|^2$; the complex case is then built from this.
Step 1: symmetry, and $\varphi(u,u)=\|u\|^2$.
Since $\|v-u\|=\|-(u-v)\|=\|u-v\|$, we have $\varphi(v,u)=\varphi(u,v)$ for all $u,v\in U$. Also

$$\varphi(u,u)=\frac{\|2u\|^2-\|0\|^2}{4}=\frac{4\|u\|^2}{4}=\|u\|^2\ge 0,$$

using homogeneity. So φ satisfies positivity, and it satisfies definiteness because $\varphi(u,u)=\|u\|^2=0$ if and only if $u=0$. Moreover $\varphi(u,0)=\varphi(0,u)=(\|u\|^2-\|u\|^2)/4=0$, and

$$\varphi(-u,v)=\frac{\|-u+v\|^2-\|-u-v\|^2}{4}=\frac{\|u-v\|^2-\|u+v\|^2}{4}=-\varphi(u,v). \tag{1}$$

Step 2: additivity in the first slot.
Let $u,v,w\in U$. Applying (P) to the pair $u+w$, $v+w$, whose sum is $u+v+2w$ and whose difference is $u-v$, gives

$$\|u+v+2w\|^2+\|u-v\|^2=2\big(\|u+w\|^2+\|v+w\|^2\big).$$

Applying (P) to the pair $u-w$, $v-w$, whose sum is $u+v-2w$ and whose difference is again $u-v$, gives

$$\|u+v-2w\|^2+\|u-v\|^2=2\big(\|u-w\|^2+\|v-w\|^2\big).$$

Subtracting the second equation from the first, the terms $\|u-v\|^2$ cancel and we obtain

$$\|u+v+2w\|^2-\|u+v-2w\|^2=2\Big(\big(\|u+w\|^2-\|u-w\|^2\big)+\big(\|v+w\|^2-\|v-w\|^2\big)\Big).$$

By the definition of φ, the left side equals $4\varphi(u+v,2w)$ and the right side equals $2\big(4\varphi(u,w)+4\varphi(v,w)\big)$. Hence

$$\varphi(u+v,2w)=2\big(\varphi(u,w)+\varphi(v,w)\big)\qquad\text{for all } u,v,w\in U. \tag{2}$$

Now set $v=0$ in (2). Since $\varphi(0,w)=0$, we get

$$\varphi(u,2w)=2\varphi(u,w)\qquad\text{for all } u,w\in U. \tag{3}$$

Applying (3) with u replaced by $u+v$ and comparing with (2) gives

$$2\varphi(u+v,w)=2\big(\varphi(u,w)+\varphi(v,w)\big),$$

so

$$\varphi(u+v,w)=\varphi(u,w)+\varphi(v,w)\qquad\text{for all } u,v,w\in U. \tag{4}$$

Thus φ is additive in the first slot, and by symmetry also in the second slot.
Step 3: real homogeneity in the first slot.

Fix $v \in U$. We first show that

$$\varphi(kw, v) = k\,\varphi(w, v) \qquad \text{for all } w \in U \text{ and all } k \in \mathbb{Z}. \tag{I}$$

For k a positive integer this follows from (4) by induction on k: the case $k = 1$ is trivial, and $\varphi(kw, v) = \varphi((k-1)w + w, v) = \varphi((k-1)w, v) + \varphi(w, v)$. The case $k = 0$ is $\varphi(0, v) = 0$ from Step 1, and the case of negative k follows from the positive case together with (1).
Now let $m \in \mathbb{Z}$ and let n be a positive integer, and fix $u \in U$. Applying (I) with $w = \frac{1}{n}u$ and $k = n$ gives $\varphi(u, v) = n\,\varphi(\frac{1}{n}u, v)$, so $\varphi(\frac{1}{n}u, v) = \frac{1}{n}\varphi(u, v)$. Applying (I) with $w = \frac{1}{n}u$ and $k = m$ then gives

$$\varphi\Big(\tfrac{m}{n}u, v\Big) = m\,\varphi\Big(\tfrac{1}{n}u, v\Big) = \frac{m}{n}\varphi(u, v).$$

Thus, setting $f(t) = \varphi(tu, v)$ for $t \in \mathbb{R}$, we have shown $f(t) = t\,\varphi(u, v)$ for every rational t.
Next, f is continuous. Indeed, by (R), for $s, t \in \mathbb{R}$,

$$\big|\,\|tu \pm v\| - \|su \pm v\|\,\big| \le \|(t-s)u\| = |t-s|\,\|u\|,$$

so $t \mapsto \|tu + v\|$ and $t \mapsto \|tu - v\|$ are continuous, hence so are their squares, hence so is f. Since f and the continuous function $t \mapsto t\,\varphi(u, v)$ agree on the dense subset $\mathbb{Q}$ of $\mathbb{R}$, they agree everywhere:

$$\varphi(tu, v) = t\,\varphi(u, v) \qquad \text{for all } t \in \mathbb{R} \text{ and all } u, v \in U. \tag{5}$$

If $\mathbb{F} = \mathbb{R}$, then Steps 1, 2, 3 show that φ satisfies positivity, definiteness, additivity in the first slot, homogeneity in the first slot, and (conjugate) symmetry, so φ is an inner product on U; and $\varphi(u, u)^{1/2} = \|u\|$ by Step 1. This finishes the real case.
Step 4: the complex case.
Now suppose $\mathbb{F} = \mathbb{C}$. Let $U_{\mathbb{R}}$ denote U regarded as a real vector space, that is, with scalar multiplication restricted to $\mathbb{R}$. Then $\|\cdot\|$ is still a norm on $U_{\mathbb{R}}$ (homogeneity for real scalars is a special case of homogeneity for complex scalars) and (P) still holds. Hence by Steps 1, 2, 3 the function φ is symmetric, is additive in each slot, satisfies (5), and satisfies $\varphi(u, u) = \|u\|^2$. In particular φ is $\mathbb{R}$-bilinear.
Define

$$\langle u, v\rangle = \varphi(u, v) + \varphi(u, iv)\,i \qquad \text{for } u, v \in U.$$

Since $\varphi(u, iv) = \big(\|u + iv\|^2 - \|u - iv\|^2\big)/4$, this is precisely the expression in Exercise 27:

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4}.$$

We verify the five axioms of 6.2.
First, an observation. For every $u, v \in U$,

$$\varphi(iu, iv) = \frac{\|iu + iv\|^2 - \|iu - iv\|^2}{4} = \frac{|i|^2\|u+v\|^2 - |i|^2\|u-v\|^2}{4} = \varphi(u, v), \tag{6}$$

using complex homogeneity of the norm. Consequently

$$\varphi(v, iu) = \varphi(iv, i \cdot iu) = \varphi(iv, -u) = -\varphi(iv, u) = -\varphi(u, iv), \tag{7}$$

where the first equality is (6) applied to the pair v, iu, the second is $i \cdot iu = -u$, the third is (5) with $t = -1$ in the second slot (legitimate because φ is symmetric), and the fourth is symmetry of φ.
Positivity and definiteness. By homogeneity of the norm, $\|u + iu\| = |1 + i|\,\|u\| = \sqrt{2}\,\|u\|$ and $\|u - iu\| = |1 - i|\,\|u\| = \sqrt{2}\,\|u\|$, so

$$\varphi(u, iu) = \frac{2\|u\|^2 - 2\|u\|^2}{4} = 0.$$

Hence $\langle u, u\rangle = \varphi(u, u) = \|u\|^2$, which is a nonnegative real number, and it equals 0 if and only if $u = 0$. This also establishes the required identity $\|u\| = \langle u, u\rangle^{1/2}$.

Additivity in the first slot. Both $u \mapsto \varphi(u, v)$ and $u \mapsto \varphi(u, iv)$ are additive by (4), so $\langle \cdot, \cdot \rangle$ is additive in the first slot.
Conjugate symmetry. Using symmetry of φ and (7),

$$\langle v, u \rangle = \varphi(v, u) + \varphi(v, iu)\, i = \varphi(u, v) - \varphi(u, iv)\, i = \overline{\varphi(u, v) + \varphi(u, iv)\, i} = \overline{\langle u, v \rangle},$$

where the middle step uses that φ takes real values.
Homogeneity in the first slot. For $t \in \mathbb{R}$, applying (5) to each of the two terms gives

$$\langle tu, v \rangle = \varphi(tu, v) + \varphi(tu, iv)\, i = t\varphi(u, v) + t\varphi(u, iv)\, i = t\langle u, v \rangle.$$

So it suffices to prove $\langle iu, v \rangle = i\langle u, v \rangle$: granting this, for $\alpha = a + bi$ with $a, b \in \mathbb{R}$ we get, using additivity in the first slot,

$$\langle \alpha u, v \rangle = \langle au, v \rangle + \langle b(iu), v \rangle = a\langle u, v \rangle + b\langle iu, v \rangle = a\langle u, v \rangle + bi\langle u, v \rangle = \alpha\langle u, v \rangle.$$

To prove $\langle iu, v \rangle = i\langle u, v \rangle$, compute using (6) and then (7) together with symmetry:

$$\langle iu, v \rangle = \varphi(iu, v) + \varphi(iu, iv)\, i = \varphi(iu, v) + \varphi(u, v)\, i,$$

while

$$i\langle u, v \rangle = \varphi(u, v)\, i + \varphi(u, iv)\, i^2 = \varphi(u, v)\, i - \varphi(u, iv).$$

These two expressions agree precisely when $\varphi(iu, v) = -\varphi(u, iv)$, which is exactly (7) combined with the symmetry $\varphi(iu, v) = \varphi(v, iu)$.
All five conditions of 6.2 hold, so $\langle \cdot, \cdot \rangle$ is an inner product on U, and we showed above that $\langle u, u \rangle^{1/2} = \|u\|$ for all $u \in U$.
Therefore, in both the real and the complex case, a norm satisfying the parallelogram equality is the norm associated with an inner product. (This result is due to Jordan and von Neumann, 1935.)

Problem 6A.29. Suppose $V_1, \ldots, V_m$ are inner product spaces. Show that the equation

$$\langle (u_1, \ldots, u_m), (v_1, \ldots, v_m) \rangle = \langle u_1, v_1 \rangle + \cdots + \langle u_m, v_m \rangle$$

defines an inner product on $V_1 \times \cdots \times V_m$.
In the expression above on the right, for each $k = 1, \ldots, m$, the inner product $\langle u_k, v_k \rangle$ denotes the inner product on V_k. Each of the spaces $V_1, \ldots, V_m$ may have a different inner product, even though the same notation is used here.

Solution. All of $V_1, \ldots, V_m$ are vector spaces over the same field $\mathbf{F}$ (this is implicit in forming the product space $V = V_1 \times \cdots \times V_m$, whose vector space structure is the coordinatewise one). Write $\langle \cdot, \cdot \rangle_k$ for the inner product on V_k, and define

$$\langle u, v \rangle = \sum_{k=1}^{m} \langle u_k, v_k \rangle_k$$

for $u = (u_1, \ldots, u_m)$ and $v = (v_1, \ldots, v_m)$ in V. Since each $\langle u_k, v_k \rangle_k$ lies in $\mathbf{F}$, so does the sum; thus this is a function from $V \times V$ to $\mathbf{F}$, as required. We check the five conditions in the definition 6.2 of an inner product.
Positivity. If $v = (v_1, \ldots, v_m) \in V$, then

$$\langle v, v \rangle = \sum_{k=1}^{m} \langle v_k, v_k \rangle_k.$$

Each summand is a nonnegative real number, by the positivity of the inner product on V_k. Hence $\langle v, v \rangle$ is a sum of nonnegative real numbers and so $\langle v, v \rangle \geq 0$.

Definiteness. Suppose $v = (v_1, \dots, v_m) \in V$. As just noted, $\langle v, v\rangle = \sum_k \langle v_k, v_k\rangle_k$ is a sum of nonnegative real numbers. A finite sum of nonnegative real numbers equals 0 if and only if every term equals 0. Thus $\langle v, v\rangle = 0$ if and only if $\langle v_k, v_k\rangle_k = 0$ for each k, which by the definiteness of the inner product on V_k holds if and only if $v_k = 0$ for each k, which holds if and only if $v = (0, \dots, 0)$, the zero vector of V.
Additivity in the first slot. Suppose $u = (u_1, \dots, u_m)$, $v = (v_1, \dots, v_m)$, and $w = (w_1, \dots, w_m)$ lie in V. Because addition in V is coordinatewise, $u + v = (u_1 + v_1, \dots, u_m + v_m)$. Hence

$$\begin{aligned}\langle u + v, w\rangle &= \sum_{k=1}^{m} \langle u_k + v_k, w_k\rangle_k \\ &= \sum_{k=1}^{m} \big(\langle u_k, w_k\rangle_k + \langle v_k, w_k\rangle_k\big) \\ &= \sum_{k=1}^{m} \langle u_k, w_k\rangle_k + \sum_{k=1}^{m} \langle v_k, w_k\rangle_k \\ &= \langle u, w\rangle + \langle v, w\rangle,\end{aligned}$$

where the second equality uses additivity in the first slot for each $\langle \cdot, \cdot\rangle_k$.
Homogeneity in the first slot. Suppose $\lambda \in \mathbf{F}$ and $u, v \in V$ are as above. Because scalar multiplication in V is coordinatewise, $\lambda u = (\lambda u_1, \dots, \lambda u_m)$, and therefore

$$\langle \lambda u, v\rangle = \sum_{k=1}^{m} \langle \lambda u_k, v_k\rangle_k = \sum_{k=1}^{m} \lambda \langle u_k, v_k\rangle_k = \lambda \sum_{k=1}^{m} \langle u_k, v_k\rangle_k = \lambda \langle u, v\rangle,$$

using homogeneity in the first slot for each $\langle \cdot, \cdot\rangle_k$.
Conjugate symmetry. Suppose $u, v \in V$ are as above. Using conjugate symmetry for each $\langle \cdot, \cdot\rangle_k$ and the fact that complex conjugation is additive,

$$\langle u, v\rangle = \sum_{k=1}^{m} \langle u_k, v_k\rangle_k = \sum_{k=1}^{m} \overline{\langle v_k, u_k\rangle_k} = \overline{\sum_{k=1}^{m} \langle v_k, u_k\rangle_k} = \overline{\langle v, u\rangle}.$$

All five conditions hold, so the displayed equation defines an inner product on $V_1 \times \cdots \times V_m$.
Note that the case $V_1 = \cdots = V_m = \mathbf{F}$, each with the inner product $\langle w, z\rangle = w\overline{z}$, recovers the Euclidean inner product on $\mathbf{F}^m$ from 6.3(a).

Problem 6A.30. Suppose V is a real inner product space. For $u, v, w, x \in V$, define

$$\langle u + iv, w + ix\rangle_{\mathbf{C}} = \langle u, w\rangle + \langle v, x\rangle + \big(\langle v, w\rangle - \langle u, x\rangle\big)i.$$

(a) Show that $\langle \cdot, \cdot\rangle_{\mathbf{C}}$ makes $V_{\mathbf{C}}$ into a complex inner product space.
(b) Show that if $u, v \in V$, then

$$\langle u, v\rangle_{\mathbf{C}} = \langle u, v\rangle \quad \text{and} \quad \|u + iv\|_{\mathbf{C}}^2 = \|u\|^2 + \|v\|^2.$$

See Exercise 8 in Section 1B for the definition of the complexification $V_{\mathbf{C}}$.

Solution. Recall from Exercise 8 in Section 1B that $V_{\mathbf{C}} = V \times V$ as a set, that the ordered pair (u, v) is written $u + iv$, that addition is

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2),$$

and that complex scalar multiplication is

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

for $a, b \in \mathbf{R}$ and $u, v \in V$. That exercise shows $V_{\mathbf{C}}$ is a complex vector space; its zero vector is $0 + i0$. Because $V_{\mathbf{C}}$ is literally the set $V \times V$, every element of $V_{\mathbf{C}}$ has exactly one representation in the form $u + iv$ with $u, v \in V$. Hence the displayed formula is a well-defined function from $V_{\mathbf{C}} \times V_{\mathbf{C}}$ to $\mathbf{C}$.
Throughout, $\langle \cdot, \cdot \rangle$ denotes the given inner product on the real inner product space V; it is symmetric, so $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$, and it is linear in each slot, by the definition 6.2 of an inner product together with 6.6.
Solution to (a). We verify the five conditions of 6.2 for $\langle \cdot, \cdot \rangle_{\mathbf{C}}$ on the complex vector space $V_{\mathbf{C}}$.
Positivity. Let $z = u + iv \in V_{\mathbf{C}}$. Taking $w = u$ and $x = v$ in the defining formula,

$$\langle z, z \rangle_{\mathbf{C}} = \langle u, u \rangle + \langle v, v \rangle + \big(\langle v, u \rangle - \langle u, v \rangle\big) i = \|u\|^2 + \|v\|^2,$$

because $\langle v, u \rangle = \langle u, v \rangle$ by the symmetry of the real inner product, so the imaginary part vanishes. Thus $\langle z, z \rangle_{\mathbf{C}}$ is a sum of two nonnegative real numbers, hence $\langle z, z \rangle_{\mathbf{C}} \geq 0$.
Definiteness. With $z = u + iv$ as above, $\langle z, z \rangle_{\mathbf{C}} = \|u\|^2 + \|v\|^2$, and this equals 0 if and only if $\|u\| = \|v\| = 0$, which by 6.9(a) holds if and only if $u = v = 0$, which holds if and only if z is the zero vector of $V_{\mathbf{C}}$.
Additivity in the first slot. Let $z_1 = u_1 + iv_1$, $z_2 = u_2 + iv_2$, and $z = w + ix$ be elements of $V_{\mathbf{C}}$. Since $z_1 + z_2 = (u_1 + u_2) + i(v_1 + v_2)$,

$$\begin{aligned}
\langle z_1 + z_2, z \rangle_{\mathbf{C}} &= \langle u_1 + u_2, w \rangle + \langle v_1 + v_2, x \rangle + \big(\langle v_1 + v_2, w \rangle - \langle u_1 + u_2, x \rangle\big) i \\
&= \big(\langle u_1, w \rangle + \langle v_1, x \rangle + (\langle v_1, w \rangle - \langle u_1, x \rangle) i\big) \\
&\quad + \big(\langle u_2, w \rangle + \langle v_2, x \rangle + (\langle v_2, w \rangle - \langle u_2, x \rangle) i\big) \\
&= \langle z_1, z \rangle_{\mathbf{C}} + \langle z_2, z \rangle_{\mathbf{C}},
\end{aligned}$$

where the second equality uses additivity in the first slot of $\langle \cdot, \cdot \rangle$ four times and then regroups.
Homogeneity in the first slot. First suppose $a \in \mathbf{R}$ and $z = u + iv$, $z' = w + ix$. Then $az = (au) + i(av)$, so

$$\begin{aligned}
\langle az, z' \rangle_{\mathbf{C}} &= \langle au, w \rangle + \langle av, x \rangle + \big(\langle av, w \rangle - \langle au, x \rangle\big) i \\
&= a\langle u, w \rangle + a\langle v, x \rangle + a\big(\langle v, w \rangle - \langle u, x \rangle\big) i \\
&= a\langle z, z' \rangle_{\mathbf{C}},
\end{aligned}$$

by homogeneity in the first slot of $\langle \cdot, \cdot \rangle$. Next, $iz = (-v) + iu$ (take $a = 0$ and $b = 1$ in the formula for complex scalar multiplication), so

$$\begin{aligned}
\langle iz, z' \rangle_{\mathbf{C}} &= \langle -v, w \rangle + \langle u, x \rangle + \big(\langle u, w \rangle - \langle -v, x \rangle\big) i \\
&= \big(-\langle v, w \rangle + \langle u, x \rangle\big) + \big(\langle u, w \rangle + \langle v, x \rangle\big) i \\
&= i\Big(\langle u, w \rangle + \langle v, x \rangle + \big(\langle v, w \rangle - \langle u, x \rangle\big) i\Big) \\
&= i\langle z, z' \rangle_{\mathbf{C}},
\end{aligned}$$

where the third equality holds because multiplying $\alpha + \beta i$ by i produces $-\beta + \alpha i$: here $\alpha = \langle u, w \rangle + \langle v, x \rangle$ and $\beta = \langle v, w \rangle - \langle u, x \rangle$, giving exactly the second line. Now let $\lambda = a + bi$ with $a, b \in \mathbf{R}$. Since $V_{\mathbf{C}}$ is a complex vector space, $\lambda z = az + b(iz)$. Using the additivity already proved and then the two homogeneity computations,

$$\langle \lambda z, z' \rangle_{\mathbf{C}} = \langle az, z' \rangle_{\mathbf{C}} + \langle b(iz), z' \rangle_{\mathbf{C}} = a\langle z, z' \rangle_{\mathbf{C}} + b\, i\langle z, z' \rangle_{\mathbf{C}} = \lambda \langle z, z' \rangle_{\mathbf{C}}.$$

Conjugate symmetry. Let $z = u + iv$ and $z' = w + ix$. Then

$$\begin{aligned}
\overline{\langle z', z \rangle_{\mathbf{C}}} &= \overline{\langle w, u \rangle + \langle x, v \rangle + \big(\langle x, u \rangle - \langle w, v \rangle\big) i} \\
&= \langle w, u \rangle + \langle x, v \rangle - \big(\langle x, u \rangle - \langle w, v \rangle\big) i \\
&= \langle u, w \rangle + \langle v, x \rangle + \big(\langle v, w \rangle - \langle u, x \rangle\big) i \\
&= \langle z, z' \rangle_{\mathbf{C}},
\end{aligned}$$

where the third equality uses the symmetry of the real inner product in all four terms.

All five conditions hold, so $\langle\cdot,\cdot\rangle_{\mathbf{C}}$ is an inner product on the complex vector space $V_{\mathbf{C}}$, making $V_{\mathbf{C}}$ into a complex inner product space.
Solution to (b). We identify $u \in V$ with $u + i0 \in V_{\mathbf{C}}$, as in Exercise 8 in Section 1B. Thus for $u, v \in V$, applying the defining formula to the four vectors $u, 0, v, 0$ and using $\langle 0, y\rangle = \langle y, 0\rangle = 0$ from 6.6(b) and 6.6(c),

$$\langle u, v\rangle_{\mathbf{C}} = \langle u + i0, v + i0\rangle_{\mathbf{C}} = \langle u, v\rangle + \langle 0, 0\rangle + \big(\langle 0, v\rangle - \langle u, 0\rangle\big)i = \langle u, v\rangle.$$

So the inner product on $V_{\mathbf{C}}$ restricts on V to the original inner product; in particular the norm on $V_{\mathbf{C}}$ restricts to the original norm on V.
For the second assertion, we computed in the proof of positivity above that

$$\|u + iv\|_{\mathbf{C}}^2 = \langle u + iv, u + iv\rangle_{\mathbf{C}} = \langle u, u\rangle + \langle v, v\rangle = \|u\|^2 + \|v\|^2$$

for all $u, v \in V$, as desired. This is the expected generalization of $|a + bi|^2 = a^2 + b^2$ for $a, b \in \mathbf{R}$. It is not an instance of the Pythagorean theorem 6.12: the defining formula gives $\langle u, iv\rangle_{\mathbf{C}} = -\langle u, v\rangle i$, which is generally nonzero, so u and iv need not be orthogonal in $V_{\mathbf{C}}$. What happens is that this cross term is purely imaginary, so when $\|u+iv\|_{\mathbf{C}}^2$ is expanded the cross terms $\langle u, iv\rangle_{\mathbf{C}}$ and $\langle iv, u\rangle_{\mathbf{C}}$ are conjugates of each other with zero real part and hence cancel.

Problem 6A.31. Suppose $u, v, w \in V$. Prove that

$$\Big\|w - \tfrac{1}{2}(u + v)\Big\|^2 = \frac{\|w - u\|^2 + \|w - v\|^2}{2} - \frac{\|u - v\|^2}{4}.$$

Solution. Set

$$a = w - u \quad \text{and} \quad b = w - v.$$

Then

$$a + b = 2w - u - v = 2\Big(w - \tfrac{1}{2}(u + v)\Big) \quad \text{and} \quad a - b = (w - u) - (w - v) = v - u.$$

By the homogeneity of the norm (6.9(b), which says $\|\lambda x\| = |\lambda|\,\|x\|$ for $\lambda \in \mathbf{F}$ and $x \in V$),

$$\|a + b\|^2 = \Big\|2\big(w - \tfrac{1}{2}(u + v)\big)\Big\|^2 = 4\Big\|w - \tfrac{1}{2}(u + v)\Big\|^2$$

and

$$\|a - b\|^2 = \|v - u\|^2 = \big|-1\big|^2\|u - v\|^2 = \|u - v\|^2.$$

Now apply the parallelogram equality 6.21 to the vectors a and b:

$$\|a + b\|^2 + \|a - b\|^2 = 2\big(\|a\|^2 + \|b\|^2\big).$$

Substituting the four expressions computed above gives

$$4\Big\|w - \tfrac{1}{2}(u + v)\Big\|^2 + \|u - v\|^2 = 2\big(\|w - u\|^2 + \|w - v\|^2\big).$$

Solving for the first term,

$$4\Big\|w - \tfrac{1}{2}(u + v)\Big\|^2 = 2\big(\|w - u\|^2 + \|w - v\|^2\big) - \|u - v\|^2,$$

and dividing both sides by 4 yields

$$\Big\|w - \tfrac{1}{2}(u + v)\Big\|^2 = \frac{\|w - u\|^2 + \|w - v\|^2}{2} - \frac{\|u - v\|^2}{4},$$

as desired.
Geometrically this is the median length formula: the point $\frac{1}{2}(u + v)$ is the midpoint of the segment joining u and v, and the identity expresses the squared distance from w to that midpoint in terms of the squared distances from w to the two endpoints and the squared length of the segment. Note in particular that the right side is always at most $\frac{1}{2}(\|w - u\|^2 + \|w - v\|^2)$, with equality exactly when $u = v$; this observation is what drives Exercise 32.

Problem 6A.32. Suppose that E is a subset of V with the property that $u, v \in E$ implies $\frac{1}{2}(u+v) \in E$. Let $w \in V$. Show that there is at most one point in E that is closest to w. In other words, show that there is at most one $u \in E$ such that

$$\|w - u\| \leq \|w - x\|$$

for all $x \in E$.

Solution. Suppose $u_1, u_2 \in E$ both have the stated property, so that

$$\|w - u_1\| \leq \|w - x\| \quad \text{and} \quad \|w - u_2\| \leq \|w - x\|$$

for every $x \in E$. We must show $u_1 = u_2$.
First, the two minimizing distances agree. Taking $x = u_2$ in the first inequality gives $\|w - u_1\| \leq \|w - u_2\|$, and taking $x = u_1$ in the second gives $\|w - u_2\| \leq \|w - u_1\|$. Hence

$$\|w - u_1\| = \|w - u_2\|.$$

Call this common value d; note $d \geq 0$.
Second, the midpoint is a competitor. Since $u_1, u_2 \in E$, the hypothesis on E gives $\frac{1}{2}(u_1 + u_2) \in E$. Applying the minimizing property of u_1 with $x = \frac{1}{2}(u_1 + u_2)$ therefore yields

$$d = \|w - u_1\| \leq \left\|w - \tfrac{1}{2}(u_1 + u_2)\right\|,$$

and hence, since both sides are nonnegative,

$$d^2 \leq \left\|w - \tfrac{1}{2}(u_1 + u_2)\right\|^2.$$

Third, we evaluate the right side using Exercise 31 (applied with $u = u_1$ and $v = u_2$):

$$\left\|w - \tfrac{1}{2}(u_1+u_2)\right\|^2 = \frac{\|w - u_1\|^2 + \|w - u_2\|^2}{2} - \frac{\|u_1 - u_2\|^2}{4} = \frac{d^2 + d^2}{2} - \frac{\|u_1 - u_2\|^2}{4} = d^2 - \frac{\|u_1 - u_2\|^2}{4}.$$

Combining the last two displays,

$$d^2 \leq d^2 - \frac{\|u_1 - u_2\|^2}{4},$$

which forces

$$\frac{\|u_1 - u_2\|^2}{4} \leq 0,$$

and hence $\|u_1 - u_2\|^2 \leq 0$. Since a norm is always nonnegative, $\|u_1 - u_2\| = 0$, so $u_1 - u_2 = 0$ by 6.9(a). Therefore $u_1 = u_2$.
Thus there is at most one point of E closest to w.
Two remarks. First, the hypothesis is genuinely weaker than convexity: E need only be closed under taking midpoints, not under all convex combinations, and indeed the proof used only the single midpoint $\frac{1}{2}(u_1 + u_2)$. Second, the conclusion asserts uniqueness only; existence can fail. For example, in $V = \mathbf{R}$ with the Euclidean inner product, take $E = (0, \infty)$, which is closed under midpoints, and $w = 0$: no point of E is closest to w, since $\inf_{x \in E} |0 - x| = 0$ is not attained. The parallelogram equality (through Exercise 31) is what makes the argument work; in a normed space whose norm does not come from an inner product, closest points to a midpoint-closed set need not be unique.

Problem 6A.33. Suppose f, g are differentiable functions from $\mathbf{R}$ to $\mathbf{R}^n$.
(a) Show that

$$\langle f(t), g(t) \rangle' = \langle f'(t), g(t) \rangle + \langle f(t), g'(t) \rangle.$$

(b) Suppose c is a positive number and $\|f(t)\| = c$ for every $t \in \mathbf{R}$. Show that $\langle f'(t), f(t)\rangle = 0$ for every $t \in \mathbf{R}$.
(c) Interpret the result in (b) geometrically in terms of the tangent vector to a curve lying on a sphere in $\mathbf{R}^n$ centered at the origin.
A function $f\colon \mathbf{R} \to \mathbf{R}^n$ is called differentiable if there exist differentiable functions $f_1, \ldots, f_n$ from $\mathbf{R}$ to $\mathbf{R}$ such that $f(t) = (f_1(t), \ldots, f_n(t))$ for each $t \in \mathbf{R}$. Furthermore, for each $t \in \mathbf{R}$, the derivative $f'(t) \in \mathbf{R}^n$ is defined by $f'(t) = (f_1{}'(t), \ldots, f_n{}'(t))$.

Solution. Here $\mathbf{R}^n$ carries the Euclidean inner product of 6.3(a), which on the real vector space $\mathbf{R}^n$ is

$$\langle (y_1, \ldots, y_n), (z_1, \ldots, z_n)\rangle = y_1 z_1 + \cdots + y_n z_n$$

(no complex conjugates appear, since all entries are real).
Solution to (a). Write $f(t) = (f_1(t), \ldots, f_n(t))$ and $g(t) = (g_1(t), \ldots, g_n(t))$ with each f_k and each g_k a differentiable function from $\mathbf{R}$ to $\mathbf{R}$. Then the function $t \mapsto \langle f(t), g(t)\rangle$ from $\mathbf{R}$ to $\mathbf{R}$ is given by

$$\langle f(t), g(t)\rangle = \sum_{k=1}^{n} f_k(t)\, g_k(t).$$

Each summand $f_k g_k$ is a product of two differentiable real-valued functions of a real variable, hence is differentiable, with $(f_k g_k)' = f_k{}' g_k + f_k g_k{}'$ by the usual product rule from one-variable calculus. A finite sum of differentiable functions is differentiable, with derivative the sum of the derivatives. Therefore $t \mapsto \langle f(t), g(t)\rangle$ is differentiable and

$$\begin{aligned}
\langle f(t), g(t)\rangle' &= \sum_{k=1}^{n} \big(f_k{}'(t)\, g_k(t) + f_k(t)\, g_k{}'(t)\big) \\
&= \sum_{k=1}^{n} f_k{}'(t)\, g_k(t) + \sum_{k=1}^{n} f_k(t)\, g_k{}'(t) \\
&= \langle f'(t), g(t)\rangle + \langle f(t), g'(t)\rangle,
\end{aligned}$$

where the last equality uses the definitions $f'(t) = (f_1{}'(t), \ldots, f_n{}'(t))$ and $g'(t) = (g_1{}'(t), \ldots, g_n{}'(t))$ together with the formula for the Euclidean inner product.
Solution to (b). Suppose $\|f(t)\| = c$ for every $t \in \mathbf{R}$. By the definition 6.7 of the norm, this means

$$\langle f(t), f(t)\rangle = \|f(t)\|^2 = c^2$$

for every $t \in \mathbf{R}$. Thus $t \mapsto \langle f(t), f(t)\rangle$ is a constant function on $\mathbf{R}$, and so its derivative is 0 at every t. On the other hand, applying (a) with $g = f$ gives

$$\langle f(t), f(t)\rangle' = \langle f'(t), f(t)\rangle + \langle f(t), f'(t)\rangle = 2\langle f'(t), f(t)\rangle,$$

where we used the symmetry $\langle f(t), f'(t)\rangle = \langle f'(t), f(t)\rangle$ of the inner product on the real inner product space $\mathbf{R}^n$. Combining the two computations gives $2\langle f'(t), f(t)\rangle = 0$, hence

$$\langle f'(t), f(t)\rangle = 0$$

for every $t \in \mathbf{R}$, as desired. (The hypothesis that c is positive was not needed; all that mattered is that $\|f(t)\|$ is constant. The positivity merely rules out the degenerate case $f \equiv 0$, where the conclusion is trivial.)
Solution to (c). The set $\{x \in \mathbf{R}^n : \|x\| = c\}$ is the sphere of radius c in $\mathbf{R}^n$ centered at the origin. The hypothesis in (b) says exactly that the curve traced out by f lies on this sphere. The vector $f(t)$ is the position vector of the point $f(t)$ on the curve, that is, the radius vector from the center of the sphere to that point, and $f'(t)$ is the tangent vector to the curve at that point. So (b) says

$$\langle f'(t), f(t)\rangle = 0,$$

which by the definition 6.10 of orthogonality means the tangent vector to the curve at $f(t)$ is orthogonal to the radius vector at $f(t)$.
In words: a curve that stays on a sphere centered at the origin can move only perpendicularly to the radius. Any component of velocity along the radius would change the distance to the center and hence push the curve off the sphere. Equivalently, at each point of the sphere the tangent vector lies in the tangent plane (more precisely, the tangent hyperplane) to the sphere at that point, which is precisely the set of vectors orthogonal to the radius vector there. This generalizes the familiar plane fact that a circle's tangent line at a point is perpendicular to the radius drawn to that point.

Problem 6A.34. Use inner products to prove Apollonius's identity: In a triangle with sides of length a, b, and c, let d be the length of the line segment from the midpoint of the side of length c to the opposite vertex. Then
$$a^2 + b^2 = \tfrac{1}{2}c^2 + 2d^2.$$

Solution. Place the triangle in $\mathbf{R}^2$ with the Euclidean inner product of 6.3(a); the argument below in fact works verbatim in any real inner product space, so it applies to a triangle sitting in $\mathbf{R}^n$ for any n. Label the vertices u, v, and w so that the side of length c is the segment joining u and v, and w is the vertex opposite that side. The lengths of the three sides are the distances between the pairs of vertices, so
$$c = \|u - v\|, \qquad a = \|w - v\|, \qquad b = \|w - u\|,$$
after naming the two remaining sides a and b in this way (the identity to be proved is symmetric in a and b, so this choice loses no generality).
The midpoint of the segment joining u and v is the point $\frac{1}{2}(u+v)$: indeed it lies on that segment, and its distances to the two endpoints are
$$\left\|\tfrac{1}{2}(u+v) - u\right\| = \left\|\tfrac{1}{2}(v-u)\right\| = \tfrac{1}{2}\|u - v\| = \left\|\tfrac{1}{2}(u+v) - v\right\|,$$
using 6.9(b). The segment described in the statement therefore joins $\frac{1}{2}(u+v)$ to the opposite vertex w, so its length is
$$d = \left\|w - \tfrac{1}{2}(u+v)\right\|.$$
Now apply Exercise 31 to the vectors u, v, w:
$$\left\|w - \tfrac{1}{2}(u+v)\right\|^2 = \frac{\|w-u\|^2 + \|w-v\|^2}{2} - \frac{\|u-v\|^2}{4}.$$
Substituting the labels above, this reads
$$d^2 = \frac{b^2 + a^2}{2} - \frac{c^2}{4}.$$
Multiplying both sides by 2 gives
$$2d^2 = a^2 + b^2 - \frac{c^2}{2},$$
and rearranging yields
$$a^2 + b^2 = \tfrac{1}{2}c^2 + 2d^2,$$
which is Apollonius's identity.
For completeness, here is the underlying inner product computation without citing Exercise 31. Translating the whole triangle does not change any of the four lengths, so we may assume the midpoint $\frac{1}{2}(u+v)$ is the origin, that is, $v = -u$. Then $c = \|u - v\| = \|2u\| = 2\|u\|$, $b = \|w - u\|$, $a = \|w + u\|$, and $d = \|w\|$. Expanding,
$$a^2 + b^2 = \|w+u\|^2 + \|w-u\|^2 = 2\big(\|w\|^2 + \|u\|^2\big) = 2d^2 + 2\Big(\frac{c}{2}\Big)^2 = 2d^2 + \tfrac{1}{2}c^2,$$

where the middle equality is the parallelogram equality 6.21. This is the same identity.
As a sanity check, consider the isosceles case $a = b$. The identity then reads $2a^2 = \frac{1}{2}c^2 + 2d^2$, that is, $a^2 = (c/2)^2 + d^2$. This agrees with the Pythagorean theorem 6.12 applied to the right triangle with legs of length $c/2$ and d and hypotenuse of length a, which is exactly the configuration one gets in an isosceles triangle, where the median to the base is perpendicular to the base.

Problem 6A.35. Fix a positive integer n. The Laplacian Δp of a twice differentiable real-valued function p on $\mathbf{R}^n$ is the function on $\mathbf{R}^n$ defined by

$$\Delta p = \frac{\partial^2 p}{\partial x_1^2} + \cdots + \frac{\partial^2 p}{\partial x_n^2}.$$

The function p is called harmonic if $\Delta p = 0$.
A polynomial on $\mathbf{R}^n$ is a linear combination (with coefficients in $\mathbf{R}$) of functions of the form $x_1^{m_1} \cdots x_n^{m_n}$, where $m_1, \ldots, m_n$ are nonnegative integers.
Suppose q is a polynomial on $\mathbf{R}^n$. Prove that there exists a harmonic polynomial p on $\mathbf{R}^n$ such that $p(x) = q(x)$ for every $x \in \mathbf{R}^n$ with $\|x\| = 1$.
The only fact about harmonic functions that you need for this exercise is that if p is a harmonic function on $\mathbf{R}^n$ and $p(x) = 0$ for all $x \in \mathbf{R}^n$ with $\|x\| = 1$, then $p = 0$.
Hint: A reasonable guess is that the desired harmonic polynomial p is of the form $q + (1 - \|x\|^2)r$ for some polynomial r. Prove that there is a polynomial r on $\mathbf{R}^n$ such that $q + (1 - \|x\|^2)r$ is harmonic by defining an operator T on a suitable vector space by

$$Tr = \Delta\big((1 - \|x\|^2)r\big)$$

and then showing that T is injective and hence surjective.

Solution. Throughout, $\mathbf{R}^n$ has the Euclidean inner product, so $\|x\|^2 = x_1^2 + \cdots + x_n^2$ for $x = (x_1, \ldots, x_n)$; in particular the function $x \mapsto 1 - \|x\|^2$ is itself a polynomial on $\mathbf{R}^n$, namely $1 - x_1^2 - \cdots - x_n^2$. Write $\mathcal{P}$ for the real vector space of all polynomials on $\mathbf{R}^n$ (a subspace of the vector space of all real-valued functions on $\mathbf{R}^n$).
Step 1: a finite-dimensional space to work in. For each integer $k \geq 0$ let

$$\mathcal{P}_k = \operatorname{span}\big\{x_1^{m_1} \cdots x_n^{m_n} : m_1, \ldots, m_n \text{ are nonnegative integers with } m_1 + \cdots + m_n \leq k\big\},$$

and set $\mathcal{P}_k = \{0\}$ for $k < 0$. Since there are only finitely many n-tuples of nonnegative integers with sum at most k, the space $\mathcal{P}_k$ is spanned by a finite list and hence is finite-dimensional. Every polynomial on $\mathbf{R}^n$ is by definition a linear combination of finitely many monomials, so every polynomial lies in $\mathcal{P}_k$ for all sufficiently large k. Fix $m \geq 0$ with $q \in \mathcal{P}_m$.
Step 2: how Δ and multiplication by $1 - \|x\|^2$ move degrees. Both operations are clearly linear on $\mathcal{P}$ (differentiation and multiplication by a fixed function are linear). We record their effect on the spaces $\mathcal{P}_k$.
First, Δ maps $\mathcal{P}_k$ into $\mathcal{P}_{k-2}$. It suffices to check this on a spanning monomial $x_1^{m_1} \cdots x_n^{m_n}$ with $m_1 + \cdots + m_n \leq k$, since Δ is linear. For each j,

$$\frac{\partial^2}{\partial x_j^2}\big(x_1^{m_1} \cdots x_n^{m_n}\big) = m_j(m_j - 1)\, x_1^{m_1} \cdots x_j^{m_j - 2} \cdots x_n^{m_n},$$

which is 0 if $m_j \leq 1$ and otherwise is a scalar times a monomial whose exponents sum to $m_1 + \cdots + m_n - 2 \leq k - 2$. Summing over j shows $\Delta\big(x_1^{m_1} \cdots x_n^{m_n}\big) \in \mathcal{P}_{k-2}$.
Second, multiplication by $1 - \|x\|^2$ maps $\mathcal{P}_k$ into $\mathcal{P}_{k+2}$, because multiplying a monomial of total degree at most k by 1 or by $-x_j^2$ produces a monomial of total degree at most $k + 2$.
Step 3: the operator T. Define $T\colon \mathcal{P}_{m-2} \to \mathcal{P}_{m-2}$ by

$$Tr = \Delta\big((1 - \|x\|^2)r\big).$$

By Step 2, if $r \in \mathcal{P}_{m-2}$ then $(1-\|x\|^2)r \in \mathcal{P}_m$ and hence $\Delta\big((1-\|x\|^2)r\big) \in \mathcal{P}_{m-2}$. So T does map $\mathcal{P}_{m-2}$ into itself, and T is linear because it is the composition of the two linear maps of Step 2. Thus T is an operator on the finite-dimensional vector space $\mathcal{P}_{m-2}$.
Step 4: T is injective. Suppose $r \in \mathcal{P}_{m-2}$ and $Tr = 0$. Put

$$h(x) = (1-\|x\|^2)\, r(x).$$

Then h is a polynomial on $\mathbf{R}^n$, hence infinitely differentiable, and $\Delta h = Tr = 0$, so h is harmonic. Moreover, if $x \in \mathbf{R}^n$ satisfies $\|x\| = 1$, then $1-\|x\|^2 = 0$ and hence $h(x) = 0$. By the fact quoted in the statement of the exercise, $h = 0$; that is,

$$(1-\|x\|^2)\, r(x) = 0 \quad \text{for every } x \in \mathbf{R}^n.$$

For every x with $\|x\| \neq 1$ the factor $1-\|x\|^2$ is a nonzero real number, so $r(x) = 0$ for all such x. Thus r vanishes on the complement of the unit sphere. Since r is a polynomial it is continuous, and every point of the unit sphere is a limit of points not on the unit sphere (for x with $\|x\| = 1$ we have $(1+1/k)x \to x$ and $\|(1+1/k)x\| = 1 + 1/k \neq 1$). Hence $r(x) = 0$ there too, and therefore $r = 0$ as a function on $\mathbf{R}^n$, which is the zero vector of $\mathcal{P}_{m-2}$. So $\operatorname{null} T = \{0\}$, and hence T is injective by 3.15.
Step 5: T is surjective. The operator T maps the finite-dimensional vector space $\mathcal{P}_{m-2}$ to itself, and it is injective by Step 4. Hence T is surjective by 3.65 (injectivity is equivalent to surjectivity for a linear map between finite-dimensional spaces of equal dimension).
Step 6: constructing p. Since $q \in \mathcal{P}_m$, Step 2 gives $\Delta q \in \mathcal{P}_{m-2}$, and therefore $-\Delta q \in \mathcal{P}_{m-2}$. By Step 5 there exists $r \in \mathcal{P}_{m-2}$ with

$$Tr = -\Delta q.$$

Define

$$p(x) = q(x) + (1-\|x\|^2)\, r(x).$$

Then p is a polynomial on $\mathbf{R}^n$ (a sum of q with a product of polynomials; indeed $p \in \mathcal{P}_m$). Using the linearity of Δ,

$$\Delta p = \Delta q + \Delta\big((1-\|x\|^2)r\big) = \Delta q + Tr = \Delta q - \Delta q = 0,$$

so p is harmonic. Finally, if $x \in \mathbf{R}^n$ satisfies $\|x\| = 1$, then $1-\|x\|^2 = 0$ and hence

$$p(x) = q(x) + 0 \cdot r(x) = q(x).$$

Thus p is a harmonic polynomial on $\mathbf{R}^n$ agreeing with q on the unit sphere, completing the proof.
Two remarks. First, the case $m \leq 1$ is covered by the argument above with no special treatment: then $\mathcal{P}_{m-2} = \{0\}$, the operator T on the zero space is trivially injective and surjective, and $\Delta q = 0$ (every polynomial of total degree at most 1 is harmonic), so the construction returns $r = 0$ and $p = q$.
Second, the harmonic polynomial p produced here is unique. If p_1 and p_2 are harmonic polynomials agreeing with q on the unit sphere, then $p_1 - p_2$ is a polynomial, is harmonic (by linearity of Δ), and vanishes on the unit sphere; the quoted fact then gives $p_1 - p_2 = 0$. So the exercise says that every polynomial's restriction to the unit sphere extends in exactly one way to a harmonic polynomial on $\mathbf{R}^n$. This is the polynomial case of the solution of the Dirichlet problem for the ball.

6.2 Exercises 6B

Problem 6B.1. Suppose $e_1, \dots, e_m$ is a list of vectors in V such that

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

for all $a_1, \dots, a_m \in \mathbb{F}$. Show that $e_1, \dots, e_m$ is an orthonormal list.
This exercise provides a converse to 6.24.

Solution. We must show that $\|e_k\| = 1$ for each k and that $\langle e_j, e_k\rangle = 0$ whenever $j \neq k$.
First the norms. Fix $k \in \{1, \dots, m\}$ and apply the hypothesis with $a_k = 1$ and $a_j = 0$ for $j \neq k$. The left side becomes $\|e_k\|^2$ and the right side becomes 1. Hence $\|e_k\|^2 = 1$, so $\|e_k\| = 1$.
Now orthogonality. Fix $j \neq k$ in $\{1, \dots, m\}$ and let $a \in \mathbb{F}$ be arbitrary. Apply the hypothesis with the coefficient of e_j equal to 1, the coefficient of e_k equal to a, and all other coefficients equal to 0. This gives

$$\|e_j + ae_k\|^2 = 1 + |a|^2.$$

On the other hand, expanding the left side using additivity and homogeneity in the first slot and conjugate additivity and conjugate homogeneity in the second slot [6.6],

$$\begin{aligned}\|e_j + ae_k\|^2 &= \langle e_j + ae_k,\ e_j + ae_k\rangle \\ &= \langle e_j, e_j\rangle + \overline{a}\,\langle e_j, e_k\rangle + a\,\langle e_k, e_j\rangle + a\overline{a}\,\langle e_k, e_k\rangle \\ &= 1 + \overline{a}\,\langle e_j, e_k\rangle + \overline{\overline{a}\,\langle e_j, e_k\rangle} + |a|^2,\end{aligned}$$

where we used $\|e_j\| = \|e_k\| = 1$ from the first paragraph and the conjugate symmetry $\langle e_k, e_j\rangle = \overline{\langle e_j, e_k\rangle}$.
Comparing the two computations of $\|e_j + ae_k\|^2$ and cancelling $1 + |a|^2$ from both sides yields

$$\overline{a}\,\langle e_j, e_k\rangle + \overline{\overline{a}\,\langle e_j, e_k\rangle} = 0, \qquad \text{that is,} \qquad 2\,\mathrm{Re}\big(\overline{a}\,\langle e_j, e_k\rangle\big) = 0,$$

for every $a \in \mathbb{F}$.
Taking $a = 1$ gives $\mathrm{Re}\langle e_j, e_k\rangle = 0$. If $\mathbb{F} = \mathbb{R}$, then $\langle e_j, e_k\rangle$ is real and equals its own real part, so $\langle e_j, e_k\rangle = 0$ and we are done. If $\mathbb{F} = \mathbb{C}$, write $\langle e_j, e_k\rangle = x + iy$ with $x, y \in \mathbb{R}$; the case $a = 1$ gives $x = 0$, and taking $a = i$ (so $\overline{a} = -i$) gives

$$0 = \mathrm{Re}\big(-i(x + iy)\big) = \mathrm{Re}(y - ix) = y.$$

Hence $\langle e_j, e_k\rangle = 0$ in this case too.
Since every vector in the list has norm 1 and every two distinct vectors in the list are orthogonal, $e_1, \dots, e_m$ is an orthonormal list.

Problem 6B.2. (a) Suppose $\theta \in \mathbb{R}$. Show that both

$$(\cos\theta, \sin\theta),\ (-\sin\theta, \cos\theta) \qquad \text{and} \qquad (\cos\theta, \sin\theta),\ (\sin\theta, -\cos\theta)$$

are orthonormal bases of $\mathbb{R}^2$.
(b) Show that each orthonormal basis of $\mathbb{R}^2$ is of the form given by one of the two possibilities in (a).

Solution. Throughout, $\mathbb{R}^2$ carries the Euclidean inner product $\langle (x_1, x_2), (y_1, y_2)\rangle = x_1y_1 + x_2y_2$.
(a) Using the identity $\cos^2\theta + \sin^2\theta = 1$ we get

$$\|(\cos\theta, \sin\theta)\|^2 = \cos^2\theta + \sin^2\theta = 1,$$

and likewise $\|(-\sin\theta, \cos\theta)\|^2 = \sin^2\theta + \cos^2\theta = 1$ and $\|(\sin\theta, -\cos\theta)\|^2 = 1$. So all three vectors have norm 1. For orthogonality,

$$\langle (\cos\theta, \sin\theta), (-\sin\theta, \cos\theta)\rangle = -\cos\theta\sin\theta + \sin\theta\cos\theta = 0$$

and

$$\langle (\cos\theta, \sin\theta), (\sin\theta, -\cos\theta)\rangle = \cos\theta\sin\theta - \sin\theta\cos\theta = 0.$$

Hence each of the two lists is an orthonormal list in $\mathbb{R}^2$. Each has length $2 = \dim\mathbb{R}^2$, so each is an orthonormal basis of $\mathbb{R}^2$ by 6.28.
(b) Suppose u, w is an orthonormal basis of $\mathbb{R}^2$ (every basis of $\mathbb{R}^2$ has length 2, so an orthonormal basis is indeed a list of two vectors). Write $u = (a, b)$ and $w = (c, d)$ with $a, b, c, d \in \mathbb{R}$. Orthonormality says

$$a^2 + b^2 = 1, \qquad c^2 + d^2 = 1, \qquad ac + bd = 0.$$

First we produce θ. Since $a^2+b^2=1$ we have $|a|\le 1$, so we may set $\theta_0=\arccos a\in[0,\pi]$, which satisfies $\cos\theta_0=a$ and $\sin\theta_0=\sqrt{1-a^2}=|b|$ (the sine is nonnegative on $[0,\pi]$, and $b^2=1-a^2$). If $b\ge 0$, put $\theta=\theta_0$; if $b<0$, put $\theta=-\theta_0$, and then $\cos\theta=\cos\theta_0=a$ and $\sin\theta=-\sin\theta_0=-|b|=b$. Either way there is $\theta\in\mathbb{R}$ with

$$u=(a,b)=(\cos\theta,\sin\theta).$$

Next we pin down w. Set $s=bc-ad$. Multiplying the relation $ac+bd=0$ by a gives $a^2c+abd=0$, hence $abd=-a^2c$, and therefore

$$-bs=-b^2c+abd=-b^2c-a^2c=-(a^2+b^2)c=-c,$$

so $c=bs$. Multiplying $ac+bd=0$ by b gives $abc+b^2d=0$, hence $abc=-b^2d$, and therefore

$$as=abc-a^2d=-b^2d-a^2d=-(a^2+b^2)d=-d,$$

so $d=-as$. Thus $w=(c,d)=s(b,-a)$. Taking norms,

$$1=c^2+d^2=s^2(a^2+b^2)=s^2,$$

so $s=1$ or $s=-1$.
If $s=-1$, then $w=(-b,a)=(-\sin\theta,\cos\theta)$, so u,w is the first list in (a). If $s=1$, then $w=(b,-a)=(\sin\theta,-\cos\theta)$, so u,w is the second list in (a). Hence every orthonormal basis of $\mathbb{R}^2$ has one of the two forms in (a).

Problem 6B.3. Suppose $e_1,\dots,e_m$ is an orthonormal list in V and $v\in V$. Prove that

$$\|v\|^2=\big|\langle v,e_1\rangle\big|^2+\cdots+\big|\langle v,e_m\rangle\big|^2 \iff v\in\operatorname{span}(e_1,\dots,e_m).$$

Solution. Following the proof of Bessel's inequality (6.26), decompose v as $v=u+w$, where

$$u=\langle v,e_1\rangle e_1+\cdots+\langle v,e_m\rangle e_m \qquad\text{and}\qquad w=v-u.$$

For each $k\in\{1,\dots,m\}$, orthonormality of $e_1,\dots,e_m$ gives

$$\langle w,e_k\rangle=\langle v,e_k\rangle-\sum_{j=1}^{m}\langle v,e_j\rangle\langle e_j,e_k\rangle=\langle v,e_k\rangle-\langle v,e_k\rangle=0.$$

Hence w is orthogonal to each e_k, and therefore (by additivity and homogeneity of the inner product in its first slot, applied to a linear combination of the e_k) w is orthogonal to every vector in $\operatorname{span}(e_1,\dots,e_m)$. In particular $\langle u,w\rangle=0$ since $u\in\operatorname{span}(e_1,\dots,e_m)$, so the Pythagorean theorem (6.12) gives

$$\|v\|^2=\|u\|^2+\|w\|^2.$$

Moreover $\|u\|^2=\big|\langle v,e_1\rangle\big|^2+\cdots+\big|\langle v,e_m\rangle\big|^2$ by 6.24. Combining the last two sentences,

$$\|v\|^2-\Big(\big|\langle v,e_1\rangle\big|^2+\cdots+\big|\langle v,e_m\rangle\big|^2\Big)=\|w\|^2.$$

So the asserted equality holds if and only if $\|w\|=0$, which by definiteness of the norm happens if and only if $w=0$, that is, if and only if $v=u$.
Now both implications follow.
Suppose the equality $\|v\|^2=|\langle v,e_1\rangle|^2+\cdots+|\langle v,e_m\rangle|^2$ holds. By the previous paragraph $v=u$, and u is by construction a linear combination of $e_1,\dots,e_m$. Hence $v\in\operatorname{span}(e_1,\dots,e_m)$.
Conversely, suppose $v\in\operatorname{span}(e_1,\dots,e_m)$. Since $u\in\operatorname{span}(e_1,\dots,e_m)$ as well and a span is a subspace, $w=v-u\in\operatorname{span}(e_1,\dots,e_m)$. But we showed above that w is orthogonal to every vector in that span, so $\langle w,w\rangle=0$ and thus $w=0$. Hence $v=u$, and the equality holds.

Problem 6B.4. Suppose n is a positive integer. Prove that

$$\frac{1}{\sqrt{2\pi}},\ \frac{\cos x}{\sqrt{\pi}},\ \frac{\cos 2x}{\sqrt{\pi}},\ \ldots,\ \frac{\cos nx}{\sqrt{\pi}},\ \frac{\sin x}{\sqrt{\pi}},\ \frac{\sin 2x}{\sqrt{\pi}},\ \ldots,\ \frac{\sin nx}{\sqrt{\pi}}$$

is an orthonormal list of vectors in $C[-\pi,\pi]$, the vector space of continuous real-valued functions on $[-\pi,\pi]$ with inner product

$$\langle f,g\rangle = \int_{-\pi}^{\pi} fg.$$

Hint: The following formulas should help.

$$(\sin x)(\cos y) = \frac{\sin(x-y)+\sin(x+y)}{2}$$
$$(\sin x)(\sin y) = \frac{\cos(x-y)-\cos(x+y)}{2}$$
$$(\cos x)(\cos y) = \frac{\cos(x-y)+\cos(x+y)}{2}$$

Solution. Everything reduces to one elementary computation. For an integer m,

$$\int_{-\pi}^{\pi} \sin(mx)\,dx = 0, \qquad \text{and} \qquad \int_{-\pi}^{\pi} \cos(mx)\,dx = 0 \ \text{ if } m \neq 0.$$

Indeed, if $m = 0$ the first integrand is identically 0. If $m \neq 0$, then

$$\int_{-\pi}^{\pi} \sin(mx)\,dx = \left[\frac{-\cos(mx)}{m}\right]_{-\pi}^{\pi} = \frac{-\cos(m\pi)+\cos(-m\pi)}{m} = 0$$

because cosine is an even function, and

$$\int_{-\pi}^{\pi} \cos(mx)\,dx = \left[\frac{\sin(mx)}{m}\right]_{-\pi}^{\pi} = \frac{\sin(m\pi)-\sin(-m\pi)}{m} = 0$$

because $\sin(m\pi) = 0$ for every integer m. Call these facts $(*)$.
Below, j and k denote integers in $\{1,\ldots,n\}$. Note that then $j+k \neq 0$, and $j-k = 0$ exactly when $j = k$.
Norms. First,

$$\left\|\frac{1}{\sqrt{2\pi}}\right\|^2 = \int_{-\pi}^{\pi} \frac{1}{2\pi}\,dx = \frac{2\pi}{2\pi} = 1.$$

Next, taking both x and y in the third hint formula to be kx gives $\cos^2(kx) = \frac{\cos 0+\cos(2kx)}{2} = \frac{1+\cos(2kx)}{2}$, so by $(*)$ (with $m = 2k \neq 0$),

$$\left\|\frac{\cos kx}{\sqrt{\pi}}\right\|^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2(kx)\,dx = \frac{1}{\pi}\left(\frac{2\pi}{2} + \frac{1}{2}\int_{-\pi}^{\pi} \cos(2kx)\,dx\right) = \frac{1}{\pi}\cdot\pi = 1.$$

Similarly the second hint formula, with both x and y taken to be kx, gives $\sin^2(kx) = \frac{1-\cos(2kx)}{2}$, and the same computation gives $\left\|\frac{\sin kx}{\sqrt{\pi}}\right\|^2 = 1$. So every vector in the list has norm 1.
Orthogonality of the constant with the rest. By $(*)$, for each k,

$$\left\langle \frac{1}{\sqrt{2\pi}}, \frac{\cos kx}{\sqrt{\pi}} \right\rangle = \frac{1}{\pi\sqrt{2}}\int_{-\pi}^{\pi} \cos(kx)\,dx = 0, \qquad \left\langle \frac{1}{\sqrt{2\pi}}, \frac{\sin kx}{\sqrt{\pi}} \right\rangle = \frac{1}{\pi\sqrt{2}}\int_{-\pi}^{\pi} \sin(kx)\,dx = 0.$$

Cosine against sine. By the first hint formula, $(\sin kx)(\cos jx) = \frac{\sin((k-j)x)+\sin((k+j)x)}{2}$, so

$$\left\langle \frac{\cos jx}{\sqrt{\pi}}, \frac{\sin kx}{\sqrt{\pi}} \right\rangle = \frac{1}{2\pi}\left(\int_{-\pi}^{\pi} \sin\big((k-j)x\big)dx + \int_{-\pi}^{\pi} \sin\big((k+j)x\big)dx\right) = 0$$

by $(*)$, which kills every sine integral regardless of the integer multiplier. This holds for all j, k, including $j = k$.
Cosine against cosine and sine against sine, with $j \neq k$. By the third hint formula,

$$\left\langle \frac{\cos jx}{\sqrt{\pi}}, \frac{\cos kx}{\sqrt{\pi}} \right\rangle = \frac{1}{2\pi}\left(\int_{-\pi}^{\pi} \cos\big((j-k)x\big)\,dx + \int_{-\pi}^{\pi} \cos\big((j+k)x\big)\,dx\right) = 0,$$

since $j - k \neq 0$ and $j + k \neq 0$, so $(*)$ applies to both integrals. By the second hint formula,

$$\left\langle \frac{\sin jx}{\sqrt{\pi}}, \frac{\sin kx}{\sqrt{\pi}} \right\rangle = \frac{1}{2\pi}\left(\int_{-\pi}^{\pi} \cos\big((j-k)x\big)\,dx - \int_{-\pi}^{\pi} \cos\big((j+k)x\big)\,dx\right) = 0$$

for the same reason.
We have checked that each of the $2n + 1$ functions in the list has norm 1 and that any two distinct functions in the list are orthogonal. Hence the list is an orthonormal list in $C[-\pi, \pi]$.

Problem 6B.5. Suppose $f\colon [-\pi, \pi] \to \mathbb{R}$ is continuous. For each nonnegative integer k, define

$$a_k = \frac{1}{\sqrt{\pi}}\int_{-\pi}^{\pi} f(x)\cos(kx)\,dx \qquad \text{and} \qquad b_k = \frac{1}{\sqrt{\pi}}\int_{-\pi}^{\pi} f(x)\sin(kx)\,dx.$$

Prove that

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty}\big(a_k^2 + b_k^2\big) \leq \int_{-\pi}^{\pi} f^2.$$

The inequality above is actually an equality for all continuous functions $f\colon [-\pi, \pi] \to \mathbb{R}$. However, proving that this inequality is an equality involves Fourier series techniques beyond the scope of this book.

Solution. Work in the inner product space $V = C[-\pi, \pi]$ of continuous real-valued functions on $[-\pi, \pi]$ with

$$\langle g, h\rangle = \int_{-\pi}^{\pi} gh,$$

as in Exercise 6B.4. This is an inner product on V by the same verification as in 6.3(c), which treats the interval $[-1, 1]$ in place of $[-\pi, \pi]$: bilinearity and symmetry are immediate from linearity of the integral, and if g is continuous with $\int_{-\pi}^{\pi} g^2 = 0$ then $g = 0$, since a continuous function that is nonzero at some point is bounded away from 0 on a nondegenerate subinterval and hence has strictly positive integral of its square. In particular $\|f\|^2 = \int_{-\pi}^{\pi} f^2$.
Fix a positive integer n and consider the list

$$e_0 = \frac{1}{\sqrt{2\pi}}, \qquad c_k = \frac{\cos kx}{\sqrt{\pi}}, \qquad s_k = \frac{\sin kx}{\sqrt{\pi}} \qquad (k = 1, \ldots, n),$$

which is an orthonormal list in V by Exercise 6B.4.
Compute the inner products of f with these vectors. First,

$$\langle f, e_0\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} f(x)\,dx = \frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{\pi}}\int_{-\pi}^{\pi} f(x)\cos(0\cdot x)\,dx = \frac{a_0}{\sqrt{2}},$$

so $\big|\langle f, e_0\rangle\big|^2 = \dfrac{a_0^2}{2}$. Next, for $k = 1, \ldots, n$,

$$\langle f, c_k\rangle = \frac{1}{\sqrt{\pi}}\int_{-\pi}^{\pi} f(x)\cos(kx)\,dx = a_k, \qquad \langle f, s_k\rangle = \frac{1}{\sqrt{\pi}}\int_{-\pi}^{\pi} f(x)\sin(kx)\,dx = b_k.$$

Now Bessel's inequality (6.26), applied to the orthonormal list $e_0, c_1, \ldots, c_n, s_1, \ldots, s_n$ and the vector f, gives

$$\frac{a_0^2}{2} + \sum_{k=1}^{n}\big(a_k^2 + b_k^2\big) = \big|\langle f, e_0\rangle\big|^2 + \sum_{k=1}^{n}\Big(\big|\langle f, c_k\rangle\big|^2 + \big|\langle f, s_k\rangle\big|^2\Big) \leq \|f\|^2 = \int_{-\pi}^{\pi} f^2.$$

This holds for every positive integer n. Let

$$S_n = \frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2).$$

The terms $a_k^2 + b_k^2$ are nonnegative, so (S_n) is an increasing sequence of real numbers, and we have just shown that it is bounded above by $\int_{-\pi}^{\pi} f^2$. Hence (S_n) converges, which is exactly the statement that the infinite series $\frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2)$ converges, and its sum equals $\lim_{n\to\infty} S_n = \sup_n S_n$. Since every S_n is at most $\int_{-\pi}^{\pi} f^2$, so is the supremum. Therefore

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2) \le \int_{-\pi}^{\pi} f^2,$$

as desired.

Problem 6B.6. Suppose $e_1, \dots, e_n$ is an orthonormal basis of V.
(a) Prove that if $v_1, \dots, v_n$ are vectors in V such that

$$\|e_k - v_k\| < \frac{1}{\sqrt{n}}$$

for each k, then $v_1, \dots, v_n$ is a basis of V.
(b) Show that there exist $v_1, \dots, v_n \in V$ such that

$$\|e_k - v_k\| \le \frac{1}{\sqrt{n}}$$

for each k, but $v_1, \dots, v_n$ is not linearly independent.
This exercise states in (a) that an appropriately small perturbation of an orthonormal basis is a basis. Then (b) shows that the number $1/\sqrt{n}$ on the right side of the inequality in (a) cannot be higher.

Solution. Since $e_1, \dots, e_n$ is a basis of V, we have $\dim V = n$; in particular V is finite-dimensional.
(a) We show that $v_1, \dots, v_n$ is linearly independent; because this list has length $n = \dim V$, it is then a basis of V by 2.38.
Suppose $a_1, \dots, a_n \in \mathbb{F}$ are such that

$$a_1 v_1 + \cdots + a_n v_n = 0,$$

and suppose, seeking a contradiction, that not all of $a_1, \dots, a_n$ equal 0. Then

$$a_1 e_1 + \cdots + a_n e_n = (a_1 e_1 + \cdots + a_n e_n) - (a_1 v_1 + \cdots + a_n v_n) = \sum_{k=1}^{n} a_k(e_k - v_k).$$

Taking norms and applying the triangle inequality (6.17) repeatedly, together with the homogeneity $\|aw\| = |a|\,\|w\|$ of the norm,

$$\|a_1 e_1 + \cdots + a_n e_n\| \le \sum_{k=1}^{n} |a_k|\,\|e_k - v_k\|.$$

For each k we have $|a_k|\,\|e_k - v_k\| \le |a_k|/\sqrt{n}$, with strict inequality whenever $a_k \neq 0$. Since at least one a_k is nonzero, summing gives the strict inequality

$$\sum_{k=1}^{n} |a_k|\,\|e_k - v_k\| < \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |a_k|.$$

Now apply the Cauchy–Schwarz inequality (6.14) in $\mathbb{R}^n$ with its Euclidean inner product to the vectors $(|a_1|, \dots, |a_n|)$ and $(1, \dots, 1)$:

$$\sum_{k=1}^{n} |a_k| \le \left(\sum_{k=1}^{n} |a_k|^2 \right)^{1/2} \sqrt{n}.$$

Also, since $e_1, \dots, e_n$ is orthonormal, 6.24 gives

$$\|a_1 e_1 + \cdots + a_n e_n\| = \left(\sum_{k=1}^{n} |a_k|^2 \right)^{1/2}.$$

Writing $c = \left(\sum_{k=1}^{n} |a_k|^2\right)^{1/2}$ and chaining the four displayed relations,

$$c \le \sum_{k=1}^{n} |a_k| \, \|e_k - v_k\| < \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |a_k| \le \frac{1}{\sqrt{n}} \cdot \sqrt{n}\, c = c,$$

so $c < c$, which is impossible. Hence all the a_k equal 0, so $v_1, \dots, v_n$ is linearly independent and therefore a basis of V.
(b) Let $e = e_1 + \cdots + e_n$ and define

$$v_k = e_k - \frac{1}{n} e \qquad \text{for } k = 1, \dots, n.$$

By 6.24 (all coefficients equal 1),

$$\|e\| = \sqrt{\underbrace{1 + \cdots + 1}_{n \text{ terms}}} = \sqrt{n},$$

so for each k,

$$\|e_k - v_k\| = \left\| \frac{1}{n} e \right\| = \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}},$$

which satisfies the required inequality (with equality). On the other hand,

$$v_1 + \cdots + v_n = (e_1 + \cdots + e_n) - n \cdot \frac{1}{n} e = e - e = 0,$$

which is a linear combination of $v_1, \dots, v_n$ equal to 0 with all coefficients equal to 1, hence not all zero. Therefore $v_1, \dots, v_n$ is not linearly independent.
Thus the strict inequality in (a) cannot be weakened to $\le$, and in particular the constant $1/\sqrt{n}$ cannot be replaced by any larger number.

Problem 6B.7. Suppose $T \in \mathcal{L}(\mathbb{R}^3)$ has an upper-triangular matrix with respect to the basis $(1, 0, 0)$, $(1, 1, 1)$, $(1, 1, 2)$. Find an orthonormal basis of $\mathbb{R}^3$ with respect to which T has an upper-triangular matrix.

Solution. The answer is

$$e_1 = (1, 0, 0), \qquad e_2 = \left(0, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right), \qquad e_3 = \left(0, -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right),$$

obtained by applying the Gram–Schmidt procedure to the given basis. Here is the computation and the justification, which is exactly the argument in the proof of 6.37.
Put $v_1 = (1, 0, 0)$, $v_2 = (1, 1, 1)$, $v_3 = (1, 1, 2)$, and use the Euclidean inner product on $\mathbb{R}^3$. Following 6.32, set $f_1 = v_1 = (1, 0, 0)$, so $\|f_1\|^2 = 1$ and

$$e_1 = \frac{f_1}{\|f_1\|} = (1, 0, 0).$$

Since $\langle v_2, f_1\rangle = 1$,

$$f_2 = v_2 - \frac{\langle v_2, f_1\rangle}{\|f_1\|^2} f_1 = (1,1,1) - (1,0,0) = (0,1,1),$$

so $\|f_2\|^2 = 2$, $\|f_2\| = \sqrt{2}$, and

$$e_2 = \frac{f_2}{\|f_2\|} = \left(0, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right).$$

Since $\langle v_3, f_1\rangle = 1$ and $\langle v_3, f_2\rangle = 0 + 1 + 2 = 3$,

$$f_3 = v_3 - \frac{\langle v_3, f_1\rangle}{\|f_1\|^2} f_1 - \frac{\langle v_3, f_2\rangle}{\|f_2\|^2} f_2 = (1,1,2) - (1,0,0) - \tfrac{3}{2}(0,1,1) = \left(0, -\tfrac{1}{2}, \tfrac{1}{2}\right),$$

so $\|f_3\|^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, $\|f_3\| = \frac{1}{\sqrt{2}}$, and

$$e_3 = \frac{f_3}{\|f_3\|} = \sqrt{2}\left(0, -\tfrac{1}{2}, \tfrac{1}{2}\right) = \left(0, -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right).$$

As a check, each of e_1, e_2, e_3 has norm 1, and

$$\langle e_1, e_2\rangle = 0, \qquad \langle e_1, e_3\rangle = 0, \qquad \langle e_2, e_3\rangle = -\tfrac{1}{2} + \tfrac{1}{2} = 0,$$

so e_1, e_2, e_3 is an orthonormal list of length $3 = \dim \mathbb{R}^3$ and hence an orthonormal basis of $\mathbb{R}^3$ by 6.28. It remains to verify that T has an upper-triangular matrix with respect to e_1, e_2, e_3. The Gram–Schmidt procedure guarantees that

$$\operatorname{span}(e_1, \ldots, e_k) = \operatorname{span}(v_1, \ldots, v_k) \qquad \text{for } k = 1, 2, 3$$

(6.32); this is also visible directly in the computation above, since e_1 is a multiple of v_1, and e_2, e_3 were built from v_2, v_3 by subtracting off vectors already in the earlier spans. Because T has an upper-triangular matrix with respect to v_1, v_2, v_3, the subspace $\operatorname{span}(v_1, \ldots, v_k)$ is invariant under T for each k, by 5.39. Hence $\operatorname{span}(e_1, \ldots, e_k)$ is invariant under T for each $k = 1, 2, 3$. Applying 5.39 in the other direction, now to the basis e_1, e_2, e_3, we conclude that the matrix of T with respect to e_1, e_2, e_3 is upper triangular.
Therefore

$$(1,0,0), \qquad \left(0, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right), \qquad \left(0, -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right)$$

is an orthonormal basis of $\mathbb{R}^3$ with respect to which T has an upper-triangular matrix.

Problem 6B.8. Make $\mathcal{P}_2(\mathbf{R})$ into an inner product space by defining $\langle p, q\rangle = \int_0^1 pq$ for all $p, q \in \mathcal{P}_2(\mathbf{R})$.
(a) Apply the Gram–Schmidt procedure to the basis $1, x, x^2$ to produce an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$.
(b) The differentiation operator (the operator that takes p to p') on $\mathcal{P}_2(\mathbf{R})$ has an upper-triangular matrix with respect to the basis $1, x, x^2$, which is not an orthonormal basis. Find the matrix of the differentiation operator on $\mathcal{P}_2(\mathbf{R})$ with respect to the orthonormal basis produced in (a) and verify that this matrix is upper triangular, as expected from the proof of 6.37.

Solution. (a) We run the Gram–Schmidt procedure 6.32 on $v_1 = 1$, $v_2 = x$, $v_3 = x^2$.
Take $f_1 = v_1 = 1$. Then $\|f_1\|^2 = \int_0^1 1\, dt = 1$, so $e_1 = 1$.
Next, $\langle v_2, f_1\rangle = \int_0^1 t\, dt = \frac{1}{2}$, so

$$f_2 = x - \frac{\langle v_2, f_1\rangle}{\|f_1\|^2} f_1 = x - \tfrac{1}{2}.$$

Then

$$\|f_2\|^2 = \int_0^1 \left(t - \tfrac{1}{2}\right)^2 dt = \left[\frac{(t-\frac{1}{2})^3}{3}\right]_0^1 = \frac{1}{24} + \frac{1}{24} = \frac{1}{12},$$

so $\|f_2\| = 1/(2\sqrt{3})$ and

$$e_2 = 2\sqrt{3}\left(x - \tfrac{1}{2}\right) = \sqrt{3}\,(2x-1).$$

Finally, $\langle v_3, f_1\rangle = \int_0^1 t^2\,dt = \frac{1}{3}$ and

$$\langle v_3, f_2\rangle = \int_0^1 \left(t^3 - \tfrac{1}{2}t^2\right) dt = \tfrac{1}{4} - \tfrac{1}{6} = \tfrac{1}{12},$$

so

$$f_3 = x^2 - \frac{1/3}{1}\cdot 1 - \frac{1/12}{1/12}\left(x - \tfrac{1}{2}\right) = x^2 - x + \tfrac{1}{6}.$$

To compute $\|f_3\|^2$, expand

$$\left(t^2 - t + \tfrac{1}{6}\right)^2 = t^4 - 2t^3 + \tfrac{4}{3}t^2 - \tfrac{1}{3}t + \tfrac{1}{36},$$

where the t^2 coefficient is $1 + 2\cdot\frac{1}{6} = \frac{4}{3}$. Hence

$$\|f_3\|^2 = \tfrac{1}{5} - \tfrac{1}{2} + \tfrac{4}{9} - \tfrac{1}{6} + \tfrac{1}{36} = \frac{36 - 90 + 80 - 30 + 5}{180} = \frac{1}{180}.$$

Thus $\|f_3\| = 1/(6\sqrt{5})$ and

$$e_3 = 6\sqrt{5}\left(x^2 - x + \tfrac{1}{6}\right) = \sqrt{5}\,(6x^2 - 6x + 1).$$

So the Gram–Schmidt procedure produces the orthonormal list

$$e_1 = 1, \qquad e_2 = \sqrt{3}\,(2x-1), \qquad e_3 = \sqrt{5}\,(6x^2 - 6x + 1),$$

which has length $3 = \dim \mathcal{P}_2(\mathbf{R})$ and hence is an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$ by 6.28.
(b) Let $D \in \mathcal{L}(\mathcal{P}_2(\mathbf{R}))$ denote differentiation. We compute D on each basis vector and express the result in terms of e_1, e_2, e_3.
First, $De_1 = 0$.
Second, $De_2 = \sqrt{3}\cdot 2 = 2\sqrt{3}$, which equals $2\sqrt{3}\,e_1$ since $e_1 = 1$.
Third, $De_3 = \sqrt{5}\,(12x - 6) = 6\sqrt{5}\,(2x-1)$. Since $2x - 1 = e_2/\sqrt{3}$, this equals

$$\frac{6\sqrt{5}}{\sqrt{3}}\,e_2 = 2\sqrt{15}\,e_2.$$

Writing the coefficients of De_k in the k-th column, we obtain

$$\mathcal{M}\big(D, (e_1, e_2, e_3)\big) = \begin{pmatrix} 0 & 2\sqrt{3} & 0 \\ 0 & 0 & 2\sqrt{15} \\ 0 & 0 & 0 \end{pmatrix}.$$

Every entry below the diagonal equals 0, so this matrix is upper triangular, as the proof of 6.37 predicts: differentiation has an upper-triangular matrix with respect to $1, x, x^2$, so $\operatorname{span}(v_1, \ldots, v_k)$ is invariant under D for each k (see 5.39); because $\operatorname{span}(e_1, \ldots, e_k) = \operatorname{span}(v_1, \ldots, v_k)$ for each k by 6.32, those spans are also invariant under D, and 5.39 again gives an upper-triangular matrix with respect to e_1, e_2, e_3.

Problem 6B.9. Suppose $e_1, \ldots, e_m$ is the result of applying the Gram–Schmidt procedure to a linearly independent list $v_1, \ldots, v_m$ in V. Prove that $\langle v_k, e_k\rangle > 0$ for each $k = 1, \ldots, m$.

Solution. Let $f_1, \dots, f_m$ be the vectors produced by the Gram–Schmidt formulas 6.32, so that $f_1 = v_1$,

$$f_k = v_k - \frac{\langle v_k, f_1\rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1}\rangle}{\|f_{k-1}\|^2} f_{k-1}$$

for $k = 2, \dots, m$, and $e_k = f_k/\|f_k\|$. As shown in the proof of 6.32, the linear independence of $v_1, \dots, v_m$ guarantees $f_k \neq 0$ for each k, so $\|f_k\| > 0$ and each e_k is defined.
Fix $k \in \{1, \dots, m\}$. Rearranging the formula above (and reading it as $v_1 = f_1$ when $k = 1$) gives

$$v_k = f_k + u_k, \qquad \text{where } u_k = \sum_{j=1}^{k-1} \frac{\langle v_k, f_j\rangle}{\|f_j\|^2} f_j \in \operatorname{span}(f_1, \dots, f_{k-1}).$$

Since e_j is a scalar multiple of f_j and f_j is a scalar multiple of e_j for each j, we have $\operatorname{span}(f_1, \dots, f_{k-1}) = \operatorname{span}(e_1, \dots, e_{k-1})$. Because $e_1, \dots, e_m$ is an orthonormal list (by 6.32), the vector e_k is orthogonal to each of $e_1, \dots, e_{k-1}$, hence orthogonal to every vector in their span. In particular $\langle u_k, e_k\rangle = 0$. Therefore

$$\langle v_k, e_k\rangle = \langle f_k, e_k\rangle + \langle u_k, e_k\rangle = \langle f_k, e_k\rangle.$$

Now $e_k = f_k/\|f_k\|$ and $\|f_k\|$ is a positive real number, so homogeneity in the second slot (which introduces a complex conjugate, harmless here since the scalar is real and positive) gives

$$\langle f_k, e_k\rangle = \left\langle f_k, \frac{1}{\|f_k\|} f_k \right\rangle = \frac{1}{\|f_k\|}\langle f_k, f_k\rangle = \frac{\|f_k\|^2}{\|f_k\|} = \|f_k\|.$$

Thus $\langle v_k, e_k\rangle = \|f_k\| > 0$, because $f_k \neq 0$ and the definiteness axiom for the inner product forces $\|f_k\| > 0$. This holds for each $k = 1, \dots, m$, as desired.

Problem 6B.10. Suppose $v_1, \dots, v_m$ is a linearly independent list in V. Explain why the orthonormal list produced by the formulas of the Gram–Schmidt procedure (6.32) is the only orthonormal list $e_1, \dots, e_m$ in V such that $\langle v_k, e_k\rangle > 0$ and $\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k)$ for each $k = 1, \dots, m$. [The result in this exercise is used in the proof of 7.58.]

Solution. Let $e_1, \dots, e_m$ denote the orthonormal list produced by the Gram–Schmidt formulas 6.32 applied to $v_1, \dots, v_m$.
First, this list does have the two stated properties: it is orthonormal and satisfies $\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k)$ for each k by 6.32, and $\langle v_k, e_k\rangle > 0$ for each k by Exercise 6B.9. So such a list exists.
Now suppose $g_1, \dots, g_m$ is any orthonormal list in V with $\langle v_k, g_k\rangle > 0$ and

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(g_1, \dots, g_k)$$

for each $k = 1, \dots, m$. We show by induction on k that $g_k = e_k$.
Base case $k = 1$. We have $\operatorname{span}(g_1) = \operatorname{span}(v_1) = \operatorname{span}(e_1)$, so $g_1 = ae_1$ for some scalar a. Orthonormality gives $1 = \|g_1\| = |a|\,\|e_1\| = |a|$. Also

$$\langle v_1, g_1\rangle = \langle v_1, ae_1\rangle = \overline{a}\,\langle v_1, e_1\rangle.$$

Both $\langle v_1, g_1\rangle$ and $\langle v_1, e_1\rangle$ are positive real numbers (the latter by 6B.9), so $\overline{a}$ is a positive real number; combined with $|a| = 1$ this forces $a = 1$. Hence $g_1 = e_1$.
Induction step. Let $1 < k \leq m$ and suppose $g_j = e_j$ for all $j < k$. Since

$$\operatorname{span}(g_1, \dots, g_k) = \operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k),$$

we have $g_k \in \operatorname{span}(e_1, \dots, e_k)$. Because $e_1, \dots, e_k$ is an orthonormal basis of that subspace, 6.30(a) applied inside it gives

$$g_k = \langle g_k, e_1\rangle e_1 + \cdots + \langle g_k, e_k\rangle e_k.$$

For $j < k$ we have $e_j = g_j$ by the induction hypothesis, and $\langle g_k, g_j\rangle = 0$ because $g_1, \dots, g_m$ is orthonormal; hence $\langle g_k, e_j\rangle = 0$ for each $j < k$. Therefore $g_k = ae_k$, where $a = \langle g_k, e_k\rangle$. As in the base case, $1 = \|g_k\| = |a|$, and

$$\langle v_k, g_k\rangle = \overline{a}\,\langle v_k, e_k\rangle$$

with both $\langle v_k, g_k\rangle$ and $\langle v_k, e_k\rangle$ positive real numbers. Thus $\overline{a} > 0$ and $|a| = 1$, giving $a = 1$ and $g_k = e_k$, completing the induction.
Hence every orthonormal list with the two stated properties equals the Gram–Schmidt list $e_1, \dots, e_m$, which is what was to be explained.

Problem 6B.11. Find a polynomial $q \in \mathcal{P}_2(\mathbf{R})$ such that $p\left(\frac{1}{2}\right) = \int_0^1 pq$ for every $p \in \mathcal{P}_2(\mathbf{R})$.

Solution. Give $\mathcal{P}_2(\mathbf{R})$ the inner product $\langle p, q\rangle = \int_0^1 pq$, and let φ be the linear functional on $\mathcal{P}_2(\mathbf{R})$ defined by $\varphi(p) = p\left(\frac{1}{2}\right)$. We seek q with $\varphi(p) = \langle p, q\rangle$ for all p; the Riesz representation theorem 6.42 says such a q exists and is unique, and formula 6.43 tells us how to compute it: if e_1, e_2, e_3 is an orthonormal basis, then

$$q = \varphi(e_1)e_1 + \varphi(e_2)e_2 + \varphi(e_3)e_3$$

(no complex conjugates appear, since the scalar field is $\mathbf{R}$).
By Exercise 6B.8(a), an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$ for this inner product is

$$e_1 = 1, \qquad e_2 = \sqrt{3}\,(2x - 1), \qquad e_3 = \sqrt{5}\,(6x^2 - 6x + 1).$$

Evaluating at $x = \frac{1}{2}$:

$$\varphi(e_1) = 1, \qquad \varphi(e_2) = \sqrt{3}\,(1 - 1) = 0, \qquad \varphi(e_3) = \sqrt{5}\left(\tfrac{3}{2} - 3 + 1\right) = -\frac{\sqrt{5}}{2}.$$

Therefore

$$q = 1 \cdot 1 + 0 \cdot e_2 - \frac{\sqrt{5}}{2} \cdot \sqrt{5}\,(6x^2 - 6x + 1) = 1 - \frac{5}{2}\,(6x^2 - 6x + 1),$$

that is,

$$q(x) = -15x^2 + 15x - \tfrac{3}{2}.$$

Verification. It suffices to check the identity $p\left(\frac{1}{2}\right) = \int_0^1 pq$ for $p = 1, x, x^2$, since both sides are linear in p and these three polynomials span $\mathcal{P}_2(\mathbf{R})$. We have

$$\int_0^1 q = -5 + \tfrac{15}{2} - \tfrac{3}{2} = 1 = 1\big|_{x=1/2},$$

$$\int_0^1 x\,q(x)\,dx = -\tfrac{15}{4} + 5 - \tfrac{3}{4} = \tfrac{1}{2} = x\big|_{x=1/2},$$

$$\int_0^1 x^2 q(x)\,dx = -3 + \tfrac{15}{4} - \tfrac{1}{2} = \tfrac{1}{4} = x^2\big|_{x=1/2}.$$

So $q(x) = -15x^2 + 15x - \frac{3}{2}$ works, and by the uniqueness half of 6.42 it is the only such polynomial.

Problem 6B.12. Find a polynomial $q \in \mathcal{P}_2(\mathbf{R})$ such that

$$\int_0^1 p(x)\cos(\pi x)\,dx = \int_0^1 pq$$

for every $p \in \mathcal{P}_2(\mathbf{R})$.

Solution. Give $\mathcal{P}_2(\mathbf{R})$ the inner product $\langle p, q\rangle = \int_0^1 pq$ and define the linear functional

$$\varphi(p) = \int_0^1 p(x)\cos(\pi x)\,dx.$$

By the Riesz representation theorem 6.42 there is a unique $q \in \mathcal{P}_2(\mathbf{R})$ with $\varphi(p) = \langle p, q\rangle$ for all p, and by 6.43 it is given by $q = \varphi(e_1)e_1 + \varphi(e_2)e_2 + \varphi(e_3)e_3$ for any orthonormal basis e_1, e_2, e_3 (the field is $\mathbf{R}$, so no conjugates appear). Use the orthonormal basis found in Exercise 6B.8(a):

$$e_1 = 1, \qquad e_2 = \sqrt{3}\,(2x-1), \qquad e_3 = \sqrt{5}\,(6x^2 - 6x + 1).$$

We need three integrals. First,

$$\int_0^1 \cos(\pi x)\,dx = \left[\frac{\sin(\pi x)}{\pi}\right]_0^1 = 0.$$

Next, integrating by parts twice,

$$\int_0^1 x\sin(\pi x)\,dx = \left[\frac{-x\cos(\pi x)}{\pi}\right]_0^1 + \frac{1}{\pi}\int_0^1 \cos(\pi x)\,dx = \frac{1}{\pi},$$

$$\int_0^1 x\cos(\pi x)\,dx = \left[\frac{x\sin(\pi x)}{\pi}\right]_0^1 - \frac{1}{\pi}\int_0^1 \sin(\pi x)\,dx = 0 - \frac{1}{\pi}\cdot\frac{2}{\pi} = -\frac{2}{\pi^2},$$

using $\int_0^1 \sin(\pi x)\,dx = [-\cos(\pi x)/\pi]_0^1 = 2/\pi$. Finally,

$$\int_0^1 x^2\cos(\pi x)\,dx = \left[\frac{x^2\sin(\pi x)}{\pi}\right]_0^1 - \frac{2}{\pi}\int_0^1 x\sin(\pi x)\,dx = 0 - \frac{2}{\pi}\cdot\frac{1}{\pi} = -\frac{2}{\pi^2}.$$

Now compute the three coefficients:

$$\varphi(e_1) = \int_0^1 \cos(\pi x)\,dx = 0,$$

$$\varphi(e_2) = \sqrt{3}\left(2\int_0^1 x\cos(\pi x)\,dx - \int_0^1 \cos(\pi x)\,dx\right) = \sqrt{3}\left(-\frac{4}{\pi^2}\right) = -\frac{4\sqrt{3}}{\pi^2},$$

$$\varphi(e_3) = \sqrt{5}\left(6\left(-\frac{2}{\pi^2}\right) - 6\left(-\frac{2}{\pi^2}\right) + 0\right) = 0.$$

Therefore

$$q = -\frac{4\sqrt{3}}{\pi^2}\cdot\sqrt{3}\,(2x-1) = -\frac{12}{\pi^2}\,(2x-1) = \frac{12}{\pi^2}\,(1-2x).$$

Verification. Both sides of the desired identity are linear in p, so it suffices to check $p = 1, x, x^2$. With $q(x) = \frac{12}{\pi^2}(1-2x)$,

$$\int_0^1 q = \frac{12}{\pi^2}\,(1-1) = 0 = \int_0^1 \cos(\pi x)\,dx,$$

$$\int_0^1 x\,q(x)\,dx = \frac{12}{\pi^2}\left(\tfrac{1}{2} - \tfrac{2}{3}\right) = -\frac{2}{\pi^2} = \int_0^1 x\cos(\pi x)\,dx,$$

$$\int_0^1 x^2 q(x)\,dx = \frac{12}{\pi^2}\left(\tfrac{1}{3} - \tfrac{1}{2}\right) = -\frac{2}{\pi^2} = \int_0^1 x^2\cos(\pi x)\,dx.$$

So $q(x) = \dfrac{12(1-2x)}{\pi^2}$ is the (unique, by 6.42) polynomial with the required property.

Problem 6B.13. Show that a list $v_1, \dots, v_m$ of vectors in V is linearly dependent if and only if the Gram–Schmidt formula in 6.32 produces $f_k = 0$ for some $k \in \{1, \dots, m\}$.
[This exercise gives an alternative to Gaussian elimination techniques for determining whether a list of vectors in an inner product space is linearly dependent.]

Solution. Recall the formulas 6.32: $f_1 = v_1$ and, for $k \geq 2$,

$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}.$$

The formula for f_k makes sense only when none of $f_1, \dots, f_{k-1}$ equals 0, since $\|f_j\|^2$ appears in a denominator. So say that f_k is defined when $f_1, \dots, f_{k-1}$ are all defined and all nonzero; the procedure thus produces $f_1, f_2, \dots$ in turn and halts at the first index where the output is 0. The assertion that the formula produces $f_k = 0$ for some $k \in \{1, \dots, m\}$ thus means that the procedure halts at some index $k \leq m$, rather than running through all m steps with nonzero output. Everything below refers only to vectors f_k that are defined in this sense.
First, a computation valid whenever $f_1, \dots, f_k$ are defined: we claim $f_1, \dots, f_k$ is an orthogonal list. Argue by induction on k; for $k = 1$ there is nothing to prove. Assume $f_1, \dots, f_{k-1}$ are pairwise orthogonal and let $j \in \{1, \dots, k-1\}$. Then, since $\langle f_i, f_j \rangle = 0$ for $i \neq j$ with $i, j \leq k-1$,

$$\langle f_k, f_j \rangle = \langle v_k, f_j \rangle - \sum_{i=1}^{k-1} \frac{\langle v_k, f_i \rangle}{\|f_i\|^2} \langle f_i, f_j \rangle = \langle v_k, f_j \rangle - \frac{\langle v_k, f_j \rangle}{\|f_j\|^2} \|f_j\|^2 = 0.$$

This proves the claim. Note that this computation nowhere used linear independence of the v's; only $f_i \neq 0$ for $i < k$ was needed.
Second, whenever $f_1, \dots, f_k$ are defined we have

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(f_1, \dots, f_k).$$

Indeed, by induction: $f_1 = v_1$; and if the spans agree through index $k-1$, then the displayed formula shows $f_k \in \operatorname{span}(v_k, f_1, \dots, f_{k-1}) \subseteq \operatorname{span}(v_1, \dots, v_k)$ and, rearranged, that $v_k \in \operatorname{span}(f_1, \dots, f_k)$; combining these two inclusions with the inductive hypothesis gives the claim.
Now we prove both directions.
Suppose $f_k = 0$ for some k, and take k to be the smallest such index (so $f_1, \dots, f_k$ are all defined). If $k = 1$, then $v_1 = f_1 = 0$, and a list containing the zero vector is linearly dependent. If $k > 1$, then the formula for f_k rearranges to

$$v_k = \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 + \cdots + \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1} \in \operatorname{span}(f_1, \dots, f_{k-1}) = \operatorname{span}(v_1, \dots, v_{k-1}),$$

using the second claim above. Thus v_k is a linear combination of $v_1, \dots, v_{k-1}$, so the list $v_1, \dots, v_m$ is linearly dependent (a nontrivial linear combination equal to 0 is obtained by moving v_k to the other side, with coefficient -1 on v_k).
Conversely, suppose the procedure produces no zero vector, so $f_1, \dots, f_m$ are all defined and all nonzero. By the first claim, $f_1, \dots, f_m$ is an orthogonal list of nonzero vectors; setting $e_k = f_k / \|f_k\|$ gives an orthonormal list $e_1, \dots, e_m$, which is linearly independent by 6.25. Hence $\operatorname{span}(f_1, \dots, f_m) = \operatorname{span}(e_1, \dots, e_m)$ has dimension m. By the second claim this span equals $\operatorname{span}(v_1, \dots, v_m)$, so the list $v_1, \dots, v_m$ spans a subspace U with $\dim U = m$. A list of length $\dim U$ that spans U is a basis of U (by 2.42), hence $v_1, \dots, v_m$ is linearly independent.
Combining the two paragraphs: $v_1, \dots, v_m$ is linearly dependent if and only if $f_k = 0$ for some $k \in \{1, \dots, m\}$.

Problem 6B.14. Suppose V is a real inner product space and $v_1, \dots, v_m$ is a linearly independent list of vectors in V. Prove that there exist exactly 2^m orthonormal lists $e_1, \dots, e_m$ of vectors in V such that

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k)$$

for all $k \in \{1, \dots, m\}$.

Solution. Let $g_1, \dots, g_m$ be the orthonormal list produced by applying the Gram–Schmidt procedure 6.32 to $v_1, \dots, v_m$; by 6.32 it satisfies

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(g_1, \dots, g_k) \qquad \text{for } k = 1, \dots, m.$$

We show that the lists with the required property are precisely the lists

$$\varepsilon_1 g_1, \ \dots, \ \varepsilon_m g_m \qquad \text{with } \varepsilon_1, \dots, \varepsilon_m \in \{1, -1\},$$

and that these 2^m sign choices give 2^m distinct lists. This proves the assertion.

Step 1: each such list works. Fix signs $\varepsilon_1, \dots, \varepsilon_m \in \{1, -1\}$ and set $e_k = \varepsilon_k g_k$. Then $\|e_k\| = |\varepsilon_k| \, \|g_k\| = 1$, and for $j \neq k$ we have $\langle e_j, e_k \rangle = \varepsilon_j \varepsilon_k \langle g_j, g_k \rangle = 0$; so $e_1, \dots, e_m$ is orthonormal. Moreover, since $\varepsilon_k \neq 0$, each e_k is a nonzero scalar multiple of g_k and vice versa, so

$$\operatorname{span}(e_1, \dots, e_k) = \operatorname{span}(g_1, \dots, g_k) = \operatorname{span}(v_1, \dots, v_k)$$

for each k, as required.

Step 2: there are no others. Suppose $e_1, \dots, e_m$ is an orthonormal list in V with $\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k)$ for each k. We show that $e_k = \pm g_k$ for every k; no induction is needed, because the argument for a given index uses only the span identities.

Fix $k \in \{1, \dots, m\}$. We have

$$\operatorname{span}(e_1, \dots, e_k) = \operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(g_1, \dots, g_k),$$

so in particular $e_k \in \operatorname{span}(g_1, \dots, g_k)$. Since $g_1, \dots, g_k$ is an orthonormal basis of that subspace, 6.30(a) applied within it gives

$$e_k = \langle e_k, g_1 \rangle g_1 + \cdots + \langle e_k, g_k \rangle g_k.$$

Now let $j < k$. Then $g_j \in \operatorname{span}(g_1, \dots, g_{k-1}) = \operatorname{span}(e_1, \dots, e_{k-1})$, so g_j is a linear combination of $e_1, \dots, e_{k-1}$. Since $e_1, \dots, e_m$ is orthonormal, e_k is orthogonal to each of $e_1, \dots, e_{k-1}$ and hence to every vector in their span; therefore $\langle e_k, g_j \rangle = 0$ for each $j < k$. (For $k = 1$ this step is vacuous.) The displayed expansion thus collapses to

$$e_k = a_k g_k, \qquad \text{where } a_k = \langle e_k, g_k \rangle \in \mathbf{R}.$$

Taking norms, $1 = \|e_k\| = |a_k| \, \|g_k\| = |a_k|$. The scalar field is $\mathbf{R}$, so the only real numbers of absolute value 1 are 1 and -1; hence $a_k \in \{1, -1\}$ and $e_k = \pm g_k$. This is exactly where the hypothesis that V is a real inner product space is used: over $\mathbf{C}$ the scalar a_k would only be forced to satisfy $|a_k| = 1$, allowing infinitely many choices at each step.

Step 3: the 2^m lists are distinct. If $(\varepsilon_1, \dots, \varepsilon_m)$ and $(\delta_1, \dots, \delta_m)$ are distinct sign vectors, pick k with $\varepsilon_k \neq \delta_k$. Then $\varepsilon_k g_k - \delta_k g_k = \pm 2 g_k \neq 0$ because $\|g_k\| = 1$, so the k-th entries of the two lists differ and the lists are distinct.

By Steps 1–3 the set of orthonormal lists with the stated span property is in bijection with $\{1, -1\}^m$, which has exactly 2^m elements. Hence there are exactly 2^m such lists.

Remark. This is consistent with Exercise 6B.10: among these 2^m lists, exactly one also satisfies $\langle v_k, e_k \rangle > 0$ for every k, namely the Gram–Schmidt list itself (for each k, exactly one of $\pm g_k$ makes $\langle v_k, e_k \rangle$ positive, since $\langle v_k, g_k \rangle > 0$ by 6B.9).

Problem 6B.15. Suppose $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ are inner products on V such that $\langle u, v \rangle_1 = 0$ if and only if $\langle u, v \rangle_2 = 0$. Prove that there is a positive number c such that $\langle u, v \rangle_1 = c\langle u, v \rangle_2$ for every $u, v \in V$. *This exercise shows that if two inner products have the same pairs of orthogonal vectors, then each of the inner products is a scalar multiple of the other inner product.*

Solution. If $V = \{0\}$ then both inner products are the zero function on $V \times V$ and $c = 1$ works. So assume $V \neq \{0\}$.
For $x \in V$ with $x \neq 0$ define

$$f(x) = \frac{\langle x, x \rangle_1}{\langle x, x \rangle_2},$$

which makes sense and is a positive real number because positive definiteness of an inner product gives $\langle x, x \rangle_1 > 0$ and $\langle x, x \rangle_2 > 0$. The whole proof consists of showing that f is constant and that this constant does the job.
Step 1: two-dimensional subspaces. Suppose U is a two-dimensional subspace of V. Restricting $\langle \cdot, \cdot \rangle_2$ to U makes U an inner product space, so by 6.35 there is a list e_1, e_2 that is an orthonormal basis of U with respect to $\langle \cdot, \cdot \rangle_2$; thus

$$\langle e_1, e_1 \rangle_2 = \langle e_2, e_2 \rangle_2 = 1, \qquad \langle e_1, e_2 \rangle_2 = 0.$$

Because $\langle e_1, e_2 \rangle_2 = 0$, the hypothesis gives

$$\langle e_1, e_2 \rangle_1 = 0,$$

and hence also $\langle e_2, e_1 \rangle_1 = \overline{\langle e_1, e_2 \rangle_1} = 0$ by conjugate symmetry.
Next,

$$\langle e_1 + e_2, e_1 - e_2 \rangle_2 = \langle e_1, e_1 \rangle_2 - \langle e_1, e_2 \rangle_2 + \langle e_2, e_1 \rangle_2 - \langle e_2, e_2 \rangle_2 = 1 - 0 + 0 - 1 = 0.$$

So the hypothesis applies again and yields $\langle e_1 + e_2, e_1 - e_2 \rangle_1 = 0$. Expanding this with the sesquilinearity of $\langle \cdot, \cdot \rangle_1$ and using $\langle e_1, e_2 \rangle_1 = \langle e_2, e_1 \rangle_1 = 0$ gives

$$0 = \langle e_1, e_1 \rangle_1 - \langle e_1, e_2 \rangle_1 + \langle e_2, e_1 \rangle_1 - \langle e_2, e_2 \rangle_1 = \langle e_1, e_1 \rangle_1 - \langle e_2, e_2 \rangle_1.$$

Write $c_U = \langle e_1, e_1 \rangle_1$. We have just shown $\langle e_2, e_2 \rangle_1 = c_U$ as well, and $c_U > 0$.
Now let $x = a_1 e_1 + a_2 e_2$ and $y = b_1 e_1 + b_2 e_2$ be arbitrary elements of U. Expanding both inner products and using the three computed values $\langle e_1, e_1 \rangle_1 = \langle e_2, e_2 \rangle_1 = c_U$ and $\langle e_1, e_2 \rangle_1 = \langle e_2, e_1 \rangle_1 = 0$ gives

$$\langle x, y \rangle_1 = c_U(a_1 \overline{b_1} + a_2 \overline{b_2}) = c_U \langle x, y \rangle_2.$$

In particular, taking $y = x \neq 0$ shows $f(x) = c_U$ for every nonzero $x \in U$. So:
(i) on every two-dimensional subspace U the two inner products satisfy $\langle \cdot, \cdot \rangle_1 = c_U \langle \cdot, \cdot \rangle_2$, and
(ii) f is constant on $U \setminus \{0\}$.
Step 2: the function f is constant on $V \setminus \{0\}$. Let $x, y \in V$ be nonzero. If x, y is linearly dependent, then $y = \lambda x$ for some $\lambda \neq 0$ (or $x = \lambda y$, handled the same way), and then

$$f(y) = \frac{\langle \lambda x, \lambda x \rangle_1}{\langle \lambda x, \lambda x \rangle_2} = \frac{|\lambda|^2 \langle x, x \rangle_1}{|\lambda|^2 \langle x, x \rangle_2} = f(x).$$

If x, y is linearly independent, then $U = \operatorname{span}(x, y)$ is a two-dimensional subspace containing both, so $f(x) = f(y)$ by (ii). Hence f takes a single value on $V \setminus \{0\}$; call it c. As noted above, $c > 0$.
Step 3: conclusion. Let $u, v \in V$; we show $\langle u, v \rangle_1 = c\langle u, v \rangle_2$.
If $u = 0$ or $v = 0$, both sides are 0.
If u, v are both nonzero and linearly dependent, write $v = \lambda u$ with $\lambda \neq 0$ (the case $u = \lambda v$ is symmetric). Then, using $\langle u, u \rangle_1 = f(u)\langle u, u \rangle_2 = c\langle u, u \rangle_2$,

$$\langle u, v \rangle_1 = \overline{\lambda} \langle u, u \rangle_1 = \overline{\lambda}\, c \langle u, u \rangle_2 = c\langle u, \lambda u \rangle_2 = c\langle u, v \rangle_2.$$

If u, v is linearly independent, let $U = \operatorname{span}(u, v)$, a two-dimensional subspace. By (i), $\langle u, v\rangle_1 = c_U \langle u, v\rangle_2$, and by (ii) applied to the nonzero vector $u \in U$ we get $c_U = f(u) = c$. Hence $\langle u, v\rangle_1 = c\langle u, v\rangle_2$.
In all cases $\langle u, v\rangle_1 = c\langle u, v\rangle_2$, with $c > 0$, as desired.
Two remarks. First, only one direction of the hypothesis was used, namely that $\langle u, v\rangle_2 = 0$ implies $\langle u, v\rangle_1 = 0$; the reverse implication then follows automatically from the conclusion, since $c \neq 0$. Second, V was not assumed to be finite-dimensional: every application of 6.35 above was made to a subspace of dimension 2.

Problem 6B.16. Suppose V is finite-dimensional. Suppose $\langle \cdot, \cdot\rangle_1$, $\langle \cdot, \cdot\rangle_2$ are inner products on V with corresponding norms $\| \cdot \|_1$ and $\| \cdot \|_2$. Prove that there exists a positive number c such that $\|v\|_1 \leq c\|v\|_2$ for every $v \in V$.

Solution. If $V = \{0\}$, then $\|v\|_1 = 0 = \|v\|_2$ for the only $v \in V$, and $c = 1$ works. So assume $n = \dim V \geq 1$.
Regard V as an inner product space using $\langle \cdot, \cdot\rangle_2$. By 6.35 there is a basis $e_1, \ldots, e_n$ of V that is orthonormal with respect to $\langle \cdot, \cdot\rangle_2$.
Set

$$M = \max\{\|e_1\|_1, \ldots, \|e_n\|_1\}.$$

Because $e_1 \neq 0$ we have $\|e_1\|_1 > 0$, so $M > 0$. Define $c = \sqrt{n}\, M$, a positive number.
Now let $v \in V$ and write $v = a_1 e_1 + \cdots + a_n e_n$ with $a_1, \ldots, a_n \in \mathbf{F}$. Because $e_1, \ldots, e_n$ is orthonormal with respect to $\langle \cdot, \cdot\rangle_2$, the formula 6.24 for the norm of an orthonormal linear combination gives

$$\|v\|_2^2 = |a_1|^2 + \cdots + |a_n|^2.$$

For the other norm we use only the triangle inequality and homogeneity, both of which hold for $\| \cdot \|_1$ since it comes from an inner product (see 6.9 and 6.17):

$$\|v\|_1 = \|a_1 e_1 + \cdots + a_n e_n\|_1 \leq |a_1|\, \|e_1\|_1 + \cdots + |a_n|\, \|e_n\|_1 \leq M\left(|a_1| + \cdots + |a_n|\right).$$

Finally apply the Cauchy–Schwarz inequality (6.14) in $\mathbf{R}^n$ with its usual inner product, to the vectors $(|a_1|, \ldots, |a_n|)$ and $(1, \ldots, 1)$:

$$|a_1| + \cdots + |a_n| \leq \sqrt{n}\left(|a_1|^2 + \cdots + |a_n|^2\right)^{1/2} = \sqrt{n}\, \|v\|_2.$$

Combining the last two displays,

$$\|v\|_1 \leq M\sqrt{n}\, \|v\|_2 = c\|v\|_2,$$

which is the desired inequality. Since $v \in V$ was arbitrary and $c > 0$ does not depend on v, the proof is complete.
Note that the roles of the two inner products may be interchanged, so there is also a positive c' with $\|v\|_2 \leq c'\|v\|_1$ for all $v \in V$; that is, any two norms arising from inner products on a finite-dimensional vector space are equivalent. Finite-dimensionality is essential: on the space of sequences in $\mathbf{F}^\infty$ with only finitely many nonzero terms, the inner products $\langle x, y\rangle_1 = \sum_k k^2 x_k \overline{y_k}$ and $\langle x, y\rangle_2 = \sum_k x_k \overline{y_k}$ admit no such c, as testing on the standard basis vectors shows.

Problem 6B.17. Suppose $\mathbf{F} = \mathbf{C}$ and V is finite-dimensional. Prove that if T is an operator on V such that 1 is the only eigenvalue of T and $\|Tv\| \leq \|v\|$ for all $v \in V$, then T is the identity operator.

Solution. If $V = \{0\}$ then the only operator on V is the identity operator and there is nothing to prove, so assume $V \neq \{0\}$ and let $n = \dim V \geq 1$. (In fact the hypothesis already forces $V \neq \{0\}$ under the reading that T has 1 as an eigenvalue, since an operator on the zero space has no eigenvalues; but we do not need this.)

Since $\mathbf{F} = \mathbf{C}$ and V is a finite-dimensional complex inner product space, Schur's theorem (6.38) provides an orthonormal basis $e_1, \dots, e_n$ of V with respect to which T has an upper-triangular matrix. Write

$$A = \mathcal{M}(T, (e_1, \dots, e_n)),$$

so $A_{j,k} = 0$ whenever $j > k$, and

$$Te_k = A_{1,k}e_1 + \cdots + A_{k,k}e_k$$

for each $k \in \{1, \dots, n\}$.
By 5.41, the eigenvalues of T are exactly the diagonal entries $A_{1,1}, \dots, A_{n,n}$ of this upper-triangular matrix. By hypothesis 1 is the only eigenvalue of T, so

$$A_{k,k} = 1 \quad \text{for every } k \in \{1, \dots, n\}.$$

Now fix k. Because $e_1, \dots, e_k$ is an orthonormal list, the formula 6.24 for the norm of an orthonormal linear combination gives

$$\|Te_k\|^2 = \left\| \sum_{j=1}^{k} A_{j,k} e_j \right\|^2 = \sum_{j=1}^{k} |A_{j,k}|^2 = 1 + \sum_{j=1}^{k-1} |A_{j,k}|^2,$$

where the last equality uses $A_{k,k} = 1$.
On the other hand, the hypothesis $\|Tv\| \le \|v\|$ applied to $v = e_k$ gives

$$\|Te_k\|^2 \le \|e_k\|^2 = 1.$$

Combining the two displays,

$$1 + \sum_{j=1}^{k-1} |A_{j,k}|^2 \le 1,$$

so $\sum_{j=1}^{k-1} |A_{j,k}|^2 \le 0$. A sum of nonnegative real numbers that is at most 0 must have every term equal to 0, so $A_{j,k} = 0$ for all $j < k$.
Thus for each k we have $Te_k = A_{k,k}e_k = e_k$. Two linear maps that agree on a basis are equal (this is the uniqueness half of the linear map lemma 3.4), and the identity operator I also satisfies $Ie_k = e_k$ for each k. Hence $T = I$.
Both hypotheses are needed. The operator on $\mathbf{C}^2$ with matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ with respect to the standard basis has 1 as its only eigenvalue but is not the identity; it fails the norm condition, since it sends $(0, 1)$, of norm 1, to $(1, 1)$, of norm $\sqrt{2}$. Conversely, the operator on $\mathbf{C}^2$ given by multiplication by $\frac{1}{2}$ satisfies $\|Tv\| \le \|v\|$ but has $\frac{1}{2}$, not 1, as its only eigenvalue. The assumption $\mathbf{F} = \mathbf{C}$ is what makes Schur's theorem available, and the proof used no other feature of the complex case.

Problem 6B.18. Suppose $u_1, \dots, u_m$ is a linearly independent list in V. Show that there exists $v \in V$ such that $\langle u_k, v \rangle = 1$ for all $k \in \{1, \dots, m\}$.

Solution. Let

$$U = \operatorname{span}(u_1, \dots, u_m).$$

Then U is a subspace of V, and $u_1, \dots, u_m$ is a basis of U: it spans U by definition and is linearly independent by hypothesis. In particular U is finite-dimensional, with $\dim U = m$. Restricting the inner product of V to $U \times U$ makes U an inner product space, since all the defining properties in 6.2 are inherited by a subspace.
By the linear map lemma (3.4) there is a unique linear map $\varphi \colon U \to \mathbf{F}$ with

$$\varphi(u_k) = 1 \quad \text{for each } k \in \{1, \dots, m\};$$

the lemma applies because $u_1, \dots, u_m$ is a basis of U, so the values of a linear map on this list may be

prescribed arbitrarily. Thus φ is a linear functional on the finite-dimensional inner product space U. By the Riesz representation theorem (6.42) applied to U and φ, there is a vector $v \in U$ such that

$$\varphi(u) = \langle u, v\rangle \quad \text{for every } u \in U,$$

where the inner product is the one U inherits from V, so this is also the inner product of V.
Since $v \in U \subseteq V$, this v is a vector in V, and for each $k \in \{1, \dots, m\}$ we have $u_k \in U$, so

$$\langle u_k, v\rangle = \varphi(u_k) = 1,$$

as required.
Note that V itself need not be finite-dimensional; the Riesz representation theorem was applied only to the finite-dimensional subspace U. Note also that linear independence is essential: if, say, $u_2 = 2u_1$, then $\langle u_2, v\rangle = 2\langle u_1, v\rangle$ for every v, so $\langle u_1, v\rangle$ and $\langle u_2, v\rangle$ cannot both equal 1.
Concretely, the vector v produced above can be described using the Gram–Schmidt procedure: if $e_1, \dots, e_m$ is an orthonormal basis of U obtained from $u_1, \dots, u_m$ by 6.32, then formula 6.43 in the proof of the Riesz representation theorem gives

$$v = \overline{\varphi(e_1)}\, e_1 + \cdots + \overline{\varphi(e_m)}\, e_m.$$

Problem 6B.19. Suppose $v_1, \dots, v_n$ is a basis of V. Prove that there exists a basis $u_1, \dots, u_n$ of V such that

$$\langle v_j, u_k\rangle = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k. \end{cases}$$

Solution. Because V has a basis of length n, the space V is finite-dimensional with $\dim V = n$.
Construction of the list. Fix $k \in \{1, \dots, n\}$. By the linear map lemma (3.4) there is a unique linear map $\varphi_k \colon V \to \mathbf{F}$ satisfying

$$\varphi_k(v_j) = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k \end{cases}$$

for each $j \in \{1, \dots, n\}$; the lemma applies because $v_1, \dots, v_n$ is a basis of V, so the values on this list may be prescribed arbitrarily. Thus $\varphi_1, \dots, \varphi_n$ are linear functionals on V (in the language of 3F they form the dual basis of $v_1, \dots, v_n$).
Since V is finite-dimensional, the Riesz representation theorem (6.42) gives, for each k, a unique vector $u_k \in V$ with

$$\varphi_k(u) = \langle u, u_k\rangle \quad \text{for every } u \in V.$$

Taking $u = v_j$ we obtain exactly the required relations:

$$\langle v_j, u_k\rangle = \varphi_k(v_j) = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k. \end{cases}$$

The list $u_1, \dots, u_n$ is a basis. Suppose $a_1, \dots, a_n \in \mathbf{F}$ satisfy

$$a_1 u_1 + \cdots + a_n u_n = 0.$$

Fix $j \in \{1, \dots, n\}$ and take the inner product of both sides with v_j in the first slot. Using conjugate linearity of the inner product in the second slot [6.6(d) and 6.6(e)],

$$0 = \langle v_j, a_1 u_1 + \cdots + a_n u_n\rangle = \overline{a_1}\langle v_j, u_1\rangle + \cdots + \overline{a_n}\langle v_j, u_n\rangle = \overline{a_j},$$

because every term with index different from j vanishes and the j-th term equals $\overline{a_j} \cdot 1$. Hence $a_j = 0$. As j was arbitrary, $a_1 = \cdots = a_n = 0$, so $u_1, \dots, u_n$ is linearly independent.
A linearly independent list of length $\dim V$ in V is a basis of V (2.38). Since $u_1, \dots, u_n$ has length $n = \dim V$, it is a basis of V, completing the proof.

Remark. The list $u_1, \dots, u_n$ is uniquely determined by $v_1, \dots, v_n$, because the Riesz representation theorem produces a unique u_k for each φ_k, and each φ_k is determined by its values on the basis $v_1, \dots, v_n$. The list $u_1, \dots, u_n$ is sometimes called the dual basis of $v_1, \dots, v_n$ with respect to the inner product. If $v_1, \dots, v_n$ happens to be orthonormal, then $u_k = v_k$ for each k, since the required relations are then satisfied by the v_k themselves and the list is unique.

Problem 6B.20. Suppose $\mathbf{F} = \mathbf{C}$, V is finite-dimensional, and $\mathcal{E} \subseteq \mathcal{L}(V)$ is such that

$$ST = TS$$

for all $S, T \in \mathcal{E}$. Prove that there is an orthonormal basis of V with respect to which every element of $\mathcal{E}$ has an upper-triangular matrix.
This exercise strengthens Exercise 9(b) in Section 5E (in the context of inner product spaces) by asserting that the basis in that exercise can be chosen to be orthonormal.

Solution. If $V = \{0\}$ the empty list is an orthonormal basis of V and every operator on V has the (empty) upper-triangular matrix with respect to it, so assume $V \neq \{0\}$ and set $n = \dim V \geq 1$.
Step 1: an ordinary basis that works for all of $\mathcal{E}$. By Exercise 9(b) in Section 5E, applied to the finite-dimensional nonzero complex vector space V and the commuting family $\mathcal{E}$, there is a basis

$$v_1, \dots, v_n$$

of V with respect to which every element of $\mathcal{E}$ has an upper-triangular matrix.
For completeness, here is the argument for that exercise. First one shows that some nonzero $v \in V$ is an eigenvector of every element of $\mathcal{E}$, by induction on $\dim V$: if every element of $\mathcal{E}$ is a scalar multiple of the identity, any nonzero vector works; otherwise choose $T \in \mathcal{E}$ that is not a scalar multiple of the identity, let λ be an eigenvalue of T (which exists by 5.19 because $\mathbf{F} = \mathbf{C}$ and $V \neq \{0\}$), and put $U = \operatorname{null}(T - \lambda I)$, so that $U \neq \{0\}$ and $U \neq V$. For $S \in \mathcal{E}$ and $u \in U$ we have $(T - \lambda I)(Su) = S(T - \lambda I)u = 0$, so U is invariant under S; thus $\{S|_U : S \in \mathcal{E}\}$ is a commuting family of operators on U, and since $0 < \dim U < \dim V$ the induction hypothesis supplies a nonzero $v \in U$ that is an eigenvector of every $S|_U$, hence of every $S \in \mathcal{E}$. Now induct on $\dim V$ again to get the basis: take such a common eigenvector v_1, put $W = \operatorname{span}(v_1)$, and pass to the quotient space V/W. Each $T \in \mathcal{E}$ satisfies $Tv_1 = \lambda_T v_1$ for some scalar λ_T, so W is invariant under T and the quotient operator $T/W \in \mathcal{L}(V/W)$ is defined (Exercise 38 in Section 5A); these quotient operators commute, because $(S/W)(T/W)(v + W) = STv + W = TSv + W = (T/W)(S/W)(v + W)$ for all $v \in V$. Since $\dim(V/W) = n - 1 < n$ by 3.105, the induction hypothesis gives a basis of V/W, of length $n - 1$, with respect to which every T/W is upper triangular; write it as $v_2 + W, \dots, v_n + W$ with $v_2, \dots, v_n \in V$. Then $v_1, \dots, v_n$ is a basis of V: if $a_1 v_1 + \cdots + a_n v_n = 0$, applying the quotient map $v \mapsto v + W$ and using $v_1 + W = W$ gives $a_2(v_2 + W) + \cdots + a_n(v_n + W) = W$, so $a_2 = \cdots = a_n = 0$ by linear independence in V/W, whence $a_1 v_1 = 0$ and $a_1 = 0$ because $v_1 \neq 0$; a linearly independent list of length $n = \dim V$ is a basis by 2.38. Now fix $T \in \mathcal{E}$ and $k \in \{1, \dots, n\}$. For $k = 1$ the subspace $\operatorname{span}(v_1) = W$ is invariant under T, as already noted. For $k \geq 2$, the subspace $\operatorname{span}(v_1, \dots, v_k)$ is exactly the inverse image under the quotient map $\pi\colon v \mapsto v + W$ of $\operatorname{span}(v_2 + W, \dots, v_k + W)$: the inverse image contains W and $v_2, \dots, v_k$, hence contains $\operatorname{span}(v_1, \dots, v_k)$, and conversely if $\pi(v) = a_2(v_2 + W) + \cdots + a_k(v_k + W)$ then $v - (a_2 v_2 + \cdots + a_k v_k) \in W = \operatorname{span}(v_1)$, so $v \in \operatorname{span}(v_1, \dots, v_k)$. Since $\operatorname{span}(v_2 + W, \dots, v_k + W)$ is invariant under T/W by 5.39, and $\pi(Tv) = (T/W)(\pi(v))$ for all v, this inverse image is invariant under T. So $\operatorname{span}(v_1, \dots, v_k)$ is invariant under T for every k, and by the criterion recalled in Step 2, every $T \in \mathcal{E}$ is upper triangular with respect to $v_1, \dots, v_n$.
Step 2: the invariant-subspace reformulation. By 5.39, for an operator $T \in \mathcal{L}(V)$ and a basis $v_1, \dots, v_n$ of V, the matrix of T with respect to $v_1, \dots, v_n$ is upper triangular if and only if

$$\operatorname{span}(v_1, \dots, v_k) \text{ is invariant under } T \quad \text{for each } k \in \{1, \dots, n\}.$$

Applying this to each element of $\mathcal{E}$ and the basis from Step 1, we conclude that

$$\operatorname{span}(v_1, \dots, v_k) \text{ is invariant under every } T \in \mathcal{E}, \quad \text{for each } k \in \{1, \dots, n\}.$$

Step 3: orthonormalize. Apply the Gram–Schmidt procedure (6.32) to the linearly independent list $v_1, \dots, v_n$, producing an orthonormal list $e_1, \dots, e_n$ in V with

$$\operatorname{span}(e_1, \dots, e_k) = \operatorname{span}(v_1, \dots, v_k) \quad \text{for each } k \in \{1, \dots, n\};$$

this equality of spans is part of the statement of 6.32. Because $e_1, \dots, e_n$ is an orthonormal list of length $n = \dim V$, it is an orthonormal basis of V (orthonormal lists are linearly independent by 6.25, and a linearly independent list of length $\dim V$ is a basis by 2.38; this is 6.28).
Step 4: conclusion. Let $T \in \mathcal{E}$ and let $k \in \{1, \dots, n\}$. By Step 2 the subspace $\operatorname{span}(v_1, \dots, v_k)$ is invariant under T, and by Step 3 this subspace equals $\operatorname{span}(e_1, \dots, e_k)$. Hence $\operatorname{span}(e_1, \dots, e_k)$ is invariant under T for each k, and so by 5.39 again the matrix of T with respect to the basis $e_1, \dots, e_n$ is upper triangular.
Since $T \in \mathcal{E}$ was arbitrary, every element of $\mathcal{E}$ has an upper-triangular matrix with respect to the single orthonormal basis $e_1, \dots, e_n$ of V, as desired.
Note that the argument places no restriction on the size of $\mathcal{E}$: the family may be infinite, because Step 1 produces one basis serving all of $\mathcal{E}$ at once and Steps 2–4 are applied to each element separately. Taking $\mathcal{E} = \{T\}$ for a single operator T recovers Schur's theorem (6.38).

Problem 6B.21. Suppose $\mathbf{F} = \mathbf{C}$, V is finite-dimensional, $T \in \mathcal{L}(V)$, and all eigenvalues of T have absolute value less than 1. Let $\epsilon > 0$. Prove that there exists a positive integer m such that $\|T^m v\| \le \epsilon \|v\|$ for every $v \in V$.

Solution. If $V = \{0\}$ then every $v \in V$ is 0 and $m = 1$ works. So assume $n = \dim V \ge 1$.
Step 1: an orthonormal basis making T upper triangular. By Schur's theorem (6.38) there is an orthonormal basis $e_1, \dots, e_n$ of V with respect to which T has an upper-triangular matrix A; so $A_{j,k} = 0$ when $j > k$, and

$$Te_k = \sum_{j \le k} A_{j,k} e_j.$$

By 5.41 the diagonal entries $A_{1,1}, \dots, A_{n,n}$ are precisely the eigenvalues of T. Set

$$r = \max\{|A_{1,1}|, \dots, |A_{n,n}|\}.$$

By hypothesis every eigenvalue of T has absolute value less than 1, so $r < 1$. Also set $M = \max\{|A_{j,k}| : 1 \le j < k \le n\}$ if $n \ge 2$, and $M = 0$ if $n = 1$.
Step 2: rescale the basis to shrink the entries above the diagonal. Let t be a real number with $0 < t \le 1$, to be chosen shortly, and define

$$f_k = t^k e_k \quad \text{for } k \in \{1, \dots, n\}.$$

Since each f_k is a nonzero scalar multiple of e_k, the list $f_1, \dots, f_n$ is a basis of V. Computing,

$$Tf_k = t^k Te_k = t^k \sum_{j \le k} A_{j,k} e_j = \sum_{j \le k} A_{j,k}\, t^{k-j} f_j,$$

using $e_j = t^{-j} f_j$. So the matrix B of T with respect to $f_1, \dots, f_n$ is upper triangular with

$$B_{j,k} = A_{j,k}\, t^{k-j}.$$

In particular $B_{k,k} = A_{k,k}$, while for $j < k$ we have $|B_{j,k}| \le Mt$, because $0 < t \le 1$ and $k - j \ge 1$.
Step 3: a new inner product. Define an inner product $\langle \cdot, \cdot \rangle'$ on V by declaring $f_1, \dots, f_n$ to be orthonormal, that is,

$$\Big\langle \sum_{k=1}^n a_k f_k,\ \sum_{k=1}^n b_k f_k \Big\rangle' = \sum_{k=1}^n a_k \overline{b_k}.$$

This is an inner product: transporting the standard inner product of $\mathbf{C}^n$ through the isomorphism $V \to \mathbf{C}^n$ that sends $\sum a_k f_k$ to $(a_1, \dots, a_n)$ verifies all the properties in 6.2. Write $\|\cdot\|'$ for the associated norm, so $\big\| \sum a_k f_k \big\|' = \big(\sum |a_k|^2 \big)^{1/2}$.

Step 4: T is a strict contraction for $\|\cdot\|'$. Split $B = D + E$, where D is the diagonal matrix with entries $A_{1,1}, \dots, A_{n,n}$ and E is strictly upper triangular with entries $E_{j,k} = B_{j,k}$ for $j < k$ and $E_{j,k} = 0$ otherwise. Let $S, N \in \mathcal{L}(V)$ be the operators whose matrices with respect to $f_1, \dots, f_n$ are D and E; then $T = S + N$.
Let $v = \sum_{k=1}^{n} a_k f_k$. Then $Sv = \sum_k A_{k,k} a_k f_k$, so

$$\|Sv\|' = \Big(\sum_{k=1}^{n} |A_{k,k}|^2 |a_k|^2 \Big)^{1/2} \le r \Big(\sum_{k=1}^{n} |a_k|^2 \Big)^{1/2} = r \|v\|'.$$

For N, the j-th coordinate of Nv with respect to $f_1, \dots, f_n$ is $\sum_{k>j} E_{j,k} a_k$, so by the triangle inequality and then the Cauchy–Schwarz inequality (6.14) in $\mathbf{R}^n$,

$$\Big| \sum_{k>j} E_{j,k} a_k \Big| \le Mt \sum_{k=1}^{n} |a_k| \le Mt\sqrt{n}\, \|v\|'.$$

Squaring and summing over the n values of j,

$$\big(\|Nv\|' \big)^2 \le n \cdot M^2 t^2 n \, \big(\|v\|' \big)^2, \qquad \text{so} \qquad \|Nv\|' \le nMt \, \|v\|'.$$

Everything above holds for each fixed t with $0 < t \le 1$ (the basis $f_1, \dots, f_n$, the inner product $\langle \cdot, \cdot \rangle'$, and the operators S and N all depend on t, and the two displayed bounds hold for every such t). Now specialize to

$$t = \min \left\{ 1, \ \frac{1 - r}{2n(M+1)} \right\},$$

which is a positive number at most 1 (positive because $r < 1$ and $M \ge 0$). By definition of the minimum, $t \le \frac{1-r}{2n(M+1)}$ in either case, so

$$nMt \le nM \cdot \frac{1-r}{2n(M+1)} = \frac{1-r}{2} \cdot \frac{M}{M+1} \le \frac{1-r}{2}.$$

With this choice, using the triangle inequality (6.17) for $\|\cdot\|'$, for every $v \in V$,

$$\|Tv\|' \le \|Sv\|' + \|Nv\|' \le \left(r + \frac{1-r}{2} \right) \|v\|' = \rho \|v\|', \qquad \text{where } \rho = \frac{1+r}{2}.$$

Since $0 \le r < 1$, we have $\frac{1}{2} \le \rho < 1$. Iterating the inequality gives, for every positive integer m and every $v \in V$,

$$\|T^m v\|' \le \rho^m \|v\|'.$$

Step 5: return to the original norm. Both $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle'$ are inner products on the finite-dimensional space V, so Exercise 16 of this section supplies positive numbers c_1 and c_2 with

$$\|w\| \le c_1 \|w\|' \quad \text{and} \quad \|w\|' \le c_2 \|w\| \quad \text{for every } w \in V$$

(the first inequality is Exercise 16 with the roles of the two inner products as written, the second with the roles interchanged). Hence for every positive integer m and every $v \in V$,

$$\|T^m v\| \le c_1 \|T^m v\|' \le c_1 \rho^m \|v\|' \le c_1 c_2 \, \rho^m \|v\|.$$

Because $0 \le \rho < 1$, we have $\rho^m \to 0$ as $m \to \infty$, so there is a positive integer m with

$$c_1 c_2 \, \rho^m \le \epsilon$$

(here $\rho = (1+r)/2 \ge 1/2 > 0$, so $\log \rho$ is a negative real number and any integer $m > \log\big(\epsilon/(c_1 c_2)\big) / \log \rho$ will do). For this m,

$$\|T^m v\| \le \epsilon \|v\| \quad \text{for every } v \in V,$$

completing the proof.

Remark. The heart of the matter is Step 2: an operator all of whose eigenvalues are small need not itself shrink vectors in the given norm, as the matrix $\begin{pmatrix} 0 & 100 \\ 0 & 0 \end{pmatrix}$ on $\mathbf{C}^2$ shows. But rescaling the basis by increasing powers of a small t makes the entries above the diagonal as small as we like without changing the diagonal, and then the operator genuinely contracts with respect to the inner product for which the rescaled basis is orthonormal. Exercise 16 converts that contraction back into the original norm at the cost of one fixed constant, which the factor ρ^m eventually absorbs.

Problem 6B.22. Suppose $C[-1,1]$ is the vector space of continuous real-valued functions on the interval $[-1,1]$ with inner product given by

$$\langle f, g \rangle = \int_{-1}^{1} fg$$

for all $f, g \in C[-1,1]$. Let φ be the linear functional on $C[-1,1]$ defined by $\varphi(f) = f(0)$. Show that there does not exist $g \in C[-1,1]$ such that

$$\varphi(f) = \langle f, g \rangle$$

for every $f \in C[-1,1]$.
This exercise shows that the Riesz representation theorem (6.42) does not hold on infinite-dimensional vector spaces without additional hypotheses on V and φ.

Solution. We first isolate the standard fact about integrals of nonnegative continuous functions that the argument needs.
Lemma: suppose $h \in C[-1,1]$, $h(x) \geq 0$ for all $x \in [-1,1]$, and $\int_{-1}^{1} h = 0$. Then $h(x) = 0$ for all $x \in [-1,1]$.
To prove the lemma, suppose not, so $h(c) > 0$ for some $c \in [-1,1]$. Because h is continuous at c, there exists $\delta > 0$ such that

$$h(x) > \tfrac{1}{2}h(c) \quad \text{for all } x \in [-1,1] \text{ with } |x - c| < \delta.$$

The set $I = [-1,1] \cap (c-\delta, c+\delta)$ contains a closed interval $[\alpha, \beta] \subseteq [-1,1]$ of positive length $\beta - \alpha > 0$ (if c is an endpoint of $[-1,1]$, take the half of the neighborhood lying inside $[-1,1]$; here we use that $[-1,1]$ has positive length). Because $h \geq 0$ on all of $[-1,1]$ and $h > \frac{1}{2}h(c)$ on $[\alpha, \beta]$, monotonicity of the integral gives

$$\int_{-1}^{1} h \;\geq\; \int_{\alpha}^{\beta} h \;\geq\; \tfrac{1}{2}h(c)(\beta - \alpha) \;>\; 0,$$

contradicting $\int_{-1}^{1} h = 0$. This proves the lemma. □
Now suppose, for the sake of contradiction, that there exists $g \in C[-1,1]$ such that

$$f(0) = \varphi(f) = \langle f, g \rangle = \int_{-1}^{1} fg$$

for every $f \in C[-1,1]$.
Define $f\colon [-1,1] \to \mathbb{R}$ by

$$f(x) = x^2 g(x).$$

Then f is continuous (a product of continuous functions), so $f \in C[-1,1]$, and $f(0) = 0$. Applying the displayed identity to this f gives

$$0 = f(0) = \int_{-1}^{1} fg = \int_{-1}^{1} x^2 g(x)^2 \, dx.$$

The function $h(x) = x^2 g(x)^2$ is continuous on $[-1,1]$ and satisfies $h(x) \geq 0$ for all x, and we have just shown $\int_{-1}^{1} h = 0$. By the lemma, $h \equiv 0$ on $[-1,1]$; that is, $x^2 g(x)^2 = 0$ for all $x \in [-1,1]$. For $x \neq 0$

we have $x^2 \neq 0$, and hence $g(x) = 0$. Thus g vanishes on $[-1, 1] \setminus \{0\}$, and by continuity of g at 0,

$$g(0) = \lim_{x \to 0} g(x) = 0.$$

Hence $g = 0$, the zero function.
But then for every $f \in C[-1, 1]$ we would have

$$f(0) = \varphi(f) = \langle f, 0 \rangle = 0,$$

which is false: taking f to be the constant function 1 (which lies in $C[-1, 1]$) gives $f(0) = 1 \neq 0$.
This contradiction shows that no such $g \in C[-1, 1]$ exists.
Remark: the infinite-dimensional space $C[-1, 1]$ is genuinely infinite-dimensional (it contains the linearly independent list $1, x, x^2, \ldots$), so the finite-dimensionality hypothesis of the Riesz representation theorem (6.42) is not satisfied, and the conclusion of that theorem indeed fails here. Intuitively, a representing g would have to be a spike at 0 of unit mass – a Dirac delta – which is not a continuous function. The failure is not merely one of dimension: the linear functional φ is unbounded with respect to the norm coming from this inner product. Indeed, choosing continuous f_n with $f_n(0) = 1$, $0 \leq f_n \leq 1$, and $f_n = 0$ outside $[-1/n, 1/n]$, one gets $\varphi(f_n) = 1$ while $\|f_n\|^2 = \int_{-1}^{1} f_n^2 \leq 2/n \to 0$. A Riesz-type theorem does hold in infinite dimensions once one assumes completeness of V and continuity (boundedness) of φ.

Problem 6B.23. For all $u, v \in V$, define $d(u, v) = \|u - v\|$.

- (a) Show that d is a metric on V.
- (b) Show that if V is finite-dimensional, then d is a complete metric on V (meaning that every Cauchy sequence converges).
- (c) Show that every finite-dimensional subspace of V is a closed subset of V (with respect to the metric d).

This exercise requires familiarity with metric spaces.

Solution. Throughout, V is an inner product space over $\mathbf{F}$, where $\mathbf{F}$ is $\mathbb{R}$ or $\mathbb{C}$, and $\|v\| = \sqrt{\langle v, v \rangle}$.
(a) We verify the three metric axioms for $d(u, v) = \|u - v\|$ on $V \times V$. Note first that $\|w\| = \sqrt{\langle w, w \rangle}$ is a well-defined nonnegative real number for each $w \in V$, because positivity of the inner product gives $\langle w, w \rangle \geq 0$. Hence $d(u, v) \geq 0$ for all $u, v \in V$.
Positive definiteness. If $u = v$ then $d(u, v) = \|0\| = 0$. Conversely, if $d(u, v) = 0$ then $\|u - v\| = 0$, so $u - v = 0$ by 6.9(a), which gives $u = v$. Thus $d(u, v) = 0$ if and only if $u = v$.
Symmetry. For all $u, v \in V$, using 6.9(b) with $\lambda = -1$,

$$d(u, v) = \|u - v\| = \|(-1)(v - u)\| = |-1| \, \|v - u\| = \|v - u\| = d(v, u).$$

Triangle inequality. Let $u, v, w \in V$. Then $u - w = (u - v) + (v - w)$, so the triangle inequality for the norm (6.17) gives

$$d(u, w) = \|u - w\| = \|(u - v) + (v - w)\| \leq \|u - v\| + \|v - w\| = d(u, v) + d(v, w).$$

Hence d is a metric on V.
(b) Suppose V is finite-dimensional, say $\dim V = n$. If $n = 0$ then $V = \{0\}$ and every sequence in V is constant, hence convergent; so assume $n \geq 1$. By 6.35 there is an orthonormal basis $e_1, \ldots, e_n$ of V. Let $v_1, v_2, \ldots$ be a Cauchy sequence in (V, d). For each $k \in \{1, \ldots, n\}$ and each m, set

$$a_{k,m} = \langle v_m, e_k \rangle \in \mathbf{F},$$

so that $v_m = a_{1,m} e_1 + \cdots + a_{n,m} e_n$ by 6.30(a).
Fix k. For all m, j, the Cauchy-Schwarz inequality (6.14) together with $\|e_k\| = 1$ gives

$$|a_{k,m} - a_{k,j}| = |\langle v_m - v_j, e_k \rangle| \leq \|v_m - v_j\| \, \|e_k\| = d(v_m, v_j),$$

where the first equality uses additivity of the inner product in its first slot. Because (v_m) is Cauchy in V, the right side is small for large m, j; precisely, given $\varepsilon > 0$ there is M with $d(v_m, v_j) < \varepsilon$ for all $m, j \geq M$, and then $|a_{k,m} - a_{k,j}| < \varepsilon$ for all $m, j \geq M$. Thus $(a_{k,m})_{m=1}^{\infty}$ is a Cauchy sequence of scalars. Since $\mathbb{R}$ and $\mathbb{C}$ are complete metric spaces, this sequence converges: there exists $a_k \in \mathbf{F}$ with

$$\lim_{m \to \infty} a_{k,m} = a_k.$$

Define $v = a_1 e_1 + \cdots + a_n e_n \in V$. Because $e_1, \ldots, e_n$ is orthonormal, taking the inner product of this equation with e_k gives $\langle v, e_k \rangle = a_k$ for each k. Hence for each m,

$$v_m - v = (a_{1,m} - a_1)e_1 + \cdots + (a_{n,m} - a_n)e_n,$$

and $\langle v_m - v, e_k \rangle = a_{k,m} - a_k$ for each k. Thus 6.30(b), applied to the vector $v_m - v$ and the orthonormal basis $e_1, \ldots, e_n$, yields

$$d(v_m, v)^2 = \|v_m - v\|^2 = \sum_{k=1}^{n} |a_{k,m} - a_k|^2.$$

Each of the n terms on the right tends to 0 as $m \to \infty$, and n is a fixed finite number, so the whole sum tends to 0. Explicitly: given $\varepsilon > 0$, choose for each k an index M_k with $|a_{k,m} - a_k| < \varepsilon/\sqrt{n}$ for all $m \geq M_k$, and let $M = \max\{M_1, \ldots, M_n\}$; then $d(v_m, v)^2 < n \cdot \varepsilon^2/n = \varepsilon^2$ for all $m \geq M$, so $d(v_m, v) < \varepsilon$.
Hence $v_m \to v$ in (V, d). Every Cauchy sequence in V converges, so d is a complete metric on V.
(c) Let U be a finite-dimensional subspace of V. Restricting the inner product of V to $U \times U$ makes U an inner product space in its own right (all the required properties are inherited, and U is closed under the vector space operations), and the norm on U so obtained is the restriction of the norm on V. Consequently the metric that U inherits as a subset of the metric space (V, d) is exactly the metric $d_U(u, u') = \|u - u'\|$ furnished by part (a) for the inner product space U. Since U is finite-dimensional, part (b) says (U, d_U) is complete.
To show U is a closed subset of V, let $v \in V$ belong to the closure of U; we must show $v \in U$. Because v is in the closure of U, for each positive integer m there exists $u_m \in U$ with $d(u_m, v) < 1/m$; thus $u_m \to v$ in V.
The sequence $u_1, u_2, \ldots$ is Cauchy: given $\varepsilon > 0$, choose M with $1/M < \varepsilon/2$; then for all $m, j \geq M$ the triangle inequality gives

$$d(u_m, u_j) \leq d(u_m, v) + d(v, u_j) < \tfrac{1}{m} + \tfrac{1}{j} \leq \tfrac{2}{M} < \varepsilon.$$

All the terms u_m lie in U, so this is a Cauchy sequence in the complete metric space (U, d_U). Hence there exists $u \in U$ with $d(u_m, u) \to 0$.
Finally, for every m,

$$d(u, v) \leq d(u, u_m) + d(u_m, v),$$

and both terms on the right tend to 0 as $m \to \infty$. Since $d(u, v)$ does not depend on m and is bounded above by quantities tending to 0, we get $d(u, v) = 0$, hence $u = v$ by part (a). Therefore $v = u \in U$.
Thus the closure of U is contained in U, so U is a closed subset of V.
Remark: the finite-dimensionality hypotheses in (b) and (c) cannot be dropped. In the inner product space $C[-1, 1]$ of 6B.22, the subspace consisting of all polynomial functions is infinite-dimensional and is not a closed subset, since by the Weierstrass approximation theorem its closure is all of $C[-1, 1]$. Likewise the conclusion of (b) can fail for an infinite-dimensional inner product space: $C[-1, 1]$ with the integral inner product is not complete.

6.3 Exercises 6C

Problem 6C.1. Suppose $v_1, \dots, v_m \in V$. Prove that

$$\{v_1, \dots, v_m\}^\perp = \big(\operatorname{span}(v_1, \dots, v_m)\big)^\perp.$$

Solution. Recall that for a subset S of V, Axler defines $S^\perp = \{v \in V : \langle u, v\rangle = 0 \text{ for every } u \in S\}$. Write $U = \operatorname{span}(v_1, \dots, v_m)$.
First we prove $U^\perp \subseteq \{v_1, \dots, v_m\}^\perp$. Each v_k lies in U, so $\{v_1, \dots, v_m\} \subseteq U$; hence 6.48(e) (which states that $G \subseteq H$ implies $H^\perp \subseteq G^\perp$) gives $U^\perp \subseteq \{v_1, \dots, v_m\}^\perp$.
Now we prove the reverse inclusion. Suppose $w \in \{v_1, \dots, v_m\}^\perp$, which means

$$\langle v_k, w\rangle = 0 \quad \text{for each } k = 1, \dots, m.$$

Let $u \in U$. Then there exist $a_1, \dots, a_m \in \mathbf{F}$ with $u = a_1 v_1 + \cdots + a_m v_m$. Using additivity in the first slot and homogeneity in the first slot of the inner product,

$$\langle u, w\rangle = \langle a_1 v_1 + \cdots + a_m v_m,\, w\rangle = a_1\langle v_1, w\rangle + \cdots + a_m\langle v_m, w\rangle = a_1 \cdot 0 + \cdots + a_m \cdot 0 = 0.$$

Because $u \in U$ was arbitrary, $w \in U^\perp$. Thus $\{v_1, \dots, v_m\}^\perp \subseteq U^\perp$.
Combining the two inclusions gives $\{v_1, \dots, v_m\}^\perp = \big(\operatorname{span}(v_1, \dots, v_m)\big)^\perp$, as desired.
The content of this exercise is the useful principle that to check that a vector is orthogonal to a subspace, it suffices to check orthogonality against the vectors of any spanning list of that subspace.

Problem 6C.2. Suppose U is a subspace of V with basis $u_1, \dots, u_m$ and

$$u_1, \dots, u_m, v_1, \dots, v_n$$

is a basis of V. Prove that if the Gram–Schmidt procedure is applied to the basis of V above, producing a list $e_1, \dots, e_m, f_1, \dots, f_n$, then $e_1, \dots, e_m$ is an orthonormal basis of U and $f_1, \dots, f_n$ is an orthonormal basis of $U^\perp$.

Solution. Because V has the basis $u_1, \dots, u_m, v_1, \dots, v_n$, the space V is finite-dimensional with $\dim V = m + n$; also $\dim U = m$.
Write the given basis of V as the single list $w_1, \dots, w_{m+n}$, where $w_k = u_k$ for $k = 1, \dots, m$ and $w_{m+j} = v_j$ for $j = 1, \dots, n$. Applying the Gram–Schmidt procedure (6.32) to this linearly independent list produces the orthonormal list

$$g_1, \dots, g_{m+n}, \qquad \text{where } g_k = e_k \text{ for } k \le m \text{ and } g_{m+j} = f_j,$$

and 6.32 guarantees that

$$\operatorname{span}(g_1, \dots, g_k) = \operatorname{span}(w_1, \dots, w_k) \quad \text{for each } k = 1, \dots, m + n.$$

Step 1: the list $e_1, \dots, e_m$ is an orthonormal basis of U. Taking $k = m$ in the span equality above gives

$$\operatorname{span}(e_1, \dots, e_m) = \operatorname{span}(u_1, \dots, u_m) = U,$$

the last equality because $u_1, \dots, u_m$ is a basis of U. The list $e_1, \dots, e_m$ is orthonormal, hence linearly independent by 6.25. A linearly independent spanning list of U is a basis of U; thus $e_1, \dots, e_m$ is an orthonormal basis of U.
Step 2: each f_j lies in $U^\perp$. Because $g_1, \dots, g_{m+n}$ is an orthonormal list, we have $\langle e_k, f_j\rangle = 0$ for all $k \in \{1, \dots, m\}$ and all $j \in \{1, \dots, n\}$. Thus $f_j \in \{e_1, \dots, e_m\}^\perp$, and by Exercise 6C.1 together with Step 1,

$$\{e_1, \dots, e_m\}^\perp = \big(\operatorname{span}(e_1, \dots, e_m)\big)^\perp = U^\perp.$$

Hence $f_j \in U^\perp$ for each $j = 1, \dots, n$.
Step 3: the list $f_1, \dots, f_n$ is an orthonormal basis of $U^\perp$. This list is an orthonormal list of vectors in the inner product space $U^\perp$ (with the inner product inherited from V), by Step 2. By 6.51,

$$\dim U^\perp = \dim V - \dim U = (m+n) - m = n,$$

so $f_1, \dots, f_n$ is an orthonormal list in $U^\perp$ of length $\dim U^\perp$. By 6.28 applied to the inner product space $U^\perp$, this list is an orthonormal basis of $U^\perp$.
(Alternatively, for Step 3 one can argue without 6.51: $f_1, \dots, f_n$ is linearly independent by 6.25, so it suffices to show it spans $U^\perp$. If $w \in U^\perp \subseteq V$, then since $g_1, \dots, g_{m+n}$ is an orthonormal basis of V by 6.28, we have $w = \sum_{k=1}^{m} \langle w, e_k\rangle e_k + \sum_{j=1}^{n} \langle w, f_j\rangle f_j$ by 6.30(a); each $e_k \in U$ and $w \in U^\perp$ force $\langle e_k, w\rangle = 0$ and hence $\langle w, e_k\rangle = 0$ by conjugate symmetry, so $w \in \operatorname{span}(f_1, \dots, f_n)$.)

Problem 6C.3. Suppose U is the subspace of $\mathbf{R}^4$ defined by

$$U = \operatorname{span}\big((1,2,3,-4),\ (-5,4,3,2)\big).$$

Find an orthonormal basis of U and an orthonormal basis of $U^\perp$.

Solution. Throughout, $\mathbf{R}^4$ carries the Euclidean inner product $\langle x, y\rangle = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4$.
Put $v_1 = (1,2,3,-4)$ and $v_2 = (-5,4,3,2)$. Neither is a scalar multiple of the other, so v_1, v_2 is linearly independent and hence a basis of U; thus $\dim U = 2$ and, by 6.51, $\dim U^\perp = 4 - 2 = 2$.
An orthonormal basis of U. Apply the Gram–Schmidt procedure (6.32) to v_1, v_2. First,

$$\|v_1\|^2 = 1 + 4 + 9 + 16 = 30, \qquad \langle v_2, v_1\rangle = -5 + 8 + 9 - 8 = 4.$$

Set $g_1 = v_1$ and

$$g_2 = v_2 - \frac{\langle v_2, g_1\rangle}{\|g_1\|^2} g_1 = (-5,4,3,2) - \frac{4}{30}(1,2,3,-4) = \frac{1}{15}(-77, 56, 39, 38).$$

Indeed $-5 - \frac{2}{15} = -\frac{77}{15}$, $4 - \frac{4}{15} = \frac{56}{15}$, $3 - \frac{6}{15} = \frac{39}{15}$, and $2 + \frac{8}{15} = \frac{38}{15}$. As a check, $\langle (-77,56,39,38), v_1\rangle = -77 + 112 + 117 - 152 = 0$. Also

$$\|(-77,56,39,38)\|^2 = 5929 + 3136 + 1521 + 1444 = 12030.$$

Normalizing g_1 and g_2 (a positive scalar multiple does not change the direction, so we may normalize $(-77,56,39,38)$ itself) gives the orthonormal basis of U

$$\frac{1}{\sqrt{30}}(1,2,3,-4), \qquad \frac{1}{\sqrt{12030}}(-77,56,39,38).$$

These two vectors are orthonormal and lie in $U = \operatorname{span}(v_1, v_2)$, and a list of two orthonormal vectors is linearly independent (6.25); since $\dim U = 2$, it is an orthonormal basis of U by 6.28.
An orthonormal basis of $U^\perp$. By Exercise 6C.1, $U^\perp = \{v_1, v_2\}^\perp$, so $x = (x_1, x_2, x_3, x_4) \in U^\perp$ exactly when

$$\begin{aligned} x_1 + 2x_2 + 3x_3 - 4x_4 &= 0, \\ -5x_1 + 4x_2 + 3x_3 + 2x_4 &= 0. \end{aligned}$$

Subtracting the second equation from the first gives $6x_1 - 2x_2 - 6x_4 = 0$, that is, $x_2 = 3x_1 - 3x_4$. Substituting into the first equation gives $7x_1 + 3x_3 - 10x_4 = 0$, that is, $x_3 = \frac{1}{3}(10x_4 - 7x_1)$. Writing $x_1 = 3a$ and $x_4 = 3b$ (with $a, b \in \mathbf{R}$ arbitrary) yields

$$x = a(3,9,-7,0) + b(0,-9,10,3).$$

Thus $U^\perp = \operatorname{span}(w_1, w_2)$ with $w_1 = (3,9,-7,0)$ and $w_2 = (0,-9,10,3)$; these are visibly linearly independent, matching $\dim U^\perp = 2$.

Now apply Gram–Schmidt to w_1, w_2. We compute

$$\|w_1\|^2 = 9 + 81 + 49 = 139, \qquad \langle w_2, w_1\rangle = 0 - 81 - 70 + 0 = -151.$$

Set $h_1 = w_1$ and

$$h_2 = w_2 + \frac{151}{139}w_1 = \frac{1}{139}(453, 108, 333, 417) = \frac{3}{139}(151, 36, 111, 139).$$

(Here $-9 \cdot 139 + 151 \cdot 9 = 108$, $10 \cdot 139 - 151 \cdot 7 = 333$, and $3 \cdot 139 = 417$.) As checks, $\langle (151, 36, 111, 139), w_1\rangle = 453+324-777+0 = 0$, while $\langle (151, 36, 111, 139), v_1\rangle = 151+72+333-556 = 0$ and $\langle (151, 36, 111, 139), v_2\rangle = -755+144+333+278 = 0$, confirming that this vector lies in $U^\perp$. Finally,

$$\|(151, 36, 111, 139)\|^2 = 22801 + 1296 + 12321 + 19321 = 55739.$$

Normalizing gives the orthonormal basis of $U^\perp$

$$\frac{1}{\sqrt{139}}(3, 9, -7, 0), \qquad \frac{1}{\sqrt{55739}}(151, 36, 111, 139).$$

As before, these are two orthonormal vectors in the 2-dimensional space $U^\perp$, hence an orthonormal basis of $U^\perp$ by 6.28.
(For the curious: $12030 = 30 \cdot 401$ and $55739 = 139 \cdot 401$, and 401 is the quotient of the Gram determinant $30 \cdot 54 - 4^2 = 1604$ by 4.)

Problem 6C.4. Suppose $e_1, \dots, e_n$ is a list of vectors in V with $\|e_k\| = 1$ for each $k = 1, \dots, n$ and

$$\|v\|^2 = \big|\langle v, e_1\rangle\big|^2 + \cdots + \big|\langle v, e_n\rangle\big|^2$$

for all $v \in V$. Prove that $e_1, \dots, e_n$ is an orthonormal basis of V.
[This exercise provides a converse to 6.30(b).]

Solution. There are two things to prove: that the list $e_1, \dots, e_n$ is orthonormal, and that it is a basis of V. Note that V is not assumed to be finite-dimensional; finite-dimensionality will come out of the hypothesis.
The list is orthonormal. By hypothesis $\|e_k\| = 1$ for each k, so only the orthogonality remains. Fix $j \in \{1, \dots, n\}$ and apply the hypothesis with $v = e_j$:

$$1 = \|e_j\|^2 = \sum_{k=1}^{n}\big|\langle e_j, e_k\rangle\big|^2 = \big|\langle e_j, e_j\rangle\big|^2 + \sum_{k\neq j}\big|\langle e_j, e_k\rangle\big|^2 = 1 + \sum_{k\neq j}\big|\langle e_j, e_k\rangle\big|^2,$$

where we used $\langle e_j, e_j\rangle = \|e_j\|^2 = 1$. Subtracting 1 from both ends gives

$$\sum_{k\neq j}\big|\langle e_j, e_k\rangle\big|^2 = 0.$$

This is a sum of nonnegative real numbers equal to 0, so every term is 0; hence $\langle e_j, e_k\rangle = 0$ whenever $k \neq j$. Since j was arbitrary, $e_1, \dots, e_n$ is an orthonormal list.
The list spans V. Let $U = \operatorname{span}(e_1, \dots, e_n)$, a finite-dimensional subspace of V. Suppose $v \in U^\perp$. By Exercise 6C.1, $U^\perp = \{e_1, \dots, e_n\}^\perp$, so $\langle e_k, v\rangle = 0$ for each k, and hence $\langle v, e_k\rangle = \overline{\langle e_k, v\rangle} = 0$ for each k by conjugate symmetry. The hypothesis then gives

$$\|v\|^2 = \sum_{k=1}^{n}\big|\langle v, e_k\rangle\big|^2 = 0,$$

so $v = 0$ by the definiteness of the inner product. Therefore $U^\perp = \{0\}$. Since U is a finite-dimensional subspace of V, 6.54 gives $U = V$. Thus $e_1, \dots, e_n$ spans V (and in particular V is finite-dimensional, with $\dim V \leq n$).

Conclusion. The list $e_1, \dots, e_n$ is orthonormal, hence linearly independent by 6.25, and it spans V. Therefore it is a basis of V, and being orthonormal it is an orthonormal basis of V.
Note that 6.30(b) says that for an orthonormal basis $e_1, \dots, e_n$ of V the displayed identity (Parseval's identity) holds for all $v \in V$; this exercise shows that, for a list of unit vectors, Parseval's identity conversely forces the list to be an orthonormal basis.

Problem 6C.5. Suppose that V is finite-dimensional and U is a subspace of V. Show that $P_{U^\perp} = I - P_U$, where I is the identity operator on V.

Solution. Because V is finite-dimensional, so are the subspaces U and $U^\perp$ (by 2.25). Hence both orthogonal projections P_U and $P_{U^\perp}$ are defined (6.55).
Let $v \in V$. By 6.49 applied to U, there are unique vectors $u \in U$ and $w \in U^\perp$ with

$$v = u + w,$$

and by the definition 6.55 of the orthogonal projection onto U we have $P_U v = u$.
Now rewrite this same decomposition so that it computes $P_{U^\perp} v$. Because U is finite-dimensional, 6.52 gives

$$(U^\perp)^\perp = U.$$

Therefore $u \in (U^\perp)^\perp$, and the equation

$$v = w + u \quad \text{with } w \in U^\perp \text{ and } u \in (U^\perp)^\perp$$

is exactly the decomposition of v required by definition 6.55 for the subspace $U^\perp$ (this decomposition is unique by 6.49 applied to $U^\perp$, which is legitimate since $U^\perp$ is finite-dimensional). Consequently

$$P_{U^\perp} v = w = v - u = v - P_U v = (I - P_U)v.$$

Since $v \in V$ was arbitrary, $P_{U^\perp} = I - P_U$.
Two remarks. First, what the proof actually used is that U and $U^\perp$ are both finite-dimensional (needed for 6.49, for 6.52, and for the definition 6.55 of each of the two projections); the hypothesis that V is finite-dimensional supplies this by 2.25. This is not genuinely more general, however: if U and $U^\perp$ are both finite-dimensional, then $V = U \oplus U^\perp$ by 6.49 and so V is finite-dimensional anyway. Second, applying the result twice and using 6.52 gives the expected symmetry: $P_U = I - P_{U^\perp} = P_{(U^\perp)^\perp}$.

Problem 6C.6. Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Show that

$$T = T P_{(\operatorname{null} T)^\perp} = P_{\operatorname{range} T}\, T.$$

Solution. Because V is finite-dimensional, the subspaces $\operatorname{null} T$ and $(\operatorname{null} T)^\perp$ of V are finite-dimensional, and $\operatorname{range} T$ is a finite-dimensional subspace of W (the fundamental theorem of linear maps, 3.21, explicitly asserts that $\operatorname{range} T$ is finite-dimensional). Hence the orthogonal projections $P_{(\operatorname{null} T)^\perp} \in \mathcal{L}(V)$ and $P_{\operatorname{range} T} \in \mathcal{L}(W)$ are both defined, and all three maps in the asserted equalities belong to $\mathcal{L}(V, W)$.
The first equality: $T = T P_{(\operatorname{null} T)^\perp}$. Let $v \in V$. By 6.49 applied to the finite-dimensional subspace $\operatorname{null} T$, we may write

$$v = u + x \quad \text{with } u \in \operatorname{null} T \text{ and } x \in (\operatorname{null} T)^\perp.$$

By 6.52 we have $\big((\operatorname{null} T)^\perp\big)^\perp = \operatorname{null} T$, so the rewriting $v = x + u$ with $x \in (\operatorname{null} T)^\perp$ and $u \in \big((\operatorname{null} T)^\perp\big)^\perp$ is precisely the decomposition used in definition 6.55 for the subspace $(\operatorname{null} T)^\perp$. Hence

$$P_{(\operatorname{null} T)^\perp} v = x.$$

(Equivalently, this is Exercise 6C.5 with $U = \operatorname{null} T$.) Since $Tu = 0$, we get

$$T\big(P_{(\operatorname{null} T)^\perp} v\big) = Tx = Tx + Tu = T(x + u) = Tv.$$

As $v \in V$ was arbitrary, $TP_{(\operatorname{null} T)^\perp} = T$.
The second equality: $T = P_{\operatorname{range} T}\, T$. Let $v \in V$. Then $Tv \in \operatorname{range} T$, so by 6.57(b), applied to the finite-dimensional subspace $\operatorname{range} T$ of W,

$$P_{\operatorname{range} T}(Tv) = Tv.$$

As $v \in V$ was arbitrary, $P_{\operatorname{range} T}\, T = T$.
Combining the two displayed conclusions gives $T = TP_{(\operatorname{null} T)^\perp} = P_{\operatorname{range} T}\, T$, as desired.
These identities say that T loses no information by first discarding the part of its input lying in $\operatorname{null} T$, and that the output of T already lies in $\operatorname{range} T$ so projecting onto $\operatorname{range} T$ changes nothing. They are the reason the restriction $T|_{(\operatorname{null} T)^\perp}$ of 6.67 captures all of T, which in turn underlies the definition 6.68 of the pseudoinverse $T^\dagger$.

Problem 6C.7. Suppose that X and Y are finite-dimensional subspaces of V. Prove that $P_X P_Y = 0$ if and only if $\langle x, y\rangle = 0$ for all $x \in X$ and all $y \in Y$.

Solution. Because X and Y are finite-dimensional subspaces of V, the orthogonal projections $P_X, P_Y \in \mathcal{L}(V)$ are defined (6.55). Recall from 6.57 that $P_Y y = y$ for every $y \in Y$ [part (b)], that $\operatorname{range} P_Y = Y$ [part (d)], and that $\operatorname{null} P_X = X^\perp$ [part (e)], where $X^\perp = \{v \in V : \langle x, v\rangle = 0 \text{ for all } x \in X\}$.
First suppose $\langle x, y\rangle = 0$ for all $x \in X$ and all $y \in Y$. This says exactly that $Y \subseteq X^\perp$. Let $v \in V$. Then $P_Y v \in \operatorname{range} P_Y = Y \subseteq X^\perp = \operatorname{null} P_X$, and therefore

$$(P_X P_Y)v = P_X(P_Y v) = 0.$$

Since $v \in V$ was arbitrary, $P_X P_Y = 0$.
Conversely, suppose $P_X P_Y = 0$. Let $y \in Y$. By 6.57(b) we have $P_Y y = y$, so

$$P_X y = P_X(P_Y y) = (P_X P_Y)y = 0.$$

Hence $y \in \operatorname{null} P_X = X^\perp$ by 6.57(e), which means $\langle x, y\rangle = 0$ for every $x \in X$. Since $y \in Y$ was arbitrary, $\langle x, y\rangle = 0$ for all $x \in X$ and all $y \in Y$.
This proves the equivalence.
Two remarks. First, the condition in the exercise is symmetric in X and Y: by conjugate symmetry, $\langle x, y\rangle = 0$ for all $x \in X, y \in Y$ if and only if $\langle y, x\rangle = 0$ for all such x, y. Hence $P_X P_Y = 0$ if and only if $P_Y P_X = 0$, even though the products themselves need not be equal in general. Second, the condition is equivalent to each of $Y \subseteq X^\perp$ and $X \subseteq Y^\perp$.

Problem 6C.8. Suppose U is a finite-dimensional subspace of V and $v \in V$. Define a linear functional $\varphi\colon U \to \mathbf{F}$ by

$$\varphi(u) = \langle u, v\rangle$$

for all $u \in U$. By the Riesz representation theorem (6.42) as applied to the inner product space U, there exists a unique vector $w \in U$ such that

$$\varphi(u) = \langle u, w\rangle$$

for all $u \in U$. Show that $w = P_U v$.

Solution. The vector w lies in U and satisfies $\langle u, w\rangle = \varphi(u) = \langle u, v\rangle$ for every $u \in U$. Hence for every $u \in U$ we have, using additivity in the second slot [6.6(d)] and $\langle u, -w\rangle = -\langle u, w\rangle$ [6.6(e)],

$$\langle u, v - w\rangle = \langle u, v\rangle - \langle u, w\rangle = 0.$$

Because $u \in U$ was arbitrary, this says exactly that $v - w \in U^\perp$.
Thus we have written

$$v = w + (v - w),$$

where $w \in U$ and $v - w \in U^\perp$. Since U is finite-dimensional, 6.49 gives $V = U \oplus U^\perp$, so the decomposition of v into a vector in U plus a vector in $U^\perp$ is unique, and by the definition of the orthogonal projection (6.55) the U-component of v is $P_U v$. Therefore

$$P_U v = w,$$

as desired.
Note the pleasant consequence: the Riesz vector representing $u \mapsto \langle u, v\rangle$ on the subspace U is not v itself (which need not lie in U) but its orthogonal projection onto U.

Problem 6C.9. Suppose V is finite-dimensional. Suppose $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and every vector in $\operatorname{null} P$ is orthogonal to every vector in $\operatorname{range} P$. Prove that there exists a subspace U of V such that $P = P_U$.

Solution. Let

$$U = \operatorname{range} P.$$

Then U is a subspace of V, and it is finite-dimensional because V is, so the orthogonal projection $P_U \in \mathcal{L}(V)$ is defined (6.55). We show $P = P_U$.
Let $v \in V$. Write

$$v = Pv + (v - Pv).$$

The first term Pv lies in $\operatorname{range} P = U$. For the second term,

$$P(v - Pv) = Pv - P^2 v = Pv - Pv = 0,$$

using $P^2 = P$; hence $v - Pv \in \operatorname{null} P$. By hypothesis every vector in $\operatorname{null} P$ is orthogonal to every vector in $\operatorname{range} P = U$, so

$$v - Pv \in U^\perp.$$

Thus $v = Pv + (v - Pv)$ with $Pv \in U$ and $v - Pv \in U^\perp$. Because $V = U \oplus U^\perp$ (6.49), this is the unique such decomposition of v, and therefore the definition of the orthogonal projection (6.55) gives

$$P_U v = Pv.$$

Since $v \in V$ was arbitrary, $P = P_U$, as desired.
(As a remark, the hypotheses in fact force $\operatorname{null} P = U^\perp$ exactly: the inclusion $\operatorname{null} P \subseteq U^\perp$ was just used, and conversely $\operatorname{null} P = \operatorname{null} P_U = U^\perp$ by 6.57(e) once $P = P_U$ is known.)

Problem 6C.10. Suppose V is finite-dimensional and $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and

$$\|Pv\| \le \|v\|$$

for every $v \in V$. Prove that there exists a subspace U of V such that $P = P_U$.

Solution. By Exercise 6C.9 it suffices to show that every vector in $\operatorname{null} P$ is orthogonal to every vector in $\operatorname{range} P$; then $P = P_U$ with $U = \operatorname{range} P$.
So suppose $u \in \operatorname{range} P$ and $w \in \operatorname{null} P$. If $w = 0$ then $\langle u, w\rangle = 0$ by 6.6(c) and we are done, so assume $w \neq 0$.
First note that $Pu = u$: indeed $u = Px$ for some $x \in V$, and then

$$Pu = P^2 x = Px = u$$

because $P^2 = P$. Hence for every $\lambda \in \mathbf{F}$,

$$P(u + \lambda w) = Pu + \lambda Pw = u + 0 = u.$$

Applying the hypothesis $\|Pv\| \le \|v\|$ to the vector $v = u + \lambda w$ therefore gives

$$\|u\| \le \|u + \lambda w\| \qquad \text{for every } \lambda \in \mathbf{F}.$$

Now expand the right side. Using additivity in each slot, homogeneity in the first slot, 6.6(e), and conjugate symmetry,

$$\begin{aligned}\|u + \lambda w\|^2 &= \langle u, u\rangle + \langle u, \lambda w\rangle + \langle \lambda w, u\rangle + \langle \lambda w, \lambda w\rangle \\ &= \|u\|^2 + \overline{\lambda}\,\langle u, w\rangle + \overline{\overline{\lambda}\,\langle u, w\rangle} + |\lambda|^2\|w\|^2 \\ &= \|u\|^2 + 2\,\mathrm{Re}\big(\overline{\lambda}\,\langle u, w\rangle\big) + |\lambda|^2\|w\|^2.\end{aligned}$$

Choose the specific scalar

$$\lambda = -\frac{\langle u, w\rangle}{\|w\|^2},$$

which is legitimate because $w \neq 0$ and hence $\|w\| \neq 0$. For this λ,

$$\overline{\lambda}\,\langle u, w\rangle = -\frac{\overline{\langle u, w\rangle}\,\langle u, w\rangle}{\|w\|^2} = -\frac{|\langle u, w\rangle|^2}{\|w\|^2},$$

which is a real number, and

$$|\lambda|^2\|w\|^2 = \frac{|\langle u, w\rangle|^2}{\|w\|^4}\,\|w\|^2 = \frac{|\langle u, w\rangle|^2}{\|w\|^2}.$$

Substituting into the expansion above,

$$\|u + \lambda w\|^2 = \|u\|^2 - 2\,\frac{|\langle u, w\rangle|^2}{\|w\|^2} + \frac{|\langle u, w\rangle|^2}{\|w\|^2} = \|u\|^2 - \frac{|\langle u, w\rangle|^2}{\|w\|^2}.$$

Combining this with $\|u\|^2 \le \|u + \lambda w\|^2$ yields

$$\|u\|^2 \le \|u\|^2 - \frac{|\langle u, w\rangle|^2}{\|w\|^2},$$

so $|\langle u, w\rangle|^2/\|w\|^2 \le 0$. Since this quantity is nonnegative, it equals 0, and hence $\langle u, w\rangle = 0$.
Thus every vector in range P is orthogonal to every vector in null P. By Exercise 6C.9 (or by repeating its two-line argument: for $v \in V$ we have $v = Pv + (v - Pv)$ with $Pv \in U$ and $v - Pv \in \mathrm{null}\,P \subseteq U^\perp$, so $P_U v = Pv$), we conclude that

$$P = P_U \qquad \text{where } U = \mathrm{range}\,P.$$

Geometrically: an idempotent operator is a projection onto its range along its null space, and the norm-shrinking condition forces that null space to be exactly the orthogonal complement of the range, making the projection orthogonal. The converse of this exercise is 6.57(g) together with 6.57(h).

Problem 6C.11. Suppose $T \in \mathcal{L}(V)$ and U is a finite-dimensional subspace of V. Prove that

$$U \text{ is invariant under } T \iff P_U T P_U = T P_U.$$

Solution. Because U is finite-dimensional, $P_U \in \mathcal{L}(V)$ is defined (6.55, 6.57(a)), and both sides of the claimed identity are operators on V; we prove that they agree on every $v \in V$ exactly when U is invariant under T.
First suppose U is invariant under T. Let $v \in V$. Then $P_U v \in \mathrm{range}\,P_U = U$ by 6.57(d). Since U is

invariant under T, we get $T(P_U v) \in U$. Now 6.57(b) says that P_U fixes each vector of U, so

$$P_U\big(TP_U v\big) = TP_U v.$$

As $v \in V$ was arbitrary, $P_U T P_U = TP_U$.
Conversely, suppose $P_U T P_U = TP_U$. Let $u \in U$. By 6.57(b) we have $P_U u = u$, so

$$Tu = TP_U u = P_U T P_U u = P_U(Tu).$$

Thus $Tu \in \operatorname{range} P_U = U$ by 6.57(d). As $u \in U$ was arbitrary, U is invariant under T.
This proves both implications, so the equivalence holds.

Problem 6C.12. Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is a subspace of V. Prove that

$$U \text{ and } U^\perp \text{ are both invariant under } T \iff P_U T = TP_U.$$

Solution. Since V is finite-dimensional, so is U, and thus $P_U \in \mathcal{L}(V)$ is defined, with $\operatorname{range} P_U = U$ and $\operatorname{null} P_U = U^\perp$ by 6.57(d) and 6.57(e). Also $V = U \oplus U^\perp$ by 6.49.
First suppose U and $U^\perp$ are both invariant under T. Let $v \in V$. By 6.49 write

$$v = u + w \quad \text{with } u \in U,\ w \in U^\perp,$$

so that $P_U v = u$ by the definition 6.55. Then

$$TP_U v = Tu.$$

On the other hand, $Tv = Tu + Tw$. Invariance of U gives $Tu \in U$, and invariance of $U^\perp$ gives $Tw \in U^\perp$. Hence $Tv = Tu + Tw$ is precisely the decomposition of Tv into a vector in U plus a vector in $U^\perp$, and therefore

$$P_U T v = Tu.$$

Comparing the two displays, $P_U T v = TP_U v$ for every $v \in V$, i.e. $P_U T = TP_U$.
Conversely, suppose $P_U T = TP_U$. First let $u \in U$. Then $P_U u = u$ by 6.57(b), so

$$Tu = TP_U u = P_U T u \in \operatorname{range} P_U = U.$$

Hence U is invariant under T.
Next let $w \in U^\perp = \operatorname{null} P_U$ [6.57(e)]. Then $P_U w = 0$, so

$$P_U(Tw) = TP_U w = T0 = 0,$$

which says $Tw \in \operatorname{null} P_U = U^\perp$. Hence $U^\perp$ is invariant under T.
Both implications are proved, so the equivalence holds.
Remark: comparing with Exercise 6C.11, the condition $P_U T P_U = TP_U$ captures invariance of U alone, while the stronger condition $P_U T = TP_U$ captures invariance of U and $U^\perp$ simultaneously. This can also be seen by applying 6C.11 twice. That exercise applied to $U^\perp$, together with $P_{U^\perp} = I - P_U$ (Exercise 6C.5), says that $U^\perp$ is invariant under T if and only if

$$(I - P_U)T(I - P_U) = T(I - P_U),$$

and expanding both sides reduces this to $P_U T P_U = P_U T$. So 6C.11 gives

$$U \text{ invariant} \iff P_U T P_U = TP_U, \qquad U^\perp \text{ invariant} \iff P_U T P_U = P_U T,$$

and these two conditions hold together exactly when $TP_U = P_U T$.

Problem 6C.13. Suppose $\mathbf{F} = \mathbf{R}$ and V is finite-dimensional. For each $v \in V$, let φ_v denote the linear functional on V defined by

$$\varphi_v(u) = \langle u, v \rangle$$

for all $u \in V$.

(a) Show that $v \mapsto \varphi_v$ is an injective linear map from V to V'.

(b) Use (a) and a dimension-counting argument to show that $v \mapsto \varphi_v$ is an isomorphism from V onto V'.

[The purpose of this exercise is to give an alternative proof of the Riesz representation theorem (6.42 and 6.58) when $\mathbf{F} = \mathbf{R}$. Thus you should not use the Riesz representation theorem as a tool in your solution.]

Solution. Throughout, $\mathbf{F} = \mathbf{R}$, so every scalar equals its own complex conjugate; in particular conjugate symmetry becomes symmetry, $\langle u, v \rangle = \langle v, u \rangle$, and 6.6(e) reads $\langle u, \lambda v \rangle = \lambda \langle u, v \rangle$. Write Φ for the function $v \mapsto \varphi_v$. No use is made of 6.42 or 6.58 anywhere below.

(a) First, each φ_v really is an element of V': by 6.6(a), for fixed $v \in V$ the function $u \mapsto \langle u, v \rangle$ is a linear map from V to $\mathbf{F}$, which is exactly the statement that $\varphi_v \in V'$. So $\Phi \colon V \to V'$ is a well-defined function.

Next, Φ is linear. For additivity, let $v_1, v_2 \in V$. For every $u \in V$, additivity in the second slot [6.6(d)] gives

$$\varphi_{v_1+v_2}(u) = \langle u, v_1 + v_2 \rangle = \langle u, v_1 \rangle + \langle u, v_2 \rangle = \varphi_{v_1}(u) + \varphi_{v_2}(u) = \big(\varphi_{v_1} + \varphi_{v_2}\big)(u).$$

Since two functions on V agreeing at every $u \in V$ are equal, $\Phi(v_1 + v_2) = \Phi(v_1) + \Phi(v_2)$.

For homogeneity, let $\lambda \in \mathbf{R}$ and $v \in V$. For every $u \in V$, 6.6(e) gives

$$\varphi_{\lambda v}(u) = \langle u, \lambda v \rangle = \overline{\lambda}\, \langle u, v \rangle = \lambda\, \langle u, v \rangle = \big(\lambda \varphi_v\big)(u),$$

where $\overline{\lambda} = \lambda$ because λ is real. Hence $\Phi(\lambda v) = \lambda \Phi(v)$. (This is precisely the step that fails when $\mathbf{F} = \mathbf{C}$: there one gets $\varphi_{\lambda v} = \overline{\lambda} \varphi_v$, so Φ is only conjugate-linear.) Thus $\Phi \in \mathcal{L}(V, V')$.

Finally, Φ is injective. Suppose $v \in \operatorname{null} \Phi$, that is, $\varphi_v = 0$. Evaluating at $u = v$ gives

$$0 = \varphi_v(v) = \langle v, v \rangle = \|v\|^2,$$

and positive definiteness of the inner product forces $v = 0$. Hence $\operatorname{null} \Phi = \{0\}$, so Φ is injective by 3.15.

(b) Since V is finite-dimensional, 3.111 gives that V' is finite-dimensional with

$$\dim V' = \dim V.$$

By the fundamental theorem of linear maps (3.21) applied to $\Phi \in \mathcal{L}(V, V')$, together with $\dim \operatorname{null} \Phi = 0$ from part (a),

$$\dim \operatorname{range} \Phi = \dim V - \dim \operatorname{null} \Phi = \dim V = \dim V'.$$

Thus $\operatorname{range} \Phi$ is a subspace of V' whose dimension equals $\dim V'$, so $\operatorname{range} \Phi = V'$ by 2.39. Hence Φ is surjective as well as injective, i.e. Φ is an isomorphism from V onto V'. (Alternatively, quote 3.65 directly: $\dim V = \dim V' < \infty$ and Φ injective imply Φ invertible.)

This yields the Riesz representation theorem for $\mathbf{F} = \mathbf{R}$: given $\varphi \in V'$, surjectivity of Φ produces $v \in V$ with $\varphi = \varphi_v$, that is,

$$\varphi(u) = \langle u, v \rangle \quad \text{for all } u \in V,$$

and injectivity of Φ makes this v unique.

Problem 6C.14. Suppose that $e_1, \ldots, e_n$ is an orthonormal basis of V. Explain why the dual basis (see 3.112) of $e_1, \ldots, e_n$ is $e_1, \ldots, e_n$ under the identification of V' with V provided by the Riesz representation theorem (6.58).

Solution. The identification of V' with V given by 6.58 is the bijection

$$v \longleftrightarrow \varphi_v, \qquad \text{where } \varphi_v(u) = \langle u, v\rangle \text{ for all } u \in V.$$

Under this identification, saying that "the dual basis of $e_1, \dots, e_n$ is $e_1, \dots, e_n$" means: if $\varphi_1, \dots, \varphi_n$ denotes the dual basis of $e_1, \dots, e_n$ in the sense of 3.112, then the vector of V corresponding to φ_j is e_j; equivalently

$$\varphi_j = \varphi_{e_j} \quad \text{for each } j = 1, \dots, n.$$

To verify this, recall that by 3.112 the dual basis $\varphi_1, \dots, \varphi_n \in V'$ is determined by the conditions

$$\varphi_j(e_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

Now compute φ_{e_j} on the basis vectors. For all $j, k \in \{1, \dots, n\}$,

$$\varphi_{e_j}(e_k) = \langle e_k, e_j\rangle = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j, \end{cases}$$

where the last equality is exactly the statement that $e_1, \dots, e_n$ is an orthonormal list: distinct vectors of the list are orthogonal, and each has norm 1, so $\langle e_j, e_j\rangle = \|e_j\|^2 = 1$.
Thus the linear functionals φ_j and φ_{e_j} take the same value at each of $e_1, \dots, e_n$. Two linear maps that agree on a basis of their domain are equal, because their values on a basis determine them on all of V by linearity. Since $e_1, \dots, e_n$ is a basis of V, we conclude

$$\varphi_j = \varphi_{e_j} \quad \text{for each } j = 1, \dots, n.$$

Hence under the identification $v \leftrightarrow \varphi_v$ of 6.58, the dual basis $\varphi_1, \dots, \varphi_n$ of $e_1, \dots, e_n$ corresponds to the list $e_1, \dots, e_n$ itself. In other words, an orthonormal basis is its own dual basis once V' is identified with V via the inner product.
(Two remarks. First, the argument does not require $\mathbf{F} = \mathbf{R}$: even when $\mathbf{F} = \mathbf{C}$, where $v \mapsto \varphi_v$ is not linear, 6.58 still supplies a bijection between V and V', and the computation above is unaffected. Second, this is a genuinely special feature of orthonormal bases; for a general basis $v_1, \dots, v_n$ of V the dual basis corresponds under the identification to a different list, the so-called dual or reciprocal list of $v_1, \dots, v_n$.)

Problem 6C.15. In $\mathbb{R}^4$, let

$$U = \operatorname{span}\big((1,1,0,0), (1,1,1,2)\big).$$

Find $u \in U$ such that $\|u - (1,2,3,4)\|$ is as small as possible.

Solution. Write $v = (1,2,3,4)$. By 6.61, the unique vector $u \in U$ minimizing $\|u - v\|$ is $u = P_U v$, the orthogonal projection of v onto U. So we must compute $P_U v$.
First apply the Gram–Schmidt procedure (6.32) to the basis $(1,1,0,0), (1,1,1,2)$ of U (these two vectors are not scalar multiples of each other, so they are linearly independent and hence form a basis of U). Because $\|(1,1,0,0)\| = \sqrt{2}$, the first vector produced is

$$e_1 = \tfrac{1}{\sqrt{2}}(1,1,0,0).$$

Next, $\langle (1,1,1,2), e_1\rangle = \frac{1}{\sqrt{2}}(1+1) = \sqrt{2}$, and

$$(1,1,1,2) - \sqrt{2}\, e_1 = (1,1,1,2) - (1,1,0,0) = (0,0,1,2),$$

which has norm $\sqrt{5}$. Hence

$$e_2 = \tfrac{1}{\sqrt{5}}(0,0,1,2),$$

and e_1, e_2 is an orthonormal basis of U.
Now 6.57(i) gives $P_U v = \langle v, e_1\rangle e_1 + \langle v, e_2\rangle e_2$. We have

$$\langle v, e_1\rangle = \tfrac{1}{\sqrt{2}}(1+2) = \tfrac{3}{\sqrt{2}}, \qquad \langle v, e_2\rangle = \tfrac{1}{\sqrt{5}}(3+8) = \tfrac{11}{\sqrt{5}}.$$

Therefore

$$P_U v = \tfrac{3}{2}(1,1,0,0) + \tfrac{11}{5}(0,0,1,2) = \left(\tfrac{3}{2},\ \tfrac{3}{2},\ \tfrac{11}{5},\ \tfrac{22}{5}\right).$$

Thus the answer is

$$u = \left(\tfrac{3}{2},\ \tfrac{3}{2},\ \tfrac{11}{5},\ \tfrac{22}{5}\right).$$

As a check, $v - u = \left(-\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{4}{5}, -\tfrac{2}{5}\right)$, and indeed

$$\langle v-u, (1,1,0,0)\rangle = -\tfrac{1}{2} + \tfrac{1}{2} = 0, \qquad \langle v-u, (1,1,1,2)\rangle = -\tfrac{1}{2} + \tfrac{1}{2} + \tfrac{4}{5} - \tfrac{4}{5} = 0,$$

so $v - u \in U^\perp$, confirming that $u = P_U v$. The minimum value of the distance is

$$\|v - u\| = \sqrt{\tfrac{1}{4} + \tfrac{1}{4} + \tfrac{16}{25} + \tfrac{4}{25}} = \sqrt{\tfrac{13}{10}}.$$

Problem 6C.16. Suppose $C[-1,1]$ is the vector space of continuous real-valued functions on the interval $[-1,1]$ with inner product given by

$$\langle f, g\rangle = \int_{-1}^{1} fg$$

for all $f, g \in C[-1,1]$. Let U be the subspace of $C[-1,1]$ defined by

$$U = \{f \in C[-1,1] : f(0) = 0\}.$$

(a) Show that $U^\perp = \{0\}$.
(b) Show that 6.49 and 6.52 do not hold without the finite-dimensional hypothesis.

Solution. Throughout, $V = C[-1,1]$ with the inner product above. Note first that U really is a subspace of V: the zero function lies in U, and if $f, g \in U$ and $\lambda \in \mathbb{R}$ then $(f+g)(0) = 0$ and $(\lambda f)(0) = 0$.
(a) Certainly $0 \in U^\perp$. For the reverse inclusion, suppose $g \in U^\perp$. Define $f \in C[-1,1]$ by

$$f(x) = x^2 g(x).$$

Then f is continuous (a product of continuous functions) and $f(0) = 0$, so $f \in U$. Because $g \in U^\perp$, we have $\langle f, g\rangle = 0$, which means

$$\int_{-1}^{1} x^2 g(x)^2\, dx = 0.$$

The integrand $x \mapsto x^2 g(x)^2$ is continuous and nonnegative on $[-1,1]$. A continuous nonnegative function whose integral over an interval is 0 must be identically 0 on that interval. (Indeed, if h is continuous, $h \geq 0$, and $h(x_0) > 0$ for some $x_0 \in [-1,1]$, then by continuity there is a nondegenerate subinterval $I \subseteq [-1,1]$ containing x_0 on which $h > h(x_0)/2$, and then $\int_{-1}^{1} h \geq \frac{1}{2} h(x_0)\,\text{length}(I) > 0$.) Hence $x^2 g(x)^2 = 0$ for every $x \in [-1,1]$, which forces $g(x) = 0$ for every $x \neq 0$. Since g is continuous, $g(0) = \lim_{x\to 0} g(x) = 0$ as well. Thus $g = 0$, and therefore $U^\perp = \{0\}$.
(b) Note that U is infinite-dimensional: the functions $x, x^2, x^3, \ldots$ all lie in U and, being polynomials of distinct degrees, every finite sublist of them is linearly independent, so U contains linearly independent lists of arbitrary length. Thus the finite-dimensionality hypothesis of 6.49 and 6.52 fails for this U, and we now check that both conclusions fail as well.
Result 6.49 asserts that $V = U \oplus U^\perp$. Here, by (a),

$$U + U^\perp = U + \{0\} = U.$$

But $U \neq V$: the constant function 1 belongs to $C[-1,1]$ and does not belong to U, since its value at 0 is $1 \neq 0$. Hence $U \oplus U^\perp = U \subsetneq V$, so the conclusion of 6.49 fails for this subspace.
Result 6.52 asserts that $U = (U^\perp)^\perp$. Here, by (a),

$$(U^\perp)^\perp = \{0\}^\perp = V,$$

because every vector of V is orthogonal to 0. Since $U \neq V$ (again because the constant function 1 lies in $V \setminus U$), we get $(U^\perp)^\perp \neq U$, so the conclusion of 6.52 also fails for this subspace.
Thus neither 6.49 nor 6.52 remains valid if the hypothesis that U is finite-dimensional is dropped. (This example also shows that 6.54 fails without that hypothesis: $U^\perp = \{0\}$ but $U \neq V$.)

Problem 6C.17. Find $p \in \mathcal{P}_3(\mathbb{R})$ such that $p(0) = 0$, $p'(0) = 0$, and

$$\int_0^1 \left|2 + 3x - p(x)\right|^2 dx$$

is as small as possible.

Solution. Work in the real inner product space $V = C[0,1]$ of continuous real-valued functions on $[0,1]$ with

$$\langle f, g\rangle = \int_0^1 fg.$$

Let $v \in V$ be defined by $v(x) = 2 + 3x$, and let

$$U = \{p \in \mathcal{P}_3(\mathbb{R}) : p(0) = 0 \text{ and } p'(0) = 0\}.$$

If $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, then $p(0) = a_0$ and $p'(0) = a_1$; hence the conditions $p(0) = p'(0) = 0$ say exactly that $a_0 = a_1 = 0$. Thus

$$U = \operatorname{span}(x^2, x^3),$$

a two-dimensional subspace of V (the functions x^2 and x^3 are linearly independent). The problem asks us to minimize $\|v - p\|^2$ over $p \in U$, and by 6.61 the unique minimizer is $p = P_U v$.
Apply Gram–Schmidt (6.32) to x^2, x^3. Since $\int_0^1 x^4\,dx = \frac{1}{5}$, we get

$$e_1(x) = \sqrt{5}\,x^2.$$

Next, $\langle x^3, e_1\rangle = \sqrt{5}\int_0^1 x^5\,dx = \frac{\sqrt{5}}{6}$, so

$$x^3 - \langle x^3, e_1\rangle e_1 = x^3 - \tfrac{5}{6}x^2,$$

and

$$\int_0^1 \Big(x^3 - \tfrac{5}{6}x^2\Big)^2 dx = \int_0^1 \Big(x^6 - \tfrac{5}{3}x^5 + \tfrac{25}{36}x^4\Big)dx = \tfrac{1}{7} - \tfrac{5}{18} + \tfrac{5}{36} = \tfrac{1}{252}.$$

Hence

$$e_2(x) = \sqrt{252}\,\Big(x^3 - \tfrac{5}{6}x^2\Big),$$

and e_1, e_2 is an orthonormal basis of U.
Now compute the two inner products needed in 6.57(i). First,

$$\int_0^1 (2+3x)x^2\,dx = \tfrac{2}{3} + \tfrac{3}{4} = \tfrac{17}{12}, \qquad \int_0^1 (2+3x)x^3\,dx = \tfrac{1}{2} + \tfrac{3}{5} = \tfrac{11}{10}.$$

Therefore

$$\langle v, e_1\rangle = \tfrac{17\sqrt{5}}{12}, \qquad \langle v, e_2\rangle = \sqrt{252}\Big(\tfrac{11}{10} - \tfrac{5}{6}\cdot\tfrac{17}{12}\Big) = \sqrt{252}\Big(\tfrac{11}{10} - \tfrac{85}{72}\Big) = -\tfrac{29\sqrt{252}}{360}.$$

Consequently

$$\begin{aligned} P_U v &= \tfrac{17\sqrt{5}}{12} \cdot \sqrt{5}x^2 \;-\; \tfrac{29\sqrt{252}}{360} \cdot \sqrt{252}\Big(x^3 - \tfrac{5}{6}x^2\Big) \\ &= \tfrac{85}{12}x^2 - \tfrac{203}{10}\Big(x^3 - \tfrac{5}{6}x^2\Big) \\ &= \Big(\tfrac{85}{12} + \tfrac{203}{12}\Big)x^2 - \tfrac{203}{10}x^3 \\ &= 24x^2 - \tfrac{203}{10}x^3, \end{aligned}$$

where we used $\frac{29\cdot 252}{360} = \frac{203}{10}$ and $\frac{203}{10} \cdot \frac{5}{6} = \frac{203}{12}$.
Thus the answer is

$$p(x) = 24x^2 - \tfrac{203}{10}x^3.$$

As a check, with p as above,

$$\int_0^1 p(x)x^2\,dx = \tfrac{24}{5} - \tfrac{203}{60} = \tfrac{288}{60} - \tfrac{203}{60} = \tfrac{85}{60} = \tfrac{17}{12},$$

and

$$\int_0^1 p(x)x^3\,dx = \tfrac{24}{6} - \tfrac{203}{70} = 4 - \tfrac{29}{10} = \tfrac{11}{10}.$$

These agree with $\int_0^1 (2+3x)x^2dx$ and $\int_0^1 (2+3x)x^3dx$ computed above, so $v - p$ is orthogonal to both x^2 and x^3, hence to all of U. This confirms $p = P_U v$.

Problem 6C.18. Find $p \in \mathcal{P}_5(\mathbb{R})$ that makes

$$\int_{-\pi}^{\pi} \big|\sin x - p(x)\big|^2\,dx$$

as small as possible.
The polynomial 6.65 is an excellent approximation to the answer to this exercise, but here you are asked to find the exact solution, which involves powers of π. A computer that can perform symbolic integration should help.

Solution. Let $V = C[-\pi, \pi]$ with the inner product 6.64, namely $\langle f, g\rangle = \int_{-\pi}^{\pi} fg$, let $v \in V$ be given by $v(x) = \sin x$, and let $U \subseteq V$ be the subspace consisting of the polynomials with real coefficients of degree at most 5. We must minimize $\|v - p\|^2$ over $p \in U$, so by 6.61 the answer is $p = P_U v$.
Rather than running Gram–Schmidt, we use the characterization of $P_U v$ by orthogonality. If $p \in U$ and $v - p \in U^\perp$, then $v = p + (v - p)$ is the decomposition of v coming from $V = U \oplus U^\perp$ (6.49), so $p = P_U v$; conversely $v - P_U v \in U^\perp$ by 6.57(f). Since $1, x, x^2, x^3, x^4, x^5$ spans U, the condition $v - p \in U^\perp$ says precisely

$$\int_{-\pi}^{\pi} x^k\big(\sin x - p(x)\big)\,dx = 0 \qquad \text{for } k = 0, 1, 2, 3, 4, 5.$$

Two elementary integrals are needed. First, for every integer $m \geq 0$,

$$\int_{-\pi}^{\pi} x^m\,dx = \begin{cases} 0 & \text{if } m \text{ is odd,} \\ \dfrac{2\pi^{m+1}}{m+1} & \text{if } m \text{ is even,} \end{cases}$$

since x^m is an odd function when m is odd. Second, $x^k \sin x$ is an odd function when k is even, so $\int_{-\pi}^{\pi} x^k \sin x\,dx = 0$ for k even; and for k odd, $x^k \sin x$ is even, so $\int_{-\pi}^{\pi} x^k \sin x\,dx = 2I_k$ where $I_k = \int_0^{\pi} x^k \sin x\,dx$. Integrating by parts twice,

$$I_n = \big[-x^n \cos x\big]_0^{\pi} + n\int_0^{\pi} x^{n-1}\cos x\,dx = \pi^n + n\big[x^{n-1}\sin x\big]_0^{\pi} - n(n-1)I_{n-2} = \pi^n - n(n-1)I_{n-2}$$

for $n \geq 2$. Since $I_1 = \big[\sin x - x\cos x\big]_0^{\pi} = \pi$, we get

$$I_1 = \pi, \qquad I_3 = \pi^3 - 6\pi, \qquad I_5 = \pi^5 - 20\pi^3 + 120\pi.$$

Write $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5$.
Step 1: the even coefficients vanish. Take k even. Then $\int_{-\pi}^{\pi} x^k \sin x\, dx = 0$, and $\int_{-\pi}^{\pi} x^{k+j}dx = 0$ for j odd, so the equation for that k reduces to

$$\int_{-\pi}^{\pi} x^k\, q(x)\, dx = 0, \qquad \text{where } q(x) = a_0 + a_2x^2 + a_4x^4.$$

Thus q is orthogonal to each of $1, x^2, x^4$, hence to their span, which contains q itself. So $\langle q, q\rangle = 0$, giving $q = 0$ and therefore $a_0 = a_2 = a_4 = 0$ (the functions $1, x^2, x^4$ are linearly independent).
Step 2: the odd coefficients. Write $a = a_1$, $b = a_3$, $c = a_5$, so $p(x) = ax + bx^3 + cx^5$. For k odd, $\int_{-\pi}^{\pi} x^{k+j}dx = 0$ whenever j is even, so the three remaining equations ($k = 1, 3, 5$) are

$$\begin{aligned}
\tfrac{2\pi^3}{3}a + \tfrac{2\pi^5}{5}b + \tfrac{2\pi^7}{7}c &= 2\pi,\\
\tfrac{2\pi^5}{5}a + \tfrac{2\pi^7}{7}b + \tfrac{2\pi^9}{9}c &= 2(\pi^3 - 6\pi),\\
\tfrac{2\pi^7}{7}a + \tfrac{2\pi^9}{9}b + \tfrac{2\pi^{11}}{11}c &= 2(\pi^5 - 20\pi^3 + 120\pi).
\end{aligned}$$

Divide the three equations by $2\pi^3$, $2\pi^5$, $2\pi^7$ respectively, and set

$$t = \pi^2, \qquad A = a, \qquad B = \pi^2 b, \qquad C = \pi^4 c.$$

The system becomes

$$\begin{aligned}
\tfrac{A}{3} + \tfrac{B}{5} + \tfrac{C}{7} &= \tfrac{1}{t},\\
\tfrac{A}{5} + \tfrac{B}{7} + \tfrac{C}{9} &= \tfrac{1}{t} - \tfrac{6}{t^2},\\
\tfrac{A}{7} + \tfrac{B}{9} + \tfrac{C}{11} &= \tfrac{1}{t} - \tfrac{20}{t^2} + \tfrac{120}{t^3}.
\end{aligned}$$

The coefficient matrix is a 3×3 matrix of rationals with determinant $\frac{256}{22920975} \neq 0$; solving (by elimination, or by inverting that matrix) gives

$$A = \tfrac{105}{8t} - \tfrac{16065}{8t^2} + \tfrac{155925}{8t^3}, \quad B = -\tfrac{315}{4t} + \tfrac{39375}{4t^2} - \tfrac{363825}{4t^3}, \quad C = \tfrac{693}{8t} - \tfrac{72765}{8t^2} + \tfrac{654885}{8t^3}.$$

Verification of this solution is a matter of collecting powers of $1/t$. For the first equation,

$$\tfrac{A}{3} + \tfrac{B}{5} + \tfrac{C}{7} = \tfrac{35-126+99}{8t} + \tfrac{-5355+15750-10395}{8t^2} + \tfrac{51975-145530+93555}{8t^3} = \tfrac{1}{t},$$

for the second,

$$\tfrac{A}{5} + \tfrac{B}{7} + \tfrac{C}{9} = \tfrac{21-90+77}{8t} + \tfrac{-3213+11250-8085}{8t^2} + \tfrac{31185-103950+72765}{8t^3} = \tfrac{1}{t} - \tfrac{6}{t^2},$$

and for the third,

$$\tfrac{A}{7} + \tfrac{B}{9} + \tfrac{C}{11} = \tfrac{15-70+63}{8t} + \tfrac{-2295+8750-6615}{8t^2} + \tfrac{22275-80850+59535}{8t^3} = \tfrac{1}{t} - \tfrac{20}{t^2} + \tfrac{120}{t^3},$$

exactly as required.
Undoing the substitutions ($a = A$, $b = B/\pi^2$, $c = C/\pi^4$, $t = \pi^2$) gives

$$a = \frac{105\big(\pi^4 - 153\pi^2 + 1485\big)}{8\pi^6}, \quad b = -\frac{315\big(\pi^4 - 125\pi^2 + 1155\big)}{4\pi^8}, \quad c = \frac{693\big(\pi^4 - 105\pi^2 + 945\big)}{8\pi^{10}}.$$

Hence the polynomial that minimizes the integral is

$$p(x) = \frac{105\big(\pi^4 - 153\pi^2 + 1485\big)}{8\pi^6}\, x \;-\; \frac{315\big(\pi^4 - 125\pi^2 + 1155\big)}{4\pi^8}\, x^3 \;+\; \frac{693\big(\pi^4 - 105\pi^2 + 945\big)}{8\pi^{10}}\, x^5.$$

Numerically the coefficients are $0.987862\ldots$, $-0.155271\ldots$, and $0.00564312\ldots$, which agrees with the decimal approximation 6.65 given in the text.

Problem 6C.19. Suppose V is finite-dimensional and $P \in \mathcal{L}(V)$ is an orthogonal projection of V onto some subspace of V. Prove that $P^\dagger = P$.

Solution. Suppose U is a subspace of V with $P = P_U$. Because V is finite-dimensional, so is U, so P_U is defined and 6.57 and 6.52 apply to U.
By 6.57(d) and 6.57(e),

$$\operatorname{range} P = U \qquad \text{and} \qquad \operatorname{null} P = U^\perp.$$

Hence, using 6.52 (valid because U is finite-dimensional),

$$(\operatorname{null} P)^\perp = (U^\perp)^\perp = U.$$

Thus $P|_{(\operatorname{null} P)^\perp} = P|_U$. By 6.57(b) we have $Pu = u$ for every $u \in U$, so $P|_U$ is the identity map of U onto $U = \operatorname{range} P$. Its inverse (as a map from $\operatorname{range} P$ onto $(\operatorname{null} P)^\perp$, which is how it is used in 6.68) is again the identity map of U:

$$\big(P|_{(\operatorname{null} P)^\perp}\big)^{-1} u = u \qquad \text{for every } u \in U.$$

Now let $v \in V$. Since $\operatorname{range} P = U$, we have $P_{\operatorname{range} P} = P_U = P$. Therefore the definition 6.68 of the pseudoinverse gives

$$P^\dagger v = \big(P|_{(\operatorname{null} P)^\perp}\big)^{-1} P_{\operatorname{range} P}\, v = \big(P|_{(\operatorname{null} P)^\perp}\big)^{-1} (Pv) = Pv,$$

where the last equality holds because $Pv \in U$ and the inverse map above is the identity on U.
Because $P^\dagger v = Pv$ for every $v \in V$, we conclude that $P^\dagger = P$.

Problem 6C.20. Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Show that

$$\operatorname{null} T^\dagger = (\operatorname{range} T)^\perp \qquad \text{and} \qquad \operatorname{range} T^\dagger = (\operatorname{null} T)^\perp.$$

Solution. Because V is finite-dimensional, $\operatorname{null} T$, $(\operatorname{null} T)^\perp$, and $\operatorname{range} T$ are all finite-dimensional, so the orthogonal projections used below are defined, and 6.67 tells us that

$$S := T|_{(\operatorname{null} T)^\perp}$$

is an invertible linear map from $(\operatorname{null} T)^\perp$ onto $\operatorname{range} T$. By 6.68,

$$T^\dagger w = S^{-1} P_{\operatorname{range} T}\, w \qquad \text{for every } w \in W.$$

First equality. Suppose $w \in W$. Since S^{-1} is injective and $S^{-1}0 = 0$, we have

$$T^\dagger w = 0 \iff S^{-1}\big(P_{\operatorname{range} T} w\big) = 0 \iff P_{\operatorname{range} T} w = 0 \iff w \in \big(\operatorname{range} T\big)^\perp,$$

where the last equivalence is 6.57(e) applied to the finite-dimensional subspace $\operatorname{range} T$ of W, which says $\operatorname{null} P_{\operatorname{range} T} = (\operatorname{range} T)^\perp$. Hence

$$\operatorname{null} T^\dagger = (\operatorname{range} T)^\perp.$$

Second equality. For the inclusion $\subseteq$: if $w \in W$, then $T^\dagger w = S^{-1}(P_{\operatorname{range} T} w)$ lies in the range of S^{-1}, which is $(\operatorname{null} T)^\perp$ because S maps $(\operatorname{null} T)^\perp$ onto $\operatorname{range} T$. Hence

$$\operatorname{range} T^\dagger \subseteq (\operatorname{null} T)^\perp.$$

For the inclusion $\supseteq$: suppose $v \in (\operatorname{null} T)^\perp$. Put $w = Tv$. Then $w \in \operatorname{range} T$, so $P_{\operatorname{range} T} w = w$ by 6.57(b). Also $w = Tv = Sv$ because $v \in (\operatorname{null} T)^\perp$. Therefore

$$T^\dagger w = S^{-1}\big(P_{\operatorname{range} T} w\big) = S^{-1} w = S^{-1}(Sv) = v,$$

so $v \in \operatorname{range} T^\dagger$. Hence $(\operatorname{null} T)^\perp \subseteq \operatorname{range} T^\dagger$.
Combining the two inclusions gives $\operatorname{range} T^\dagger = (\operatorname{null} T)^\perp$, completing the proof.
(Alternatively, the second equality follows from 6.69(c): that result gives $T^\dagger T = P_{(\operatorname{null} T)^\perp}$, so

$$(\operatorname{null} T)^\perp = \operatorname{range} P_{(\operatorname{null} T)^\perp} = \operatorname{range}\big(T^\dagger T\big) \subseteq \operatorname{range} T^\dagger,$$

using 6.57(d), and the reverse inclusion is the first paragraph above.)

Problem 6C.21. Suppose $T \in \mathcal{L}(\mathbf{F}^3, \mathbf{F}^2)$ is defined by

$$T(a, b, c) = (a + b + c,\ 2b + 3c).$$

(a) For $(x, y) \in \mathbf{F}^2$, find a formula for $T^\dagger(x, y)$.
(b) Verify that the equation $TT^\dagger = P_{\operatorname{range} T}$ from 6.69(b) holds with the formula for $T^\dagger$ obtained in (a).
(c) Verify that the equation $T^\dagger T = P_{(\operatorname{null} T)^\perp}$ from 6.69(c) holds with the formula for $T^\dagger$ obtained in (a).

Solution. Throughout, $\mathbf{F}^n$ carries the Euclidean inner product $\langle u, v\rangle = u_1\overline{v_1} + \cdots + u_n\overline{v_n}$.
(a) First compute $\operatorname{range} T$ and $\operatorname{null} T$.
Since $T(1, 0, 0) = (1, 0)$ and $T(0, 1, 0) = (1, 2)$, the range of T contains two linearly independent vectors of $\mathbf{F}^2$, so

$$\operatorname{range} T = \mathbf{F}^2.$$

In particular T is surjective and $P_{\operatorname{range} T}$ is the identity operator on $\mathbf{F}^2$.
Next, $(a, b, c) \in \operatorname{null} T$ exactly when $a + b + c = 0$ and $2b + 3c = 0$. The second equation gives $b = -\frac{3}{2}c$, and then $a = -b - c = \frac{1}{2}c$. Taking $c = 2$ shows

$$\operatorname{null} T = \operatorname{span}\big((1, -3, 2)\big),$$

a one-dimensional subspace (consistent with the fundamental theorem of linear maps 3.21: $3 = 1 + 2$). Because $(1, -3, 2)$ has real entries, a vector (a, b, c) lies in $(\operatorname{null} T)^\perp$ if and only if

$$\langle (a, b, c), (1, -3, 2)\rangle = a - 3b + 2c = 0.$$

Now fix $(x, y) \in \mathbf{F}^2$. Write $S = T|_{(\operatorname{null} T)^\perp}$, which by 6.67 is an injective map of $(\operatorname{null} T)^\perp$ onto $\operatorname{range} T = \mathbf{F}^2$. The definition 6.68 gives

$$T^\dagger(x, y) = S^{-1}P_{\operatorname{range} T}(x, y) = S^{-1}(x, y),$$

the second equality holding because $P_{\operatorname{range} T}$ is the identity operator on $\mathbf{F}^2$, as noted above. Thus $T^\dagger(x, y)$ is the unique vector $(a, b, c) \in (\operatorname{null} T)^\perp$ satisfying $T(a, b, c) = (x, y)$. So a, b, c are determined by the system

$$\begin{aligned} a + b + c &= x, \\ 2b + 3c &= y, \\ a - 3b + 2c &= 0. \end{aligned}$$

Subtracting the third equation from the first gives $4b - c = x$, so $c = 4b - x$. Substituting into the second equation yields $2b + 12b - 3x = y$, hence

$$b = \frac{3x + y}{14}, \qquad c = 4b - x = \frac{12x + 4y - 14x}{14} = \frac{-2x + 4y}{14},$$

and then

$$a = x - b - c = \frac{14x - (3x + y) - (-2x + 4y)}{14} = \frac{13x - 5y}{14}.$$

Therefore

$$T^\dagger(x, y) = \tfrac{1}{14}\big(13x - 5y,\ 3x + y,\ -2x + 4y\big).$$

As a check on this solution of the system: $a+b+c=\frac{1}{14}(13x-5y+3x+y-2x+4y)=\frac{14x}{14}=x$; $2b+3c=\frac{1}{14}(6x+2y-6x+12y)=\frac{14y}{14}=y$; and $a-3b+2c=\frac{1}{14}(13x-5y-9x-3y-4x+8y)=0$.
(b) Using the formula from (a),

$$TT^\dagger(x,y)=T\Big(\tfrac{13x-5y}{14},\tfrac{3x+y}{14},\tfrac{-2x+4y}{14}\Big)=\big(x,\ y\big),$$

by the two verifications at the end of (a). Thus $TT^\dagger$ is the identity operator on $\mathbf{F}^2$. On the other hand $\operatorname{range} T=\mathbf{F}^2$, so $P_{\operatorname{range} T}$ is also the identity operator on $\mathbf{F}^2$ (by 6.57(b)). Hence $TT^\dagger=P_{\operatorname{range} T}$, as 6.69(b) asserts.
(c) Let $(a,b,c)\in\mathbf{F}^3$ and set $x=a+b+c$ and $y=2b+3c$, so that $T(a,b,c)=(x,y)$. Then, by (a),

$$\begin{aligned}T^\dagger T(a,b,c)&=\tfrac{1}{14}\big(13x-5y,\ 3x+y,\ -2x+4y\big)\\&=\tfrac{1}{14}\Big(13(a+b+c)-5(2b+3c),\ 3(a+b+c)+(2b+3c),\ -2(a+b+c)+4(2b+3c)\Big)\\&=\tfrac{1}{14}\big(13a+3b-2c,\ 3a+5b+6c,\ -2a+6b+10c\big).\end{aligned}$$

Now compute $P_{(\operatorname{null} T)^\perp}$. By Exercise 6C.5, $P_{(\operatorname{null} T)^\perp}=I-P_{\operatorname{null} T}$. Since $\operatorname{null} T=\operatorname{span}((1,-3,2))$ and $\|(1,-3,2)\|^2=1+9+4=14$, the formula 6.56 for the orthogonal projection onto a one-dimensional subspace gives

$$P_{\operatorname{null} T}(a,b,c)=\frac{\langle(a,b,c),(1,-3,2)\rangle}{14}(1,-3,2)=\frac{a-3b+2c}{14}(1,-3,2).$$

Hence

$$\begin{aligned}P_{(\operatorname{null} T)^\perp}(a,b,c)&=(a,b,c)-\frac{a-3b+2c}{14}(1,-3,2)\\&=\tfrac{1}{14}\Big(14a-(a-3b+2c),\ 14b+3(a-3b+2c),\ 14c-2(a-3b+2c)\Big)\\&=\tfrac{1}{14}\big(13a+3b-2c,\ 3a+5b+6c,\ -2a+6b+10c\big).\end{aligned}$$

The two displayed formulas agree for every $(a,b,c)\in\mathbf{F}^3$, so $T^\dagger T=P_{(\operatorname{null} T)^\perp}$, as 6.69(c) asserts.

Problem 6C.22. Suppose V is finite-dimensional and $T\in\mathcal{L}(V,W)$. Prove that

$$TT^\dagger T=T\quad\text{and}\quad T^\dagger TT^\dagger=T^\dagger.$$

Both formulas above clearly hold if T is invertible because in that case we can replace $T^\dagger$ with T^{-1}.

Solution. Throughout, recall the definition 6.68 of the pseudoinverse: since V is finite-dimensional, both $\operatorname{null} T$ and $\operatorname{range} T$ are finite-dimensional subspaces (of V and of W respectively), the restriction $T|_{(\operatorname{null} T)^\perp}$ is an invertible linear map from $(\operatorname{null} T)^\perp$ onto $\operatorname{range} T$ (by 6.67), and

$$T^\dagger w=\big(T|_{(\operatorname{null} T)^\perp}\big)^{-1}P_{\operatorname{range} T}\,w$$

for every $w\in W$.
Step 1. We first show $TP_{(\operatorname{null} T)^\perp}=T$.
Write $U=(\operatorname{null} T)^\perp$, a finite-dimensional subspace of V; since $\operatorname{null} T$ is also finite-dimensional, 6.52 gives $U^\perp=\big((\operatorname{null} T)^\perp\big)^\perp=\operatorname{null} T$. Let $v\in V$. By 6.49 applied to $\operatorname{null} T$ we may write $v=x+u$ with $x\in U$ and $u\in\operatorname{null} T=U^\perp$. This is precisely the decomposition used in the definition 6.55 of P_U, so $P_{(\operatorname{null} T)^\perp}v=P_Uv=x$. Hence

$$TP_{(\operatorname{null} T)^\perp}v=Tx=Tx+Tu=T(x+u)=Tv,$$

where the second equality holds because $Tu=0$. As $v\in V$ was arbitrary, $TP_{(\operatorname{null} T)^\perp}=T$.
Step 2. Proof that $TT^\dagger T=T$.

By 6.69(c) we have $T^\dagger T = P_{(\operatorname{null} T)^\perp}$. Therefore

$$TT^\dagger T = T\big(T^\dagger T\big) = TP_{(\operatorname{null} T)^\perp} = T,$$

the last equality being Step 1.
Step 3. Proof that $T^\dagger T T^\dagger = T^\dagger$.
By 6.69(b) we have $TT^\dagger = P_{\operatorname{range} T}$. Therefore, using the definition of $T^\dagger$ and the idempotence $P^2_{\operatorname{range} T} = P_{\operatorname{range} T}$ of an orthogonal projection [6.57(g)], we get for every $w \in W$

$$\begin{aligned} T^\dagger T T^\dagger w &= T^\dagger\big(P_{\operatorname{range} T}\, w\big) \\ &= \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1} P_{\operatorname{range} T}\big(P_{\operatorname{range} T}\, w\big) \\ &= \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1} P_{\operatorname{range} T}\, w \\ &= T^\dagger w. \end{aligned}$$

Hence $T^\dagger T T^\dagger = T^\dagger$, completing the proof.
As a remark, the two identities just proved are the first two of the four Moore–Penrose conditions; the remaining two say that $TT^\dagger$ and $T^\dagger T$ are self-adjoint. That holds here too: by 6.69(b) and 6.69(c) each of these maps is an orthogonal projection P_U onto a finite-dimensional subspace U, and any such P_U satisfies $\langle P_U x, y\rangle = \langle x, P_U y\rangle$, since writing $x = u + w$ and $y = u' + w'$ with $u, u' \in U$ and $w, w' \in U^\perp$ gives $\langle P_U x, y\rangle = \langle u, u'\rangle = \langle x, P_U y\rangle$. (This phrasing avoids assuming that W is finite-dimensional, which is not part of the hypotheses here and which the book's definition of the adjoint would require.)

Problem 6C.23. Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that

$$\big(T^\dagger\big)^\dagger = T.$$

The equation above is analogous to the equation $(T^{-1})^{-1} = T$ that holds if T is invertible.

Solution. Since V is finite-dimensional, the pseudoinverse $T^\dagger \in \mathcal{L}(W, V)$ is defined (6.68); since W is finite-dimensional, the pseudoinverse of $T^\dagger$ is in turn defined, and $\big(T^\dagger\big)^\dagger \in \mathcal{L}(V, W)$. Thus both sides of the asserted equation are linear maps from V to W.
Write $N = \operatorname{null} T$ and $R = \operatorname{range} T$, and let

$$S = T|_{N^\perp},$$

which by 6.67 is an invertible linear map from $N^\perp$ onto R. The definition 6.68 of the pseudoinverse reads

$$T^\dagger w = S^{-1} P_R\, w \quad \text{for every } w \in W.$$

Step 1. We compute $\operatorname{null} T^\dagger$ and $\operatorname{range} T^\dagger$ (this is Exercise 20 of this section, proved here so that the argument is self-contained).
First, $\operatorname{null} T^\dagger = R^\perp$. Indeed, if $w \in R^\perp$ then $P_R w = 0$ by 6.57(c), so $T^\dagger w = S^{-1}0 = 0$. Conversely, if $T^\dagger w = 0$, then $S^{-1} P_R w = 0$; and since S^{-1} is injective this forces $P_R w = 0$, so $w \in \operatorname{null} P_R = R^\perp$ by 6.57(e).
Second, $\operatorname{range} T^\dagger = N^\perp$. For every $w \in W$ the vector $T^\dagger w = S^{-1} P_R w$ lies in $N^\perp$, because S^{-1} maps into $N^\perp$; hence $\operatorname{range} T^\dagger \subseteq N^\perp$. Conversely, if $x \in N^\perp$, put $w = Tx = Sx \in R$; then $P_R w = w$ by 6.57(b), so

$$T^\dagger w = S^{-1} P_R w = S^{-1} S x = x,$$

and therefore $x \in \operatorname{range} T^\dagger$. Thus $\operatorname{range} T^\dagger = N^\perp$.
Step 2. Identification of the data entering the definition of $\big(T^\dagger\big)^\dagger$.
By definition 6.68 applied to the map $T^\dagger \in \mathcal{L}(W, V)$ (whose domain W is finite-dimensional),

$$\big(T^\dagger\big)^\dagger v = \Big(T^\dagger\big|_{(\operatorname{null} T^\dagger)^\perp}\Big)^{-1} P_{\operatorname{range} T^\dagger}\, v$$

for every $v \in V$. By Step 1, $\operatorname{range} T^\dagger = N^\perp$, so the projection appearing above is $P_{N^\perp}$. Also by Step 1, $\operatorname{null} T^\dagger = R^\perp$, and since R is a finite-dimensional subspace of W, 6.52 gives

$$\big(\operatorname{null} T^\dagger\big)^\perp = \big(R^\perp\big)^\perp = R.$$

Hence the map to be inverted is $T^\dagger|_R$.
Step 3. The map $T^\dagger|_R$ equals S^{-1}.
If $w \in R$, then $P_R w = w$ by 6.57(b), so $T^\dagger w = S^{-1} P_R w = S^{-1} w$. Thus $T^\dagger|_R = S^{-1}$, viewed as a map from R onto $N^\perp$, and consequently

$$\Big(T^\dagger\big|_{(\operatorname{null} T^\dagger)^\perp}\Big)^{-1} = \big(S^{-1}\big)^{-1} = S = T|_{N^\perp}.$$

(Both sides are maps from $N^\perp$ onto R; note that this is consistent with Step 2, which identified the domain of $T^\dagger|_{(\operatorname{null} T^\dagger)^\perp}$ as R and, by Step 1, its range as $N^\perp$.)
Step 4. Conclusion.
Combining Steps 2 and 3, for every $v \in V$ we have

$$\big(T^\dagger\big)^\dagger v = S\big(P_{N^\perp} v\big) = T\big(P_{N^\perp} v\big),$$

where the second equality holds because $P_{N^\perp} v \in N^\perp$ and S is the restriction of T to $N^\perp$. It remains to check that $T P_{N^\perp} = T$. Since $N = \operatorname{null} T$ is finite-dimensional, 6.52 gives $\big(N^\perp\big)^\perp = N$, and 6.49 lets us write any $v \in V$ as $v = x + u$ with $x \in N^\perp$ and $u \in N = \big(N^\perp\big)^\perp$; this is exactly the decomposition used in the definition 6.55 of $P_{N^\perp}$, so $P_{N^\perp} v = x$ and

$$T P_{N^\perp} v = Tx = Tx + Tu = Tv,$$

because $Tu = 0$. (This is Exercise 6 of this section, and it also appeared as Step 1 of the solution to Exercise 22.)
Therefore $\big(T^\dagger\big)^\dagger v = T P_{N^\perp} v = Tv$ for every $v \in V$, which is to say

$$\big(T^\dagger\big)^\dagger = T,$$

as desired.
As a sanity check, note the consistency with 6.69(a): if T is invertible, then $T^\dagger = T^{-1}$ is invertible as well, so $\big(T^\dagger\big)^\dagger = \big(T^{-1}\big)^{-1} = T$.

7 Operators on Inner Product Spaces

Contents

7.1 Exercises 7A 362
7.2 Exercises 7B 388
7.3 Exercises 7C 407
7.4 Exercises 7D 425
7.5 Exercises 7E 444
7.6 Exercises 7F 459

7.1 Exercises 7A

Problem 7A.1. Suppose n is a positive integer. Define $T \in \mathcal{L}(\mathbf{F}^n)$ by

$$T(z_1, \dots, z_n) = (0, z_1, \dots, z_{n-1}).$$

Find a formula for $T^*(z_1, \dots, z_n)$.

Solution. The answer is

$$T^*(z_1, \dots, z_n) = (z_2, \dots, z_n, 0).$$

To derive this, equip $\mathbf{F}^n$ with its standard (Euclidean) inner product

$$\langle (a_1, \dots, a_n), (b_1, \dots, b_n) \rangle = a_1\overline{b_1} + \cdots + a_n\overline{b_n}.$$

Following the technique described after 7.3, we start with $\langle Tz, w \rangle$ and manipulate it until z sits alone in the first slot. Let $z = (z_1, \dots, z_n)$ and $w = (w_1, \dots, w_n)$ in $\mathbf{F}^n$. Because the first coordinate of Tz is 0 and its $(k+1)^{\text{st}}$ coordinate is z_k for $k \in \{1, \dots, n-1\}$, we have

$$\langle Tz, w \rangle = 0 \cdot \overline{w_1} + z_1\overline{w_2} + z_2\overline{w_3} + \cdots + z_{n-1}\overline{w_n} = \sum_{k=1}^{n-1} z_k \overline{w_{k+1}}.$$

Now put $u = (w_2, w_3, \dots, w_n, 0) \in \mathbf{F}^n$, so that $u_k = w_{k+1}$ for $k \in \{1, \dots, n-1\}$ and $u_n = 0$. Then

$$\langle z, u \rangle = \sum_{k=1}^{n} z_k \overline{u_k} = \sum_{k=1}^{n-1} z_k \overline{w_{k+1}} + z_n \cdot \overline{0} = \sum_{k=1}^{n-1} z_k \overline{w_{k+1}}.$$

Comparing the two displays gives

$$\langle Tz, w \rangle = \langle z, (w_2, \dots, w_n, 0) \rangle$$

for every $z \in \mathbf{F}^n$. On the other hand, the definition 7.1 of the adjoint says $\langle Tz, w \rangle = \langle z, T^*w \rangle$ for every $z \in \mathbf{F}^n$. Subtracting, $\langle z, T^*w - u \rangle = 0$ for every $z \in \mathbf{F}^n$; taking $z = T^*w - u$ gives $\|T^*w - u\|^2 = 0$, hence $T^*w = u$. That is,

$$T^*(w_1, \dots, w_n) = (w_2, \dots, w_n, 0),$$

which is the claimed formula (after renaming w as z).
As a check, one may instead use 7.9: with respect to the standard basis of $\mathbf{F}^n$, which is orthonormal, the matrix $\mathcal{M}(T)$ has 1 in each entry of the subdiagonal (row $k+1$, column k, for $k \in \{1, \dots, n-1\}$) and 0 elsewhere. Its conjugate transpose has 1 in row k, column $k+1$, for $k \in \{1, \dots, n-1\}$, and 0 elsewhere; that is the matrix of the map sending $(z_1, \dots, z_n)$ to $(z_2, \dots, z_n, 0)$. Since all entries are real, conjugation has no effect, and 7.9 confirms the formula for T^*.
In words: T is the forward shift, and its adjoint is the backward shift.

Problem 7A.2. Suppose $T \in \mathcal{L}(V, W)$. Prove that

$$T = 0 \iff T^* = 0 \iff T^*T = 0 \iff TT^* = 0.$$

Solution. Recall that $T^* \in \mathcal{L}(W, V)$ by 7.4, so $T^*T \in \mathcal{L}(V)$ and $TT^* \in \mathcal{L}(W)$; all four statements make sense. We prove three separate equivalences and then assemble them.
Step 1: $T = 0 \iff T^* = 0$.
Suppose $T = 0$. Then for every $v \in V$ and every $w \in W$,

$$\langle v, T^*w \rangle = \langle Tv, w \rangle = \langle 0, w \rangle = 0.$$

Fixing $w \in W$ and letting v range over V, we may in particular take $v = T^*w$, which gives $\|T^*w\|^2 =$

$\langle T^*w, T^*w\rangle = 0$ and hence $T^*w = 0$. As $w \in W$ was arbitrary, $T^* = 0$.
For the converse, apply what we just proved with T^* in place of T: if $T^* = 0$, then $(T^*)^* = 0$. By 7.5(c) we have $(T^*)^* = T$, so $T = 0$.
Step 2: $T = 0 \iff T^*T = 0$.
If $T = 0$, then $T^*T = T^*0 = 0$ trivially. Conversely, suppose $T^*T = 0$. Let $v \in V$. Using the definition 7.1 of the adjoint with $w = Tv$,

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle v, T^*(Tv)\rangle = \langle v, (T^*T)v\rangle = \langle v, 0\rangle = 0.$$

By the definiteness of the norm, $Tv = 0$. As $v \in V$ was arbitrary, $T = 0$.
Step 3: $T^* = 0 \iff TT^* = 0$.
Apply Step 2 with T^* in place of T. That step gives $T^* = 0 \iff (T^*)^*T^* = 0$, and $(T^*)^* = T$ by 7.5(c), so $T^* = 0 \iff TT^* = 0$.
Conclusion. Combining the three steps,

$$T = 0 \iff T^* = 0 \iff T^*T = 0 \iff TT^* = 0,$$

since Step 1 gives the first equivalence, Step 2 gives $T = 0 \iff T^*T = 0$ (hence $T^* = 0 \iff T^*T = 0$ via Step 1), and Steps 1 and 3 give $T^* = 0 \iff TT^* = 0$. Thus all four conditions are equivalent, as desired.

Problem 7A.3. Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Prove that

$$\lambda \text{ is an eigenvalue of } T \iff \overline{\lambda} \text{ is an eigenvalue of } T^*.$$

Solution. First note the identity

$$(T - \lambda I)^* = T^* - \overline{\lambda}\, I.$$

Indeed, by 7.5(a) we have $\big(T + (-\lambda)I\big)^* = T^* + \big((-\lambda)I\big)^*$, and by 7.5(b) followed by 7.5(e),

$$\big((-\lambda)I\big)^* = \overline{(-\lambda)}\, I^* = -\overline{\lambda}\, I.$$

Now recall that a scalar $\mu \in \mathbf{F}$ is an eigenvalue of an operator $S \in \mathcal{L}(V)$ exactly when $S - \mu I$ is not injective, equivalently when $\operatorname{null}(S - \mu I) \neq \{0\}$.
Using this, we compute a chain of equivalences. Because V is finite-dimensional (a standing assumption of this chapter),

$$\begin{aligned}
\lambda \text{ is an eigenvalue of } T &\iff T - \lambda I \text{ is not injective} \\
&\iff T - \lambda I \text{ is not surjective} \\
&\iff \operatorname{range}(T - \lambda I) \neq V \\
&\iff \big(\operatorname{range}(T - \lambda I)\big)^{\perp} \neq \{0\} \\
&\iff \operatorname{null}\big((T - \lambda I)^*\big) \neq \{0\} \\
&\iff \operatorname{null}\big(T^* - \overline{\lambda}\, I\big) \neq \{0\} \\
&\iff \overline{\lambda} \text{ is an eigenvalue of } T^*.
\end{aligned}$$

Justifications for the steps, in order. The second equivalence is 3.65, applied to the operator $T - \lambda I$ on the finite-dimensional space V (here $\dim V = \dim V$, so injectivity and surjectivity are equivalent). The fourth equivalence is 6.54 applied to the subspace $U = \operatorname{range}(T - \lambda I)$ of V, which is finite-dimensional because V is: that result says $U^{\perp} = \{0\} \iff U = V$, and we have taken the contrapositive of both sides. The fifth equivalence is 7.6(a), which states $\operatorname{null} S^* = (\operatorname{range} S)^{\perp}$, applied with $S = T - \lambda I$. The sixth equivalence is the identity established at the start.
This proves the desired statement in both directions.
Remark. The complex conjugate is essential and is easy to lose sight of. For example, take $V = \mathbf{C}$ and $T \in \mathcal{L}(\mathbf{C})$ defined by $Tz = iz$. Then $T^*z = \overline{i}\, z = -iz$ by 7.5(b), and indeed the eigenvalue i of T corresponds to the eigenvalue $-i = \overline{i}$ of T^*, not to i. Note also that the exercise says nothing

about eigenvectors: an operator and its adjoint may have entirely different eigenvectors. However, if T is normal, then 7.21(e) shows that $Tv = \lambda v$ if and only if $T^*v = \overline{\lambda}v$, so in that case the eigenvectors match up as well.

Problem 7A.4. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that

$$U \text{ is invariant under } T \iff U^\perp \text{ is invariant under } T^*.$$

Solution. Proof of $\Longrightarrow$. Suppose U is invariant under T; thus $Tu \in U$ for every $u \in U$. Let $w \in U^\perp$. To show $T^*w \in U^\perp$, we must show that $\langle u, T^*w\rangle = 0$ for every $u \in U$. So let $u \in U$. Then by the definition 7.1 of the adjoint,

$$\langle u, T^*w\rangle = \langle Tu, w\rangle = 0,$$

where the last equality holds because $Tu \in U$ (by the invariance of U under T) and $w \in U^\perp$, so Tu and w are orthogonal. Since $u \in U$ was arbitrary, T^*w is orthogonal to every vector in U, which means $T^*w \in U^\perp$. Since $w \in U^\perp$ was arbitrary, $U^\perp$ is invariant under T^*.
Proof of $\Longleftarrow$. Suppose $U^\perp$ is invariant under T^*. The implication already proved holds for every operator on V and every subspace of V; apply it with the subspace $U^\perp$ in place of U and the operator T^* in place of T. It yields that $(U^\perp)^\perp$ is invariant under $(T^*)^*$.
Now V is finite-dimensional (a standing assumption of this chapter), hence so is its subspace U, and therefore 6.52 gives

$$(U^\perp)^\perp = U.$$

Also 7.5(c) gives $(T^*)^* = T$. Substituting these two identities into the conclusion of the previous paragraph shows that U is invariant under T, as desired.
Remark. The two directions are genuinely the same statement read through the involutions $U \mapsto U^\perp$ and $T \mapsto T^*$; the finite-dimensionality of V is what makes $U \mapsto U^\perp$ an involution and hence makes the converse free. Combined with 6.49, which gives $V = U \oplus U^\perp$, this exercise is the tool behind statements such as: if U is invariant under a self-adjoint T, then so is $U^\perp$ (take $T^* = T$), so T decomposes as a direct sum of its restrictions to U and to $U^\perp$.

Problem 7A.5. Suppose $T \in \mathcal{L}(V, W)$. Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $f_1, \dots, f_m$ is an orthonormal basis of W. Prove that

$$\|Te_1\|^2 + \cdots + \|Te_n\|^2 = \|T^*f_1\|^2 + \cdots + \|T^*f_m\|^2.$$

[The numbers $\|Te_1\|^2, \dots, \|Te_n\|^2$ in the equation above depend on the orthonormal basis $e_1, \dots, e_n$, but the right side of the equation does not depend on $e_1, \dots, e_n$. Thus the equation above shows that the sum on the left side does not depend on which orthonormal basis $e_1, \dots, e_n$ is used.]

Solution. The strategy is to expand each side into the same double sum of squared absolute values of the numbers $\langle Te_j, f_k\rangle$, which are the entries of the matrix of T with respect to these bases.
Expanding the left side. Fix $j \in \{1, \dots, n\}$. The vector Te_j lies in W, and $f_1, \dots, f_m$ is an orthonormal basis of W, so 6.30(b) gives

$$\|Te_j\|^2 = \big|\langle Te_j, f_1\rangle\big|^2 + \cdots + \big|\langle Te_j, f_m\rangle\big|^2 = \sum_{k=1}^{m} \big|\langle Te_j, f_k\rangle\big|^2.$$

Summing over j,

$$\sum_{j=1}^{n} \|Te_j\|^2 = \sum_{j=1}^{n}\sum_{k=1}^{m} \big|\langle Te_j, f_k\rangle\big|^2.$$

Expanding the right side. By 7.4 we have $T^* \in \mathcal{L}(W, V)$. Fix $k \in \{1, \dots, m\}$. The vector T^*f_k lies in

V, and $e_1, \dots, e_n$ is an orthonormal basis of V, so 6.30(b) gives

$$\|T^* f_k\|^2 = \sum_{j=1}^{n} \big|\langle T^* f_k, e_j\rangle\big|^2.$$

Summing over k,

$$\sum_{k=1}^{m} \|T^* f_k\|^2 = \sum_{k=1}^{m}\sum_{j=1}^{n} \big|\langle T^* f_k, e_j\rangle\big|^2.$$

Matching the two double sums. For each j and k, the definition 7.1 of the adjoint and the conjugate symmetry of the inner product give

$$\langle T^* f_k, e_j\rangle = \overline{\langle e_j, T^* f_k\rangle} = \overline{\langle Te_j, f_k\rangle}.$$

Since a complex number and its conjugate have the same absolute value,

$$\big|\langle T^* f_k, e_j\rangle\big|^2 = \big|\langle Te_j, f_k\rangle\big|^2$$

for all $j \in \{1, \dots, n\}$ and $k \in \{1, \dots, m\}$.
Both double sums are finite sums of nonnegative real numbers indexed by the same finite set of pairs (j, k), and by the display above the summands agree term by term; hence the sums are equal (the order of summation is immaterial for a finite sum). Therefore

$$\|Te_1\|^2 + \cdots + \|Te_n\|^2 = \sum_{j=1}^{n}\sum_{k=1}^{m} \big|\langle Te_j, f_k\rangle\big|^2 = \sum_{k=1}^{m}\sum_{j=1}^{n} \big|\langle T^* f_k, e_j\rangle\big|^2 = \|T^* f_1\|^2 + \cdots + \|T^* f_m\|^2,$$

as desired.
Remark on the bracketed comment. The left side of the identity is built from $e_1, \dots, e_n$ alone and the right side from $f_1, \dots, f_m$ alone. Since the identity holds for every choice of the two orthonormal bases, the common value of the two sides is independent of both. In particular, fixing $f_1, \dots, f_m$ and varying $e_1, \dots, e_n$ shows that $\|Te_1\|^2 + \cdots + \|Te_n\|^2$ is the same for every orthonormal basis $e_1, \dots, e_n$ of V.
Concretely, in the language of 7.9, if $A = \mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_m))$, then the entry of A in row k, column j, equals $\langle Te_j, f_k\rangle$, so the common value is $\sum_{j,k} |A_{k,j}|^2$, the sum of the squared absolute values of all entries of A. This quantity is the square of the Frobenius norm of A; the exercise says it is a basis-independent invariant of T, and that A and its conjugate transpose A^* have the same value.

Problem 7A.6. Suppose $T \in \mathcal{L}(V, W)$. Prove that
(a) T is injective $\iff$ T^* is surjective;
(b) T is surjective $\iff$ T^* is injective.

Solution. Throughout, recall from 7.4 that $T^* \in \mathcal{L}(W, V)$, so that surjectivity of T^* means $\operatorname{range} T^* = V$ and injectivity of T^* means $\operatorname{null} T^* = \{0\}$ (using 3.15, which says a linear map is injective if and only if its null space is $\{0\}$). Both V and W are finite-dimensional by the standing assumptions of this chapter.
We first record a small lemma.
Lemma. If U is a subspace of a finite-dimensional inner product space X, then

$$U^\perp = X \iff U = \{0\} \qquad \text{and} \qquad U^\perp = \{0\} \iff U = X.$$

For the first: by 6.51, $\dim U^\perp = \dim X - \dim U$. Since $U^\perp$ is a subspace of X, result 2.39 (a subspace of full dimension equals the whole space) gives $U^\perp = X$ if and only if $\dim U^\perp = \dim X$, which by the dimension formula holds if and only if $\dim U = 0$, that is, if and only if $U = \{0\}$. The second equivalence is exactly 6.54.
(a) T is injective $\iff$ T^* is surjective.

By 7.6(b),

$$\operatorname{range} T^* = (\operatorname{null} T)^\perp,$$

where $\operatorname{null} T$ is a subspace of V. Hence

$$\begin{aligned} T^* \text{ is surjective} &\iff \operatorname{range} T^* = V \\ &\iff (\operatorname{null} T)^\perp = V \\ &\iff \operatorname{null} T = \{0\} \\ &\iff T \text{ is injective,} \end{aligned}$$

where the third equivalence is the first part of the Lemma applied with $X = V$ and $U = \operatorname{null} T$, and the fourth is 3.15.
(b) T is surjective $\iff$ T^* is injective.
By 7.6(a),

$$\operatorname{null} T^* = (\operatorname{range} T)^\perp,$$

where $\operatorname{range} T$ is a subspace of W. Hence

$$\begin{aligned} T^* \text{ is injective} &\iff \operatorname{null} T^* = \{0\} \\ &\iff (\operatorname{range} T)^\perp = \{0\} \\ &\iff \operatorname{range} T = W \\ &\iff T \text{ is surjective,} \end{aligned}$$

where the first equivalence is 3.15, and the third is the second part of the Lemma applied with $X = W$ and $U = \operatorname{range} T$.
Remark. Part (b) also follows from part (a) by duality: applying (a) to T^* in place of T gives that T^* is injective if and only if $(T^*)^*$ is surjective, and $(T^*)^* = T$ by 7.5(c). Either route is fine; the direct argument above is symmetric to the one for (a) and uses the companion formula 7.6(a).

Problem 7A.7. Prove that if $T \in \mathcal{L}(V, W)$, then
(a) $\dim \operatorname{null} T^* = \dim \operatorname{null} T + \dim W - \dim V$;
(b) $\dim \operatorname{range} T^* = \dim \operatorname{range} T$.

Solution. By 7.4 we have $T^* \in \mathcal{L}(W, V)$, and both V and W are finite-dimensional by the standing assumptions of this chapter, so every dimension appearing below is a well-defined nonnegative integer.
(a) By 7.6(a),

$$\operatorname{null} T^* = (\operatorname{range} T)^\perp,$$

and $\operatorname{range} T$ is a subspace of W. Applying 6.51 (the dimension of an orthogonal complement) inside the space W gives

$$\dim \operatorname{null} T^* = \dim (\operatorname{range} T)^\perp = \dim W - \dim \operatorname{range} T.$$

By the fundamental theorem of linear maps 3.21 applied to $T \in \mathcal{L}(V, W)$,

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T, \qquad \text{so} \qquad \dim \operatorname{range} T = \dim V - \dim \operatorname{null} T.$$

Substituting this into the previous display,

$$\dim \operatorname{null} T^* = \dim W - \big(\dim V - \dim \operatorname{null} T\big) = \dim \operatorname{null} T + \dim W - \dim V,$$

which is (a).
(b) By 7.6(b),

$$\operatorname{range} T^* = (\operatorname{null} T)^\perp,$$

and $\operatorname{null} T$ is a subspace of V. Applying 6.51 inside the space V gives

$$\dim \operatorname{range} T^* = \dim (\operatorname{null} T)^\perp = \dim V - \dim \operatorname{null} T.$$

By 3.21 applied to T, the right side equals $\dim \operatorname{range} T$. Hence

$$\dim \operatorname{range} T^* = \dim \operatorname{range} T,$$

which is (b).
Alternative derivation of (b) from (a). Apply 3.21 to $T^* \in \mathcal{L}(W, V)$, whose domain is W:

$$\dim W = \dim \operatorname{null} T^* + \dim \operatorname{range} T^*.$$

Substituting the formula from part (a) for $\dim \operatorname{null} T^*$,

$$\dim \operatorname{range} T^* = \dim W - \dim \operatorname{null} T^* = \dim W - \big(\dim \operatorname{null} T + \dim W - \dim V\big) = \dim V - \dim \operatorname{null} T,$$

and by 3.21 applied to T this equals $\dim \operatorname{range} T$, as before.
Remark. Part (b) says that T and T^* have the same rank. Since 7.9 identifies the matrix of T^* with respect to orthonormal bases as the conjugate transpose of the matrix of T, part (b) is precisely the statement that a matrix and its conjugate transpose have the same rank; Exercise 8 of this section uses exactly this to obtain another proof that the row rank of a matrix equals its column rank, a result previously proved in 3.57 and 3.133. Note also that (a) shows $\dim \operatorname{null} T^*$ and $\dim \operatorname{null} T$ differ by the constant $\dim W - \dim V$, independent of T; in particular, when $\dim V = \dim W$ (for instance when T is an operator, $W = V$) the two null spaces have the same dimension.

Problem 7A.8. Suppose A is an m-by-n matrix with entries in $\mathbf{F}$. Use (b) in Exercise 7 to prove that the row rank of A equals the column rank of A.
[This exercise asks for yet another alternative proof of a result that was previously proved in 3.57 and 3.133.]

Solution. Recall that part (b) of Exercise 7 states that if $T \in \mathcal{L}(V, W)$, then

$$\dim \operatorname{range} T^* = \dim \operatorname{range} T.$$

Give $\mathbf{F}^{n,1}$ and $\mathbf{F}^{m,1}$ their standard inner products, so that

$$\langle x, y \rangle = x_1 \overline{y_1} + \cdots + x_n \overline{y_n}.$$

With respect to these inner products the standard bases of $\mathbf{F}^{n,1}$ and of $\mathbf{F}^{m,1}$ are orthonormal.
Define $T\colon \mathbf{F}^{n,1} \to \mathbf{F}^{m,1}$ by

$$Tx = Ax.$$

Then $\mathcal{M}(T) = A$, where the matrix is computed with respect to the standard bases of $\mathbf{F}^{n,1}$ and $\mathbf{F}^{m,1}$; indeed, the k^{th} column of A is $Ae_k = Te_k$. Because those bases are orthonormal, 7.9 tells us that

$$\mathcal{M}(T^*) = \big(\mathcal{M}(T)\big)^* = A^*,$$

where A^* denotes the conjugate transpose of A.
Now we compute:

$$\begin{aligned} \text{column rank of } A &= \text{column rank of } \mathcal{M}(T) \\ &= \dim \operatorname{range} T \\ &= \dim \operatorname{range} T^* \\ &= \text{column rank of } \mathcal{M}(T^*) \\ &= \text{column rank of } A^*, \end{aligned}$$

where the second and fourth equalities come from 3.78 and the third equality comes from (b) in Exercise 7.
It remains only to observe that the column rank of A^* equals the row rank of A. Let $r_1, \ldots, r_m \in \mathbf{F}^{1,n}$ denote the rows of A. By the definition 7.7 of the conjugate transpose, the columns of A^* are exactly

$\overline{r_1}^{\mathrm{t}}, \ldots, \overline{r_m}^{\mathrm{t}}$, where the bar denotes entrywise complex conjugation. Complex conjugation converts a linear dependence relation

$$\lambda_1 r_1 + \cdots + \lambda_m r_m = 0$$

into the relation $\overline{\lambda_1}\,\overline{r_1} + \cdots + \overline{\lambda_m}\,\overline{r_m} = 0$, and conversely (conjugation is its own inverse). Hence if $r_{k_1}, \ldots, r_{k_c}$ is a maximal linearly independent sublist of $r_1, \ldots, r_m$, then $\overline{r_{k_1}}, \ldots, \overline{r_{k_c}}$ is linearly independent, and every $\overline{r_j}$ is in its span (conjugate the expression of r_j as a linear combination of $r_{k_1}, \ldots, r_{k_c}$). Thus

$$\dim \operatorname{span}(\overline{r_1}, \ldots, \overline{r_m}) = c = \dim \operatorname{span}(r_1, \ldots, r_m).$$

The right side is the row rank of A by definition. For the left side, note that transposition is an isomorphism from $\mathbf{F}^{1,n}$ onto $\mathbf{F}^{n,1}$, so it preserves dimensions of spans; since the columns of A^* are $\overline{r_1}^{\mathrm{t}}, \ldots, \overline{r_m}^{\mathrm{t}}$, the column rank of A^* equals $\dim \operatorname{span}(\overline{r_1}, \ldots, \overline{r_m})$, which is the left side. (If $\mathbf{F} = \mathbf{R}$, then $A^* = A^{\mathrm{t}}$ and this paragraph is immediate.)
Combining the two displays above gives

$$\text{column rank of } A = \text{column rank of } A^* = \text{row rank of } A,$$

as desired.

Problem 7A.9. Prove that the product of two self-adjoint operators on V is self-adjoint if and only if the two operators commute.

Solution. Suppose $S, T \in \mathcal{L}(V)$ are self-adjoint, so $S^* = S$ and $T^* = T$. By 7.5(d),

$$(ST)^* = T^* S^* = TS.$$

First suppose ST is self-adjoint. Then

$$ST = (ST)^* = TS,$$

so S and T commute.
Conversely, suppose S and T commute, meaning $ST = TS$. Then

$$(ST)^* = TS = ST,$$

so ST is self-adjoint.
Thus ST is self-adjoint if and only if $ST = TS$, as desired.

Problem 7A.10. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that T is self-adjoint if and only if

$$\langle Tv, v\rangle = \langle T^* v, v\rangle$$

for all $v \in V$.

Solution. Because $\mathbf{F} = \mathbf{C}$, the space V is a complex inner product space, so 7.13 is available.
Since the inner product is additive in its first slot, for every $v \in V$ we have

$$\langle Tv, v\rangle - \langle T^* v, v\rangle = \big\langle (T - T^*)v,\, v\big\rangle.$$

Hence

$$\langle Tv, v\rangle = \langle T^* v, v\rangle \text{ for all } v \in V \quad \Longleftrightarrow \quad \big\langle (T - T^*)v,\, v\big\rangle = 0 \text{ for all } v \in V.$$

Now apply 7.13 to the operator $T - T^* \in \mathcal{L}(V)$: because V is a complex inner product space,

$$\big\langle (T - T^*)v,\, v\big\rangle = 0 \text{ for all } v \in V \quad \Longleftrightarrow \quad T - T^* = 0.$$

And $T - T^* = 0$ means exactly that $T^* = T$, which is the definition of T being self-adjoint.

Chaining the two equivalences gives: T is self-adjoint if and only if $\langle Tv, v\rangle = \langle T^*v, v\rangle$ for all $v \in V$, as desired.
[Alternative route. By 7.15 we have $\langle T^*v, v\rangle = \overline{\langle Tv, v\rangle}$ for every $v \in V$. So the displayed condition says precisely that $\langle Tv, v\rangle = \overline{\langle Tv, v\rangle}$, i.e., that $\langle Tv, v\rangle \in \mathbf{R}$, for every $v \in V$. By 7.14 (valid because $\mathbf{F} = \mathbf{C}$), that happens if and only if T is self-adjoint.]
Note that the hypothesis $\mathbf{F} = \mathbf{C}$ is essential. If $\mathbf{F} = \mathbf{R}$ and $T \in \mathcal{L}(V)$ is arbitrary, then for all $v \in V$ we have $\langle T^*v, v\rangle = \langle v, Tv\rangle = \langle Tv, v\rangle$, since a real inner product is symmetric. So over $\mathbf{R}$ the displayed condition holds for every operator, self-adjoint or not.

Problem 7A.11. Define an operator $S\colon \mathbf{F}^2 \to \mathbf{F}^2$ by $S(w, z) = (-z, w)$.
(a) Find a formula for S^*.
(b) Show that S is normal but not self-adjoint.
(c) Find all eigenvalues of S.
[If $\mathbf{F} = \mathbf{R}$, then S is the operator on $\mathbf{R}^2$ of counterclockwise rotation by 90°.]

Solution. Throughout, $\mathbf{F}^2$ carries its standard inner product $\langle (w, z), (u, v)\rangle = w\overline{u} + z\overline{v}$, for which the standard basis $e_1 = (1, 0)$, $e_2 = (0, 1)$ is orthonormal.
(a) The claim is that

$$S^*(w, z) = (z, -w).$$

One way to see this is to compute the matrix. We have $Se_1 = S(1, 0) = (0, 1) = e_2$ and $Se_2 = S(0, 1) = (-1, 0) = -e_1$, so

$$\mathcal{M}(S) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Because the standard basis is orthonormal, 7.9 gives

$$\mathcal{M}(S^*) = \big(\mathcal{M}(S)\big)^* = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

which is the matrix of the map $(w, z) \mapsto (z, -w)$.
We can also verify the formula directly from the definition of the adjoint. Let $R(w, z) = (z, -w)$. For all $(w, z), (u, v) \in \mathbf{F}^2$,

$$\big\langle S(w, z), (u, v)\big\rangle = \big\langle (-z, w), (u, v)\big\rangle = -z\overline{u} + w\overline{v},$$

while

$$\big\langle (w, z), R(u, v)\big\rangle = \big\langle (w, z), (v, -u)\big\rangle = w\overline{v} + z\overline{(-u)} = w\overline{v} - z\overline{u}.$$

These agree for all vectors, so $S^* = R$; that is, $S^*(w, z) = (z, -w)$.
(b) For every $(w, z) \in \mathbf{F}^2$,

$$SS^*(w, z) = S(z, -w) = (w, z)$$

and

$$S^*S(w, z) = S^*(-z, w) = (w, z).$$

Thus $SS^* = I = S^*S$; in particular $SS^* = S^*S$, so S is normal (7.18).
However S is not self-adjoint: $S(1, 0) = (0, 1)$ but $S^*(1, 0) = (0, -1)$, and $(0, 1) \neq (0, -1)$ in $\mathbf{F}^2$ for $\mathbf{F} = \mathbf{R}$ and for $\mathbf{F} = \mathbf{C}$. Hence $S \neq S^*$.
(c) Suppose $\lambda \in \mathbf{F}$ and $(w, z) \in \mathbf{F}^2$ satisfy $S(w, z) = \lambda(w, z)$, that is,

$$-z = \lambda w \quad \text{and} \quad w = \lambda z.$$

Substituting the second equation into the first gives $-z = \lambda^2 z$, so $(\lambda^2 + 1)z = 0$. If $z = 0$, then $w = \lambda z = 0$, so (w, z) is not an eigenvector. Hence for an eigenvector we must have $z \neq 0$ and therefore

$$\lambda^2 = -1.$$

If $\mathbf{F} = \mathbf{R}$, no real λ satisfies $\lambda^2 = -1$, so S has no eigenvalues. (This is as expected: a counterclockwise rotation of $\mathbf{R}^2$ by 90° maps no nonzero vector to a scalar multiple of itself.)

If $\mathbf{F} = \mathbf{C}$, then $\lambda^2 = -1$ forces $\lambda = i$ or $\lambda = -i$, and both indeed occur:

$$S(i,1) = (-1,i) = i\,(i,1), \qquad S(-i,1) = (-1,-i) = -i\,(-i,1).$$

So the eigenvalues of S are i and $-i$, with corresponding eigenvectors $(i,1)$ and $(-i,1)$.
(As a check on 7.21(e): $S^*(i,1) = (1,-i) = -i\,(i,1) = \bar{i}\,(i,1)$, consistent with S being normal. Note also that the eigenvectors $(i,1)$ and $(-i,1)$ are orthogonal, as 7.22 predicts: $\langle (i,1),(-i,1)\rangle = i\overline{(-i)}+1 = i\cdot i + 1 = 0$.)

Problem 7A.12. An operator $B \in \mathcal{L}(V)$ is called *skew* if

$$B^* = -B.$$

Suppose that $T \in \mathcal{L}(V)$. Prove that T is normal if and only if there exist commuting operators A and B such that A is self-adjoint, B is a skew operator, and $T = A + B$.

Solution. First suppose T is normal, so $TT^* = T^*T$. Define

$$A = \frac{T+T^*}{2} \qquad \text{and} \qquad B = \frac{T-T^*}{2}.$$

Clearly $A + B = T$. Using 7.5(a), 7.5(b), and 7.5(c) (which give $(S+R)^* = S^* + R^*$, $(\lambda S)^* = \overline{\lambda} S^*$, and $(T^*)^* = T$; note $\frac{1}{2}$ and $-\frac{1}{2}$ are real, so the conjugation is harmless), we get

$$A^* = \frac{T^*+T}{2} = A \quad \text{and} \quad B^* = \frac{T^*-T}{2} = -\frac{T-T^*}{2} = -B.$$

So A is self-adjoint and B is skew.
It remains to check that A and B commute. Expanding,

$$AB = \frac{(T+T^*)(T-T^*)}{4} = \frac{T^2 - TT^* + T^*T - (T^*)^2}{4},$$
$$BA = \frac{(T-T^*)(T+T^*)}{4} = \frac{T^2 + TT^* - T^*T - (T^*)^2}{4}.$$

Subtracting,

$$AB - BA = \frac{2\,(T^*T - TT^*)}{4} = \frac{T^*T - TT^*}{2} = 0,$$

where the last equality holds because T is normal. Thus $AB = BA$.
Conversely, suppose $A, B \in \mathcal{L}(V)$ satisfy $A^* = A$, $B^* = -B$, $AB = BA$, and $T = A + B$. Then by 7.5(a) and 7.5(b),

$$T^* = A^* + B^* = A - B.$$

Hence

$$TT^* = (A+B)(A-B) = A^2 - AB + BA - B^2 = A^2 - B^2,$$
$$T^*T = (A-B)(A+B) = A^2 + AB - BA - B^2 = A^2 - B^2,$$

where in both lines the middle two terms cancel because $AB = BA$. Therefore $TT^* = T^*T$, so T is normal.
Two remarks. First, the decomposition is unique: if $T = A + B$ with A self-adjoint and B skew (commuting or not), then $T^* = A - B$, and adding and subtracting these two equations forces $A = (T+T^*)/2$ and $B = (T-T^*)/2$. So the content of the theorem is that T is normal exactly when its self-adjoint part and its skew part commute.
Second, this is the version of 7.23 that makes sense for both $\mathbf{F} = \mathbf{R}$ and $\mathbf{F} = \mathbf{C}$. When $\mathbf{F} = \mathbf{C}$, an operator B is skew if and only if $B = iC$ for some self-adjoint C (take $C = -iB$, and note $C^* = \overline{-i}\,B^* = i(-B) = -iB = C$; conversely $(iC)^* = -iC^* = -iC$), and B commutes with A if and only if C does. So over $\mathbf{C}$ this exercise reduces exactly to 7.23.

Problem 7A.13. Suppose $\mathbf{F} = \mathbf{R}$. Define $\mathcal{A} \in \mathcal{L}(\mathcal{L}(V))$ by $\mathcal{A}T = T^*$ for all $T \in \mathcal{L}(V)$.
(a) Find all eigenvalues of $\mathcal{A}$.
(b) Find the minimal polynomial of $\mathcal{A}$.

Solution. Write $n = \dim V$; by the standing assumptions of this chapter, $n \geq 1$.
First note that $\mathcal{A}$ really is a linear map on $\mathcal{L}(V)$. Indeed 7.5(a) gives $(S+T)^* = S^* + T^*$ and 7.5(b) gives $(\lambda T)^* = \overline{\lambda} T^* = \lambda T^*$ for $\lambda \in \mathbf{R}$; the hypothesis $\mathbf{F} = \mathbf{R}$ is what makes the second of these hold, since over $\mathbf{C}$ the map $T \mapsto T^*$ is conjugate-linear rather than linear. Also $\mathcal{L}(V)$ is finite-dimensional (of dimension n^2), so it makes sense to speak of the minimal polynomial of $\mathcal{A}$.
The key observation is that 7.5(c) gives $(T^*)^* = T$ for every $T \in \mathcal{L}(V)$, so

$$\mathcal{A}^2 = I,$$

where here I denotes the identity operator on $\mathcal{L}(V)$.
(a) Suppose $\lambda \in \mathbf{R}$ is an eigenvalue of $\mathcal{A}$, with eigenvector $T \neq 0$. Then

$$T = \mathcal{A}^2 T = \mathcal{A}(\lambda T) = \lambda^2 T,$$

and since $T \neq 0$ this forces $\lambda^2 = 1$, so $\lambda \in \{1, -1\}$.
The eigenvectors of $\mathcal{A}$ corresponding to 1 are exactly the nonzero T with $T^* = T$, i.e., the nonzero self-adjoint operators; the eigenvectors corresponding to -1 are exactly the nonzero T with $T^* = -T$, i.e., the nonzero skew operators (in the terminology of Exercise 12).
The identity operator on V is self-adjoint and nonzero (because $V \neq \{0\}$), so 1 is an eigenvalue of $\mathcal{A}$.
Now suppose $n \geq 2$. Let $e_1, \ldots, e_n$ be an orthonormal basis of V and define $B \in \mathcal{L}(V)$ by

$$Be_1 = e_2, \qquad Be_2 = -e_1, \qquad Be_k = 0 \ \text{ for } k \geq 3.$$

Then $\mathcal{M}(B)$ with respect to this orthonormal basis has 1 in row 2, column 1, has -1 in row 1, column 2, and has 0 elsewhere. Its conjugate transpose is its negative, so by 7.9 we get $\mathcal{M}(B^*) = -\mathcal{M}(B)$ and hence $B^* = -B$. Since $B \neq 0$, the number -1 is an eigenvalue of $\mathcal{A}$.
Thus if $\dim V \geq 2$, the eigenvalues of $\mathcal{A}$ are exactly 1 and -1.
The case $n = 1$ is degenerate: then every $T \in \mathcal{L}(V)$ is of the form λI with $\lambda \in \mathbf{R}$, and $(\lambda I)^* = \lambda I$ by 7.5(b) and 7.5(e), so every operator on V is self-adjoint and there are no nonzero skew operators. In that case $\mathcal{A}$ is the identity operator on the one-dimensional space $\mathcal{L}(V)$, and its only eigenvalue is 1.
(b) Let $q(z) = z^2 - 1$. Because $\mathcal{A}^2 = I$, we have $q(\mathcal{A}) = \mathcal{A}^2 - I = 0$, so by 5.29 the minimal polynomial of $\mathcal{A}$ divides $q(z) = (z-1)(z+1)$.
Suppose $n \geq 2$. By (a) both 1 and -1 are eigenvalues of $\mathcal{A}$, and by 5.27 every eigenvalue of $\mathcal{A}$ is a zero of the minimal polynomial. Hence both $z-1$ and $z+1$ divide the minimal polynomial. Since the minimal polynomial also divides $(z-1)(z+1)$ and is monic, we conclude that the minimal polynomial of $\mathcal{A}$ is

$$z^2 - 1.$$

(Equivalently, without citing 5.27: the minimal polynomial is monic of degree at most 2; it is not $z-1$ because $\mathcal{A} \neq I$, as witnessed by the skew operator B above with $\mathcal{A}B = -B \neq B$; and it is not $z+1$ because $\mathcal{A} \neq -I$, as witnessed by $\mathcal{A}I_V = I_V \neq -I_V$. It cannot have degree 1 in any other way and cannot have degree 0, so it is $z^2 - 1$.)
Again the case $n = 1$ is exceptional: there $\mathcal{A}$ is the identity on $\mathcal{L}(V)$, so its minimal polynomial is $z - 1$.
Finally, note that the answer to (b) is consistent with the structure of $\mathcal{A}$: since $z^2 - 1$ has distinct real zeros, 5.62 says $\mathcal{A}$ is diagonalizable, and indeed $\mathcal{L}(V)$ is the direct sum of the subspace of self-adjoint operators (the eigenspace for 1) and the subspace of skew operators (the eigenspace for -1), via $T = \frac{T+T^*}{2} + \frac{T-T^*}{2}$ as in Exercise 12.

Problem 7A.14. Define an inner product on $\mathcal{P}_2(\mathbf{R})$ by $\langle p, q\rangle = \int_0^1 pq$. Define an operator $T \in \mathcal{L}(\mathcal{P}_2(\mathbf{R}))$ by

$$T(ax^2 + bx + c) = bx.$$

(a) Show that with this inner product, the operator T is not self-adjoint.
(b) The matrix of T with respect to the basis $1, x, x^2$ is

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This matrix equals its conjugate transpose, even though T is not self-adjoint. Explain why this is not a contradiction.

Solution. (a) By the definition 7.10 of self-adjoint together with the definition 7.1 of the adjoint, T is self-adjoint if and only if $\langle Tp, q\rangle = \langle p, Tq\rangle$ for all $p, q \in \mathcal{P}_2(\mathbf{R})$. So it suffices to exhibit one pair for which this fails. Take

$$p = x \quad \text{and} \quad q = 1.$$

Then $Tp = Tx = x$ (here $a = 0$, $b = 1$, $c = 0$) and $Tq = T1 = 0$ (here $a = b = 0$, $c = 1$). Hence

$$\langle Tp, q\rangle = \langle x, 1\rangle = \int_0^1 x\,dx = \frac{1}{2},$$

whereas

$$\langle p, Tq\rangle = \langle x, 0\rangle = 0.$$

Since $\frac{1}{2} \neq 0$, we have $\langle Tp, q\rangle \neq \langle p, Tq\rangle$ for this choice of p and q. Therefore T is not self-adjoint.
(Equivalently: $\langle x, T^*1\rangle = \langle Tx, 1\rangle = \frac{1}{2} \neq 0 = \langle x, T1\rangle$, so $T^*1 \neq T1$ and thus $T^* \neq T$.)
(b) There is no contradiction because the result 7.9, which says that $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$, requires the matrices to be computed with respect to an *orthonormal* basis. The basis $1, x, x^2$ is not orthonormal with respect to the inner product $\langle p, q\rangle = \int_0^1 pq$. For example,

$$\langle 1, x\rangle = \int_0^1 x\,dx = \frac{1}{2} \neq 0,$$

so 1 and x are not orthogonal, and

$$\langle x, x\rangle = \int_0^1 x^2\,dx = \frac{1}{3} \neq 1,$$

so x is not even a unit vector. Indeed no two of $1, x, x^2$ are orthogonal, since $\int_0^1 x^{j+k}\,dx > 0$ for all nonnegative integers j, k.
As the caution accompanying 7.9 states: with respect to a nonorthonormal basis, the matrix of T^* need not equal the conjugate transpose of the matrix of T. So the symmetry of the displayed matrix tells us nothing about whether T is self-adjoint.
To make this fully concrete, one can compute $\mathcal{M}(T^*)$ with respect to the basis $1, x, x^2$. Let $A = \mathcal{M}(T)$ be the displayed matrix and let G be the Gram matrix of the basis, so

$$G_{j,k} = \langle x^{j-1}, x^{k-1}\rangle = \int_0^1 x^{j+k-2}\,dx = \frac{1}{j+k-1}$$

for $j, k \in \{1, 2, 3\}$; thus G is the 3-by-3 Hilbert matrix, which is invertible (being the Gram matrix of a basis). If v and w have coordinate columns u and y with respect to $1, x, x^2$, then $\langle v, w\rangle = u^{\mathrm{t}} G y$, so

$$\langle Tv, w\rangle = (Au)^{\mathrm{t}} G y = u^{\mathrm{t}} A^{\mathrm{t}} G y \quad \text{and} \quad \langle v, T^*w\rangle = u^{\mathrm{t}} G\, \mathcal{M}(T^*)\, y.$$

Since u and y are arbitrary, $A^{\mathrm{t}} G = G\, \mathcal{M}(T^*)$, that is,

$$\mathcal{M}(T^*) = G^{-1} A^{\mathrm{t}} G.$$

Now $\mathcal{M}(T^*) = A$ would require $AG = GA$ (using $A^{\mathrm{t}} = A$). But AG has row 2 equal to row 2 of G and all other rows zero, while GA has column 2 equal to column 2 of G and all other columns zero; comparing the entries in row 2, column 1 gives $\frac{1}{2}$ versus 0. So $AG \neq GA$ and therefore $\mathcal{M}(T^*) \neq \mathcal{M}(T)$, confirming again that T is not self-adjoint.

Problem 7A.15. Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that
- (a) T is self-adjoint $\iff$ T^{-1} is self-adjoint;
- (b) T is normal $\iff$ T^{-1} is normal.

Solution. Throughout we use 7.5(f): if T is invertible, then T^* is invertible and

$$(T^*)^{-1} = (T^{-1})^*.$$

Note also that T^{-1} is itself an invertible operator on V, with $(T^{-1})^{-1} = T$. This lets us obtain each reverse implication by applying the forward implication to T^{-1} in place of T.
(a) Suppose T is self-adjoint, so $T^* = T$. Then

$$(T^{-1})^* = (T^*)^{-1} = T^{-1},$$

where the first equality is 7.5(f) and the second uses $T^* = T$. Hence T^{-1} is self-adjoint.
Conversely, suppose T^{-1} is self-adjoint. Applying the implication just proved to the invertible operator T^{-1} shows that $(T^{-1})^{-1}$ is self-adjoint. Because $(T^{-1})^{-1} = T$, we conclude that T is self-adjoint.
(b) Suppose T is normal, so that

$$TT^* = T^*T.$$

Both sides of this equation are invertible operators (each is a composition of invertible operators), and equal operators have equal inverses. Taking inverses of both sides and using $(AB)^{-1} = B^{-1}A^{-1}$ gives

$$(T^*)^{-1}T^{-1} = T^{-1}(T^*)^{-1}.$$

Now substitute $(T^*)^{-1} = (T^{-1})^*$, which holds by 7.5(f). The displayed equation becomes

$$(T^{-1})^*T^{-1} = T^{-1}(T^{-1})^*,$$

which says precisely that T^{-1} commutes with its adjoint. Hence T^{-1} is normal.
Conversely, suppose T^{-1} is normal. Applying the implication just proved to the invertible operator T^{-1} shows that $(T^{-1})^{-1} = T$ is normal.

Problem 7A.16. Suppose $\mathbf{F} = \mathbf{R}$.
- (a) Show that the set of self-adjoint operators on V is a subspace of $\mathcal{L}(V)$.
- (b) What is the dimension of the subspace of $\mathcal{L}(V)$ in (a) [in terms of $\dim V$]?

Solution. Let $n = \dim V$ and let

$$\mathcal{S} = \{T \in \mathcal{L}(V) : T^* = T\}$$

denote the set of self-adjoint operators on V.
(a) The zero operator satisfies $0^* = 0$ (immediately from the definition of the adjoint, since $\langle 0v, w\rangle = 0 = \langle v, 0w\rangle$ for all $v, w \in V$), so $0 \in \mathcal{S}$.
Suppose $S, T \in \mathcal{S}$. Then 7.5(a) gives

$$(S+T)^* = S^* + T^* = S + T,$$

so $S + T \in \mathcal{S}$. Thus $\mathcal{S}$ is closed under addition.
Suppose $T \in \mathcal{S}$ and $\lambda \in \mathbf{F} = \mathbf{R}$. Then 7.5(b) gives

$$(\lambda T)^* = \overline{\lambda}\, T^* = \lambda T,$$

where the last equality uses $\overline{\lambda} = \lambda$ (valid because λ is real) and $T^* = T$. Thus $\lambda T \in \mathcal{S}$, so $\mathcal{S}$ is closed under scalar multiplication.
Hence $\mathcal{S}$ is a subspace of $\mathcal{L}(V)$. (Note that the real hypothesis was used only in the last step; see 7A.17 for what goes wrong when $\mathbf{F} = \mathbf{C}$.)
(b) We claim that

$$\dim \mathcal{S} = \frac{n(n+1)}{2} = \frac{(\dim V)(\dim V + 1)}{2}.$$

By 6.35 there is an orthonormal basis $e_1, \dots, e_n$ of V. All matrices below are taken with respect to this basis, and we write $\mathcal{M}(T) = \mathcal{M}(T, (e_1, \dots, e_n))$. By 3.71 the map

$$\mathcal{M} : \mathcal{L}(V) \to \mathbf{R}^{n,n}$$

is an isomorphism of vector spaces.
Because $e_1, \dots, e_n$ is orthonormal, 7.9 gives $\mathcal{M}(T^*) = \mathcal{M}(T)^*$, where A^* denotes the conjugate transpose. Since $\mathbf{F} = \mathbf{R}$, all entries are real and the conjugate transpose is just the transpose A^{t}. Hence, using injectivity of $\mathcal{M}$,

$$T \in \mathcal{S} \iff \mathcal{M}(T^*) = \mathcal{M}(T) \iff \mathcal{M}(T)^{\mathrm{t}} = \mathcal{M}(T).$$

Thus $\mathcal{M}$ restricts to an isomorphism from $\mathcal{S}$ onto the subspace

$$\mathcal{Y} = \{A \in \mathbf{R}^{n,n} : A^{\mathrm{t}} = A\}$$

of symmetric n-by-n real matrices. Isomorphisms preserve dimension, so $\dim \mathcal{S} = \dim \mathcal{Y}$.
It remains to compute $\dim \mathcal{Y}$. For $j, k \in \{1, \dots, n\}$ let $E_{j,k} \in \mathbf{R}^{n,n}$ be the matrix with 1 in row j, column k, and 0 elsewhere; these matrices form the standard basis of $\mathbf{R}^{n,n}$. Consider the list

$$E_{j,j} \ (1 \le j \le n), \qquad E_{j,k} + E_{k,j} \ (1 \le j < k \le n).$$

Each of these matrices is symmetric, so the list lies in $\mathcal{Y}$.
The list spans $\mathcal{Y}$: if $A \in \mathcal{Y}$ has entries $A_{j,k}$, then symmetry gives $A_{k,j} = A_{j,k}$, so

$$A = \sum_{j=1}^{n} A_{j,j} E_{j,j} + \sum_{1 \le j < k \le n} A_{j,k} \, (E_{j,k} + E_{k,j}).$$

The list is linearly independent: if a linear combination

$$\sum_{j=1}^{n} c_j E_{j,j} + \sum_{1 \le j < k \le n} c_{j,k} \, (E_{j,k} + E_{k,j})$$

equals the zero matrix, then reading off the entry in row j, column j gives $c_j = 0$, and reading off the entry in row j, column k with $j < k$ gives $c_{j,k} = 0$.
Hence the list is a basis of $\mathcal{Y}$. Its length is

$$n + \#\{(j,k) : 1 \le j < k \le n\} = n + \binom{n}{2} = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}.$$

Therefore the subspace of self-adjoint operators on V has dimension $\dfrac{(\dim V)(\dim V + 1)}{2}$.
As a check, $\dim \mathcal{L}(V) = n^2$ by 3.72, and indeed $\frac{n(n+1)}{2} \le n^2$ for every positive integer n, with equality only when $n = 1$ (when every operator is multiplication by a real scalar, hence self-adjoint).

Problem 7A.17. Suppose $\mathbf{F} = \mathbf{C}$. Show that the set of self-adjoint operators on V is not a subspace of $\mathcal{L}(V)$.

Solution. Let $\mathcal{S} = \{T \in \mathcal{L}(V) : T^* = T\}$ be the set of self-adjoint operators on V. Recall the standing assumption of this chapter that V is a nonzero finite-dimensional inner product space, so $V \neq \{0\}$.
Let I denote the identity operator on V. By 7.5(e) we have $I^* = I$, so $I \in \mathcal{S}$.
We show that $\mathcal{S}$ is not closed under scalar multiplication by exhibiting a scalar $\lambda \in \mathbf{C}$ with $\lambda I \notin \mathcal{S}$. Take $\lambda = i$. By 7.5(b) and 7.5(e),

$$(iI)^* = \bar{i}\, I^* = -iI.$$

Thus iI is self-adjoint if and only if $-iI = iI$, that is, if and only if $2i\, I = 0$, that is, if and only if $I = 0$. Because $V \neq \{0\}$, we have $I \neq 0$; indeed, choosing any nonzero $v \in V$ gives $Iv = v \neq 0$. Hence $iI \neq (iI)^*$, so $iI \notin \mathcal{S}$.
Since $I \in \mathcal{S}$, $i \in \mathbf{C} = \mathbf{F}$, and $iI \notin \mathcal{S}$, the set $\mathcal{S}$ is not closed under multiplication by scalars in $\mathbf{F}$. Therefore $\mathcal{S}$ is not a subspace of $\mathcal{L}(V)$.
Remark. The failure is exactly the complex conjugate appearing in 7.5(b): when $\mathbf{F} = \mathbf{C}$, the map $T \mapsto T^*$ is not linear, and $\mathcal{S}$ is only a real subspace (it is closed under addition and under multiplication by real scalars, by the argument in 7A.16(a)). This matches the analogy in which self-adjoint operators play the role of the real numbers inside $\mathbf{C}$: the set $\mathbf{R}$ is not a subspace of the complex vector space $\mathbf{C}$ either, for the same reason.

Problem 7A.18. Suppose $\dim V \geq 2$. Show that the set of normal operators on V is not a subspace of $\mathcal{L}(V)$.

Solution. Let $\mathcal{N} = \{T \in \mathcal{L}(V) : TT^* = T^*T\}$ be the set of normal operators on V. We produce two normal operators whose sum is not normal; this shows $\mathcal{N}$ is not closed under addition, hence is not a subspace of $\mathcal{L}(V)$.
Let $n = \dim V \geq 2$ and, using 6.35, let $e_1, \ldots, e_n$ be an orthonormal basis of V. All matrices below are with respect to this basis; since the basis is orthonormal, 7.9 applies: $\mathcal{M}(T^*) = \mathcal{M}(T)^*$, the conjugate transpose.
Define $S, T \in \mathcal{L}(V)$ by specifying them on the basis (legitimate by the linear map lemma 3.4):

$$Se_1 = e_2, \quad Se_2 = e_1, \quad Se_j = 0 \ \text{ for } j \geq 3,$$
$$Te_1 = -e_2, \quad Te_2 = e_1, \quad Te_j = 0 \ \text{ for } j \geq 3.$$

In block form, $\mathcal{M}(S)$ and $\mathcal{M}(T)$ are the n-by-n matrices whose upper-left 2-by-2 blocks are

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

respectively, with all other entries 0. (Note that all entries are real, so these operators are defined whether $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$.)
S is normal. The matrix $\mathcal{M}(S)$ has real entries and is symmetric, so $\mathcal{M}(S)^* = \mathcal{M}(S)$. By 7.9 this gives $\mathcal{M}(S^*) = \mathcal{M}(S)$, and since $\mathcal{M}$ is injective (3.71) we get $S^* = S$. Thus S is self-adjoint, hence normal, since a self-adjoint operator trivially commutes with its adjoint (as noted after 7.18).
T is normal. The matrix $\mathcal{M}(T)$ has real entries, and its transpose is $-\mathcal{M}(T)$. So $\mathcal{M}(T^*) = \mathcal{M}(T)^* = -\mathcal{M}(T) = \mathcal{M}(-T)$, and injectivity of $\mathcal{M}$ gives $T^* = -T$. Hence

$$TT^* = T(-T) = -T^2 = (-T)T = T^*T,$$

so T is normal. (That is, T is a skew operator, in the terminology of Exercise 12 in this section, and every skew operator is normal.)
$S + T$ is not normal. From the definitions,

$$(S+T)e_1 = e_2 - e_2 = 0, \qquad (S+T)e_2 = e_1 + e_1 = 2e_1,$$

and $(S+T)e_j = 0$ for $j \geq 3$. Thus $\mathcal{M}(S+T)$ has upper-left 2-by-2 block $\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$ and all other entries 0. Its conjugate transpose has upper-left block $\begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}$, so by 7.9,

$$(S+T)^* e_1 = 2e_2.$$

Therefore

$$\|(S+T)e_1\| = 0 \qquad \text{while} \qquad \|(S+T)^* e_1\| = 2\|e_2\| = 2.$$

Since these are unequal, 7.20 shows that $S+T$ is not normal.
Thus $S, T \in \mathcal{N}$ but $S+T \notin \mathcal{N}$, so $\mathcal{N}$ is not closed under addition and hence is not a subspace of $\mathcal{L}(V)$. (The hypothesis $\dim V \geq 2$ is needed: if $\dim V = 1$ then every operator on V is multiplication by a scalar, and any two such operators commute, so every operator on V is normal and the set of normal operators equals $\mathcal{L}(V)$, which is a subspace.)

Problem 7A.19. Suppose $T \in \mathcal{L}(V)$ and $\|T^* v\| \leq \|Tv\|$ for every $v \in V$. Prove that T is normal. [This exercise fails on infinite-dimensional inner product spaces, leading to what are called hyponormal operators, which have a well-developed theory.]

Solution. By 7.20, it suffices to prove that $\|Tv\| = \|T^* v\|$ for every $v \in V$. Given the hypothesis, we must upgrade the inequality to an equality; the key is that the two quantities have the same total over any orthonormal basis, so no slack is available.
Step 1: for every orthonormal basis $e_1, \ldots, e_n$ of V,

$$\sum_{k=1}^{n} \|Te_k\|^2 = \sum_{k=1}^{n} \|T^* e_k\|^2.$$

To see this, apply Parseval's identity 6.30(b) to each vector Te_k and to each vector $T^* e_k$:

$$\sum_{k=1}^{n} \|Te_k\|^2 = \sum_{k=1}^{n}\sum_{j=1}^{n} \big|\langle Te_k, e_j\rangle\big|^2, \qquad \sum_{k=1}^{n} \|T^* e_k\|^2 = \sum_{k=1}^{n}\sum_{j=1}^{n} \big|\langle T^* e_k, e_j\rangle\big|^2.$$

Now for all j, k,

$$\langle T^* e_k, e_j\rangle = \overline{\langle e_j, T^* e_k\rangle} = \overline{\langle Te_j, e_k\rangle},$$

where the first equality is conjugate symmetry of the inner product and the second is the defining property of the adjoint. Taking absolute values, $|\langle T^* e_k, e_j\rangle| = |\langle Te_j, e_k\rangle|$. Hence

$$\sum_{k=1}^{n}\sum_{j=1}^{n} \big|\langle T^* e_k, e_j\rangle\big|^2 = \sum_{k=1}^{n}\sum_{j=1}^{n} \big|\langle Te_j, e_k\rangle\big|^2 = \sum_{j=1}^{n}\sum_{k=1}^{n} \big|\langle Te_j, e_k\rangle\big|^2,$$

the last step being merely an interchange of the order of summation in a finite double sum. Relabelling the summation indices ($j \leftrightarrow k$) turns the right side into $\sum_k \sum_j |\langle Te_k, e_j\rangle|^2$, which is the first displayed sum. This proves Step 1.
Step 2: for every orthonormal basis $e_1, \ldots, e_n$ of V we have $\|Te_k\| = \|T^* e_k\|$ for each k.
Indeed, set $c_k = \|Te_k\|^2 - \|T^* e_k\|^2$. The hypothesis $\|T^* e_k\| \leq \|Te_k\|$ gives $c_k \geq 0$ for each k (both norms are nonnegative reals, so squaring preserves the inequality). Step 1 says $\sum_{k=1}^{n} c_k = 0$. A finite sum of nonnegative real numbers is zero only if every term is zero, so $c_k = 0$ for each k, that is, $\|Te_k\| = \|T^* e_k\|$.
Step 3: $\|Tv\| = \|T^* v\|$ for every $v \in V$.
If $v = 0$ this is clear, since $T0 = 0 = T^* 0$. So suppose $v \neq 0$ and set $e = v/\|v\|$. Then $\|e\| = 1$, so the one-element list e is an orthonormal list in V. By 6.36 this list extends to an orthonormal basis $e, e_2, \ldots, e_n$ of V. Applying Step 2 to this basis and to its first vector gives

$$\|Te\| = \|T^* e\|.$$

Multiplying both sides by the positive scalar $\|v\|$ and using homogeneity of the norm together with linearity of T and of T^*,

$$\|Tv\| = \big\|T(\|v\|e)\big\| = \|v\|\,\|Te\| = \|v\|\,\|T^* e\| = \big\|T^*(\|v\|e)\big\| = \|T^* v\|.$$

By 7.20, T is normal.

Remark. Finite-dimensionality is essential in Step 1, which amounts to the statement that T^*T and TT^* have the same trace; the double sums must be finite for the interchange and the cancellation to be legitimate. (Step 3 also uses it, via 6.36.) On an infinite-dimensional space the conclusion genuinely fails. For instance, let T be the unilateral shift on the space of square-summable sequences, $T(a_1, a_2, a_3, \dots) = (0, a_1, a_2, \dots)$, whose adjoint is $T^*(a_1, a_2, a_3, \dots) = (a_2, a_3, \dots)$. Then $\|Tv\| = \|v\|$ and $\|T^*v\| \le \|v\|$ for every v, so $\|T^*v\| \le \|Tv\|$ holds; but $T^*Tv = v$ while $TT^*(1, 0, 0, \dots) = 0$, so T is not normal. Operators satisfying this inequality are the hyponormal operators of the book's remark.

Problem 7A.20. Suppose $P \in \mathcal{L}(V)$ is such that $P^2 = P$. Prove that the following are equivalent.

- (a) P is self-adjoint.
- (b) P is normal.
- (c) There is a subspace U of V such that $P = P_U$.

Solution. We prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a).
(a) $\Rightarrow$ (b). Suppose P is self-adjoint, so $P^* = P$. Then $PP^* = PP = P^*P$, so P commutes with its adjoint. Hence P is normal. (This is the general remark following 7.18: every self-adjoint operator is normal.) Note that the hypothesis $P^2 = P$ was not needed here.
(b) $\Rightarrow$ (c). Suppose P is normal. Set

$$U = \operatorname{range} P,$$

a subspace of V. We claim $P = P_U$.
First we identify $\operatorname{null} P$. Because P is normal, 7.21(a) gives $\operatorname{null} P = \operatorname{null} P^*$. Because $\operatorname{null} P^* = (\operatorname{range} P)^\perp$ by 7.6(a), we conclude

$$\operatorname{null} P = (\operatorname{range} P)^\perp = U^\perp.$$

Next we use $P^2 = P$ to see how P acts on U. If $u \in U = \operatorname{range} P$, then $u = Px$ for some $x \in V$, so

$$Pu = P(Px) = P^2x = Px = u.$$

Thus P is the identity on U.
Now let $v \in V$ be arbitrary. By 6.49 we may write

$$v = u + w \quad \text{with} \quad u \in U,\ w \in U^\perp,$$

and by the definition 6.55 of the orthogonal projection, $P_U v = u$. Applying P and using $Pu = u$ together with $w \in U^\perp = \operatorname{null} P$,

$$Pv = Pu + Pw = u + 0 = u = P_U v.$$

Since $v \in V$ was arbitrary, $P = P_U$, proving (c).
(c) $\Rightarrow$ (a). Suppose U is a subspace of V with $P = P_U$. (Since V is finite-dimensional, U is finite-dimensional and P_U is defined.) We show P_U is self-adjoint by verifying $\langle P_U v_1, v_2\rangle = \langle v_1, P_U v_2\rangle$ for all $v_1, v_2 \in V$.
Let $v_1, v_2 \in V$ and use 6.49 to write

$$v_1 = u_1 + w_1, \qquad v_2 = u_2 + w_2,$$

with $u_1, u_2 \in U$ and $w_1, w_2 \in U^\perp$; so $P_U v_1 = u_1$ and $P_U v_2 = u_2$. Then

$$\langle P_U v_1, v_2\rangle = \langle u_1, u_2 + w_2\rangle = \langle u_1, u_2\rangle + \langle u_1, w_2\rangle = \langle u_1, u_2\rangle,$$

because $u_1 \in U$ and $w_2 \in U^\perp$ force $\langle u_1, w_2\rangle = 0$. Similarly,

$$\langle v_1, P_U v_2\rangle = \langle u_1 + w_1, u_2\rangle = \langle u_1, u_2\rangle + \langle w_1, u_2\rangle = \langle u_1, u_2\rangle,$$

because $\langle w_1, u_2\rangle = \overline{\langle u_2, w_1\rangle} = \overline{0} = 0$. Hence

$$\langle P_U v_1, v_2\rangle = \langle v_1, P_U v_2\rangle \quad \text{for all } v_1, v_2 \in V.$$

By the definition of the adjoint, $\langle P_U v_1, v_2\rangle = \langle v_1, P_U^* v_2\rangle$ for all v_1, v_2. Comparing, $\langle v_1, (P_U^* - P_U)v_2\rangle = 0$ for all $v_1, v_2 \in V$; taking $v_1 = (P_U^* - P_U)v_2$ gives $(P_U^* - P_U)v_2 = 0$ for every v_2. Hence $P_U^* = P_U$, so $P = P_U$ is self-adjoint, proving (a).
This completes the cycle, so (a), (b), (c) are equivalent.
Remark. The hypothesis $P^2 = P$ is consistent with (c): by 6.57(g) every orthogonal projection satisfies $P_U^2 = P_U$. The hypothesis is genuinely needed for (b) $\Rightarrow$ (c): a rotation of $\mathbf{R}^2$ by 90° is normal but is not an orthogonal projection.

Problem 7A.21. Suppose $D\colon \mathcal{P}_8(\mathbf{R}) \to \mathcal{P}_8(\mathbf{R})$ is the differentiation operator defined by $Dp = p'$. Prove that there does not exist an inner product on $\mathcal{P}_8(\mathbf{R})$ that makes D a normal operator.

Solution. The point of the proof is that $\operatorname{null} D$ and $\operatorname{range} D$ are defined without reference to any inner product, and they overlap nontrivially. But 7.21(c) forces these two subspaces to intersect only in $\{0\}$ for a normal operator, no matter which inner product is used.
First note that $\mathcal{P}_8(\mathbf{R})$ is a nonzero finite-dimensional real vector space (its dimension is 9, with basis $1, x, x^2, \ldots, x^8$), so equipping it with any inner product produces an inner product space satisfying the standing assumptions of this chapter, and D is an operator on it.
Compute $\operatorname{null} D$. If $p \in \mathcal{P}_8(\mathbf{R})$ and $p' = 0$, then p is a constant polynomial; conversely every constant polynomial has derivative 0. Thus

$$\operatorname{null} D = \{\text{constant polynomials}\} = \mathcal{P}_0(\mathbf{R}).$$

In particular $\operatorname{null} D$ contains the constant polynomial 1.
Compute $\operatorname{range} D$. If $p \in \mathcal{P}_8(\mathbf{R})$ has degree at most 8, then p' has degree at most 7, so $\operatorname{range} D \subseteq \mathcal{P}_7(\mathbf{R})$. Conversely, if $q \in \mathcal{P}_7(\mathbf{R})$, say $q(x) = \sum_{k=0}^{7} a_k x^k$, then the polynomial

$$p(x) = \sum_{k=0}^{7} \frac{a_k}{k+1}\, x^{k+1}$$

lies in $\mathcal{P}_8(\mathbf{R})$ and satisfies $Dp = q$. Hence

$$\operatorname{range} D = \mathcal{P}_7(\mathbf{R}).$$

In particular $\operatorname{range} D$ also contains the constant polynomial 1; explicitly, $D(x) = 1$.
Therefore

$$1 \in (\operatorname{null} D) \cap (\operatorname{range} D),$$

so this intersection is not $\{0\}$.
Now suppose, seeking a contradiction, that there is an inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{P}_8(\mathbf{R})$ with respect to which D is normal. Then 7.21(c) gives

$$\mathcal{P}_8(\mathbf{R}) = \operatorname{null} D \oplus \operatorname{range} D.$$

By the definition of a direct sum of two subspaces (see 1.46), this requires

$$(\operatorname{null} D) \cap (\operatorname{range} D) = \{0\},$$

contradicting the fact that the nonzero polynomial 1 lies in both subspaces.
Hence no inner product on $\mathcal{P}_8(\mathbf{R})$ makes D a normal operator.
Remark. The same argument applies to any nonzero nilpotent operator. If T is normal, then $\operatorname{null} T^k = \operatorname{null} T$ for every positive integer k: indeed, if $T^2 v = 0$ then Tv lies in $(\operatorname{null} T) \cap (\operatorname{range} T) = \{0\}$ by 7.21(c), so $Tv = 0$, giving $\operatorname{null} T^2 = \operatorname{null} T$, and induction extends this to all k. Since $D^9 = 0$ on $\mathcal{P}_8(\mathbf{R})$, normality of D would give $\operatorname{null} D = \operatorname{null} D^9 = \mathcal{P}_8(\mathbf{R})$, that is $D = 0$, which is false because $D(x) = 1 \neq 0$.

Problem 7A.22. Give an example of an operator $T \in \mathcal{L}(\mathbf{R}^3)$ such that T is normal but not self-adjoint.

Solution. Define $T \in \mathcal{L}(\mathbf{R}^3)$ by

$$T(x, y, z) = (-y, x, z).$$

Geometrically T is the counterclockwise rotation by $90°$ in the x, y-plane, combined with the identity on the z-axis.
The matrix of T with respect to the standard basis e_1, e_2, e_3 of $\mathbf{R}^3$, which is an orthonormal basis with respect to the Euclidean inner product, is

$$\mathcal{M}(T) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

By 7.9, the matrix of T^* with respect to this orthonormal basis is the conjugate transpose of $\mathcal{M}(T)$; because all entries are real, this is just the transpose:

$$\mathcal{M}(T^*) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \text{so} \qquad T^*(x, y, z) = (y, -x, z).$$

Now compute both products. For every $(x, y, z) \in \mathbf{R}^3$,

$$TT^*(x, y, z) = T(y, -x, z) = (x, y, z),$$

$$T^*T(x, y, z) = T^*(-y, x, z) = (x, y, z).$$

Thus $TT^* = I = T^*T$, so T is normal (by the definition 7.18).
However T is not self-adjoint: we have $Te_1 = (0, 1, 0) = e_2$ while $T^*e_1 = (0, -1, 0) = -e_2$, so $T \neq T^*$. Equivalently, the entry in row 2, column 1 of $\mathcal{M}(T)$ equals 1, which does not equal the entry in row 1, column 2, which equals -1; hence $\mathcal{M}(T) \neq \mathcal{M}(T)^*$ and so T is not self-adjoint.
Thus T is normal but not self-adjoint, as desired.

Problem 7A.23. Suppose T is a normal operator on V. Suppose also that $v, w \in V$ satisfy the equations

$$\|v\| = \|w\| = 2, \quad Tv = 3v, \quad Tw = 4w.$$

Show that $\|T(v + w)\| = 10$.

Solution. Because $\|v\| = 2 \neq 0$ and $\|w\| = 2 \neq 0$, the vectors v and w are nonzero. Thus $Tv = 3v$ says that v is an eigenvector of T corresponding to the eigenvalue 3, and $Tw = 4w$ says that w is an eigenvector of T corresponding to the eigenvalue 4.
The eigenvalues 3 and 4 are distinct. Because T is normal, 7.22 tells us that eigenvectors of T corresponding to distinct eigenvalues are orthogonal. Hence

$$\langle v, w \rangle = 0.$$

(For completeness, here is the one-line reason: 7.21(e) gives $T^*w = \bar{4}w = 4w$, and therefore

$$(3 - 4)\langle v, w \rangle = \langle 3v, w \rangle - \langle v, 4w \rangle = \langle Tv, w \rangle - \langle v, T^*w \rangle = 0,$$

which forces $\langle v, w \rangle = 0$.)
Now

$$T(v + w) = Tv + Tw = 3v + 4w.$$

Since $\langle v, w \rangle = 0$, the vectors $3v$ and $4w$ are also orthogonal, so the Pythagorean theorem (6.12) gives

$$\|T(v + w)\|^2 = \|3v + 4w\|^2 = \|3v\|^2 + \|4w\|^2 = 9\|v\|^2 + 16\|w\|^2 = 9 \cdot 4 + 16 \cdot 4 = 100.$$

Taking nonnegative square roots, we conclude that $\|T(v+w)\| = 10$, as desired.

Problem 7A.24. Suppose $T \in \mathcal{L}(V)$ and

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

is the minimal polynomial of T. Prove that the minimal polynomial of T^* is

$$\overline{a_0} + \overline{a_1} z + \overline{a_2} z^2 + \cdots + \overline{a_{m-1}} z^{m-1} + z^m.$$

[This exercise shows that the minimal polynomial of T^* equals the minimal polynomial of T if $\mathbf{F} = \mathbf{R}$.]

Solution. For a polynomial $q(z) = c_0 + c_1 z + \cdots + c_n z^n \in \mathcal{P}(\mathbf{F})$, write

$$\overline{q}(z) = \overline{c_0} + \overline{c_1} z + \cdots + \overline{c_n} z^n$$

for the polynomial obtained by conjugating each coefficient. Note that $\overline{\overline{q}} = q$, that $\deg \overline{q} = \deg q$, and that $\overline{q}$ is monic if and only if q is monic (because $\overline{1} = 1$). If $\mathbf{F} = \mathbf{R}$, then $\overline{q} = q$ for every q.
Step 1: for every $S \in \mathcal{L}(V)$ and every $q \in \mathcal{P}(\mathbf{F})$ we have

$$\big(q(S)\big)^* = \overline{q}(S^*).$$

To see this, first note that $(S^k)^* = (S^*)^k$ for every nonnegative integer k: the case $k = 0$ is 7.5(e), which says $I^* = I$, and the inductive step follows from 7.5(d), since

$$(S^{k+1})^* = (S\, S^k)^* = (S^k)^* S^* = (S^*)^k S^* = (S^*)^{k+1}.$$

Now if $q(z) = c_0 + c_1 z + \cdots + c_n z^n$, then repeated use of 7.5(a) and then 7.5(b) gives

$$\big(q(S)\big)^* = \Big(\sum_{k=0}^{n} c_k S^k\Big)^* = \sum_{k=0}^{n} \big(c_k S^k\big)^* = \sum_{k=0}^{n} \overline{c_k} (S^k)^* = \sum_{k=0}^{n} \overline{c_k} (S^*)^k = \overline{q}(S^*),$$

which proves the claim.
Also note that $0^* = 0$, since $\langle v, 0^* w\rangle = \langle 0v, w\rangle = 0$ for all $v, w \in V$, which forces $0^* w = 0$ for all $w \in V$.
Step 2: let

$$p(z) = a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + z^m$$

denote the minimal polynomial of T, and let q denote the minimal polynomial of T^*. Thus p is the unique monic polynomial of smallest degree with $p(T) = 0$, and q is the unique monic polynomial of smallest degree with $q(T^*) = 0$ (both by 5.22).
Applying Step 1 with $S = T$ and the polynomial p, we get

$$\overline{p}(T^*) = \big(p(T)\big)^* = 0^* = 0.$$

Because $\overline{p}$ is monic of degree m, the minimality of $\deg q$ gives

$$\deg q \le m.$$

Applying Step 1 with $S = T^*$ and the polynomial q, and using $(T^*)^* = T$ from 7.5(c), we get

$$\overline{q}(T) = \big(q(T^*)\big)^* = 0^* = 0.$$

Because $\overline{q}$ is monic of degree $\deg q$, the minimality of $\deg p = m$ gives

$$m \le \deg \overline{q} = \deg q.$$

Combining the two displayed inequalities yields $\deg q = m$. Hence $\overline{q}$ is a monic polynomial of degree m satisfying $\overline{q}(T) = 0$; since m is the smallest degree of a monic polynomial annihilating T, the uniqueness assertion in 5.22 forces

$$\overline{q} = p.$$

Conjugating the coefficients of both sides gives $q = \overline{p}$, that is,

$$q(z) = \overline{a_0} + \overline{a_1}z + \overline{a_2}z^2 + \cdots + \overline{a_{m-1}}z^{m-1} + z^m,$$

as desired.
Finally, if $\mathbf{F} = \mathbf{R}$, then $\overline{a_k} = a_k$ for each k, so the minimal polynomial of T^* equals the minimal polynomial of T.

Problem 7A.25. Suppose $T \in \mathcal{L}(V)$. Prove that T is diagonalizable if and only if T^* is diagonalizable.

Solution. As in Exercise 7A.24, for $q(z) = c_0 + c_1 z + \cdots + c_n z^n \in \mathcal{P}(\mathbf{F})$ let $\overline{q}(z) = \overline{c_0} + \overline{c_1}z + \cdots + \overline{c_n}z^n$ denote the polynomial obtained by conjugating each coefficient of q.
First we record that coefficient-conjugation preserves products: if $q(z) = \sum_j b_j z^j$ and $r(z) = \sum_k c_k z^k$, then the coefficient of z^n in qr equals $\sum_{j+k=n} b_j c_k$, whose complex conjugate equals $\sum_{j+k=n} \overline{b_j}\,\overline{c_k}$, which is exactly the coefficient of z^n in $\overline{q}\,\overline{r}$. Hence

$$\overline{qr} = \overline{q}\,\overline{r},$$

and by induction the same holds for a product of any finite number of polynomials.
Now suppose T is diagonalizable. By 5.62, the minimal polynomial p of T has the form

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m)$$

for some list of distinct numbers $\lambda_1, \ldots, \lambda_m \in \mathbf{F}$. Applying the multiplicativity just proved, together with $\overline{z - \lambda_k} = z - \overline{\lambda_k}$, we obtain

$$\overline{p}(z) = (z - \overline{\lambda_1}) \cdots (z - \overline{\lambda_m}).$$

Each $\overline{\lambda_k}$ lies in $\mathbf{F}$ (this is clear if $\mathbf{F} = \mathbf{C}$, and if $\mathbf{F} = \mathbf{R}$ then $\overline{\lambda_k} = \lambda_k$). Moreover $\overline{\lambda_1}, \ldots, \overline{\lambda_m}$ are distinct, because complex conjugation is injective: if $\overline{\lambda_j} = \overline{\lambda_k}$, then conjugating gives $\lambda_j = \lambda_k$, hence $j = k$.
By Exercise 7A.24, the minimal polynomial of T^* equals $\overline{p}$. Thus the minimal polynomial of T^* equals $(z - \overline{\lambda_1}) \cdots (z - \overline{\lambda_m})$ for a list of distinct numbers $\overline{\lambda_1}, \ldots, \overline{\lambda_m} \in \mathbf{F}$. Hence 5.62 implies that T^* is diagonalizable.
For the other direction, suppose T^* is diagonalizable. Applying what we have just proved with T^* in place of T shows that $(T^*)^*$ is diagonalizable. Because $(T^*)^* = T$ by 7.5(c), we conclude that T is diagonalizable.
Therefore T is diagonalizable if and only if T^* is diagonalizable.

Problem 7A.26. Fix $u, x \in V$. Define $T \in \mathcal{L}(V)$ by $Tv = \langle v, u\rangle x$ for every $v \in V$.
(a) Prove that if V is a real vector space, then T is self-adjoint if and only if the list u, x is linearly dependent.
(b) Prove that T is normal if and only if the list u, x is linearly dependent.

Solution. We begin with two preliminary observations that are used in both parts.
Formula for T^*. For all $v, w \in V$ we have

$$\langle Tv, w\rangle = \langle \langle v, u\rangle x, w\rangle = \langle v, u\rangle\langle x, w\rangle = \langle v, u\rangle\overline{\langle w, x\rangle} = \langle v, \langle w, x\rangle u\rangle.$$

Because T^* is by definition (7.1) the function satisfying $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for all $v, w \in V$, and because the vector in the second slot is determined by this identity, we conclude (as in Example 7.3)

that

$$T^*w = \langle w, x\rangle u \quad \text{for every } w \in V.$$

Dependence of a two-vector list. The list u, x is linearly dependent if and only if $u = 0$ or x is a scalar multiple of u. Indeed, if $au + bx = 0$ with a, b not both 0, then either $b \neq 0$, in which case $x = -(a/b)u$, or $b = 0$, in which case $a \neq 0$ and $au = 0$, so $u = 0$. Conversely, if $u = 0$ then $1 \cdot u + 0 \cdot x = 0$ exhibits dependence, and if $x = cu$ then $cu + (-1)x = 0$ does.
(a) Assume V is a real inner product space, so all scalars and all inner products are real and $\langle v, cu\rangle = c\langle v, u\rangle$ for $c \in \mathbf{R}$.
Suppose first that the list u, x is linearly dependent. If $u = 0$, then $Tv = \langle v, 0\rangle x = 0$ and $T^*w = \langle w, x\rangle \cdot 0 = 0$ for all $v, w \in V$, so $T = 0 = T^*$ and T is self-adjoint. Otherwise $x = cu$ for some $c \in \mathbf{R}$, and then for every $v \in V$,

$$Tv = \langle v, u\rangle x = c\langle v, u\rangle u \quad \text{and} \quad T^*v = \langle v, x\rangle u = \langle v, cu\rangle u = c\langle v, u\rangle u.$$

Thus $T = T^*$, so T is self-adjoint.
Conversely, suppose T is self-adjoint. If $u = 0$, the list u, x is linearly dependent and we are done. So assume $u \neq 0$, which gives $\|u\|^2 \neq 0$. Evaluating T and T^* at u yields

$$Tu = \langle u, u\rangle x = \|u\|^2 x \quad \text{and} \quad T^*u = \langle u, x\rangle u.$$

Since $Tu = T^*u$, we get $\|u\|^2 x = \langle u, x\rangle u$, hence

$$x = \frac{\langle u, x\rangle}{\|u\|^2}\, u,$$

so x is a scalar multiple of u and the list u, x is linearly dependent.
(This is where the hypothesis $\mathbf{F} = \mathbf{R}$ matters, and only in the first direction: over $\mathbf{C}$ dependence need not give self-adjointness. For instance take $V = \mathbf{C}$ with $\langle v, w\rangle = v\overline{w}$, and let $u = 1$, $x = i$. Then u, x is linearly dependent, but $Tv = iv$ while $T^*v = \langle v, i\rangle \cdot 1 = -iv$, so $T \neq T^*$. In general, for $u \neq 0$ and $x = cu$ the computation above gives $Tv = c\langle v, u\rangle u$ and $T^*v = \overline{c}\langle v, u\rangle u$, so T is self-adjoint exactly when $c \in \mathbf{R}$.)
(b) Now $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$.
Suppose first that the list u, x is linearly dependent. If $u = 0$, then as in (a) we have $T = 0 = T^*$, so $TT^* = 0 = T^*T$ and T is normal. Otherwise $x = cu$ for some $c \in \mathbf{F}$, and then for every $v \in V$,

$$T^*Tv = T^*\big(\langle v, u\rangle x\big) = \langle v, u\rangle\langle x, x\rangle u = |c|^2\|u\|^2\langle v, u\rangle u,$$

$$TT^*v = T\big(\langle v, x\rangle u\big) = \langle v, x\rangle\langle u, u\rangle x = \|u\|^2\langle v, cu\rangle\, cu = \|u\|^2\overline{c}c\langle v, u\rangle u = |c|^2\|u\|^2\langle v, u\rangle u,$$

where we used $\langle x, x\rangle = \|cu\|^2 = |c|^2\|u\|^2$. Thus $T^*T = TT^*$, so T is normal.
Conversely, suppose T is normal. If $u = 0$, the list u, x is linearly dependent and we are done. So assume $u \neq 0$. By 7.20, $\|Tv\| = \|T^*v\|$ for every $v \in V$. Taking $v = u$ gives

$$\|u\|^2\|x\| = \big\|\langle u, u\rangle x\big\| = \|Tu\| = \|T^*u\| = \big\|\langle u, x\rangle u\big\| = |\langle u, x\rangle|\,\|u\|.$$

Dividing by $\|u\| \neq 0$ gives

$$|\langle u, x\rangle| = \|u\|\,\|x\|.$$

This is equality in the Cauchy-Schwarz inequality, and by the equality condition in 6.14 this happens only if one of u, x is a scalar multiple of the other. Hence the list u, x is linearly dependent.
Therefore T is normal if and only if the list u, x is linearly dependent.

Problem 7A.27. Suppose $T \in \mathcal{L}(V)$ is normal. Prove that

$$\operatorname{null} T^k = \operatorname{null} T \quad \text{and} \quad \operatorname{range} T^k = \operatorname{range} T$$

for every positive integer k.

Solution. Because T is normal, 7.21(c) gives

$$V = \operatorname{null} T \oplus \operatorname{range} T.$$

By 1.46, a sum of two subspaces is a direct sum if and only if their intersection is $\{0\}$. Hence

$$\operatorname{null} T \cap \operatorname{range} T = \{0\}.$$

We refer to this last equation as the intersection property.
Null spaces. We prove $\operatorname{null} T^k = \operatorname{null} T$ by induction on k.
The case $k = 1$ is trivial.
Next consider $k = 2$. If $v \in \operatorname{null} T$, then $T^2 v = T(Tv) = T0 = 0$, so $v \in \operatorname{null} T^2$; thus $\operatorname{null} T \subseteq \operatorname{null} T^2$. Conversely, suppose $v \in \operatorname{null} T^2$. Then $T(Tv) = 0$, so $Tv \in \operatorname{null} T$; also $Tv \in \operatorname{range} T$. By the intersection property we get $Tv = 0$, so $v \in \operatorname{null} T$. Hence

$$\operatorname{null} T^2 = \operatorname{null} T.$$

Now let $k \geq 1$ and suppose $\operatorname{null} T^k = \operatorname{null} T$. The inclusion $\operatorname{null} T \subseteq \operatorname{null} T^{k+1}$ is clear, because $Tv = 0$ implies $T^{k+1} v = T^k(Tv) = 0$. Conversely, suppose $v \in \operatorname{null} T^{k+1}$. Then $T^k(Tv) = T^{k+1} v = 0$, so

$$Tv \in \operatorname{null} T^k = \operatorname{null} T,$$

using the induction hypothesis. Thus $T^2 v = T(Tv) = 0$, so $v \in \operatorname{null} T^2 = \operatorname{null} T$ by the case $k = 2$ proved above. Hence $\operatorname{null} T^{k+1} = \operatorname{null} T$, completing the induction.
Ranges. Let k be a positive integer. If $w \in \operatorname{range} T^k$, then $w = T^k v = T(T^{k-1} v)$ for some $v \in V$, so $w \in \operatorname{range} T$. Thus

$$\operatorname{range} T^k \subseteq \operatorname{range} T.$$

The fundamental theorem of linear maps (3.21), applied to T^k and to T, gives

$$\dim \operatorname{range} T^k = \dim V - \dim \operatorname{null} T^k = \dim V - \dim \operatorname{null} T = \dim \operatorname{range} T,$$

where the middle equality uses the result on null spaces proved above. So $\operatorname{range} T^k$ is a subspace of the finite-dimensional vector space $\operatorname{range} T$ having the same dimension as $\operatorname{range} T$; by 2.39 this forces

$$\operatorname{range} T^k = \operatorname{range} T,$$

as desired.

Problem 7A.28. Suppose $T \in \mathcal{L}(V)$ is normal. Prove that if $\lambda \in \mathbf{F}$, then the minimal polynomial of T is not a polynomial multiple of $(x - \lambda)^2$.

Solution. Let p denote the minimal polynomial of T; thus p is the monic polynomial of smallest degree such that $p(T) = 0$ (5.22). Fix $\lambda \in \mathbf{F}$ and suppose, seeking a contradiction, that p is a polynomial multiple of $(z - \lambda)^2$. (We write the variable as z; nothing depends on the name of the variable.) So there exists $q \in \mathcal{P}(\mathbf{F})$ with

$$p(z) = (z - \lambda)^2 q(z).$$

Because p is monic, $p \neq 0$, so $q \neq 0$ and $\deg p = \deg q + 2 \geq 2$. Comparing leading coefficients in the equation above shows that q is monic. Define

$$r(z) = (z - \lambda) q(z),$$

so r is monic and $\deg r = \deg p - 1$.

Set $S = T - \lambda I$. By 7.21(d), S is normal, since T is normal. Hence Exercise 7A.27, applied to the normal operator S with $k = 2$, gives

$$\operatorname{null} S^2 = \operatorname{null} S.$$

Now let $v \in V$ be arbitrary. Because $p(z) = (z - \lambda)^2 q(z)$, the multiplicative property 5.17(a) gives $p(T) = (T - \lambda I)^2 q(T) = S^2 q(T)$, and likewise $r(T) = (T - \lambda I) q(T) = S\, q(T)$. Because $p(T) = 0$,

$$S^2\big(q(T)v\big) = p(T)v = 0,$$

so $q(T)v \in \operatorname{null} S^2 = \operatorname{null} S$. Therefore

$$r(T)v = S\big(q(T)v\big) = 0.$$

As $v \in V$ was arbitrary, $r(T) = 0$.
But r is a monic polynomial with $\deg r = \deg p - 1 < \deg p$, contradicting the fact that p is a monic polynomial of smallest degree with $p(T) = 0$. (Equivalently: by 5.29, $r(T) = 0$ implies that r is a polynomial multiple of p, which is impossible for the nonzero polynomial r of degree less than $\deg p$.)
This contradiction shows that the minimal polynomial of T is not a polynomial multiple of $(z - \lambda)^2$ for any $\lambda \in \mathbf{F}$.
Remark. Combined with 5.62, this gives another proof that a normal operator on a complex inner product space is diagonalizable: over $\mathbf{C}$ the minimal polynomial factors as a product of linear factors (by the fundamental theorem of algebra), and what we just proved shows that no factor $(z - \lambda)$ is repeated, so the minimal polynomial is a product of distinct linear factors.

Problem 7A.29. Prove or give a counterexample: If $T \in \mathcal{L}(V)$ and there is an orthonormal basis $e_1, \dots, e_n$ of V such that $\|Te_k\| = \|T^* e_k\|$ for each $k = 1, \dots, n$, then T is normal.

Solution. The statement is **false**. A counterexample is given below.
Before producing it, here is why the hypothesis is genuinely weaker than normality. By 7.20, T is normal if and only if $\|Tv\| = \|T^* v\|$ for **every** $v \in V$, and the hypothesis only imposes this for the n vectors $e_1, \dots, e_n$. The condition $\|Tv\| = \|T^* v\|$ is not linear in v, so it does not propagate from a basis to all of V. In matrix terms: let $A = \mathcal{M}(T, (e_1, \dots, e_n))$. Then Te_k has coordinate vector equal to the k-th column of A, so $\|Te_k\|^2$ is the sum of the squares of the absolute values of the entries of column k. By 7.9, $\mathcal{M}(T^*, (e_1, \dots, e_n)) = A^*$, whose k-th column is the conjugate of the k-th row of A; hence $\|T^* e_k\|^2$ is the sum of the squares of the absolute values of the entries of row k. So the hypothesis says exactly that each column of A has the same norm as the corresponding row. Normality, on the other hand, says $A^* A = A A^*$, which forces agreement of **all** the inner products between columns and between rows, not merely the diagonal ones.
Counterexample: Let $V = \mathbf{F}^2$ with the standard inner product and standard (orthonormal) basis $e_1 = (1, 0)$, $e_2 = (0, 1)$. Define $T \in \mathcal{L}(\mathbf{F}^2)$ by

$$T(x, y) = (x + y, -x).$$

Thus $\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$.
First we compute T^*. For $(x, y), (u, v) \in \mathbf{F}^2$ we have

$$\langle T(x, y), (u, v) \rangle = (x + y)\overline{u} + (-x)\overline{v} = x\overline{(u - v)} + y\overline{u} = \big\langle (x, y), (u - v, u) \big\rangle.$$

Since the adjoint is the unique map satisfying $\langle Tv, w \rangle = \langle v, T^* w \rangle$ for all v, w (definition 7.1), we conclude

$$T^*(u, v) = (u - v, u).$$

Now check the hypothesis:

$$Te_1 = (1, -1), \qquad Te_2 = (1, 0), \qquad T^* e_1 = (1, 1), \qquad T^* e_2 = (-1, 0).$$

Hence $\|Te_1\| = \sqrt{2} = \|T^*e_1\|$ and $\|Te_2\| = 1 = \|T^*e_2\|$. So $\|Te_k\| = \|T^*e_k\|$ for $k = 1, 2$.
But T is not normal:

$$T^*Te_1 = T^*(1,-1) = (1-(-1),1) = (2,1), \qquad TT^*e_1 = T(1,1) = (1+1,-1) = (2,-1).$$

Since $(2,1) \neq (2,-1)$, we have $T^*T \neq TT^*$, so T is not normal.
(Equivalently, one may see the failure of normality from

$$A^*A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = AA^*,$$

where $A = \mathcal{M}(T, (e_1, e_2))$ as above; the off-diagonal entries disagree even though the diagonal entries agree, which is precisely the phenomenon described in the first paragraph. A concrete vector witnessing the failure of the criterion 7.20 is $v = (1,1)$: here $Tv = (2,-1)$ and $T^*v = (0,1)$, so $\|Tv\| = \sqrt{5} \neq 1 = \|T^*v\|$.)
Note that this counterexample works over $\mathbf{F} = \mathbf{R}$ and over $\mathbf{F} = \mathbf{C}$ alike.

Problem 7A.30. Suppose that $T \in \mathcal{L}(\mathbf{F}^3)$ is normal and $T(1,1,1) = (2,2,2)$. Suppose $(z_1, z_2, z_3) \in \operatorname{null} T$. Prove that $z_1 + z_2 + z_3 = 0$.

Solution. Give $\mathbf{F}^3$ its standard inner product, so that $\langle (w_1,w_2,w_3), (u_1,u_2,u_3) \rangle = w_1\overline{u_1} + w_2\overline{u_2} + w_3\overline{u_3}$.
Write $u = (1,1,1)$. The hypothesis says $Tu = (2,2,2) = 2u$, so u is an eigenvector of T corresponding to the eigenvalue 2.
Because T is normal, 7.21(e) applies (with $v = u$ and $\lambda = 2$): $Tu = 2u$ if and only if $T^*u = \overline{2}\,u$. Since $\overline{2} = 2$, we get

$$T^*u = 2u.$$

Now let $z = (z_1, z_2, z_3) \in \operatorname{null} T$, so $Tz = 0$. Using the defining property of the adjoint (7.1), namely $\langle Tz, u \rangle = \langle z, T^*u \rangle$, we compute

$$0 = \langle 0, u \rangle = \langle Tz, u \rangle = \langle z, T^*u \rangle = \langle z, 2u \rangle = \overline{2}\,\langle z, u \rangle = 2\langle z, u \rangle.$$

Hence $\langle z, u \rangle = 0$. But

$$\langle z, u \rangle = z_1\overline{1} + z_2\overline{1} + z_3\overline{1} = z_1 + z_2 + z_3.$$

Therefore $z_1 + z_2 + z_3 = 0$, as desired.
Remark: The argument above is a direct verification, but it is worth recording the conceptual statement behind it. If $z \neq 0$, then z is an eigenvector of T corresponding to the eigenvalue 0, while $u = (1,1,1)$ is an eigenvector corresponding to the eigenvalue 2. Since $0 \neq 2$ and T is normal, 7.22 says that z and u are orthogonal, which is exactly the equation $z_1 + z_2 + z_3 = 0$. (The case $z = 0$ is trivial.)
Note also that the conclusion is exactly the statement $\operatorname{null} T \subseteq \{u\}^\perp$, and normality forces this: since $u = T(u/2)$ we have $u \in \operatorname{range} T$, while 7.6(c) and 7.21(b) give $\operatorname{null} T = (\operatorname{range} T^*)^\perp = (\operatorname{range} T)^\perp$.
Normality is essential. Define $S \in \mathcal{L}(\mathbf{F}^3)$ by $S(x,y,z) = (x+y,\ x+y,\ x+y)$. Then $S(1,1,1) = (2,2,2)$ and $(0,0,1) \in \operatorname{null} S$, yet $0+0+1 \neq 0$. Of course S is not normal: $\operatorname{range} S = \operatorname{span}\big((1,1,1)\big)$, and if S were normal then 7.6(d) together with 7.21(a) would give $\operatorname{range} S = (\operatorname{null} S^*)^\perp = (\operatorname{null} S)^\perp$, forcing $(1,1,1)$ to be orthogonal to $(0,0,1)$, which it is not.

Problem 7A.31. Fix a positive integer n. In the inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product $\langle f, g \rangle = \int_{-\pi}^{\pi} fg$, let

$$V = \operatorname{span}(1, \cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin nx).$$

(a) Define $D \in \mathcal{L}(V)$ by $Df = f'$. Show that $D^* = -D$. Conclude that D is normal but not self-adjoint.
(b) Define $T \in \mathcal{L}(V)$ by $Tf = f''$. Show that T is self-adjoint.

Solution. Throughout, $C[-\pi,\pi]$ denotes the real inner product space of continuous real-valued functions on $[-\pi,\pi]$ with $\langle f,g\rangle = \int_{-\pi}^{\pi} fg$ (this is the space 6.64). Because $\mathbf{F} = \mathbf{R}$ here, the inner product is symmetric and no complex conjugates appear below.

Two preliminary observations.

First, V is a finite-dimensional subspace of $C[-\pi,\pi]$: it is the span of the $2n+1$ functions listed, each of which is continuous on $[-\pi,\pi]$. Finite-dimensionality is what guarantees that adjoints of operators on V exist (the construction of the adjoint in 7.1 rests on the Riesz representation theorem, which requires finite-dimensionality).

Second, and this is the crucial point:

$$f(-\pi) = f(\pi) \quad \text{for every } f \in V.$$

Indeed, the constant function 1 obviously satisfies this; $\cos(-k\pi) = \cos(k\pi)$ for every integer k because cosine is even; and $\sin(-k\pi) = 0 = \sin(k\pi)$ for every integer k. So each of the $2n+1$ spanning functions of V takes the same value at $-\pi$ as at π, and since the condition $f(-\pi) = f(\pi)$ is preserved under linear combinations, every $f \in V$ satisfies it. (Concretely: every element of V is a trigonometric polynomial, hence the restriction to $[-\pi,\pi]$ of a 2π-periodic function on $\mathbf{R}$.)

Also note that V is invariant under differentiation, so D and T really are operators on V: we have $1' = 0 \in V$, $(\cos kx)' = -k\sin kx \in V$, and $(\sin kx)' = k\cos kx \in V$ for $k = 1,\dots,n$. Since differentiation is linear, $Df = f' \in V$ for every $f \in V$. Applying this twice shows $Tf = f'' = D(Df) \in V$ for every $f \in V$; that is,

$$T = D^2.$$

(a) Let $f, g \in V$. Each element of V is a finite linear combination of sines and cosines, hence is infinitely differentiable on $[-\pi,\pi]$, so integration by parts is legitimate:

$$\langle Df, g\rangle = \int_{-\pi}^{\pi} f'g = \Big[fg\Big]_{-\pi}^{\pi} - \int_{-\pi}^{\pi} fg' = f(\pi)g(\pi) - f(-\pi)g(-\pi) - \langle f, Dg\rangle.$$

By the preliminary observation, $f(\pi) = f(-\pi)$ and $g(\pi) = g(-\pi)$, so the boundary term $f(\pi)g(\pi) - f(-\pi)g(-\pi)$ equals 0. Therefore

$$\langle Df, g\rangle = -\langle f, Dg\rangle = \langle f, (-D)g\rangle \quad \text{for all } f, g \in V.$$

The adjoint D^* is by definition (7.1) the unique operator on V satisfying $\langle Df, g\rangle = \langle f, D^*g\rangle$ for all $f, g \in V$; uniqueness holds because if $\langle f, h_1\rangle = \langle f, h_2\rangle$ for all $f \in V$ then $\langle f, h_1 - h_2\rangle = 0$ for all f, and taking $f = h_1 - h_2$ gives $h_1 = h_2$. The displayed equation exhibits $-D$ as such an operator. Hence

$$D^* = -D.$$

D is normal: using $D^* = -D$,

$$DD^* = D(-D) = -D^2 \qquad \text{and} \qquad D^*D = (-D)D = -D^2,$$

so $DD^* = D^*D$, which is exactly the definition (7.18) of D being normal.

D is not self-adjoint: if $D^* = D$, then combining with $D^* = -D$ gives $D = -D$, hence $2D = 0$, hence $D = 0$. But $D \neq 0$, because $n \geq 1$ ensures $\sin x \in V$, and $D(\sin x) = \cos x$, which is not the zero function on $[-\pi,\pi]$ (its value at 0 is 1). So $D^* \neq D$, i.e. D is not self-adjoint (7.10).

Thus D is an operator that is normal but not self-adjoint. (This is a function-space analogue of 7.19.)

(b) Since $T = D^2 = D \circ D$, we may apply 7.5(d) with $S = D$ and the map D, obtaining

$$T^* = (DD)^* = D^*D^* = (-D)(-D) = D^2 = T.$$

Hence T is self-adjoint.

Alternatively, argue directly. Applying the identity $\langle Dh, g\rangle = -\langle h, Dg\rangle$ from part (a) twice, once with $h = Df$ and once with g replaced by Dg, we get for all $f, g \in V$:

$$\langle Tf, g\rangle = \langle D(Df), g\rangle = -\langle Df, Dg\rangle$$

and

$$\langle f, Tg\rangle = \langle f, D(Dg)\rangle = -\langle Df, Dg\rangle.$$

The right-hand sides agree, so $\langle Tf, g\rangle = \langle f, Tg\rangle$ for all $f, g \in V$, which by the uniqueness of the adjoint gives $T^* = T$.

A third viewpoint makes the self-adjointness of T transparent. The functions $1, \cos x, \dots, \cos nx, \sin x, \dots, \sin nx$ are pairwise orthogonal with respect to $\langle f, g\rangle = \int_{-\pi}^{\pi} fg$ and each is nonzero, so after dividing each by its norm we obtain an orthonormal basis of V. Each of these basis functions is an eigenvector of T with a real eigenvalue: $T(1) = 0$, $T(\cos kx) = -k^2 \cos kx$, and $T(\sin kx) = -k^2 \sin kx$. Hence the matrix of T with respect to this orthonormal basis is a real diagonal matrix, which equals its own conjugate transpose; by 7.9 this matrix is also the matrix of T^* with respect to the same orthonormal basis, and therefore $T^* = T$. (Note by contrast that D interchanges $\cos kx$ and $\sin kx$ up to scalars, so its matrix with respect to that orthonormal basis is skew-symmetric, not symmetric, in agreement with $D^* = -D$.)

Problem 7A.32. Suppose $T\colon V \to W$ is a linear map. Show that under the standard identification of V with V' (see 6.58) and the corresponding identification of W with W', the adjoint map $T^*\colon W \to V$ corresponds to the dual map $T'\colon W' \to V'$. More precisely, show that

$$T'(\varphi_w) = \varphi_{T^*w}$$

for all $w \in W$, where φ_w and φ_{T^*w} are defined as in 6.58.

Solution. As always in this chapter, V and W are finite-dimensional inner product spaces over $\mathbf{F}$; finite-dimensionality is needed both for 6.58 and for the existence of T^*.

Recall the three definitions in play.

By 6.58, for each $v \in V$ the linear functional $\varphi_v \in V'$ is defined by $\varphi_v(u) = \langle u, v\rangle$ for $u \in V$, and $v \mapsto \varphi_v$ is a bijection from V onto V'. The same construction applied to W gives, for each $w \in W$, the linear functional $\varphi_w \in W'$ defined by $\varphi_w(x) = \langle x, w\rangle$ for $x \in W$, and $w \mapsto \varphi_w$ is a bijection from W onto W'. (We use the same symbol φ for both constructions, as the exercise does; which one is meant is determined by whether the subscript lies in V or in W.)

By 3.118, the dual map $T' \in \mathcal{L}(W', V')$ is defined by $T'(\psi) = \psi \circ T$ for each $\psi \in W'$.

By 7.1, the adjoint $T^*\colon W \to V$ is the unique map satisfying

$$\langle Tv, w\rangle = \langle v, T^*w\rangle \qquad \text{for all } v \in V \text{ and all } w \in W,$$

where the inner product on the left is that of W and the one on the right is that of V.

Now fix $w \in W$. Note first that both $T'(\varphi_w)$ and φ_{T^*w} are elements of V': indeed $\varphi_w \in W'$, so $T'(\varphi_w) \in V'$; and $T^*w \in V$, so $\varphi_{T^*w} \in V'$. Two linear functionals on V are equal exactly when they agree at every $v \in V$, so let $v \in V$ be arbitrary. Then

$$\begin{aligned}
\big(T'(\varphi_w)\big)(v) &= (\varphi_w \circ T)(v) && \text{by the definition of the dual map (3.118)}\\
&= \varphi_w(Tv) && \text{by the definition of composition}\\
&= \langle Tv, w\rangle && \text{by the definition of } \varphi_w \text{ (6.58, applied in } W)\\
&= \langle v, T^*w\rangle && \text{by the definition of the adjoint (7.1)}\\
&= \varphi_{T^*w}(v) && \text{by the definition of } \varphi_{T^*w} \text{ (6.58, applied in } V).
\end{aligned}$$

Since $v \in V$ was arbitrary, the two linear functionals agree everywhere on V, so

$$T'(\varphi_w) = \varphi_{T^*w}.$$

As $w \in W$ was arbitrary, this holds for all $w \in W$, as required.

Interpretation. Write $\Phi_V\colon V \to V'$ for the map $v \mapsto \varphi_v$ and $\Phi_W\colon W \to W'$ for the map $w \mapsto \varphi_w$. By 6.58 both are bijections, and these are the standard identifications referred to in the exercise. The equation just proved says precisely that

$$T' \circ \Phi_W = \Phi_V \circ T^*,$$

that is, the diagram with $T^*\colon W \to V$ along one side, $T'\colon W' \to V'$ along the opposite side, and the identifications Φ_W, Φ_V joining them, commutes. Since Φ_V and Φ_W are bijections, this can be rewritten as

$$T^* = \Phi_V^{-1} \circ T' \circ \Phi_W,$$

which is the assertion that the adjoint map is nothing other than the dual map read through the identifications of V with V' and of W with W'. In particular, all of the properties of the adjoint proved in Chapter 7 that do not mention inner products directly can be transported to statements about dual maps, and conversely: for example, 7.6(a), which says $\operatorname{null} T^* = (\operatorname{range} T)^\perp$, becomes under this correspondence the duality statement 3.128(a) that $\operatorname{null} T' = (\operatorname{range} T)^0$, the annihilator of the range. One caution is worth recording, since it explains an apparent discrepancy between the two theories. When $\mathbf{F} = \mathbf{C}$, the identification Φ_V is not linear: as the margin note accompanying 6.58 points out, $\varphi_{\lambda v} = \overline{\lambda}\,\varphi_v$ for $\lambda \in \mathbf{C}$, so Φ_V is conjugate-linear (it is linear when $\mathbf{F} = \mathbf{R}$). This is exactly what reconciles 7.5(b), which says $(\lambda T)^* = \overline{\lambda} T^*$, with 3.120(b), which says $(\lambda T)' = \lambda T'$. Indeed, applying the identity just proved to λT and then using conjugate-linearity of Φ_V,

$$(\lambda T)' \circ \Phi_W = \Phi_V \circ (\lambda T)^* = \Phi_V \circ (\overline{\lambda} T^*) = \lambda\,(\Phi_V \circ T^*) = \lambda\,(T' \circ \Phi_W) = (\lambda T') \circ \Phi_W,$$

and since Φ_W is surjective this yields $(\lambda T)' = \lambda T'$, in agreement with 3.120(b). The two conjugations cancel. Nothing in the proof of the exercise itself is affected, because the identity $T'(\varphi_w) = \varphi_{T^* w}$ is a pointwise statement that never requires linearity of Φ_V or Φ_W.

7.2 Exercises 7B

Problem 7B.1. Prove that a normal operator on a complex inner product space is self-adjoint if and only if all its eigenvalues are real.
This exercise strengthens the analogy (for normal operators) between self-adjoint operators and real numbers.

Solution. Throughout, V is a finite-dimensional complex inner product space (the standing assumption of Chapter 7) and $T \in \mathcal{L}(V)$ is normal.
First suppose T is self-adjoint. Then every eigenvalue of T is real by 7.12 (this direction does not use normality, and indeed self-adjoint operators are automatically normal). To recall the short argument: if $Tv = \lambda v$ with $v \neq 0$, then

$$\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \|v\|^2,$$

and dividing by $\|v\|^2 \neq 0$ gives $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$.
Conversely, suppose every eigenvalue of T is real. Because $\mathbf{F} = \mathbf{C}$ and T is normal, the complex spectral theorem 7.31 provides an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_n \in \mathbf{C}$ be such that

$$Te_j = \lambda_j e_j \quad \text{for each } j = 1, \ldots, n.$$

Each e_j is nonzero, so each λ_j is an eigenvalue of T and hence $\lambda_j \in \mathbb{R}$, meaning $\overline{\lambda_j} = \lambda_j$. Because T is normal, 7.21(e) applied to the eigenvector e_j gives

$$T^* e_j = \overline{\lambda_j}\, e_j = \lambda_j e_j = Te_j$$

for each j. Thus the linear maps T^* and T agree on the basis $e_1, \ldots, e_n$ of V, and therefore $T^* = T$. Hence T is self-adjoint.

Problem 7B.2. Suppose $\mathbf{F} = \mathbf{C}$. Suppose $T \in \mathcal{L}(V)$ is normal and has only one eigenvalue. Prove that T is a scalar multiple of the identity operator.

Solution. Let $\lambda \in \mathbf{C}$ be the unique eigenvalue of T. We claim $T = \lambda I$.
Because $\mathbf{F} = \mathbf{C}$ and T is normal, the complex spectral theorem 7.31 gives an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T. For each j there is a scalar $\lambda_j \in \mathbf{C}$ with $Te_j = \lambda_j e_j$. Since $e_j \neq 0$ (it has norm 1), each λ_j is an eigenvalue of T. But T has only one eigenvalue, namely λ, so $\lambda_j = \lambda$ for every $j = 1, \dots, n$.
Hence $Te_j = \lambda e_j = (\lambda I)e_j$ for each j. The operators T and λI agree on the basis $e_1, \dots, e_n$ of V, so $T = \lambda I$, a scalar multiple of the identity operator.
Two remarks. First, the existence of λ is guaranteed by the hypothesis that T has exactly one eigenvalue; that hypothesis is consistent because, when $V \neq \{0\}$ and $\mathbf{F} = \mathbf{C}$, the operator T has at least one eigenvalue by 5.19. Second, normality is essential: the operator $S \in \mathcal{L}(\mathbf{C}^2)$ defined by $S(w, z) = (z, 0)$ satisfies $S^2 = 0$, so its only eigenvalue is 0, yet S is not a scalar multiple of the identity. Of course S is not normal.

Problem 7B.3. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Prove that the set of eigenvalues of T is contained in $\{0, 1\}$ if and only if there is a subspace U of V such that $T = P_U$.

Solution. Recall that P_U denotes the orthogonal projection of V onto the subspace U (see 6.55); because V is finite-dimensional, every subspace U of V is finite-dimensional and so P_U is defined.
First suppose there is a subspace U of V with $T = P_U$. Let λ be an eigenvalue of T and let $v \in V$ be a nonzero vector with $Tv = \lambda v$. Applying T to both sides gives

$$T^2 v = \lambda\, Tv = \lambda^2 v.$$

But $T^2 = P_U^2 = P_U = T$ by 6.57(g), so $T^2 v = Tv = \lambda v$. Hence $\lambda^2 v = \lambda v$, and since $v \neq 0$ we get $\lambda^2 = \lambda$, which forces $\lambda(\lambda - 1) = 0$ and therefore $\lambda \in \{0, 1\}$. Thus the set of eigenvalues of T is contained in $\{0, 1\}$. (Normality was not needed for this direction.)
Conversely, suppose every eigenvalue of T lies in $\{0, 1\}$. Because $\mathbf{F} = \mathbf{C}$ and T is normal, the complex spectral theorem 7.31 gives an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T, say $Te_j = \lambda_j e_j$. Each e_j is nonzero, so each λ_j is an eigenvalue of T and hence $\lambda_j \in \{0, 1\}$. After reordering the basis we may assume that there is an integer m with $0 \leq m \leq n$ such that

$$\lambda_1 = \cdots = \lambda_m = 1 \quad \text{and} \quad \lambda_{m+1} = \cdots = \lambda_n = 0.$$

Let $U = \operatorname{span}(e_1, \dots, e_m)$, interpreted as $\{0\}$ if $m = 0$. Then $e_1, \dots, e_m$ is an orthonormal basis of U. We claim $T = P_U$.
Let $v \in V$. Because $e_1, \dots, e_n$ is an orthonormal basis of V, we have

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n.$$

Applying T and using $Te_j = \lambda_j e_j$ together with the values of the λ_j gives

$$Tv = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m,$$

because the terms with index greater than m are multiplied by $\lambda_j = 0$. On the other hand, 6.57(i) applied to the orthonormal basis $e_1, \dots, e_m$ of U gives

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m.$$

Hence $Tv = P_U v$ for every $v \in V$, so $T = P_U$, as desired.

Problem 7B.4. Prove that a normal operator on a complex inner product space is skew (meaning it equals the negative of its adjoint) if and only if all its eigenvalues are purely imaginary (meaning that they have real part equal to 0).

Solution. Let V be a finite-dimensional complex inner product space and let $T \in \mathcal{L}(V)$ be normal. Recall from Exercise 12 in Section 7A that T is called skew if $T^* = -T$.
First suppose T is skew, so $T^* = -T$. Let λ be an eigenvalue of T, with $Tv = \lambda v$ for some nonzero $v \in V$. Then

$$\lambda\|v\|^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, -Tv\rangle = -\langle v, \lambda v\rangle = -\overline{\lambda}\|v\|^2.$$

Because $v \neq 0$ we may divide by $\|v\|^2$, obtaining $\lambda = -\overline{\lambda}$. Writing $\lambda = a + bi$ with $a, b \in \mathbb{R}$, this says $a + bi = -a + bi$, so $a = 0$. Thus λ is purely imaginary. (This direction did not use normality.)
Conversely, suppose every eigenvalue of T is purely imaginary. Because $\mathbf{F} = \mathbf{C}$ and T is normal, the complex spectral theorem 7.31 gives an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T, say $Te_j = \lambda_j e_j$. Each e_j is nonzero, so each λ_j is an eigenvalue of T and hence has real part 0; writing $\lambda_j = ib_j$ with $b_j \in \mathbb{R}$ we get

$$\overline{\lambda_j} = \overline{ib_j} = -ib_j = -\lambda_j.$$

Because T is normal, 7.21(e) gives $T^*e_j = \overline{\lambda_j}\, e_j$ for each j. Therefore

$$T^*e_j = \overline{\lambda_j}\, e_j = -\lambda_j e_j = -Te_j = (-T)e_j$$

for each $j = 1, \ldots, n$. The operators T^* and $-T$ thus agree on the basis $e_1, \ldots, e_n$ of V, so $T^* = -T$. Hence T is skew.

Problem 7B.5. Prove or give a counterexample: If $T \in \mathcal{L}(\mathbf{C}^3)$ is a diagonalizable operator, then T is normal (with respect to the usual inner product).

Solution. The statement is false. Here is a counterexample.
Define $T \in \mathcal{L}(\mathbf{C}^3)$ by

$$T(z_1, z_2, z_3) = (z_1 + z_2,\ 2z_2,\ 3z_3).$$

First we check that T is diagonalizable. Set

$$u_1 = (1, 0, 0), \qquad u_2 = (1, 1, 0), \qquad u_3 = (0, 0, 1).$$

Then

$$Tu_1 = (1, 0, 0) = 1 \cdot u_1, \qquad Tu_2 = (2, 2, 0) = 2u_2, \qquad Tu_3 = (0, 0, 3) = 3u_3.$$

The list u_1, u_2, u_3 is linearly independent (the matrix with these vectors as columns is upper triangular with nonzero diagonal entries $1, 1, 1$; concretely, if $au_1 + bu_2 + cu_3 = 0$ then the second coordinate gives $b = 0$, the third gives $c = 0$, and then the first gives $a = 0$). Because $\dim \mathbf{C}^3 = 3$, the list u_1, u_2, u_3 is a basis of $\mathbf{C}^3$ consisting of eigenvectors of T. Hence T is diagonalizable (see 5.55); with respect to this basis the matrix of T is the diagonal matrix with diagonal entries $1, 2, 3$.
Now we check that T is not normal, using the usual inner product on $\mathbf{C}^3$. The vectors u_1 and u_2 are eigenvectors of T corresponding to the distinct eigenvalues 1 and 2, but

$$\langle u_1, u_2\rangle = 1 \cdot \overline{1} + 0 \cdot \overline{1} + 0 \cdot \overline{0} = 1 \neq 0,$$

so u_1 and u_2 are not orthogonal. By 7.22, eigenvectors of a normal operator corresponding to distinct eigenvalues are orthogonal. Hence T is not normal.
As a direct verification: with respect to the standard basis of $\mathbf{C}^3$, which is orthonormal, the matrix of T^* is the conjugate transpose of the matrix of T, so $T^*(z_1, z_2, z_3) = (z_1,\ z_1 + 2z_2,\ 3z_3)$. Taking $v = (1, 0, 0)$ gives $\|Tv\| = \|(1, 0, 0)\| = 1$ while $\|T^*v\| = \|(1, 1, 0)\| = \sqrt{2}$. Since $\|Tv\| \neq \|T^*v\|$, the operator T is not normal by 7.20.
Thus a diagonalizable operator on $\mathbf{C}^3$ need not be normal. What the complex spectral theorem 7.31 says is the stronger statement that T is normal exactly when $\mathbf{C}^3$ has an orthonormal basis of eigenvectors of T; diagonalizability only supplies a basis of eigenvectors, with no orthogonality.

Problem 7B.6. Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$ is a normal operator such that $T^9 = T^8$. Prove that T is self-adjoint and $T^2 = T$.

Solution. Because $\mathbf{F} = \mathbf{C}$ and T is normal, the complex spectral theorem 7.31 gives an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \dots, \lambda_n \in \mathbf{C}$ satisfy $Te_j = \lambda_j e_j$ for each j. An easy induction shows $T^k e_j = \lambda_j^k e_j$ for every nonnegative integer k: it holds for $k = 0$, and if $T^k e_j = \lambda_j^k e_j$ then $T^{k+1} e_j = T(\lambda_j^k e_j) = \lambda_j^k T e_j = \lambda_j^{k+1} e_j$.
Applying the hypothesis $T^9 = T^8$ to e_j therefore gives

$$\lambda_j^9 e_j = \lambda_j^8 e_j.$$

Since $e_j \neq 0$, we conclude $\lambda_j^9 = \lambda_j^8$, that is, $\lambda_j^8(\lambda_j - 1) = 0$. Because $\mathbf{C}$ has no zero divisors, either $\lambda_j = 0$ or $\lambda_j = 1$. Hence

$$\lambda_j \in \{0, 1\} \quad \text{for each } j = 1, \dots, n.$$

T is self-adjoint. Each λ_j lies in $\{0, 1\} \subseteq \mathbb{R}$, so $\overline{\lambda_j} = \lambda_j$. Because T is normal, 7.21(e) gives $T^* e_j = \overline{\lambda_j}\, e_j = \lambda_j e_j = T e_j$ for each j. Thus T^* and T agree on the basis $e_1, \dots, e_n$, so $T^* = T$. (Alternatively: every eigenvalue of T is one of the λ_j, hence real, so Exercise 1 of this section applies.)
$T^2 = T$. Each $\lambda_j \in \{0, 1\}$ satisfies $\lambda_j^2 = \lambda_j$. Hence for each j,

$$T^2 e_j = \lambda_j^2 e_j = \lambda_j e_j = T e_j.$$

Thus T^2 and T agree on the basis $e_1, \dots, e_n$ of V, so $T^2 = T$.
Combining the two conclusions with Exercise 3 of this section, T is in fact the orthogonal projection P_U onto the subspace $U = \operatorname{span}\{e_j : \lambda_j = 1\}$, which equals $\operatorname{range} T$.

Problem 7B.7. Give an example of an operator T on a complex vector space such that $T^9 = T^8$ but $T^2 \neq T$.

Solution. Take $V = \mathbf{C}^2$ and define $T \in \mathcal{L}(\mathbf{C}^2)$ by

$$T(w, z) = (z, 0).$$

This is clearly a linear map from $\mathbf{C}^2$ to $\mathbf{C}^2$.
Then for all $(w, z) \in \mathbf{C}^2$,

$$T^2(w, z) = T(z, 0) = (0, 0),$$

so $T^2 = 0$. Consequently $T^k = T^{k-2} T^2 = 0$ for every integer $k \geq 2$; in particular

$$T^8 = 0 = T^9,$$

so $T^9 = T^8$.
However $T^2 \neq T$, because $T \neq 0$: indeed $T(0, 1) = (1, 0) \neq (0, 0)$, while $T^2(0, 1) = (0, 0)$.
This shows that the conclusion of Exercise 6 genuinely requires the normality hypothesis. Here $\mathbf{C}^2$ with its usual inner product makes T non-normal: $\|T(1, 0)\| = 0$ but $\|T^*(1, 0)\| = \|(0, 1)\| = 1$, so T is not normal by 7.20. More conceptually, T is nilpotent and nonzero, so it is not diagonalizable, whereas 7.31 forces every normal operator on a complex inner product space to be diagonalizable.

Problem 7B.8. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that T is normal if and only if every eigenvector of T is also an eigenvector of T^*.

Solution. First suppose T is normal. Let v be an eigenvector of T, say $Tv = \lambda v$ with $\lambda \in \mathbf{C}$. By 7.21(e), $T^* v = \overline{\lambda} v$. Since $v \neq 0$, this says precisely that v is an eigenvector of T^* (with eigenvalue $\overline{\lambda}$). Thus every eigenvector of T is an eigenvector of T^*.

Conversely, suppose every eigenvector of T is also an eigenvector of T^*. Because $\mathbf{F} = \mathbf{C}$, Schur's theorem (6.38) provides an orthonormal basis $e_1, \dots, e_n$ of V with respect to which T has an upper-triangular matrix. Write

$$A = \mathcal{M}\big(T, (e_1, \dots, e_n)\big), \qquad A_{j,k} = a_{j,k},$$

so $a_{j,k} = 0$ whenever $j > k$. By 7.9 the matrix of T^* with respect to this same orthonormal basis is the conjugate transpose A^*, and hence

$$T^* e_k = \sum_{j=1}^{n} \overline{a_{k,j}}\, e_j \qquad \text{for each } k,$$

because the k-th column of A^* has j-th entry $(A^*)_{j,k} = \overline{a_{k,j}}$.
We prove by induction on k that $a_{k,j} = 0$ for all $j \neq k$; that is, that row k of A has no nonzero entry off the diagonal. So let $k \in \{1, \dots, n\}$ and suppose $a_{i,j} = 0$ for all $j \neq i$ whenever $i < k$ (for $k = 1$ this hypothesis is vacuous).
Consider $Te_k = \sum_{j=1}^{n} a_{j,k} e_j$. If $j > k$ then $a_{j,k} = 0$ because A is upper triangular. If $j < k$ then $a_{j,k} = 0$ by the induction hypothesis applied to row j (note $k \neq j$). Hence

$$Te_k = a_{k,k} e_k,$$

so e_k is an eigenvector of T (it is nonzero, being a basis vector). By our assumption, e_k is therefore an eigenvector of T^*: there exists $\mu \in \mathbf{C}$ with $T^* e_k = \mu e_k$. Comparing with the displayed formula for $T^* e_k$ and using that $e_1, \dots, e_n$ is a basis, we conclude $\overline{a_{k,j}} = 0$ for every $j \neq k$, that is, $a_{k,j} = 0$ for every $j \neq k$. This completes the induction.
Thus A is a diagonal matrix: we have shown $Te_k = a_{k,k} e_k$ and $T^* e_k = \overline{a_{k,k}}\, e_k$ for every k. Hence for each k,

$$TT^* e_k = |a_{k,k}|^2 e_k = T^* T e_k.$$

The operators TT^* and T^*T therefore agree on a basis of V, so $TT^* = T^*T$. Hence T is normal.
(Equivalently: T has a diagonal matrix with respect to the orthonormal basis $e_1, \dots, e_n$, so T is normal by the complex spectral theorem 7.31.)

Problem 7B.9. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that T is normal if and only if there exists a polynomial $p \in \mathcal{P}(\mathbf{C})$ such that $T^* = p(T)$.

Solution. The easy direction first. Suppose there is a polynomial $p \in \mathcal{P}(\mathbf{C})$ with $T^* = p(T)$. Every polynomial in T commutes with T: if $p(z) = c_0 + c_1 z + \cdots + c_d z^d$, then

$$T\, p(T) = c_0 T + c_1 T^2 + \cdots + c_d T^{d+1} = p(T)\, T.$$

Hence $TT^* = T\, p(T) = p(T)\, T = T^*T$, so T is normal.
Now suppose T is normal. Since $\mathbf{F} = \mathbf{C}$, the complex spectral theorem (7.31) gives an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T; say $Te_j = \alpha_j e_j$ with $\alpha_j \in \mathbf{C}$. By 7.21(e),

$$T^* e_j = \overline{\alpha_j}\, e_j \qquad \text{for } j = 1, \dots, n.$$

Let $\lambda_1, \dots, \lambda_m$ be the distinct elements of the list $\alpha_1, \dots, \alpha_n$ (these are exactly the distinct eigenvalues of T; since V is nonzero by the standing assumptions for this chapter, we have $n \geq 1$ and hence $m \geq 1$). Define $p \in \mathcal{P}(\mathbf{C})$ by the Lagrange interpolation formula

$$p(z) = \sum_{k=1}^{m} \overline{\lambda_k} \prod_{\substack{j=1 \\ j \neq k}}^{m} \frac{z - \lambda_j}{\lambda_k - \lambda_j},$$

where an empty product is interpreted as 1 (this happens when $m = 1$, giving the constant polynomial $p(z) = \overline{\lambda_1}$). The denominators are nonzero because $\lambda_1, \dots, \lambda_m$ are distinct. For each $i \in \{1, \dots, m\}$,

every term of the sum with $k \neq i$ vanishes at $z = \lambda_i$ (its product contains the factor $z - \lambda_i$), while the $k = i$ term equals $\overline{\lambda_i}$ at $z = \lambda_i$. Hence

$$p(\lambda_i) = \overline{\lambda_i} \qquad \text{for } i = 1, \dots, m.$$

Now fix $j \in \{1, \dots, n\}$. Since $Te_j = \alpha_j e_j$, an easy induction gives $T^r e_j = \alpha_j^r e_j$ for every nonnegative integer r, and therefore $q(T)e_j = q(\alpha_j)e_j$ for every polynomial $q \in \mathcal{P}(\mathbf{C})$. In particular, since α_j is one of $\lambda_1, \dots, \lambda_m$,

$$p(T)e_j = p(\alpha_j)e_j = \overline{\alpha_j}\, e_j = T^* e_j.$$

Thus the operators $p(T)$ and T^* agree on the basis $e_1, \dots, e_n$ of V, so $p(T) = T^*$, as desired.

Problem 7B.10. Suppose V is a complex inner product space. Prove that every normal operator on V has a square root.
[An operator $S \in \mathcal{L}(V)$ is called a square root of $T \in \mathcal{L}(V)$ if $S^2 = T$. We will discuss more about square roots of operators in Sections 7C and 8C.]

Solution. As in the standing assumptions for this chapter, V is a nonzero finite-dimensional inner product space; here its scalar field is $\mathbf{C}$. Write $n = \dim V \geq 1$.
Let $T \in \mathcal{L}(V)$ be normal. By the complex spectral theorem (7.31), V has an orthonormal basis $e_1, \dots, e_n$ consisting of eigenvectors of T; say

$$Te_j = \lambda_j e_j, \qquad \lambda_j \in \mathbf{C}, \quad j = 1, \dots, n.$$

Every complex number has a complex square root. Indeed, write λ_j in polar form as $\lambda_j = r_j(\cos\theta_j + i\sin\theta_j)$ with $r_j \geq 0$ and $\theta_j \in \mathbf{R}$, and set

$$\mu_j = \sqrt{r_j}\Big(\cos\tfrac{\theta_j}{2} + i\sin\tfrac{\theta_j}{2}\Big).$$

Then by the double-angle formulas,

$$\mu_j^2 = r_j\Big(\cos^2\tfrac{\theta_j}{2} - \sin^2\tfrac{\theta_j}{2} + 2i\sin\tfrac{\theta_j}{2}\cos\tfrac{\theta_j}{2}\Big) = r_j(\cos\theta_j + i\sin\theta_j) = \lambda_j.$$

Now define $S \in \mathcal{L}(V)$ to be the unique linear map with

$$Se_j = \mu_j e_j \qquad \text{for } j = 1, \dots, n$$

(a linear map is determined by, and may be freely prescribed on, a basis; see 3.4). Then for each j,

$$S^2 e_j = S(\mu_j e_j) = \mu_j S e_j = \mu_j^2 e_j = \lambda_j e_j = Te_j.$$

Since S^2 and T agree on the basis $e_1, \dots, e_n$ of V, we conclude $S^2 = T$. Thus S is a square root of T. Note that S has a diagonal matrix with respect to the orthonormal basis $e_1, \dots, e_n$, so by 7.31 the square root S produced here is itself normal.

Problem 7B.11. Prove that every self-adjoint operator on V has a cube root.
[An operator $S \in \mathcal{L}(V)$ is called a cube root of $T \in \mathcal{L}(V)$ if $S^3 = T$.]

Solution. Here $\mathbf{F}$ may be either $\mathbf{R}$ or $\mathbf{C}$; we treat both cases at once by first producing an orthonormal basis of eigenvectors with real eigenvalues.
Let $T \in \mathcal{L}(V)$ be self-adjoint. By the standing assumptions for this chapter, V is a nonzero finite-dimensional inner product space; write $n = \dim V \geq 1$.
If $\mathbf{F} = \mathbf{R}$, then the real spectral theorem (7.29) gives an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T, and the corresponding eigenvalues lie in $\mathbf{R}$ because the scalar field is $\mathbf{R}$.
If $\mathbf{F} = \mathbf{C}$, then T is normal (indeed $TT^* = TT = T^*T$), so the complex spectral theorem (7.31) gives an

orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T; and every eigenvalue of a self-adjoint operator is real by 7.12.
So in both cases we have an orthonormal basis $e_1, \dots, e_n$ of V and real numbers $\lambda_1, \dots, \lambda_n$ with

$$Te_j = \lambda_j e_j \qquad \text{for } j = 1, \dots, n.$$

Every real number has a real cube root: the function $t \mapsto t^3$ maps $\mathbf{R}$ onto $\mathbf{R}$ (it is continuous, and $t^3 \to \pm\infty$ as $t \to \pm\infty$, so the intermediate value theorem applies). For each j choose $\mu_j \in \mathbf{R}$ with $\mu_j^3 = \lambda_j$. By the linear map lemma (3.4) there is a unique $S \in \mathcal{L}(V)$ with

$$Se_j = \mu_j e_j \qquad \text{for } j = 1, \dots, n.$$

Then for each j,

$$S^3 e_j = \mu_j^3 e_j = \lambda_j e_j = Te_j,$$

so S^3 and T agree on a basis of V and therefore $S^3 = T$. Thus S is a cube root of T.
Moreover S is itself self-adjoint: with respect to the orthonormal basis $e_1, \dots, e_n$, the matrix of S is the diagonal matrix with real diagonal entries $\mu_1, \dots, \mu_n$, and by 7.9 the matrix of S^* with respect to this basis is its conjugate transpose, which is the same matrix (the μ_j are real). Hence $S^* = S$. So every self-adjoint operator even has a self-adjoint cube root.

Problem 7B.12. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ is normal. Prove that if S is an operator on V that commutes with T, then S commutes with T^*.
[The result in this exercise is called Fuglede's theorem.]

Solution. As throughout this chapter, V is a finite-dimensional inner product space (this is needed for T^* to be defined), and here the scalar field is $\mathbf{C}$.
Quickest proof, using Exercise 9 of this section: since T is normal and $\mathbf{F} = \mathbf{C}$, there is a polynomial $p \in \mathcal{P}(\mathbf{C})$ with $T^* = p(T)$. Now $ST = TS$ implies $ST^k = T^k S$ for every nonnegative integer k (induct on k: $ST^{k+1} = (ST^k)T = T^k ST = T^k TS = T^{k+1}S$), and hence S commutes with every linear combination of powers of T, that is, with $q(T)$ for every $q \in \mathcal{P}(\mathbf{C})$. Taking $q = p$ gives $ST^* = T^* S$.
Here is a self-contained proof that does not invoke Exercise 9. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T. By the complex spectral theorem (7.31), V has a basis consisting of eigenvectors of T, so T is diagonalizable and therefore

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$$

by the equivalence of (b) and (c) in 5.55. (Here $m \geq 1$, since V is nonzero by the standing assumptions for this chapter.)
Fix k and let $v \in E(\lambda_k, T)$, so $Tv = \lambda_k v$. Two observations.
First, $E(\lambda_k, T)$ is invariant under S: indeed

$$T(Sv) = (TS)v = (ST)v = S(Tv) = S(\lambda_k v) = \lambda_k (Sv),$$

so $Sv \in E(\lambda_k, T)$.
Second, T^* acts on $E(\lambda_k, T)$ as multiplication by $\overline{\lambda_k}$: since T is normal and $Tv = \lambda_k v$, part (e) of 7.21 gives $T^* v = \overline{\lambda_k} v$ for every $v \in E(\lambda_k, T)$.
Combining these, for $v \in E(\lambda_k, T)$ we get

$$ST^* v = S(\overline{\lambda_k} v) = \overline{\lambda_k}\, Sv \qquad \text{and} \qquad T^* Sv = \overline{\lambda_k}\, Sv,$$

the second equality because $Sv \in E(\lambda_k, T)$ and T^* is multiplication by $\overline{\lambda_k}$ there. Hence $ST^* v = T^* Sv$ for every $v \in E(\lambda_k, T)$ and every k.
Since V is the sum of the subspaces $E(\lambda_1, T), \dots, E(\lambda_m, T)$, every $v \in V$ can be written $v = v_1 + \cdots + v_m$ with $v_k \in E(\lambda_k, T)$; by linearity

$$ST^* v = \sum_{k=1}^{m} ST^* v_k = \sum_{k=1}^{m} T^* S v_k = T^* Sv.$$

Thus $ST^* = T^*S$, so S commutes with T^*.

Problem 7B.13. Without using the complex spectral theorem, use the version of Schur's theorem that applies to two commuting operators (take $\mathcal{E} = \{T, T^*\}$ in Exercise 20 in Section 6B) to give a different proof that if $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal, then T has a diagonal matrix with respect to some orthonormal basis of V.

Solution. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Set

$$\mathcal{E} = \{T, T^*\} \subseteq \mathcal{L}(V).$$

We first check that any two elements of $\mathcal{E}$ commute. Every operator commutes with itself, so $TT = TT$ and $T^*T^* = T^*T^*$; and $TT^* = T^*T$ precisely because T is normal. Hence $AB = BA$ for all $A, B \in \mathcal{E}$. Therefore Exercise 20 in Section 6B applies: there is an orthonormal basis $e_1, \dots, e_n$ of V with respect to which every element of $\mathcal{E}$ has an upper-triangular matrix. In particular both

$$A = \mathcal{M}\big(T, (e_1, \dots, e_n)\big) \qquad \text{and} \qquad B = \mathcal{M}\big(T^*, (e_1, \dots, e_n)\big)$$

are upper-triangular matrices.
Because $e_1, \dots, e_n$ is an orthonormal basis, 7.9 tells us that B is the conjugate transpose of A:

$$B = A^*, \qquad \text{that is,} \qquad B_{j,k} = \overline{A_{k,j}} \quad \text{for all } j, k.$$

Now fix indices j and k with $j \neq k$, and consider the two cases.
If $j > k$, then $A_{j,k} = 0$ because A is upper triangular.
If $j < k$, look instead at the entry $B_{k,j}$ of B. Its row index k exceeds its column index j, so $B_{k,j} = 0$ because B is upper triangular. Since $B_{k,j} = \overline{A_{j,k}}$, this gives $A_{j,k} = 0$.
Hence $A_{j,k} = 0$ for all $j \neq k$, that is, A is a diagonal matrix.
Thus T has a diagonal matrix with respect to the orthonormal basis $e_1, \dots, e_n$ of V, which is the conclusion of the complex spectral theorem, obtained here without using it.
For completeness, here is the argument for the case of Exercise 20 in Section 6B that is used above. The claim proved by induction on n is the following statement, which is exactly the instance $\mathcal{E} = \{T, T^*\}$ of that exercise:
If W is a complex inner product space with $\dim W = n$ and $R \in \mathcal{L}(W)$ is normal, then W has an orthonormal basis with respect to which both R and R^* have upper-triangular matrices.
(Stating the claim for an arbitrary normal operator on an arbitrary space, rather than for the fixed T on the fixed V, is what makes the induction step below legitimate.) So let R be a normal operator on a complex inner product space W with $\dim W = n$, and rename R as T and W as V for readability. If $n = 0$ there is nothing to prove. Suppose $n \geq 1$ and the claim holds for all complex inner product spaces of dimension $n - 1$. Since $\mathbf{F} = \mathbf{C}$, the operator T^* has an eigenvalue μ (by 5.19), and $E(\mu, T^*) \neq \{0\}$. This eigenspace is invariant under T: if $T^*u = \mu u$, then $T^*(Tu) = T(T^*u) = \mu Tu$, using $TT^* = T^*T$. Restricting T to the nonzero invariant subspace $E(\mu, T^*)$ and again using 5.19, we obtain a common eigenvector u of T and T^*; normalize so that $\|u\| = 1$. Let $U = (\operatorname{span}(u))^\perp$, a subspace of dimension $n-1$. For $w \in U$ we have $\langle Tw, u\rangle = \langle w, T^*u\rangle = 0$ (as T^*u is a multiple of u) and $\langle T^*w, u\rangle = \langle w, Tu\rangle = 0$, so U is invariant under both T and T^*. Now U, with the inner product inherited from V, is a complex inner product space of dimension $n - 1$, and $T|_U$ is an operator on it. Moreover the adjoint of $T|_U$ is $(T^*)|_U$: for all $w, x \in U$ we have $\langle (T|_U)w, x\rangle = \langle Tw, x\rangle = \langle w, T^*x\rangle = \langle w, (T^*)|_U x\rangle$, and $(T^*)|_U$ does map U into U because U is invariant under T^*; since the adjoint is unique, $(T|_U)^* = (T^*)|_U$. Consequently $T|_U$ is normal, because

$$(T|_U)(T|_U)^* = (TT^*)|_U = (T^*T)|_U = (T|_U)^*(T|_U).$$

So the induction hypothesis applies to the normal operator $T|_U$ on the $(n-1)$-dimensional space U, giving an orthonormal basis $e_2, \dots, e_n$ of U with respect to which both $T|_U$ and $(T|_U)^* = (T^*)|_U$ have upper-triangular matrices. Then $u, e_2, \dots, e_n$ is an orthonormal basis of V with respect to which both T and T^* have upper-triangular matrices, since Tu and T^*u are multiples of u and, for $k \geq 2$, Te_k and T^*e_k lie in $\operatorname{span}(e_2, \dots, e_k)$.

Problem 7B.14. Suppose $\mathbf{F} = \mathbf{R}$ and $T \in \mathcal{L}(V)$. Prove that T is self-adjoint if and only if all pairs of eigenvectors corresponding to distinct eigenvalues of T are orthogonal and

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.

Solution. Throughout, $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T; since $\mathbf{F} = \mathbf{R}$, these are real numbers. By the standing assumptions for this chapter, V is a nonzero finite-dimensional inner product space. Note that on either side of the equivalence the list $\lambda_1, \ldots, \lambda_m$ is nonempty: if T is self-adjoint then T has an eigenvalue by the real spectral theorem 7.29 (a diagonal matrix with respect to a basis of the nonzero space V exhibits one), and if $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$ with $V \neq \{0\}$ then $m \geq 1$.
First suppose T is self-adjoint.
Orthogonality of eigenvectors for distinct eigenvalues: suppose $Tu = \alpha u$ and $Tv = \beta v$ with $\alpha \neq \beta$ and u, v nonzero. Since $T^* = T$ and all scalars are real (so the inner product is symmetric and bilinear),

$$\alpha\langle u, v\rangle = \langle \alpha u, v\rangle = \langle Tu, v\rangle = \langle u, T^*v\rangle = \langle u, Tv\rangle = \langle u, \beta v\rangle = \beta\langle u, v\rangle.$$

Hence $(\alpha - \beta)\langle u, v\rangle = 0$, and since $\alpha \neq \beta$ we get $\langle u, v\rangle = 0$. (This is exactly 7.22, which applies because every self-adjoint operator is normal.)
The direct sum decomposition: by the real spectral theorem (7.29), V has an orthonormal basis consisting of eigenvectors of T. In particular V has a basis consisting of eigenvectors of T, so T is diagonalizable, and therefore

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$$

by the equivalence of (b) and (c) in 5.55.
Conversely, suppose that eigenvectors of T corresponding to distinct eigenvalues are orthogonal and that $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.
Each $E(\lambda_k, T)$ is a subspace of the inner product space V, hence is itself a finite-dimensional inner product space, so by 6.35 it has an orthonormal basis; call it

$$e_1^{(k)}, \ldots, e_{d_k}^{(k)}, \qquad d_k = \dim E(\lambda_k, T).$$

Concatenate these lists to form the list

$$e_1^{(1)}, \ldots, e_{d_1}^{(1)},\ e_1^{(2)}, \ldots, e_{d_2}^{(2)},\ \ldots,\ e_1^{(m)}, \ldots, e_{d_m}^{(m)}.$$

This list is orthonormal. Each vector has norm 1. Two distinct vectors from the same block are orthogonal, since each block is an orthonormal list. And if $e_i^{(k)}$ and $e_j^{(l)}$ come from different blocks, that is $k \neq l$, then $e_i^{(k)}$ is a nonzero vector of $E(\lambda_k, T)$ and hence an eigenvector of T with eigenvalue λ_k, while $e_j^{(l)}$ is an eigenvector of T with eigenvalue λ_l; as $\lambda_k \neq \lambda_l$, our hypothesis gives $\langle e_i^{(k)}, e_j^{(l)}\rangle = 0$.
This list spans V. Indeed, given $v \in V$, the hypothesis $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$ lets us write $v = v_1 + \cdots + v_m$ with $v_k \in E(\lambda_k, T)$, and each v_k is a linear combination of $e_1^{(k)}, \ldots, e_{d_k}^{(k)}$.
Hence the list is an orthonormal spanning list. Every orthonormal list is linearly independent (6.25), so the list is a linearly independent spanning list, that is, an orthonormal basis of V.
Finally, every vector in the list is an eigenvector of T: $Te_i^{(k)} = \lambda_k e_i^{(k)}$.
Thus V has an orthonormal basis consisting of eigenvectors of T, so condition (c) of the real spectral theorem 7.29 holds, and therefore T is self-adjoint.
(One can also finish directly, without quoting 7.29: with respect to the orthonormal basis just constructed, the matrix of T is the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_m$, all of which are real. By 7.9 the matrix of T^* with respect to the same orthonormal basis is the conjugate transpose of that diagonal real matrix, which is the matrix itself. Since T and T^* have the same matrix with respect to a basis of V, they agree on that basis and hence $T^* = T$.)

Problem 7B.15. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that T is normal if and only if all pairs of eigenvectors corresponding to distinct eigenvalues of T are orthogonal and

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.

Solution. Throughout, V is a nonzero finite-dimensional complex inner product space (the standing convention of Chapter 7) and $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T; there is at least one such eigenvalue by 5.19, so $m \geq 1$.

Suppose first that T is normal. Then 7.22 says immediately that eigenvectors of T corresponding to distinct eigenvalues are orthogonal, which is the first assertion.

For the second assertion, the complex spectral theorem 7.31 (implication (a) $\Rightarrow$ (c)) gives an orthonormal basis of V consisting of eigenvectors of T. In particular T is diagonalizable, and hence 5.55 gives

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$

Now suppose conversely that eigenvectors of T corresponding to distinct eigenvalues are orthogonal and that $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$. For each $j \in \{1, \ldots, m\}$, the eigenspace $E(\lambda_j, T)$ is a subspace of the inner product space V, so applying the Gram–Schmidt procedure 6.32 to any basis of $E(\lambda_j, T)$ produces an orthonormal basis

$$e_1^{(j)}, \ldots, e_{d_j}^{(j)}$$

of $E(\lambda_j, T)$, where $d_j = \dim E(\lambda_j, T)$.

Consider the list obtained by concatenating these m lists:

$$e_1^{(1)}, \ldots, e_{d_1}^{(1)},\ e_1^{(2)}, \ldots, e_{d_2}^{(2)},\ \ldots,\ e_1^{(m)}, \ldots, e_{d_m}^{(m)}.$$

Because V is the direct sum of the subspaces $E(\lambda_j, T)$, concatenating bases of the summands produces a basis of V; thus the list above is a basis of V.

This basis is orthonormal. Indeed, each vector in it has norm 1. Two distinct vectors from the same block, say $e_r^{(j)}$ and $e_s^{(j)}$ with $r \neq s$, are orthogonal because $e_1^{(j)}, \ldots, e_{d_j}^{(j)}$ is an orthonormal list. Two vectors from different blocks, say $e_r^{(i)}$ and $e_s^{(j)}$ with $i \neq j$, are nonzero vectors lying in $E(\lambda_i, T)$ and $E(\lambda_j, T)$ respectively; hence they are eigenvectors of T corresponding to the distinct eigenvalues λ_i and λ_j, and so they are orthogonal by hypothesis.

Finally, every vector in this basis is an eigenvector of T (each $e_r^{(j)}$ is a nonzero element of $E(\lambda_j, T)$). Thus V has an orthonormal basis consisting of eigenvectors of T, and the complex spectral theorem 7.31 (implication (c) $\Rightarrow$ (a)) shows that T is normal.

Problem 7B.16. Suppose $\mathbf{F} = \mathbf{C}$ and $\mathcal{E} \subseteq \mathcal{L}(V)$. Prove that there is an orthonormal basis of V with respect to which every element of $\mathcal{E}$ has a diagonal matrix if and only if S and T are commuting normal operators for all $S, T \in \mathcal{E}$.

[This exercise extends the complex spectral theorem to the context of a collection of commuting normal operators.]

Solution. Suppose first that there is an orthonormal basis $e_1, \ldots, e_n$ of V with respect to which every element of $\mathcal{E}$ has a diagonal matrix. If $T \in \mathcal{E}$, then T has a diagonal matrix with respect to an orthonormal basis, so T is normal by the complex spectral theorem 7.31 (implication (b) $\Rightarrow$ (a)). If $S, T \in \mathcal{E}$, then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$ and $\mathcal{M}(TS) = \mathcal{M}(T)\mathcal{M}(S)$ with respect to $e_1, \ldots, e_n$ (by 3.43); since $\mathcal{M}(S)$ and $\mathcal{M}(T)$ are diagonal matrices, they commute (the product of two diagonal matrices is the diagonal matrix whose k-th diagonal entry is the product of the k-th diagonal entries, in either order). Hence $\mathcal{M}(ST) = \mathcal{M}(TS)$; since an operator is determined by its matrix with respect to a fixed basis (the matrix records the images of the basis vectors), $ST = TS$. Thus all elements of $\mathcal{E}$ are normal

and pairwise commuting.
For the converse, suppose every element of $\mathcal{E}$ is normal and $ST = TS$ for all $S, T \in \mathcal{E}$. We prove by induction on $n = \dim V$ that V has an orthonormal basis with respect to which every element of $\mathcal{E}$ has a diagonal matrix. (If $\mathcal{E}$ is empty the assertion is trivial, so assume $\mathcal{E} \neq \emptyset$.)
If $n = 0$ the empty basis works, and if $n = 1$ then every operator on V is a scalar multiple of the identity and every 1-by-1 matrix is diagonal, so any vector of norm 1 gives the required basis. Now let $n \geq 1$ and suppose the assertion holds for all complex inner product spaces of dimension less than n (and all collections of commuting normal operators on them).
Case 1: every $T \in \mathcal{E}$ is a scalar multiple of the identity operator. Then with respect to any orthonormal basis of V (one exists by 6.35) each $T \in \mathcal{E}$ has matrix a scalar multiple of the identity matrix, which is diagonal. Done.
Case 2: some $T \in \mathcal{E}$ is not a scalar multiple of the identity operator. Because T is normal and $\mathbf{F} = \mathbf{C}$, the complex spectral theorem 7.31 gives an orthonormal basis of V consisting of eigenvectors of T; in particular T is diagonalizable, so by 5.55

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T. If $m = 1$ then $V = E(\lambda_1, T)$, which would mean $T = \lambda_1 I$, contrary to the assumption of this case. Hence $m \geq 2$, and therefore

$$\dim E(\lambda_j, T) < \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) = n$$

for each j (each eigenspace has dimension at least 1).
Each eigenspace is invariant under every element of $\mathcal{E}$: if $S \in \mathcal{E}$ and $v \in E(\lambda_j, T)$, then

$$T(Sv) = (TS)v = (ST)v = S(Tv) = S(\lambda_j v) = \lambda_j (Sv),$$

so $Sv \in E(\lambda_j, T)$.
Fix j and let $U_j = E(\lambda_j, T)$, a complex inner product space with the inner product inherited from V. Set

$$\mathcal{E}_j = \{\, S|_{U_j} : S \in \mathcal{E} \,\} \subseteq \mathcal{L}(U_j),$$

which makes sense by the invariance just proved. Each element of $\mathcal{E}_j$ is normal: S is normal and U_j is invariant under S, so $S|_{U_j}$ is normal by Exercise 20 in this section (7B.20(d)), whose proof is independent of the present exercise. The elements of $\mathcal{E}_j$ commute pairwise: for $S, S' \in \mathcal{E}$ and $u \in U_j$ we have $S'u \in U_j$ and $Su \in U_j$, so

$$(S|_{U_j})(S'|_{U_j})u = S(S'u) = (SS')u = (S'S)u = S'(Su) = (S'|_{U_j})(S|_{U_j})u.$$

Since $\dim U_j < n$, the induction hypothesis applied to U_j and $\mathcal{E}_j$ produces an orthonormal basis

$$e_1^{(j)}, \ldots, e_{d_j}^{(j)}$$

of U_j with respect to which every element of $\mathcal{E}_j$ has a diagonal matrix; equivalently, each $e_r^{(j)}$ is an eigenvector of $S|_{U_j}$, and hence of S, for every $S \in \mathcal{E}$.
Concatenate these lists over $j = 1, \ldots, m$. Because $V = U_1 \oplus \cdots \oplus U_m$, the concatenation is a basis of V. It is orthonormal: within one block this holds by construction, and a vector from block i and a vector from block j with $i \neq j$ are eigenvectors of the normal operator T corresponding to the distinct eigenvalues λ_i and λ_j, hence orthogonal by 7.22.
Finally, every vector of this concatenated basis is an eigenvector of every $S \in \mathcal{E}$. An operator whose every basis vector is an eigenvector has a diagonal matrix with respect to that basis. Hence every element of $\mathcal{E}$ has a diagonal matrix with respect to this orthonormal basis of V, completing the induction and the proof.

Problem 7B.17. Suppose $\mathbf{F} = \mathbf{R}$ and $\mathcal{E} \subseteq \mathcal{L}(V)$. Prove that there is an orthonormal basis of V with respect to which every element of $\mathcal{E}$ has a diagonal matrix if and only if S and T are commuting self-adjoint operators for all $S, T \in \mathcal{E}$.
[This exercise extends the real spectral theorem to the context of a collection of commuting self-adjoint operators.]

Solution. Suppose first that there is an orthonormal basis $e_1, \dots, e_n$ of V with respect to which every element of $\mathcal{E}$ has a diagonal matrix. If $T \in \mathcal{E}$, then T is self-adjoint by the real spectral theorem 7.29 (implication (b) $\Rightarrow$ (a)). If $S, T \in \mathcal{E}$, then with respect to $e_1, \dots, e_n$ we have $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T) = \mathcal{M}(T)\mathcal{M}(S) = \mathcal{M}(TS)$, using 3.43 and the fact that diagonal matrices commute. Since an operator is determined by its matrix with respect to a fixed basis, $ST = TS$. Thus all elements of $\mathcal{E}$ are self-adjoint and pairwise commuting.
For the converse, suppose every element of $\mathcal{E}$ is self-adjoint and $ST = TS$ for all $S, T \in \mathcal{E}$. We induct on $n = \dim V$. (If $\mathcal{E} = \emptyset$ any orthonormal basis works, so assume $\mathcal{E} \neq \emptyset$.)
If $n = 0$ the empty basis works; if $n = 1$ every operator on V is a scalar multiple of the identity and every 1-by-1 matrix is diagonal, so any vector of norm 1 works. Now let $n \geq 1$ and suppose the assertion holds for all real inner product spaces of dimension less than n.
Case 1: every $T \in \mathcal{E}$ is a scalar multiple of the identity operator. Then with respect to any orthonormal basis of V (one exists by 6.35) every $T \in \mathcal{E}$ has a diagonal matrix, and we are done.
Case 2: some $T \in \mathcal{E}$ is not a scalar multiple of the identity operator. Because T is self-adjoint and $\mathbf{F} = \mathbf{R}$, the real spectral theorem 7.29 gives an orthonormal basis of V consisting of eigenvectors of T; thus T is diagonalizable and 5.55 gives

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T. If $m = 1$ we would have $V = E(\lambda_1, T)$ and hence $T = \lambda_1 I$, contrary to the assumption of this case. So $m \geq 2$, and since each eigenspace has dimension at least 1 and the dimensions add up to n, we get $\dim E(\lambda_j, T) < n$ for every j.
Every eigenspace of T is invariant under every $S \in \mathcal{E}$: if $v \in E(\lambda_j, T)$ then

$$T(Sv) = (TS)v = (ST)v = S(Tv) = \lambda_j(Sv),$$

so $Sv \in E(\lambda_j, T)$.
Fix j, put $U_j = E(\lambda_j, T)$ with the inner product inherited from V, and set

$$\mathcal{E}_j = \{\, S|_{U_j} : S \in \mathcal{E} \,\} \subseteq \mathcal{L}(U_j).$$

Each $S|_{U_j}$ is self-adjoint: S is self-adjoint and U_j is invariant under S, so Exercise 19(b) in this section (7B.19(b)) applies. Alternatively, directly: for $u, w \in U_j$ we have $\langle S|_{U_j} u, w\rangle = \langle Su, w\rangle = \langle u, Sw\rangle = \langle u, S|_{U_j} w\rangle$. The elements of $\mathcal{E}_j$ commute pairwise, since for $S, S' \in \mathcal{E}$ and $u \in U_j$ (using that U_j is invariant under both)

$$(S|_{U_j})(S'|_{U_j})u = (SS')u = (S'S)u = (S'|_{U_j})(S|_{U_j})u.$$

Because $\dim U_j < n$, the induction hypothesis gives an orthonormal basis $e_1^{(j)}, \dots, e_{d_j}^{(j)}$ of U_j with respect to which every element of $\mathcal{E}_j$ has a diagonal matrix; that is, each $e_r^{(j)}$ is an eigenvector of $S|_{U_j}$, and hence of S, for every $S \in \mathcal{E}$.
Concatenating these lists over $j = 1, \dots, m$ gives a basis of V, because $V = U_1 \oplus \cdots \oplus U_m$. It is orthonormal: within a block this holds by construction, while a vector from block i and a vector from block j with $i \neq j$ are eigenvectors of T corresponding to the distinct eigenvalues λ_i and λ_j, hence orthogonal by 7.22 (every self-adjoint operator is normal, since $T^*T = T^2 = TT^*$).
Every vector of the resulting orthonormal basis is an eigenvector of every $S \in \mathcal{E}$, so every element of $\mathcal{E}$ has a diagonal matrix with respect to it. This completes the induction and the proof.

Problem 7B.18. Give an example of a real inner product space V, an operator $T \in \mathcal{L}(V)$, and real numbers b, c with $b^2 < 4c$ such that

$$T^2 + bT + cI$$

is not invertible.
[This exercise shows that the hypothesis that T is self-adjoint cannot be deleted in 7.26, even for real vector spaces.]

Solution. Take $V = \mathbf{R}^2$ with the usual (Euclidean) inner product

$$\langle (x_1, x_2), (y_1, y_2) \rangle = x_1 y_1 + x_2 y_2,$$

and define $T \in \mathcal{L}(\mathbf{R}^2)$ by

$$T(x, y) = (-y, x)$$

(counterclockwise rotation by 90°). Take $b = 0$ and $c = 1$, so that

$$b^2 = 0 < 4 = 4c.$$

Then for every $(x, y) \in \mathbf{R}^2$,

$$T^2(x, y) = T(-y, x) = (-x, -y) = -(x, y),$$

so $T^2 = -I$. Therefore

$$T^2 + bT + cI = T^2 + I = -I + I = 0.$$

The zero operator on $\mathbf{R}^2$ is not invertible (it is not injective: it sends the nonzero vector $(1, 0)$ to 0; alternatively, $\dim \mathbf{R}^2 = 2 > 0$ so $0 \neq I$ and the zero operator has no inverse). Hence $T^2 + bT + cI$ is not invertible, as desired.
This is consistent with 7.26 because T is not self-adjoint. Indeed,

$$\langle T(1, 0), (0, 1) \rangle = \langle (0, 1), (0, 1) \rangle = 1,$$

whereas

$$\langle (1, 0), T(0, 1) \rangle = \langle (1, 0), (-1, 0) \rangle = -1.$$

Since these differ, $T^* \neq T$. (In fact $T^* = -T$, so T is a skew operator; it is normal but not self-adjoint. Thus the example also shows that normality is not enough in 7.26.)
More generally, whenever $b, c \in \mathbf{R}$ satisfy $b^2 < 4c$ one can produce such an example: the polynomial $z^2 + bz + c$ has the two nonreal conjugate roots $\alpha = \frac{-b + i\sqrt{4c - b^2}}{2}$ and $\bar{\alpha}$, and the operator T on $\mathbf{R}^2$ with matrix

$$\begin{pmatrix} -b/2 & -\sqrt{4c - b^2}/2 \\ \sqrt{4c - b^2}/2 & -b/2 \end{pmatrix}$$

with respect to the standard basis satisfies $T^2 + bT + cI = 0$. The displayed case is $b = 0$, $c = 1$.

Problem 7B.19. Suppose $T \in \mathcal{L}(V)$ is self-adjoint and U is a subspace of V that is invariant under T.
(a) Prove that $U^\perp$ is invariant under T.
(b) Prove that $T|_U \in \mathcal{L}(U)$ is self-adjoint.
(c) Prove that $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint.

Solution. Throughout, U and $U^\perp$ carry the inner product inherited from V. Recall that an operator R on an inner product space W is self-adjoint exactly when $\langle Rw, w' \rangle = \langle w, Rw' \rangle$ for all $w, w' \in W$; this is the defining property of the adjoint (7.1) together with the uniqueness of the adjoint.

(a) Let $v \in U^\perp$. To show $Tv \in U^\perp$, take an arbitrary $u \in U$. Then

$$\langle Tv, u\rangle = \langle v, T^*u\rangle = \langle v, Tu\rangle = 0,$$

where the first equality is the definition of the adjoint, the second uses $T^* = T$, and the third holds because $Tu \in U$ (as U is invariant under T) while $v \in U^\perp$. Since $\langle Tv, u\rangle = 0$ for all $u \in U$, we conclude $Tv \in U^\perp$. Hence $U^\perp$ is invariant under T.
(b) Because U is invariant under T, the restriction $T|_U$ maps U into U, and it is clearly linear, so $T|_U \in \mathcal{L}(U)$. For all $u, w \in U$,

$$\langle (T|_U)u, w\rangle = \langle Tu, w\rangle = \langle u, T^*w\rangle = \langle u, Tw\rangle = \langle u, (T|_U)w\rangle,$$

where the inner products are those of V, which restrict to the inner product of U. Thus $(T|_U)^* = T|_U$, so $T|_U$ is self-adjoint.
(c) By part (a), $U^\perp$ is a subspace of V invariant under T. Since T is self-adjoint, part (b) applied with U replaced by the invariant subspace $U^\perp$ shows that $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint. Explicitly, for all $v, w \in U^\perp$,

$$\langle (T|_{U^\perp})v, w\rangle = \langle Tv, w\rangle = \langle v, Tw\rangle = \langle v, (T|_{U^\perp})w\rangle.$$

Problem 7B.20. Suppose $T \in \mathcal{L}(V)$ is normal and U is a subspace of V that is invariant under T.
(a) Prove that $U^\perp$ is invariant under T.
(b) Prove that U is invariant under T^*.
(c) Prove that $(T|_U)^* = (T^*)|_U$.
(d) Prove that $T|_U \in \mathcal{L}(U)$ and $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ are normal operators.
[This exercise can be used to give yet another proof of the complex spectral theorem (use induction on $\dim V$ and the result that T has an eigenvector).]

Solution. Let $m = \dim U$ and $n = \dim V$. Choose an orthonormal basis $e_1, \dots, e_m$ of U and extend it to an orthonormal basis $e_1, \dots, e_n$ of V (possible by 6.36). Then $e_{m+1}, \dots, e_n$ is an orthonormal list in $U^\perp$ (each of these vectors is orthogonal to $e_1, \dots, e_m$ and hence to $U = \operatorname{span}(e_1, \dots, e_m)$) of length $n - m = \dim U^\perp$ (by 6.51), so $e_{m+1}, \dots, e_n$ is an orthonormal basis of $U^\perp$.
Let $A = \mathcal{M}(T, (e_1, \dots, e_n))$, so that $Te_k = \sum_{j=1}^n A_{j,k} e_j$ for each k. Because the basis is orthonormal,

$$\|Te_k\|^2 = \sum_{j=1}^n |A_{j,k}|^2.$$

Because U is invariant under T, for $k \le m$ we have $Te_k \in U = \operatorname{span}(e_1, \dots, e_m)$, so $A_{j,k} = 0$ whenever $j > m$ and $k \le m$. In block form,

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix},$$

where B is m-by-m, D is $(n-m)$-by-$(n-m)$, and C is m-by-$(n-m)$. Note that $B = \mathcal{M}(T|_U, (e_1, \dots, e_m))$.
The key point is that $C = 0$. To see this, recall that because $e_1, \dots, e_n$ is orthonormal, $\mathcal{M}(T^*, (e_1, \dots, e_n)) = A^*$, the conjugate transpose of A (by 7.9). Hence for each k,

$$\|T^*e_k\|^2 = \sum_{j=1}^n |(A^*)_{j,k}|^2 = \sum_{j=1}^n |A_{k,j}|^2,$$

the sum of the squared absolute values of the entries in row k of A. Now sum over $k = 1, \dots, m$:

$$\sum_{k=1}^m \|Te_k\|^2 = \sum_{k=1}^m \sum_{j=1}^m |A_{j,k}|^2 = \sum_{j,k \le m} |B_{j,k}|^2,$$

using that the entries below B in the first m columns vanish, while

$$\sum_{k=1}^{m} \|T^* e_k\|^2 = \sum_{k=1}^{m} \sum_{j=1}^{n} |A_{k,j}|^2 = \sum_{j,k \le m} |B_{k,j}|^2 + \sum_{k \le m,\, j>m} |C_{k,j-m}|^2.$$

Because T is normal, $\|Te_k\| = \|T^* e_k\|$ for every k (by 7.20). Comparing the two displayed sums (which have the same first term, since $\sum_{j,k\le m} |B_{j,k}|^2 = \sum_{j,k\le m} |B_{k,j}|^2$) yields

$$\sum_{k \le m,\, j>m} |C_{k,j-m}|^2 = 0,$$

so every entry of C is 0. Thus

$$A = \begin{pmatrix} B & 0 \\ 0 & D \end{pmatrix}, \qquad A^* = \begin{pmatrix} B^* & 0 \\ 0 & D^* \end{pmatrix}.$$

(a) For $k > m$, column k of A has zero entries in rows $1, \dots, m$ (these are the entries of C), so $Te_k \in \operatorname{span}(e_{m+1}, \dots, e_n) = U^\perp$. Since $e_{m+1}, \dots, e_n$ spans $U^\perp$ and T is linear, $Tv \in U^\perp$ for every $v \in U^\perp$. Hence $U^\perp$ is invariant under T.
(b) For $k \le m$, column k of A^* has zero entries in rows $m+1, \dots, n$, so $T^* e_k \in \operatorname{span}(e_1, \dots, e_m) = U$. Since $e_1, \dots, e_m$ spans U, linearity gives $T^* u \in U$ for every $u \in U$. Hence U is invariant under T^*.
(c) By (b), $(T^*)|_U$ is an operator on U; by hypothesis $T|_U$ is an operator on U. For all $u, w \in U$,

$$\langle (T|_U)u, w\rangle = \langle Tu, w\rangle = \langle u, T^* w\rangle = \langle u, (T^*)|_U w\rangle,$$

where the inner product of V restricts to that of U. Because the adjoint of an operator on U is the unique operator on U satisfying this identity for all $u, w \in U$ (7.1), we conclude that $(T|_U)^* = (T^*)|_U$.
(d) First consider $T|_U$. For $u \in U$ we have $(T^*)|_U u = T^* u \in U$ by (b) and $(T|_U)u = Tu \in U$, so using (c),

$$(T|_U)(T|_U)^* u = T(T^* u) = (TT^*)u, \qquad (T|_U)^*(T|_U)u = T^*(Tu) = (T^*T)u.$$

Since T is normal, $TT^* = T^*T$, so the two right-hand sides agree for every $u \in U$. Hence

$$(T|_U)(T|_U)^* = (T|_U)^*(T|_U),$$

so $T|_U$ is normal.
Now consider $T|_{U^\perp}$. By part (a), $U^\perp$ is a subspace of V invariant under the normal operator T. Everything proved above applies to any T-invariant subspace, so applying the previous paragraph with U replaced by $U^\perp$ shows that $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is normal.
(Equivalently, in matrix terms: with respect to the orthonormal bases above, $\mathcal{M}(T|_U) = B$ and $\mathcal{M}(T|_{U^\perp}) = D$, and the normality equation $A^*A = AA^*$ becomes, blockwise, $B^*B = BB^*$ and $D^*D = DD^*$.)

Problem 7B.21. Suppose that T is a self-adjoint operator on a finite-dimensional inner product space and that 2 and 3 are the only eigenvalues of T. Prove that

$$T^2 - 5T + 6I = 0.$$

Solution. Let V be the finite-dimensional inner product space on which T acts, over $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$. Note $V \neq \{0\}$, since T has eigenvalues.
The first step is to produce a basis of V consisting of eigenvectors of T.
If $\mathbf{F} = \mathbf{R}$, then the real spectral theorem 7.29 (implication (a) $\Rightarrow$ (c)) gives an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T.
If $\mathbf{F} = \mathbf{C}$, note first that T is normal, because

$$T^*T = TT = TT^*,$$

using $T^* = T$. Hence the complex spectral theorem 7.31 (implication (a) $\Rightarrow$ (c)) gives an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T.
So in either case there are a basis $e_1, \dots, e_n$ of V and scalars $\lambda_1, \dots, \lambda_n$ with $Te_k = \lambda_k e_k$ for each k. Each e_k is nonzero, so each λ_k is an eigenvalue of T; by hypothesis the only eigenvalues of T are 2 and 3, so $\lambda_k \in \{2, 3\}$ for every k.
Now compute, for each k,

$$\begin{aligned}(T^2 - 5T + 6I)e_k &= T^2 e_k - 5Te_k + 6e_k \\ &= \lambda_k^2 e_k - 5\lambda_k e_k + 6e_k \\ &= (\lambda_k^2 - 5\lambda_k + 6)\, e_k \\ &= (\lambda_k - 2)(\lambda_k - 3)\, e_k \\ &= 0,\end{aligned}$$

where the last equality holds because λ_k equals 2 or 3, so one of the two factors vanishes.
Thus the operator $T^2 - 5T + 6I$ sends every vector of the basis $e_1, \dots, e_n$ to 0. A linear map that vanishes on a basis is the zero map: if $v \in V$, write $v = a_1 e_1 + \cdots + a_n e_n$ and then

$$(T^2 - 5T + 6I)v = \sum_{k=1}^{n} a_k (T^2 - 5T + 6I)e_k = 0.$$

Hence $T^2 - 5T + 6I = 0$, as desired.
An alternative argument uses minimal polynomials. As shown above, T is diagonalizable, so by 5.62 the minimal polynomial of T equals $(z-\mu_1)\cdots(z-\mu_M)$ for some distinct $\mu_1, \dots, \mu_M \in \mathbf{F}$. The zeros of the minimal polynomial of T are precisely the eigenvalues of T (by 5.27(a)), so $\{\mu_1, \dots, \mu_M\} = \{2, 3\}$ and hence the minimal polynomial of T is $(z-2)(z-3) = z^2 - 5z + 6$. Applying the minimal polynomial to T gives $T^2 - 5T + 6I = 0$.
Note that self-adjointness cannot be dropped. For the operator T on $\mathbf{C}^3$ whose matrix with respect to the standard basis is

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix},$$

the matrix is upper triangular, so by 5.41 the eigenvalues of T are exactly the diagonal entries, namely 2 and 3; but $(T-2I)(T-3I)$ sends the first standard basis vector to 0 and the second to $(-1, 0, 0) \neq 0$, so $T^2 - 5T + 6I \neq 0$. This answers Exercise 22 of this section.

Problem 7B.22. Give an example of an operator $T \in \mathcal{L}(\mathbf{C}^3)$ such that 2 and 3 are the only eigenvalues of T and $T^2 - 5T + 6I \neq 0$.

Solution. Define $T \in \mathcal{L}(\mathbf{C}^3)$ by

$$T(z_1, z_2, z_3) = (2z_1 + z_2,\ 2z_2,\ 3z_3).$$

With respect to the standard basis of $\mathbf{C}^3$, the matrix of T is

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix},$$

which is upper triangular with diagonal entries $2, 2, 3$. Hence by 5.41 the eigenvalues of T are exactly the numbers appearing on that diagonal, namely 2 and 3. So 2 and 3 are the only eigenvalues of T.
Because $z^2 - 5z + 6 = (z-2)(z-3)$, we have

$$T^2 - 5T + 6I = (T - 2I)(T - 3I).$$

Now $(T-2I)(z_1, z_2, z_3) = (z_2, 0, z_3)$ and $(T-3I)(z_1, z_2, z_3) = (-z_1 + z_2,\ -z_2,\ 0)$. Therefore

$$(T-2I)(T-3I)(z_1, z_2, z_3) = (T-2I)(-z_1 + z_2,\ -z_2,\ 0) = (-z_2,\ 0,\ 0).$$

Taking $(z_1, z_2, z_3) = (0, 1, 0)$ gives $(T^2 - 5T + 6I)(0, 1, 0) = (-1, 0, 0) \neq 0$. Thus $T^2 - 5T + 6I \neq 0$, as desired.
Note that this operator is not diagonalizable: if it were, then $\mathbf{C}^3$ would have a basis of eigenvectors of T (by 5.55), and since each such basis vector is annihilated by $T - 2I$ or by $T - 3I$ and these two operators commute, we would get $(T - 2I)(T - 3I) = 0$, contradicting the computation above. Hence by the complex spectral theorem (7.31) T is not normal; in particular it is not self-adjoint. This is consistent with Exercise 21, which shows that a self-adjoint operator whose only eigenvalues are 2 and 3 must satisfy $T^2 - 5T + 6I = 0$. Thus the self-adjointness hypothesis in Exercise 21 cannot be deleted.

Problem 7B.23. Suppose $T \in \mathcal{L}(V)$ is self-adjoint, $\lambda \in \mathbf{F}$, and $\varepsilon > 0$. Suppose there exists $v \in V$ such that $\|v\| = 1$ and

$$\|Tv - \lambda v\| < \varepsilon.$$

Prove that T has an eigenvalue λ' such that $|\lambda - \lambda'| < \varepsilon$.
[This exercise shows that for a self-adjoint operator, a number that is close to satisfying an equation that would make it an eigenvalue is close to an eigenvalue.]

Solution. Because T is self-adjoint, the spectral theorem applies: if $\mathbf{F} = \mathbf{R}$ this is the real spectral theorem (7.29), and if $\mathbf{F} = \mathbf{C}$ then T is normal (every self-adjoint operator commutes with its adjoint, which equals itself) and so the complex spectral theorem (7.31) applies. In either case there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \dots, \lambda_n \in \mathbf{F}$ be the corresponding eigenvalues, so

$$Te_k = \lambda_k e_k \qquad \text{for each } k \in \{1, \dots, n\}.$$

The hypothesis provides a vector $v \in V$ with $\|v\| = 1$, so $V \neq \{0\}$ and hence $n = \dim V \geq 1$; in particular the list $\lambda_1, \dots, \lambda_n$ is nonempty, so T does have eigenvalues. In fact $\lambda_1, \dots, \lambda_n$ is a list of all the eigenvalues of T (with repetitions), since T has a diagonal matrix with respect to $e_1, \dots, e_n$ and 5.41 identifies the eigenvalues of an operator with the diagonal entries of any upper-triangular matrix representing it.
Choose $k_0 \in \{1, \dots, n\}$ so that $|\lambda - \lambda_{k_0}|$ is as small as possible among $|\lambda - \lambda_1|, \dots, |\lambda - \lambda_n|$, and set

$$\lambda' = \lambda_{k_0}, \qquad d = |\lambda - \lambda'| = \min\{|\lambda - \lambda_k| : k \in \{1, \dots, n\}\}.$$

Then λ' is an eigenvalue of T. We must show $d < \varepsilon$.
Write v in terms of the orthonormal basis: by 6.30(a),

$$v = a_1 e_1 + \cdots + a_n e_n, \qquad \text{where } a_k = \langle v, e_k \rangle,$$

and by 6.30(b),

$$|a_1|^2 + \cdots + |a_n|^2 = \|v\|^2 = 1.$$

Applying $T - \lambda I$ gives

$$Tv - \lambda v = \sum_{k=1}^{n} a_k (T - \lambda I) e_k = \sum_{k=1}^{n} (\lambda_k - \lambda) a_k e_k.$$

The vectors $(\lambda_1 - \lambda)a_1 e_1, \dots, (\lambda_n - \lambda)a_n e_n$ are pairwise orthogonal, so the Pythagorean theorem (6.12), or equivalently 6.30(b) applied to $Tv - \lambda v$, yields

$$\|Tv - \lambda v\|^2 = \sum_{k=1}^{n} |\lambda_k - \lambda|^2 |a_k|^2 \geq d^2 \sum_{k=1}^{n} |a_k|^2 = d^2,$$

where the inequality holds because $|\lambda_k - \lambda| \geq d$ for every k, by the choice of d as the minimum.
Hence

$$|\lambda - \lambda'| = d \leq \|Tv - \lambda v\| < \varepsilon.$$

Thus λ' is an eigenvalue of T with $|\lambda - \lambda'| < \varepsilon$, as desired.

Self-adjointness is essential. For a concrete failure without it, let N be a large positive number and let $T \in \mathcal{L}(\mathbf{C}^2)$ have matrix $\begin{pmatrix} 0 & N \\ 0 & 0 \end{pmatrix}$ with respect to the standard basis, so that 0 is the only eigenvalue of T. Take $\lambda = 1$ and

$$v = \frac{1}{\sqrt{N^2+1}}\,(N, 1),$$

which has norm 1. Then $Tv = \frac{1}{\sqrt{N^2+1}}(N, 0)$, so

$$\|Tv - \lambda v\| = \frac{1}{\sqrt{N^2+1}}\,\|(0, -1)\| = \frac{1}{\sqrt{N^2+1}},$$

which is less than $\varepsilon = \frac{1}{2}$ once $N \geq 2$. Yet the only eigenvalue λ' of T is 0, and $|\lambda - \lambda'| = 1 > \varepsilon$.

Problem 7B.24. Suppose U is a finite-dimensional vector space and $T \in \mathcal{L}(U)$.
(a) Suppose $\mathbf{F} = \mathbf{R}$. Prove that T is diagonalizable if and only if there is a basis of U such that the matrix of T with respect to this basis equals its transpose.
(b) Suppose $\mathbf{F} = \mathbf{C}$. Prove that T is diagonalizable if and only if there is a basis of U such that the matrix of T with respect to this basis commutes with its conjugate transpose.
[This exercise adds another equivalence to the list of conditions equivalent to diagonalizability in 5.55.]

Solution. Here U is merely a vector space; no inner product on U is given in advance. The point of the exercise is that we may manufacture one.
If $U = \{0\}$ then $T = 0$ is diagonalizable and the only basis of U is the empty list, whose 0-by-0 matrix equals its own transpose and commutes with its conjugate transpose; both equivalences hold vacuously. So assume from now on that $U \neq \{0\}$, and let $n = \dim U \geq 1$.
Throughout, for an n-by-n matrix A with entries in $\mathbf{F}$ we write A^{t} for the transpose of A and A^* for the conjugate transpose of A (so $A^* = A^{\mathrm{t}}$ when $\mathbf{F} = \mathbf{R}$).
Construction of an inner product. Suppose $u_1, \ldots, u_n$ is a basis of U. Define

$$\Big\langle \sum_{j=1}^{n} a_j u_j,\ \sum_{j=1}^{n} b_j u_j \Big\rangle = \sum_{j=1}^{n} a_j \overline{b_j}$$

for all $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbf{F}$. This is well defined because each element of U has a unique representation as a linear combination of $u_1, \ldots, u_n$. It is exactly the pullback of the Euclidean inner product on $\mathbf{F}^n$ under the isomorphism $U \to \mathbf{F}^n$ sending $\sum a_j u_j$ to $(a_1, \ldots, a_n)$; since that map is a bijective linear map and the Euclidean inner product satisfies the axioms in 6.2, so does the form just defined. (Positivity: $\langle u, u\rangle = \sum |a_j|^2 \geq 0$, with equality only if every $a_j = 0$, that is, only if $u = 0$.) With respect to this inner product, $u_1, \ldots, u_n$ is by construction an orthonormal basis of U.
The key consequence is 7.9: if $u_1, \ldots, u_n$ is an orthonormal basis of an inner product space U and $S \in \mathcal{L}(U)$, then

$$\mathcal{M}\big(S^*, (u_1, \ldots, u_n)\big) = \Big(\mathcal{M}\big(S, (u_1, \ldots, u_n)\big)\Big)^*.$$

(a) Suppose $\mathbf{F} = \mathbf{R}$.
First suppose T is diagonalizable. By 5.55 there is a basis $u_1, \ldots, u_n$ of U consisting of eigenvectors of T, and with respect to that basis $\mathcal{M}(T)$ is a diagonal matrix. A diagonal matrix equals its own transpose. So the matrix of T with respect to this basis equals its transpose.
Conversely, suppose $u_1, \ldots, u_n$ is a basis of U such that $A = \mathcal{M}(T, (u_1, \ldots, u_n))$ satisfies $A = A^{\mathrm{t}}$. Equip U with the inner product constructed above, making $u_1, \ldots, u_n$ an orthonormal basis. By 7.9 and because $\mathbf{F} = \mathbf{R}$,

$$\mathcal{M}(T^*, (u_1, \ldots, u_n)) = A^* = A^{\mathrm{t}} = A = \mathcal{M}(T, (u_1, \ldots, u_n)).$$

Two operators with the same matrix with respect to the same basis are equal, so $T^* = T$; that is, T is self-adjoint on this inner product space. By the real spectral theorem (7.29), U has an orthonormal

basis consisting of eigenvectors of T. In particular U has a basis consisting of eigenvectors of T, so T is diagonalizable by 5.55.
(b) Suppose $\mathbf{F} = \mathbf{C}$.
First suppose T is diagonalizable. By 5.55 there is a basis $u_1, \dots, u_n$ of U consisting of eigenvectors of T, so $D = \mathcal{M}(T, (u_1, \dots, u_n))$ is a diagonal matrix, say with diagonal entries $\lambda_1, \dots, \lambda_n$. Then D^* is the diagonal matrix with diagonal entries $\overline{\lambda_1}, \dots, \overline{\lambda_n}$, and both DD^* and D^*D equal the diagonal matrix with diagonal entries $|\lambda_1|^2, \dots, |\lambda_n|^2$. Hence D commutes with its conjugate transpose.
Conversely, suppose $u_1, \dots, u_n$ is a basis of U such that $A = \mathcal{M}(T, (u_1, \dots, u_n))$ satisfies $AA^* = A^*A$. Equip U with the inner product constructed above, so that $u_1, \dots, u_n$ is an orthonormal basis. By 7.9, $\mathcal{M}(T^*) = A^*$ with respect to this basis. Using 3.43 (the matrix of a product of linear maps is the product of the matrices), we get

$$\mathcal{M}(TT^*) = \mathcal{M}(T)\,\mathcal{M}(T^*) = AA^* \quad \text{and} \quad \mathcal{M}(T^*T) = \mathcal{M}(T^*)\,\mathcal{M}(T) = A^*A,$$

all matrices being taken with respect to $u_1, \dots, u_n$. Since $AA^* = A^*A$, the operators TT^* and T^*T have the same matrix with respect to this basis and hence are equal. Thus T is normal. By the complex spectral theorem (7.31), U has an orthonormal basis consisting of eigenvectors of T; in particular U has a basis of eigenvectors of T, so T is diagonalizable by 5.55.
[Remark. In each direction the inner product is a tool that we are free to choose after seeing the basis, which is why a condition on the matrix in a single basis suffices. Note also that the forward implications produce a basis whose matrix is diagonal, which is stronger than what is required.]

Problem 7B.25. Suppose that $T \in \mathcal{L}(V)$ and there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T, with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Show that if $k \in \{1, \dots, n\}$, then the pseudoinverse $T^\dagger$ satisfies the equation

$$T^\dagger e_k = \begin{cases} \dfrac{1}{\lambda_k} e_k & \text{if } \lambda_k \neq 0, \\ 0 & \text{if } \lambda_k = 0. \end{cases}$$

Solution. Recall the definition of the pseudoinverse (6.68): $T^\dagger w = \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1} P_{\operatorname{range} T}\, w$ for $w \in V$, where $T|_{(\operatorname{null} T)^\perp}$ is an invertible linear map from $(\operatorname{null} T)^\perp$ onto $\operatorname{range} T$ by 6.67. In particular, if $w \in (\operatorname{range} T)^\perp$ then $T^\dagger w = 0$, and if $w \in \operatorname{range} T$ then $T^\dagger w$ is the unique vector in $(\operatorname{null} T)^\perp$ whose image under T is w.
Partition the index set. Let

$$N = \operatorname{span}\{e_k : \lambda_k = 0\}, \qquad R = \operatorname{span}\{e_k : \lambda_k \neq 0\},$$

with the convention that the span of the empty list is $\{0\}$.
Step 1: $\operatorname{null} T = N$ and $\operatorname{range} T = R$.
Every $v \in V$ can be written uniquely as $v = a_1 e_1 + \cdots + a_n e_n$ with $a_1, \dots, a_n \in \mathbf{F}$, and then

$$Tv = a_1 \lambda_1 e_1 + \cdots + a_n \lambda_n e_n.$$

Because $e_1, \dots, e_n$ is a basis, $Tv = 0$ if and only if $a_k \lambda_k = 0$ for every k, that is, if and only if $a_k = 0$ for every k with $\lambda_k \neq 0$. The set of such v is precisely N. Hence $\operatorname{null} T = N$.
For the range: $\operatorname{range} T$ is spanned by $Te_1, \dots, Te_n$, that is, by $\lambda_1 e_1, \dots, \lambda_n e_n$. Deleting the zero vectors from this list (those with $\lambda_k = 0$) and rescaling the survivors by the nonzero scalars $1/\lambda_k$ does not change the span, so $\operatorname{range} T = \operatorname{span}\{e_k : \lambda_k \neq 0\} = R$.
Step 2: $R = N^\perp$ and $N = R^\perp$; equivalently, $\operatorname{range} T = (\operatorname{null} T)^\perp$.
Each e_j with $\lambda_j \neq 0$ is orthogonal to each e_k with $\lambda_k = 0$, because $e_1, \dots, e_n$ is orthonormal and $j \neq k$. Hence every vector in R is orthogonal to every vector in N, giving $R \subseteq N^\perp$. The lists $\{e_k : \lambda_k = 0\}$ and $\{e_k : \lambda_k \neq 0\}$ are linearly independent (being sublists of a basis), so $\dim N + \dim R = n = \dim V$. By 6.51, $\dim N^\perp = \dim V - \dim N = \dim R$. A subspace contained in a subspace of the same dimension equals it (2.39), so $R = N^\perp$. Applying 6.52 then gives $N = (N^\perp)^\perp = R^\perp$.

Combining with Step 1: $\operatorname{range} T = (\operatorname{null} T)^\perp$ and $\operatorname{null} T = (\operatorname{range} T)^\perp$.
Step 3: computation of $T^\dagger e_k$.
Fix $k \in \{1, \dots, n\}$.
Case 1: $\lambda_k = 0$. Then $e_k \in N = (\operatorname{range} T)^\perp$ by Step 2. Hence $P_{\operatorname{range} T}\, e_k = 0$, and therefore

$$T^\dagger e_k = \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1} 0 = 0.$$

Case 2: $\lambda_k \neq 0$. Then $e_k \in R = \operatorname{range} T$ by Step 1, so $P_{\operatorname{range} T}\, e_k = e_k$ and

$$T^\dagger e_k = \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1} e_k.$$

Now $\frac{1}{\lambda_k} e_k \in R = (\operatorname{null} T)^\perp$ by Step 2, and

$$T\Big(\tfrac{1}{\lambda_k} e_k\Big) = \tfrac{1}{\lambda_k}\, \lambda_k e_k = e_k.$$

Thus $\frac{1}{\lambda_k} e_k$ is a vector in $(\operatorname{null} T)^\perp$ that T maps to e_k, and since $T|_{(\operatorname{null} T)^\perp}$ is injective it is the only such vector. Hence

$$T^\dagger e_k = \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1} e_k = \frac{1}{\lambda_k} e_k.$$

This proves the asserted formula in both cases.
[Remark. Consequently $T^\dagger$ is itself diagonal with respect to the same orthonormal basis $e_1, \dots, e_n$: its eigenvalues are $1/\lambda_k$ for the nonzero λ_k and 0 for the zero ones. In particular, by the spectral theorems 7.29 and 7.31 this applies whenever T is self-adjoint (for $\mathbf{F} = \mathbf{R}$) or normal (for $\mathbf{F} = \mathbf{C}$), and it shows that the pseudoinverse inverts T exactly on $\operatorname{range} T$ while killing $(\operatorname{range} T)^\perp$.]

7.3 Exercises 7C

Problem 7C.1. Suppose $T \in \mathcal{L}(V)$. Prove that if both T and $-T$ are positive operators, then $T = 0$.

Solution. Suppose T and $-T$ are both positive operators. Let $v \in V$. Because T is positive,

$$\langle Tv, v\rangle \geq 0.$$

Because $-T$ is positive,

$$0 \leq \langle (-T)v, v\rangle = \langle -(Tv), v\rangle = -\langle Tv, v\rangle,$$

using the homogeneity of the inner product in the first slot. Hence $\langle Tv, v\rangle \leq 0$. Combining the two inequalities gives

$$\langle Tv, v\rangle = 0.$$

Since T is a positive operator and $\langle Tv, v\rangle = 0$, result 7.43 implies that $Tv = 0$.
As $v \in V$ was arbitrary, we conclude that $Tv = 0$ for every $v \in V$, which means $T = 0$. (Instead of 7.43 one may quote 7.16: a positive operator is self-adjoint by definition, and a self-adjoint operator with $\langle Tv, v\rangle = 0$ for all $v \in V$ equals 0.)
Alternative argument avoiding 7.43: a positive operator is by definition self-adjoint, so T is self-adjoint, and by 7.38 every eigenvalue of T is nonnegative. Likewise every eigenvalue of $-T$ is nonnegative; since the eigenvalues of $-T$ are exactly the negatives of the eigenvalues of T, every eigenvalue of T is also nonpositive. Hence 0 is the only possible eigenvalue of T. Because T is self-adjoint, the spectral theorem (7.29 and 7.31) provides an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T; each corresponding eigenvalue equals 0, so $Te_k = 0$ for each k. An operator that vanishes on a basis is the zero operator, so $T = 0$.

Problem 7C.2. Suppose $T \in \mathcal{L}(\mathbf{F}^4)$ is the operator whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}.$$

Show that T is an invertible positive operator.

Solution. Call the displayed matrix A, so $A = \mathcal{M}(T)$ with respect to the standard basis e_1, e_2, e_3, e_4 of $\mathbf{F}^4$, which is an orthonormal basis with respect to the Euclidean inner product.
Step 1: T is self-adjoint. All entries of A are real and A equals its transpose, so A equals its conjugate transpose: $A = A^*$. By 7.9, the matrix of T^* with respect to an orthonormal basis is the conjugate transpose of the matrix of T; hence $\mathcal{M}(T^*) = A^* = A = \mathcal{M}(T)$, and therefore $T^* = T$ (definition 7.10).
Step 2: computation of $\langle Tv, v\rangle$. Let $v = (v_1, v_2, v_3, v_4) \in \mathbf{F}^4$. Reading off the rows of A,

$$Tv = (2v_1 - v_2,\ -v_1 + 2v_2 - v_3,\ -v_2 + 2v_3 - v_4,\ -v_3 + 2v_4).$$

Therefore

$$\begin{aligned} \langle Tv, v\rangle &= (2v_1 - v_2)\overline{v_1} + (-v_1 + 2v_2 - v_3)\overline{v_2} + (-v_2 + 2v_3 - v_4)\overline{v_3} + (-v_3 + 2v_4)\overline{v_4} \\ &= 2\big(|v_1|^2 + |v_2|^2 + |v_3|^2 + |v_4|^2\big) - \sum_{k=1}^{3}\big(v_{k+1}\overline{v_k} + v_k\overline{v_{k+1}}\big) \\ &= 2\sum_{k=1}^{4}|v_k|^2 - 2\sum_{k=1}^{3}\operatorname{Re}\big(v_k\overline{v_{k+1}}\big). \end{aligned}$$

(The last equality uses $z + \overline{z} = 2\operatorname{Re} z$ with $z = v_k\overline{v_{k+1}}$; when $\mathbf{F} = \mathbf{R}$ all conjugations and real parts are vacuous.)
Now observe the telescoping identity

$$|v_1|^2 + \sum_{k=1}^{3}|v_k - v_{k+1}|^2 + |v_4|^2 = 2\sum_{k=1}^{4}|v_k|^2 - 2\sum_{k=1}^{3}\operatorname{Re}\big(v_k\overline{v_{k+1}}\big),$$

which holds because $|v_k - v_{k+1}|^2 = |v_k|^2 + |v_{k+1}|^2 - 2\operatorname{Re}(v_k\overline{v_{k+1}})$, and because in the sum $|v_1|^2 + \big(|v_1|^2 + |v_2|^2\big) + \big(|v_2|^2 + |v_3|^2\big) + \big(|v_3|^2 + |v_4|^2\big) + |v_4|^2$ each $|v_k|^2$ occurs exactly twice. Hence

$$\langle Tv, v\rangle = |v_1|^2 + |v_1 - v_2|^2 + |v_2 - v_3|^2 + |v_3 - v_4|^2 + |v_4|^2.$$

Step 3: T is positive. The right side above is a sum of squares of absolute values, so $\langle Tv, v\rangle \geq 0$ for every $v \in \mathbf{F}^4$. Together with Step 1, this shows that T is a positive operator.
Step 4: T is invertible. Suppose $v \in \mathbf{F}^4$ and $Tv = 0$. Then $\langle Tv, v\rangle = 0$, so by the formula in Step 2 every term vanishes:

$$v_1 = 0, \quad v_1 = v_2, \quad v_2 = v_3, \quad v_3 = v_4, \quad v_4 = 0.$$

Hence $v_1 = v_2 = v_3 = v_4 = 0$, so $v = 0$. Thus T is injective, and therefore invertible because $\mathbf{F}^4$ is finite-dimensional (by 3.65, an injective operator on a finite-dimensional vector space is invertible).

Problem 7C.3. Suppose n is a positive integer and $T \in \mathcal{L}(\mathbf{F}^n)$ is the operator whose matrix (with respect to the standard basis) consists of all 1's. Show that T is a positive operator.

Solution. Let $A = \mathcal{M}(T)$ with respect to the standard basis $e_1, \ldots, e_n$ of $\mathbf{F}^n$, so $A_{j,k} = 1$ for all j, k. The standard basis is orthonormal with respect to the Euclidean inner product.

Step 1: T is self-adjoint. Every entry of A is the real number 1, so the conjugate transpose A^* also has all entries equal to 1; thus $A^* = A$. By 7.9, the matrix of T^* with respect to an orthonormal basis is the conjugate transpose of the matrix of T, so $\mathcal{M}(T^*) = A^* = A = \mathcal{M}(T)$, which gives $T^* = T$; thus T is self-adjoint (definition 7.10).
Step 2: $\langle Tv, v\rangle \geq 0$. Let $v = (v_1, \ldots, v_n) \in \mathbf{F}^n$ and set

$$s = v_1 + \cdots + v_n.$$

Every row of A consists of 1's, so every coordinate of Tv equals s; that is, $Tv = (s, s, \ldots, s)$. Hence

$$\langle Tv, v\rangle = \sum_{j=1}^{n} s\,\overline{v_j} = s\sum_{j=1}^{n}\overline{v_j} = s\,\overline{s} = |s|^2 \geq 0.$$

Because T is self-adjoint and $\langle Tv, v\rangle \geq 0$ for every $v \in \mathbf{F}^n$, the operator T is positive.
Remark. Writing $u = (1, \ldots, 1)$, the computation above says $Tv = \langle v, u\rangle u$ for all v, so $T = nP_U$ where $U = \operatorname{span}(u)$ and P_U is the orthogonal projection onto U (note $\|u\|^2 = n$, so $P_U v = \frac{\langle v,u\rangle}{n}u$). Thus T is a nonnegative multiple of an orthogonal projection, and orthogonal projections are positive operators by 7.35(b). Alternatively, one can verify $T = R^*R$, where $R \in \mathcal{L}(\mathbf{F}^n)$ is defined by $Rv = (s, 0, \ldots, 0)$ with $s = v_1 + \cdots + v_n$ as above; indeed $\langle R^*Rv, w\rangle = \langle Rv, Rw\rangle = s\,\overline{t}$ where $t = w_1 + \cdots + w_n$, and $\langle Tv, w\rangle = \sum_{j=1}^{n} s\,\overline{w_j} = s\,\overline{t}$ as well. Then apply 7.38, condition (f).

Problem 7C.4. Suppose n is an integer with $n > 1$. Show that there exists an n-by-n matrix A such that all of the entries of A are positive numbers and $A = A^*$, but the operator on $\mathbf{F}^n$ whose matrix (with respect to the standard basis) equals A is not a positive operator.

Solution. Define an n-by-n matrix A by

$$A_{j,k} = \begin{cases} 2 & \text{if } \{j,k\} = \{1,2\}, \\ 1 & \text{otherwise,} \end{cases}$$

so that A has 2 in the two positions $(1,2)$ and $(2,1)$ and 1 everywhere else. For example, when $n = 2$,

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

This makes sense because $n > 1$, so the entries in positions $(1,2)$ and $(2,1)$ exist.
Every entry of A is either 1 or 2, hence is a positive number. Also $A_{j,k} = A_{k,j}$ for all j, k by the symmetry of the defining condition $\{j,k\} = \{1,2\}$, and all entries are real; hence the conjugate transpose of A equals A, so $A = A^*$.
Let $T \in \mathcal{L}(\mathbf{F}^n)$ be the operator whose matrix with respect to the standard basis $e_1, \ldots, e_n$ equals A. Take

$$v = e_1 - e_2 = (1, -1, 0, \ldots, 0).$$

The first coordinate of Tv is $A_{1,1}\cdot 1 + A_{1,2}\cdot(-1) = 1 - 2 = -1$, the second coordinate is $A_{2,1}\cdot 1 + A_{2,2}\cdot(-1) = 2 - 1 = 1$, and for $j \geq 3$ the j-th coordinate is $A_{j,1}\cdot 1 + A_{j,2}\cdot(-1) = 1 - 1 = 0$. Thus $Tv = (-1, 1, 0, \ldots, 0)$, and therefore

$$\langle Tv, v\rangle = (-1)\cdot\overline{1} + 1\cdot\overline{(-1)} = -1 - 1 = -2 < 0.$$

Hence T is not a positive operator, even though $A = A^*$ (so T is self-adjoint: the standard basis is orthonormal, so by 7.9 we have $\mathcal{M}(T^*) = A^* = A = \mathcal{M}(T)$) and every entry of A is a positive number.
Remark. This shows that positivity of an operator has nothing to do with positivity of the entries of its matrix. Indeed, by 7.38 the operator T above has a negative eigenvalue: for $n = 2$ the eigenvalues of A are 3 and -1.

Problem 7C.5. Suppose $T \in \mathcal{L}(V)$ is self-adjoint. Prove that T is a positive operator if and only if for every orthonormal basis $e_1, \dots, e_n$ of V, all entries on the diagonal of $\mathcal{M}(T, (e_1, \dots, e_n))$ are nonnegative numbers.

Solution. First we record what the diagonal entries are. Let $e_1, \dots, e_n$ be an orthonormal basis of V and let $A = \mathcal{M}(T, (e_1, \dots, e_n))$. By 6.30, every vector of V is the sum of its orthonormal-basis expansion; applying this to Te_k gives

$$Te_k = \langle Te_k, e_1\rangle e_1 + \cdots + \langle Te_k, e_n\rangle e_n.$$

Since the k-th column of A consists of the coefficients of Te_k in this basis, we get $A_{j,k} = \langle Te_k, e_j\rangle$ for all j, k. In particular, the diagonal entries are

$$A_{k,k} = \langle Te_k, e_k\rangle \qquad \text{for } k = 1, \dots, n.$$

Now suppose T is a positive operator. Let $e_1, \dots, e_n$ be any orthonormal basis of V. Then for each k we have $A_{k,k} = \langle Te_k, e_k\rangle \geq 0$ by the definition 7.34 of a positive operator. Thus all diagonal entries of $\mathcal{M}(T, (e_1, \dots, e_n))$ are nonnegative, proving one direction. (Note that $\langle Te_k, e_k\rangle$ is a real number: this is automatic when $\mathbf{F} = \mathbf{R}$, and when $\mathbf{F} = \mathbf{C}$ it follows from 7.14 because T is self-adjoint. So calling these entries nonnegative makes sense.)

Conversely, suppose that for every orthonormal basis $e_1, \dots, e_n$ of V, all diagonal entries of $\mathcal{M}(T, (e_1, \dots, e_n))$ are nonnegative. Because T is self-adjoint, the spectral theorem (7.29 and 7.31) provides an orthonormal basis $f_1, \dots, f_n$ of V consisting of eigenvectors of T; let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues, so $Tf_k = \lambda_k f_k$. With respect to this particular orthonormal basis, the matrix of T is the diagonal matrix with $\lambda_1, \dots, \lambda_n$ on the diagonal (as the computation above confirms: $\langle Tf_k, f_k\rangle = \lambda_k\langle f_k, f_k\rangle = \lambda_k$). By hypothesis applied to the orthonormal basis $f_1, \dots, f_n$, each λ_k is nonnegative.

Hence, with respect to the orthonormal basis $f_1, \dots, f_n$ of V, the matrix of T is a diagonal matrix with only nonnegative numbers on the diagonal. That is exactly condition (c) of 7.38, so by the equivalence of the conditions in 7.38 the operator T is a positive operator, completing the proof of the converse.

Remark. The proof of the converse used the hypothesis for only one cleverly chosen orthonormal basis, namely a basis of eigenvectors. Also, the assumption that T is self-adjoint cannot be dropped when $\mathbf{F} = \mathbf{R}$: the operator on $\mathbf{R}^2$ with matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has $\langle Tv, v\rangle = 0$ for all v, hence has nonnegative (indeed zero) diagonal entries with respect to every orthonormal basis, but it is not self-adjoint and so is not positive.

Problem 7C.6. Prove that the sum of two positive operators on V is a positive operator.

Solution. Suppose $S, T \in \mathcal{L}(V)$ are positive operators. We must show that $S + T$ is self-adjoint and that $\langle (S + T)v, v\rangle \geq 0$ for every $v \in V$.

Self-adjointness: by definition 7.34, $S^* = S$ and $T^* = T$. By 7.5(a),

$$(S + T)^* = S^* + T^* = S + T,$$

so $S + T$ is self-adjoint.

Nonnegativity: let $v \in V$. Using additivity of the inner product in its first slot,

$$\langle (S + T)v, v\rangle = \langle Sv + Tv, v\rangle = \langle Sv, v\rangle + \langle Tv, v\rangle \geq 0,$$

because $\langle Sv, v\rangle \geq 0$ and $\langle Tv, v\rangle \geq 0$ (both S and T are positive) and the sum of two nonnegative real numbers is nonnegative.

Hence $S + T$ is a positive operator.

Remark. The same argument shows that $aS+bT$ is positive whenever a, b are nonnegative real numbers, using 7.5(b) for self-adjointness. Thus the positive operators on V form a cone in $\mathcal{L}(V)$: closed under addition and under multiplication by nonnegative scalars, but (by Exercise 7C.1) containing no nonzero operator together with its negative.

Problem 7C.7. Suppose $S \in \mathcal{L}(V)$ is an invertible positive operator and $T \in \mathcal{L}(V)$ is a positive operator. Prove that $S + T$ is invertible.

Solution. Because V is finite-dimensional, it suffices to prove that $S + T$ is injective; then $S + T$ is invertible by 3.65 (for an operator on a finite-dimensional vector space, injectivity, surjectivity, and invertibility are equivalent).
So suppose $v \in V$ and $(S + T)v = 0$. Taking the inner product with v gives

$$0 = \langle (S + T)v, v\rangle = \langle Sv, v\rangle + \langle Tv, v\rangle.$$

Since S and T are positive operators, both $\langle Sv, v\rangle$ and $\langle Tv, v\rangle$ are nonnegative real numbers. Two nonnegative numbers summing to 0 must each equal 0, so

$$\langle Sv, v\rangle = 0 \quad \text{and} \quad \langle Tv, v\rangle = 0.$$

Now S is positive and $\langle Sv, v\rangle = 0$, so 7.43 gives $Sv = 0$. Because S is invertible, it is injective, and therefore $v = 0$.
Thus $\operatorname{null}(S + T) = \{0\}$, so $S + T$ is injective and hence invertible.
Remark. The hypothesis that S is invertible cannot be weakened to $S \neq 0$: taking $V = \mathbf{F}^2$ with S the orthogonal projection onto the first coordinate axis and $T = 0$ gives positive operators whose sum is not invertible. Also note that by Exercise 7C.6 the operator $S + T$ is positive, so this exercise says that a positive operator plus an invertible positive operator is an invertible positive operator.

Problem 7C.8. Suppose $T \in \mathcal{L}(V)$. Prove that T is a positive operator if and only if the pseudoinverse $T^\dagger$ is a positive operator.

Solution. Throughout, V is finite-dimensional (the standing assumption of this chapter), so the pseudoinverse $T^\dagger \in \mathcal{L}(V)$ is defined by 6.68:

$$T^\dagger w \;=\; \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1} P_{\operatorname{range} T}\, w \qquad \text{for } w \in V,$$

where $T|_{(\operatorname{null} T)^\perp}$ is regarded as an invertible linear map from $(\operatorname{null} T)^\perp$ onto $\operatorname{range} T$ (this is legitimate by 6.67).
The proof has three steps: first we show that the pseudoinverse is an involution, then we prove one direction, and then the other direction follows formally.
Step 1: $(T^\dagger)^\dagger = T$.
First we compute the null space of $T^\dagger$. If $w \in (\operatorname{range} T)^\perp$, then $P_{\operatorname{range} T} w = 0$ and hence $T^\dagger w = 0$. Conversely, if $T^\dagger w = 0$, then because $(T|_{(\operatorname{null} T)^\perp})^{-1}$ is injective we get $P_{\operatorname{range} T} w = 0$; since $w - P_{\operatorname{range} T} w \in (\operatorname{range} T)^\perp$ (by 6.55 and the definition of orthogonal projection), this gives $w \in (\operatorname{range} T)^\perp$. Thus

$$\operatorname{null} T^\dagger = (\operatorname{range} T)^\perp, \qquad \text{and hence} \qquad (\operatorname{null} T^\dagger)^\perp = \operatorname{range} T$$

by 6.52.
Next we compute the range of $T^\dagger$. Every value of $T^\dagger$ lies in $(\operatorname{null} T)^\perp$, because $(T|_{(\operatorname{null} T)^\perp})^{-1}$ maps into $(\operatorname{null} T)^\perp$. Conversely, if $v \in (\operatorname{null} T)^\perp$, then $T^\dagger(Tv) = P_{(\operatorname{null} T)^\perp} v = v$ by 6.69(c), so $v \in \operatorname{range} T^\dagger$. Thus

$$\operatorname{range} T^\dagger = (\operatorname{null} T)^\perp.$$

Finally, if $w \in \operatorname{range} T$, then $P_{\operatorname{range} T} w = w$ and hence $T^\dagger w = (T|_{(\operatorname{null} T)^\perp})^{-1} w$. In other words, the restriction of $T^\dagger$ to $(\operatorname{null} T^\dagger)^\perp = \operatorname{range} T$ equals the invertible map $(T|_{(\operatorname{null} T)^\perp})^{-1}$ from $\operatorname{range} T$ onto $(\operatorname{null} T)^\perp$, whose inverse is $T|_{(\operatorname{null} T)^\perp}$. Therefore, for each $v \in V$,

$$(T^\dagger)^\dagger v = \big(T^\dagger|_{(\operatorname{null} T^\dagger)^\perp}\big)^{-1} P_{\operatorname{range} T^\dagger} v = T|_{(\operatorname{null} T)^\perp}\big(P_{(\operatorname{null} T)^\perp} v\big) = T\big(P_{(\operatorname{null} T)^\perp} v\big).$$

Because $v - P_{(\operatorname{null} T)^\perp} v \in \operatorname{null} T$, we have $T(P_{(\operatorname{null} T)^\perp} v) = Tv$. Hence $(T^\dagger)^\dagger = T$.

Step 2: if T is a positive operator, then so is $T^\dagger$.
Suppose T is positive. Then T is self-adjoint, so by the spectral theorem (7.29 if $\mathbf{F} = \mathbf{R}$, and 7.31 if $\mathbf{F} = \mathbf{C}$, using that self-adjoint operators are normal) there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T, say $Te_k = \lambda_k e_k$. By 7.38(b), each $\lambda_k \geq 0$. Let

$$\Omega = \{k : \lambda_k > 0\}.$$

Writing $v = a_1 e_1 + \cdots + a_n e_n$, we have $Tv = \sum_k a_k \lambda_k e_k$, which is 0 if and only if $a_k = 0$ for every $k \in \Omega$. Hence

$$\operatorname{null} T = \operatorname{span}\{e_k : k \notin \Omega\}, \qquad \operatorname{range} T = \operatorname{span}\{e_k : k \in \Omega\},$$

the second equality because $Tv = \sum_{k\in\Omega} a_k \lambda_k e_k$ and each e_k with $k \in \Omega$ equals $T(e_k/\lambda_k)$. Because $e_1, \dots, e_n$ is orthonormal, $(\operatorname{null} T)^\perp = \operatorname{span}\{e_k : k \in \Omega\} = \operatorname{range} T$.
Now compute $T^\dagger$ on the basis. If $k \in \Omega$, then $e_k \in \operatorname{range} T$, so $P_{\operatorname{range} T} e_k = e_k$; and $e_k/\lambda_k \in (\operatorname{null} T)^\perp$ satisfies $T(e_k/\lambda_k) = e_k$, so

$$T^\dagger e_k = \frac{1}{\lambda_k}\, e_k.$$

If $k \notin \Omega$, then e_k is orthogonal to every e_j with $j \in \Omega$, so $e_k \in (\operatorname{range} T)^\perp$ and hence $T^\dagger e_k = 0$.
Thus the matrix of $T^\dagger$ with respect to the orthonormal basis $e_1, \dots, e_n$ is the diagonal matrix whose k^{th} diagonal entry is $1/\lambda_k$ (for $k \in \Omega$) or 0 (for $k \notin \Omega$). All these numbers are nonnegative, so $T^\dagger$ is a positive operator by the implication (c) $\Rightarrow$ (a) of 7.38.
Step 3: if $T^\dagger$ is a positive operator, then so is T.
Apply Step 2 to the operator $T^\dagger$: since $T^\dagger$ is positive, $(T^\dagger)^\dagger$ is positive. By Step 1, $(T^\dagger)^\dagger = T$, so T is a positive operator.
Combining Steps 2 and 3 proves that T is positive if and only if $T^\dagger$ is positive.

Problem 7C.9. Suppose $T \in \mathcal{L}(V)$ is a positive operator and $S \in \mathcal{L}(W, V)$. Prove that S^*TS is a positive operator on W.

Solution. Because $S \in \mathcal{L}(W, V)$, we have $S^* \in \mathcal{L}(V, W)$, and thus the composition

$$S^*TS : W \longrightarrow W$$

makes sense and is an operator on W; explicitly, S maps W to V, then T maps V to V, then S^* maps V back to W.
First we show that S^*TS is self-adjoint. Because T is a positive operator, T is self-adjoint, so $T^* = T$. Using 7.5(d) twice and then 7.5(c),

$$(S^*TS)^* = S^*T^*(S^*)^* = S^*TS.$$

Hence S^*TS is self-adjoint.
Now let $w \in W$. By the definition of the adjoint (applied to S), for every $v \in V$ we have $\langle Sw, v\rangle = \langle w, S^*v\rangle$, and hence $\langle S^*v, w\rangle = \overline{\langle w, S^*v\rangle} = \overline{\langle Sw, v\rangle} = \langle v, Sw\rangle$. Taking $v = TSw$ gives

$$\langle (S^*TS)w, w\rangle = \langle S^*(TSw), w\rangle = \langle TSw,\, Sw\rangle.$$

Because T is a positive operator and $Sw \in V$, the right side is nonnegative:

$$\langle TSw, Sw\rangle \geq 0.$$

Thus $\langle (S^*TS)w, w\rangle \geq 0$ for every $w \in W$.
Since S^*TS is self-adjoint and $\langle (S^*TS)w, w\rangle \geq 0$ for all $w \in W$, the operator S^*TS is positive by 7.34.
Remark: the special case $T = I$ shows that S^*S is a positive operator on W for every $S \in \mathcal{L}(W, V)$. When $W = V$ this is exactly the implication (f) $\Rightarrow$ (a) of 7.38; for $W \neq V$ it is a mild generalization of that implication, since 7.38(f) is stated only for operators.

Problem 7C.10. Suppose T is a positive operator on V. Suppose $v, w \in V$ are such that

$$Tv = w \quad \text{and} \quad Tw = v.$$

Prove that $v = w$.

Solution. Apply T to the vector $v - w$ and use linearity together with the two hypotheses:

$$T(v - w) = Tv - Tw = w - v = -(v - w).$$

Now take the inner product of both sides with $v - w$:

$$\langle T(v - w),\, v - w\rangle = \langle -(v - w),\, v - w\rangle = -\langle v - w,\, v - w\rangle = -\|v - w\|^2.$$

Because T is a positive operator, the left side is nonnegative (by 7.34). Hence

$$-\|v - w\|^2 \geq 0,$$

which forces $\|v - w\|^2 \leq 0$. Since norms are nonnegative, $\|v - w\| = 0$, and therefore $v - w = 0$ by the positive-definiteness of the norm. Thus $v = w$, as desired.
Remark: the same conclusion can be reached spectrally. The computation $T(v - w) = -(v - w)$ says that if $v \neq w$, then $v - w$ is an eigenvector of T with eigenvalue -1. But every eigenvalue of a positive operator is nonnegative, by 7.38(b). This contradiction again gives $v = w$.

Problem 7C.11. Suppose T is a positive operator on V and U is a subspace of V invariant under T. Prove that $T|_U \in \mathcal{L}(U)$ is a positive operator on U.

Solution. Regard U as an inner product space, using the inner product of V restricted to $U \times U$; all the inner product axioms hold on U because they hold on V. Because U is invariant under T, the restriction $T|_U$ maps U into U, and it is clearly linear; thus $T|_U \in \mathcal{L}(U)$.
First we check that $T|_U$ is self-adjoint as an operator on U. Its adjoint $(T|_U)^* \in \mathcal{L}(U)$ is the unique operator on U satisfying

$$\langle (T|_U)u_1,\, u_2\rangle = \langle u_1,\, (T|_U)^* u_2\rangle \qquad \text{for all } u_1, u_2 \in U.$$

Now let $u_1, u_2 \in U$. Since T is positive it is self-adjoint, so

$$\langle (T|_U)u_1,\, u_2\rangle = \langle Tu_1,\, u_2\rangle = \langle u_1,\, Tu_2\rangle = \langle u_1,\, (T|_U)u_2\rangle,$$

where the last equality uses $u_2 \in U$. Since the adjoint is uniquely determined by the displayed identity (see 7.1), we conclude that $(T|_U)^* = T|_U$. Hence $T|_U$ is self-adjoint.
Second, let $u \in U$. Then

$$\langle (T|_U)u,\, u\rangle = \langle Tu,\, u\rangle \geq 0,$$

because T is a positive operator on V and $u \in V$.
Thus $T|_U$ is a self-adjoint operator on U with $\langle (T|_U)u, u\rangle \geq 0$ for every $u \in U$, so $T|_U$ is a positive operator on U by 7.34.

Problem 7C.12. Suppose $T \in \mathcal{L}(V)$ is a positive operator. Prove that T^k is a positive operator for every positive integer k.

Solution. Because T is positive, T is self-adjoint, so $T^* = T$.
First, T^k is self-adjoint for every positive integer k. Indeed, repeated use of 7.5(d) gives

$$(T^k)^* = (T \cdots T)^* = T^* \cdots T^* = (T^*)^k = T^k,$$

the last equality because $T^* = T$. (Formally this is an induction on k: the case $k = 1$ is the hypothesis, and if $(T^k)^* = T^k$, then $(T^{k+1})^* = (T\,T^k)^* = (T^k)^* T^* = T^k T = T^{k+1}$.)
Second, we show that $\langle T^k v, v\rangle \geq 0$ for every $v \in V$. There are two cases, according to the parity of k.
Case 1: $k = 2m$ for some positive integer m. Let $v \in V$. Using $(T^m)^* = T^m$ from the first paragraph,

$$\langle T^{2m} v,\, v\rangle = \langle T^m(T^m v),\, v\rangle = \langle T^m v,\, (T^m)^* v\rangle = \langle T^m v,\, T^m v\rangle = \|T^m v\|^2 \geq 0.$$

Case 2: $k = 2m + 1$ for some nonnegative integer m. Let $v \in V$. Then, again using $(T^m)^* = T^m$ (which holds trivially when $m = 0$, since $T^0 = I$),

$$\langle T^{2m+1} v,\, v\rangle = \langle T^m\big(T(T^m v)\big),\, v\rangle = \langle T(T^m v),\, (T^m)^* v\rangle = \langle T(T^m v),\, T^m v\rangle \geq 0,$$

where the final inequality holds because T is a positive operator and $T^m v \in V$.
In both cases $\langle T^k v, v\rangle \geq 0$ for all $v \in V$. Combined with the self-adjointness of T^k, this shows that T^k is a positive operator, by 7.34.
Remark: alternatively, use the spectral theorem (7.29 or 7.31) to choose an orthonormal basis $e_1, \ldots, e_n$ of V with $Te_j = \lambda_j e_j$; by 7.38(b) each $\lambda_j \geq 0$. Then $T^k e_j = \lambda_j^k e_j$, so the matrix of T^k with respect to this orthonormal basis is diagonal with the nonnegative numbers $\lambda_1^k, \ldots, \lambda_n^k$ on the diagonal. Hence T^k is positive by the implication (c) $\Rightarrow$ (a) of 7.38.

Problem 7C.13. Suppose $T \in \mathcal{L}(V)$ is self-adjoint and $\alpha \in \mathbf{R}$.
(a) Prove that $T - \alpha I$ is a positive operator if and only if α is less than or equal to every eigenvalue of T.
(b) Prove that $\alpha I - T$ is a positive operator if and only if α is greater than or equal to every eigenvalue of T.

Solution. We first record two observations that are used in both parts.
Observation 1: both $T - \alpha I$ and $\alpha I - T$ are self-adjoint. Indeed, $I^* = I$ by 7.5(e), and $(\alpha I)^* = \overline{\alpha}\, I^* = \alpha I$ by 7.5(b), because α is real. Hence, using 7.5(a) and 7.5(b),

$$(T - \alpha I)^* = T^* - \alpha I = T - \alpha I, \qquad (\alpha I - T)^* = \alpha I - T^* = \alpha I - T.$$

Observation 2 (shifting eigenvalues): for $\mu \in \mathbf{F}$ and $v \in V$ with $v \neq 0$,

$$(T - \alpha I)v = \mu v \iff Tv = (\mu + \alpha)v,$$

and likewise

$$(\alpha I - T)v = \mu v \iff Tv = (\alpha - \mu)v.$$

Hence the eigenvalues of $T - \alpha I$ are exactly the numbers $\lambda - \alpha$ with λ an eigenvalue of T, and the eigenvalues of $\alpha I - T$ are exactly the numbers $\alpha - \lambda$ with λ an eigenvalue of T. Note also that all eigenvalues of T are real, by 7.12, so all these differences are real numbers and the inequalities below make sense.
Proof of (a).
By Observation 1, the operator $T - \alpha I$ is self-adjoint. Hence the equivalence (a) $\Leftrightarrow$ (b) of 7.38 says

$$T - \alpha I \text{ is a positive operator} \iff \text{every eigenvalue of } T - \alpha I \text{ is nonnegative.}$$

By Observation 2, the eigenvalues of $T - \alpha I$ are precisely the numbers $\lambda - \alpha$ as λ ranges over the eigenvalues of T. Thus

$$T - \alpha I \text{ is positive} \iff \lambda - \alpha \geq 0 \text{ for every eigenvalue } \lambda \text{ of } T \iff \alpha \leq \lambda \text{ for every eigenvalue } \lambda \text{ of } T,$$

which is exactly the assertion that α is less than or equal to every eigenvalue of T.
Proof of (b).
By Observation 1, the operator $\alpha I - T$ is self-adjoint, so again by the equivalence (a) $\Leftrightarrow$ (b) of 7.38,

$$\alpha I - T \text{ is a positive operator} \iff \text{every eigenvalue of } \alpha I - T \text{ is nonnegative.}$$

By Observation 2, the eigenvalues of $\alpha I - T$ are precisely the numbers $\alpha - \lambda$ as λ ranges over the eigenvalues of T. Thus

$$\alpha I - T \text{ is positive} \iff \alpha - \lambda \geq 0 \text{ for every eigenvalue } \lambda \text{ of } T \iff \alpha \geq \lambda \text{ for every eigenvalue } \lambda \text{ of } T,$$

which is exactly the assertion that α is greater than or equal to every eigenvalue of T.
Remark 1: part (b) also follows formally from part (a). Apply (a) to the self-adjoint operator $-T$ and the real number $-\alpha$: the operator $\alpha I - T = (-T) - (-\alpha)I$ is positive if and only if $-\alpha$ is less than or equal to every eigenvalue of $-T$. Since the eigenvalues of $-T$ are exactly the numbers $-\lambda$ with λ an eigenvalue of T, this says $-\alpha \leq -\lambda$ for every eigenvalue λ of T, that is, $\alpha \geq \lambda$ for every eigenvalue λ of T.
Remark 2: unwinding the definition, part (a) says that $\langle Tv, v\rangle \geq \alpha\|v\|^2$ for all $v \in V$ if and only if α is a lower bound for the eigenvalues of T; part (b) says that $\langle Tv, v\rangle \leq \alpha\|v\|^2$ for all $v \in V$ if and only if α is an upper bound for the eigenvalues of T. Because V is nonzero and finite-dimensional and T is self-adjoint, T does have eigenvalues: if $\mathbf{F} = \mathbf{C}$ this is 5.19, and if $\mathbf{F} = \mathbf{R}$ the real spectral theorem 7.29 supplies an orthonormal basis of the nonzero space V with respect to which the matrix of T is diagonal, so each basis vector is an eigenvector. There are only finitely many eigenvalues (by 5.12), so a smallest and a largest eigenvalue exist. Taking α to be each of these in turn shows that the extreme eigenvalues of T are the best possible bounds of the two displayed kinds.

Problem 7C.14. Suppose T is a positive operator on V and $v_1, \dots, v_m \in V$. Prove that

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\langle Tv_k, v_j\rangle \geq 0.$$

Solution. Set

$$v = v_1 + \cdots + v_m \in V.$$

The whole double sum collapses to a single inner product evaluated at v.
First fix $j \in \{1, \dots, m\}$ and sum over k. Because T is linear and because the inner product is additive in its first slot (6.6(a)),

$$\sum_{k=1}^{m}\langle Tv_k, v_j\rangle = \Big\langle \sum_{k=1}^{m} Tv_k,\ v_j \Big\rangle = \Big\langle T\Big(\sum_{k=1}^{m} v_k\Big),\ v_j \Big\rangle = \langle Tv,\ v_j\rangle.$$

Now sum over j. The inner product is additive in its second slot by 6.6(d), so

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\langle Tv_k, v_j\rangle = \sum_{j=1}^{m}\langle Tv,\ v_j\rangle = \Big\langle Tv,\ \sum_{j=1}^{m} v_j \Big\rangle = \langle Tv,\ v\rangle.$$

(Only additivity in the second slot is used here, not homogeneity, so no complex conjugates appear: all the coefficients are 1.)
Because T is a positive operator and $v \in V$, the definition of positive operator (7.34) gives $\langle Tv, v\rangle \geq 0$. In particular this number is real, so the inequality in the statement makes sense. Therefore

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\langle Tv_k, v_j\rangle = \langle Tv, v\rangle \geq 0,$$

as desired.
Remark: the argument works just as well for arbitrary scalar multiples. If $e_1, \dots, e_n$ is an orthonormal basis of V and $A = \mathcal{M}(T, (e_1, \dots, e_n))$, so that $A_{j,k} = \langle Te_k, e_j\rangle$, then applying the computation above to $v_k = c_k e_k$ (where $c_1, \dots, c_n \in \mathbf{F}$) gives

$$\sum_{j=1}^{n}\sum_{k=1}^{n} A_{j,k}\, c_k\overline{c_j} = \langle Tv, v\rangle \geq 0,$$

where $v = c_1 e_1 + \cdots + c_n e_n$. This is the familiar statement that the matrix of a positive operator with respect to an orthonormal basis is positive semidefinite; the exercise as stated is the case in which all the coefficients equal 1, but with arbitrary vectors in place of basis vectors.

Problem 7C.15. Suppose $T \in \mathcal{L}(V)$ is self-adjoint. Prove that there exist positive operators $A, B \in \mathcal{L}(V)$ such that

$$T = A - B \quad \text{and} \quad \sqrt{T^*T} = A + B \quad \text{and} \quad AB = BA = 0.$$

Solution. Because T is self-adjoint, the spectral theorem (7.29 if $\mathbf{F} = \mathbf{R}$, 7.31 if $\mathbf{F} = \mathbf{C}$) provides an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, all of which are real (by 7.12).
For a real number λ set

$$\lambda^+ = \max\{\lambda, 0\}, \qquad \lambda^- = \max\{-\lambda, 0\}.$$

Then $\lambda^+ \geq 0$, $\lambda^- \geq 0$, and for every real λ,

$$\lambda^+ - \lambda^- = \lambda, \qquad \lambda^+ + \lambda^- = |\lambda|, \qquad \lambda^+ \lambda^- = 0$$

(the last equality holds because at least one of λ^+, λ^- is 0).
By the linear map lemma (3.4) there exist $A, B \in \mathcal{L}(V)$ with

$$Ae_k = \lambda_k^+ e_k \quad \text{and} \quad Be_k = \lambda_k^- e_k \qquad (k = 1, \ldots, n).$$

Both operators are positive: the matrix of A with respect to the orthonormal basis $e_1, \ldots, e_n$ is the diagonal matrix with the nonnegative numbers $\lambda_1^+, \ldots, \lambda_n^+$ on the diagonal, so A is a positive operator by 7.38 (condition (c) implies condition (a)). The same argument applies to B.
Now the three required identities. For each k,

$$(A - B)e_k = (\lambda_k^+ - \lambda_k^-)e_k = \lambda_k e_k = Te_k,$$

and since $e_1, \ldots, e_n$ is a basis of V this gives $T = A - B$.
Next,

$$ABe_k = \lambda_k^- Ae_k = \lambda_k^- \lambda_k^+ e_k = 0 \quad \text{and} \quad BAe_k = \lambda_k^+ \lambda_k^- e_k = 0$$

for each k, so $AB = BA = 0$.
Finally, put $C = A + B$; thus $Ce_k = (\lambda_k^+ + \lambda_k^-)e_k = |\lambda_k| e_k$. The matrix of C with respect to $e_1, \ldots, e_n$ is diagonal with the nonnegative numbers $|\lambda_1|, \ldots, |\lambda_n|$ on the diagonal, so C is a positive operator by 7.38. Moreover

$$C^2 e_k = |\lambda_k|^2 e_k = \lambda_k^2 e_k = T^2 e_k = T^*T e_k$$

for each k, where the last equality uses $T^* = T$. Hence $C^2 = T^*T$, so C is a positive square root of the positive operator T^*T. By the uniqueness of positive square roots (7.39) and the notation 7.40,

$$A + B = C = \sqrt{T^*T},$$

as required.

Problem 7C.16. Suppose T is a positive operator on V. Prove that

$$\operatorname{null} \sqrt{T} = \operatorname{null} T \quad \text{and} \quad \operatorname{range} \sqrt{T} = \operatorname{range} T.$$

Solution. Write $R = \sqrt{T}$, the unique positive square root of T (7.39, 7.40); thus R is positive, $R^* = R$, and $R^2 = T$.
Null spaces. Suppose $v \in \operatorname{null} R$. Then

$$Tv = R^2 v = R(Rv) = R0 = 0,$$

so $v \in \operatorname{null} T$. Hence $\operatorname{null} \sqrt{T} \subseteq \operatorname{null} T$.
Conversely, suppose $v \in \operatorname{null} T$. Then

$$0 = \langle Tv, v\rangle = \langle R^2 v, v\rangle = \langle Rv, R^* v\rangle = \langle Rv, Rv\rangle = \|Rv\|^2,$$

using $R^* = R$. By the positive definiteness of the norm, $Rv = 0$, so $v \in \operatorname{null} R$. Hence $\operatorname{null} T \subseteq \operatorname{null} \sqrt{T}$, and the two null spaces are equal.
Ranges. First, if $w \in \operatorname{range} T$, say $w = Tv$, then $w = R(Rv) \in \operatorname{range} R$. Thus

$$\operatorname{range} T \subseteq \operatorname{range} \sqrt{T}.$$

Second, by the fundamental theorem of linear maps (3.21) applied to T and to R,

$$\dim \operatorname{range} T = \dim V - \dim \operatorname{null} T = \dim V - \dim \operatorname{null} R = \dim \operatorname{range} R,$$

where the middle equality is the statement about null spaces just proved. A subspace of a finite-dimensional vector space that has the same dimension as the containing space equals it (2.39). Hence

$$\operatorname{range} \sqrt{T} = \operatorname{range} T.$$

Alternatively, the range statement follows from the null space statement without any dimension count: both T and R are self-adjoint, so 7.6(b) gives $\operatorname{range} T = \operatorname{range} T^* = (\operatorname{null} T)^\perp$ and likewise $\operatorname{range} R = (\operatorname{null} R)^\perp$; since $\operatorname{null} T = \operatorname{null} R$, the two ranges coincide.

Problem 7C.17. Suppose that $T \in \mathcal{L}(V)$ is a positive operator. Prove that there exists a polynomial p with real coefficients such that $\sqrt{T} = p(T)$.

Solution. Because T is positive, it is self-adjoint, so the spectral theorem (7.29 or 7.31, according to whether $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$) provides an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues, so $Te_k = \lambda_k e_k$ for each k. By 7.38 (condition (a) implies condition (b)), each λ_k is a nonnegative real number.
First we identify $\sqrt{T}$ on this basis. Let $R \in \mathcal{L}(V)$ be the operator determined by the linear map lemma (3.4) with

$$Re_k = \sqrt{\lambda_k}\, e_k \qquad (k = 1, \dots, n).$$

The matrix of R with respect to the orthonormal basis $e_1, \dots, e_n$ is diagonal with the nonnegative numbers $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$ on the diagonal, so R is a positive operator by 7.38. Also $R^2 e_k = \lambda_k e_k = Te_k$ for each k, so $R^2 = T$. Thus R is a positive square root of T, and by the uniqueness statement 7.39 we conclude

$$\sqrt{T}\, e_k = \sqrt{\lambda_k}\, e_k \qquad (k = 1, \dots, n).$$

Now let $\mu_1, \dots, \mu_m$ be the distinct numbers among $\lambda_1, \dots, \lambda_n$; these are exactly the eigenvalues of T, and each is a nonnegative real number. Define the Lagrange interpolation polynomial

$$p(x) = \sum_{r=1}^{m} \sqrt{\mu_r} \prod_{\substack{l=1 \\ l \neq r}}^{m} \frac{x - \mu_l}{\mu_r - \mu_l},$$

with the convention that an empty product equals 1 (so that p is the constant $\sqrt{\mu_1}$ when $m = 1$). The denominators $\mu_r - \mu_l$ are nonzero because $\mu_1, \dots, \mu_m$ are distinct, and every number appearing in the formula is real; hence p is a polynomial with real coefficients.
Fix $i \in \{1, \dots, m\}$ and evaluate p at μ_i. If $r \neq i$, then the index $l = i$ occurs in the product attached to the summand indexed by r, and that factor has numerator $\mu_i - \mu_i = 0$; hence that whole summand equals 0. The summand indexed by $r = i$ has product $\prod_{l \neq i} \frac{\mu_i - \mu_l}{\mu_i - \mu_l} = 1$, so it contributes $\sqrt{\mu_i}$. Therefore

$$p(\mu_i) = \sqrt{\mu_i} \qquad (i = 1, \dots, m).$$

Finally, for each $k \in \{1, \dots, n\}$ we have $\lambda_k = \mu_i$ for some i, and since $Te_k = \lambda_k e_k$ it follows that $q(T)e_k = q(\lambda_k)e_k$ for every polynomial q (apply T repeatedly to get $T^s e_k = \lambda_k^s e_k$ for every nonnegative integer s, then use linearity). Therefore

$$p(T)e_k = p(\lambda_k)e_k = \sqrt{\lambda_k}\, e_k = \sqrt{T}\, e_k.$$

The two operators $p(T)$ and $\sqrt{T}$ agree on the basis $e_1, \dots, e_n$ of V, hence they are equal. Thus $\sqrt{T} = p(T)$ for a polynomial p with real coefficients.
A consequence used elsewhere: $\sqrt{T}$ commutes with every operator that commutes with T, since $p(T)$ does.

Problem 7C.18. Suppose S and T are positive operators on V. Prove that ST is a positive operator if and only if S and T commute.

Solution. Throughout, S and T are positive, hence self-adjoint: $S^* = S$ and $T^* = T$.
First suppose ST is a positive operator. Then ST is self-adjoint, so using 7.5(d) (the adjoint of a product is the product of the adjoints in reverse order),

$$ST = (ST)^* = T^* S^* = TS.$$

Thus S and T commute.
Conversely, suppose $ST = TS$. We first check that ST is self-adjoint:

$$(ST)^* = T^* S^* = TS = ST.$$

Now we show $\langle STv, v\rangle \geq 0$ for every $v \in V$. Let $R = \sqrt{S}$, the positive square root of S; so $R^* = R$ and $R^2 = S$. By Exercise 7C.17 there is a polynomial p with real coefficients such that $R = p(S)$. Since T commutes with S, it commutes with every power of S and hence with every polynomial in S; therefore

$$RT = p(S)T = Tp(S) = TR.$$

Let $v \in V$. Then

$$\begin{aligned}
\langle STv, v\rangle &= \langle R^2 Tv, v\rangle \\
&= \langle R(RT)v, v\rangle \\
&= \langle R(TR)v, v\rangle \\
&= \langle T(Rv), R^* v\rangle \\
&= \langle T(Rv), Rv\rangle,
\end{aligned}$$

where the fourth equality moves the outer R across the inner product as its adjoint, and the fifth uses $R^* = R$. Since T is a positive operator, $\langle T(Rv), Rv\rangle \geq 0$. Hence $\langle STv, v\rangle \geq 0$ for all $v \in V$.
So ST is self-adjoint and $\langle STv, v\rangle \geq 0$ for all v, which is exactly the statement that ST is a positive operator (7.34).

Problem 7C.19. Show that the identity operator on $\mathbf{F}^2$ has infinitely many self-adjoint square roots.

Solution. Give $\mathbf{F}^2$ its usual inner product, for which the standard basis $e_1 = (1, 0)$, $e_2 = (0, 1)$ is orthonormal.
For each $t \in \mathbf{R}$ let $R_t \in \mathcal{L}(\mathbf{F}^2)$ be the operator whose matrix with respect to the standard basis is

$$\mathcal{M}(R_t) = \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix},$$

that is,

$$R_t(w, z) = \big((\cos t)w + (\sin t)z,\ (\sin t)w - (\cos t)z\big).$$

Each R_t is self-adjoint. Indeed, the matrix above has real entries and equals its own transpose, hence equals its own conjugate transpose. Since the standard basis is orthonormal, the matrix of R_t^* is the conjugate transpose of the matrix of R_t (by 7.9), so $\mathcal{M}(R_t^*) = \mathcal{M}(R_t)$ and therefore $R_t^* = R_t$.
Each R_t is a square root of the identity. Computing the square of the matrix,

$$\begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix}^2 = \begin{pmatrix} \cos^2 t + \sin^2 t & \cos t \sin t - \sin t \cos t \\ \sin t \cos t - \cos t \sin t & \sin^2 t + \cos^2 t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Because taking matrices with respect to a fixed basis turns products of operators into products of matrices (3.43), it follows that $\mathcal{M}(R_t^2) = \mathcal{M}(I)$ and hence $R_t^2 = I$.
Finally, these operators are pairwise distinct for t in the interval $[0, \pi)$: the first entry of $R_t e_1$ is $\cos t$ and the second is $\sin t$, so $R_s = R_t$ forces $\cos s = \cos t$ and $\sin s = \sin t$, which for $s, t \in [0, \pi)$ forces $s = t$. Since $[0, \pi)$ is infinite, the family $\{R_t : t \in [0, \pi)\}$ consists of infinitely many distinct self-adjoint square roots of the identity operator on $\mathbf{F}^2$.
Geometrically, when $\mathbf{F} = \mathbf{R}$ each R_t is the reflection of $\mathbf{R}^2$ across the line through the origin at angle $t/2$; equivalently $R_t = 2P_{U_t} - I$ where U_t is that line. Reflections are self-adjoint and square to the identity, which is exactly what was verified above.
Note that none of the operators R_t is a positive operator. Indeed, put

$$u_t = \big(-\sin(t/2), \cos(t/2)\big) \in \mathbf{F}^2,$$

a vector of norm 1. Using the subtraction formulas for sine and cosine,

$$R_t u_t = \big(-\cos t \sin(t/2) + \sin t \cos(t/2),\ -\sin t \sin(t/2) - \cos t \cos(t/2)\big) = \big(\sin(t/2), -\cos(t/2)\big) = -u_t.$$

Hence $\langle R_t u_t, u_t \rangle = -\|u_t\|^2 = -1 < 0$, so R_t is not a positive operator (7.34). This is consistent with 7.39, which says the identity operator has exactly one positive square root, namely I itself; the operators above are self-adjoint square roots, not positive ones.

Problem 7C.20. Suppose $T \in \mathcal{L}(V)$ and $e_1, \dots, e_n$ is an orthonormal basis of V. Prove that T is a positive operator if and only if there exist $v_1, \dots, v_n \in V$ such that

$$\langle Te_k, e_j \rangle = \langle v_k, v_j \rangle$$

for all $j, k = 1, \dots, n$.
[The numbers $\{\langle Te_k, e_j \rangle\}_{j,k=1,\dots,n}$ are the entries in the matrix of T with respect to the orthonormal basis $e_1, \dots, e_n$.]

Solution. Throughout we use the expansion valid for any $u \in V$ with respect to the orthonormal basis $e_1, \dots, e_n$ (6.30(a)):

$$u = \sum_{j=1}^{n} \langle u, e_j \rangle e_j.$$

Consequently, two operators $A, B \in \mathcal{L}(V)$ are equal as soon as $\langle Ae_k, e_j \rangle = \langle Be_k, e_j \rangle$ for all j, k, because then $Ae_k = \sum_j \langle Ae_k, e_j \rangle e_j = \sum_j \langle Be_k, e_j \rangle e_j = Be_k$ for each k, and $e_1, \dots, e_n$ is a basis.
First suppose T is a positive operator. Let $R = \sqrt{T}$, so $R^* = R$ and $R^2 = T$ (7.39, 7.40). Define

$$v_k = Re_k \qquad (k = 1, \dots, n).$$

Then for all j, k,

$$\langle v_k, v_j \rangle = \langle Re_k, Re_j \rangle = \langle R^* Re_k, e_j \rangle = \langle R^2 e_k, e_j \rangle = \langle Te_k, e_j \rangle,$$

where the second equality is the defining property of the adjoint and the third uses $R^* = R$. So vectors of the required kind exist.

Conversely, suppose $v_1, \dots, v_n \in V$ satisfy $\langle Te_k, e_j \rangle = \langle v_k, v_j \rangle$ for all j, k. By the linear map lemma (3.4) there exists $R \in \mathcal{L}(V)$ with

$$Re_k = v_k \qquad (k = 1, \dots, n).$$

Then for all j, k,

$$\langle R^* Re_k, e_j \rangle = \langle Re_k, Re_j \rangle = \langle v_k, v_j \rangle = \langle Te_k, e_j \rangle.$$

By the remark in the first paragraph, $T = R^*R$. Hence T is a positive operator by 7.38 (condition (f) implies condition (a)).
Note that the second half of the argument delivers self-adjointness of T for free, even when $\mathbf{F} = \mathbf{R}$: an operator of the form R^*R always satisfies $(R^*R)^* = R^*(R^*)^* = R^*R$, using 7.5(c) and 7.5(d).

Problem 7C.21. Suppose n is a positive integer. The n-by-n Hilbert matrix is the n-by-n matrix whose entry in row j, column k is $\frac{1}{j+k-1}$. Suppose $T \in \mathcal{L}(V)$ is an operator whose matrix with respect to some orthonormal basis of V is the n-by-n Hilbert matrix. Prove that T is a positive invertible operator.
Example: The 4-by-4 Hilbert matrix is

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix}.$$

Solution. Let $e_1, \dots, e_n$ be an orthonormal basis of V with respect to which the matrix of T is the Hilbert matrix; thus $\dim V = n$ and

$$\langle Te_k, e_j \rangle = \frac{1}{j+k-1} \qquad (j, k = 1, \dots, n),$$

since the entry in row j, column k of the matrix of T with respect to an orthonormal basis is $\langle Te_k, e_j \rangle$ (this is 6.30(a) applied to Te_k).
The key identity is that for all $j, k \in \{1, \dots, n\}$,

$$\frac{1}{j+k-1} = \int_0^1 x^{j+k-2}\, dx = \int_0^1 x^{k-1} x^{j-1}\, dx,$$

which is valid because $j + k - 2 \geq 0$.
The operator is self-adjoint. The Hilbert matrix has real entries and is symmetric, since the entry in row j, column k and the entry in row k, column j are both $\frac{1}{j+k-1}$. So the matrix equals its own conjugate transpose. Since $e_1, \dots, e_n$ is orthonormal, the matrix of T^* is the conjugate transpose of the matrix of T (7.9), hence the matrices of T^* and T agree and therefore $T^* = T$.
The operator is positive, and in fact $\langle Tv, v \rangle > 0$ for $v \neq 0$. Let $v \in V$ and write $v = a_1 e_1 + \cdots + a_n e_n$ with $a_1, \dots, a_n \in \mathbf{F}$. Associate to v the polynomial function

$$q(x) = a_1 + a_2 x + \cdots + a_n x^{n-1} \qquad (x \in [0, 1]),$$

with coefficients in $\mathbf{F}$. Using linearity of T and of the inner product in its first slot, and conjugate-

linearity in its second slot,

$$\begin{aligned}\langle Tv, v\rangle &= \sum_{k=1}^{n}\sum_{j=1}^{n} a_k\overline{a_j}\,\langle Te_k, e_j\rangle \\ &= \sum_{k=1}^{n}\sum_{j=1}^{n} a_k\overline{a_j}\int_0^1 x^{\,k-1}x^{\,j-1}\,dx \\ &= \int_0^1 \Big(\sum_{k=1}^{n} a_k x^{\,k-1}\Big)\Big(\sum_{j=1}^{n}\overline{a_j}\,x^{\,j-1}\Big)dx \\ &= \int_0^1 q(x)\,\overline{q(x)}\,dx \\ &= \int_0^1 |q(x)|^2\,dx,\end{aligned}$$

where the third equality interchanges the finite sums with the integral (legitimate by linearity of the integral), and the fourth uses that x is real, so $\overline{q(x)} = \sum_j \overline{a_j} x^{\,j-1}$.
Since $|q(x)|^2 \geq 0$ for all $x \in [0,1]$, we get $\langle Tv, v\rangle \geq 0$ for every $v \in V$. Combined with the self-adjointness proved above, this shows T is a positive operator (7.34).
The operator is invertible. Suppose $v \in \operatorname{null} T$. Then $\langle Tv, v\rangle = 0$, so by the computation above

$$\int_0^1 |q(x)|^2\,dx = 0.$$

The function $x \mapsto |q(x)|^2$ is continuous and nonnegative on $[0,1]$, and a continuous nonnegative function on an interval whose integral over that interval equals 0 must be identically 0 there. Indeed, if $|q(x_0)|^2 = c > 0$ for some $x_0 \in [0,1]$, then by continuity $|q(x)|^2 > c/2$ on a subinterval $J \subseteq [0,1]$ of some positive length δ, giving $\int_0^1 |q|^2 \geq c\delta/2 > 0$, a contradiction. Hence $q(x) = 0$ for every $x \in [0,1]$.
So the polynomial $a_1 + a_2 x + \cdots + a_n x^{\,n-1}$ vanishes at every point of $[0,1]$, hence has infinitely many zeros. If its coefficients were not all 0, it would be a polynomial of some degree $m \geq 0$; a nonzero constant has no zeros, and for $m \geq 1$ the polynomial has at most m zeros by 4.8. Either way we contradict having infinitely many zeros. Hence $a_1 = \cdots = a_n = 0$, so $v = 0$.
Therefore $\operatorname{null} T = \{0\}$, so T is injective (3.15). An injective operator on a finite-dimensional vector space is also surjective (3.65), hence bijective, hence invertible (3.63). Thus T is a positive invertible operator.
Remark. The identity $\frac{1}{j+k-1} = \int_0^1 x^{k-1}x^{j-1}\,dx$ says exactly that the Hilbert matrix records the inner products $\langle x^{k-1}, x^{j-1}\rangle$ of the linearly independent list $1, x, \ldots, x^{n-1}$ in the real inner product space of polynomials of degree less than n on $[0,1]$ with $\langle f, g\rangle = \int_0^1 fg$. That is the conceptual reason behind the result, and it is the shape of the criterion in Exercise 7C.20. That exercise cannot be quoted verbatim here, since it asks for vectors lying in V itself rather than in an auxiliary space; the computation above carries out the same idea directly inside V.

Problem 7C.22. Suppose $T \in \mathcal{L}(V)$ is a positive operator and $u \in V$ is such that $\|u\| = 1$ and $\|Tu\| \geq \|Tv\|$ for all $v \in V$ with $\|v\| = 1$. Show that u is an eigenvector of T corresponding to the largest eigenvalue of T.

Solution. Because a vector u with $\|u\| = 1$ exists, we have $V \neq \{0\}$.
Since T is positive, T is self-adjoint and all eigenvalues of T are nonnegative (by 7.38). By the spectral theorem (7.29 if $\mathbf{F} = \mathbf{R}$; if $\mathbf{F} = \mathbf{C}$ use 7.31, which applies because a self-adjoint operator is normal), there is an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T; let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues, so each $\lambda_k \geq 0$. Every eigenvalue of T appears in the list $\lambda_1, \ldots, \lambda_n$: indeed, if $Tw = \lambda w$ with $w \neq 0$, then writing $w = \sum_k a_k e_k$ and applying T gives $\sum_k a_k \lambda_k e_k = \sum_k a_k \lambda e_k$, so

$a_k(\lambda_k - \lambda) = 0$ for each k; since some $a_k \neq 0$ we get $\lambda = \lambda_k$ for that k. Relabelling, we may assume

$$\lambda_1 = \max\{\lambda_1, \dots, \lambda_n\},$$

so that λ_1 is the largest eigenvalue of T.
Write

$$u = a_1 e_1 + \cdots + a_n e_n$$

with $a_1, \dots, a_n \in \mathbf{F}$. Because $e_1, \dots, e_n$ is orthonormal and $\|u\| = 1$, we have $|a_1|^2 + \cdots + |a_n|^2 = 1$ (by 6.30). Also $Tu = a_1\lambda_1 e_1 + \cdots + a_n\lambda_n e_n$, hence

$$\|Tu\|^2 = \sum_{k=1}^{n} \lambda_k^2 |a_k|^2.$$

Now e_1 is a unit vector, so the hypothesis on u gives $\|Tu\| \geq \|Te_1\| = \|\lambda_1 e_1\| = \lambda_1$ (using $\lambda_1 \geq 0$). Therefore

$$\lambda_1^2 \leq \|Tu\|^2 = \sum_{k=1}^{n} \lambda_k^2 |a_k|^2 \leq \sum_{k=1}^{n} \lambda_1^2 |a_k|^2 = \lambda_1^2,$$

where the middle inequality holds because $0 \leq \lambda_k \leq \lambda_1$ for each k, hence $\lambda_k^2 \leq \lambda_1^2$. Thus both inequalities above are equalities, and in particular

$$\sum_{k=1}^{n} (\lambda_1^2 - \lambda_k^2)|a_k|^2 = 0.$$

Every term of this sum is nonnegative, so every term is 0. Hence $a_k = 0$ for each k with $\lambda_k^2 \neq \lambda_1^2$. Because λ_k and λ_1 are both nonnegative, $\lambda_k^2 = \lambda_1^2$ if and only if $\lambda_k = \lambda_1$. Therefore

$$u = \sum_{\{k \,:\, \lambda_k = \lambda_1\}} a_k e_k,$$

which is a linear combination of eigenvectors of T all corresponding to the eigenvalue λ_1. Applying T termwise gives $Tu = \lambda_1 u$. Since $\|u\| = 1$, the vector u is nonzero, so u is an eigenvector of T corresponding to λ_1, the largest eigenvalue of T.
Note that positivity (or at least nonnegativity of the eigenvalues) is essential: for the self-adjoint operator on $\mathbf{R}^2$ with matrix $\begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$ with respect to the standard basis, the unit vector maximizing $\|Tv\|$ is $(0, 1)$, whose eigenvalue -2 is the smallest eigenvalue rather than the largest.

Problem 7C.23. For $T \in \mathcal{L}(V)$ and $u, v \in V$, define $\langle u, v\rangle_T$ by $\langle u, v\rangle_T = \langle Tu, v\rangle$.
(a) Suppose $T \in \mathcal{L}(V)$. Prove that $\langle \cdot, \cdot \rangle_T$ is an inner product on V if and only if T is an invertible positive operator (with respect to the original inner product $\langle \cdot, \cdot \rangle$).
(b) Prove that every inner product on V is of the form $\langle \cdot, \cdot \rangle_T$ for some positive invertible operator $T \in \mathcal{L}(V)$.

Solution. Throughout, $\langle \cdot, \cdot \rangle$ denotes the original inner product on V, and adjoints, positivity and self-adjointness refer to it.
(a) First suppose T is an invertible positive operator. We verify the five conditions in the definition of an inner product (6.2).
Positivity: for $v \in V$ we have $\langle v, v\rangle_T = \langle Tv, v\rangle \geq 0$, because T is positive.
Definiteness: if $v = 0$ then $\langle v, v\rangle_T = \langle T0, 0\rangle = 0$. Conversely, if $\langle v, v\rangle_T = 0$, then $\langle Tv, v\rangle = 0$, so $Tv = 0$ by 7.43; because T is invertible (hence injective), this gives $v = 0$.
Additivity in the first slot: for $u_1, u_2, v \in V$,

$$\langle u_1 + u_2, v\rangle_T = \langle T(u_1 + u_2), v\rangle = \langle Tu_1, v\rangle + \langle Tu_2, v\rangle = \langle u_1, v\rangle_T + \langle u_2, v\rangle_T,$$

using linearity of T and additivity of $\langle \cdot, \cdot \rangle$ in its first slot.
Homogeneity in the first slot: for $\lambda \in \mathbf{F}$ and $u, v \in V$,

$$\langle \lambda u, v \rangle_T = \langle T(\lambda u), v \rangle = \langle \lambda T u, v \rangle = \lambda \langle T u, v \rangle = \lambda \langle u, v \rangle_T.$$

Conjugate symmetry: because T is positive, T is self-adjoint, so for $u, v \in V$,

$$\langle v, u \rangle_T = \langle Tv, u \rangle = \langle v, Tu \rangle = \overline{\langle Tu, v \rangle} = \overline{\langle u, v \rangle_T}.$$

Hence $\langle \cdot, \cdot \rangle_T$ is an inner product on V.
Conversely, suppose $\langle \cdot, \cdot \rangle_T$ is an inner product on V. Conjugate symmetry of $\langle \cdot, \cdot \rangle_T$ and of $\langle \cdot, \cdot \rangle$ give, for all $u, v \in V$,

$$\langle Tu, v \rangle = \langle u, v \rangle_T = \overline{\langle v, u \rangle_T} = \overline{\langle Tv, u \rangle} = \langle u, Tv \rangle.$$

On the other hand $\langle Tu, v \rangle = \langle u, T^* v \rangle$ by the definition of the adjoint (7.1). Hence $\langle u, T^* v - Tv \rangle = 0$ for all $u, v \in V$; taking $u = T^* v - Tv$ shows $T^* v = Tv$ for every $v \in V$. Thus $T^* = T$, so T is self-adjoint.
Positivity of $\langle \cdot, \cdot \rangle_T$ gives $\langle Tv, v \rangle = \langle v, v \rangle_T \geq 0$ for every $v \in V$. Together with self-adjointness this says exactly that T is a positive operator (7.34).
Finally, T is injective: if $Tv = 0$, then $\langle v, v \rangle_T = \langle Tv, v \rangle = 0$, so $v = 0$ by the definiteness of the inner product $\langle \cdot, \cdot \rangle_T$. An injective operator on a finite-dimensional vector space is invertible (3.65). Hence T is an invertible positive operator.
(b) Let $(\cdot, \cdot)$ denote an arbitrary inner product on V. Fix $u \in V$ and consider the function $\psi_u \colon V \to \mathbf{F}$ defined by

$$\psi_u(v) = (v, u).$$

Because an inner product is linear in its first slot, ψ_u is a linear functional on V. By the Riesz representation theorem (6.42) applied to the original inner product $\langle \cdot, \cdot \rangle$, there is a unique vector, which we call $Tu \in V$, such that

$$(v, u) = \langle v, Tu \rangle \quad \text{for all } v \in V.$$

This defines a function $T \colon V \to V$.
T is linear. For $u_1, u_2 \in V$ and every $v \in V$,

$$\langle v, T(u_1 + u_2) \rangle = (v, u_1 + u_2) = (v, u_1) + (v, u_2) = \langle v, Tu_1 \rangle + \langle v, Tu_2 \rangle = \langle v, Tu_1 + Tu_2 \rangle,$$

where the second equality holds because an inner product is additive in its second slot (this follows from additivity in the first slot and conjugate symmetry). Since $v \in V$ is arbitrary, taking $v = T(u_1 + u_2) - Tu_1 - Tu_2$ shows $T(u_1 + u_2) = Tu_1 + Tu_2$. Similarly, for $\lambda \in \mathbf{F}$ and every $v \in V$,

$$\langle v, T(\lambda u) \rangle = (v, \lambda u) = \overline{\lambda}\,(v, u) = \overline{\lambda}\,\langle v, Tu \rangle = \langle v, \lambda Tu \rangle,$$

using conjugate homogeneity in the second slot of both inner products; hence $T(\lambda u) = \lambda Tu$. Thus $T \in \mathcal{L}(V)$.
Now for all $u, v \in V$,

$$\langle u, v \rangle_T = \langle Tu, v \rangle = \overline{\langle v, Tu \rangle} = \overline{(v, u)} = (u, v).$$

So $\langle \cdot, \cdot \rangle_T$ is precisely the given inner product $(\cdot, \cdot)$. In particular $\langle \cdot, \cdot \rangle_T$ is an inner product on V, so by part (a) the operator T is positive and invertible. Hence every inner product on V has the form $\langle \cdot, \cdot \rangle_T$ for some positive invertible $T \in \mathcal{L}(V)$.

Problem 7C.24. Suppose S and T are positive operators on V. Prove that

$$\operatorname{null}(S + T) = \operatorname{null} S \cap \operatorname{null} T.$$

Solution. First suppose $v \in \operatorname{null} S \cap \operatorname{null} T$. Then $Sv = 0$ and $Tv = 0$, so $(S+T)v = Sv + Tv = 0$. Hence $v \in \operatorname{null}(S+T)$. This proves the inclusion

$$\operatorname{null} S \cap \operatorname{null} T \subseteq \operatorname{null}(S+T),$$

and note that this inclusion holds for arbitrary operators S and T.
Now suppose $v \in \operatorname{null}(S+T)$, so $Sv + Tv = 0$. Taking the inner product with v gives

$$0 = \langle Sv + Tv, v\rangle = \langle Sv, v\rangle + \langle Tv, v\rangle.$$

Because S and T are positive operators, $\langle Sv, v\rangle \geq 0$ and $\langle Tv, v\rangle \geq 0$. Two nonnegative numbers summing to 0 must both equal 0, so

$$\langle Sv, v\rangle = 0 \quad \text{and} \quad \langle Tv, v\rangle = 0.$$

By 7.43 (a positive operator R with $\langle Rv, v\rangle = 0$ satisfies $Rv = 0$), applied to S and then to T, we conclude $Sv = 0$ and $Tv = 0$. Hence $v \in \operatorname{null} S \cap \operatorname{null} T$, which gives the reverse inclusion

$$\operatorname{null}(S+T) \subseteq \operatorname{null} S \cap \operatorname{null} T.$$

Combining the two inclusions yields $\operatorname{null}(S+T) = \operatorname{null} S \cap \operatorname{null} T$, as desired.

Problem 7C.25. Let T be the second derivative operator in Exercise 31(b) in Section 7A. Show that $-T$ is a positive operator.

Solution. Recall the setting of Exercise 31 in Section 7A. Fix a positive integer n. In the inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} fg,$$

let

$$V = \operatorname{span}(1, \cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin nx),$$

and let $T \in \mathcal{L}(V)$ be defined by $Tf = f''$.
Let $D \in \mathcal{L}(V)$ be the differentiation operator $Df = f'$ of Exercise 31(a). This does map V into V, because

$$1' = 0, \qquad (\cos kx)' = -k \sin kx, \qquad (\sin kx)' = k \cos kx$$

for $k = 1, \ldots, n$, and each of these lies in V; linearity of differentiation then gives $Df \in V$ for every $f \in V$. Clearly $T = D^2$.
Next we verify that $D^* = -D$, which is the content of Exercise 31(a). Every element of V is a linear combination of 1, $\cos kx$, $\sin kx$ for $k \leq n$, each of which is 2π-periodic and continuously differentiable; hence every $f \in V$ satisfies $f(-\pi) = f(\pi)$ and $f'(-\pi) = f'(\pi)$. For $f, g \in V$, integration by parts gives

$$\langle Df, g\rangle = \int_{-\pi}^{\pi} f'g = \Big[fg\Big]_{-\pi}^{\pi} - \int_{-\pi}^{\pi} fg' = f(\pi)g(\pi) - f(-\pi)g(-\pi) - \langle f, Dg\rangle = -\langle f, Dg\rangle,$$

where the boundary term vanishes by the periodicity noted above. Thus $\langle Df, g\rangle = \langle f, (-D)g\rangle$ for all $f, g \in V$, which by the definition of the adjoint (7.1) says exactly that $D^* = -D$; the adjoint is uniquely determined by this condition, because if $\langle f, w_1\rangle = \langle f, w_2\rangle$ for all $f \in V$, then taking $f = w_1 - w_2$ gives $w_1 = w_2$.
Now

$$-T = -D^2 = (-D)D = D^*D.$$

By 7.38, an operator of the form R^*R with $R \in \mathcal{L}(V)$ is positive; taking $R = D$ shows that $-T$ is a positive operator.

For a direct verification: $-T$ is self-adjoint because

$$(-T)^* = (D^*D)^* = D^*(D^*)^* = D^*D = -T,$$

using 7.5, and for every $f \in V$,

$$\langle -Tf, f\rangle = \langle D^*Df, f\rangle = \langle Df, Df\rangle = \int_{-\pi}^{\pi} (f')^2 \geq 0.$$

Hence $-T$ is self-adjoint with $\langle (-T)f, f\rangle \geq 0$ for all $f \in V$, so $-T$ is positive by definition (7.34). Alternatively one can see this from the eigenvalues. The functions

$$1,\ \cos x,\ \ldots,\ \cos nx,\ \sin x,\ \ldots,\ \sin nx$$

are pairwise orthogonal with respect to $\langle f, g\rangle = \int_{-\pi}^{\pi} fg$ (by the standard orthogonality relations for trigonometric functions) and nonzero, so they form an orthogonal basis of V; dividing each by its norm gives an orthonormal basis of V. Each of these functions is an eigenvector of $-T$: we have $-T(1) = 0$, and

$$-T(\cos kx) = k^2 \cos kx, \qquad -T(\sin kx) = k^2 \sin kx$$

for $k = 1, \ldots, n$. Thus the matrix of $-T$ with respect to this orthonormal basis is diagonal with the nonnegative numbers $0, 1, 4, \ldots, n^2, 1, 4, \ldots, n^2$ on the diagonal, so $-T$ is a positive operator by 7.38.

7.4 Exercises 7D

Problem 7D.1. Suppose $\dim V \geq 2$ and $S \in \mathcal{L}(V, W)$. Prove that S is an isometry if and only if Se_1, Se_2 is an orthonormal list in W for every orthonormal list e_1, e_2 of length two in V.

Solution. Throughout, V is a finite-dimensional inner product space with $\dim V = n \geq 2$.
First suppose S is an isometry. By the equivalence of (a) and (c) in 7.49, S preserves inner products:

$$\langle Su, Sv\rangle = \langle u, v\rangle \quad \text{for all } u, v \in V.$$

Now let e_1, e_2 be any orthonormal list of length two in V. Then for $j, k \in \{1, 2\}$,

$$\langle Se_j, Se_k\rangle = \langle e_j, e_k\rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Thus Se_1, Se_2 is an orthonormal list in W, as desired.
Conversely, suppose that Se_1, Se_2 is an orthonormal list in W for every orthonormal list e_1, e_2 of length two in V. Because V is finite-dimensional, V has an orthonormal basis $e_1, \ldots, e_n$ (by 6.35). We claim that $Se_1, \ldots, Se_n$ is an orthonormal list in W.
To see this, let $j, k \in \{1, \ldots, n\}$ with $j \neq k$. Then e_j, e_k is an orthonormal list of length two in V, so by hypothesis Se_j, Se_k is an orthonormal list in W. In particular

$$\|Se_j\| = \|Se_k\| = 1 \quad \text{and} \quad \langle Se_j, Se_k\rangle = 0.$$

Here we use the hypothesis $n \geq 2$: it guarantees that for each index j there is at least one index $k \neq j$, so that every Se_j occurs in some such pair and hence has norm 1. Since the displayed equations hold for all $j \neq k$, the list $Se_1, \ldots, Se_n$ is orthonormal in W.
Because $e_1, \ldots, e_n$ is an orthonormal basis of V and $Se_1, \ldots, Se_n$ is an orthonormal list in W, the equivalence of (d) and (a) in 7.49 shows that S is an isometry.
Remark: the hypothesis $\dim V \geq 2$ cannot be dropped. If $\dim V = 1$ there are no orthonormal lists of length two in V, so the right-hand condition holds vacuously for every $S \in \mathcal{L}(V, W)$, including $S = 0$, which is not an isometry.

Problem 7D.2. Suppose $T \in \mathcal{L}(V, W)$ and $T \neq 0$. Prove that T is a scalar multiple of an isometry if and only if T preserves orthogonality.
[The phrase "T preserves orthogonality" means that $\langle Tu, Tv\rangle = 0$ for all $u, v \in V$ such that $\langle u, v\rangle = 0$.]

Solution. First suppose T is a scalar multiple of an isometry, say $T = cS$ where $c \in \mathbf{F}$ and $S \in \mathcal{L}(V, W)$ is an isometry. Suppose $u, v \in V$ with $\langle u, v\rangle = 0$. Because S is an isometry, S preserves inner products [by the equivalence of (a) and (c) in 7.49], so

$$\langle Tu, Tv\rangle = \langle cSu, cSv\rangle = c\overline{c}\,\langle Su, Sv\rangle = |c|^2\langle u, v\rangle = 0.$$

Thus T preserves orthogonality.
For the converse, suppose T preserves orthogonality. The heart of the proof is the following claim.
Claim: there is a constant $c \geq 0$ such that $\|Tu\| = c$ for every $u \in V$ with $\|u\| = 1$.
To prove the claim, we first show that any two orthonormal vectors have images of equal norm. So suppose $u, v \in V$ with $\|u\| = \|v\| = 1$ and $\langle u, v\rangle = 0$. Then

$$\langle u + v, u - v\rangle = \|u\|^2 - \langle u, v\rangle + \langle v, u\rangle - \|v\|^2 = 1 - 0 + 0 - 1 = 0,$$

so $u + v$ and $u - v$ are orthogonal. Because T preserves orthogonality, $\langle T(u + v), T(u - v)\rangle = 0$. Also $\langle Tu, Tv\rangle = 0$ and hence $\langle Tv, Tu\rangle = \overline{\langle Tu, Tv\rangle} = 0$, again because T preserves orthogonality. Expanding by additivity of the inner product in each slot now gives

$$0 = \langle Tu + Tv,\, Tu - Tv\rangle = \|Tu\|^2 - \langle Tu, Tv\rangle + \langle Tv, Tu\rangle - \|Tv\|^2 = \|Tu\|^2 - \|Tv\|^2.$$

Thus $\|Tu\| = \|Tv\|$ whenever u, v is an orthonormal list of length two in V.
Because $T \neq 0$ we have $V \neq \{0\}$, so V contains a nonzero vector and hence (after dividing by its norm) a vector of norm 1. Fix such a vector $u \in V$ and set $c = \|Tu\|$. Let $v \in V$ with $\|v\| = 1$; we show $\|Tv\| = c$.
If v is a scalar multiple of u, say $v = \lambda u$, then $1 = \|v\| = |\lambda|\,\|u\| = |\lambda|$, so

$$\|Tv\| = \|\lambda Tu\| = |\lambda|\,\|Tu\| = c.$$

If v is not a scalar multiple of u, then u, v is linearly independent. Put

$$e = \frac{v - \langle v, u\rangle u}{\|v - \langle v, u\rangle u\|},$$

which makes sense because $v - \langle v, u\rangle u \neq 0$ (otherwise v would be a scalar multiple of u). A direct computation gives $\langle v - \langle v, u\rangle u,\, u\rangle = \langle v, u\rangle - \langle v, u\rangle\|u\|^2 = 0$, so $\langle e, u\rangle = 0$ and $\|e\| = 1$. Thus u, e is an orthonormal list of length two, and the previous paragraph's conclusion gives $\|Te\| = \|Tu\| = c$; moreover $\langle Tu, Te\rangle = 0$ because T preserves orthogonality.
By construction $e \in \operatorname{span}(u, v)$, and rearranging the definition of e gives $v = \langle v, u\rangle u + \|v - \langle v, u\rangle u\|\, e$, so $v \in \operatorname{span}(u, e)$; hence $\operatorname{span}(u, e) = \operatorname{span}(u, v)$. Because u, e is an orthonormal basis of this subspace and v lies in it, we can write $v = au + be$ with $a, b \in \mathbf{F}$, and then $1 = \|v\|^2 = |a|^2 + |b|^2$ by the Pythagorean theorem (6.12). Applying T and using $Tu \perp Te$ together with the Pythagorean theorem again,

$$\|Tv\|^2 = \|a\,Tu + b\,Te\|^2 = |a|^2\|Tu\|^2 + |b|^2\|Te\|^2 = (|a|^2 + |b|^2)c^2 = c^2.$$

Hence $\|Tv\| = c$, which proves the claim. (If $\dim V = 1$ only the first case occurs, so the claim holds there too.)
Now $c > 0$: if $c = 0$ then $Tu = 0$ for every unit vector u, and since every nonzero $v \in V$ equals $\|v\|$ times the unit vector $v/\|v\|$, we would get $Tv = 0$ for all v, contradicting $T \neq 0$.
Define $S = \frac{1}{c}T \in \mathcal{L}(V, W)$. If $v \in V$ and $v \neq 0$, then $v/\|v\|$ is a unit vector, so

$$\|Sv\| = \frac{1}{c}\|Tv\| = \frac{\|v\|}{c}\left\|T\Big(\frac{v}{\|v\|}\Big)\right\| = \frac{\|v\|}{c} \cdot c = \|v\|.$$

Also $\|S0\| = 0 = \|0\|$. Thus S is an isometry and $T = cS$ is a scalar multiple of an isometry, completing the proof.

Problem 7D.3. (a) Show that the product of two unitary operators on V is a unitary operator.
(b) Show that the inverse of a unitary operator on V is a unitary operator.
[This exercise shows that the set of unitary operators on V is a group, where the group operation is the usual product of two operators.]

Solution. Recall (7.51) that an operator $S \in \mathcal{L}(V)$ is unitary if S is an invertible isometry.
(a) Suppose $S, T \in \mathcal{L}(V)$ are unitary. Then S and T are invertible, so their product ST is invertible, with $(ST)^{-1} = T^{-1}S^{-1}$. Furthermore, if $v \in V$ then

$$\|(ST)v\| = \|S(Tv)\| = \|Tv\| = \|v\|,$$

where the second equality holds because S is an isometry (applied to the vector Tv) and the third because T is an isometry. Thus ST is an isometry. Being an invertible isometry, ST is a unitary operator.
Alternatively, one can argue with adjoints: by the equivalence of (a) and (b) in 7.53 we have $S^*S = I$ and $T^*T = I$, so

$$(ST)^*(ST) = T^*S^*ST = T^*IT = T^*T = I,$$

where we used 7.5(d) for the adjoint of a product. Hence ST is unitary, again by 7.53.
(b) Suppose $S \in \mathcal{L}(V)$ is unitary. Then S is invertible, and S^{-1} is invertible (its inverse is S). To see that S^{-1} is an isometry, let $v \in V$ and set $w = S^{-1}v$. Because S is an isometry,

$$\|v\| = \|Sw\| = \|w\| = \|S^{-1}v\|.$$

Thus S^{-1} is an isometry, and being invertible it is a unitary operator.
Alternatively: by the equivalence of (a) and (c) in 7.53, $S^{-1} = S^*$, and by the equivalence of (a) and (f) in 7.53 the operator S^* is unitary. Hence S^{-1} is unitary.
Together with the observations that the identity operator I is unitary (it is invertible and $\|Iv\| = \|v\|$ for all v) and that composition of operators is associative, parts (a) and (b) show that the set of unitary operators on V is a group under composition.

Problem 7D.4. Suppose $\mathbf{F} = \mathbf{C}$ and $A, B \in \mathcal{L}(V)$ are self-adjoint. Show that $A + iB$ is unitary if and only if $AB = BA$ and $A^2 + B^2 = I$.

Solution. Let $T = A + iB$. Because A and B are self-adjoint, $A^* = A$ and $B^* = B$, so 7.5(a) and 7.5(b) give

$$T^* = A^* + \bar{i}\, B^* = A - iB.$$

Expanding the compositions by distributivity, we obtain

$$T^*T = (A - iB)(A + iB) = A^2 + iAB - iBA + B^2 = (A^2 + B^2) + i(AB - BA),$$
$$TT^* = (A + iB)(A - iB) = A^2 - iAB + iBA + B^2 = (A^2 + B^2) - i(AB - BA).$$

(Here we used $(-i)(i) = 1$ and $(i)(-i) = 1$ for the B^2 terms.)
First suppose $AB = BA$ and $A^2 + B^2 = I$. Then $AB - BA = 0$, so the first displayed equation gives

$$T^*T = A^2 + B^2 = I.$$

By the equivalence of (a) and (b) in 7.49, T is an isometry. Because V is finite-dimensional, every isometry from V to V is injective and hence invertible (3.65). Thus T is an invertible isometry, which is to say T is a unitary operator (7.51).
Conversely, suppose $T = A + iB$ is unitary. By the equivalence of (a) and (b) in 7.53,

$$T^*T = TT^* = I.$$

Substituting the two displayed formulas above gives

$$(A^2 + B^2) + i(AB - BA) = I \quad \text{and} \quad (A^2 + B^2) - i(AB - BA) = I.$$

Adding these two equations yields $2(A^2 + B^2) = 2I$, hence

$$A^2 + B^2 = I.$$

Subtracting the second from the first yields $2i(AB - BA) = 0$, hence $AB - BA = 0$, that is,

$$AB = BA.$$

This completes the proof.
Remark: this exercise is the operator analogue of the statement that a complex number $z = a + ib$ (with a, b real) satisfies $|z| = 1$ if and only if $a^2 + b^2 = 1$; the commutativity condition $AB = BA$ is exactly the condition from 7.23 which makes $A + iB$ normal, and the unit-circle condition then upgrades normality to unitarity.

Problem 7D.5. Suppose $S \in \mathcal{L}(V)$. Prove that the following are equivalent.
(a) S is a self-adjoint unitary operator.
(b) $S = 2P - I$ for some orthogonal projection P on V.
(c) There exists a subspace U of V such that $Su = u$ for every $u \in U$ and $Sw = -w$ for every $w \in U^\perp$.

Solution. We prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a).
(a) $\Rightarrow$ (b). Suppose S is a self-adjoint unitary operator. Because S is unitary, $S^*S = I$ by the equivalence of (a) and (b) in 7.53; because S is self-adjoint, $S^* = S$. Hence

$$S^2 = S^*S = I.$$

Define

$$P = \frac{S + I}{2} \in \mathcal{L}(V).$$

Then $P^* = \frac{S^* + I^*}{2} = \frac{S+I}{2} = P$ by 7.5(a), 7.5(b), and 7.5(e), so P is self-adjoint. Also

$$P^2 = \frac{(S + I)^2}{4} = \frac{S^2 + 2S + I}{4} = \frac{I + 2S + I}{4} = \frac{S + I}{2} = P.$$

Let $U = \operatorname{range} P$, a subspace of V. We claim that $P = P_U$, the orthogonal projection of V onto U. To verify this using the definition 6.55, let $v \in V$ and write

$$v = Pv + (v - Pv).$$

Certainly $Pv \in U$. To see that $v - Pv \in U^\perp$, let $u \in U$; then $u = Pw$ for some $w \in V$, and

$$\langle v - Pv, u\rangle = \langle v - Pv, Pw\rangle = \langle P(v - Pv), w\rangle = \langle Pv - P^2v, w\rangle = \langle 0, w\rangle = 0,$$

where the second equality uses $P^* = P$ and the fourth uses $P^2 = P$. Thus $v - Pv \in U^\perp$, and the decomposition displayed above is the decomposition of v coming from $V = U \oplus U^\perp$ (6.49). By the definition of P_U in 6.55, we conclude $P_U v = Pv$ for every $v \in V$, so $P = P_U$ is an orthogonal projection on V. Finally,

$$2P - I = (S + I) - I = S,$$

which proves (b).
(b) $\Rightarrow$ (c). Suppose $S = 2P - I$ where P is an orthogonal projection on V, say $P = P_U$ for some subspace U of V. If $u \in U$, then $P_U u = u$ by 6.57(b), so

$$Su = 2P_U u - u = 2u - u = u.$$

If $w \in U^\perp$, then $P_U w = 0$ by 6.57(c), so

$$Sw = 2P_U w - w = 0 - w = -w.$$

Thus the subspace U has the properties required in (c).
(c) $\Rightarrow$ (a). Suppose U is a subspace of V with $Su = u$ for every $u \in U$ and $Sw = -w$ for every $w \in U^\perp$. By 6.49 we have $V = U \oplus U^\perp$, so each $v \in V$ can be written uniquely as $v = u + w$ with $u \in U$ and $w \in U^\perp$, and then linearity of S gives

$$Sv = Su + Sw = u - w.$$

S is an isometry: since u and $-w$ are orthogonal (as are u and w), the Pythagorean theorem (6.12) gives

$$\|Sv\|^2 = \|u - w\|^2 = \|u\|^2 + \|w\|^2 = \|u + w\|^2 = \|v\|^2,$$

so $\|Sv\| = \|v\|$ for every $v \in V$.
S is invertible: note that $u \in U$ and $-w \in U^\perp$, so applying the formula for S to $Sv = u - w$ gives

$$S^2 v = S(u - w) = u - (-w) = u + w = v.$$

Thus $S^2 = I$, so S is invertible with $S^{-1} = S$. Being an invertible isometry, S is a unitary operator (7.51).
S is self-adjoint: let $v, v' \in V$ and write $v = u + w$, $v' = u' + w'$ with $u, u' \in U$ and $w, w' \in U^\perp$. Because $\langle u, w' \rangle = \langle w, u' \rangle = 0$,

$$\langle Sv, v' \rangle = \langle u - w,\ u' + w' \rangle = \langle u, u' \rangle - \langle w, w' \rangle$$

and

$$\langle v, Sv' \rangle = \langle u + w,\ u' - w' \rangle = \langle u, u' \rangle - \langle w, w' \rangle.$$

Hence $\langle Sv, v' \rangle = \langle v, Sv' \rangle$ for all $v, v' \in V$, which by the definition of the adjoint (7.1) means $S^* = S$.
Thus S is a self-adjoint unitary operator, proving (a) and completing the cycle of implications.

Problem 7D.6. Suppose T_1, T_2 are both normal operators on $\mathbf{F}^3$ with $2, 5, 7$ as eigenvalues. Prove that there exists a unitary operator $S \in \mathcal{L}(\mathbf{F}^3)$ such that $T_1 = S^* T_2 S$.

Solution. Throughout, $\mathbf{F}^3$ carries the Euclidean inner product, so $\dim \mathbf{F}^3 = 3$.
Step 1: each of T_1, T_2 has an orthonormal basis of eigenvectors, with eigenvalues in the order $2, 5, 7$.
Let T denote either T_1 or T_2. By hypothesis T is normal and $2, 5, 7$ are eigenvalues of T. Choose eigenvectors v_1, v_2, v_3 of T corresponding to $2, 5, 7$ respectively, and normalize them: setting $g_k = v_k / \|v_k\|$ (legitimate since eigenvectors are nonzero) we get $\|g_k\| = 1$ and $T g_k = \lambda_k g_k$, where

$$\lambda_1 = 2, \qquad \lambda_2 = 5, \qquad \lambda_3 = 7.$$

Because T is normal and $\lambda_1, \lambda_2, \lambda_3$ are distinct, 7.22 (orthogonal eigenvectors for normal operators) shows that g_1, g_2, g_3 are pairwise orthogonal. Hence g_1, g_2, g_3 is an orthonormal list in $\mathbf{F}^3$ of length $3 = \dim \mathbf{F}^3$, and therefore an orthonormal basis of $\mathbf{F}^3$ by 6.28.
Note that this argument works whether $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$: we never invoked a spectral theorem, only 7.22 together with the fact that three distinct eigenvalues in a three-dimensional space force the corresponding eigenvectors to span.
Applying this to T_1 and to T_2, we obtain orthonormal bases e_1, e_2, e_3 and f_1, f_2, f_3 of $\mathbf{F}^3$ with

$$T_1 e_k = \lambda_k e_k \quad \text{and} \quad T_2 f_k = \lambda_k f_k \qquad \text{for } k = 1, 2, 3.$$

Step 2: construct S.
Because e_1, e_2, e_3 is a basis of $\mathbf{F}^3$, there is a unique $S \in \mathcal{L}(\mathbf{F}^3)$ with

$$S e_k = f_k \qquad \text{for } k = 1, 2, 3$$

(by 3.4, the linear map lemma). Then S maps the orthonormal basis e_1, e_2, e_3 to the orthonormal basis f_1, f_2, f_3, so the equivalence of (d) and (a) in 7.53 shows that S is a unitary operator. In particular $S^* = S^{-1}$ by the equivalence of (a) and (c) in 7.53, and $S^{-1} f_k = e_k$.

Step 3: verify $T_1 = S^* T_2 S$.
For each $k \in \{1, 2, 3\}$,

$$S^* T_2 S e_k = S^* T_2 f_k = S^*(\lambda_k f_k) = \lambda_k S^* f_k = \lambda_k S^{-1} f_k = \lambda_k e_k = T_1 e_k.$$

Thus the linear maps $S^* T_2 S$ and T_1 agree on the basis e_1, e_2, e_3 of $\mathbf{F}^3$, and hence they are equal:

$$T_1 = S^* T_2 S,$$

with S unitary, as desired.

Problem 7D.7. Give an example of two self-adjoint operators $T_1, T_2 \in \mathcal{L}(\mathbf{F}^4)$ such that the eigenvalues of both operators are $2, 5, 7$ but there does not exist a unitary operator $S \in \mathcal{L}(\mathbf{F}^4)$ such that $T_1 = S^* T_2 S$. Be sure to explain why there is no unitary operator with the required property.

Solution. Let e_1, e_2, e_3, e_4 denote the standard basis of $\mathbf{F}^4$, which is an orthonormal basis with respect to the Euclidean inner product. Define $T_1, T_2 \in \mathcal{L}(\mathbf{F}^4)$ by

$$T_1(z_1, z_2, z_3, z_4) = (2z_1,\, 2z_2,\, 5z_3,\, 7z_4)$$

and

$$T_2(z_1, z_2, z_3, z_4) = (2z_1,\, 5z_2,\, 5z_3,\, 7z_4).$$

Equivalently, with respect to e_1, e_2, e_3, e_4 the matrices of T_1 and T_2 are the diagonal matrices with diagonals $2, 2, 5, 7$ and $2, 5, 5, 7$ respectively.
Both operators are self-adjoint. Indeed, the matrix of T_j with respect to the orthonormal basis e_1, e_2, e_3, e_4 is a diagonal matrix with real entries, hence equals its own conjugate transpose; by 7.9 (the matrix of T^* is the conjugate transpose of the matrix of T, computed with respect to orthonormal bases) this says $\mathcal{M}(T_j) = \mathcal{M}(T_j^*)$, and therefore $T_j = T_j^*$.
The eigenvalues of both operators are exactly $2, 5, 7$. For T_1 we have $T_1 e_1 = 2e_1$, $T_1 e_3 = 5e_3$, and $T_1 e_4 = 7e_4$, so $2, 5, 7$ are eigenvalues of T_1; similarly $T_2 e_1 = 2e_1$, $T_2 e_2 = 5e_2$, and $T_2 e_4 = 7e_4$. Conversely, if $\lambda \in \mathbf{F}$ with $\lambda \notin \{2, 5, 7\}$, then $T_j - \lambda I$ is diagonal with respect to e_1, e_2, e_3, e_4 with all diagonal entries nonzero, so it maps each e_k to a nonzero multiple of e_k; hence $\operatorname{range}(T_j - \lambda I) = \mathbf{F}^4$, so $T_j - \lambda I$ is surjective and thus invertible (3.65), so λ is not an eigenvalue of T_j. Thus the eigenvalues of T_1 are $2, 5, 7$ and the eigenvalues of T_2 are $2, 5, 7$.
Now we explain why no unitary $S \in \mathcal{L}(\mathbf{F}^4)$ satisfies $T_1 = S^* T_2 S$.
First compute the eigenspaces for the eigenvalue 2. Since $T_1 e_1 = 2e_1$ and $T_1 e_2 = 2e_2$, we have $\operatorname{span}(e_1, e_2) \subseteq \operatorname{null}(T_1 - 2I)$; and if $(z_1, z_2, z_3, z_4) \in \operatorname{null}(T_1 - 2I)$ then $3z_3 = 0$ and $5z_4 = 0$, so $z_3 = z_4 = 0$. Hence

$$\operatorname{null}(T_1 - 2I) = \operatorname{span}(e_1, e_2), \qquad \text{so} \qquad \dim \operatorname{null}(T_1 - 2I) = 2.$$

Similarly, if $(z_1, z_2, z_3, z_4) \in \operatorname{null}(T_2 - 2I)$ then $3z_2 = 3z_3 = 5z_4 = 0$, so

$$\operatorname{null}(T_2 - 2I) = \operatorname{span}(e_1), \qquad \text{so} \qquad \dim \operatorname{null}(T_2 - 2I) = 1.$$

Suppose, for the sake of contradiction, that $S \in \mathcal{L}(\mathbf{F}^4)$ is unitary and $T_1 = S^* T_2 S$. Because S is unitary, $S^* = S^{-1}$ by the equivalence of (a) and (c) in 7.53. Hence

$$T_1 - 2I = S^{-1} T_2 S - 2S^{-1} S = S^{-1}(T_2 - 2I) S.$$

Now let $v \in \operatorname{null}(T_1 - 2I)$. Then $S^{-1}(T_2 - 2I) S v = 0$, and applying S to both sides gives $(T_2 - 2I) S v = 0$, that is, $Sv \in \operatorname{null}(T_2 - 2I)$. Thus S restricts to a linear map

$$S|_{\operatorname{null}(T_1 - 2I)} : \operatorname{null}(T_1 - 2I) \longrightarrow \operatorname{null}(T_2 - 2I),$$

and this restriction is injective because S is injective. Injectivity forces

$$\dim \operatorname{null}(T_1 - 2I) \leq \dim \operatorname{null}(T_2 - 2I)$$

because an injective linear map has null space $\{0\}$, so the fundamental theorem of linear maps (3.21) applied to the restriction gives $\dim \operatorname{null}(T_1 - 2I) = \dim \operatorname{range}\big(S|_{\operatorname{null}(T_1-2I)}\big) \le \dim \operatorname{null}(T_2 - 2I)$. But we computed these dimensions to be 2 and 1, and $2 \le 1$ is false. This contradiction shows that no such unitary S exists.
Remark: the point is that the list of eigenvalues alone does not determine a self-adjoint operator up to unitary equivalence once the dimension exceeds the number of distinct eigenvalues; the multiplicities $\dim \operatorname{null}(T-\lambda I)$ must also match. In Exercise 7D.6 the space had dimension 3 and there were 3 distinct eigenvalues, so each multiplicity was forced to equal 1 and no such obstruction could arise.

Problem 7D.8. Prove or give a counterexample: If $S \in \mathcal{L}(V)$ and there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $\|Se_k\| = 1$ for each e_k, then S is a unitary operator.

Solution. The statement is false.
Counterexample. Let $V = \mathbf{F}^2$ with the Euclidean inner product, let e_1, e_2 be the standard basis (which is orthonormal), and define $S \in \mathcal{L}(\mathbf{F}^2)$ by

$$S(x_1, x_2) = (x_1 + x_2,\, 0).$$

Then $Se_1 = (1, 0)$ and $Se_2 = (1, 0)$, so

$$\|Se_1\| = \|Se_2\| = 1,$$

as required by the hypothesis.
However S is not a unitary operator. Indeed $S(1, -1) = (0, 0)$ with $(1, -1) \neq 0$, so S is not injective and hence not invertible; a unitary operator is by definition an invertible isometry (7.51). Alternatively, $\|S(1, -1)\| = 0 \neq \sqrt{2} = \|(1, -1)\|$, so S is not an isometry.
What goes wrong is instructive. By the equivalence of (a) and (d) in 7.53, S is unitary exactly when $Se_1, \dots, Se_n$ is an orthonormal basis of V. The hypothesis $\|Se_k\| = 1$ supplies only the normalization half of that condition; it says nothing about orthogonality of the vectors $Se_1, \dots, Se_n$. In the counterexample above $Se_1 = Se_2$, so the list Se_1, Se_2 is as far from orthogonal as possible.
More generally, for any $n \ge 2$ and any unit vector $u \in V$, the operator defined on an orthonormal basis $e_1, \dots, e_n$ by $Se_k = u$ for every k satisfies $\|Se_k\| = 1$ for each k while having a one-dimensional range, hence is not invertible when $n \ge 2$. (For $n = 1$ the statement is true, since then $\|Se_1\| = 1 = \|e_1\|$ forces S to be an isometry on the one-dimensional space V.)

Problem 7D.9. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Suppose every eigenvalue of T has absolute value 1 and $\|Tv\| \le \|v\|$ for every $v \in V$. Prove that T is a unitary operator.

Solution. Because $\mathbf{F} = \mathbf{C}$ and V is a finite-dimensional complex inner product space, Schur's theorem (6.38) provides an orthonormal basis $e_1, \dots, e_n$ of V with respect to which T has an upper-triangular matrix. Write

$$A = \mathcal{M}\big(T, (e_1, \dots, e_n)\big),$$

so $A_{j,k} = 0$ whenever $j > k$. By 5.41 the eigenvalues of T are precisely the diagonal entries $A_{1,1}, \dots, A_{n,n}$ of A. Hence, by hypothesis,

$$|A_{k,k}| = 1 \qquad \text{for each } k = 1, \dots, n.$$

Now fix $k \in \{1, \dots, n\}$. Since A is upper triangular, the k-th column of A gives

$$Te_k = A_{1,k}e_1 + \cdots + A_{k,k}e_k.$$

Because $e_1, \dots, e_n$ is orthonormal, the Pythagorean-type formula 6.30(b) yields

$$\|Te_k\|^2 = \sum_{j=1}^{k} |A_{j,k}|^2 = |A_{k,k}|^2 + \sum_{j=1}^{k-1} |A_{j,k}|^2 = 1 + \sum_{j=1}^{k-1} |A_{j,k}|^2.$$

On the other hand the hypothesis $\|Tv\| \le \|v\|$ applied to $v = e_k$ gives

$$\|Te_k\|^2 \le \|e_k\|^2 = 1.$$

Combining the two displays,

$$1 + \sum_{j=1}^{k-1} |A_{j,k}|^2 \le 1,$$

so $\sum_{j=1}^{k-1} |A_{j,k}|^2 \le 0$. A sum of nonnegative real numbers that is at most 0 must have every term equal to 0; hence $A_{j,k} = 0$ for all $j < k$.
Since k was arbitrary, all entries of A off the diagonal above the diagonal vanish, and the entries below the diagonal already vanish because A is upper triangular. Thus A is a diagonal matrix, which means

$$Te_k = \lambda_k e_k \quad \text{where } \lambda_k = A_{k,k}, \qquad |\lambda_k| = 1,$$

for each $k = 1, \dots, n$.
So $e_1, \dots, e_n$ is an orthonormal basis of V consisting of eigenvectors of T whose corresponding eigenvalues all have absolute value 1. By the equivalence of (a) and (b) in 7.55, T is a unitary operator.
(Alternatively, one can finish without invoking 7.55: for $j \ne k$ we have $\langle Te_j, Te_k\rangle = \lambda_j \overline{\lambda_k} \langle e_j, e_k\rangle = 0$, and $\|Te_k\| = |\lambda_k| = 1$, so $Te_1, \dots, Te_n$ is an orthonormal list of length $n = \dim V$, hence an orthonormal basis of V; now apply the equivalence of (a) and (d) in 7.53.)

Problem 7D.10. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is a self-adjoint operator such that $\|Tv\| \le \|v\|$ for all $v \in V$.
(a) Show that $I - T^2$ is a positive operator.
(b) Show that $T + i\sqrt{I - T^2}$ is a unitary operator.

Solution. (a) First, $I - T^2$ is self-adjoint: the identity is self-adjoint, and

$$(T^2)^* = (T^*)^2 = T^2$$

because $T^* = T$ (here $(ST)^* = T^*S^*$ is 7.5(d)). Hence, writing $I - T^2 = I + (-1)T^2$ and using 7.5(a), 7.5(b), and 7.5(e),

$$(I - T^2)^* = I^* + \overline{(-1)}\,(T^2)^* = I - T^2.$$

Second, for every $v \in V$,

$$\big\langle (I - T^2)v, v\big\rangle = \langle v, v\rangle - \langle T(Tv), v\rangle = \|v\|^2 - \langle Tv, T^*v\rangle = \|v\|^2 - \langle Tv, Tv\rangle = \|v\|^2 - \|Tv\|^2 \ge 0,$$

where we used $T^* = T$ in the third equality and the hypothesis $\|Tv\| \le \|v\|$ in the last step.
Being self-adjoint with $\langle (I - T^2)v, v\rangle \ge 0$ for all $v \in V$, the operator $I - T^2$ is positive by the definition 7.34.
(b) By (a) the operator $I - T^2$ is positive, so by 7.39 it has a unique positive square root, denoted $\sqrt{I - T^2}$. Write $R = \sqrt{I - T^2}$; thus R is positive and $R^2 = I - T^2$.
We first identify R concretely. Since $\mathbf{F} = \mathbf{C}$ and T is self-adjoint, T is normal, so the complex spectral theorem (7.31) gives an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T, say

$$Te_k = \lambda_k e_k \qquad (k = 1, \dots, n).$$

Each λ_k is real, because eigenvalues of a self-adjoint operator are real (7.12). Moreover

$$|\lambda_k| = |\lambda_k|\,\|e_k\| = \|\lambda_k e_k\| = \|Te_k\| \le \|e_k\| = 1,$$

so $\lambda_k \in [-1, 1]$ and therefore $1 - \lambda_k^2 \ge 0$ for each k.
Define $R' \in \mathcal{L}(V)$ by

$$R'e_k = \sqrt{1 - \lambda_k^2}\, e_k \qquad (k = 1, \dots, n),$$

extended linearly. With respect to the orthonormal basis $e_1, \dots, e_n$ the matrix of R' is diagonal with real entries, hence equals its own conjugate transpose, so R' is self-adjoint by 7.9 (which says that with respect to an orthonormal basis the matrix of the adjoint is the conjugate transpose of the matrix of the operator); and for $v = \sum_k a_k e_k$ we get

$$\langle R'v, v\rangle = \sum_{k=1}^{n} \sqrt{1-\lambda_k^2}\,|a_k|^2 \geq 0.$$

Hence R' is a positive operator. Also

$$(R')^2 e_k = (1-\lambda_k^2)e_k = e_k - \lambda_k^2 e_k = e_k - T^2 e_k = (I-T^2)e_k$$

for each k, so $(R')^2 = I - T^2$. Thus R' is a positive square root of $I - T^2$, and the uniqueness statement in 7.39 forces

$$R = \sqrt{I-T^2} = R'.$$

Now set $S = T + iR$. For each k,

$$Se_k = Te_k + iRe_k = \Big(\lambda_k + i\sqrt{1-\lambda_k^2}\Big)e_k,$$

so e_k is an eigenvector of S with eigenvalue $\mu_k = \lambda_k + i\sqrt{1-\lambda_k^2}$. Since λ_k and $\sqrt{1-\lambda_k^2}$ are real,

$$|\mu_k|^2 = \lambda_k^2 + \big(1-\lambda_k^2\big) = 1,$$

so $|\mu_k| = 1$.
Therefore $e_1, \dots, e_n$ is an orthonormal basis of V consisting of eigenvectors of S whose eigenvalues all have absolute value 1. By the equivalence of (b) and (a) in 7.55, $S = T + i\sqrt{I-T^2}$ is a unitary operator.
(Remark: the argument also shows directly that $S^*S = I$. Indeed $S^* = T^* + \overline{i}\,R^* = T - iR$ by 7.5(a), 7.5(b) and the self-adjointness of T and R; moreover R commutes with T, since both are diagonal with respect to $e_1, \dots, e_n$. Hence

$$S^*S = (T-iR)(T+iR) = T^2 + R^2 + i(TR - RT) = T^2 + (I-T^2) = I,$$

and similarly $SS^* = I$, so S is unitary by the equivalence of (a) and (b) in 7.53. This is Exercise 4 of this section with $A = T$ and $B = R$.)

Problem 7D.11. Suppose $S \in \mathcal{L}(V)$. Prove that S is a unitary operator if and only if

$$\{Sv : v \in V \text{ and } \|v\| \leq 1\} = \{v \in V : \|v\| \leq 1\}.$$

Solution. Write $B = \{v \in V : \|v\| \leq 1\}$ for the closed unit ball of V, so that the displayed condition reads $S(B) = B$.
First suppose S is a unitary operator. If $v \in B$ then $\|Sv\| = \|v\| \leq 1$, so $Sv \in B$; this proves $S(B) \subseteq B$. For the reverse inclusion, let $w \in B$. Since S is unitary it is invertible, and $S^{-1} = S^*$ by the equivalence of (a) and (c) in 7.53; moreover S^* is a unitary operator by the equivalence of (a) and (f) in 7.53. Hence S^{-1} is an isometry, so

$$\|S^{-1}w\| = \|w\| \leq 1,$$

which says $S^{-1}w \in B$. Since $w = S(S^{-1}w)$, we get $w \in S(B)$. Thus $B \subseteq S(B)$, and therefore $S(B) = B$.
Conversely, suppose $S(B) = B$. We prove S is an invertible isometry.
Step 1: S is invertible. Every $v \in V$ with $v \neq 0$ satisfies $v/\|v\| \in B = S(B) \subseteq \operatorname{range} S$, hence $v = \|v\| \cdot \big(v/\|v\|\big) \in \operatorname{range} S$ because the range is a subspace. Also $0 \in \operatorname{range} S$. Thus $\operatorname{range} S = V$, so S is surjective. Since V is finite-dimensional, a surjective operator on V is invertible (3.65).

Step 2: $\|Sv\| \le \|v\|$ for all $v \in V$. This is clear for $v = 0$. If $v \neq 0$, then $v/\|v\| \in B$, so $S(v/\|v\|) \in S(B) = B$, giving

$$\Big\|S\big(v/\|v\|\big)\Big\| \le 1, \quad \text{i.e.} \quad \frac{\|Sv\|}{\|v\|} \le 1,$$

by homogeneity of the norm and linearity of S. Hence $\|Sv\| \le \|v\|$.
Step 3: $\|S^{-1}w\| \le \|w\|$ for all $w \in V$. Since S is a bijection of V and $S(B) = B$, applying S^{-1} to both sides gives $B = S^{-1}(S(B)) = S^{-1}(B)$. So $S^{-1}(B) = B$, which means $\|S^{-1}w\| \le 1$ whenever $\|w\| \le 1$. Exactly the homogeneity argument of Step 2, now applied to S^{-1}, yields $\|S^{-1}w\| \le \|w\|$ for all $w \in V$.
Step 4: S is an isometry. Let $v \in V$ and apply Step 3 to $w = Sv$:

$$\|v\| = \|S^{-1}(Sv)\| \le \|Sv\|.$$

Combining with Step 2 gives $\|v\| \le \|Sv\| \le \|v\|$, so $\|Sv\| = \|v\|$.
Thus S is an invertible isometry, that is, a unitary operator (7.51). (In fact Step 1 was not strictly needed for invertibility once we know S is an isometry, since every isometry is injective and every injective operator on a finite-dimensional space is invertible; but Step 1 is what makes Step 3, and hence Step 4, available.)

Problem 7D.12. Prove or give a counterexample: If $S \in \mathcal{L}(V)$ is invertible and $\|S^{-1}v\| = \|Sv\|$ for every $v \in V$, then S is unitary.

Solution. The statement is false.
Counterexample. Let $V = \mathbf{F}^2$ with the Euclidean inner product and define $S \in \mathcal{L}(\mathbf{F}^2)$ by

$$S(x_1, x_2) = \Big(2x_2,\, \tfrac{1}{2}x_1\Big),$$

that is, S has matrix

$$\mathcal{M}(S) = \begin{pmatrix} 0 & 2 \\ \frac{1}{2} & 0 \end{pmatrix}$$

with respect to the standard basis.
First, S is invertible and $S^{-1} = S$. Indeed, for all $(x_1, x_2) \in \mathbf{F}^2$,

$$S\big(S(x_1, x_2)\big) = S\Big(2x_2, \tfrac{1}{2}x_1\Big) = \Big(2 \cdot \tfrac{1}{2}x_1,\, \tfrac{1}{2} \cdot 2x_2\Big) = (x_1, x_2),$$

so $S^2 = I$. Hence S is invertible with $S^{-1} = S$.
Consequently

$$\|S^{-1}v\| = \|Sv\| \qquad \text{for every } v \in V,$$

trivially, because S^{-1} and S are the same operator. So S satisfies the hypothesis.
But S is not unitary: taking $v = (1, 0)$ we get $Sv = (0, \frac{1}{2})$, so

$$\|Sv\| = \tfrac{1}{2} \neq 1 = \|v\|,$$

and hence S is not an isometry. Since a unitary operator is by definition an invertible isometry (7.51), S is not unitary. (Equivalently, the columns $(0, \frac{1}{2})$ and $(2, 0)$ of $\mathcal{M}(S)$ are orthogonal but not of norm 1, so S fails the criterion in 7.49(e).)
Remark. The failure is not an accident of this one example. Any invertible S with $S^2 = I$ satisfies the hypothesis vacuously, and there are many such non-unitary operators: for each scalar $c \neq 0$ the operator $S_c(x_1, x_2) = (cx_2,\, c^{-1}x_1)$ satisfies $S_c^2 = I$, yet $\|S_c(1, 0)\| = |c|^{-1}$, so S_c is unitary only when $|c| = 1$. The hypothesis $\|S^{-1}v\| = \|Sv\|$ for all v says (using 7.16, since $S^*S - (S^{-1})^*S^{-1}$ is self-adjoint) exactly that $S^*S = (S^{-1})^*S^{-1}$, which is a genuine restriction on S but a far weaker one than $S^*S = I$.

Problem 7D.13. Explain why the columns of a square matrix of complex numbers form an orthonormal list in $\mathbf{C}^n$ if and only if the rows of the matrix form an orthonormal list in $\mathbf{C}^n$.

Solution. Let Q be an n-by-n matrix of complex numbers, and equip $\mathbf{C}^n$ with the Euclidean inner product

$$\langle u, w\rangle = u_1\overline{w_1} + \cdots + u_n\overline{w_n}.$$

Let Q^* denote the conjugate transpose of Q, so $(Q^*)_{j,k} = \overline{Q_{k,j}}$.
Translate both conditions into matrix equations.
Columns. For $k, r \in \{1, \ldots, n\}$, the inner product in $\mathbf{C}^n$ of column k of Q with column r of Q is

$$\sum_{j=1}^{n} Q_{j,k}\,\overline{Q_{j,r}} = \sum_{j=1}^{n} (Q^*)_{r,j}\,Q_{j,k} = (Q^*Q)_{r,k}.$$

Hence the columns of Q form an orthonormal list in $\mathbf{C}^n$ if and only if $(Q^*Q)_{r,k}$ equals 1 when $r = k$ and 0 when $r \neq k$, that is, if and only if

$$Q^*Q = I.$$

Rows. Similarly, for $j, k \in \{1, \ldots, n\}$, the inner product of row j of Q with row k of Q is

$$\sum_{r=1}^{n} Q_{j,r}\,\overline{Q_{k,r}} = \sum_{r=1}^{n} Q_{j,r}\,(Q^*)_{r,k} = (QQ^*)_{j,k}.$$

Hence the rows of Q form an orthonormal list in $\mathbf{C}^n$ if and only if

$$QQ^* = I.$$

So it suffices to explain why $Q^*Q = I$ if and only if $QQ^* = I$, for square Q. Suppose $Q^*Q = I$. Then Q^* is a left inverse of the square matrix Q. Identifying Q with the operator $S \in \mathcal{L}(\mathbf{C}^n)$ whose matrix with respect to the standard basis is Q, the equation $Q^*Q = I$ says $S^*S = I$, so S is injective (if $Sv = 0$ then $v = S^*Sv = 0$). Because $\mathbf{C}^n$ is finite-dimensional, an injective operator on it is invertible (3.65). Multiplying $S^*S = I$ on the right by S^{-1} gives $S^* = S^{-1}$, hence $SS^* = I$, that is, $QQ^* = I$. The reverse implication follows by applying what was just proved to Q^* in place of Q, since $(Q^*)^* = Q$.
Therefore

$$\text{columns of } Q \text{ orthonormal} \iff Q^*Q = I \iff QQ^* = I \iff \text{rows of } Q \text{ orthonormal},$$

which is what we were asked to explain.
Equivalently, in the language of the section: by the equivalence of (a) and (e) in 7.49, the columns of Q form an orthonormal list in $\mathbf{C}^n$ exactly when the associated operator S on $\mathbf{C}^n$ is an isometry, hence (since $\mathbf{C}^n$ is finite-dimensional) exactly when S is a unitary operator; and by the equivalence of (a) and (e) in 7.53, S is a unitary operator exactly when the rows of Q form an orthonormal basis of $\mathbf{C}^n$. This is precisely the remark, made after 7.56, that "columns" may be replaced by "rows" in the definition of a unitary matrix.
Note that nothing in the argument used that the entries are complex rather than real; the same statement holds over $\mathbf{R}$, with Q^* the transpose. Note also that squareness is essential: the 2-by-1 matrix with entries 1 and 0 has an orthonormal list of columns but its rows (the lists 1 and 0, viewed in $\mathbf{C}^1$) do not form an orthonormal list.

Problem 7D.14. Suppose $v \in V$ with $\|v\| = 1$ and $b \in \mathbf{F}$. Also suppose $\dim V \geq 2$. Prove that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $\langle Sv, v\rangle = b$ if and only if $|b| \leq 1$.

Solution. Let $n = \dim V$, so $n \geq 2$.
First suppose there exists a unitary operator $S \in \mathcal{L}(V)$ with $\langle Sv, v\rangle = b$. Since S is an isometry,

$\|Sv\| = \|v\| = 1$. The Cauchy–Schwarz inequality (6.14) then gives

$$|b| = |\langle Sv, v\rangle| \le \|Sv\|\,\|v\| = 1 \cdot 1 = 1.$$

Thus $|b| \le 1$.
Conversely, suppose $b \in \mathbf{F}$ with $|b| \le 1$. Since $\|v\| = 1$, the list consisting of the single vector v is orthonormal, so by 6.36 it extends to an orthonormal basis

$$e_1, e_2, \dots, e_n \quad \text{of } V, \qquad \text{with } e_1 = v.$$

Here we use the hypothesis $\dim V \ge 2$, which guarantees that a second basis vector e_2 is present.
Because $|b| \le 1$, the real number $1 - |b|^2$ is nonnegative; set

$$c = \sqrt{1 - |b|^2} \in \mathbf{R}, \qquad c \ge 0, \qquad |b|^2 + c^2 = 1.$$

Note that $c \in \mathbf{F}$ in either case $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$, and that $\overline{b} \in \mathbf{F}$ as well (if $\mathbf{F} = \mathbf{R}$ then $\overline{b} = b$).
Define $S \in \mathcal{L}(V)$ to be the unique linear map satisfying

$$Se_1 = b\,e_1 + c\,e_2, \qquad Se_2 = -c\,e_1 + \overline{b}\,e_2, \qquad Se_k = e_k \ \text{ for } k = 3, \dots, n.$$

(Such an S exists and is unique because a linear map may be defined arbitrarily on a basis; see 3.4.)
We claim $Se_1, \dots, Se_n$ is an orthonormal basis of V. Compute, using that $e_1, \dots, e_n$ is orthonormal:

$$\|Se_1\|^2 = |b|^2 + c^2 = 1, \qquad \|Se_2\|^2 = |-c|^2 + |\overline{b}|^2 = c^2 + |b|^2 = 1,$$

and

$$\langle Se_1, Se_2\rangle = \langle be_1 + ce_2,\ -ce_1 + \overline{b}e_2\rangle = b\,\overline{(-c)} + c\,\overline{\overline{b}} = -bc + cb = 0,$$

where we used that c is real, so $\overline{c} = c$, and that $\overline{\overline{b}} = b$. For $k \ge 3$ we have $\|Se_k\| = \|e_k\| = 1$, and

$$\langle Se_1, Se_k\rangle = \langle be_1 + ce_2, e_k\rangle = 0, \qquad \langle Se_2, Se_k\rangle = \langle -ce_1 + \overline{b}e_2, e_k\rangle = 0,$$

since e_k is orthogonal to e_1 and to e_2; and for $3 \le j < k$ we have $\langle Se_j, Se_k\rangle = \langle e_j, e_k\rangle = 0$. Hence $Se_1, \dots, Se_n$ is an orthonormal list of length $n = \dim V$ in V, and therefore an orthonormal basis of V (an orthonormal list is linearly independent by 6.25, and a linearly independent list of length $\dim V$ is a basis by 2.38).
By the equivalence of (d) and (a) in 7.53, S is a unitary operator.
Finally,

$$\langle Sv, v\rangle = \langle Se_1, e_1\rangle = \langle be_1 + ce_2,\ e_1\rangle = b\langle e_1, e_1\rangle + c\langle e_2, e_1\rangle = b \cdot 1 + c \cdot 0 = b.$$

Thus a unitary S with $\langle Sv, v\rangle = b$ exists precisely when $|b| \le 1$.
Remark. The hypothesis $\dim V \ge 2$ cannot be dropped. If $\dim V = 1$ and $\|v\| = 1$, then $Sv = \lambda v$ for some $\lambda \in \mathbf{F}$, and S is unitary if and only if $|\lambda| = 1$; in that case $\langle Sv, v\rangle = \lambda$, so the attainable values of b are exactly those with $|b| = 1$, not all b with $|b| \le 1$.

Problem 7D.15. Suppose T is a unitary operator on V such that $T - I$ is invertible.
(a) Prove that $(T+I)(T-I)^{-1}$ is a skew operator (meaning that it equals the negative of its adjoint).
(b) Prove that if $\mathbb{F} = \mathbb{C}$, then $i(T + I)(T - I)^{-1}$ is a self-adjoint operator.
[The function $z \mapsto i(z + 1)(z - 1)^{-1}$ maps the unit circle in $\mathbb{C}$ (except for the point 1) to $\mathbb{R}$. Thus (b) illustrates the analogy between the unitary operators and the unit circle in $\mathbb{C}$, along with the analogy between the self-adjoint operators and $\mathbb{R}$.]

Solution. Throughout, write $S = (T + I)(T - I)^{-1}$. We use the standard algebraic properties of the adjoint (7.5): $(A + B)^* = A^* + B^*$, $(AB)^* = B^*A^*$, $I^* = I$, and $(\lambda A)^* = \overline{\lambda}A^*$. We also use the fact that if A is invertible then A^* is invertible with

$$(A^*)^{-1} = (A^{-1})^*,$$

which holds because $(A^{-1})^*A^* = (AA^{-1})^* = I^* = I$ and likewise $A^*(A^{-1})^* = (A^{-1}A)^* = I$.
Step 1: $T + I$ and $(T - I)^{-1}$ commute.
Expanding both sides gives

$$(T + I)(T - I) = T^2 - I = (T - I)(T + I).$$

Multiplying this equation on the left by $(T - I)^{-1}$ and on the right by $(T - I)^{-1}$ yields

$$(T - I)^{-1}(T + I) = (T + I)(T - I)^{-1}.$$

So $S = (T + I)(T - I)^{-1} = (T - I)^{-1}(T + I)$.
Step 2: computation of S^*.
Because T is unitary, T is invertible and $T^* = T^{-1}$, by the equivalence of (a) and (c) in 7.53. Using this,

$$T^* - I = T^{-1} - I = T^{-1}(I - T) = -T^{-1}(T - I),$$

and

$$T^* + I = T^{-1} + I = T^{-1}(I + T) = T^{-1}(T + I).$$

In particular $T^* - I$ is invertible, being the product of the invertible operators $-T^{-1}$ and $T - I$, and

$$(T^* - I)^{-1} = \big(-T^{-1}(T - I)\big)^{-1} = -(T - I)^{-1}T.$$

Now compute, using $(A^*)^{-1} = (A^{-1})^*$ with $A = T - I$:

$$\begin{aligned} S^* &= \big((T + I)(T - I)^{-1}\big)^* \\ &= \big((T - I)^{-1}\big)^*(T + I)^* \\ &= \big((T - I)^*\big)^{-1}(T^* + I) \\ &= (T^* - I)^{-1}(T^* + I) \\ &= \big(-(T - I)^{-1}T\big)\big(T^{-1}(T + I)\big) \\ &= -(T - I)^{-1}(T + I). \end{aligned}$$

By Step 1 the last expression equals $-(T+I)(T-I)^{-1} = -S$. Hence $S^* = -S$, so S is a skew operator, proving (a).
(b) Now suppose $\mathbb{F} = \mathbb{C}$, and set $A = iS = i(T+I)(T-I)^{-1}$. Using $(\lambda B)^* = \overline{\lambda}B^*$ with $\lambda = i$ and part (a),

$$A^* = (iS)^* = \bar{i}\, S^* = (-i)(-S) = iS = A.$$

Thus A is self-adjoint, proving (b).
Remark (consistency with the analogy described in the book, still in the case $\mathbb{F} = \mathbb{C}$). By 7.55 there is an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of the unitary operator T, say $Te_k = \lambda_k e_k$, with $|\lambda_k| = 1$ by 7.54; since $T - I$ is invertible, no λ_k equals 1. Each e_k is then an eigenvector of A: from $(T - I)e_k = (\lambda_k - 1)e_k$ we get $(T - I)^{-1}e_k = (\lambda_k - 1)^{-1}e_k$, hence

$$Ae_k = i\,\frac{\lambda_k + 1}{\lambda_k - 1}\,e_k.$$

Writing $\lambda_k = e^{i\theta}$ with θ not an integer multiple of 2π, we have $\lambda_k + 1 = 2e^{i\theta/2}\cos(\theta/2)$ and $\lambda_k - 1 = 2ie^{i\theta/2}\sin(\theta/2)$, so

$$i\,\frac{\lambda_k + 1}{\lambda_k - 1} = i \cdot \frac{\cos(\theta/2)}{i\sin(\theta/2)} = \cot(\theta/2),$$

a real number. So the eigenvalues of A are exactly the images of the eigenvalues of T under the map $z \mapsto i(z + 1)(z - 1)^{-1}$, and they are real, as they must be for a self-adjoint operator by 7.12.

Problem 7D.16. Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$ is self-adjoint. Prove that $(T + iI)(T - iI)^{-1}$ is a unitary operator and 1 is not an eigenvalue of this operator.

Solution. The key computation is the following identity. Let $v \in V$. Since T is self-adjoint, $\langle Tv, v\rangle = \langle v, Tv\rangle$. Hence

$$\begin{aligned}\|(T - iI)v\|^2 &= \langle Tv - iv,\ Tv - iv\rangle \\ &= \langle Tv, Tv\rangle - \langle Tv, iv\rangle - \langle iv, Tv\rangle + \langle iv, iv\rangle \\ &= \|Tv\|^2 + i\langle Tv, v\rangle - i\langle v, Tv\rangle + \|v\|^2 \\ &= \|Tv\|^2 + \|v\|^2,\end{aligned}$$

where we used $\langle Tv, iv\rangle = \overline{i}\,\langle Tv, v\rangle = -i\langle Tv, v\rangle$ and $\langle iv, Tv\rangle = i\langle v, Tv\rangle$, and then the cancellation $i\langle Tv, v\rangle - i\langle v, Tv\rangle = 0$. The identical computation with $-i$ replaced by $+i$ (the two cross terms again cancel, now as $-i\langle Tv, v\rangle + i\langle v, Tv\rangle$) gives

$$\|(T + iI)v\|^2 = \|Tv\|^2 + \|v\|^2.$$

Therefore

$$\|(T - iI)v\| = \|(T + iI)v\| = \sqrt{\|Tv\|^2 + \|v\|^2}$$

for every $v \in V$. Call this identity $(*)$.

Invertibility of $T - iI$ and of $T + iI$. If $(T - iI)v = 0$, then by $(*)$ we get $\|v\|^2 \le \|Tv\|^2 + \|v\|^2 = 0$, so $v = 0$. Thus $T - iI$ is injective, and since V is finite-dimensional, $T - iI$ is invertible (3.65). The same argument shows $T + iI$ is invertible. (Alternatively: every eigenvalue of a self-adjoint operator is real by 7.12, so neither i nor $-i$ is an eigenvalue of T.)

Now set

$$S = (T + iI)(T - iI)^{-1}.$$

As a product of two invertible operators, S is invertible.

S is an isometry. Let $w \in V$. Because $T - iI$ is invertible, there is a unique $v \in V$ with $w = (T - iI)v$, namely $v = (T - iI)^{-1}w$. Then $Sw = (T + iI)v$, so by $(*)$,

$$\|Sw\| = \|(T + iI)v\| = \|(T - iI)v\| = \|w\|.$$

Since $w \in V$ was arbitrary, S is an isometry. Being an invertible isometry from V to itself, S is a unitary operator by definition 7.51.

1 is not an eigenvalue of S. Suppose $w \in V$ satisfies $Sw = w$. Put $v = (T - iI)^{-1}w$, so $w = (T - iI)v$ and $Sw = (T + iI)v$. Then

$$(T + iI)v = (T - iI)v,$$

so subtracting gives $2iv = 0$, hence $v = 0$ and therefore $w = (T - iI)0 = 0$. Thus the only vector fixed by S is 0, which means 1 is not an eigenvalue of S.

Remark. This operator S is the Cayley transform of T, and it is precisely the inverse construction to that of Exercise 7D.15: the condition that 1 not be an eigenvalue of S says exactly that $S - I$ is invertible, which is the hypothesis of that exercise. Indeed, one checks that $i(S + I)(S - I)^{-1} = T$, matching the fact that the fractional linear map $z \mapsto (z + i)(z - i)^{-1}$, which carries $\mathbb{R}$ onto the unit circle minus $\{1\}$, is inverted by $z \mapsto i(z + 1)(z - 1)^{-1}$.

Problem 7D.17. Explain why the characterizations of unitary matrices given by 7.57 hold.

[7.57: Suppose Q is an n-by-n matrix. Then the following are equivalent.
(a) Q is a unitary matrix.
(b) The rows of Q form an orthonormal list in $\mathbb{F}^n$.
(c) $\|Qv\| = \|v\|$ for every $v \in \mathbb{F}^n$.
(d) $Q^*Q = QQ^* = I$, the n-by-n matrix with 1's on the diagonal and 0's elsewhere.]

Solution. The whole point is to transfer 7.49 and 7.53 from operators to matrices. Give $\mathbb{F}^n$ the Euclidean inner product and let $e_1, \dots, e_n$ denote the standard basis of $\mathbb{F}^n$, which is an orthonormal basis. Let $S \in \mathcal{L}(\mathbb{F}^n)$ be the unique operator whose matrix with respect to this basis is Q; that is,

$$\mathcal{M}\big(S, (e_1, \dots, e_n)\big) = Q.$$

Every n-by-n matrix arises this way, and S is determined by Q.
Two identifications make the translation work.
First, if we identify each $v \in \mathbb{F}^n$ with the n-by-1 column matrix $\mathcal{M}(v)$ of its coordinates with respect to $e_1, \dots, e_n$ (which are just the entries of v), then 3.76 gives

$$\mathcal{M}(Sv) = \mathcal{M}(S)\,\mathcal{M}(v) = Qv.$$

In other words, under this identification the operator S acts as multiplication by the matrix Q: we have $Sv = Qv$ for all $v \in \mathbb{F}^n$. In particular $\|Qv\| = \|Sv\|$, where the norm on the left is the Euclidean norm on $\mathbb{F}^n$.
Second, because $e_1, \dots, e_n$ is an orthonormal basis, 7.9 gives

$$\mathcal{M}(S^*) = \big(\mathcal{M}(S)\big)^* = Q^*,$$

where Q^* denotes the conjugate transpose. Combining this with 3.43 (which says $\mathcal{M}(AB) = \mathcal{M}(A)\mathcal{M}(B)$) and with the fact that $\mathcal{M}(I) = I$ and that $A \mapsto \mathcal{M}(A)$ is injective on $\mathcal{L}(\mathbb{F}^n)$, we get

$$S^*S = SS^* = I \iff Q^*Q = QQ^* = I.$$

We now prove the four conditions are equivalent by showing each is equivalent to the statement that S is a unitary operator.
(a) $\iff$ S unitary. By definition 7.56, Q is a unitary matrix exactly when its columns form an orthonormal list in $\mathbb{F}^n$. Since $Q = \mathcal{M}(S, (e_1, \dots, e_n))$ and $e_1, \dots, e_n$ is orthonormal, the equivalence of (a) and (e) in 7.49 (applied with $V = W = \mathbb{F}^n$ and both bases equal to $e_1, \dots, e_n$) says that the columns of Q form an orthonormal list if and only if S is an isometry. Finally, an isometry is injective, and an injective operator on the finite-dimensional space $\mathbb{F}^n$ is invertible (3.65), so S is an isometry if and only if S is a unitary operator (7.51). Hence (a) holds if and only if S is unitary.
(b) $\iff$ S unitary. This is immediate from the equivalence of (a) and (e) in 7.53: the rows of $\mathcal{M}(S, (e_1, \dots, e_n)) = Q$ form an orthonormal basis of $\mathbb{F}^n$ if and only if S is a unitary operator. (A list of n orthonormal vectors in the n-dimensional space $\mathbb{F}^n$ is automatically a basis, by 6.28, so "orthonormal list" and "orthonormal basis" say the same thing here.)
(c) $\iff$ S unitary. By the first identification above, $\|Qv\| = \|Sv\|$ for every $v \in \mathbb{F}^n$. Thus condition (c) says precisely that $\|Sv\| = \|v\|$ for all v, i.e. that S is an isometry (7.44), which as noted in the proof of (a) is equivalent to S being unitary.
(d) $\iff$ S unitary. By the second identification above, $Q^*Q = QQ^* = I$ if and only if $S^*S = SS^* = I$, and the latter is equivalent to S being a unitary operator by the equivalence of (a) and (b) in 7.53.
Since each of (a), (b), (c), (d) is equivalent to the single statement "S is a unitary operator", all four are equivalent to one another, which is what 7.57 asserts.

Problem 7D.18. A square matrix A is called *symmetric* if it equals its transpose. Prove that if A is a symmetric matrix with real entries, then there exists a unitary matrix Q with real entries such that Q^*AQ is a diagonal matrix.

Solution. Suppose A is an n-by-n symmetric matrix with real entries, so $A^{\mathrm{t}} = A$.
Set $V = \mathbb{R}^n$ with the Euclidean inner product, and let $e_1, \dots, e_n$ be the standard basis of $\mathbb{R}^n$, which is an orthonormal basis of V. Let $T \in \mathcal{L}(V)$ be the operator determined by

$$\mathcal{M}\big(T, (e_1, \dots, e_n)\big) = A;$$

concretely, T sends v to the vector whose coordinate column is Av, by 3.76. Note that this makes

sense: all entries of A are real, so A does define an operator on the real vector space $\mathbb{R}^n$.
T is self-adjoint. Because $e_1, \dots, e_n$ is an orthonormal basis, 7.9 gives

$$\mathcal{M}\big(T^*, (e_1, \dots, e_n)\big) = \big(\mathcal{M}(T, (e_1, \dots, e_n))\big)^* = A^*.$$

Since A has only real entries, its conjugate transpose A^* equals its transpose A^{t}, which equals A by symmetry. Hence $\mathcal{M}(T^*) = A = \mathcal{M}(T)$, and since an operator is determined by its matrix with respect to a fixed basis, $T^* = T$. Thus T is self-adjoint.
Diagonalization. Here $\mathbb{F} = \mathbb{R}$, so the real spectral theorem (7.29) applies: since T is self-adjoint, there is an orthonormal basis $f_1, \dots, f_n$ of $V = \mathbb{R}^n$ consisting of eigenvectors of T, and with respect to that basis the matrix of T is diagonal. Write

$$D = \mathcal{M}\big(T, (f_1, \dots, f_n)\big),$$

a diagonal matrix (its diagonal entries are the corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, which are real by 7.12).
Construction of Q. Let Q be the n-by-n matrix whose k-th column is the coordinate column of f_k with respect to the standard basis, i.e. whose columns are $f_1, \dots, f_n$ themselves regarded as elements of $\mathbb{R}^n$. Then:

- The entries of Q are real, because each $f_k \in \mathbb{R}^n$.
- The columns of Q form an orthonormal list in $\mathbb{R}^n$, because $f_1, \dots, f_n$ is an orthonormal basis of $\mathbb{R}^n$. Hence Q is a unitary matrix, by definition 7.56. (This does not depend on whether one reads "unitary" over $\mathbb{R}$ or over $\mathbb{C}$: the entries of Q are real, so for columns $Q_{\cdot,k}$ and $Q_{\cdot,r}$ the complex Euclidean inner product $\sum_j Q_{j,k}\overline{Q_{j,r}}$ equals the real one $\sum_j Q_{j,k}Q_{j,r}$. Thus the columns of Q are orthonormal in $\mathbb{C}^n$ as well.)
- Q is exactly the change-of-basis matrix $\mathcal{M}\big(I, (f_1, \dots, f_n), (e_1, \dots, e_n)\big)$, since the k-th column of that matrix is by definition the coordinate column of $If_k = f_k$ with respect to $e_1, \dots, e_n$.

Now apply the change-of-basis formula 3.84 with $u_1, \dots, u_n$ equal to $f_1, \dots, f_n$ and $v_1, \dots, v_n$ equal to $e_1, \dots, e_n$. It gives

$$D = \mathcal{M}\big(T, (f_1, \dots, f_n)\big) = Q^{-1}\,\mathcal{M}\big(T, (e_1, \dots, e_n)\big)\,Q = Q^{-1}AQ.$$

Finally, because Q is a unitary matrix, 7.57(d) gives $Q^*Q = QQ^* = I$, so $Q^{-1} = Q^*$. Therefore

$$Q^*AQ = Q^{-1}AQ = D,$$

which is a diagonal matrix. Thus Q is a unitary matrix with real entries such that Q^*AQ is diagonal, as desired.
Remark. Since Q is real, $Q^* = Q^{\mathrm{t}}$, so the conclusion can be restated as: every real symmetric matrix A can be written as $A = QDQ^{\mathrm{t}}$ with D real diagonal and Q real orthogonal. The hypothesis that A be real is essential: the complex symmetric matrix $\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$ is not diagonalizable at all, since it is nonzero but its square is 0.

Problem 7D.19. Suppose n is a positive integer. For this exercise, we adopt the notation that a typical element z of $\mathbb{C}^n$ is denoted by $z = (z_0, z_1, \dots, z_{n-1})$. Define linear functionals $\omega_0, \omega_1, \dots, \omega_{n-1}$ on $\mathbb{C}^n$ by

$$\omega_j(z_0, z_1, \dots, z_{n-1}) = \frac{1}{\sqrt{n}} \sum_{m=0}^{n-1} z_m\, e^{-2\pi i j m/n}.$$

The *discrete Fourier transform* is the operator $\mathcal{F}\colon \mathbb{C}^n \to \mathbb{C}^n$ defined by

$$\mathcal{F}z = \big(\omega_0(z), \omega_1(z), \dots, \omega_{n-1}(z)\big).$$

(a) Show that $\mathcal{F}$ is a unitary operator on $\mathbb{C}^n$.

(b) Show that if $(z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$ and z_n is defined to equal z_0, then

$$\mathcal{F}^{-1}(z_0, z_1, \ldots, z_{n-1}) = \mathcal{F}(z_n, z_{n-1}, \ldots, z_1).$$

(c) Show that $\mathcal{F}^4 = I$.
[The discrete Fourier transform has many important applications in data analysis. The usual Fourier transform involves expressions of the form $\int_{-\infty}^{\infty} f(x)e^{-2\pi itx}\,dx$ for complex-valued integrable functions f defined on $\mathbb{R}$.]

Solution. Throughout put

$$\zeta = e^{2\pi i/n},$$

a primitive n-th root of unity, so that $e^{-2\pi ijm/n} = \zeta^{-jm}$ and $\zeta^n = 1$. With this notation

$$(\mathcal{F}z)_j = \omega_j(z) = \frac{1}{\sqrt{n}} \sum_{m=0}^{n-1} z_m \zeta^{-jm}, \qquad j = 0, 1, \ldots, n-1.$$

Each ω_j is clearly linear, so $\mathcal{F}$ is linear.
We record the basic root-of-unity identity that drives everything. For an integer k,

$$\sum_{j=0}^{n-1} \zeta^{jk} = n \ \text{ if } n \text{ divides } k, \qquad \sum_{j=0}^{n-1} \zeta^{jk} = 0 \ \text{ otherwise.}$$

Indeed, if $n \mid k$ then $\zeta^{jk} = (\zeta^n)^{jk/n} = 1$ for every j and the sum is n. If $n \nmid k$ then $\zeta^k \neq 1$ (as ζ is a primitive n-th root of unity), and the finite geometric sum gives

$$\sum_{j=0}^{n-1} (\zeta^k)^j = \frac{1 - (\zeta^k)^n}{1 - \zeta^k} = \frac{1 - (\zeta^n)^k}{1 - \zeta^k} = \frac{1 - 1}{1 - \zeta^k} = 0.$$

(a) Let $e_0, \ldots, e_{n-1}$ be the standard basis of $\mathbb{C}^n$, an orthonormal basis for the Euclidean inner product. From the displayed formula, the matrix $F = \mathcal{M}(\mathcal{F}, (e_0, \ldots, e_{n-1}))$ has entries

$$F_{j,m} = \frac{1}{\sqrt{n}}\, \zeta^{-jm}, \qquad j, m \in \{0, 1, \ldots, n-1\}.$$

Fix $m, r \in \{0, \ldots, n-1\}$ and compute the Euclidean inner product of column m with column r of F:

$$\sum_{j=0}^{n-1} F_{j,m}\, \overline{F_{j,r}} = \frac{1}{n} \sum_{j=0}^{n-1} \zeta^{-jm}\, \overline{\zeta^{-jr}} = \frac{1}{n} \sum_{j=0}^{n-1} \zeta^{-jm} \zeta^{jr} = \frac{1}{n} \sum_{j=0}^{n-1} \zeta^{j(r-m)},$$

where we used $\overline{\zeta^{-jr}} = \zeta^{jr}$ because $|\zeta| = 1$. Since $|r - m| \leq n - 1$, the integer $r - m$ is a multiple of n if and only if $r = m$. So by the identity above this inner product equals 1 if $m = r$ and 0 if $m \neq r$.
Thus the columns of F form an orthonormal list in $\mathbb{C}^n$. By the equivalence of (a) and (e) in 7.49, $\mathcal{F}$ is an isometry. An isometry is injective, and an injective operator on the finite-dimensional space $\mathbb{C}^n$ is invertible (3.65). Hence $\mathcal{F}$ is an invertible isometry from $\mathbb{C}^n$ to itself, i.e. a unitary operator (7.51).
(b) Since $\mathcal{F}$ is unitary, 7.53(c) gives $\mathcal{F}^{-1} = \mathcal{F}^*$. By 7.9 the matrix of $\mathcal{F}^*$ with respect to the standard basis is the conjugate transpose F^*, whose entry in row j, column m is

$$(F^*)_{j,m} = \overline{F_{m,j}} = \overline{\tfrac{1}{\sqrt{n}} \zeta^{-mj}} = \frac{1}{\sqrt{n}}\, \zeta^{jm}.$$

Therefore, for $z = (z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$,

$$\left(\mathcal{F}^{-1}z\right)_j = \frac{1}{\sqrt{n}} \sum_{m=0}^{n-1} z_m\, \zeta^{jm}, \qquad j = 0, \ldots, n-1.$$

Now let $w = (z_n, z_{n-1}, \dots, z_1)$, where $z_n := z_0$. Indexing w from 0 as usual, this says

$$w_k = z_{n-k} \qquad \text{for } k = 0, 1, \dots, n-1$$

(and $w_0 = z_n = z_0$). Fix $j \in \{0, \dots, n-1\}$ and compute, substituting $m = n - k$ (so $k = 0, 1, \dots, n-1$ corresponds to $m = n, n-1, \dots, 1$) and using $\zeta^{-jn} = 1$:

$$\begin{aligned}(\mathcal{F}w)_j &= \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} w_k \zeta^{-jk} = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} z_{n-k}\, \zeta^{-jk} \\ &= \frac{1}{\sqrt{n}} \sum_{m=1}^{n} z_m\, \zeta^{-j(n-m)} = \frac{1}{\sqrt{n}} \sum_{m=1}^{n} z_m\, \zeta^{jm}.\end{aligned}$$

In the last sum the term with $m = n$ is $z_n \zeta^{jn} = z_0 \cdot 1 = z_0 \zeta^{j \cdot 0}$, which is exactly the term that would appear for $m = 0$. Replacing the $m = n$ term by the $m = 0$ term therefore leaves the sum unchanged, giving

$$(\mathcal{F}w)_j = \frac{1}{\sqrt{n}} \sum_{m=0}^{n-1} z_m\, \zeta^{jm} = \left(\mathcal{F}^{-1} z\right)_j.$$

Since this holds for every j, we conclude

$$\mathcal{F}^{-1}(z_0, z_1, \dots, z_{n-1}) = \mathcal{F}(z_n, z_{n-1}, \dots, z_1),$$

as desired.
(c) Define the *reversal operator* $R \in \mathcal{L}(\mathbb{C}^n)$ by

$$R(z_0, z_1, z_2, \dots, z_{n-1}) = (z_0, z_{n-1}, z_{n-2}, \dots, z_1);$$

equivalently, $(Rz)_k = z_{(n-k) \bmod n}$ for $k = 0, \dots, n-1$. This is linear.
Two observations.
First, $R^2 = I$. Indeed, for each k,

$$(R^2 z)_k = (Rz)_{(n-k) \bmod n} = z_{(n-((n-k) \bmod n)) \bmod n} = z_{k \bmod n} = z_k,$$

since $n - (n-k) = k$ and everything is read modulo n.
Second, since z_n was defined to equal z_0, the vector $(z_n, z_{n-1}, \dots, z_1)$ appearing in part (b) is precisely $(z_0, z_{n-1}, \dots, z_1) = Rz$. So part (b) says exactly

$$\mathcal{F}^{-1} = \mathcal{F}R.$$

Applying $\mathcal{F}^{-1}$ on the left of this last equation gives

$$\mathcal{F}^{-1}\mathcal{F}^{-1} = \mathcal{F}^{-1}\mathcal{F}R = R,$$

that is, $\mathcal{F}^{-2} = R$. Squaring and using $R^2 = I$,

$$\mathcal{F}^{-4} = \left(\mathcal{F}^{-2}\right)^2 = R^2 = I.$$

Multiplying both sides by $\mathcal{F}^4$ gives $\mathcal{F}^4 = I$, as desired.
Remark. Because $\mathcal{F}^4 = I$, the polynomial $x^4 - 1$ is a multiple of the minimal polynomial of $\mathcal{F}$, so every eigenvalue of $\mathcal{F}$ lies in $\{1, -1, i, -i\}$ and $\mathcal{F}$ is diagonalizable (consistent with 7.55, since these four numbers all have absolute value 1).

Problem 7D.20. Suppose A is a square matrix with linearly independent columns. Prove that there exist unique matrices R and Q such that R is lower triangular with only positive numbers on

its diagonal, Q is unitary, and $A = RQ$.

Solution. The idea is to reduce to the QR factorization 7.58 by taking conjugate transposes, which converts a product RQ into a product Q^*R^* and converts lower triangular into upper triangular.
Preliminaries on the conjugate transpose. For matrices M, N of compatible sizes we have $(MN)^* = N^*M^*$ and $(M^*)^* = M$, and $I^* = I$. Consequently, if M is invertible then so is M^*, with $(M^*)^{-1} = (M^{-1})^*$, since $(M^{-1})^*M^* = (MM^{-1})^* = I$ and $M^*(M^{-1})^* = (M^{-1}M)^* = I$. Also:

- M is lower triangular if and only if M^* is upper triangular. Indeed $(M^*)_{j,k} = \overline{M_{k,j}}$, so $(M^*)_{j,k} = 0$ for all $j > k$ exactly when $M_{k,j} = 0$ for all $j > k$, i.e. exactly when M vanishes above its diagonal.
- The diagonal entries of M^* are the complex conjugates of the diagonal entries of M. Hence M has only positive numbers on its diagonal if and only if M^* does.
- M is a unitary matrix if and only if M^* is a unitary matrix. This follows from 7.57: by 7.57(d), M unitary means $M^*M = MM^* = I$, and this condition is symmetric under replacing M by M^* (using $(M^*)^* = M$).

A^* has linearly independent columns. Let n be the size of the square matrix A. Regard A as the matrix, with respect to the standard basis $e_1, \dots, e_n$ of $\mathbb{F}^n$, of an operator $T \in \mathcal{L}(\mathbb{F}^n)$. The k-th column of A is $\mathcal{M}(Te_k)$, so the columns of A are linearly independent if and only if $Te_1, \dots, Te_n$ is linearly independent, which (as this list has length $n = \dim \mathbb{F}^n$) happens if and only if T is surjective, hence if and only if T is invertible. Thus A having linearly independent columns is equivalent to A being an invertible matrix. By the preliminaries, A^* is then invertible too, so by the same equivalence applied to A^*, the columns of A^* are linearly independent.
Existence. Apply the QR factorization 7.58 to the square matrix A^*, whose columns are linearly independent. It produces matrices $\widetilde{Q}$ and $\widetilde{R}$ with $\widetilde{Q}$ unitary, $\widetilde{R}$ upper triangular with only positive numbers on its diagonal, and

$$A^* = \widetilde{Q}\,\widetilde{R}.$$

Take conjugate transposes of both sides:

$$A = (A^*)^* = \big(\widetilde{Q}\widetilde{R}\big)^* = \widetilde{R}^*\,\widetilde{Q}^*.$$

Set

$$R = \widetilde{R}^*, \qquad Q = \widetilde{Q}^*.$$

By the preliminaries, R is lower triangular (being the conjugate transpose of an upper-triangular matrix) with only positive numbers on its diagonal (the conjugates of the positive diagonal entries of $\widetilde{R}$), and Q is unitary. And $A = RQ$. This proves existence.
Uniqueness. Suppose R, Q and R', Q' both have the stated properties, so

$$A = RQ = R'Q'$$

with R, R' lower triangular with positive diagonal entries and Q, Q' unitary. Taking conjugate transposes,

$$A^* = Q^*R^* = (Q')^*(R')^*.$$

Here Q^* and $(Q')^*$ are unitary, and R^* and $(R')^*$ are upper triangular with only positive numbers on their diagonals, again by the preliminaries. So these are two QR factorizations of A^*. The uniqueness assertion of 7.58 forces

$$Q^* = (Q')^* \qquad \text{and} \qquad R^* = (R')^*.$$

Taking conjugate transposes once more (and using $(M^*)^* = M$) gives $Q = Q'$ and $R = R'$. This proves uniqueness.
Remark. Concretely, this factorization is a statement about the *rows* of A. Because R is lower triangular, the entry of $A = RQ$ in row k, column j is $\sum_{m \le k} R_{k,m}Q_{m,j}$, so

$$\operatorname{row}_k(A) = \sum_{m=1}^{k} R_{k,m} \operatorname{row}_m(Q),$$

with $R_{k,k} > 0$. Thus the rows of Q are an orthonormal list obtained from the rows of A by the Gram-Schmidt procedure run from the top row downward, and R records the coefficients. (In the proof above this is the Gram-Schmidt procedure of 7.58 applied to the columns of A^*, which are the complex conjugates of the rows of A, taken in the order $1, \dots, n$.)

7.5 Exercises 7E

Problem 7E.1. Suppose $T \in \mathcal{L}(V, W)$. Show that $T = 0$ if and only if all singular values of T are 0.

Solution. Recall the definition 7.65: the singular values of T are the nonnegative square roots of the eigenvalues of T^*T, each listed as many times as the dimension of the corresponding eigenspace of T^*T (so the list has length $\dim V$).
First suppose $T = 0$. Then $T^*T = 0$, so every vector in V is an eigenvector of T^*T with eigenvalue 0. Hence 0 is the only eigenvalue of T^*T, and therefore every singular value of T equals $\sqrt{0} = 0$.
Conversely, suppose all singular values of T are 0. If $V = \{0\}$ then trivially $T = 0$, so assume $\dim V \geq 1$. Because the nonnegative square roots of the eigenvalues of T^*T are all 0, the only eigenvalue of T^*T is 0. By 7.64(a) the operator T^*T is positive, hence self-adjoint, so the spectral theorem (7.29 if $\mathbf{F} = \mathbf{R}$, 7.31 if $\mathbf{F} = \mathbf{C}$) provides an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T^*T. Each corresponding eigenvalue is 0, so $T^*Te_k = 0$ for every k. Since $e_1, \dots, e_n$ spans V, this gives $T^*T = 0$, i.e. $\operatorname{null} T^*T = V$. By 7.64(b) we have $\operatorname{null} T = \operatorname{null} T^*T = V$, and therefore $T = 0$.
Alternatively, the converse direction follows immediately from 7.68(b): the number of positive singular values of T equals $\dim \operatorname{range} T$. If all singular values are 0, then T has no positive singular values, so $\dim \operatorname{range} T = 0$, which forces $T = 0$.

Problem 7E.2. Suppose $T \in \mathcal{L}(V, W)$ and $s > 0$. Prove that s is a singular value of T if and only if there exist nonzero vectors $v \in V$ and $w \in W$ such that

$$Tv = sw \quad \text{and} \quad T^*w = sv.$$

[The vectors v, w satisfying both equations above are called a Schmidt pair. Erhard Schmidt introduced the concept of singular values in 1907.]

Solution. By definition 7.65, and because $s > 0$, the number s is a singular value of T if and only if s^2 is an eigenvalue of T^*T.
First suppose s is a singular value of T. Then s^2 is an eigenvalue of T^*T, so there exists a nonzero $v \in V$ with

$$T^*Tv = s^2 v.$$

Set $w = \frac{Tv}{s}$, which makes sense because $s > 0$. Then $Tv = sw$, which is the first equation. Furthermore

$$T^*w = \frac{1}{s} T^*Tv = \frac{1}{s}\, s^2 v = sv,$$

which is the second equation. It remains to check that $w \neq 0$. We have

$$\|Tv\|^2 = \langle Tv, Tv \rangle = \langle T^*Tv, v \rangle = \langle s^2 v, v \rangle = s^2 \|v\|^2.$$

Because $s > 0$ and $v \neq 0$, the right side is positive, so $Tv \neq 0$ and hence $w = Tv/s \neq 0$. Thus v, w are nonzero vectors satisfying both equations.
Conversely, suppose $v \in V$ and $w \in W$ are nonzero vectors with $Tv = sw$ and $T^*w = sv$. Then

$$T^*Tv = T^*(sw) = s\, T^*w = s(sv) = s^2 v.$$

Because $v \neq 0$, this shows that s^2 is an eigenvalue of T^*T. Since $s > 0$, the number s is the nonnegative square root of s^2, and hence s is a singular value of T.

Note that the two equations are symmetric under interchanging the roles of T, v and T^*, w (using $(T^*)^* = T$), so the criterion just proved also shows that T and T^* have the same positive singular values; this is consistent with 7.75.

Problem 7E.3. Give an example of $T \in \mathcal{L}(\mathbf{C}^2)$ such that 0 is the only eigenvalue of T and the singular values of T are $5, 0$.

Solution. Define $T \in \mathcal{L}(\mathbf{C}^2)$ by

$$T(z_1, z_2) = (5z_2, 0).$$

With respect to the standard basis $(1,0), (0,1)$ of $\mathbf{C}^2$ (which is orthonormal for the usual inner product), the matrix of T is

$$\mathcal{M}(T) = \begin{pmatrix} 0 & 5 \\ 0 & 0 \end{pmatrix}.$$

Eigenvalues of T. Suppose $T(z_1, z_2) = \lambda(z_1, z_2)$ with $(z_1, z_2) \neq 0$. Then $5z_2 = \lambda z_1$ and $0 = \lambda z_2$. If $\lambda \neq 0$, the second equation gives $z_2 = 0$, and then the first gives $\lambda z_1 = 0$, so $z_1 = 0$, contradicting $(z_1, z_2) \neq 0$. Hence $\lambda = 0$. Conversely 0 is an eigenvalue, since $T(1,0) = (0,0)$ and $(1,0) \neq 0$. Thus 0 is the only eigenvalue of T.

Singular values of T. Because the standard basis is orthonormal, the matrix of T^* is the conjugate transpose of $\mathcal{M}(T)$ (by 7.9), so

$$\mathcal{M}(T^*T) = \mathcal{M}(T^*)\,\mathcal{M}(T) = \begin{pmatrix} 0 & 0 \\ 5 & 0 \end{pmatrix}\begin{pmatrix} 0 & 5 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 25 \end{pmatrix}.$$

Equivalently, $T^*T(z_1, z_2) = (0, 25z_2)$. Thus the standard basis diagonalizes T^*T, whose eigenvalues are 0 and 25, with

$$\dim E(25, T^*T) = 1 \quad \text{and} \quad \dim E(0, T^*T) = 1.$$

Taking nonnegative square roots and listing in decreasing order, the singular values of T are $5, 0$, as required.

This example illustrates the point of 7.66: the list of eigenvalues of T (here only 0) can completely miss the number 5 that governs how much T stretches vectors, while the singular values record it.

Problem 7E.4. Suppose that $T \in \mathcal{L}(V, W)$, s_1 is the largest singular value of T, and s_n is the smallest singular value of T. Prove that

$$\{\|Tv\| : v \in V \text{ and } \|v\| = 1\} = [s_n, s_1].$$

Solution. Here $n = \dim V$, and we assume $n \geq 1$ (otherwise there are no unit vectors and no singular values). Let $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$ be the singular values of T, listed in decreasing order as in 7.65; thus s_1 is the largest and s_n the smallest.

By 7.64(a) the operator T^*T is positive, hence self-adjoint, so the spectral theorem (7.29 or 7.31) gives an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T^*T; moreover, by the definition of the singular values we may index this basis so that

$$T^*Te_k = s_k^2 e_k \quad \text{for } k = 1, \ldots, n,$$

exactly as in the proof of 7.70. (Indeed, listing each eigenvalue of T^*T as many times as the dimension of its eigenspace produces the list $s_1^2, \ldots, s_n^2$, and choosing an orthonormal basis of each eigenspace and concatenating produces such a basis of V.)

Now let $v \in V$ with $\|v\| = 1$. Write $v = a_1e_1 + \cdots + a_ne_n$ with $a_k = \langle v, e_k \rangle$. By 6.30(b) we have $|a_1|^2 + \cdots + |a_n|^2 = \|v\|^2 = 1$. Then

$$\|Tv\|^2 = \langle Tv, Tv \rangle = \langle T^*Tv, v \rangle = \Big\langle \sum_{k=1}^{n} s_k^2 a_k e_k,\ \sum_{j=1}^{n} a_j e_j \Big\rangle = \sum_{k=1}^{n} s_k^2 |a_k|^2,$$

using orthonormality of $e_1, \dots, e_n$. Since $s_n^2 \le s_k^2 \le s_1^2$ for every k and $\sum_k |a_k|^2 = 1$, this gives

$$s_n^2 = s_n^2 \sum_{k=1}^{n} |a_k|^2 \le \|Tv\|^2 \le s_1^2 \sum_{k=1}^{n} |a_k|^2 = s_1^2.$$

All quantities are nonnegative, so taking square roots gives $s_n \le \|Tv\| \le s_1$. Hence the set on the left is contained in $[s_n, s_1]$.
For the reverse inclusion, let $c \in [s_n, s_1]$. If $s_1 = s_n$, then $c = s_1$ and $v = e_1$ is a unit vector with $\|Te_1\|^2 = s_1^2$ by the displayed formula, so $\|Te_1\| = c$. Now suppose $s_1 > s_n$ (so $n \ge 2$). Set

$$t = \frac{c^2 - s_n^2}{s_1^2 - s_n^2},$$

and note that $s_n \le c \le s_1$ implies $s_n^2 \le c^2 \le s_1^2$, hence $t \in [0, 1]$. Define

$$v = \sqrt{t}\, e_1 + \sqrt{1 - t}\, e_n.$$

Then $\|v\|^2 = t + (1 - t) = 1$, so v is a unit vector, and by the formula above

$$\|Tv\|^2 = s_1^2\, t + s_n^2(1 - t) = s_n^2 + t\,(s_1^2 - s_n^2) = s_n^2 + (c^2 - s_n^2) = c^2.$$

Since $c \ge 0$, this gives $\|Tv\| = c$. Thus every $c \in [s_n, s_1]$ lies in the set on the left.
Combining the two inclusions yields $\{\|Tv\| : v \in V,\ \|v\| = 1\} = [s_n, s_1]$, as desired.

Problem 7E.5. Suppose $T \in \mathcal{L}(\mathbf{C}^2)$ is defined by $T(x, y) = (-4y, x)$. Find the singular values of T.

Solution. The standard basis $e_1 = (1, 0)$, $e_2 = (0, 1)$ of $\mathbf{C}^2$ is orthonormal with respect to the usual inner product. Since $Te_1 = (0, 1)$ and $Te_2 = (-4, 0)$, the matrix of T with respect to this basis is

$$\mathcal{M}(T) = \begin{pmatrix} 0 & -4 \\ 1 & 0 \end{pmatrix}.$$

Because the basis is orthonormal, 7.9 gives that $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$. Hence

$$\mathcal{M}(T^*T) = \begin{pmatrix} 0 & 1 \\ -4 & 0 \end{pmatrix} \begin{pmatrix} 0 & -4 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 16 \end{pmatrix}.$$

In other words, $T^*T(x, y) = (x, 16y)$; one can verify this directly, since $T^*(x, y) = (y, -4x)$ and therefore $T^*T(x, y) = T^*(-4y, x) = (x, 16y)$.
Thus the standard basis diagonalizes T^*T, and the eigenvalues of T^*T are 16 and 1, with

$$\dim E(16, T^*T) = 1 \quad \text{and} \quad \dim E(1, T^*T) = 1.$$

Taking nonnegative square roots and listing in decreasing order, the singular values of T are

$$4,\ 1.$$

As a check, the eigenvalues of T are the roots of $\lambda^2 + 4 = 0$, namely $2i$ and $-2i$, each of absolute value 2; these are different from the singular values, which is consistent with 7E.7 because T is not normal (indeed $TT^*(x, y) = (16x, y) \ne T^*T(x, y)$ in general).

Problem 7E.6. Find the singular values of the differentiation operator $D \in \mathcal{L}(\mathcal{P}_2(\mathbf{R}))$ defined by $Dp = p'$, where the inner product on $\mathcal{P}_2(\mathbf{R})$ is as in Example 6.34.

Solution. The inner product of Example 6.34 is

$$\langle p, q\rangle = \int_{-1}^{1} pq$$

for $p, q \in \mathcal{P}_2(\mathbf{R})$, and Example 6.34 produces (via Gram–Schmidt applied to $1, x, x^2$) the orthonormal basis

$$e_1 = \sqrt{\tfrac{1}{2}}, \qquad e_2 = \sqrt{\tfrac{3}{2}}\, x, \qquad e_3 = \sqrt{\tfrac{45}{8}}\left(x^2 - \tfrac{1}{3}\right)$$

of $\mathcal{P}_2(\mathbf{R})$.
We compute the action of D on this basis. First, $De_1 = 0$ since e_1 is constant. Next,

$$De_2 = \sqrt{\tfrac{3}{2}} = \sqrt{\tfrac{3}{2}} \cdot \sqrt{2}\, e_1 = \sqrt{3}\, e_1,$$

where we used $1 = \sqrt{2}\, e_1$. Finally,

$$De_3 = \sqrt{\tfrac{45}{8}} \cdot 2x = 2\sqrt{\tfrac{45}{8}} \cdot \sqrt{\tfrac{2}{3}}\, e_2 = 2\sqrt{\tfrac{45 \cdot 2}{8 \cdot 3}}\, e_2 = 2\sqrt{\tfrac{15}{4}}\, e_2 = \sqrt{15}\, e_2,$$

where we used $x = \sqrt{\frac{2}{3}}\, e_2$.
Hence the matrix of D with respect to the orthonormal basis e_1, e_2, e_3 is

$$\mathcal{M}(D) = \begin{pmatrix} 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{15} \\ 0 & 0 & 0 \end{pmatrix}.$$

Because this basis is orthonormal, 7.9 tells us that $\mathcal{M}(D^*)$ is the conjugate transpose (here just the transpose, as all entries are real) of $\mathcal{M}(D)$. Therefore

$$\mathcal{M}(D^*D) = \begin{pmatrix} 0 & 0 & 0 \\ \sqrt{3} & 0 & 0 \\ 0 & \sqrt{15} & 0 \end{pmatrix} \begin{pmatrix} 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{15} \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 15 \end{pmatrix}.$$

Equivalently, $D^*De_1 = 0$, $D^*De_2 = 3e_2$, and $D^*De_3 = 15e_3$.
Thus e_1, e_2, e_3 is an orthonormal basis of eigenvectors of D^*D, the eigenvalues of D^*D are 15, 3, 0, and each of the three eigenspaces is one-dimensional. Taking nonnegative square roots and listing in decreasing order, the singular values of D are

$$\sqrt{15},\ \sqrt{3},\ 0.$$

Note that D is nilpotent, so 0 is its only eigenvalue; the singular values carry considerably more information. Also, 0 is a singular value, in accordance with 7.68(a), since D is not injective.

Problem 7E.7. Suppose that $T \in \mathcal{L}(V)$ is self-adjoint or that $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, each included in this list as many times as the dimension of the corresponding eigenspace. Show that the singular values of T are $|\lambda_1|, \ldots, |\lambda_n|$, after these numbers have been sorted into decreasing order.

Solution. In either hypothesis the spectral theorem applies. If $\mathbf{F} = \mathbf{R}$, then T is self-adjoint and 7.29 applies. If $\mathbf{F} = \mathbf{C}$, then T is normal (either by hypothesis, or because T is self-adjoint and every self-adjoint operator is normal) and 7.31 applies. In particular V has an orthonormal basis consisting of eigenvectors of T, and V is the direct sum of the eigenspaces of T, which are mutually orthogonal by 7.22 (note T is normal in every case).
Let $\alpha_1, \ldots, \alpha_m$ denote the distinct eigenvalues of T. Because V has a basis consisting of eigenvectors of T, the operator T is diagonalizable, so 5.55(c) gives $V = E(\alpha_1, T) \oplus \cdots \oplus E(\alpha_m, T)$; hence $\dim E(\alpha_1, T) +$

$\cdots + \dim E(\alpha_m, T) = n$, where $n = \dim V$; this confirms that the list $\lambda_1, \ldots, \lambda_n$ in the statement has length n. Choosing an orthonormal basis of each $E(\alpha_j, T)$ and concatenating produces an orthonormal basis $e_1, \ldots, e_n$ of V (orthonormal across the blocks by 7.22) such that

$$Te_k = \lambda_k e_k \quad \text{for } k = 1, \ldots, n,$$

where $\lambda_1, \ldots, \lambda_n$ is precisely the list in the statement, up to reordering. (Reordering does not matter, since the conclusion involves sorting the numbers $|\lambda_k|$ anyway.)
Next we compute T^*Te_k. Since T is normal in every case, 7.21(e) gives

$$T^*e_k = \overline{\lambda_k}\, e_k \quad \text{for } k = 1, \ldots, n.$$

(If T is self-adjoint one can see this directly: $T^* = T$ and each λ_k is real by 7.12, so $T^*e_k = \lambda_k e_k = \overline{\lambda_k}\, e_k$.) Consequently,

$$T^*Te_k = T^*(\lambda_k e_k) = \lambda_k \overline{\lambda_k}\, e_k = |\lambda_k|^2 e_k$$

for each $k = 1, \ldots, n$.
Thus $e_1, \ldots, e_n$ is an orthonormal basis of V consisting of eigenvectors of T^*T, with corresponding eigenvalue $|\lambda_k|^2$ for e_k. We now identify the eigenvalues of T^*T and the dimensions of the corresponding eigenspaces. Let $\mu \in \mathbf{F}$ and write an arbitrary $v \in V$ as $v = a_1e_1 + \cdots + a_ne_n$. Then

$$T^*Tv - \mu v = \sum_{k=1}^{n} \big(|\lambda_k|^2 - \mu\big) a_k e_k,$$

and since $e_1, \ldots, e_n$ is linearly independent, $T^*Tv = \mu v$ if and only if $a_k = 0$ for every k with $|\lambda_k|^2 \neq \mu$. Hence

$$E(\mu, T^*T) = \operatorname{span}\{e_k : |\lambda_k|^2 = \mu\},$$

which is nonzero exactly when $\mu = |\lambda_k|^2$ for some k. Therefore the eigenvalues of T^*T are precisely the distinct numbers among $|\lambda_1|^2, \ldots, |\lambda_n|^2$, and for each such eigenvalue μ,

$$\dim E(\mu, T^*T) = \#\{k : |\lambda_k|^2 = \mu\} = \#\{k : |\lambda_k| = \sqrt{\mu}\},$$

the last equality because all the $|\lambda_k|$ and $\sqrt{\mu}$ are nonnegative.
By definition 7.65, the singular values of T are the nonnegative square roots of the eigenvalues of T^*T, each repeated as many times as the dimension of the corresponding eigenspace, listed in decreasing order. By the previous paragraph, this list consists of the distinct values among $|\lambda_1|, \ldots, |\lambda_n|$, each repeated exactly the number of times it occurs in the list $|\lambda_1|, \ldots, |\lambda_n|$, arranged in decreasing order. That is exactly the list $|\lambda_1|, \ldots, |\lambda_n|$ sorted into decreasing order, as desired.
For example, if T is self-adjoint with eigenvalues -3 and 2, with $\dim E(-3, T) = 2$ and $\dim E(2, T) = 1$, then the singular values of T are $3, 3, 2$.

Problem 7E.8. Suppose $T \in \mathcal{L}(V, W)$. Suppose $s_1 \geq s_2 \geq \cdots \geq s_m > 0$ and $e_1, \ldots, e_m$ is an orthonormal list in V and $f_1, \ldots, f_m$ is an orthonormal list in W such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

for every $v \in V$.
(a) Prove that $f_1, \ldots, f_m$ is an orthonormal basis of $\operatorname{range} T$.
(b) Prove that $e_1, \ldots, e_m$ is an orthonormal basis of $(\operatorname{null} T)^\perp$.
(c) Prove that $s_1, \ldots, s_m$ are the positive singular values of T.
(d) Prove that if $k \in \{1, \ldots, m\}$, then e_k is an eigenvector of T^*T with corresponding eigenvalue s_k^2.
(e) Prove that

$$TT^*w = s_1^2\langle w, f_1\rangle f_1 + \cdots + s_m^2\langle w, f_m\rangle f_m$$

for all $w \in W$.

Solution. Throughout, write $n = \dim V$ and let

$$E = \operatorname{span}(e_1, \dots, e_m), \qquad F = \operatorname{span}(f_1, \dots, f_m).$$

Because $e_1, \dots, e_m$ is orthonormal, $\langle e_j, e_k \rangle$ equals 1 if $j = k$ and 0 otherwise; hence applying the displayed formula to $v = e_j$ gives

$$Te_j = s_j f_j \qquad \text{for each } j \in \{1, \dots, m\}.$$

We record one computation that will be used repeatedly. Define $S \in \mathcal{L}(W, V)$ by

$$Sw = s_1 \langle w, f_1 \rangle e_1 + \cdots + s_m \langle w, f_m \rangle e_m.$$

If $v \in V$ and $w \in W$, then

$$\begin{aligned} \langle Tv, w \rangle &= \Big\langle \sum_{k=1}^{m} s_k \langle v, e_k \rangle f_k,\ w \Big\rangle = \sum_{k=1}^{m} s_k \langle v, e_k \rangle \langle f_k, w \rangle \\ &= \Big\langle v,\ \sum_{k=1}^{m} s_k \overline{\langle f_k, w \rangle}\, e_k \Big\rangle = \Big\langle v,\ \sum_{k=1}^{m} s_k \langle w, f_k \rangle e_k \Big\rangle = \langle v, Sw \rangle. \end{aligned}$$

By the definition of the adjoint (and its uniqueness), this shows that $T^* = S$, i.e. that the adjoint formula

$$T^* w = s_1 \langle w, f_1 \rangle e_1 + \cdots + s_m \langle w, f_m \rangle e_m$$

holds for all $w \in W$.
We also record the following computation, valid for each $k \in \{1, \dots, m\}$: using $Te_k = s_k f_k$ and then the adjoint formula with $w = f_k$ (together with the orthonormality of $f_1, \dots, f_m$),

$$T^* T e_k = T^*(s_k f_k) = s_k\, T^* f_k = s_k \Big(\sum_{j=1}^{m} s_j \langle f_k, f_j \rangle e_j \Big) = s_k \cdot s_k e_k = s_k^2 e_k.$$

(a) The displayed formula for T exhibits every vector Tv as a linear combination of $f_1, \dots, f_m$, so $\operatorname{range} T \subseteq F$. Conversely, for each j we have $f_j = \frac{1}{s_j} T e_j \in \operatorname{range} T$ (legitimate because $s_j > 0$), so $F \subseteq \operatorname{range} T$. Hence $\operatorname{range} T = F$. An orthonormal list is linearly independent, so $f_1, \dots, f_m$ is a linearly independent list spanning $\operatorname{range} T$, i.e. an orthonormal basis of $\operatorname{range} T$.
(b) Suppose $v \in V$. Since $f_1, \dots, f_m$ is linearly independent and each s_k is nonzero,

$$Tv = 0 \iff \sum_{k=1}^{m} s_k \langle v, e_k \rangle f_k = 0 \iff s_k \langle v, e_k \rangle = 0 \text{ for all } k \iff \langle v, e_k \rangle = 0 \text{ for all } k.$$

The last condition says exactly that v is orthogonal to each of $e_1, \dots, e_m$, which holds if and only if $v \in E^\perp$: indeed, if $\langle v, e_k \rangle = 0$ for every k and $u = a_1 e_1 + \cdots + a_m e_m \in E$, then by conjugate homogeneity and additivity of the inner product in its second slot $\langle v, u \rangle = \sum_{k=1}^{m} \overline{a_k} \langle v, e_k \rangle = 0$; the converse is immediate since each $e_k \in E$. Thus $\operatorname{null} T = E^\perp$. Since E is a subspace of the finite-dimensional space V, applying 6.52 gives

$$(\operatorname{null} T)^\perp = (E^\perp)^\perp = E.$$

Therefore $e_1, \dots, e_m$ is an orthonormal (hence linearly independent) list spanning $(\operatorname{null} T)^\perp$, i.e. an orthonormal basis of $(\operatorname{null} T)^\perp$.
(c) Extend the orthonormal list $e_1, \dots, e_m$ to an orthonormal basis $e_1, \dots, e_n$ of V: extend E's orthonormal basis by an orthonormal basis of $E^\perp$, which exists by 6.35, and use $V = E \oplus E^\perp$ (6.49). For $k > m$ we have $e_k \in E^\perp = \operatorname{null} T$ by part (b), so $Te_k = 0$ and hence $T^* T e_k = 0$. Combining with the recorded computation $T^* T e_k = s_k^2 e_k$ for $k \le m$, and setting

$$d_k = \begin{cases} s_k^2 & \text{if } 1 \le k \le m, \\ 0 & \text{if } m < k \le n, \end{cases}$$

we get $T^*Te_k = d_k e_k$ for every $k \in \{1, \dots, n\}$, and $d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$ because $s_1 \geq \cdots \geq s_m > 0$. Now we identify the eigenvalues of T^*T together with the dimensions of the eigenspaces. Fix $\lambda \in \mathbf{F}$ and write $v = a_1 e_1 + \cdots + a_n e_n$. Then

$$T^*Tv - \lambda v = \sum_{k=1}^{n} a_k (d_k - \lambda) e_k,$$

which is 0 if and only if $a_k(d_k - \lambda) = 0$ for every k, i.e. if and only if $a_k = 0$ for every k with $d_k \neq \lambda$. Hence

$$E(\lambda, T^*T) = \operatorname{span}\{e_k : d_k = \lambda\}, \qquad \dim E(\lambda, T^*T) = \#\{k : d_k = \lambda\}.$$

In particular the eigenvalues of T^*T are exactly the distinct numbers appearing in the list $d_1, \dots, d_n$, and each appears in that list as many times as the dimension of its eigenspace. Therefore the list of eigenvalues of T^*T, in decreasing order and with each repeated according to the dimension of the corresponding eigenspace, is precisely $d_1, \dots, d_n$. Taking nonnegative square roots, the singular values of T (definition 7.65) are

$$s_1, \dots, s_m, \underbrace{0, \dots, 0}_{n-m}.$$

Thus the positive singular values of T are exactly $s_1, \dots, s_m$.
(d) Let $k \in \{1, \dots, m\}$. The computation recorded at the start of this solution gives $T^*Te_k = s_k^2 e_k$. Because $\|e_k\| = 1$ we have $e_k \neq 0$, so e_k is an eigenvector of T^*T with corresponding eigenvalue s_k^2, as desired.
(e) Let $w \in W$. Apply T to the adjoint formula and use $Te_k = s_k f_k$:

$$TT^*w = T\Big(\sum_{k=1}^{m} s_k \langle w, f_k \rangle e_k\Big) = \sum_{k=1}^{m} s_k \langle w, f_k \rangle Te_k = \sum_{k=1}^{m} s_k^2 \langle w, f_k \rangle f_k,$$

which is the asserted formula

$$TT^*w = s_1^2 \langle w, f_1 \rangle f_1 + \cdots + s_m^2 \langle w, f_m \rangle f_m.$$

Problem 7E.9. Suppose $T \in \mathcal{L}(V, W)$. Show that T and T^* have the same positive singular values.

Solution. Let m denote the number of positive singular values of T.
First dispose of the degenerate case $m = 0$. If T has no positive singular values, then by 7.68(b) we have $\dim \operatorname{range} T = 0$, so $T = 0$ and hence $T^* = 0$. Then $\dim \operatorname{range} T^* = 0$, so 7.68(b) applied to T^* shows that T^* also has no positive singular values. Thus in this case T and T^* have the same (empty) list of positive singular values.
Now suppose $m \geq 1$ and let $s_1 \geq s_2 \geq \cdots \geq s_m > 0$ be the positive singular values of T, listed in decreasing order as in definition 7.65. By the singular value decomposition 7.70, there exist orthonormal lists $e_1, \dots, e_m$ in V and $f_1, \dots, f_m$ in W such that

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m$$

for every $v \in V$. By 7.75 (see 7.77), the adjoint is then given by

$$T^*w = s_1 \langle w, f_1 \rangle e_1 + \cdots + s_m \langle w, f_m \rangle e_m$$

for every $w \in W$.
Now read the last display as a statement about the linear map $T^* \in \mathcal{L}(W, V)$: we have numbers $s_1 \geq s_2 \geq \cdots \geq s_m > 0$, an orthonormal list $f_1, \dots, f_m$ in the domain W, and an orthonormal list $e_1, \dots, e_m$ in the codomain V, satisfying exactly the hypothesis of Exercise 7E.8 for the map T^* (with the roles of the e's and the f's interchanged). Hence Exercise 7E.8(c), applied to T^*, shows that $s_1, \dots, s_m$ are the positive singular values of T^*.
Since $s_1, \dots, s_m$ are also by construction the positive singular values of T, the linear maps T and T^* have the same positive singular values.

A remark on the full lists: T and T^* need not have the same lists of singular values, because the list of singular values of T has length $\dim V$ while that of T^* has length $\dim W$. The two lists agree except for the number of zeros appended at the end.

Problem 7E.10. Suppose $T \in \mathcal{L}(V,W)$ has singular values $s_1,\dots,s_n$. Prove that if T is an invertible linear map, then T^{-1} has singular values

$$\frac{1}{s_n},\dots,\frac{1}{s_1}.$$

Solution. The list of singular values of T has length $\dim V$, so $n = \dim V$, and by definition 7.65 the list is in decreasing order: $s_1 \ge s_2 \ge \cdots \ge s_n \ge 0$.
Assume T is invertible. Then T is injective, so 0 is not a singular value of T by 7.68(a); hence

$$s_1 \ge s_2 \ge \cdots \ge s_n > 0.$$

Also T is surjective, so $\operatorname{range} T = W$ and therefore $\dim W = \dim V = n$.
All n singular values of T are positive, so the singular value decomposition 7.70 provides orthonormal lists $e_1,\dots,e_n$ in V and $f_1,\dots,f_n$ in W with

$$Tv = s_1\langle v,e_1\rangle f_1 + \cdots + s_n\langle v,e_n\rangle f_n$$

for every $v \in V$. An orthonormal list is linearly independent, and these lists have length $n = \dim V = \dim W$; hence $e_1,\dots,e_n$ is an orthonormal basis of V and $f_1,\dots,f_n$ is an orthonormal basis of W. Taking $v = e_k$ in the display above gives

$$Te_k = s_k f_k \qquad \text{for } k = 1,\dots,n.$$

Define $S \in \mathcal{L}(W,V)$ by

$$Sw = \frac{\langle w,f_1\rangle}{s_1}e_1 + \cdots + \frac{\langle w,f_n\rangle}{s_n}e_n.$$

Then for each k,

$$S(Te_k) = S(s_k f_k) = s_k\cdot\frac{1}{s_k}e_k = e_k \qquad \text{and} \qquad T(Sf_k) = T\Big(\frac{1}{s_k}e_k\Big) = \frac{1}{s_k}\,s_k f_k = f_k,$$

using orthonormality of $f_1,\dots,f_n$. Since ST and the identity on V agree on the basis $e_1,\dots,e_n$, and TS and the identity on W agree on the basis $f_1,\dots,f_n$, we conclude $ST = I_V$ and $TS = I_W$. Hence

$$T^{-1}w = \frac{\langle w,f_1\rangle}{s_1}e_1 + \cdots + \frac{\langle w,f_n\rangle}{s_n}e_n \qquad \text{for every } w \in W.$$

(Alternatively, this formula is 7.78 together with 6.69(a), which says $T^\dagger = T^{-1}$ for invertible T.)
Now reindex so that the coefficients appear in decreasing order. For $k = 1,\dots,n$ set

$$t_k = \frac{1}{s_{n+1-k}}, \qquad g_k = f_{n+1-k}, \qquad h_k = e_{n+1-k}.$$

Since $s_1 \ge \cdots \ge s_n > 0$, taking reciprocals reverses the inequalities, so

$$t_1 = \frac{1}{s_n} \ge \frac{1}{s_{n-1}} = t_2 \ge \cdots \ge t_n = \frac{1}{s_1} > 0.$$

Moreover $g_1,\dots,g_n$ is an orthonormal list in W (a reordering of $f_1,\dots,f_n$) and $h_1,\dots,h_n$ is an orthonormal list in V, and substituting $j = n+1-k$ in the formula for T^{-1} gives

$$T^{-1}w = t_1\langle w,g_1\rangle h_1 + \cdots + t_n\langle w,g_n\rangle h_n \qquad \text{for every } w \in W.$$

This is exactly the hypothesis of Exercise 7E.8 for the linear map $T^{-1} \in \mathcal{L}(W,V)$. Hence by 7E.8(c) the positive singular values of T^{-1} are $t_1,\dots,t_n$. Finally, the full list of singular values of T^{-1} has length $\dim W = n$, and we have just found n positive singular values, so there are no zeros left over: the singular values of T^{-1} are

$$t_1,\dots,t_n = \frac{1}{s_n},\dots,\frac{1}{s_1},$$

as desired.

Problem 7E.11. Suppose that $T \in \mathcal{L}(V, W)$ and $v_1, \dots, v_n$ is an orthonormal basis of V. Let $s_1, \dots, s_n$ denote the singular values of T.
(a) Prove that $\|Tv_1\|^2 + \cdots + \|Tv_n\|^2 = s_1^2 + \cdots + s_n^2$.
(b) Prove that if $W = V$ and T is a positive operator, then

$$\langle Tv_1, v_1\rangle + \cdots + \langle Tv_n, v_n\rangle = s_1 + \cdots + s_n.$$

See the comment after Exercise 5 in Section 7A.

Solution. Note first that the list of singular values of T has length $\dim V$, so indeed $n = \dim V$, consistent with $v_1, \dots, v_n$ being an orthonormal basis of V.
We will use Parseval's identity 6.30(b) repeatedly: if $u_1, \dots, u_n$ is an orthonormal basis of V and $v \in V$, then $\|v\|^2 = \sum_{i=1}^{n} |\langle v, u_i\rangle|^2$.
(a) Let $s_1 \geq \cdots \geq s_m > 0$ be the positive singular values of T (so $s_k = 0$ for $m < k \leq n$; if $m = 0$ then $T = 0$ by 7.68(b) and both sides of the claimed identity are 0, so assume $m \geq 1$). By the singular value decomposition 7.70 there are orthonormal lists $e_1, \dots, e_m$ in V and $f_1, \dots, f_m$ in W with

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

for every $v \in V$. Because $f_1, \dots, f_m$ is orthonormal, 6.24 (the norm of an orthonormal linear combination) gives, for every $v \in V$,

$$\|Tv\|^2 = \sum_{k=1}^{m} \left|s_k\langle v, e_k\rangle\right|^2 = \sum_{k=1}^{m} s_k^2\,|\langle v, e_k\rangle|^2.$$

Apply this with $v = v_i$ and sum over i, interchanging the two finite sums:

$$\sum_{i=1}^{n} \|Tv_i\|^2 = \sum_{i=1}^{n}\sum_{k=1}^{m} s_k^2|\langle v_i, e_k\rangle|^2 = \sum_{k=1}^{m} s_k^2 \sum_{i=1}^{n} |\langle v_i, e_k\rangle|^2.$$

For each fixed k we have $|\langle v_i, e_k\rangle| = |\overline{\langle e_k, v_i\rangle}| = |\langle e_k, v_i\rangle|$, so Parseval's identity applied to e_k and the orthonormal basis $v_1, \dots, v_n$ gives

$$\sum_{i=1}^{n} |\langle v_i, e_k\rangle|^2 = \sum_{i=1}^{n} |\langle e_k, v_i\rangle|^2 = \|e_k\|^2 = 1.$$

Therefore

$$\|Tv_1\|^2 + \cdots + \|Tv_n\|^2 = \sum_{k=1}^{m} s_k^2 = s_1^2 + \cdots + s_n^2,$$

the last equality holding because $s_k = 0$ for $k > m$. Note that the answer does not depend on the choice of orthonormal basis $v_1, \dots, v_n$, which is the point of the comment after Exercise 5 in Section 7A.
(b) Now suppose $W = V$ and T is a positive operator.
Step 1: identify the singular values of T as its eigenvalues. Since T is positive it is self-adjoint and all its eigenvalues are nonnegative (7.38). By the spectral theorem (7.29 if $\mathbf{F} = \mathbf{R}$, 7.31 if $\mathbf{F} = \mathbf{C}$) there is an orthonormal basis $u_1, \dots, u_n$ of V consisting of eigenvectors of T, say $Tu_k = \lambda_k u_k$ with each $\lambda_k \geq 0$; relabel the u_k so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Because T is self-adjoint, $T^*T = T^2$, and $T^2u_k = \lambda_k^2 u_k$ for each k.
We now use the following elementary observation. If $A \in \mathcal{L}(V)$ satisfies $Au_k = d_k u_k$ for an orthonormal basis $u_1, \dots, u_n$ of V and scalars $d_1, \dots, d_n$, then writing $v = a_1u_1 + \cdots + a_nu_n$ we have $Av - \lambda v = \sum_k a_k(d_k - \lambda)u_k$, which vanishes precisely when $a_k = 0$ for every k with $d_k \neq \lambda$. Hence $E(\lambda, A) = \operatorname{span}\{u_k : d_k = \lambda\}$ and so $\dim E(\lambda, A)$ equals the number of indices k with $d_k = \lambda$. Consequently the eigenvalues of A, listed in decreasing order with each repeated as many times as the dimension of its eigenspace, form exactly the list $d_1, \dots, d_n$ (when the d_k are already in decreasing order).

Applying this to $A = T^*T = T^2$ with $d_k = \lambda_k^2$ (a decreasing list, since $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$), the eigenvalues of T^*T listed with multiplicities are $\lambda_1^2, \ldots, \lambda_n^2$. Taking nonnegative square roots as in definition 7.65,

$$s_k = \lambda_k \qquad \text{for } k = 1, \ldots, n.$$

Step 2: compute the sum. Let $v_1, \ldots, v_n$ be any orthonormal basis of V. By 6.30(a), $v_i = \sum_{k=1}^n \langle v_i, u_k\rangle u_k$, so

$$Tv_i = \sum_{k=1}^n \lambda_k \langle v_i, u_k\rangle u_k,$$

and therefore, using orthonormality of $u_1, \ldots, u_n$,

$$\langle Tv_i, v_i\rangle = \Big\langle \sum_{k=1}^n \lambda_k \langle v_i, u_k\rangle u_k,\ \sum_{j=1}^n \langle v_i, u_j\rangle u_j \Big\rangle = \sum_{k=1}^n \lambda_k \langle v_i, u_k\rangle \overline{\langle v_i, u_k\rangle} = \sum_{k=1}^n \lambda_k |\langle v_i, u_k\rangle|^2.$$

Summing over i and interchanging the sums,

$$\sum_{i=1}^n \langle Tv_i, v_i\rangle = \sum_{k=1}^n \lambda_k \sum_{i=1}^n |\langle v_i, u_k\rangle|^2 = \sum_{k=1}^n \lambda_k \|u_k\|^2 = \sum_{k=1}^n \lambda_k,$$

where the middle equality is Parseval's identity applied to u_k with respect to the orthonormal basis $v_1, \ldots, v_n$ (again using $|\langle v_i, u_k\rangle| = |\langle u_k, v_i\rangle|$). By Step 1 this equals $s_1 + \cdots + s_n$, which completes the proof:

$$\langle Tv_1, v_1\rangle + \cdots + \langle Tv_n, v_n\rangle = s_1 + \cdots + s_n.$$

Problem 7E.12. (a) Give an example of a finite-dimensional vector space and an operator T on it such that the singular values of T^2 do not equal the squares of the singular values of T.
(b) Suppose $T \in \mathcal{L}(V)$ is normal. Prove that the singular values of T^2 equal the squares of the singular values of T.

Solution. Before starting, we isolate a fact used in both parts (it was also established in the solution to Exercise 7E.8). Suppose $A \in \mathcal{L}(V)$, $u_1, \ldots, u_n$ is an orthonormal basis of V, and $Au_k = d_k u_k$ for each k. Writing $v = a_1u_1 + \cdots + a_nu_n$, we have $Av - \lambda v = \sum_k a_k(d_k - \lambda)u_k$, which is 0 exactly when $a_k = 0$ for every k with $d_k \neq \lambda$. Hence

$$E(\lambda, A) = \operatorname{span}\{u_k : d_k = \lambda\}, \qquad \dim E(\lambda, A) = \#\{k : d_k = \lambda\}.$$

Thus if $d_1 \geq d_2 \geq \cdots \geq d_n$, then the eigenvalues of A, listed in decreasing order with each included as many times as the dimension of its eigenspace, are exactly $d_1, \ldots, d_n$. We call this the diagonal counting fact.
(a) Take $V = \mathbf{F}^2$ with the Euclidean inner product and define $T \in \mathcal{L}(\mathbf{F}^2)$ by

$$T(x, y) = (y, 0).$$

Then $T^2(x, y) = T(y, 0) = (0, 0)$, so $T^2 = 0$. All eigenvalues of $(T^2)^*T^2 = 0$ are 0, so the singular values of T^2 are $0, 0$.
To compute the singular values of T, note that for all $(x, y), (a, b) \in \mathbf{F}^2$,

$$\langle T(x, y), (a, b)\rangle = \langle (y, 0), (a, b)\rangle = y\overline{a} = \langle (x, y), (0, a)\rangle,$$

so $T^*(a, b) = (0, a)$. Hence

$$T^*T(x, y) = T^*(y, 0) = (0, y).$$

Thus $T^*Te_1 = 0$ and $T^*Te_2 = e_2$, where e_1, e_2 is the standard (orthonormal) basis of $\mathbf{F}^2$. By the diagonal counting fact with $d_1 = 1, d_2 = 0$ (after reordering the basis as e_2, e_1), the eigenvalues of T^*T listed with multiplicities are $1, 0$, so the singular values of T are $1, 0$.

The squares of the singular values of T are $1, 0$, while the singular values of T^2 are $0, 0$. These lists differ, as required.
(b) Now suppose $T \in \mathcal{L}(V)$ is normal, so $T^*T = TT^*$. Set $P = T^*T$, which is a positive operator by 7.64(a). Using $(T^2)^* = (T^*)^2$ [from 7.5(d), which gives $(ST)^* = T^*S^*$] and then normality,

$$(T^2)^*T^2 = T^*T^*TT = T^*(T^*T)T = T^*(TT^*)T = (T^*T)(T^*T) = P^2.$$

Because P is a positive operator, 7.38 and the spectral theorem (7.29 if $\mathbf{F} = \mathbf{R}$, 7.31 if $\mathbf{F} = \mathbf{C}$) give an orthonormal basis $u_1, \dots, u_n$ of V with

$$Pu_k = \mu_k u_k, \qquad \mu_k \geq 0,$$

for each k; relabel so that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \geq 0$. Applying P again,

$$P^2 u_k = \mu_k^2 u_k \qquad \text{for each } k,$$

and $\mu_1^2 \geq \mu_2^2 \geq \cdots \geq \mu_n^2 \geq 0$ because the μ_k are nonnegative and decreasing.
By the diagonal counting fact applied to $P = T^*T$, the eigenvalues of T^*T listed in decreasing order with multiplicities are $\mu_1, \dots, \mu_n$. Hence, by definition 7.65, the singular values of T are

$$s_k = \sqrt{\mu_k} \qquad (k = 1, \dots, n).$$

By the diagonal counting fact applied to $P^2 = (T^2)^*T^2$, the eigenvalues of $(T^2)^*T^2$ listed in decreasing order with multiplicities are $\mu_1^2, \dots, \mu_n^2$. Hence the singular values of T^2 are

$$\sqrt{\mu_k^2} = \mu_k = \left(\sqrt{\mu_k}\right)^2 = s_k^2 \qquad (k = 1, \dots, n),$$

where $\sqrt{\mu_k^2} = \mu_k$ because $\mu_k \geq 0$.
Thus the singular values of T^2, in order, are precisely the squares of the singular values of T.

Problem 7E.13. Suppose $T_1, T_2 \in \mathcal{L}(V)$. Prove that T_1 and T_2 have the same singular values if and only if there exist unitary operators $S_1, S_2 \in \mathcal{L}(V)$ such that $T_1 = S_1 T_2 S_2$.

Solution. Write $n = \dim V$; the list of singular values of each of T_1, T_2 has length n.
First suppose there exist unitary operators $S_1, S_2 \in \mathcal{L}(V)$ with $T_1 = S_1 T_2 S_2$. By 7.53, $S_1^* S_1 = I$ and $S_2^* = S_2^{-1}$. Hence, using 7.5(d) repeatedly,

$$T_1^*T_1 = (S_1T_2S_2)^*(S_1T_2S_2) = S_2^*T_2^*S_1^*S_1T_2S_2 = S_2^*(T_2^*T_2)S_2 = S_2^{-1}(T_2^*T_2)S_2.$$

So $T_1^*T_1$ and $T_2^*T_2$ are similar. Similar operators have the same eigenvalues with the same eigenspace dimensions: indeed, for $\lambda \in \mathbf{F}$ and $v \in V$,

$$(T_1^*T_1)v = \lambda v \iff S_2^{-1}(T_2^*T_2)S_2 v = \lambda v \iff (T_2^*T_2)(S_2 v) = \lambda(S_2 v),$$

so S_2 restricts to an injective linear map from $E(\lambda, T_1^*T_1)$ onto $E(\lambda, T_2^*T_2)$ (its inverse being the restriction of S_2^{-1}). Therefore λ is an eigenvalue of $T_1^*T_1$ if and only if it is an eigenvalue of $T_2^*T_2$, and in that case the two eigenspaces have equal dimensions. By definition 7.65, the lists of singular values of T_1 and of T_2 are obtained from these identical data in the same way, so T_1 and T_2 have the same singular values.
Conversely, suppose T_1 and T_2 have the same singular values $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$. Let m be the number of positive ones, so $s_1 \geq \cdots \geq s_m > 0$ and $s_k = 0$ for $k > m$; this number m is the same for both maps because the two lists coincide.
By the singular value decomposition 7.70 applied to T_1, there are orthonormal lists $e_1, \dots, e_m$ and $f_1, \dots, f_m$ in V with

$$T_1 v = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \qquad \text{for every } v \in V,$$

and applied to T_2, there are orthonormal lists $g_1, \dots, g_m$ and $h_1, \dots, h_m$ in V with

$$T_2 v = s_1\langle v, g_1\rangle h_1 + \cdots + s_m\langle v, g_m\rangle h_m \qquad \text{for every } v \in V.$$

(If $m = 0$ these are empty lists and both maps are 0; the argument below still applies, with the four lists below taken to be arbitrary orthonormal bases.)
Each of the four orthonormal lists spans a subspace of V whose orthogonal complement has an orthonormal basis by 6.35; adjoining such a basis and using $V = U \oplus U^{\perp}$ (6.49) extends each list to an orthonormal basis of V. Do this, obtaining orthonormal bases

$$e_1, \dots, e_n, \qquad f_1, \dots, f_n, \qquad g_1, \dots, g_n, \qquad h_1, \dots, h_n$$

of V, each extending the corresponding list of length m.
Define $S_2 \in \mathcal{L}(V)$ to be the unique linear map with $S_2 e_k = g_k$ for $k = 1, \dots, n$, and define $S_1 \in \mathcal{L}(V)$ to be the unique linear map with $S_1 h_k = f_k$ for $k = 1, \dots, n$ (each exists and is unique by 3.4, since we are prescribing the values on a basis). Each of S_1, S_2 maps an orthonormal basis of V to an orthonormal basis of V, so each is unitary by the equivalence of (a) and (d) in 7.53.
Now let $v \in V$. By 6.30(a) we have $v = \sum_{k=1}^{n}\langle v, e_k\rangle e_k$, so

$$S_2 v = \sum_{k=1}^{n}\langle v, e_k\rangle g_k,$$

and since $g_1, \dots, g_n$ is orthonormal, $\langle S_2 v, g_k\rangle = \langle v, e_k\rangle$ for each k. Therefore

$$T_2 S_2 v = \sum_{k=1}^{m} s_k\langle S_2 v, g_k\rangle h_k = \sum_{k=1}^{m} s_k\langle v, e_k\rangle h_k,$$

and applying S_1 (which sends h_k to f_k),

$$S_1 T_2 S_2 v = \sum_{k=1}^{m} s_k\langle v, e_k\rangle f_k = T_1 v.$$

Since $v \in V$ was arbitrary, $T_1 = S_1 T_2 S_2$ with S_1, S_2 unitary, completing the proof.

Problem 7E.14. Suppose $T \in \mathcal{L}(V, W)$. Let s_n denote the smallest singular value of T. Prove that $s_n\|v\| \le \|Tv\|$ for every $v \in V$.

Solution. Let $n = \dim V$ and let $s_1 \ge s_2 \ge \cdots \ge s_n \ge 0$ be the singular values of T; by definition 7.65 the list has length n and is in decreasing order, so s_n is indeed the smallest singular value and $s_n \ge 0$. By 7.64(a) the operator $T^*T \in \mathcal{L}(V)$ is positive, so by the spectral theorem (7.29 if $\mathbf{F} = \mathbf{R}$, 7.31 if $\mathbf{F} = \mathbf{C}$) there is an orthonormal basis $u_1, \dots, u_n$ of V consisting of eigenvectors of T^*T, say

$$T^*T u_k = d_k u_k,$$

with each $d_k \ge 0$ by 7.38; relabel so that $d_1 \ge d_2 \ge \cdots \ge d_n \ge 0$.
We check that $d_k = s_k^2$ for each k. Fix $\lambda \in \mathbf{F}$ and write $v = a_1 u_1 + \cdots + a_n u_n$. Then

$$T^*T v - \lambda v = \sum_{k=1}^{n} a_k(d_k - \lambda)u_k,$$

which is 0 if and only if $a_k = 0$ for every k with $d_k \ne \lambda$. Hence $E(\lambda, T^*T) = \operatorname{span}\{u_k : d_k = \lambda\}$, so $\dim E(\lambda, T^*T)$ equals the number of indices k with $d_k = \lambda$. Consequently the eigenvalues of T^*T, in decreasing order and each repeated as many times as the dimension of the corresponding eigenspace, are exactly $d_1, \dots, d_n$. Taking nonnegative square roots gives $s_k = \sqrt{d_k}$, that is, $d_k = s_k^2$.

Now let $v \in V$ be arbitrary and write $v = a_1u_1 + \cdots + a_nu_n$ with $a_k = \langle v, u_k\rangle$ (6.30(a)). Then

$$T^*Tv = \sum_{k=1}^{n} a_k s_k^2 u_k.$$

Also $\langle T^*Tv, v\rangle = \langle Tv, (T^*)^*v\rangle = \langle Tv, Tv\rangle$, where the first equality is the definition of the adjoint of T^* and the second uses $(T^*)^* = T$ (7.5(c)). Hence, using orthonormality of $u_1, \ldots, u_n$,

$$\|Tv\|^2 = \langle T^*Tv, v\rangle = \Big\langle \sum_{k=1}^{n} a_k s_k^2 u_k,\ \sum_{j=1}^{n} a_j u_j \Big\rangle = \sum_{k=1}^{n} s_k^2\, a_k\overline{a_k} = \sum_{k=1}^{n} s_k^2 |a_k|^2.$$

Since $s_k \geq s_n \geq 0$ for every k, we have $s_k^2 \geq s_n^2$ for every k, and therefore

$$\|Tv\|^2 = \sum_{k=1}^{n} s_k^2 |a_k|^2 \geq s_n^2 \sum_{k=1}^{n} |a_k|^2 = s_n^2 \|v\|^2,$$

where the last equality is 6.24 (equivalently, Parseval's identity 6.30(b)).
Both $\|Tv\|$ and $s_n\|v\|$ are nonnegative, so taking nonnegative square roots of the inequality $\|Tv\|^2 \geq (s_n\|v\|)^2$ yields

$$s_n\|v\| \leq \|Tv\|,$$

as desired.

Problem 7E.15. Suppose $T \in \mathcal{L}(V)$ and $s_1 \geq \cdots \geq s_n$ are the singular values of T. Prove that if λ is an eigenvalue of T, then $s_1 \geq |\lambda| \geq s_n$.

Solution. Let $n = \dim V$. Note that λ in the statement denotes an eigenvalue of T; the eigenvalues of T^*T will be written with a different letter to avoid confusion.
By 7.64(a) the operator T^*T is a positive operator on V; in particular T^*T is self-adjoint, hence also normal. So the spectral theorem (7.29 if $\mathbf{F} = \mathbf{R}$, 7.31 if $\mathbf{F} = \mathbf{C}$) provides an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T^*T, say

$$T^*Te_k = \mu_k e_k \qquad (k = 1, \ldots, n),$$

where each $\mu_k \geq 0$ because every eigenvalue of a positive operator is nonnegative [7.38, (a) implies (b)]. Relabel the basis so that $\mu_1 \geq \cdots \geq \mu_n$.
We claim that $s_k = \sqrt{\mu_k}$ for each k. To see this, let μ be an eigenvalue of T^*T and set $K_\mu = \{k : \mu_k = \mu\}$, so that $\operatorname{span}\{e_k : k \in K_\mu\} \subseteq E(\mu, T^*T)$. As μ ranges over the distinct eigenvalues of T^*T, the subspaces $\operatorname{span}\{e_k : k \in K_\mu\}$ decompose V as a direct sum, because $e_1, \ldots, e_n$ is a basis of V; and the eigenspaces $E(\mu, T^*T)$ also decompose V as a direct sum, by the spectral theorem. Since each summand of the first decomposition sits inside the corresponding summand of the second, and the two families of dimensions have the same total n, the inclusions must all be equalities. Hence each eigenvalue μ of T^*T occurs in the list $\mu_1, \ldots, \mu_n$ exactly $\dim E(\mu, T^*T)$ times. Thus $\sqrt{\mu_1} \geq \cdots \geq \sqrt{\mu_n}$ is precisely the list of nonnegative square roots of the eigenvalues of T^*T, in decreasing order, each included as many times as the dimension of the corresponding eigenspace. By the definition of singular values (7.65),

$$s_k = \sqrt{\mu_k} \qquad (k = 1, \ldots, n).$$

In particular there are exactly n singular values, matching the hypothesis $s_1 \geq \cdots \geq s_n$, and

$$T^*Te_k = s_k^2 e_k \qquad (k = 1, \ldots, n).$$

Now suppose $v \in V$ and write $v = a_1e_1 + \cdots + a_ne_n$ with $a_1, \ldots, a_n \in \mathbf{F}$. Then

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle = \Big\langle \sum_{k=1}^{n} a_k s_k^2 e_k,\ \sum_{j=1}^{n} a_j e_j \Big\rangle = \sum_{k=1}^{n} s_k^2 |a_k|^2,$$

where the second equality holds because $\langle T^*Tv, v\rangle = \langle Tv, (T^*)^*v\rangle = \langle Tv, Tv\rangle$ using the definition of the adjoint together with 7.5(c), the third equality uses $T^*Te_k = s_k^2 e_k$, and the last equality uses the orthonormality of $e_1, \dots, e_n$. Because $s_n^2 \le s_k^2 \le s_1^2$ for every k, and because $\|v\|^2 = \sum_{k=1}^n |a_k|^2$ (again by orthonormality), the displayed computation yields

$$s_n^2\|v\|^2 \le \|Tv\|^2 \le s_1^2\|v\|^2,$$

and hence

$$s_n\|v\| \le \|Tv\| \le s_1\|v\| \qquad \text{for every } v \in V.$$

Finally, suppose λ is an eigenvalue of T. Choose an eigenvector $v \in V$ corresponding to λ with $\|v\| = 1$ (any eigenvector may be divided by its norm). Then $Tv = \lambda v$, so

$$\|Tv\| = \|\lambda v\| = |\lambda|\,\|v\| = |\lambda|.$$

Substituting into the inequality above with this v (for which $\|v\| = 1$) gives

$$s_n \le |\lambda| \le s_1,$$

which is exactly the assertion $s_1 \ge |\lambda| \ge s_n$.

Problem 7E.16. Suppose $T \in \mathcal{L}(V, W)$. Prove that $(T^*)^\dagger = (T^\dagger)^*$.
Compare the result in this exercise to the analogous result for invertible linear maps [see 7.5(f)].

Solution. Both sides of the asserted equation are linear maps from V to W: indeed $T^* \in \mathcal{L}(W, V)$, so $(T^*)^\dagger \in \mathcal{L}(V, W)$; and $T^\dagger \in \mathcal{L}(W, V)$, so $(T^\dagger)^* \in \mathcal{L}(V, W)$. Thus the equation at least makes sense.
First dispose of the trivial case. Suppose $T = 0$. Then $T^* = 0$, because $\langle 0v, w\rangle = 0 = \langle v, 0w\rangle$ for all $v \in V$ and $w \in W$. Also $T^\dagger = 0$: indeed $\operatorname{null} T = V$, so $(\operatorname{null} T)^\perp = \{0\}$, and $\operatorname{range} T = \{0\}$, so $P_{\operatorname{range} T} = 0$; hence the defining formula $T^\dagger w = (T|_{(\operatorname{null} T)^\perp})^{-1} P_{\operatorname{range} T} w$ of 6.68 gives $T^\dagger w = 0$ for every $w \in W$. The same reasoning applied to the zero maps T^* and $T^\dagger$ yields $(T^*)^\dagger = 0$ and $(T^\dagger)^* = 0$. Hence $(T^*)^\dagger = (T^\dagger)^*$ in this case.
So assume $T \ne 0$ and let $s_1 \ge \cdots \ge s_m > 0$ be the positive singular values of T (there is at least one, by Exercise 1 in this section). By the singular value decomposition 7.70 there are orthonormal lists $e_1, \dots, e_m$ in V and $f_1, \dots, f_m$ in W such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \qquad \text{for every } v \in V.$$

This is the shape 7.76 required by the hypotheses of 7.75.
Step 1: a formula for $(T^*)^\dagger$. By 7.77 we have

$$T^*w = s_1\langle w, f_1\rangle e_1 + \cdots + s_m\langle w, f_m\rangle e_m \qquad \text{for every } w \in W.$$

This exhibits $T^* \in \mathcal{L}(W, V)$ in exactly the shape 7.76, with the decreasing positive scalars $s_1 \ge \cdots \ge s_m > 0$, the orthonormal list $f_1, \dots, f_m$ in the domain W, and the orthonormal list $e_1, \dots, e_m$ in the target V. Hence Exercise 8(c) of this section (or Exercise 9, which says that T and T^* have the same positive singular values) shows that $s_1, \dots, s_m$ are precisely the positive singular values of T^*. Therefore 7.78, applied to the map T^* with the roles of the e's and the f's interchanged, gives

$$(T^*)^\dagger v = \frac{\langle v, e_1\rangle}{s_1} f_1 + \cdots + \frac{\langle v, e_m\rangle}{s_m} f_m \qquad \text{for every } v \in V.$$

Step 2: a formula for $(T^\dagger)^*$. By 7.78 applied to T,

$$T^\dagger w = \frac{\langle w, f_1\rangle}{s_1} e_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m} e_m \qquad \text{for every } w \in W.$$

Now compute the adjoint of $T^\dagger$ directly from the definition of the adjoint (7.1), imitating the computation in the proof of 7.75. If $w \in W$ and $v \in V$, then

$$\begin{aligned}\langle T^\dagger w, v\rangle &= \Big\langle \sum_{k=1}^{m} \frac{\langle w, f_k\rangle}{s_k} e_k,\ v \Big\rangle \\ &= \sum_{k=1}^{m} \frac{\langle w, f_k\rangle}{s_k} \langle e_k, v\rangle \\ &= \sum_{k=1}^{m} \Big\langle w,\ \frac{\overline{\langle e_k, v\rangle}}{s_k} f_k \Big\rangle \\ &= \Big\langle w,\ \sum_{k=1}^{m} \frac{\langle v, e_k\rangle}{s_k} f_k \Big\rangle,\end{aligned}$$

where the second line uses homogeneity in the first slot, and the third line is justified as follows: for any scalar c we have $\langle w, cf_k\rangle = \overline{c}\,\langle w, f_k\rangle$ by conjugate homogeneity in the second slot, so taking $c = \overline{\langle e_k, v\rangle}/s_k$ (legitimate because s_k is a positive real number, so $\overline{c} = \langle e_k, v\rangle/s_k$) gives

$$\Big\langle w,\ \frac{\overline{\langle e_k, v\rangle}}{s_k} f_k \Big\rangle = \frac{\langle e_k, v\rangle}{s_k} \langle w, f_k\rangle,$$

which is the k-th term of the second line. The last line then uses $\overline{\langle e_k, v\rangle} = \langle v, e_k\rangle$ (conjugate symmetry of the inner product) together with additivity in the second slot. Because the adjoint $(T^\dagger)^*$ is the unique map satisfying $\langle T^\dagger w, v\rangle = \langle w, (T^\dagger)^* v\rangle$ for all $w \in W$ and $v \in V$ (7.1), we conclude

$$(T^\dagger)^* v = \frac{\langle v, e_1\rangle}{s_1} f_1 + \cdots + \frac{\langle v, e_m\rangle}{s_m} f_m \qquad \text{for every } v \in V.$$

Comparing the formula for $(T^*)^\dagger$ obtained in Step 1 with the formula for $(T^\dagger)^*$ obtained in Step 2 shows that $(T^*)^\dagger v = (T^\dagger)^* v$ for every $v \in V$, and hence

$$(T^*)^\dagger = (T^\dagger)^*,$$

as desired.
Comparison with 7.5(f). If T is invertible, then so is T^*, and 6.69(a) gives $T^\dagger = T^{-1}$ and $(T^*)^\dagger = (T^*)^{-1}$. In that case the equation just proved reads $(T^*)^{-1} = (T^{-1})^*$, which is exactly 7.5(f). Thus this exercise extends 7.5(f) from invertible linear maps to arbitrary linear maps, with the pseudoinverse in place of the inverse.

Problem 7E.17. Suppose $T \in \mathcal{L}(V)$. Prove that T is self-adjoint if and only if $T^\dagger$ is self-adjoint.

Solution. Because $T \in \mathcal{L}(V)$, the pseudoinverse $T^\dagger$ is also an operator on V (see 6.68), so it makes sense to ask whether $T^\dagger$ is self-adjoint.
First suppose T is self-adjoint, meaning $T^* = T$. By Exercise 16 of this section,

$$(T^\dagger)^* = (T^*)^\dagger = T^\dagger,$$

where the second equality holds because $T^* = T$. Thus $T^\dagger$ is self-adjoint.
For the other direction, suppose $T^\dagger$ is self-adjoint. Applying the implication just proved to the operator $T^\dagger$ in place of T, we conclude that $(T^\dagger)^\dagger$ is self-adjoint. But $(T^\dagger)^\dagger = T$ by Exercise 23 in Section 6C. Hence T is self-adjoint.
This completes both directions, so T is self-adjoint if and only if $T^\dagger$ is self-adjoint.
For completeness, here is a proof of the identity $(T^\dagger)^\dagger = T$ that was quoted above, phrased in the language of this section. If $T = 0$, then $T^\dagger = 0$ and the identity is clear. So assume $T \neq 0$, let $s_1 \geq \cdots \geq s_m > 0$ be the positive singular values of T, and let $e_1, \ldots, e_m$ and $f_1, \ldots, f_m$ be orthonormal

lists in V as in the singular value decomposition 7.70, so that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \qquad \text{for every } v \in V.$$

By 7.78,

$$T^\dagger w = \frac{\langle w, f_1\rangle}{s_1} e_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m} e_m \qquad \text{for every } w \in V.$$

Reindex by setting $t_k = 1/s_{m+1-k}$, $g_k = f_{m+1-k}$, and $h_k = e_{m+1-k}$ for $k = 1, \ldots, m$. Then $t_1 \geq \cdots \geq t_m > 0$, the lists $g_1, \ldots, g_m$ and $h_1, \ldots, h_m$ are orthonormal in V, and the displayed formula for $T^\dagger$ becomes

$$T^\dagger w = t_1\langle w, g_1\rangle h_1 + \cdots + t_m\langle w, g_m\rangle h_m \qquad \text{for every } w \in V,$$

which is exactly the shape 7.76 for the operator $T^\dagger$. Hence Exercise 8(c) shows that $t_1, \ldots, t_m$ are the positive singular values of $T^\dagger$, and 7.78 applied to $T^\dagger$ gives

$$(T^\dagger)^\dagger v = \frac{\langle v, h_1\rangle}{t_1} g_1 + \cdots + \frac{\langle v, h_m\rangle}{t_m} g_m = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m = Tv$$

for every $v \in V$, where the middle equality is just the reindexing read backwards. Thus $(T^\dagger)^\dagger = T$, as claimed.

7.6 Exercises 7F

Problem 7F.1. Prove that if $S, T \in \mathcal{L}(V, W)$, then $\big|\, \|S\| - \|T\| \,\big| \leq \|S - T\|$.
The inequality above is called the *reverse triangle inequality*.

Solution. Write $S = (S - T) + T$. The triangle inequality for the norm on $\mathcal{L}(V, W)$, which is 7.87(d), gives

$$\|S\| = \|(S - T) + T\| \leq \|S - T\| + \|T\|,$$

and therefore

$$\|S\| - \|T\| \leq \|S - T\|.$$

Interchanging the roles of S and T in the display above gives

$$\|T\| - \|S\| \leq \|T - S\|.$$

Now $T - S = (-1)(S - T)$, so 7.87(c) with $\lambda = -1$ gives $\|T - S\| = |-1|\,\|S - T\| = \|S - T\|$. Hence

$$\|T\| - \|S\| \leq \|S - T\|.$$

The number $\big|\, \|S\| - \|T\| \,\big|$ equals either $\|S\| - \|T\|$ or $\|T\| - \|S\|$, and both of these have just been shown to be at most $\|S - T\|$. Thus

$$\big|\, \|S\| - \|T\| \,\big| \leq \|S - T\|,$$

as desired.

Problem 7F.2. Suppose that $T \in \mathcal{L}(V)$ is self-adjoint or that $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Prove that

$$\|T\| = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } T\}.$$

Solution. In either of the two hypothesized cases the spectral theorem applies: if $\mathbf{F} = \mathbf{R}$ and T is self-adjoint, use the real spectral theorem 7.29; if $\mathbf{F} = \mathbf{C}$ and T is normal (which includes the case of a self-adjoint T on a complex inner product space), use the complex spectral theorem 7.31. Either way

there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T, say

$$Te_k = \lambda_k e_k \quad \text{for } k = 1, \dots, n.$$

First note that the set of eigenvalues of T is exactly $\{\lambda_1, \dots, \lambda_n\}$. Indeed each λ_k is an eigenvalue because $e_k \neq 0$. Conversely, if λ is an eigenvalue of T with eigenvector $v = a_1e_1 + \cdots + a_ne_n \neq 0$, then comparing coefficients in

$$\lambda_1 a_1 e_1 + \cdots + \lambda_n a_n e_n = Tv = \lambda v = \lambda a_1 e_1 + \cdots + \lambda a_n e_n$$

shows that $\lambda_k a_k = \lambda a_k$ for each k; choosing k with $a_k \neq 0$ (possible because $v \neq 0$) gives $\lambda = \lambda_k$. In particular the set of eigenvalues of T is finite and nonempty; here $n = \dim V \geq 1$ because the standing hypothesis of this chapter is that V is a nonzero finite-dimensional inner product space. Hence the maximum on the right side of the desired equation is a maximum of a finite nonempty set of nonnegative numbers and so exists. Put

$$M = \max\{|\lambda_1|, \dots, |\lambda_n|\}.$$

Now let $v \in V$ with $\|v\| \leq 1$, and write $v = a_1e_1 + \cdots + a_ne_n$. Since $e_1, \dots, e_n$ is orthonormal, $\|v\|^2 = |a_1|^2 + \cdots + |a_n|^2$ by 6.24. Also

$$Tv = \lambda_1 a_1 e_1 + \cdots + \lambda_n a_n e_n,$$

which is again a linear combination of $e_1, \dots, e_n$, now with coefficients $\lambda_1 a_1, \dots, \lambda_n a_n$. So 6.24 applies again and gives

$$\|Tv\|^2 = |\lambda_1|^2|a_1|^2 + \cdots + |\lambda_n|^2|a_n|^2 \leq M^2\big(|a_1|^2 + \cdots + |a_n|^2\big) = M^2\|v\|^2 \leq M^2.$$

Thus $\|Tv\| \leq M$ for every $v \in V$ with $\|v\| \leq 1$, and hence $\|T\| \leq M$ by the definition 7.86 of the norm of a linear map.
For the reverse inequality, choose k with $|\lambda_k| = M$. The vector e_k satisfies $\|e_k\| = 1$, so it is one of the vectors competing in the maximum defining $\|T\|$, and

$$\|Te_k\| = \|\lambda_k e_k\| = |\lambda_k|\,\|e_k\| = M,$$

so the maximum defining $\|T\|$ is at least M; that is, $\|T\| \geq M$.
Combining the two inequalities gives $\|T\| = M$, which is exactly the asserted equation.

Problem 7F.3. Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$. Prove that

$$\|Tv\| = \|T\|\,\|v\| \iff T^*Tv = \|T\|^2 v.$$

Solution. Throughout write $c = \|T\|$, a nonnegative real number.
We record one identity used in both directions. The defining property of the adjoint (7.1) reads $\langle Tx, w\rangle = \langle x, T^*w\rangle$ for $x \in V$ and $w \in W$. Taking $x = v$ and $w = Tv$ gives

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle v, T^*Tv\rangle.$$

The left side is a nonnegative real number, so the right side is real and equals its own complex conjugate; by conjugate symmetry of the inner product,

$$\langle T^*Tv, v\rangle = \overline{\langle v, T^*Tv\rangle} = \|Tv\|^2.$$

First suppose $T^*Tv = c^2 v$. Then, by the identity above and conjugate linearity of the inner product in its second slot,

$$\|Tv\|^2 = \langle v, T^*Tv\rangle = \langle v, c^2 v\rangle = \overline{c^2}\,\|v\|^2 = c^2\|v\|^2,$$

the last step because $c = \|T\|$ is a real number. Both $\|Tv\|$ and $c\|v\|$ are nonnegative, so taking square roots gives $\|Tv\| = c\|v\| = \|T\|\,\|v\|$.

Conversely, suppose $\|Tv\| = c\|v\|$. We estimate $\|T^*Tv - c^2v\|^2$ and show it is 0. Expanding,

$$\begin{aligned}\|T^*Tv - c^2v\|^2 &= \langle T^*Tv - c^2v,\ T^*Tv - c^2v\rangle \\ &= \|T^*Tv\|^2 - \langle T^*Tv, c^2v\rangle - \langle c^2v, T^*Tv\rangle + c^4\|v\|^2.\end{aligned}$$

Consider the two middle terms. Conjugate linearity in the second slot pulls out $\overline{c^2} = c^2$, since c^2 is real, and then the identity $\langle T^*Tv, v\rangle = \|Tv\|^2$ recorded above applies:

$$\langle T^*Tv, c^2v\rangle = c^2\langle T^*Tv, v\rangle = c^2\|Tv\|^2 = c^2 \cdot c^2\|v\|^2 = c^4\|v\|^2,$$

where the hypothesis $\|Tv\| = c\|v\|$ was used in the second-to-last step. This quantity is real, so its complex conjugate $\langle c^2v, T^*Tv\rangle$ equals $c^4\|v\|^2$ as well. Hence the two middle terms together contribute $-2c^4\|v\|^2$.
Now consider the first term. Note $T^* \in \mathcal{L}(W, V)$ and $Tv \in W$, so the inequality 7.89 applies to T^* and the vector Tv; then 7.91 (a linear map and its adjoint have the same norm) replaces $\|T^*\|$ by $\|T\| = c$, and the hypothesis replaces $\|Tv\|$ by $c\|v\|$:

$$\|T^*Tv\| \le \|T^*\|\,\|Tv\| = \|T\|\,\|Tv\| = c \cdot c\|v\| = c^2\|v\|.$$

Thus $\|T^*Tv\|^2 \le c^4\|v\|^2$.
Putting the three estimates together,

$$\|T^*Tv - c^2v\|^2 \le c^4\|v\|^2 - 2c^4\|v\|^2 + c^4\|v\|^2 = 0.$$

A norm is nonnegative, so $\|T^*Tv - c^2v\| = 0$, which by the definiteness of the norm gives

$$T^*Tv = c^2v = \|T\|^2v.$$

This proves both implications, so the two conditions are equivalent.

Problem 7F.4. Suppose $T \in \mathcal{L}(V, W)$, $v \in V$, and $\|Tv\| = \|T\|\,\|v\|$. Prove that if $u \in V$ and $\langle u, v\rangle = 0$, then $\langle Tu, Tv\rangle = 0$.

Solution. The hypothesis $\|Tv\| = \|T\|\,\|v\|$ is exactly the left side of the equivalence proved in Exercise 3 of this section. Hence that exercise gives

$$T^*Tv = \|T\|^2v.$$

Now suppose $u \in V$ with $\langle u, v\rangle = 0$. Using the defining property of the adjoint (7.1) and then the equation above,

$$\langle Tu, Tv\rangle = \langle u, T^*Tv\rangle = \big\langle u, \|T\|^2v\big\rangle = \overline{\|T\|^2}\,\langle u, v\rangle = \|T\|^2\langle u, v\rangle = 0,$$

where the conjugate could be dropped because $\|T\|^2$ is a nonnegative real number, and the last equality used $\langle u, v\rangle = 0$.
Thus $\langle Tu, Tv\rangle = 0$, as desired.
Geometrically, this says that a vector v at which T attains its norm (in the scaled sense $\|Tv\| = \|T\|\,\|v\|$) has the property that T carries the entire orthogonal complement of v into the orthogonal complement of Tv.

Problem 7F.5. Suppose U is a finite-dimensional inner product space, $T \in \mathcal{L}(V, U)$, and $S \in \mathcal{L}(U, W)$. Prove that

$$\|ST\| \le \|S\|\,\|T\|.$$

Solution. Note first that $ST \in \mathcal{L}(V, W)$, so $\|ST\|$ is defined.
Let $v \in V$. Applying the inequality 7.89 to the linear map $S \in \mathcal{L}(U, W)$ and the vector $Tv \in U$ gives

$$\|(ST)v\| = \|S(Tv)\| \leq \|S\| \, \|Tv\|.$$

Applying 7.89 again, this time to $T \in \mathcal{L}(V, U)$ and the vector v, gives $\|Tv\| \leq \|T\| \, \|v\|$. Since $\|S\| \geq 0$ by 7.87(a), we may multiply this inequality by $\|S\|$ and combine with the display above to obtain

$$\|(ST)v\| \leq \|S\| \, \|T\| \, \|v\|.$$

The inequality above holds for every $v \in V$. Thus the nonnegative number $c = \|S\| \, \|T\|$ satisfies $\|(ST)v\| \leq c\|v\|$ for all $v \in V$. By 7.88(c), $\|ST\|$ is the *smallest* such number c, and in particular

$$\|ST\| \leq \|S\| \, \|T\|,$$

as desired.
(Alternatively, one can argue directly from the definition 7.86: for every $v \in V$ with $\|v\| \leq 1$, the display above gives $\|(ST)v\| \leq \|S\| \, \|T\|$, so the maximum of $\|(ST)v\|$ over such v, which is $\|ST\|$, is at most $\|S\| \, \|T\|$.)

Problem 7F.6. Prove or give a counterexample: If $S, T \in \mathcal{L}(V)$, then $\|ST\| = \|TS\|$.

Solution. The statement is false. Here is a counterexample.
Take $V = \mathbf{F}^2$ with its usual inner product, and let e_1, e_2 denote the standard basis of $\mathbf{F}^2$, which is orthonormal. Define $S, T \in \mathcal{L}(\mathbf{F}^2)$ by

$$S(a_1, a_2) = (a_2, 0), \qquad T(a_1, a_2) = (a_1, 0).$$

Equivalently, $Se_1 = 0$, $Se_2 = e_1$, and $Te_1 = e_1$, $Te_2 = 0$; with respect to the standard basis the matrices are

$$\mathcal{M}(S) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \mathcal{M}(T) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Compute ST. For every $(a_1, a_2) \in \mathbf{F}^2$,

$$(ST)(a_1, a_2) = S(a_1, 0) = (0, 0).$$

Thus $ST = 0$, and hence $\|ST\| = 0$ by 7.87(b).
Compute TS. For every $(a_1, a_2) \in \mathbf{F}^2$,

$$(TS)(a_1, a_2) = T(a_2, 0) = (a_2, 0).$$

Thus $\|(TS)(a_1, a_2)\| = |a_2| \leq \|(a_1, a_2)\|$ for all (a_1, a_2), so $\|TS\| \leq 1$ by the definition 7.86. Moreover $(TS)e_2 = e_1$, so $\|(TS)e_2\| = 1$ with $\|e_2\| = 1$, giving $\|TS\| \geq 1$. Hence $\|TS\| = 1$.
Therefore

$$\|ST\| = 0 \neq 1 = \|TS\|,$$

so the assertion in the exercise is false.
The same example works over $\mathbf{R}$ and over $\mathbf{C}$, and it can be transplanted into any V with $\dim V \geq 2$: pick orthonormal $e_1, e_2 \in V$, define S and T on $\operatorname{span}(e_1, e_2)$ by the formulas above and as 0 on $(\operatorname{span}(e_1, e_2))^{\perp}$. Then $ST = 0$ while $(TS)e_2 = e_1 \neq 0$, so again $\|ST\| = 0 \neq \|TS\|$. Since V is assumed nonzero, the only case left is $\dim V = 1$, where $\mathcal{L}(V)$ is commutative and the asserted equality does hold; but as a general statement about all V it fails.

Problem 7F.7. Show that defining $d(S,T) = \|S - T\|$ for $S, T \in \mathcal{L}(V,W)$ makes d a metric on $\mathcal{L}(V,W)$.
This exercise is intended for readers who are familiar with metric spaces.

Solution. Recall that d is a metric on a set X if $d : X \times X \to \mathbf{R}$ satisfies, for all $x, y, z \in X$: (i) $d(x,y) \geq 0$; (ii) $d(x,y) = 0$ if and only if $x = y$; (iii) $d(x,y) = d(y,x)$; (iv) $d(x,z) \leq d(x,y) + d(y,z)$.
First, d really is a function into $\mathbf{R}$: for $S, T \in \mathcal{L}(V,W)$ the difference $S - T$ lies in $\mathcal{L}(V,W)$, and the discussion following 7.86 shows that the maximum defining $\|S-T\|$ exists (it equals the largest singular value of $S - T$), so $\|S - T\|$ is a well-defined real number. We now verify the four properties, each of which comes directly from 7.87.
(i) Nonnegativity. By 7.87(a) applied to the linear map $S - T$,

$$d(S,T) = \|S - T\| \geq 0.$$

(ii) Definiteness. By 7.87(b) applied to $S - T$,

$$d(S,T) = 0 \iff \|S - T\| = 0 \iff S - T = 0 \iff S = T.$$

(iii) Symmetry. Since $T - S = (-1)(S - T)$, property 7.87(c) with $\lambda = -1$ gives

$$d(T,S) = \|T - S\| = |-1|\,\|S - T\| = \|S - T\| = d(S,T).$$

(iv) Triangle inequality. Let $S, T, U \in \mathcal{L}(V,W)$. Then $S - U = (S - T) + (T - U)$, so the triangle inequality 7.87(d) for the norm on $\mathcal{L}(V,W)$ gives

$$d(S,U) = \|S - U\| = \|(S - T) + (T - U)\| \leq \|S - T\| + \|T - U\| = d(S,T) + d(T,U).$$

All four properties hold, so d is a metric on $\mathcal{L}(V,W)$.

Problem 7F.8. (a) Prove that if $T \in \mathcal{L}(V)$ and $\|I - T\| < 1$, then T is invertible.
(b) Suppose that $S \in \mathcal{L}(V)$ is invertible. Prove that if $T \in \mathcal{L}(V)$ and $\|S - T\| < 1/\|S^{-1}\|$, then T is invertible.
This exercise shows that the set of invertible operators in $\mathcal{L}(V)$ is an open subset of $\mathcal{L}(V)$, using the metric defined in Exercise 7.

Solution. Throughout, we use the inequality 7.89, which states that $\|Rv\| \leq \|R\|\,\|v\|$ for all $R \in \mathcal{L}(V)$ and all $v \in V$.
(a) Suppose $T \in \mathcal{L}(V)$ and $\|I - T\| < 1$. To show that T is invertible it suffices, because V is finite-dimensional, to show that T is injective (by 3.65, an injective operator on a finite-dimensional vector space is also surjective, hence invertible).
So suppose $v \in V$ and $Tv = 0$. Then

$$\|v\| = \|v - Tv\| = \|(I - T)v\| \leq \|I - T\|\,\|v\|.$$

If $v \neq 0$, then $\|v\| > 0$ and we may divide by $\|v\|$, obtaining $1 \leq \|I - T\|$, which contradicts the hypothesis $\|I - T\| < 1$. Hence $v = 0$. Thus $\operatorname{null} T = \{0\}$, so T is injective and therefore invertible.
(b) We first record the submultiplicativity of the operator norm (this is Exercise 5 in this section, but the proof is short enough to include). If $R, Q \in \mathcal{L}(V)$ and $v \in V$, then two applications of 7.89 give

$$\|(RQ)v\| = \|R(Qv)\| \leq \|R\|\,\|Qv\| \leq \|R\|\,\|Q\|\,\|v\|.$$

Because $\|R\|\,\|Q\|$ is thus a number c with $\|(RQ)v\| \leq c\|v\|$ for all $v \in V$, and because $\|RQ\|$ is the smallest such number (by 7.88(c)), we conclude that

$$\|RQ\| \leq \|R\|\,\|Q\|.$$

Now suppose $S \in \mathcal{L}(V)$ is invertible and $T \in \mathcal{L}(V)$ satisfies $\|S - T\| < 1/\|S^{-1}\|$. Note that this hypothesis makes sense: because $V \neq \{0\}$, the operator S^{-1} is not the zero operator, so $\|S^{-1}\| > 0$ by 7.87(b).
We have

$$I - S^{-1}T = S^{-1}S - S^{-1}T = S^{-1}(S - T).$$

Hence, by the submultiplicativity proved above,

$$\|I - S^{-1}T\| = \|S^{-1}(S - T)\| \leq \|S^{-1}\| \, \|S - T\| < \|S^{-1}\| \cdot \frac{1}{\|S^{-1}\|} = 1.$$

Thus part (a), applied to the operator $S^{-1}T$, shows that $S^{-1}T$ is invertible. Therefore

$$T = S\big(S^{-1}T\big)$$

is the composition of two invertible operators, and hence T is invertible (with $T^{-1} = (S^{-1}T)^{-1}S^{-1}$).
Finally, we note the geometric content. Fix an invertible $S \in \mathcal{L}(V)$ and set $r = 1/\|S^{-1}\| > 0$. Part (b) says that every $T \in \mathcal{L}(V)$ with $d(S, T) = \|S - T\| < r$ is invertible. In other words, the open ball of radius r centered at S, in the metric d of Exercise 7, is contained in the set of invertible operators. Because every point of that set has such a ball around it, the set of invertible operators is an open subset of $\mathcal{L}(V)$.

Problem 7F.9. Suppose $T \in \mathcal{L}(V)$. Prove that for every $\epsilon > 0$, there exists an invertible operator $S \in \mathcal{L}(V)$ such that $0 < \|T - S\| < \epsilon$.

Solution. Let $\epsilon > 0$.
By 5.12, the operator T has at most $\dim V$ distinct eigenvalues; in particular, the set of eigenvalues of T is a finite subset of $\mathbf{F}$. The open interval $(0, \epsilon)$ is an infinite subset of $\mathbf{R} \subseteq \mathbf{F}$. Hence there exists

$$\lambda \in (0, \epsilon)$$

such that λ is not an eigenvalue of T.
Set

$$S = T - \lambda I.$$

Because λ is not an eigenvalue of T, the operator $T - \lambda I$ is injective, and therefore invertible (by 3.65, injectivity implies invertibility for an operator on a finite-dimensional vector space). Thus $S \in \mathcal{L}(V)$ is invertible.
Finally we compute $\|T - S\|$. We have $T - S = \lambda I$, so by 7.87(c) and the fact that $\|I\| = 1$ (see 7.90),

$$\|T - S\| = \|\lambda I\| = |\lambda| \, \|I\| = \lambda.$$

Since $\lambda \in (0, \epsilon)$, we conclude that

$$0 < \|T - S\| < \epsilon,$$

as desired.
(For completeness: $\|I\| = 1$ because $\|Iv\| = \|v\| \leq 1$ for every $v \in V$ with $\|v\| \leq 1$, with equality attained at any v with $\|v\| = 1$; such a v exists because $V \neq \{0\}$.)
Thus every operator on V, invertible or not, has invertible operators arbitrarily close to it but distinct from it. In the language of Exercise 7, the set of invertible operators is dense in $\mathcal{L}(V)$.

Problem 7F.10. Suppose $\dim V > 1$ and $T \in \mathcal{L}(V)$ is not invertible. Prove that for every $\epsilon > 0$, there exists $S \in \mathcal{L}(V)$ such that $0 < \|T - S\| < \epsilon$ and S is not invertible.

Solution. Let $\epsilon > 0$ and set $\delta = \epsilon/2$, so $0 < \delta < \epsilon$.
Because T is not invertible and V is finite-dimensional, T is not injective (by 3.65 an injective operator

on V would be invertible). Hence $\operatorname{null} T \neq \{0\}$, so we may choose $u \in \operatorname{null} T$ with $\|u\| = 1$.
Because $\dim V > 1$, the orthogonal complement of $\operatorname{span}(u)$ satisfies

$$\dim\big(\operatorname{span}(u)\big)^{\perp} = \dim V - \dim \operatorname{span}(u) = \dim V - 1 \geq 1$$

by 6.51. Hence $(\operatorname{span}(u))^{\perp} \neq \{0\}$, so we may choose $w \in V$ with $\|w\| = 1$ and $\langle w, u\rangle = 0$ (and hence also $\langle u, w\rangle = 0$).
Now define $S \in \mathcal{L}(V)$ by

$$Sv = Tv + \delta\langle v, w\rangle w$$

for each $v \in V$. This map is linear because the inner product is linear in its first slot, so indeed $S \in \mathcal{L}(V)$.
We first check that S is not invertible. We have

$$Su = Tu + \delta\langle u, w\rangle w = 0 + \delta \cdot 0 \cdot w = 0,$$

using $u \in \operatorname{null} T$ and $\langle u, w\rangle = 0$. Because $u \neq 0$, the operator S is not injective, hence not invertible.
Next we compute the distance from T to S. For every $v \in V$ we have $(S - T)v = \delta\langle v, w\rangle w$, so

$$\|(S - T)v\| = \delta\,|\langle v, w\rangle|\,\|w\| = \delta\,|\langle v, w\rangle| \leq \delta\|v\|\,\|w\| = \delta\|v\|,$$

where the inequality is the Cauchy-Schwarz inequality 6.14. Thus $\|S - T\| \leq \delta$ by 7.88(c). On the other hand, $\|w\| = 1$ and

$$(S - T)w = \delta\langle w, w\rangle w = \delta w,$$

so $\|(S - T)w\| = \delta$, which gives $\|S - T\| \geq \delta$. Hence $\|S - T\| = \delta$.
Since $\|T - S\| = \|-(S - T)\| = \|S - T\| = \delta$ by 7.87(c), we conclude that

$$0 < \|T - S\| = \delta < \epsilon$$

and S is not invertible, as desired.
Note where the hypothesis $\dim V > 1$ was used: it guaranteed a unit vector w orthogonal to u, which is what lets the perturbation $\delta\langle \cdot\,, w\rangle w$ be nonzero while still annihilating u. The hypothesis cannot be dropped: if $\dim V = 1$, then the only operator on V that is not invertible is 0, so no S as in the exercise exists.

Problem 7F.11. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that for every $\epsilon > 0$ there exists a diagonalizable operator $S \in \mathcal{L}(V)$ such that $0 < \|T - S\| < \epsilon$.

Solution. Let $\epsilon > 0$ and let $n = \dim V$.
Because $\mathbf{F} = \mathbf{C}$, Schur's theorem 6.38 provides an orthonormal basis $e_1, \ldots, e_n$ of V with respect to which T has an upper-triangular matrix. Write

$$\mathcal{M}(T, (e_1, \ldots, e_n)) = A,$$

where A is upper triangular; denote its diagonal entries by $\lambda_1, \ldots, \lambda_n$.
We now choose perturbations of the diagonal entries. We claim there exist $\delta_1, \ldots, \delta_n \in \mathbf{C}$ such that

$$0 < |\delta_k| < \epsilon \quad \text{for each } k, \qquad \text{and} \qquad \lambda_1 + \delta_1, \ldots, \lambda_n + \delta_n \text{ are distinct.}$$

Choose them inductively. Let $D = \{\delta \in \mathbf{C} : 0 < |\delta| < \epsilon\}$, an infinite set. Suppose $1 \leq k \leq n$ and $\delta_1, \ldots, \delta_{k-1} \in D$ have already been chosen. For each $j \in \{1, \ldots, k-1\}$ there is exactly one complex number δ with $\lambda_k + \delta = \lambda_j + \delta_j$, namely $\delta = \lambda_j + \delta_j - \lambda_k$. Thus at most $k - 1$ elements of D are forbidden, and since D is infinite we may choose $\delta_k \in D$ with $\lambda_k + \delta_k \neq \lambda_j + \delta_j$ for all $j < k$. This completes the inductive construction.
Let $S \in \mathcal{L}(V)$ be the operator whose matrix with respect to $e_1, \ldots, e_n$ is obtained from A by replacing the diagonal entry λ_k with $\lambda_k + \delta_k$ for each k, leaving all other entries unchanged. (An operator is determined by its matrix with respect to a basis, so S is well defined.)

The operator S is diagonalizable. Indeed, $\mathcal{M}(S,(e_1,\dots,e_n))$ is upper triangular, because we changed only diagonal entries of the upper-triangular matrix A. Hence by 5.41 the eigenvalues of S are exactly the diagonal entries $\lambda_1+\delta_1,\dots,\lambda_n+\delta_n$ of this matrix. By construction these n numbers are distinct, so S has $n=\dim V$ distinct eigenvalues. Therefore S is diagonalizable by 5.58.
It remains to compute $\|T-S\|$. The matrix of $T-S$ with respect to $e_1,\dots,e_n$ is A minus the matrix of S; all off-diagonal entries cancel, and the k-th diagonal entry is $\lambda_k-(\lambda_k+\delta_k)=-\delta_k$. Thus

$$(T-S)e_k=-\delta_k e_k \quad \text{for } k=1,\dots,n.$$

Let $m\in\{1,\dots,n\}$ be such that $|\delta_m|=\max\{|\delta_1|,\dots,|\delta_n|\}$. If $v\in V$, write $v=a_1e_1+\cdots+a_ne_n$; then

$$\|(T-S)v\|^2=\|-\delta_1a_1e_1-\cdots-\delta_na_ne_n\|^2=\sum_{k=1}^n|\delta_k|^2|a_k|^2\le|\delta_m|^2\sum_{k=1}^n|a_k|^2=|\delta_m|^2\|v\|^2,$$

where we used that $e_1,\dots,e_n$ is orthonormal. Hence $\|T-S\|\le|\delta_m|$ by 7.88(c). Since $\|e_m\|=1$ and $\|(T-S)e_m\|=|\delta_m|$, we also have $\|T-S\|\ge|\delta_m|$. Therefore

$$\|T-S\|=|\delta_m|=\max\{|\delta_1|,\dots,|\delta_n|\}.$$

Because $0<|\delta_k|<\epsilon$ for every k, this gives

$$0<\|T-S\|<\epsilon,$$

with S diagonalizable, as desired.
Thus the diagonalizable operators are dense in $\mathcal{L}(V)$ when $\mathbf{F}=\mathbf{C}$: every operator on a finite-dimensional complex inner product space can be approximated arbitrarily well (but not exactly, by the requirement $\|T-S\|>0$) by diagonalizable operators.

Problem 7F.12. Suppose $T\in\mathcal{L}(V)$ is a positive operator. Show that $\|\sqrt{T}\|=\sqrt{\|T\|}$.

Solution. We first record a lemma that we use twice.
Lemma. Suppose $A\in\mathcal{L}(V)$ and $e_1,\dots,e_n$ is an orthonormal basis of V with $Ae_k=\mu_ke_k$ for each k. Then $\|A\|=\max\{|\mu_1|,\dots,|\mu_n|\}$.
Proof of lemma. Let $\mu=\max\{|\mu_1|,\dots,|\mu_n|\}$ and choose m with $|\mu_m|=\mu$. If $v\in V$, write $v=a_1e_1+\cdots+a_ne_n$. Then $Av=\mu_1a_1e_1+\cdots+\mu_na_ne_n$, and because $e_1,\dots,e_n$ is orthonormal,

$$\|Av\|^2=\sum_{k=1}^n|\mu_k|^2|a_k|^2\le\mu^2\sum_{k=1}^n|a_k|^2=\mu^2\|v\|^2.$$

Hence $\|A\|\le\mu$ by 7.88(c). Since $\|e_m\|=1$ and $\|Ae_m\|=|\mu_m|=\mu$, the definition 7.86 gives $\|A\|\ge\mu$. Thus $\|A\|=\mu$, proving the lemma. (This is the third bullet point of Example 7.90.)
Now suppose $T\in\mathcal{L}(V)$ is a positive operator. By 7.38, the equivalence of (a) and (c) gives an orthonormal basis $e_1,\dots,e_n$ of V with respect to which the matrix of T is diagonal with nonnegative diagonal entries. Call those entries $\lambda_1,\dots,\lambda_n$, so

$$Te_k=\lambda_ke_k \quad \text{with } \lambda_k\ge0, \qquad k=1,\dots,n.$$

Define $R\in\mathcal{L}(V)$ to be the operator determined by

$$Re_k=\sqrt{\lambda_k}\,e_k, \qquad k=1,\dots,n.$$

The matrix of R with respect to the orthonormal basis $e_1,\dots,e_n$ is diagonal with the nonnegative entries $\sqrt{\lambda_1},\dots,\sqrt{\lambda_n}$ on its diagonal, so R is a positive operator, again by the equivalence of (a) and (c) in 7.38. Moreover

$$R^2e_k=R(\sqrt{\lambda_k}\,e_k)=\lambda_ke_k=Te_k$$

for each k, and since $e_1, \dots, e_n$ is a basis of V this shows $R^2 = T$. Thus R is a positive square root of T. By 7.39 the positive square root of T is unique, so

$$\sqrt{T} = R.$$

Now apply the lemma twice. Applying it to T (whose eigenvalues along the orthonormal basis are the nonnegative numbers λ_k) gives

$$\|T\| = \max\{\lambda_1, \dots, \lambda_n\},$$

and applying it to $\sqrt{T} = R$ gives

$$\|\sqrt{T}\| = \max\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}\}.$$

Because $t \mapsto \sqrt{t}$ is increasing on $[0, \infty)$, the maximum of $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$ is the square root of the maximum of $\lambda_1, \dots, \lambda_n$. Hence

$$\|\sqrt{T}\| = \sqrt{\max\{\lambda_1, \dots, \lambda_n\}} = \sqrt{\|T\|},$$

as desired.
Alternative finish: once one knows Exercise 19 of this section, which states $\|A^*A\| = \|A\|^2$, take $A = \sqrt{T}$. Since $\sqrt{T}$ is self-adjoint and $(\sqrt{T})^2 = T$, that identity reads $\|T\| = \|\sqrt{T}\|^2$, which is the desired conclusion after taking square roots.

Problem 7F.13. Suppose $S, T \in \mathcal{L}(V)$ are positive operators. Show that

$$\|S - T\| \le \max\{\|S\|, \|T\|\} \le \|S + T\|.$$

Solution. We begin with a lemma about self-adjoint operators.
Lemma. Suppose $A \in \mathcal{L}(V)$ is self-adjoint. Then there exists $v \in V$ with $\|v\| = 1$ and $|\langle Av, v\rangle| = \|A\|$; moreover $|\langle Au, u\rangle| \le \|A\|$ for every $u \in V$ with $\|u\| = 1$.
Proof of lemma. The second assertion holds for every operator: if $\|u\| = 1$, then the Cauchy-Schwarz inequality 6.14 and 7.89 give

$$|\langle Au, u\rangle| \le \|Au\| \, \|u\| \le \|A\| \, \|u\|^2 = \|A\|.$$

For the first assertion, use the spectral theorem: if $\mathbf{F} = \mathbf{R}$ this is 7.29, and if $\mathbf{F} = \mathbf{C}$ a self-adjoint operator is normal, so 7.31 applies. Either way there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of A, say $Ae_k = \mu_k e_k$; the eigenvalues $\mu_1, \dots, \mu_n$ are real because A is self-adjoint (by 7.12 when $\mathbf{F} = \mathbf{C}$; automatically when $\mathbf{F} = \mathbf{R}$).
For $v = a_1 e_1 + \cdots + a_n e_n$ we have, by orthonormality,

$$\|Av\|^2 = \sum_{k=1}^{n} |\mu_k|^2 |a_k|^2 \le \Big(\max_k |\mu_k|\Big)^2 \|v\|^2,$$

so $\|A\| \le \max_k |\mu_k|$ by 7.88(c); and taking $v = e_m$ where $|\mu_m| = \max_k |\mu_k|$ shows $\|A\| \ge |\mu_m|$. Hence $\|A\| = \max_k |\mu_k| = |\mu_m|$. Finally $\|e_m\| = 1$ and

$$|\langle Ae_m, e_m\rangle| = |\mu_m| \, \|e_m\|^2 = \|A\|,$$

so $v = e_m$ works. This proves the lemma.
We also record that if $R \in \mathcal{L}(V)$ is positive and $u \in V$ with $\|u\| = 1$, then

$$0 \le \langle Ru, u\rangle \le \|R\|,$$

the left inequality by the definition 7.34 of a positive operator and the right inequality by the lemma (a positive operator is self-adjoint, so $\langle Ru, u\rangle$ is a nonnegative real number and equals its own absolute value).

Proof of the first inequality. The operator $S-T$ is self-adjoint, since $(S-T)^* = S^* - T^* = S - T$. By the lemma there exists $v \in V$ with $\|v\| = 1$ and

$$\|S-T\| = |\langle (S-T)v, v\rangle| = |\langle Sv, v\rangle - \langle Tv, v\rangle|.$$

Put $a = \langle Sv, v\rangle$ and $b = \langle Tv, v\rangle$. By the remark above, a and b are real numbers with $0 \le a \le \|S\|$ and $0 \le b \le \|T\|$. Hence

$$a - b \le a \le \|S\| \qquad \text{and} \qquad b - a \le b \le \|T\|,$$

so $|a-b| \le \max\{\|S\|, \|T\|\}$. Therefore

$$\|S-T\| \le \max\{\|S\|, \|T\|\}.$$

Proof of the second inequality. Both the quantity $\max\{\|S\|, \|T\|\}$ and the quantity $\|S+T\|$ are unchanged if S and T are interchanged, so we may assume without loss of generality that $\max\{\|S\|, \|T\|\} = \|S\|$.
Because S is self-adjoint, the lemma provides $v \in V$ with $\|v\| = 1$ and $|\langle Sv, v\rangle| = \|S\|$. Because S is positive, $\langle Sv, v\rangle \ge 0$, and therefore

$$\langle Sv, v\rangle = \|S\|.$$

Now, using the definition 7.86 of the operator norm (valid since $\|v\| \le 1$), then the Cauchy-Schwarz inequality 6.14, and then positivity of T:

$$\|S+T\| \ge \|(S+T)v\| \ge |\langle (S+T)v, v\rangle| = \langle Sv, v\rangle + \langle Tv, v\rangle \ge \langle Sv, v\rangle = \|S\|.$$

(The equality in the middle holds because $\langle Sv, v\rangle$ and $\langle Tv, v\rangle$ are both nonnegative reals, so their sum equals its own absolute value.) Hence

$$\max\{\|S\|, \|T\|\} = \|S\| \le \|S+T\|,$$

which completes the proof.

Problem 7F.14. Suppose U and W are subspaces of V such that $\|P_U - P_W\| < 1$. Prove that $\dim U = \dim W$.

Solution. Here P_U and P_W denote the orthogonal projections of V onto U and onto W (see 6.55); these are operators on V by 6.57(a), and they are defined because V is finite-dimensional, hence so are U and W.
First note that

$$\|P_W - P_U\| = \|(-1)(P_U - P_W)\| = |-1|\,\|P_U - P_W\| = \|P_U - P_W\| < 1$$

by 7.87(c). So the hypothesis is symmetric in U and W.
The key step is the following claim.
Claim. The restriction of P_U to W, regarded as a linear map from W to U, is injective.
Note first that this restriction really does map W into U, because $\operatorname{range} P_U = U$ by 6.57(d). To prove injectivity it suffices to show that its null space is $\{0\}$.
So suppose $w \in W$ and $P_U w = 0$. Because $w \in W$, we have $P_W w = w$ by 6.57(b). Hence

$$\|w\| = \|P_W w\| = \|P_W w - P_U w\| = \|(P_W - P_U)w\| \le \|P_W - P_U\|\,\|w\|,$$

where the last step is 7.89. If $w \ne 0$, then $\|w\| > 0$ and dividing by $\|w\|$ gives

$$1 \le \|P_W - P_U\| = \|P_U - P_W\| < 1,$$

a contradiction. Therefore $w = 0$, proving the claim.

Because $P_U|_W : W \to U$ is injective, the fundamental theorem of linear maps 3.21 gives

$$\dim W = \dim \operatorname{null}\big(P_U|_W\big) + \dim \operatorname{range}\big(P_U|_W\big) = 0 + \dim \operatorname{range}\big(P_U|_W\big) \le \dim U,$$

the last inequality because $\operatorname{range}(P_U|_W)$ is a subspace of U.
Now interchange the roles of U and W. Since $\|P_W - P_U\| < 1$ as shown above, the same argument (with U and W swapped) shows that $P_W|_U : U \to W$ is injective and hence

$$\dim U \le \dim W.$$

Combining the two inequalities yields $\dim U = \dim W$, as desired.
Remark. The hypothesis cannot be weakened to $\|P_U - P_W\| \le 1$. For example, take $V = \mathbf{F}^2$, $U = \{0\}$, and $W = \operatorname{span}\big((1,0)\big)$. Then $P_U = 0$ and P_W is the orthogonal projection onto the first coordinate axis, so $\|P_U - P_W\| = \|P_W\| = 1$, yet $\dim U = 0 \ne 1 = \dim W$.

Problem 7F.15. Define $T \in \mathcal{L}(\mathbf{F}^3)$ by

$$T(z_1, z_2, z_3) = (z_3, 2z_1, 3z_2).$$

Find (explicitly) a unitary operator $S \in \mathcal{L}(\mathbf{F}^3)$ such that $T = S\sqrt{T^*T}$.

Solution. Let e_1, e_2, e_3 denote the standard basis of $\mathbf{F}^3$, which is an orthonormal basis with respect to the usual (Euclidean) inner product. Reading off the definition of T,

$$Te_1 = (0,2,0) = 2e_2, \qquad Te_2 = (0,0,3) = 3e_3, \qquad Te_3 = (1,0,0) = e_1.$$

First we compute $\sqrt{T^*T}$. For $j, k \in \{1,2,3\}$ we have

$$\langle T^*Te_j, e_k\rangle = \langle Te_j, Te_k\rangle,$$

by the definition of the adjoint. The list Te_1, Te_2, Te_3 equals $2e_2, 3e_3, e_1$, which is an orthogonal list; hence $\langle Te_j, Te_k\rangle = 0$ when $j \ne k$, and $\langle Te_j, Te_j\rangle = \|Te_j\|^2$, which equals $4, 9, 1$ for $j = 1, 2, 3$. Because e_1, e_2, e_3 is an orthonormal basis, a vector is determined by its inner products with e_1, e_2, e_3 (see 6.30). Therefore

$$T^*Te_1 = 4e_1, \qquad T^*Te_2 = 9e_2, \qquad T^*Te_3 = e_3,$$

that is, $T^*T(z_1, z_2, z_3) = (4z_1, 9z_2, z_3)$.
Now define $R \in \mathcal{L}(\mathbf{F}^3)$ by $R(z_1, z_2, z_3) = (2z_1, 3z_2, z_3)$. Then R has a diagonal matrix with respect to the orthonormal basis e_1, e_2, e_3, with diagonal entries $2, 3, 1$, all of which are nonnegative; hence R is a positive operator, by the implication (c) $\Rightarrow$ (a) in 7.38. Also $R^2 = T^*T$, as is immediate from the formulas above. Since a positive operator has only one positive square root (7.39), we conclude

$$\sqrt{T^*T}\,(z_1, z_2, z_3) = (2z_1, 3z_2, z_3).$$

Next define $S \in \mathcal{L}(\mathbf{F}^3)$ by

$$S(z_1, z_2, z_3) = (z_3, z_1, z_2),$$

so that $Se_1 = e_2$, $Se_2 = e_3$, $Se_3 = e_1$. Then Se_1, Se_2, Se_3 is the list e_2, e_3, e_1, which is an orthonormal basis of $\mathbf{F}^3$. Hence S is a unitary operator, by the equivalence of (a) and (d) in 7.53.
Finally we verify the polar decomposition. For every $(z_1, z_2, z_3) \in \mathbf{F}^3$,

$$S\sqrt{T^*T}\,(z_1, z_2, z_3) = S(2z_1, 3z_2, z_3) = (z_3, 2z_1, 3z_2) = T(z_1, z_2, z_3).$$

Thus $T = S\sqrt{T^*T}$ with S the unitary operator $S(z_1, z_2, z_3) = (z_3, z_1, z_2)$, as required.
Remark: the construction above is exactly the one in the proof of the polar decomposition 7.93. Indeed, for every $v \in \mathbf{F}^3$ we have

$$Tv = 3\langle v, e_2\rangle e_3 + 2\langle v, e_1\rangle e_2 + 1\langle v, e_3\rangle e_1,$$

which is a singular value decomposition of T in the sense of 7.70, with singular values $3, 2, 1$ (in decreasing order, matching 7.65: the eigenvalues of T^*T are $9, 4, 1$), with e_2, e_1, e_3 playing the role of the orthonormal list in the domain and e_3, e_2, e_1 playing the role of the orthonormal list in the target. The unitary operator S produced by 7.93 is the one carrying each vector of the first list to the corresponding vector of the second list, which is exactly the S found above.

Problem 7F.16. Suppose $S \in \mathcal{L}(V)$ is a positive invertible operator. Prove that there exists $\delta > 0$ such that T is a positive operator for every self-adjoint operator $T \in \mathcal{L}(V)$ with $\|S - T\| < \delta$.

Solution. Because S is a positive operator, S is self-adjoint and, by the implication (a) $\Rightarrow$ (c) in 7.38, there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of S, say $Se_k = \lambda_k e_k$ with $\lambda_k \geq 0$ for each k. Since S is invertible, 0 is not an eigenvalue of S, so in fact $\lambda_k > 0$ for each k. Set

$$\delta = \min\{\lambda_1, \dots, \lambda_n\} > 0.$$

We first show that

$$\langle Sv, v\rangle \geq \delta\|v\|^2 \quad \text{for all } v \in V.$$

Indeed, let $v \in V$ and write $v = a_1e_1 + \cdots + a_ne_n$ with $a_k = \langle v, e_k\rangle$ (see 6.30(a)). Then $Sv = \lambda_1 a_1 e_1 + \cdots + \lambda_n a_n e_n$, and because $e_1, \dots, e_n$ is orthonormal,

$$\langle Sv, v\rangle = \sum_{k=1}^{n} \lambda_k |a_k|^2 \geq \delta \sum_{k=1}^{n} |a_k|^2 = \delta\|v\|^2,$$

the last equality being Parseval's identity 6.30(b).
Now suppose $T \in \mathcal{L}(V)$ is self-adjoint and $\|S - T\| < \delta$. Let $v \in V$. The operator $S - T$ is self-adjoint (a difference of self-adjoint operators is self-adjoint), so $\langle (S-T)v, v\rangle$ is a real number by 7.14 when $\mathbf{F} = \mathbf{C}$, and it is automatically real when $\mathbf{F} = \mathbf{R}$. Hence

$$\begin{aligned}
\langle Tv, v\rangle &= \langle Sv, v\rangle - \langle (S-T)v, v\rangle \\
&\geq \delta\|v\|^2 - \big|\langle (S-T)v, v\rangle\big| \\
&\geq \delta\|v\|^2 - \|(S-T)v\|\,\|v\| \\
&\geq \delta\|v\|^2 - \|S-T\|\,\|v\|^2 \\
&= \big(\delta - \|S-T\|\big)\|v\|^2 \\
&\geq 0,
\end{aligned}$$

where the second inequality is the Cauchy–Schwarz inequality 6.14, the third inequality uses 7.89, and the last inequality holds because $\|S - T\| < \delta$.
Thus T is self-adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$, which by definition (7.34) means that T is a positive operator. Hence $\delta = \min\{\lambda_1, \dots, \lambda_n\}$ has the required property.
Remark: the hypothesis that T be self-adjoint cannot be dropped when $\mathbf{F} = \mathbf{R}$. For example, take $V = \mathbf{R}^2$, let $S = I$, and for $\epsilon > 0$ let $T \in \mathcal{L}(\mathbf{R}^2)$ be defined by $T(x, y) = (x - \epsilon y, \epsilon x + y)$. Then $(S - T)(x, y) = (\epsilon y, -\epsilon x)$, so $\|S - T\| = \epsilon$, which is as small as we wish; but T is not self-adjoint, hence not positive under Axler's definition 7.34.

Problem 7F.17. Prove that if $u \in V$ and φ_u is the linear functional on V defined by the equation $\varphi_u(v) = \langle v, u\rangle$, then $\|\varphi_u\| = \|u\|$.
Here we are thinking of the scalar field $\mathbf{F}$ as an inner product space with $\langle \alpha, \beta\rangle = \alpha\overline{\beta}$ for all $\alpha, \beta \in \mathbf{F}$. Thus $\|\varphi_u\|$ means the norm of φ_u as a linear map from V to $\mathbf{F}$.

Solution. First note that φ_u really is a linear map from V to $\mathbf{F}$, because an inner product is linear in its first slot. Note also that the norm on $\mathbf{F}$ arising from the inner product $\langle \alpha, \beta\rangle = \alpha\overline{\beta}$ is the absolute

value: $\|\alpha\| = \sqrt{\alpha\overline{\alpha}} = |\alpha|$ for $\alpha \in \mathbf{F}$. Hence, by the definition 7.86 of the norm of a linear map,

$$\|\varphi_u\| = \max\big\{|\langle v, u\rangle| : v \in V \text{ and } \|v\| \le 1\big\}.$$

Suppose first that $u = 0$. Then $\varphi_u(v) = \langle v, 0\rangle = 0$ for every $v \in V$, so $\varphi_u = 0$ and hence $\|\varphi_u\| = 0 = \|u\|$.
Now suppose $u \neq 0$. If $v \in V$ with $\|v\| \le 1$, then the Cauchy–Schwarz inequality 6.14 gives

$$|\varphi_u(v)| = |\langle v, u\rangle| \le \|v\|\,\|u\| \le \|u\|.$$

Taking the maximum over all such v shows that $\|\varphi_u\| \le \|u\|$.
For the reverse inequality, let $v = u/\|u\|$, which makes sense because $u \neq 0$. Then $\|v\| = 1$, and

$$|\varphi_u(v)| = \Big|\Big\langle \tfrac{u}{\|u\|}, u\Big\rangle\Big| = \frac{|\langle u, u\rangle|}{\|u\|} = \frac{\|u\|^2}{\|u\|} = \|u\|.$$

Since this particular v satisfies $\|v\| \le 1$, the definition of $\|\varphi_u\|$ as a maximum gives $\|\varphi_u\| \ge \|u\|$.
Combining the two inequalities yields $\|\varphi_u\| = \|u\|$, as desired.
Remark: by the Riesz representation theorem 6.42, every linear functional on V has the form φ_u for a unique $u \in V$. The result just proved says that the map $u \mapsto \varphi_u$ from V to V' preserves norms; it is conjugate linear when $\mathbf{F} = \mathbf{C}$.

Problem 7F.18. Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $T \in \mathcal{L}(V, W)$.
(a) Prove that $\max\{\|Te_1\|, \dots, \|Te_n\|\} \le \|T\| \le (\|Te_1\|^2 + \cdots + \|Te_n\|^2)^{1/2}$.
(b) Prove that $\|T\| = (\|Te_1\|^2 + \cdots + \|Te_n\|^2)^{1/2}$ if and only if $\dim \operatorname{range} T \le 1$.
Here $e_1, \dots, e_n$ is an arbitrary orthonormal basis of V, not necessarily connected with a singular value decomposition of T. If $s_1, \dots, s_n$ is the list of singular values of T, then the right side of the inequality above equals $(s_1^2 + \cdots + s_n^2)^{1/2}$, as was shown in Exercise 11(a) in Section 7E.

Solution. Throughout, write

$$c = \big(\|Te_1\|^2 + \cdots + \|Te_n\|^2\big)^{1/2}.$$

(a) Each e_k satisfies $\|e_k\| = 1$, so 7.88(b) gives $\|Te_k\| \le \|T\|$ for each k. Taking the maximum over k gives the left inequality.
For the right inequality, let $v \in V$ with $\|v\| \le 1$. By 6.30(a) we have $v = \sum_{k=1}^n \langle v, e_k\rangle e_k$, so by linearity $Tv = \sum_{k=1}^n \langle v, e_k\rangle\, Te_k$. Hence, using the triangle inequality 6.17 and then the Cauchy–Schwarz inequality in $\mathbf{R}^n$ applied to the vectors $\big(|\langle v, e_1\rangle|, \dots, |\langle v, e_n\rangle|\big)$ and $\big(\|Te_1\|, \dots, \|Te_n\|\big)$,

$$\begin{aligned}
\|Tv\| &\le \sum_{k=1}^n |\langle v, e_k\rangle|\,\|Te_k\| \\
&\le \Big(\sum_{k=1}^n |\langle v, e_k\rangle|^2\Big)^{1/2}\Big(\sum_{k=1}^n \|Te_k\|^2\Big)^{1/2} \\
&= \|v\|\,c \\
&\le c,
\end{aligned}$$

where the equality uses Parseval's identity 6.30(b). Taking the maximum over all $v \in V$ with $\|v\| \le 1$ gives $\|T\| \le c$, completing (a).
(b) We first record a fact about the equality case of the triangle inequality for several vectors.
Claim. Suppose $x_1, \dots, x_n \in W$ and $\|x_1 + \cdots + x_n\| = \|x_1\| + \cdots + \|x_n\|$. Then there exists $w \in W$ such that every x_k is a scalar multiple of w.
Proof of claim. Let $x = x_1 + \cdots + x_n$. If $x = 0$, then $\sum_k \|x_k\| = 0$, so every $x_k = 0$ and any w works.

So assume $x \neq 0$ and put $w = x/\|x\|$, so $\|w\| = 1$. Then

$$\sum_{k=1}^{n} \|x_k\| = \|x\| = \langle x, w\rangle = \sum_{k=1}^{n} \langle x_k, w\rangle.$$

The left side is real, so taking real parts gives $\sum_k \|x_k\| = \sum_k \operatorname{Re}\langle x_k, w\rangle$. For each k we have $\operatorname{Re}\langle x_k, w\rangle \leq |\langle x_k, w\rangle| \leq \|x_k\|\|w\| = \|x_k\|$ by 6.14. Since the sums agree and no term on the right can exceed the corresponding term on the left, each of these inequalities must be an equality; in particular $|\langle x_k, w\rangle| = \|x_k\|\,\|w\|$, which is equality in the Cauchy–Schwarz inequality. By the equality condition in 6.14, one of x_k, w is a scalar multiple of the other. If x_k is a scalar multiple of w, we are done; if instead $w = \mu x_k$ for some $\mu \in \mathbf{F}$, then $\mu \neq 0$ because $w \neq 0$, so $x_k = \mu^{-1} w$ is again a scalar multiple of w. This proves the claim.

Now we prove (b).

Suppose first that $\dim \operatorname{range} T \leq 1$. If $T = 0$, then $\|T\| = 0 = c$ and we are done. Otherwise $\operatorname{range} T$ is one-dimensional; choose $w \in W$ with $\|w\| = 1$ and $\operatorname{range} T = \operatorname{span}(w)$. Then for each k there is $\alpha_k \in \mathbf{F}$ with $Te_k = \alpha_k w$, and $\|Te_k\| = |\alpha_k|$, so $c = (\sum_k |\alpha_k|^2)^{1/2} > 0$. Define

$$v = \frac{1}{c} \sum_{k=1}^{n} \overline{\alpha_k}\, e_k.$$

By Parseval's identity, $\|v\| = \frac{1}{c}(\sum_k |\alpha_k|^2)^{1/2} = 1$. Moreover

$$Tv = \frac{1}{c} \sum_{k=1}^{n} \overline{\alpha_k}\, Te_k = \frac{1}{c}\Big(\sum_{k=1}^{n} \overline{\alpha_k}\, \alpha_k\Big) w = \frac{1}{c}\Big(\sum_{k=1}^{n} |\alpha_k|^2\Big) w = c\, w,$$

so $\|Tv\| = c$. Hence $\|T\| \geq c$, which together with (a) gives $\|T\| = c$.

Conversely, suppose $\|T\| = c$. If $c = 0$, then $T = 0$ by 7.87(b) and $\dim \operatorname{range} T = 0 \leq 1$. So assume $c > 0$. By 7.88(b) the maximum defining $\|T\|$ is attained: there exists $v \in V$ with $\|v\| = 1$ and $\|Tv\| = \|T\| = c$. Set $a_k = |\langle v, e_k\rangle|$ and $b_k = \|Te_k\|$ for $k = 1, \ldots, n$; then $\sum_k a_k^2 = \|v\|^2 = 1$ by 6.30(b), and $\sum_k b_k^2 = c^2$. The chain of inequalities in part (a) applied to this v reads

$$c = \|Tv\| \leq \sum_{k=1}^{n} a_k b_k \leq \Big(\sum_{k=1}^{n} a_k^2\Big)^{1/2} \Big(\sum_{k=1}^{n} b_k^2\Big)^{1/2} = 1 \cdot c = c,$$

so both inequalities are equalities.

Consider the vectors $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ in $\mathbf{R}^n$ with the Euclidean inner product; all their entries are nonnegative, and $\|a\| = 1$ and $\|b\| = c > 0$, so both are nonzero. Equality in the second inequality above says $\langle a, b\rangle = \|a\|\,\|b\|$, which is equality in the Cauchy–Schwarz inequality 6.14. Hence one of a, b is a scalar multiple of the other, and since both are nonzero we may write $a = \lambda b$ with $\lambda \in \mathbf{R}$ and $\lambda \neq 0$. Then $1 = \|a\| = |\lambda|\,\|b\| = |\lambda| c$, so $|\lambda| = 1/c$. Choose j with $b_j > 0$ (possible since $b \neq 0$); then $a_j = \lambda b_j$ and $a_j \geq 0$ force $\lambda > 0$. Therefore $\lambda = 1/c$, that is, $a = b/c$, which says

$$|\langle v, e_k\rangle| = \frac{\|Te_k\|}{c} \quad \text{for } k = 1, \ldots, n.$$

In particular $\langle v, e_k\rangle = 0$ if and only if $Te_k = 0$.

Equality in the first inequality says

$$\Big\|\sum_{k=1}^{n} \langle v, e_k\rangle\, Te_k\Big\| = \sum_{k=1}^{n} \big\|\langle v, e_k\rangle\, Te_k\big\|.$$

By the claim, applied to the vectors $x_k = \langle v, e_k\rangle\, Te_k$, there is $w \in W$ such that each $\langle v, e_k\rangle\, Te_k$ is a scalar multiple of w. If $\langle v, e_k\rangle \neq 0$, we may divide by it and conclude that $Te_k \in \operatorname{span}(w)$. If $\langle v, e_k\rangle = 0$, then by the previous paragraph $Te_k = 0 \in \operatorname{span}(w)$. Hence $Te_k \in \operatorname{span}(w)$ for every k.

Since $e_1, \dots, e_n$ spans V, the list $Te_1, \dots, Te_n$ spans range T. Therefore range $T \subseteq \operatorname{span}(w)$, and so $\dim \operatorname{range} T \le 1$, as desired.
Remark: combining (b) with the identity $c^2 = s_1^2 + \cdots + s_n^2$ from Exercise 11(a) in Section 7E and with $\|T\| = s_1$ (7.88(a)), part (b) says that $s_1^2 = s_1^2 + \cdots + s_n^2$ exactly when $s_2 = \cdots = s_n = 0$, i.e. exactly when T has at most one positive singular value, which by the singular value decomposition 7.70 happens exactly when $\dim \operatorname{range} T \le 1$.

Problem 7F.19. Prove that if $T \in \mathcal{L}(V, W)$, then $\|T^*T\| = \|T\|^2$.
This formula for $\|T^*T\|$ leads to the important subject of C^*-algebras.

Solution. We prove the two inequalities separately.
First we show $\|T^*T\| \le \|T\|^2$. Let $v \in V$. Applying 7.89 twice,

$$\|T^*Tv\| \le \|T^*\| \, \|Tv\| \le \|T^*\| \, \|T\| \, \|v\|.$$

By 7.88(c), $\|T^*T\|$ is the smallest number c with $\|T^*Tv\| \le c\|v\|$ for all $v \in V$; hence

$$\|T^*T\| \le \|T^*\| \, \|T\| = \|T\|^2,$$

where the last equality holds because a linear map and its adjoint have the same norm (7.91).
Now we show $\|T\|^2 \le \|T^*T\|$. Suppose $v \in V$ with $\|v\| \le 1$. The defining property of the adjoint gives $\langle Tv, Tv\rangle = \langle v, T^*Tv\rangle$. Hence, using the Cauchy–Schwarz inequality 6.14 and then 7.89,

$$\|Tv\|^2 = \langle v, T^*Tv\rangle \le \|v\| \, \|T^*Tv\| \le \|T^*T\| \, \|v\|^2 \le \|T^*T\|.$$

(The first quantity is a nonnegative real number, so the inequalities above make sense.) Thus $\|Tv\| \le \|T^*T\|^{1/2}$ for every $v \in V$ with $\|v\| \le 1$. Because $\|T\|$ is by definition 7.86 the maximum of $\|Tv\|$ over all such v, we conclude that $\|T\| \le \|T^*T\|^{1/2}$, that is, $\|T\|^2 \le \|T^*T\|$.
Combining the two inequalities gives $\|T^*T\| = \|T\|^2$.
Remark: an alternative proof runs through singular values. The operator T^*T is positive, and if $s_1 \ge \cdots \ge s_n$ are the singular values of T, then by definition (7.65) the s_k^2 are the eigenvalues of T^*T, listed with multiplicity. Since T^*T is positive, its singular values are its eigenvalues, so the largest singular value of T^*T is s_1^2. Now 7.88(a) gives $\|T^*T\| = s_1^2 = \|T\|^2$.

Problem 7F.20. Suppose $T \in \mathcal{L}(V)$ is normal. Prove that $\|T^k\| = \|T\|^k$ for every positive integer k.

Solution. We give an argument valid for both $\mathbf{F} = \mathbf{R}$ and $\mathbf{F} = \mathbf{C}$ (note that when $\mathbf{F} = \mathbf{R}$ a normal operator need not be diagonalizable, so we cannot simply invoke the spectral theorem).
Step 1: submultiplicativity. If $A, B \in \mathcal{L}(V)$ and $v \in V$, then two applications of 7.89 give $\|ABv\| \le \|A\| \, \|Bv\| \le \|A\| \, \|B\| \, \|v\|$. Hence by 7.88(c),

$$\|AB\| \le \|A\| \, \|B\|.$$

Iterating, $\|A^k\| \le \|A\|^k$ for every positive integer k and every $A \in \mathcal{L}(V)$.
Step 2: powers of a normal operator are normal. Because T is normal, T commutes with T^*. Hence for every positive integer j, the operator T^j commutes with $(T^*)^j$. Since $(T^j)^* = (T^*)^j$ by repeated use of 7.5(d), the operator T^j is normal.
Step 3: $\|T^2\| = \|T\|^2$ for normal T. Using $(T^2)^* = (T^*)^2$ and then the commutativity $TT^* = T^*T$,

$$(T^2)^*T^2 = T^*T^*TT = T^*(T^*T)T = T^*(TT^*)T = (T^*T)(T^*T) = (T^*T)^2.$$

Applying Exercise 19 in this section to T^2 and then to T^*T gives

$$\|T^2\|^2 = \big\|(T^2)^*T^2\big\| = \big\|(T^*T)^2\big\|.$$

Now $A := T^*T$ is a positive operator, by 7.64(a); in particular A is self-adjoint, so $A^*A = A^2$, and Exercise 19 applied to A gives $\|A^2\| = \|A^*A\| = \|A\|^2$. Also $\|A\| = \|T^*T\| = \|T\|^2$ by Exercise 19 applied to T. Therefore

$$\|T^2\|^2 = \|A^2\| = \|A\|^2 = \big(\|T\|^2\big)^2 = \|T\|^4,$$

and taking nonnegative square roots gives $\|T^2\| = \|T\|^2$.
Step 4: $\|T^{2^m}\| = \|T\|^{2^m}$ for every nonnegative integer m. We induct on m. The case $m = 0$ is trivial. Suppose the statement holds for m. By Step 2, the operator $R = T^{2^m}$ is normal, so Step 3 applied to R gives

$$\big\|T^{2^{m+1}}\big\| = \|R^2\| = \|R\|^2 = \big(\|T\|^{2^m}\big)^2 = \|T\|^{2^{m+1}},$$

completing the induction.
Step 5: the general case. Let k be a positive integer. If $T = 0$, then both $\|T^k\|$ and $\|T\|^k$ equal 0 and there is nothing to prove; so assume $T \neq 0$, and hence $\|T\| > 0$ by 7.87(b). Choose a nonnegative integer m with $2^m \geq k$, and write $N = 2^m$. By Step 1 and Step 4,

$$\|T\|^N = \|T^N\| = \big\|T^k\, T^{N-k}\big\| \leq \|T^k\|\, \|T^{N-k}\| \leq \|T^k\|\, \|T\|^{N-k},$$

where in the last step we used $\|T^{N-k}\| \leq \|T\|^{N-k}$ from Step 1 (this also holds when $N = k$, since $T^0 = I$ has norm 1). Dividing by $\|T\|^{N-k} > 0$ gives

$$\|T\|^k \leq \|T^k\|.$$

The reverse inequality $\|T^k\| \leq \|T\|^k$ is Step 1. Hence $\|T^k\| = \|T\|^k$, as desired.

Problem 7F.21. Suppose $\dim V > 1$ and $\dim W > 1$. Prove that the norm on $\mathcal{L}(V, W)$ does not come from an inner product. In other words, prove that there does not exist an inner product on $\mathcal{L}(V, W)$ such that

$$\max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\} = \sqrt{\langle T, T\rangle}$$

for all $T \in \mathcal{L}(V, W)$.

Solution. Every norm arising from an inner product satisfies the parallelogram equality 6.21: if $\langle \cdot, \cdot \rangle$ is an inner product on a vector space U and $\|u\| = \sqrt{\langle u, u\rangle}$, then

$$\|A + B\|^2 + \|A - B\|^2 = 2\big(\|A\|^2 + \|B\|^2\big) \quad \text{for all } A, B \in U.$$

So it suffices to exhibit $A, B \in \mathcal{L}(V, W)$ for which the operator norm violates this identity.
Because $\dim V \geq 2$, the Gram–Schmidt procedure 6.32 produces an orthonormal basis of V; let e_1, e_2 be its first two vectors, so e_1, e_2 is an orthonormal list in V. Similarly, let f_1, f_2 be an orthonormal list in W, which exists because $\dim W \geq 2$. Define $A, B \in \mathcal{L}(V, W)$ by

$$Av = \langle v, e_1\rangle f_1 \quad \text{and} \quad Bv = \langle v, e_2\rangle f_2$$

for $v \in V$. Both are linear because an inner product is linear in its first slot.
We compute the four norms.
Norm of A: for $v \in V$, we have $\|Av\| = |\langle v, e_1\rangle|\, \|f_1\| = |\langle v, e_1\rangle| \leq \|v\|\, \|e_1\| = \|v\|$ by the Cauchy–Schwarz inequality 6.14. Hence $\|A\| \leq 1$ by 7.88(c). Also $\|e_1\| = 1$ and $\|Ae_1\| = |\langle e_1, e_1\rangle| = 1$, so $\|A\| \geq 1$. Thus $\|A\| = 1$. The same computation with e_2, f_2 gives $\|B\| = 1$.
Norm of $A + B$: for $v \in V$,

$$(A + B)v = \langle v, e_1\rangle f_1 + \langle v, e_2\rangle f_2,$$

and since f_1, f_2 is orthonormal, the Pythagorean theorem gives

$$\|(A + B)v\|^2 = |\langle v, e_1\rangle|^2 + |\langle v, e_2\rangle|^2 \leq \|v\|^2,$$

where the inequality is Bessel's inequality (extend e_1, e_2 to an orthonormal basis of V and apply Parseval's identity 6.30(b)). Hence $\|A + B\| \leq 1$ by 7.88(c). Taking $v = e_1$ gives $\|(A + B)e_1\| = 1$ with $\|e_1\| = 1$, so $\|A + B\| = 1$.

Norm of $A-B$: for $v \in V$ we have $(A-B)v = \langle v, e_1\rangle f_1 - \langle v, e_2\rangle f_2$, and exactly the same computation gives

$$\|(A-B)v\|^2 = |\langle v, e_1\rangle|^2 + |\langle v, e_2\rangle|^2 \le \|v\|^2,$$

with equality attained at $v = e_1$. Hence $\|A-B\| = 1$.
Now test the parallelogram equality:

$$\|A+B\|^2 + \|A-B\|^2 = 1 + 1 = 2,$$

whereas

$$2\big(\|A\|^2 + \|B\|^2\big) = 2(1+1) = 4.$$

Since $2 \neq 4$, the operator norm on $\mathcal{L}(V,W)$ does not satisfy the parallelogram equality. Therefore there is no inner product on $\mathcal{L}(V,W)$ whose associated norm is the operator norm; that is, the norm on $\mathcal{L}(V,W)$ does not come from an inner product.
Remark: the hypotheses $\dim V > 1$ and $\dim W > 1$ are genuinely needed. Suppose, for instance, that $\dim V = 1$, say $V = \operatorname{span}(e)$ with $\|e\| = 1$. Every $v \in V$ with $\|v\| \le 1$ has the form $v = \alpha e$ with $|\alpha| \le 1$, so $\|T\| = \max\{|\alpha|\,\|Te\| : |\alpha| \le 1\} = \|Te\|$ for every $T \in \mathcal{L}(V,W)$. The formula $\langle S, T\rangle = \langle Se, Te\rangle$ then defines an inner product on $\mathcal{L}(V,W)$ (it inherits all the required properties from the inner product on W, and $\langle T, T\rangle = 0$ forces $Te = 0$ and hence $T = 0$), and its associated norm is exactly the operator norm.
The case $\dim W = 1$ also admits such an inner product, though the construction needs one twist. Say $W = \operatorname{span}(f)$ with $\|f\| = 1$. For $T \in \mathcal{L}(V,W)$ we have $Tv = \langle Tv, f\rangle f$, so $\|Tv\| = |\langle Tv, f\rangle|$; thus $v \mapsto \langle Tv, f\rangle$ is a linear functional on V whose norm equals $\|T\|$. By the Riesz representation theorem 6.42 there is a unique $u_T \in V$ with $\langle Tv, f\rangle = \langle v, u_T\rangle$ for all $v \in V$, and Exercise 17 above gives $\|u_T\| = \|T\|$. The map $T \mapsto u_T$ satisfies $u_{S+T} = u_S + u_T$ and $u_{\lambda T} = \overline{\lambda}\, u_T$, so it is conjugate linear rather than linear. Hence we must reverse the slots and set

$$\langle S, T\rangle = \langle u_T, u_S\rangle,$$

which is then linear in S, satisfies $\langle T, S\rangle = \overline{\langle S, T\rangle}$, and has $\langle T, T\rangle = \|u_T\|^2 = \|T\|^2$, with $\langle T, T\rangle = 0$ forcing $u_T = 0$ and hence $T = 0$. So this is an inner product on $\mathcal{L}(V,W)$ whose associated norm is the operator norm.

Problem 7F.22. Suppose $T \in \mathcal{L}(V,W)$. Let $n = \dim V$ and let $s_1 \ge \cdots \ge s_n$ denote the singular values of T. Prove that if $1 \le k \le n$, then

$$\min\{\|T|_U\| : U \text{ is a subspace of } V \text{ with } \dim U = k\} = s_{n-k+1}.$$

Solution. Throughout, $\|T|_U\|$ denotes the norm of the linear map $T|_U \in \mathcal{L}(U,W)$, that is,

$$\|T|_U\| = \max\{\|Tv\| : v \in U \text{ and } \|v\| \le 1\},$$

by 7.86 applied to $T|_U$ (here U carries the inner product it inherits from V).
Because T^*T is a positive operator on V [by 7.64(a)], the spectral theorem gives an orthonormal basis $e_1, \ldots, e_n$ of V with

$$T^*Te_j = s_j^2 e_j \qquad \text{for } j = 1, \ldots, n,$$

where $s_1 \ge \cdots \ge s_n$ are the singular values of T. This is exactly the step 7.72 in the proof of the singular value decomposition 7.70, and it is immediate from the definition 7.65 of the singular values: they are the nonnegative square roots of the eigenvalues of T^*T, listed in decreasing order, each repeated according to the dimension of the corresponding eigenspace of T^*T. Note that $T^*T \in \mathcal{L}(V)$ even though T maps V into W, so there are exactly $n = \dim V$ singular values.
The key computation is the following. If $v \in V$, then $v = \sum_{j=1}^{n} \langle v, e_j\rangle e_j$ by 6.30(a), so $T^*Tv =$

$\sum_{j=1}^{n} s_j^2 \langle v, e_j\rangle e_j$, and therefore

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle = \sum_{j=1}^{n} s_j^2 \big|\langle v, e_j\rangle\big|^2.$$

Also $\|v\|^2 = \sum_{j=1}^{n} |\langle v, e_j\rangle|^2$ by Parseval's identity 6.30(b). Taking $v = e_j$ in the displayed formula gives $\|Te_j\| = s_j$.
Step 1: some k-dimensional subspace achieves the value s_{n-k+1}.
Let $U = \operatorname{span}(e_{n-k+1}, \dots, e_n)$. Since $e_{n-k+1}, \dots, e_n$ is an orthonormal (hence linearly independent) list of length k, we have $\dim U = k$. If $v \in U$ with $\|v\| \le 1$, then $\langle v, e_j\rangle = 0$ for $j < n-k+1$, so

$$\|Tv\|^2 = \sum_{j=n-k+1}^{n} s_j^2 \big|\langle v, e_j\rangle\big|^2 \le s_{n-k+1}^2 \sum_{j=n-k+1}^{n} \big|\langle v, e_j\rangle\big|^2 = s_{n-k+1}^2 \|v\|^2 \le s_{n-k+1}^2,$$

where the first inequality holds because $s_j \le s_{n-k+1}$ for every $j \ge n-k+1$. Thus $\|T|_U\| \le s_{n-k+1}$. On the other hand $e_{n-k+1} \in U$ and $\|e_{n-k+1}\| = 1$ and $\|Te_{n-k+1}\| = s_{n-k+1}$, so $\|T|_U\| \ge s_{n-k+1}$. Hence

$$\|T|_U\| = s_{n-k+1}.$$

Step 2: every k-dimensional subspace gives at least s_{n-k+1}.
Suppose U is a subspace of V with $\dim U = k$. Let

$$U' = \operatorname{span}(e_1, \dots, e_{n-k+1}),$$

so $\dim U' = n-k+1$. Then

$$\dim(U \cap U') = \dim U + \dim U' - \dim(U + U') \ge k + (n-k+1) - n = 1,$$

using 2.43 and the fact that $U + U'$ is a subspace of V, so $\dim(U + U') \le n$. Hence there exists $v \in U \cap U'$ with $\|v\| = 1$.
Because $v \in U'$, we have $\langle v, e_j\rangle = 0$ for $j > n-k+1$, and therefore

$$\|Tv\|^2 = \sum_{j=1}^{n-k+1} s_j^2 \big|\langle v, e_j\rangle\big|^2 \ge s_{n-k+1}^2 \sum_{j=1}^{n-k+1} \big|\langle v, e_j\rangle\big|^2 = s_{n-k+1}^2 \|v\|^2 = s_{n-k+1}^2,$$

where the inequality holds because $s_j \ge s_{n-k+1}$ for every $j \le n-k+1$. Because $v \in U$ and $\|v\| = 1$, this gives

$$\|T|_U\| \ge \|Tv\| \ge s_{n-k+1}.$$

Combining the two steps: the set $\{\|T|_U\| : U \text{ is a subspace of } V \text{ with } \dim U = k\}$ is bounded below by s_{n-k+1} by Step 2, and contains s_{n-k+1} by Step 1. Hence the minimum exists and

$$\min\{\|T|_U\| : U \text{ is a subspace of } V \text{ with } \dim U = k\} = s_{n-k+1},$$

as desired.
Note the two extreme cases as a sanity check: $k = n$ forces $U = V$ and gives $\|T\| = s_1$, agreeing with 7.88(a); $k = 1$ gives s_n, the smallest singular value, which is the smallest value of $\|Tv\|$ over unit vectors v.

Problem 7F.23. Suppose $T \in \mathcal{L}(V, W)$. Show that T is uniformly continuous with respect to the metrics on V and W that arise from the norms on those spaces (see Exercise 23 in Section 6B).

Solution. Exercise 23 in Section 6B shows that

$$d_V(u, v) = \|u - v\| \quad (u, v \in V), \qquad d_W(x, y) = \|x - y\| \quad (x, y \in W)$$

define metrics on V and on W. Recall that T is uniformly continuous with respect to these metrics means: for every $\varepsilon > 0$ there exists $\delta > 0$, depending on ε but not on the points involved, such that

$$d_W(Tu, Tv) < \varepsilon \quad \text{whenever } u, v \in V \text{ and } d_V(u, v) < \delta.$$

Because V is finite-dimensional, the number $\|T\|$ is defined (7.86), and 7.89 gives

$$\|Tv\| \le \|T\| \, \|v\| \qquad \text{for all } v \in V.$$

Now suppose $\varepsilon > 0$. Set

$$\delta = \frac{\varepsilon}{\|T\| + 1},$$

which is a positive number (the $+1$ in the denominator merely avoids division by 0 in the case $T = 0$). Suppose $u, v \in V$ satisfy $d_V(u, v) = \|u - v\| < \delta$. Then, using the linearity of T and then 7.89,

$$d_W(Tu, Tv) = \|Tu - Tv\| = \|T(u - v)\| \le \|T\| \, \|u - v\| < \|T\| \, \delta = \frac{\|T\|}{\|T\| + 1} \, \varepsilon \le \varepsilon.$$

Because δ depends only on ε and on T (and not on u or v), this shows that T is uniformly continuous with respect to d_V and d_W.
In other words, every linear map between finite-dimensional inner product spaces is Lipschitz continuous with Lipschitz constant $\|T\|$, and Lipschitz continuity implies uniform continuity. Finite-dimensionality is essential here: 7.86 requires the maximum defining $\|T\|$ to exist, and on infinite-dimensional inner product spaces there are linear maps that are not continuous at any point.

Problem 7F.24. Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that

$$\|T^{-1}\| = \|T\|^{-1} \iff \frac{T}{\|T\|} \text{ is a unitary operator.}$$

Solution. We may assume $V \neq \{0\}$: if $V = \{0\}$ then $T = 0$ and $\|T\| = 0$, so neither $\|T\|^{-1}$ nor $T/\|T\|$ is defined and the assertion has no content. So assume $\dim V \ge 1$. Because T is invertible we then have $T \neq 0$, and hence $\|T\| > 0$ by 7.87(b). Thus the operator

$$S = \frac{T}{\|T\|}$$

makes sense. By 7.87(c),

$$\|S\| = \frac{1}{\|T\|} \|T\| = 1.$$

Also S is invertible, with $S^{-1} = \|T\| \, T^{-1}$, so again by 7.87(c),

$$\|S^{-1}\| = \|T\| \, \|T^{-1}\|.$$

Consequently

$$\|T^{-1}\| = \|T\|^{-1} \iff \|S^{-1}\| = 1.$$

So the exercise reduces to proving, for an invertible $S \in \mathcal{L}(V)$ with $\|S\| = 1$:

$$\|S^{-1}\| = 1 \iff S \text{ is unitary.}$$

First suppose $\|S^{-1}\| = 1$. Let $v \in V$. Applying 7.89 to S gives

$$\|Sv\| \le \|S\| \, \|v\| = \|v\|.$$

Applying 7.89 to S^{-1} with the vector Sv gives

$$\|v\| = \|S^{-1}(Sv)\| \le \|S^{-1}\| \, \|Sv\| = \|Sv\|.$$

Combining the two inequalities yields $\|Sv\| = \|v\|$ for every $v \in V$. Thus S is an isometry (7.44), and being an invertible isometry it is a unitary operator (7.51). Hence $T/\|T\|$ is unitary.
Conversely, suppose $S = T/\|T\|$ is unitary. Then $\|Sv\| = \|v\|$ for every $v \in V$. Moreover $S^{-1} = S^*$ by 7.53(c), and S^* is unitary by 7.53(f); hence S^{-1} is also an isometry, so

$$\|S^{-1}w\| = \|w\| \qquad \text{for every } w \in V.$$

Taking w with $\|w\| = 1$ shows $\|S^{-1}w\| = 1$, and by 7.88(b) this gives $\|S^{-1}\| = 1$. Therefore

$$1 = \|S^{-1}\| = \|T\|\,\|T^{-1}\|,$$

which is exactly the statement $\|T^{-1}\| = \|T\|^{-1}$.
This completes the proof of both implications.
Remark. In terms of singular values the equivalence reads as follows. If $s_1 \geq \cdots \geq s_n > 0$ are the singular values of T, then $\|T\| = s_1$ and $\|T^{-1}\| = 1/s_n$ (the singular values of T^{-1} are the reciprocals of those of T; see Exercise 10 in Section 7E). Thus $\|T^{-1}\| = \|T\|^{-1}$ says $1/s_n = 1/s_1$, that is, all singular values of T are equal; dividing by $\|T\| = s_1$ then makes all singular values equal to 1, which characterizes unitary operators.

Problem 7F.25. Fix $u, x \in V$ with $u \neq 0$. Define $T \in \mathcal{L}(V)$ by $Tv = \langle v, u\rangle x$ for every $v \in V$. Prove that

$$\sqrt{T^*T}\, v = \frac{\|x\|}{\|u\|}\langle v, u\rangle u$$

for every $v \in V$.

Solution. Define $R \in \mathcal{L}(V)$ by

$$Rv = \frac{\|x\|}{\|u\|}\langle v, u\rangle u \qquad \text{for } v \in V.$$

(That R is linear is clear, since $v \mapsto \langle v, u\rangle$ is linear.) By 7.39 and 7.40, the positive operator T^*T has exactly one positive square root, denoted $\sqrt{T^*T}$. So it suffices to prove two things: that R is a positive operator, and that $R^2 = T^*T$.
Step 1: a formula for T^* and for T^*T.
If $v, w \in V$, then

$$\langle Tv, w\rangle = \langle\langle v, u\rangle x, w\rangle = \langle v, u\rangle\langle x, w\rangle = \big\langle v, \overline{\langle x, w\rangle}\, u\big\rangle = \big\langle v, \langle w, x\rangle u\big\rangle.$$

Because the adjoint is the unique linear map satisfying $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for all v, w, we conclude that

$$T^*w = \langle w, x\rangle u \qquad \text{for every } w \in V.$$

Hence for every $v \in V$,

$$T^*Tv = T^*\big(\langle v, u\rangle x\big) = \langle v, u\rangle\, T^*x = \langle v, u\rangle\langle x, x\rangle u = \|x\|^2\langle v, u\rangle u.$$

In the second equality we pulled the scalar $\langle v, u\rangle$ out through the linear map T^*.
Step 2: R is a positive operator.
R is self-adjoint: for $v, w \in V$, since $\|x\|/\|u\|$ is a real number,

$$\langle Rv, w\rangle = \frac{\|x\|}{\|u\|}\langle v, u\rangle\langle u, w\rangle$$

and

$$\langle v, Rw\rangle = \Big\langle v, \frac{\|x\|}{\|u\|}\langle w, u\rangle u\Big\rangle = \frac{\|x\|}{\|u\|}\overline{\langle w, u\rangle}\,\langle v, u\rangle = \frac{\|x\|}{\|u\|}\langle u, w\rangle\langle v, u\rangle.$$

These two expressions are equal, so $\langle Rv, w\rangle = \langle v, Rw\rangle$ for all $v, w \in V$, which means $R^* = R$.

R satisfies $\langle Rv, v\rangle \geq 0$: for $v \in V$,

$$\langle Rv, v\rangle = \frac{\|x\|}{\|u\|}\langle v, u\rangle\langle u, v\rangle = \frac{\|x\|}{\|u\|}\langle v, u\rangle\overline{\langle v, u\rangle} = \frac{\|x\|}{\|u\|}\big|\langle v, u\rangle\big|^2 \geq 0,$$

because $\|x\| \geq 0$ and $\|u\| > 0$. Thus R is a positive operator.
Step 3: $R^2 = T^*T$.
For $v \in V$,

$$\begin{aligned} R(Rv) &= \frac{\|x\|}{\|u\|}\big\langle Rv, u\big\rangle u \\ &= \frac{\|x\|}{\|u\|} \cdot \frac{\|x\|}{\|u\|}\langle v, u\rangle\langle u, u\rangle\, u \\ &= \frac{\|x\|^2}{\|u\|^2}\,\|u\|^2\,\langle v, u\rangle u \\ &= \|x\|^2\langle v, u\rangle u \\ &= T^*Tv, \end{aligned}$$

where the last equality is Step 1. Hence $R^2 = T^*T$.
Conclusion. By Steps 2 and 3, R is a positive square root of T^*T. By the uniqueness of the positive square root (7.39), $R = \sqrt{T^*T}$, that is,

$$\sqrt{T^*T}\,v = \frac{\|x\|}{\|u\|}\langle v, u\rangle u$$

for every $v \in V$, as desired.
Remark. Since $P_{\operatorname{span}(u)}v = \langle v, u\rangle u/\|u\|^2$, the formula just proved says

$$\sqrt{T^*T} = \|x\|\,\|u\|\,P_{\operatorname{span}(u)}.$$

Thus if $x \neq 0$ the only nonzero singular value of T is $\|x\|\,\|u\|$, in agreement with $\|T\| = \|x\|\,\|u\|$, which also follows from the Cauchy–Schwarz inequality 6.14. If $x = 0$ then $T = 0$ and both sides of the desired formula equal 0.

Problem 7F.26. Suppose $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if there exists a unique unitary operator $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$.

Solution. Write $R = \sqrt{T^*T}$, which is a well-defined positive operator on V because T^*T is positive [7.64(a)] and positive operators have a unique positive square root (7.39, 7.40). The polar decomposition 7.93 says that there exists at least one unitary $S \in \mathcal{L}(V)$ with $T = SR$. So the content of the exercise is: T is invertible if and only if this S is unique.
We first record a fact used in both directions.
Claim: $\|Rv\| = \|Tv\|$ for every $v \in V$; consequently $\operatorname{null} R = \operatorname{null} T$.
Indeed, R is self-adjoint and $R^2 = T^*T$, so for $v \in V$,

$$\|Rv\|^2 = \langle Rv, Rv\rangle = \langle R^*Rv, v\rangle = \langle R^2v, v\rangle = \langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2.$$

Taking square roots gives $\|Rv\| = \|Tv\|$, and hence $Rv = 0 \iff Tv = 0$.
Direction 1: T invertible $\implies$ S is unique.
Suppose T is invertible. Then $\operatorname{null} T = \{0\}$, so by the claim $\operatorname{null} R = \{0\}$; as R is an operator on the finite-dimensional space V, this makes R invertible (3.65). Now suppose S_1 and S_2 are unitary operators with

$$T = S_1R = S_2R.$$

Multiplying on the right by R^{-1} gives $S_1 = TR^{-1} = S_2$. Combined with the existence supplied by 7.93, there is exactly one unitary S with $T = S\sqrt{T^*T}$, namely $S = T(\sqrt{T^*T})^{-1}$.

Direction 2: S unique $\implies$ T invertible.
We prove the contrapositive: if T is not invertible, then there is more than one unitary S with $T = SR$.
So suppose T is not invertible. Then $U := \operatorname{null} T \neq \{0\}$, and by the claim $\operatorname{null} R = U$ as well. By 7.93 there is at least one unitary S with $T = SR$; fix such an S. We produce a second one.
Because R is self-adjoint, 7.6(d) applied to R gives

$$\operatorname{range} R = (\operatorname{null} R^*)^\perp = (\operatorname{null} R)^\perp = U^\perp.$$

Since $V = U \oplus U^\perp$ (by 6.49), we may define $Q \in \mathcal{L}(V)$ by

$$Q(v_1 + v_2) = -v_1 + v_2 \qquad \text{for } v_1 \in U,\ v_2 \in U^\perp.$$

This Q is well defined and linear because the decomposition $v = v_1 + v_2$ with $v_1 \in U$, $v_2 \in U^\perp$ is unique and depends linearly on v. Moreover, v_1 and v_2 are orthogonal, so by the Pythagorean theorem 6.12,

$$\|Q(v_1 + v_2)\|^2 = \|-v_1\|^2 + \|v_2\|^2 = \|v_1\|^2 + \|v_2\|^2 = \|v_1 + v_2\|^2.$$

Thus Q is an isometry (7.44) on the finite-dimensional space V, hence a unitary operator (7.51; an isometry is injective and therefore invertible by 3.65).
Now set $S' = SQ$. Then S' is unitary, being a composition of unitary operators: S' is invertible, and $\|S'v\| = \|S(Qv)\| = \|Qv\| = \|v\|$ for all v.
Next, $S'R = SR$. Indeed, if $v \in V$ then $Rv \in \operatorname{range} R = U^\perp$, so $Q(Rv) = Rv$ by the definition of Q; hence

$$S'Rv = S(Q(Rv)) = S(Rv) = SRv = Tv.$$

Thus $T = S'R$ as well.
Finally, $S' \neq S$: if $SQ = S$, then multiplying on the left by S^{-1} would give $Q = I$, which is false because $U \neq \{0\}$ and $Qv_1 = -v_1 \neq v_1$ for any nonzero $v_1 \in U$ (over both $\mathbf{R}$ and $\mathbf{C}$, $-v_1 = v_1$ forces $v_1 = 0$).
So S and S' are two distinct unitary operators with $T = S\sqrt{T^*T} = S'\sqrt{T^*T}$, showing that the unitary operator in the polar decomposition is not unique. This proves the contrapositive, and hence Direction 2.
Combining the two directions completes the proof.

Problem 7F.27. Suppose $T \in \mathcal{L}(V)$ and $s_1, \dots, s_n$ are the singular values of T. Let $e_1, \dots, e_n$ and $f_1, \dots, f_n$ be orthonormal bases of V such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for all $v \in V$. Define $S \in \mathcal{L}(V)$ by

$$Sv = \langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n.$$

(a) Show that S is unitary and $\|T - S\| = \max\{|s_1 - 1|, \dots, |s_n - 1|\}$.
(b) Show that if $E \in \mathcal{L}(V)$ is unitary, then $\|T - E\| \geq \|T - S\|$.
[This exercise finds a unitary operator S that is as close as possible (among the unitary operators) to a given operator T.]

Solution. Throughout put

$$M = \max\{|s_1 - 1|, \dots, |s_n - 1|\},$$

and fix an index $k \in \{1, \dots, n\}$ with $|s_k - 1| = M$.
(a) First we show S is unitary. Let $v \in V$. Because $f_1, \dots, f_n$ is an orthonormal list, the Pythagorean theorem (applied repeatedly, or 6.30) gives

$$\|Sv\|^2 = \big\|\langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n\big\|^2 = \big|\langle v, e_1\rangle\big|^2 + \cdots + \big|\langle v, e_n\rangle\big|^2 = \|v\|^2,$$

where the last equality is Parseval's identity 6.30(b) for the orthonormal basis $e_1, \dots, e_n$. Thus S is an isometry (7.44); since V is finite-dimensional, S is injective and hence invertible (3.65), so S is a unitary operator (7.51).
Now we compute $\|T - S\|$. Subtracting the two displayed formulas gives, for every $v \in V$,

$$(T - S)v = (s_1 - 1)\langle v, e_1\rangle f_1 + \cdots + (s_n - 1)\langle v, e_n\rangle f_n.$$

Again using orthonormality of $f_1, \dots, f_n$ and then Parseval's identity for $e_1, \dots, e_n$,

$$\|(T - S)v\|^2 = \sum_{j=1}^{n} (s_j - 1)^2 \big|\langle v, e_j\rangle\big|^2 \le M^2 \sum_{j=1}^{n} \big|\langle v, e_j\rangle\big|^2 = M^2\|v\|^2.$$

Hence $\|(T - S)v\| \le M\|v\|$ for all $v \in V$, and so $\|T - S\| \le M$ by 7.88(c).
For the reverse inequality, take $v = e_k$. Since $\langle e_k, e_j\rangle = \delta_{jk}$, the displayed formula for $(T - S)v$ gives $(T - S)e_k = (s_k - 1)f_k$, so

$$\|(T - S)e_k\| = |s_k - 1|\,\|f_k\| = |s_k - 1| = M.$$

Because $\|e_k\| = 1$, the definition 7.86 of the norm gives $\|T - S\| \ge M$. Therefore

$$\|T - S\| = M = \max\{|s_1 - 1|, \dots, |s_n - 1|\}.$$

(b) Suppose $E \in \mathcal{L}(V)$ is unitary. Then E is an isometry, so $\|Ee_k\| = \|e_k\| = 1$. Also $Te_k = s_k f_k$ by the assumed singular value decomposition, so $\|Te_k\| = s_k\|f_k\| = s_k$.
We now use the reverse triangle inequality for the norm on V: for $a, b \in V$, the triangle inequality 6.17 gives $\|a\| = \|(a - b) + b\| \le \|a - b\| + \|b\|$, hence $\|a - b\| \ge \|a\| - \|b\|$, and interchanging a and b gives $\|a - b\| \ge \big|\,\|a\| - \|b\|\,\big|$. Applying this with $a = Te_k$ and $b = Ee_k$,

$$\|(T - E)e_k\| = \|Te_k - Ee_k\| \ge \big|\,\|Te_k\| - \|Ee_k\|\,\big| = |s_k - 1| = M.$$

Since $\|e_k\| = 1$, the definition 7.86 gives

$$\|T - E\| \ge \|(T - E)e_k\| \ge M = \|T - S\|,$$

the last equality by part (a). This is precisely the desired inequality.
Together, (a) and (b) show that S minimizes the distance $\|T - E\|$ over all unitary $E \in \mathcal{L}(V)$, and that the minimum distance equals $\max\{|s_1 - 1|, \dots, |s_n - 1|\}$. Note also that this S is a unitary operator of exactly the form constructed in the proof of the polar decomposition 7.93, and indeed $T = S\sqrt{T^*T}$: applying T^* to the displayed formula for T gives $T^*Tv = \sum_j s_j^2\langle v, e_j\rangle e_j$, hence $\sqrt{T^*T}\,v = \sum_j s_j\langle v, e_j\rangle e_j$ (the operator on the right is positive and its square is T^*T), and therefore $S\sqrt{T^*T}\,v = \sum_j s_j\langle v, e_j\rangle f_j = Tv$. Thus this exercise is the precise statement of the remark made in the text just after the proof of 7.93, that the unitary operator produced there is as close to T as a unitary operator can be.

Problem 7F.28. Suppose $T \in \mathcal{L}(V)$. Prove that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $T = \sqrt{TT^*}\,S$.

Solution. First note that the expression $\sqrt{TT^*}$ makes sense: writing $A = T^*$, we have $A^*A = (T^*)^*T^* = TT^*$, so TT^* is a positive operator on V by 7.64(a), and hence has a unique positive square root $\sqrt{TT^*}$ (by 7.39 and 7.40). In particular $\sqrt{TT^*}$ is positive and therefore self-adjoint.
Apply the polar decomposition 7.93 to the operator $T^* \in \mathcal{L}(V)$. It gives a unitary operator $R \in \mathcal{L}(V)$ with

$$T^* = R\sqrt{(T^*)^*T^*} = R\sqrt{TT^*},$$

where we used $(T^*)^* = T$ [by 7.5(c), the adjoint of the adjoint is the original map]. Note that 7.93 is valid over both $\mathbf{R}$ and $\mathbf{C}$ and for every operator, invertible or not, so no extra hypothesis is needed here.

Now take adjoints of both sides. Using $(BC)^* = C^*B^*$ [by 7.5(d)] and the self-adjointness of $\sqrt{TT^*}$,

$$T = (T^*)^* = \big(R\sqrt{TT^*}\big)^* = \big(\sqrt{TT^*}\big)^* R^* = \sqrt{TT^*}\, R^*.$$

Finally, set $S = R^*$. Since R is unitary, so is R^* by the equivalence of (a) and (f) in 7.53. Therefore S is a unitary operator and

$$T = \sqrt{TT^*}\, S,$$

as desired.
Remark. This is the left polar decomposition, complementing the right polar decomposition $T = S'\sqrt{T^*T}$ of 7.93. Both factorizations mirror the polar form $z = (z/|z|)|z| = |z|(z/|z|)$ of a nonzero complex number, with the positive operator playing the role of $|z|$ and the unitary operator the role of the unimodular factor. Consistency of the two decompositions is reflected in Exercise 29, which shows that T and T^* have the same singular values, so $\sqrt{T^*T}$ and $\sqrt{TT^*}$ have the same eigenvalues with the same multiplicities.

Problem 7F.29. Suppose $T \in \mathcal{L}(V)$.
(a) Use the polar decomposition to show that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $TT^* = ST^*TS^*$.
(b) Show how (a) implies that T and T^* have the same singular values.

Solution. Part (a). By the polar decomposition 7.93 there is a unitary operator $S \in \mathcal{L}(V)$ with

$$T = S\sqrt{T^*T}.$$

Write $R = \sqrt{T^*T}$. By definition R is a positive operator, so R is self-adjoint, i.e. $R^* = R$; and $R^2 = T^*T$ because R is a square root of T^*T.
Taking adjoints in $T = SR$ gives $T^* = R^*S^* = RS^*$. Hence

$$TT^* = (SR)(RS^*) = SR^2S^* = S(T^*T)S^*,$$

which is the desired identity.
Part (b). By the definition of singular values 7.65, the singular values of T are the nonnegative square roots of the eigenvalues of T^*T, each repeated as many times as the dimension of the corresponding eigenspace of T^*T, listed in decreasing order. Likewise the singular values of T^* are the nonnegative square roots of the eigenvalues of $(T^*)^*T^* = TT^*$, with multiplicities given by the dimensions of the eigenspaces of TT^*.
So it suffices to show that T^*T and TT^* have the same eigenvalues and that corresponding eigenspaces have the same dimension.
Let S be as in (a). Because S is unitary we have $S^* = S^{-1}$ (by the equivalence of (a) and (c) in 7.53), so (a) says

$$TT^* = S(T^*T)S^{-1}.$$

Fix $\lambda \in \mathbf{F}$ and consider the map $v \mapsto Sv$ from V to V. If $v \in E(\lambda, T^*T)$, that is $T^*Tv = \lambda v$, then

$$(TT^*)(Sv) = S(T^*T)S^{-1}Sv = S(T^*Tv) = S(\lambda v) = \lambda(Sv),$$

so $Sv \in E(\lambda, TT^*)$. Thus S restricts to a linear map from $E(\lambda, T^*T)$ into $E(\lambda, TT^*)$. Reversing the roles of the two operators: from $TT^* = S(T^*T)S^{-1}$ we get $T^*T = S^{-1}(TT^*)S$, and the same computation shows that S^{-1} maps $E(\lambda, TT^*)$ into $E(\lambda, T^*T)$. Since S^{-1} and S are inverse to each other, these two restrictions are inverse bijections. Hence

$$\dim E(\lambda, T^*T) = \dim E(\lambda, TT^*) \quad \text{for every } \lambda \in \mathbf{F}.$$

In particular λ is an eigenvalue of T^*T if and only if it is an eigenvalue of TT^* (an eigenvalue is precisely a λ with $\dim E(\lambda, \cdot) > 0$), and the multiplicities agree.
Therefore the list of eigenvalues of T^*T, each repeated according to the dimension of its eigenspace, coincides with the corresponding list for TT^*. Taking nonnegative square roots and listing in decreasing order produces the same list in both cases. Thus T and T^* have the same singular values.

Problem 7F.30. Suppose $T \in \mathcal{L}(V)$, $S \in \mathcal{L}(V)$ is a unitary operator, and $R \in \mathcal{L}(V)$ is a positive operator such that $T = SR$. Prove that $R = \sqrt{T^*T}$.
[This exercise shows that if we write T as the product of a unitary operator and a positive operator (as in the polar decomposition 7.93), then the positive operator equals $\sqrt{T^*T}$.]

Solution. Recall that $\sqrt{T^*T}$ denotes the unique positive square root of the positive operator T^*T; the operator T^*T is positive by 7.64(a), and it has exactly one positive square root by 7.39. Thus to prove that $R = \sqrt{T^*T}$ it suffices to verify two things: that R is positive, and that $R^2 = T^*T$.
The first is part of the hypothesis: R is a positive operator.
For the second, note first that a positive operator is by definition self-adjoint, so $R^* = R$. Also, because S is unitary we have $S^*S = I$ (by the equivalence of (a) and (b) in 7.53). Now take adjoints in $T = SR$ to get $T^* = R^*S^* = RS^*$, and compute

$$T^*T = (RS^*)(SR) = R(S^*S)R = RIR = R^2.$$

So R is a positive operator whose square is T^*T; that is, R is a positive square root of T^*T. By the uniqueness in 7.39,

$$R = \sqrt{T^*T},$$

as desired.
Note that no such uniqueness statement holds for the unitary factor in general. For instance, if $T = 0$ then $T = SR$ with $R = 0 = \sqrt{T^*T}$ for every unitary S. Exercise 26 in this section shows that the unitary factor is unique exactly when T is invertible.

Problem 7F.31. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Prove that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$ and such that S and $\sqrt{T^*T}$ both have diagonal matrices with respect to the same orthonormal basis of V.

Solution. Because $\mathbf{F} = \mathbf{C}$ and T is normal, the complex spectral theorem 7.31 (equivalence of (a) and (c)) provides an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \dots, \lambda_n \in \mathbf{C}$ be the corresponding eigenvalues, so

$$Te_k = \lambda_k e_k \qquad \text{for } k = 1, \dots, n.$$

First we identify $\sqrt{T^*T}$ on this basis. Since T is normal, 7.21(e) gives $T^*e_k = \overline{\lambda_k}\, e_k$ for each k. Hence

$$T^*Te_k = T^*(\lambda_k e_k) = \lambda_k \overline{\lambda_k}\, e_k = |\lambda_k|^2 e_k.$$

Define $R \in \mathcal{L}(V)$ by $Re_k = |\lambda_k| e_k$ for each k (and extend linearly). The matrix of R with respect to the orthonormal basis $e_1, \dots, e_n$ is diagonal with real nonnegative diagonal entries $|\lambda_1|, \dots, |\lambda_n|$; hence R is self-adjoint, and for $v = a_1e_1 + \cdots + a_ne_n$ we have

$$\langle Rv, v\rangle = \Big\langle \sum_{k=1}^{n} a_k|\lambda_k|e_k, \sum_{j=1}^{n} a_j e_j \Big\rangle = \sum_{k=1}^{n} |\lambda_k|\,|a_k|^2 \geq 0,$$

so R is a positive operator. Moreover $R^2e_k = |\lambda_k|^2 e_k = T^*Te_k$ for every k, so $R^2 = T^*T$. Thus R is a positive square root of T^*T, and by the uniqueness of positive square roots (7.39),

$$\sqrt{T^*T} = R, \qquad \text{so} \qquad \sqrt{T^*T}\, e_k = |\lambda_k| e_k \ \text{ for each } k.$$

In particular $\sqrt{T^*T}$ has a diagonal matrix with respect to $e_1, \dots, e_n$.
Now define scalars

$$\mu_k = \begin{cases} \dfrac{\lambda_k}{|\lambda_k|} & \text{if } \lambda_k \neq 0, \\ 1 & \text{if } \lambda_k = 0, \end{cases}$$

and define $S \in \mathcal{L}(V)$ by $Se_k = \mu_k e_k$ for $k = 1, \ldots, n$ (extended linearly). Note that $|\mu_k| = 1$ for every k.
The operator S is unitary. Indeed, $Se_1, \ldots, Se_n$ is the list $\mu_1 e_1, \ldots, \mu_n e_n$, and

$$\langle \mu_j e_j, \mu_k e_k \rangle = \mu_j \overline{\mu_k} \langle e_j, e_k \rangle,$$

which equals 0 if $j \neq k$ and equals $|\mu_k|^2 = 1$ if $j = k$. So $Se_1, \ldots, Se_n$ is an orthonormal basis of V, and therefore S is unitary by the equivalence of (d) and (a) in 7.53.
Finally, for each k,

$$S\sqrt{T^*T}\, e_k = S(|\lambda_k| e_k) = \mu_k |\lambda_k| e_k = \lambda_k e_k = T e_k,$$

where the third equality holds because $\mu_k |\lambda_k| = \lambda_k$ when $\lambda_k \neq 0$ and both sides equal 0 when $\lambda_k = 0$. Since $S\sqrt{T^*T}$ and T agree on the basis $e_1, \ldots, e_n$, they are equal:

$$T = S\sqrt{T^*T}.$$

Both S and $\sqrt{T^*T}$ have diagonal matrices with respect to the single orthonormal basis $e_1, \ldots, e_n$ (with diagonal entries $\mu_1, \ldots, \mu_n$ and $|\lambda_1|, \ldots, |\lambda_n|$ respectively), which completes the proof.
This is the polar decomposition analogue of writing a nonzero complex number z as $(z/|z|)\,|z|$, carried out coordinatewise in an orthonormal eigenbasis of T. The converse of the exercise also holds, so the hypothesis that T is normal cannot be weakened: if $T = SR$ where S and R both have diagonal matrices with respect to a single orthonormal basis of V, then the product SR also has a diagonal matrix with respect to that basis, so T has a diagonal matrix with respect to an orthonormal basis of V and hence T is normal by the implication (b) $\Rightarrow$ (a) of 7.31.

Problem 7F.32. Suppose that $T \in \mathcal{L}(V, W)$ and $T \neq 0$. Let $s_1, \ldots, s_m$ denote the positive singular values of T. Show that there exists an orthonormal basis $e_1, \ldots, e_m$ of $(\operatorname{null} T)^\perp$ such that

$$T\left(E\left(\frac{e_1}{s_1}, \ldots, \frac{e_m}{s_m}\right)\right)$$

equals the ball in $\operatorname{range} T$ of radius 1 centered at 0.

Solution. Throughout, $(\operatorname{null} T)^\perp$ and $\operatorname{range} T$ are regarded as inner product spaces with the inner products they inherit from V and W respectively. The ellipsoid $E(e_1/s_1, \ldots, e_m/s_m)$ is then formed inside the inner product space $(\operatorname{null} T)^\perp$, of which $e_1, \ldots, e_m$ will be an orthonormal basis and $1/s_1, \ldots, 1/s_m$ are positive numbers, exactly as required by definition 7.96. Write

$$B_{\operatorname{range} T} = \{w \in \operatorname{range} T : \|w\| < 1\}$$

for the ball in $\operatorname{range} T$ of radius 1 centered at 0.
Since $T \neq 0$ we have $m \geq 1$ (by 7.68(b), the number of positive singular values equals $\dim \operatorname{range} T$).
Choice of the basis. By the singular value decomposition 7.70 there are orthonormal lists $e_1, \ldots, e_m$ in V and $f_1, \ldots, f_m$ in W such that

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m$$

for every $v \in V$. We claim these lists have the two properties we need.
First, $e_1, \ldots, e_m$ is an orthonormal basis of $(\operatorname{null} T)^\perp$. Let $U = \operatorname{span}(e_1, \ldots, e_m)$. Because $f_1, \ldots, f_m$ is orthonormal, the Pythagorean theorem applied to the displayed formula gives

$$\|Tv\|^2 = s_1^2 |\langle v, e_1 \rangle|^2 + \cdots + s_m^2 |\langle v, e_m \rangle|^2$$

for every $v \in V$. Since each $s_k > 0$, this shows

$$Tv = 0 \iff \langle v, e_k \rangle = 0 \text{ for } k = 1, \ldots, m \iff v \in U^\perp,$$

the last equivalence because $e_1, \dots, e_m$ spans U. Hence $\operatorname{null} T = U^\perp$, and therefore

$$(\operatorname{null} T)^\perp = (U^\perp)^\perp = U,$$

using 6.52 (the orthogonal complement of the orthogonal complement of a finite-dimensional subspace is that subspace). So the orthonormal list $e_1, \dots, e_m$ spans $(\operatorname{null} T)^\perp$ and is therefore an orthonormal basis of it.
Second, $f_1, \dots, f_m$ is an orthonormal basis of $\operatorname{range} T$. The displayed formula for Tv shows $\operatorname{range} T \subseteq \operatorname{span}(f_1, \dots, f_m)$, while taking $v = e_k$ gives $Te_k = s_k f_k$, so $f_k = Te_k/s_k \in \operatorname{range} T$ for each k. Hence $\operatorname{range} T = \operatorname{span}(f_1, \dots, f_m)$, and this list is orthonormal, so it is an orthonormal basis of $\operatorname{range} T$.
By definition 7.96, applied in the inner product space $(\operatorname{null} T)^\perp$ with orthonormal basis $e_1, \dots, e_m$ and the positive numbers $1/s_1, \dots, 1/s_m$,

$$E\left(\frac{e_1}{s_1}, \dots, \frac{e_m}{s_m}\right) = \left\{v \in (\operatorname{null} T)^\perp : s_1^2|\langle v, e_1\rangle|^2 + \cdots + s_m^2|\langle v, e_m\rangle|^2 < 1\right\},$$

because the k-th denominator is $(1/s_k)^2$, and dividing by $(1/s_k)^2$ means multiplying by s_k^2. Call this set E.
One inclusion. Suppose $v \in E$. Then $Tv \in \operatorname{range} T$, and by the norm formula established above,

$$\|Tv\|^2 = s_1^2|\langle v, e_1\rangle|^2 + \cdots + s_m^2|\langle v, e_m\rangle|^2 < 1.$$

Hence $Tv \in B_{\operatorname{range} T}$. Thus $T(E) \subseteq B_{\operatorname{range} T}$.
The other inclusion. Suppose $w \in \operatorname{range} T$ with $\|w\| < 1$. Since $f_1, \dots, f_m$ is an orthonormal basis of $\operatorname{range} T$, we have

$$w = \langle w, f_1\rangle f_1 + \cdots + \langle w, f_m\rangle f_m$$

and, by Parseval's identity 6.30(b),

$$|\langle w, f_1\rangle|^2 + \cdots + |\langle w, f_m\rangle|^2 = \|w\|^2 < 1.$$

Define

$$v = \frac{\langle w, f_1\rangle}{s_1} e_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m} e_m.$$

Then $v \in \operatorname{span}(e_1, \dots, e_m) = (\operatorname{null} T)^\perp$, and since $e_1, \dots, e_m$ is orthonormal we get $\langle v, e_k\rangle = \langle w, f_k\rangle/s_k$ for each k. Therefore

$$s_1^2|\langle v, e_1\rangle|^2 + \cdots + s_m^2|\langle v, e_m\rangle|^2 = |\langle w, f_1\rangle|^2 + \cdots + |\langle w, f_m\rangle|^2 = \|w\|^2 < 1,$$

so $v \in E$. Finally,

$$Tv = \sum_{k=1}^m s_k\langle v, e_k\rangle f_k = \sum_{k=1}^m s_k \cdot \frac{\langle w, f_k\rangle}{s_k} f_k = \sum_{k=1}^m \langle w, f_k\rangle f_k = w.$$

Hence $w \in T(E)$, and $B_{\operatorname{range} T} \subseteq T(E)$.
Combining the two inclusions,

$$T\left(E\left(\frac{e_1}{s_1}, \dots, \frac{e_m}{s_m}\right)\right) = \{w \in \operatorname{range} T : \|w\| < 1\},$$

which is the ball in $\operatorname{range} T$ of radius 1 centered at 0, as desired.
Geometrically this is the statement of 7.99 read backwards and adapted to a possibly non-invertible T: the map T collapses $\operatorname{null} T$ and acts as an invertible map from $(\operatorname{null} T)^\perp$ onto $\operatorname{range} T$, stretching by s_k along the e_k-axis; so the ellipsoid whose e_k-axis has been pre-shrunk by the factor $1/s_k$ is carried exactly onto the unit ball of $\operatorname{range} T$.

8 Operators on Complex Vector Spaces

Contents

8.1 Exercises 8A 487
8.2 Exercises 8B 504
8.3 Exercises 8C 526
8.4 Exercises 8D 539

8.1 Exercises 8A

Problem 8A.1. Suppose $T \in \mathcal{L}(V)$. Prove that if $\dim \operatorname{null} T^4 = 8$ and $\dim \operatorname{null} T^6 = 9$, then $\dim \operatorname{null} T^m = 9$ for all integers $m \geq 5$.

Solution. By 8.1 we have the chain of inclusions

$$\operatorname{null} T^4 \subseteq \operatorname{null} T^5 \subseteq \operatorname{null} T^6,$$

and hence

$$8 = \dim \operatorname{null} T^4 \leq \dim \operatorname{null} T^5 \leq \dim \operatorname{null} T^6 = 9.$$

Thus $\dim \operatorname{null} T^5$ equals either 8 or 9.
Suppose, seeking a contradiction, that $\dim \operatorname{null} T^5 = 8$. Because $\operatorname{null} T^4 \subseteq \operatorname{null} T^5$ and these two subspaces have the same (finite) dimension, they are equal (by 2.39). Now 8.2, applied with $m = 4$, gives

$$\operatorname{null} T^4 = \operatorname{null} T^5 = \operatorname{null} T^6 = \cdots,$$

and in particular $\dim \operatorname{null} T^6 = 8$, contradicting the hypothesis that $\dim \operatorname{null} T^6 = 9$. Hence

$$\dim \operatorname{null} T^5 = 9.$$

Now $\operatorname{null} T^5 \subseteq \operatorname{null} T^6$ and both subspaces have dimension 9, so again by 2.39 we get $\operatorname{null} T^5 = \operatorname{null} T^6$. Applying 8.2 with $m = 5$ now yields

$$\operatorname{null} T^5 = \operatorname{null} T^6 = \operatorname{null} T^7 = \cdots,$$

which is to say $\operatorname{null} T^m = \operatorname{null} T^5$ for every integer $m \geq 5$. Taking dimensions gives $\dim \operatorname{null} T^m = 9$ for all integers $m \geq 5$, as desired.

Problem 8A.2. Suppose $T \in \mathcal{L}(V)$, m is a positive integer, $v \in V$, and $T^{m-1}v \neq 0$ but $T^m v = 0$. Prove that $v, Tv, T^2v, \ldots, T^{m-1}v$ is linearly independent.
[The result in this exercise is used in the proof of 8.45.]

Solution. First note that

$$T^k v = 0 \quad \text{for every integer } k \geq m,$$

because $T^k v = T^{k-m}(T^m v) = T^{k-m}0 = 0$.
Suppose $a_0, a_1, \ldots, a_{m-1} \in \mathbf{F}$ satisfy

$$a_0 v + a_1 Tv + a_2 T^2 v + \cdots + a_{m-1} T^{m-1} v = 0.$$

We show by induction on j that $a_j = 0$ for $j = 0, 1, \ldots, m-1$.
Base case and induction step together: let $j \in \{0, 1, \ldots, m-1\}$ and suppose $a_0 = a_1 = \cdots = a_{j-1} = 0$ (for $j = 0$ this hypothesis is vacuous). Then the displayed equation reduces to

$$a_j T^j v + a_{j+1} T^{j+1} v + \cdots + a_{m-1} T^{m-1} v = 0.$$

Apply the operator T^{m-1-j} to both sides. For each index k with $j+1 \leq k \leq m-1$ we have $T^{m-1-j} T^k v = T^{m-1-j+k} v = 0$, because $m - 1 - j + k \geq m$. Hence every term except the first is killed, and we are left with

$$a_j T^{m-1} v = 0.$$

Since $T^{m-1}v \neq 0$ by hypothesis, this forces $a_j = 0$.
By induction, $a_0 = a_1 = \cdots = a_{m-1} = 0$. Thus the only linear combination of $v, Tv, T^2v, \ldots, T^{m-1}v$ equal to 0 is the one with all coefficients equal to 0, which means that this list is linearly independent.

Problem 8A.3. Suppose $T \in \mathcal{L}(V)$. Prove that

$$V = \operatorname{null} T \oplus \operatorname{range} T \iff \operatorname{null} T^2 = \operatorname{null} T.$$

Solution. Recall that throughout this chapter V is a finite-dimensional vector space.
First suppose $V = \operatorname{null} T \oplus \operatorname{range} T$. By 8.1 we already have $\operatorname{null} T \subseteq \operatorname{null} T^2$, so only the reverse inclusion needs proof. Let $v \in \operatorname{null} T^2$. Then

$$T(Tv) = T^2 v = 0,$$

so $Tv \in \operatorname{null} T$. Also $Tv \in \operatorname{range} T$. Because the sum $\operatorname{null} T + \operatorname{range} T$ is a direct sum, we have $(\operatorname{null} T) \cap (\operatorname{range} T) = \{0\}$ (by 1.46). Hence $Tv = 0$, which means $v \in \operatorname{null} T$. Thus $\operatorname{null} T^2 \subseteq \operatorname{null} T$, and therefore $\operatorname{null} T^2 = \operatorname{null} T$.
Conversely, suppose $\operatorname{null} T^2 = \operatorname{null} T$. We first show that

$$(\operatorname{null} T) \cap (\operatorname{range} T) = \{0\}.$$

Suppose v lies in this intersection. Then $Tv = 0$ and there exists $u \in V$ with $v = Tu$. Applying T to the last equation gives

$$0 = Tv = T^2 u,$$

so $u \in \operatorname{null} T^2 = \operatorname{null} T$. Hence $v = Tu = 0$, proving the displayed equation.
Because the intersection of $\operatorname{null} T$ and $\operatorname{range} T$ equals $\{0\}$, the sum $\operatorname{null} T + \operatorname{range} T$ is a direct sum (by 1.46). Furthermore, using 3.94 and then the fundamental theorem of linear maps (3.21),

$$\dim(\operatorname{null} T \oplus \operatorname{range} T) = \dim \operatorname{null} T + \dim \operatorname{range} T = \dim V.$$

A subspace of V whose dimension equals $\dim V$ must equal V (by 2.39). Hence

$$V = \operatorname{null} T \oplus \operatorname{range} T,$$

completing the proof of the equivalence.

Problem 8A.4. Suppose $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{F}$, and m is a positive integer such that the minimal polynomial of T is a polynomial multiple of $(z - \lambda)^m$. Prove that

$$\dim \operatorname{null}(T - \lambda I)^m \geq m.$$

Solution. Let p denote the minimal polynomial of T, and write

$$p(z) = (z - \lambda)^m q(z)$$

for some $q \in \mathcal{P}(\mathbf{F})$. Because p is monic (by 5.22) and $(z - \lambda)^m$ is monic, the polynomial q is monic as well; in particular $q \neq 0$.
Set $S = T - \lambda I$. The key step is the following claim.
Claim: there is no integer j with $0 \leq j \leq m - 1$ such that $\operatorname{null} S^j = \operatorname{null} S^{j+1}$.
To prove the claim, suppose such a j exists. By 8.2 the equality then propagates:

$$\operatorname{null} S^j = \operatorname{null} S^{j+1} = \operatorname{null} S^{j+2} = \cdots;$$

in particular, since $m \geq j$, we have $\operatorname{null} S^m = \operatorname{null} S^j$.
Now define

$$r(z) = (z - \lambda)^j q(z).$$

We show that $r(T) = 0$. Let $v \in V$ and put $u = q(T)v$. Then, using the multiplicative property 5.17(a) applied to the factorization $p(z) = (z - \lambda)^m q(z)$,

$$S^m u = (T - \lambda I)^m q(T)v = p(T)v = 0,$$

so $u \in \operatorname{null} S^m = \operatorname{null} S^j$. Hence

$$r(T)v = (T-\lambda I)^j q(T)v = S^j u = 0,$$

again using 5.17(a), this time for the factorization $r(z) = (z-\lambda)^j q(z)$. Because $v \in V$ was arbitrary, $r(T) = 0$.
But r is a monic polynomial (being a product of the monic polynomials $(z-\lambda)^j$ and q) with

$$\deg r = j + \deg q < m + \deg q = \deg p,$$

where the strict inequality uses $j \le m-1 < m$. This contradicts the fact that p is the monic polynomial of smallest degree with $p(T) = 0$ (5.22). The contradiction proves the claim.
By 8.1 we have the chain

$$\{0\} = \operatorname{null} S^0 \subseteq \operatorname{null} S^1 \subseteq \cdots \subseteq \operatorname{null} S^m,$$

and by the claim every one of these m inclusions is strict. At each strict inclusion the dimension increases by at least 1. Starting from $\dim \operatorname{null} S^0 = 0$ and taking m such steps gives

$$\dim \operatorname{null}(T-\lambda I)^m = \dim \operatorname{null} S^m \ge m,$$

as desired.

Problem 8A.5. Suppose $T \in \mathcal{L}(V)$ and m is a positive integer. Prove that

$$\dim \operatorname{null} T^m \le m \dim \operatorname{null} T.$$

[Hint: Exercise 21 in Section 3B may be useful.]

Solution. The whole proof rests on the following one-step estimate.
Step estimate: for every integer $k \ge 1$,

$$\dim \operatorname{null} T^k \le \dim \operatorname{null} T^{k-1} + \dim \operatorname{null} T.$$

To prove this, define a map

$$\varphi\colon \operatorname{null} T^k \to \operatorname{null} T, \qquad \varphi(v) = T^{k-1}v.$$

This is well defined: if $v \in \operatorname{null} T^k$, then $T(T^{k-1}v) = T^k v = 0$, so indeed $T^{k-1}v \in \operatorname{null} T$. It is clearly linear, being the restriction of the linear map T^{k-1} to the subspace $\operatorname{null} T^k$ of V.
The null space of φ is

$$\{v \in \operatorname{null} T^k : T^{k-1}v = 0\} = (\operatorname{null} T^k) \cap (\operatorname{null} T^{k-1}) = \operatorname{null} T^{k-1},$$

where the last equality holds because $\operatorname{null} T^{k-1} \subseteq \operatorname{null} T^k$ by 8.1. The range of φ is a subspace of $\operatorname{null} T$, so $\dim \operatorname{range} \varphi \le \dim \operatorname{null} T$ (by 2.37).
Applying the fundamental theorem of linear maps (3.21) to φ gives

$$\dim \operatorname{null} T^k = \dim \operatorname{null} \varphi + \dim \operatorname{range} \varphi = \dim \operatorname{null} T^{k-1} + \dim \operatorname{range} \varphi \le \dim \operatorname{null} T^{k-1} + \dim \operatorname{null} T,$$

which is the step estimate.
Remark on the hint: the step estimate is exactly Exercise 21 in Section 3B applied to the linear map T^{k-1} and the subspace $U = \operatorname{null} T$ of V. Indeed $\{v \in V : T^{k-1}v \in \operatorname{null} T\} = \operatorname{null} T^k$, so that exercise reads

$$\dim \operatorname{null} T^k = \dim \operatorname{null} T^{k-1} + \dim\big((\operatorname{null} T) \cap (\operatorname{range} T^{k-1})\big),$$

and the second term on the right is at most $\dim \operatorname{null} T$ because $(\operatorname{null} T) \cap (\operatorname{range} T^{k-1})$ is a subspace of $\operatorname{null} T$. (This intersection is precisely $\operatorname{range} \varphi$, so the two derivations agree.)

Now induct on m. For $m = 1$ the desired inequality $\dim \operatorname{null} T \leq \dim \operatorname{null} T$ holds trivially. Suppose $m \geq 2$ and that $\dim \operatorname{null} T^{m-1} \leq (m-1) \dim \operatorname{null} T$. Then the step estimate with $k = m$ gives

$$\dim \operatorname{null} T^m \leq \dim \operatorname{null} T^{m-1} + \dim \operatorname{null} T \leq (m-1) \dim \operatorname{null} T + \dim \operatorname{null} T = m \dim \operatorname{null} T.$$

By induction, $\dim \operatorname{null} T^m \leq m \dim \operatorname{null} T$ for every positive integer m.

Problem 8A.6. Suppose $T \in \mathcal{L}(V)$. Show that

$$V = \operatorname{range} T^0 \supseteq \operatorname{range} T^1 \supseteq \cdots \supseteq \operatorname{range} T^k \supseteq \operatorname{range} T^{k+1} \supseteq \cdots .$$

Solution. By definition $T^0 = I$, the identity operator on V. Every $v \in V$ satisfies $v = Iv = T^0 v$, so

$$\operatorname{range} T^0 = V.$$

Now let k be a nonnegative integer; we show that $\operatorname{range} T^{k+1} \subseteq \operatorname{range} T^k$. Suppose $w \in \operatorname{range} T^{k+1}$. Then there exists $v \in V$ with

$$w = T^{k+1} v = T^k (Tv).$$

Since $Tv \in V$, the vector w is the image under T^k of an element of V, so $w \in \operatorname{range} T^k$. Hence

$$\operatorname{range} T^{k+1} \subseteq \operatorname{range} T^k.$$

Applying this for $k = 0, 1, 2, \ldots$ and combining with $\operatorname{range} T^0 = V$ gives the displayed chain

$$V = \operatorname{range} T^0 \supseteq \operatorname{range} T^1 \supseteq \cdots \supseteq \operatorname{range} T^k \supseteq \operatorname{range} T^{k+1} \supseteq \cdots ,$$

as desired. (This is the analogue for ranges of the result 8.1 for null spaces.)

Problem 8A.7. Suppose $T \in \mathcal{L}(V)$ and m is a nonnegative integer such that

$$\operatorname{range} T^m = \operatorname{range} T^{m+1}.$$

Prove that $\operatorname{range} T^k = \operatorname{range} T^m$ for all $k > m$.

Solution. It suffices to prove that

$$\operatorname{range} T^{m+j} = \operatorname{range} T^{m+j+1} \qquad \text{for every integer } j \geq 0,$$

for then a trivial induction on j gives $\operatorname{range} T^m = \operatorname{range} T^{m+1} = \operatorname{range} T^{m+2} = \cdots$, which is the assertion that $\operatorname{range} T^k = \operatorname{range} T^m$ for all $k > m$ (and indeed for all $k \geq m$).
So fix an integer $j \geq 0$. The inclusion

$$\operatorname{range} T^{m+j+1} \subseteq \operatorname{range} T^{m+j}$$

holds by Exercise 6 in this section; directly, if $w = T^{m+j+1} u$ for some $u \in V$, then $w = T^{m+j}(Tu) \in \operatorname{range} T^{m+j}$.
For the reverse inclusion, suppose $w \in \operatorname{range} T^{m+j}$. Then there exists $u \in V$ with

$$w = T^{m+j} u = T^j \big(T^m u\big).$$

Now $T^m u \in \operatorname{range} T^m = \operatorname{range} T^{m+1}$ by hypothesis, so there exists $x \in V$ with

$$T^m u = T^{m+1} x.$$

Substituting into the previous equation gives

$$w = T^j \big(T^{m+1} x\big) = T^{m+j+1} x,$$

so $w \in \operatorname{range} T^{m+j+1}$. Hence $\operatorname{range} T^{m+j} \subseteq \operatorname{range} T^{m+j+1}$, and combining the two inclusions gives

$$\operatorname{range} T^{m+j} = \operatorname{range} T^{m+j+1}.$$

As explained above, this yields $\operatorname{range} T^k = \operatorname{range} T^m$ for all $k > m$, as desired.

Problem 8A.8. Suppose $T \in \mathcal{L}(V)$. Prove that

$$\operatorname{range} T^{\dim V} = \operatorname{range} T^{\dim V+1} = \operatorname{range} T^{\dim V+2} = \cdots.$$

Solution. Let $n = \dim V$.
First note that the ranges of the powers of T form a decreasing chain: if k is a nonnegative integer and $w \in \operatorname{range} T^{k+1}$, then $w = T^{k+1}u = T^k(Tu)$ for some $u \in V$, so $w \in \operatorname{range} T^k$. Thus

$$\operatorname{range} T^{k+1} \subseteq \operatorname{range} T^k \quad \text{for every nonnegative integer } k$$

(this is Exercise 6 in this section). In particular, for every integer $k \geq n$ we have $\operatorname{range} T^k \subseteq \operatorname{range} T^n$.
Now we compare dimensions. By 8.3 (null spaces stop growing),

$$\operatorname{null} T^n = \operatorname{null} T^{n+1} = \operatorname{null} T^{n+2} = \cdots,$$

so $\dim \operatorname{null} T^k = \dim \operatorname{null} T^n$ for every integer $k \geq n$. Applying the fundamental theorem of linear maps (3.21) to the operator $T^k \in \mathcal{L}(V)$ gives

$$\dim \operatorname{range} T^k = \dim V - \dim \operatorname{null} T^k = \dim V - \dim \operatorname{null} T^n = \dim \operatorname{range} T^n$$

for every integer $k \geq n$.
Thus for each integer $k \geq n$ the subspace $\operatorname{range} T^k$ is contained in $\operatorname{range} T^n$ and has the same dimension as $\operatorname{range} T^n$. Hence $\operatorname{range} T^k = \operatorname{range} T^n$ (by 2.39, a subspace of a finite-dimensional vector space that has the same dimension as the whole space equals the whole space, applied inside $\operatorname{range} T^n$). Therefore

$$\operatorname{range} T^{\dim V} = \operatorname{range} T^{\dim V+1} = \operatorname{range} T^{\dim V+2} = \cdots,$$

as desired.

Problem 8A.9. Suppose $T \in \mathcal{L}(V)$ and m is a nonnegative integer. Prove that

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} \iff \operatorname{range} T^m = \operatorname{range} T^{m+1}.$$

Solution. Two inclusions hold no matter what m is:

$$\operatorname{null} T^m \subseteq \operatorname{null} T^{m+1} \quad \text{and} \quad \operatorname{range} T^{m+1} \subseteq \operatorname{range} T^m.$$

The first is 8.1. The second holds because every element of $\operatorname{range} T^{m+1}$ has the form $T^{m+1}u = T^m(Tu)$ for some $u \in V$, hence lies in $\operatorname{range} T^m$ (this is Exercise 6 in this section).
Because V is finite-dimensional, each of these inclusions is an equality if and only if the two subspaces have the same dimension (by 2.39). Hence

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} \iff \dim \operatorname{null} T^m = \dim \operatorname{null} T^{m+1}$$

and

$$\operatorname{range} T^m = \operatorname{range} T^{m+1} \iff \dim \operatorname{range} T^m = \dim \operatorname{range} T^{m+1}.$$

Now apply the fundamental theorem of linear maps (3.21) to the operators T^m and T^{m+1}, each of which maps V to V:

$$\dim \operatorname{null} T^m + \dim \operatorname{range} T^m = \dim V,$$
$$\dim \operatorname{null} T^{m+1} + \dim \operatorname{range} T^{m+1} = \dim V.$$

Subtracting, we get

$$\dim \operatorname{null} T^{m+1} - \dim \operatorname{null} T^m = \dim \operatorname{range} T^m - \dim \operatorname{range} T^{m+1}.$$

Thus the left side is 0 if and only if the right side is 0; that is,

$$\dim \operatorname{null} T^m = \dim \operatorname{null} T^{m+1} \iff \dim \operatorname{range} T^m = \dim \operatorname{range} T^{m+1}.$$

Combining the three displayed equivalences gives

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} \iff \operatorname{range} T^m = \operatorname{range} T^{m+1},$$

as desired.

Problem 8A.10. Define $T \in \mathcal{L}(\mathbf{C}^2)$ by $T(w, z) = (z, 0)$. Find all generalized eigenvectors of T.

Solution. Every nonzero vector of $\mathbf{C}^2$ is a generalized eigenvector of T, and each one corresponds to the eigenvalue 0.
To see this, first find the eigenvalues of T. Suppose $\lambda \in \mathbf{C}$ and $(w, z) \in \mathbf{C}^2$ is nonzero with $T(w, z) = \lambda(w, z)$, which means

$$(z, 0) = (\lambda w, \lambda z).$$

Thus $z = \lambda w$ and $\lambda z = 0$. If $\lambda \neq 0$, then the second equation gives $z = 0$, and then the first gives $\lambda w = 0$, hence $w = 0$; this contradicts $(w, z) \neq (0, 0)$. Hence $\lambda = 0$, so 0 is the only possible eigenvalue. And 0 really is an eigenvalue: $T(1, 0) = (0, 0)$, so $(1, 0)$ is an eigenvector corresponding to 0. Thus 0 is the only eigenvalue of T.
Because the only eigenvalue of T is 0, a nonzero vector $v \in \mathbf{C}^2$ is a generalized eigenvector of T if and only if

$$(T - 0I)^k v = T^k v = 0$$

for some positive integer k, which by 8.1 and 8.3 (with $\dim \mathbf{C}^2 = 2$) happens if and only if $T^2 v = 0$.
Now compute: for all $(w, z) \in \mathbf{C}^2$,

$$T^2(w, z) = T(z, 0) = (0, 0).$$

So $T^2 = 0$, and hence $T^2 v = 0$ for every $v \in \mathbf{C}^2$.
Therefore the set of generalized eigenvectors of T is

$$\{(w, z) \in \mathbf{C}^2 : (w, z) \neq (0, 0)\},$$

all corresponding to the eigenvalue 0. (By contrast, the eigenvectors of T are only the vectors of the form $(w, 0)$ with $w \neq 0$: the equation $T(w, z) = 0$ forces $z = 0$. So T has far more generalized eigenvectors than eigenvectors; equivalently, T is nilpotent, which by 8.14 means every nonzero vector of $\mathbf{C}^2$ is a generalized eigenvector corresponding to 0.)

Problem 8A.11. Suppose that $T \in \mathcal{L}(V)$. Prove that there is a basis of V consisting of generalized eigenvectors of T if and only if the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbf{F}$.
Assume $\mathbf{F} = \mathbf{R}$ because the case $\mathbf{F} = \mathbf{C}$ follows from 5.27(b) and 8.9.
This exercise states that the condition for there to be a basis of V consisting of generalized eigenvectors of T is the same as the condition for there to be a basis with respect to which T has an upper-triangular matrix (see 5.44).
Caution: If T has an upper-triangular matrix with respect to a basis $v_1, \ldots, v_n$ of V, then v_1 is an eigenvector of T but it is not necessarily true that $v_2, \ldots, v_n$ are generalized eigenvectors of T.

Solution. Throughout, $n = \dim V$; recall the standing assumptions of this chapter, under which V is finite-dimensional and $V \neq \{0\}$, so $n \geq 1$. Note also that the minimal polynomial of an operator on a nonzero vector space is nonconstant: the only monic constant polynomial is 1, and if the minimal polynomial of an operator R were 1 we would have $I = 1(R) = 0$, forcing $V = \{0\}$. Say that a monic polynomial in $\mathcal{P}(\mathbf{F})$ splits if it equals $(z-\lambda_1)\cdots(z-\lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbf{F}$ (repetitions allowed). Recall that a nonzero $v \in V$ is a generalized eigenvector of T corresponding to an eigenvalue λ of T exactly when $(T - \lambda I)^n v = 0$ (this is the remark following 8.8, which comes from applying 8.1 and 8.3 to $T - \lambda I$). The argument below works verbatim for $\mathbf{F} = \mathbf{R}$ and for $\mathbf{F} = \mathbf{C}$.

A lemma on divisors. Suppose $p \in \mathcal{P}(\mathbf{F})$ is monic and splits, and $q \in \mathcal{P}(\mathbf{F})$ is a monic nonconstant polynomial that divides p (meaning $p = qh$ for some $h \in \mathcal{P}(\mathbf{F})$). Then q splits. Indeed, regard q as an element of $\mathcal{P}(\mathbf{C})$ and use 4.13 to write

$$q(z) = (z - \alpha_1)\cdots(z - \alpha_r)$$

with $\alpha_1, \ldots, \alpha_r \in \mathbf{C}$ (the leading coefficient is 1 because q is monic). Each α_j is a zero of q, hence a zero of $p = qh$. But writing $p(z) = (z - \lambda_1)\cdots(z - \lambda_m)$ with $\lambda_1, \ldots, \lambda_m \in \mathbf{F}$, we see that the zeros of p in $\mathbf{C}$ are exactly $\lambda_1, \ldots, \lambda_m$, all of which lie in $\mathbf{F}$. Hence every $\alpha_j \in \mathbf{F}$, so q splits.

First suppose the minimal polynomial of T splits; we produce a basis of generalized eigenvectors. We induct on $n = \dim V$, proving the following statement for every $n \geq 1$: if $\dim V = n$ and the minimal polynomial of $T \in \mathcal{L}(V)$ splits, then V has a basis consisting of generalized eigenvectors of T.

If $n = 1$, then every nonzero vector of V is an eigenvector of T, hence a generalized eigenvector, and any nonzero vector is a basis of V.

Now suppose $n > 1$ and the statement holds for all vector spaces of smaller dimension. Let $p(z) = (z - \lambda_1)\cdots(z - \lambda_m)$ be the minimal polynomial of T; here $m \geq 1$ because the minimal polynomial of an operator on a nonzero vector space is nonconstant. By 5.27(a), λ_1 is an eigenvalue of T. Applying 8.4 to the operator $T - \lambda_1 I$ gives

$$V = \operatorname{null}(T - \lambda_1 I)^n \oplus \operatorname{range}(T - \lambda_1 I)^n.$$

If $\operatorname{null}(T - \lambda_1 I)^n = V$, then every nonzero vector of V is a generalized eigenvector of T corresponding to λ_1, so every basis of V consists of generalized eigenvectors of T and we are done.

So assume $\operatorname{null}(T - \lambda_1 I)^n \neq V$, and set $U = \operatorname{range}(T - \lambda_1 I)^n$. The displayed direct sum then forces $U \neq \{0\}$. Also $\operatorname{null}(T - \lambda_1 I)^n \neq \{0\}$ because λ_1 is an eigenvalue of T, so the direct sum forces $\dim U < n$. Thus

$$0 < \dim U < n.$$

Moreover U is invariant under T, by 5.18 applied with the polynomial $p_1(z) = (z - \lambda_1)^n$. Let $S = T|_U \in \mathcal{L}(U)$.

By 5.31, the minimal polynomial p of T is a polynomial multiple of the minimal polynomial q of S; that is, q divides p. Since q is monic and nonconstant (as $U \neq \{0\}$) and p splits, the lemma shows that q splits. Because $\dim U < n$, the induction hypothesis applied to S gives a basis $u_1, \ldots, u_k$ of U consisting of generalized eigenvectors of S.

Each such u_j is also a generalized eigenvector of T: there are an eigenvalue μ of S and a positive integer r with $(S - \mu I_U)^r u_j = 0$; since S agrees with T on the T-invariant subspace U, we get $(T - \mu I)^r u_j = 0$ with $u_j \neq 0$, and μ is an eigenvalue of T because an eigenvector of S in U is an eigenvector of T.

Finally, let $w_1, \ldots, w_l$ be a basis of $\operatorname{null}(T - \lambda_1 I)^n$; each w_i is nonzero and satisfies $(T - \lambda_1 I)^n w_i = 0$, so each w_i is a generalized eigenvector of T corresponding to λ_1. Because V is the direct sum of $\operatorname{null}(T - \lambda_1 I)^n$ and U, the list

$$w_1, \ldots, w_l, u_1, \ldots, u_k$$

is a basis of V, and it consists of generalized eigenvectors of T. This completes the induction.

Conversely, suppose V has a basis $v_1, \ldots, v_n$ consisting of generalized eigenvectors of T; we show the minimal polynomial of T splits. Let $\mu_1, \ldots, \mu_k$ be the distinct eigenvalues of T; by definition these lie in $\mathbf{F}$. (There is at least one, since v_1 is a generalized eigenvector and so corresponds to some eigenvalue.) Define

$$p(z) = (z - \mu_1)^n \cdots (z - \mu_k)^n,$$

a monic polynomial in $\mathcal{P}(\mathbf{F})$ that splits.
We claim $p(T) = 0$. Fix $j \in \{1, \dots, n\}$. By 8.11, the generalized eigenvector v_j corresponds to exactly one eigenvalue of T, say μ_i, and then $(T - \mu_i I)^n v_j = 0$. Since polynomials in T commute with one another (by 5.17), we may write

$$p(T) = \Big(\prod_{l \neq i} (T - \mu_l I)^n\Big)(T - \mu_i I)^n,$$

and therefore $p(T)v_j = 0$. Thus $p(T)$ vanishes on each vector of a basis of V, so $p(T) = 0$.
By 5.29, p is a polynomial multiple of the minimal polynomial q of T; in other words q divides p. Since q is monic and nonconstant and p splits, the lemma gives that q splits, i.e. the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \dots, \lambda_m \in \mathbf{F}$, as desired.
Remark. Combining this with 5.44 (which says T has an upper-triangular matrix with respect to some basis of V if and only if the minimal polynomial of T splits over $\mathbf{F}$) shows that V has a basis of generalized eigenvectors of T if and only if T has an upper-triangular matrix with respect to some basis. As the caution in the exercise notes, the two bases need not be the same one.

Problem 8A.12. Suppose $T \in \mathcal{L}(V)$ is such that every nonzero vector in V is a generalized eigenvector of T. Prove that there exists $\lambda \in \mathbf{F}$ such that $T - \lambda I$ is nilpotent.

Solution. Let $n = \dim V$. Recall that a nonzero $v \in V$ is a generalized eigenvector of T corresponding to an eigenvalue α of T if and only if $(T - \alpha I)^n v = 0$ (the remark after 8.8, obtained by applying 8.1 and 8.3 to $T - \alpha I$).
Because $V \neq \{0\}$ (a standing assumption of this chapter), there is a nonzero vector in V, and by hypothesis it is a generalized eigenvector of T; in particular T has at least one eigenvalue.
The heart of the proof is to show that T has exactly one eigenvalue. Suppose instead that T has two distinct eigenvalues α and β. Choose corresponding eigenvectors: nonzero $u, w \in V$ with

$$Tu = \alpha u \quad \text{and} \quad Tw = \beta w.$$

By 5.11, eigenvectors corresponding to distinct eigenvalues are linearly independent, so u, w is a linearly independent list. In particular $u + w \neq 0$.
By hypothesis $u + w$ is a generalized eigenvector of T, so there is an eigenvalue γ of T with

$$(T - \gamma I)^n (u + w) = 0.$$

Now $(T - \gamma I)u = (\alpha - \gamma)u$, and iterating gives $(T - \gamma I)^n u = (\alpha - \gamma)^n u$; similarly $(T - \gamma I)^n w = (\beta - \gamma)^n w$. Hence the equation above becomes

$$(\alpha - \gamma)^n u + (\beta - \gamma)^n w = 0.$$

Because u, w is linearly independent, both coefficients vanish:

$$(\alpha - \gamma)^n = 0 \quad \text{and} \quad (\beta - \gamma)^n = 0.$$

Thus $\alpha = \gamma$ and $\beta = \gamma$, so $\alpha = \beta$, contradicting the assumption that α and β are distinct.
Therefore T has exactly one eigenvalue; call it $\lambda \in \mathbf{F}$.
Now let $v \in V$. If $v = 0$, then trivially $(T - \lambda I)^n v = 0$. If $v \neq 0$, then by hypothesis v is a generalized eigenvector of T corresponding to some eigenvalue of T, and that eigenvalue must be λ since λ is the only eigenvalue; hence $(T - \lambda I)^n v = 0$. So $(T - \lambda I)^n v = 0$ for every $v \in V$, which means

$$(T - \lambda I)^n = 0.$$

Thus $T - \lambda I$ is nilpotent, as desired.

Problem 8A.13. Suppose $S, T \in \mathcal{L}(V)$ and ST is nilpotent. Prove that TS is nilpotent.

Solution. Because ST is nilpotent, there is a positive integer k with $(ST)^k = 0$.
The key observation is the identity

$$(TS)^{k+1} = \underbrace{(TS)(TS)\cdots(TS)}_{k+1 \text{ factors}} = T\,\underbrace{(ST)(ST)\cdots(ST)}_{k \text{ factors}}\,S = T(ST)^k S,$$

which holds because composition of linear maps is associative: writing out the $k+1$ copies of TS gives the string

$$T\,S\,T\,S\,T\,S\cdots T\,S$$

of $2k+2$ factors, and regrouping it as the first T, followed by k consecutive blocks ST, followed by the last S, produces $T(ST)^k S$.
Therefore

$$(TS)^{k+1} = T(ST)^k S = T\,0\,S = 0.$$

Since $k+1$ is a positive integer, TS is nilpotent, as desired.
(As a check on the bookkeeping: by 8.16 we could also have taken $k = \dim V$ from the start, since $(ST)^{\dim V} = 0$, and then concluded $(TS)^{1+\dim V} = 0$.)

Problem 8A.14. Suppose $T \in \mathcal{L}(V)$ is nilpotent and $T \neq 0$. Prove T is not diagonalizable.

Solution. Suppose, for the sake of contradiction, that T is diagonalizable. By 5.55, there is a basis $v_1, \ldots, v_n$ of V consisting of eigenvectors of T. Thus for each $j \in \{1, \ldots, n\}$ there is an eigenvalue λ_j of T with

$$Tv_j = \lambda_j v_j.$$

Because T is nilpotent, 8.17(a) tells us that 0 is an eigenvalue of T and T has no other eigenvalues. Hence $\lambda_j = 0$ for every j, so

$$Tv_j = 0 \quad \text{for } j = 1, \ldots, n.$$

Now let $v \in V$ be arbitrary. Write $v = a_1 v_1 + \cdots + a_n v_n$ with $a_1, \ldots, a_n \in \mathbf{F}$, which is possible because $v_1, \ldots, v_n$ is a basis of V. Then by linearity

$$Tv = a_1 Tv_1 + \cdots + a_n Tv_n = 0.$$

Thus $T = 0$, contradicting the hypothesis that $T \neq 0$.
Therefore T is not diagonalizable.
Alternative proof via the minimal polynomial: because T is nilpotent, 8.18 shows that the minimal polynomial of T equals z^m for some positive integer m. If T were diagonalizable, then by 5.62 the minimal polynomial of T would equal $(z-\lambda_1)\cdots(z-\lambda_k)$ for some list of distinct numbers $\lambda_1, \ldots, \lambda_k \in \mathbf{F}$; comparing with z^m and using the uniqueness of factorization (4.13), all the λ_i would have to equal 0, and distinctness would force $k = 1$ and hence $m = 1$. But $m = 1$ means the minimal polynomial of T is z, that is, $T = 0$, contradicting $T \neq 0$.

Problem 8A.15. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that T is diagonalizable if and only if every generalized eigenvector of T is an eigenvector of T.
For $\mathbf{F} = \mathbf{C}$, this exercise adds another equivalence to the list of conditions for diagonalizability in 5.55.

Solution. Let $n = \dim V$ and let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.
First suppose T is diagonalizable. By 5.55 we then have the eigenspace decomposition

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$

Let v be a generalized eigenvector of T. By 8.11, v corresponds to exactly one eigenvalue of T; say it corresponds to λ_j. Thus $v \neq 0$ and, as noted in the discussion following 8.8 (which follows from applying 8.1 and 8.3 to the operator $T - \lambda_j I$),

$$(T - \lambda_j I)^n v = 0.$$

Write $v = u_1 + \cdots + u_m$ with $u_k \in E(\lambda_k, T)$ for each k. Because $Tu_k = \lambda_k u_k$, we have $(T - \lambda_j I)u_k = (\lambda_k - \lambda_j)u_k$ and hence $(T - \lambda_j I)^n u_k = (\lambda_k - \lambda_j)^n u_k$. Therefore

$$0 = (T - \lambda_j I)^n v = \sum_{k=1}^{m} (\lambda_k - \lambda_j)^n u_k.$$

The k-th summand above lies in $E(\lambda_k, T)$, and the sum $E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$ is direct; hence each summand equals 0 (by 1.45, the condition characterizing direct sums). For $k \neq j$ we have $(\lambda_k - \lambda_j)^n \neq 0$, and thus $u_k = 0$. Hence $v = u_j \in E(\lambda_j, T)$. Because $v \neq 0$, this means v is an eigenvector of T corresponding to λ_j, as desired.
Now suppose every generalized eigenvector of T is an eigenvector of T. Because $\mathbf{F} = \mathbf{C}$, result 8.9 tells us that there is a basis $v_1, \ldots, v_n$ of V consisting of generalized eigenvectors of T. By hypothesis each v_k is then an eigenvector of T. Thus V has a basis consisting of eigenvectors of T, which by 5.55 implies that T is diagonalizable.

Problem 8A.16. (a) Give an example of nilpotent operators S, T on the same vector space such that neither $S + T$ nor ST is nilpotent.
(b) Suppose $S, T \in \mathcal{L}(V)$ are nilpotent and $ST = TS$. Prove that $S + T$ and ST are nilpotent.

Solution. (a) Work on $\mathbf{F}^2$ and define $S, T \in \mathcal{L}(\mathbf{F}^2)$ by

$$S(x, y) = (y, 0), \qquad T(x, y) = (0, x).$$

With respect to the standard basis, S and T have matrices

$$\mathcal{M}(S) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \mathcal{M}(T) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Both S and T are nilpotent: $S^2(x, y) = S(y, 0) = (0, 0)$ and $T^2(x, y) = T(0, x) = (0, 0)$, so $S^2 = T^2 = 0$. However $(S + T)(x, y) = (y, x)$, so

$$(S + T)^2(x, y) = (S + T)(y, x) = (x, y),$$

which shows $(S+T)^2 = I$. Hence $(S+T)^{2k} = I \neq 0$ and $(S+T)^{2k+1} = S+T \neq 0$ for every nonnegative integer k, so no power of $S + T$ equals 0; thus $S + T$ is not nilpotent.
Also $(ST)(x, y) = S(0, x) = (x, 0)$, so ST is the projection onto the first coordinate. Thus $(ST)^k = ST \neq 0$ for every positive integer k, and hence ST is not nilpotent.
(b) Suppose $S, T \in \mathcal{L}(V)$ are nilpotent and $ST = TS$. Choose positive integers m and n with $S^m = 0$ and $T^n = 0$.
Because S and T commute, an easy induction shows that S^j and T^k commute for all nonnegative integers j, k: indeed $ST = TS$ implies $ST^k = T^k S$ by induction on k, and then $S^j T^k = T^k S^j$ by induction on j. Consequently

$$(ST)^m = \underbrace{(ST)(ST)\cdots(ST)}_{m \text{ times}} = S^m T^m = 0 \cdot T^m = 0,$$

where the middle equality comes from moving all the copies of S to the left past the copies of T. Hence ST is nilpotent.
For $S + T$, note that because S and T commute, the binomial theorem is valid for them (the usual induction proof of the binomial theorem uses only the ring axioms together with commutativity of the

two elements involved, and $\mathcal{L}(V)$ is a ring under addition and composition). Thus

$$(S+T)^{m+n-1} = \sum_{j=0}^{m+n-1} \binom{m+n-1}{j} S^j T^{m+n-1-j}.$$

Consider a term of this sum, indexed by j with $0 \le j \le m+n-1$. If $j \ge m$, then $S^j = S^{j-m}S^m = 0$, so the term is 0. If $j \le m-1$, then $m+n-1-j \ge n$, so $T^{m+n-1-j} = T^{m-1-j}T^n = 0$ and again the term is 0. Hence every term of the sum equals 0, so

$$(S+T)^{m+n-1} = 0,$$

and therefore $S+T$ is nilpotent.

Problem 8A.17. Suppose $T \in \mathcal{L}(V)$ is nilpotent and m is a positive integer such that $T^m = 0$.
(a) Prove that $I-T$ is invertible and that $(I-T)^{-1} = I + T + \cdots + T^{m-1}$.
(b) Explain how you would guess the formula above.

Solution. (a) Set

$$S = I + T + T^2 + \cdots + T^{m-1} \in \mathcal{L}(V).$$

Then, using distributivity and the fact that $T^0 = I$,

$$\begin{aligned}(I-T)S &= (I + T + \cdots + T^{m-1}) - (T + T^2 + \cdots + T^m) \\ &= I - T^m \\ &= I,\end{aligned}$$

where the middle equality holds because the sum telescopes (every power T^k with $1 \le k \le m-1$ appears once with a plus sign and once with a minus sign), and the last equality holds because $T^m = 0$.
Because S is a polynomial in T, the operators S and $I-T$ commute, so the same computation gives

$$S(I-T) = (I + T + \cdots + T^{m-1}) - (T + T^2 + \cdots + T^m) = I - T^m = I.$$

Thus S is both a left inverse and a right inverse of $I-T$. Hence $I-T$ is invertible and

$$(I-T)^{-1} = I + T + \cdots + T^{m-1},$$

as desired.
(b) The formula is the geometric series in disguise. For a real or complex number x with $|x| < 1$ we have

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots.$$

Replacing the number x by the operator T suggests that $I-T$ should have inverse $I+T+T^2+T^3+\cdots$. There is no convergence issue to worry about here, because T is nilpotent: all terms T^k with $k \ge m$ equal 0, so the infinite series collapses to the finite sum $I + T + \cdots + T^{m-1}$. Part (a) confirms that this guess is correct.
Alternatively, one can guess the formula from the finite geometric sum identity

$$(1-x)(1 + x + \cdots + x^{m-1}) = 1 - x^m,$$

whose operator analogue reads $(I-T)(I + T + \cdots + T^{m-1}) = I - T^m$; the hypothesis $T^m = 0$ then makes the right side equal I.

Problem 8A.18. Suppose $T \in \mathcal{L}(V)$ is nilpotent. Prove that $T^{1+\dim \operatorname{range} T} = 0$.
If $\dim \operatorname{range} T < \dim V - 1$*, then this exercise improves 8.16.*

Solution. Let $U = \operatorname{range} T$ and let $r = \dim U$.
If $U = \{0\}$, then $T = 0$ and hence $T^{1+r} = T^1 = 0$, as desired. So assume from now on that $U \neq \{0\}$.
The subspace U is invariant under T: if $u \in U \subseteq V$, then $Tu \in \operatorname{range} T = U$. Hence the restriction

$$T|_U \in \mathcal{L}(U)$$

is a well-defined operator on the vector space U.
Next, $T|_U$ is nilpotent. Indeed, because T is nilpotent there is a positive integer k with $T^k = 0$, and then $(T|_U)^k u = T^k u = 0$ for every $u \in U$, so $(T|_U)^k = 0$.
Now apply 8.16 to the nilpotent operator $T|_U$ on the vector space U, whose dimension is r. We conclude that

$$(T|_U)^r = 0,$$

which says precisely that

$$T^r u = 0 \quad \text{for every } u \in U = \operatorname{range} T.$$

Finally, let $v \in V$ be arbitrary. Then $Tv \in \operatorname{range} T = U$, so the displayed statement above gives

$$T^{1+r} v = T^r (Tv) = 0.$$

Because $v \in V$ was arbitrary, $T^{1+r} = 0$; that is,

$$T^{1+\dim \operatorname{range} T} = 0,$$

as desired.
Remark: since $\dim \operatorname{range} T \leq \dim V$ always, and in fact $\dim \operatorname{range} T \leq \dim V - 1$ for every nilpotent T on a nonzero space (a nilpotent operator is not injective, hence not surjective), this result is at least as strong as 8.16, and it is strictly stronger whenever $\dim \operatorname{range} T < \dim V - 1$.

Problem 8A.19. Suppose $T \in \mathcal{L}(V)$ is not nilpotent. Show that

$$V = \operatorname{null} T^{\dim V - 1} \oplus \operatorname{range} T^{\dim V - 1}.$$

For operators that are not nilpotent, this exercise improves 8.4.

Solution. Let $n = \dim V$. Note that $n \geq 1$, because the only operator on a 0-dimensional space is 0, which is nilpotent.
Step 1. $\operatorname{null} T^{n-1} = \operatorname{null} T^n$.
By 8.1 we have the increasing chain

$$\{0\} = \operatorname{null} T^0 \subseteq \operatorname{null} T^1 \subseteq \cdots \subseteq \operatorname{null} T^{n-1} \subseteq \operatorname{null} T^n \subseteq V.$$

Suppose, for the sake of contradiction, that all n of these inclusions (from $\operatorname{null} T^0$ up to $\operatorname{null} T^n$) are strict. At each strict inclusion the dimension increases by at least 1, so

$$\dim \operatorname{null} T^n \geq \dim \operatorname{null} T^0 + n = n.$$

Because $\operatorname{null} T^n$ is a subspace of V and $\dim V = n$, this forces $\operatorname{null} T^n = V$ (by 2.39), that is, $T^n = 0$. But then T would be nilpotent, contradicting our hypothesis.
Hence at least one of those inclusions is not strict: there is a nonnegative integer k with $k \leq n-1$ such that

$$\operatorname{null} T^k = \operatorname{null} T^{k+1}.$$

By 8.2, all later terms of the chain are then equal to $\operatorname{null} T^k$; in particular, since $k \leq n - 1 \leq n$,

$$\operatorname{null} T^{n-1} = \operatorname{null} T^k = \operatorname{null} T^n.$$

Step 2. $\operatorname{range} T^{n-1} = \operatorname{range} T^n$.
By the fundamental theorem of linear maps (3.21) applied to T^{n-1} and to T^n, together with Step 1,

$$\dim \operatorname{range} T^{n-1} = n - \dim \operatorname{null} T^{n-1} = n - \dim \operatorname{null} T^n = \dim \operatorname{range} T^n.$$

Also $\operatorname{range} T^n \subseteq \operatorname{range} T^{n-1}$, because $T^n v = T^{n-1}(Tv)$ for every $v \in V$. A subspace of the same finite dimension as the space containing it equals that space (2.39), so

$$\operatorname{range} T^{n-1} = \operatorname{range} T^n.$$

Step 3. conclusion.
By 8.4 we have $V = \operatorname{null} T^n \oplus \operatorname{range} T^n$. Substituting the equalities from Steps 1 and 2 gives

$$V = \operatorname{null} T^{n-1} \oplus \operatorname{range} T^{n-1},$$

where the sum is direct because it is the same pair of subspaces as in 8.4. This is exactly the desired conclusion, with $n = \dim V$.

Problem 8A.20. Suppose V is an inner product space and $T \in \mathcal{L}(V)$ is normal and nilpotent. Prove that $T = 0$.

Solution. The argument below works for $\mathbf{F} = \mathbf{R}$ and for $\mathbf{F} = \mathbf{C}$.
Key claim: if $v \in V$ and $T^2 v = 0$, then $Tv = 0$.
To prove the claim, apply 7.20 (which states that T is normal if and only if $\|Tu\| = \|T^* u\|$ for every $u \in V$) to the vector $u = Tv$:

$$\|T^*(Tv)\| = \|T(Tv)\| = \|T^2 v\| = 0.$$

Hence $T^*Tv = 0$. Now use the defining property of the adjoint, namely $\langle Tx, y\rangle = \langle x, T^* y\rangle$ for all $x, y \in V$, with $x = v$ and $y = Tv$:

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle v, T^*(Tv)\rangle = \langle v, 0\rangle = 0.$$

Thus $Tv = 0$, proving the claim.
Next we show that

$$\operatorname{null} T^{k+1} = \operatorname{null} T^k \quad \text{for every positive integer } k.$$

The inclusion $\operatorname{null} T^k \subseteq \operatorname{null} T^{k+1}$ holds by 8.1. For the reverse inclusion, suppose $T^{k+1} v = 0$ and set $w = T^{k-1} v$ (this makes sense because $k \geq 1$, with $T^0 = I$). Then

$$T^2 w = T^{k+1} v = 0,$$

so the claim gives $Tw = 0$, that is, $T^k v = 0$. Hence $\operatorname{null} T^{k+1} \subseteq \operatorname{null} T^k$, and the two null spaces are equal.
Applying the displayed equality repeatedly gives

$$\operatorname{null} T^m = \operatorname{null} T \quad \text{for every positive integer } m.$$

Finally, because T is nilpotent there is a positive integer m with $T^m = 0$, and hence $\operatorname{null} T^m = V$. Therefore

$$\operatorname{null} T = \operatorname{null} T^m = V,$$

which means $Tv = 0$ for every $v \in V$. In other words, $T = 0$, as desired.
Remark: when $\mathbf{F} = \mathbf{C}$ there is a shorter route. The complex spectral theorem (7.31) gives an orthonormal basis of V consisting of eigenvectors of the normal operator T, so T is diagonalizable; but by 8.17(a) the only eigenvalue of a nilpotent operator is 0, so the diagonal matrix of T with respect to that basis is the zero matrix, forcing $T = 0$. The argument given above has the advantage of covering the real case as well.

Problem 8A.21. Suppose $T \in \mathcal{L}(V)$ is such that $\operatorname{null} T^{\dim V - 1} \neq \operatorname{null} T^{\dim V}$. Prove that T is nilpotent and that $\dim \operatorname{null} T^k = k$ for every integer k with $0 \leq k \leq \dim V$.

Solution. Let $n = \dim V$. The hypothesis presupposes $n \geq 1$.
By 8.1 we have the chain of subspaces

$$\{0\} = \operatorname{null} T^0 \subseteq \operatorname{null} T^1 \subseteq \cdots \subseteq \operatorname{null} T^{n-1} \subseteq \operatorname{null} T^n \subseteq V.$$

First we show that every inclusion in this chain is strict. Suppose not: suppose $\operatorname{null} T^k = \operatorname{null} T^{k+1}$ for some integer k with $0 \leq k \leq n-1$. If $k = n-1$ this contradicts the hypothesis outright. If $k \leq n-2$, then 8.2 tells us that

$$\operatorname{null} T^k = \operatorname{null} T^{k+1} = \operatorname{null} T^{k+2} = \cdots,$$

so in particular $\operatorname{null} T^{n-1} = \operatorname{null} T^n$ (both equal $\operatorname{null} T^k$, since $k \leq n-1 \leq n$), again contradicting the hypothesis. Hence

$$\{0\} = \operatorname{null} T^0 \subsetneq \operatorname{null} T^1 \subsetneq \cdots \subsetneq \operatorname{null} T^{n-1} \subsetneq \operatorname{null} T^n,$$

where $\subsetneq$ means contained in but not equal to.
At each strict inclusion the dimension increases by at least 1, so

$$\dim \operatorname{null} T^k \geq k \quad \text{for } k = 0, 1, \ldots, n,$$

by induction on k: the case $k = 0$ is clear, and if $\dim \operatorname{null} T^k \geq k$ with $k \leq n-1$, then $\dim \operatorname{null} T^{k+1} \geq \dim \operatorname{null} T^k + 1 \geq k + 1$.
On the other hand, each $\operatorname{null} T^k$ is a subspace of V, so $\dim \operatorname{null} T^k \leq n$ (by 2.37). Taking $k = n$ gives

$$n \leq \dim \operatorname{null} T^n \leq n,$$

so $\dim \operatorname{null} T^n = n$ and hence $\operatorname{null} T^n = V$ (by 2.39). Thus $T^n = 0$, which shows that T is nilpotent.
It remains to prove that $\dim \operatorname{null} T^k = k$ for each k with $0 \leq k \leq n$. Fix such a k. We already know $\dim \operatorname{null} T^k \geq k$. For the reverse inequality, run the dimension count downward from k to n: the strict inclusions

$$\operatorname{null} T^k \subsetneq \operatorname{null} T^{k+1} \subsetneq \cdots \subsetneq \operatorname{null} T^n$$

consist of $n - k$ strict steps, each raising the dimension by at least 1, so

$$n = \dim \operatorname{null} T^n \geq \dim \operatorname{null} T^k + (n - k),$$

which gives $\dim \operatorname{null} T^k \leq k$. Combining the two inequalities yields

$$\dim \operatorname{null} T^k = k \quad \text{for every integer } k \text{ with } 0 \leq k \leq n,$$

as desired.
Remark: the two inequalities together also show that the sequence of dimensions is forced to be exactly $0, 1, 2, \ldots, n$, so each strict inclusion in the chain raises the dimension by exactly 1. In particular $\dim \operatorname{null} T = 1$, so T has a one-dimensional null space; combined with nilpotence, this says that T has a single Jordan block of size n.

Problem 8A.22. Suppose $T \in \mathcal{L}(\mathbf{C}^5)$ is such that $\operatorname{range} T^4 \neq \operatorname{range} T^5$. Prove that T is nilpotent.

Solution. Write $V = \mathbf{C}^5$, so $\dim V = 5$. (The argument below uses nothing about the field, so the same proof works on $\mathbf{R}^5$.)
Step 1: the ranges form a decreasing chain. For every nonnegative integer k we have

$$\operatorname{range} T^{k+1} \subseteq \operatorname{range} T^k,$$

because if $u = T^{k+1}v$ for some $v \in V$, then $u = T^k(Tv) \in \operatorname{range} T^k$. (This is Exercise 6 in Section 8A.)

Step 2: once the chain stabilizes, it stabilizes forever. Suppose m is a nonnegative integer with $\operatorname{range} T^m = \operatorname{range} T^{m+1}$. Applying T to these two equal subsets produces equal images:

$$\operatorname{range} T^{m+1} = T\big(\operatorname{range} T^m\big) = T\big(\operatorname{range} T^{m+1}\big) = \operatorname{range} T^{m+2},$$

where the outer two equalities hold because $T(\operatorname{range} T^j) = \{T(T^j v) : v \in V\} = \operatorname{range} T^{j+1}$ for every nonnegative integer j. Iterating this step gives $\operatorname{range} T^k = \operatorname{range} T^m$ for all integers $k \geq m$. (This is Exercise 7 in Section 8A.)
Step 3: the first six ranges strictly decrease. We are given $\operatorname{range} T^4 \neq \operatorname{range} T^5$. If we had $\operatorname{range} T^m = \operatorname{range} T^{m+1}$ for some $m \in \{0, 1, 2, 3\}$, then Step 2 would give $\operatorname{range} T^4 = \operatorname{range} T^m = \operatorname{range} T^5$, contradicting the hypothesis. Hence all the inclusions from Step 1 with $k \leq 4$ are strict:

$$V = \operatorname{range} T^0 \supsetneq \operatorname{range} T^1 \supsetneq \operatorname{range} T^2 \supsetneq \operatorname{range} T^3 \supsetneq \operatorname{range} T^4 \supsetneq \operatorname{range} T^5.$$

Step 4: dimension count. If U and W are subspaces of V with $U \subsetneq W$, then $\dim U \leq \dim W$ by 2.37, and $\dim U \neq \dim W$ by 2.39 applied inside W (a subspace of W whose dimension equals $\dim W$ equals W). Hence each of the five strict inclusions above drops the dimension by at least 1, so

$$\dim \operatorname{range} T^5 \leq \dim \operatorname{range} T^0 - 5 = 5 - 5 = 0.$$

Thus $\operatorname{range} T^5 = \{0\}$, which says exactly that $T^5 = 0$. Therefore T is nilpotent.
Alternatively one can quote earlier exercises: by Exercise 9 in Section 8A, the hypothesis $\operatorname{range} T^4 \neq \operatorname{range} T^5$ is equivalent to $\operatorname{null} T^4 \neq \operatorname{null} T^5$, that is, to $\operatorname{null} T^{\dim V - 1} \neq \operatorname{null} T^{\dim V}$, and then Exercise 21 in Section 8A gives that T is nilpotent.

Problem 8A.23. Give an example of an operator T on a finite-dimensional real vector space such that 0 is the only eigenvalue of T but T is not nilpotent.
[This exercise shows that (b) in 8.17 does not hold without the hypothesis that $\mathbf{F} = \mathbf{C}$.]

Solution. Take $V = \mathbf{R}^3$ and define $T \in \mathcal{L}(\mathbf{R}^3)$ by

$$T(x, y, z) = (-y,\, x,\, 0).$$

In other words, T rotates the xy-plane by 90° and kills the z-axis; its matrix with respect to the standard basis is

$$\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of T: suppose $\lambda \in \mathbf{R}$ and $T(x, y, z) = \lambda(x, y, z)$ with $(x, y, z) \in \mathbf{R}^3$. This means

$$-y = \lambda x, \qquad x = \lambda y, \qquad 0 = \lambda z.$$

Substituting the second equation into the first gives $-y = \lambda^2 y$, so $(\lambda^2 + 1)y = 0$. Since λ is real, $\lambda^2 + 1 > 0$, and therefore $y = 0$; then $x = \lambda y = 0$. Thus every eigenvector of T has the form $(0, 0, z)$. If $\lambda \neq 0$, the third equation forces $z = 0$ too, so the only solution is the zero vector; hence no nonzero real number is an eigenvalue of T. On the other hand, $T(0, 0, 1) = (0, 0, 0)$, so $(0, 0, 1)$ is an eigenvector of T corresponding to the eigenvalue 0. Therefore 0 is the only eigenvalue of T.
T is not nilpotent: computing directly,

$$T(1, 0, 0) = (0, 1, 0), \quad T^2(1, 0, 0) = (-1, 0, 0), \quad T^3(1, 0, 0) = (0, -1, 0), \quad T^4(1, 0, 0) = (1, 0, 0),$$

so $T^k(1, 0, 0)$ cycles through the four nonzero vectors $(1, 0, 0), (0, 1, 0), (-1, 0, 0), (0, -1, 0)$ and in particular is never 0. Hence $T^k \neq 0$ for every positive integer k, so T is not nilpotent. (Equivalently: the minimal polynomial of T is $z^3 + z = z(z^2 + 1)$, which is not of the form z^m, so T is not nilpotent by 8.18.)

This is consistent with 8.17: part (a) of that result says a nilpotent operator has 0 as its only eigenvalue, and our T indeed has 0 as its only eigenvalue while failing to be nilpotent, which shows the converse (b) genuinely needs the hypothesis $\mathbf{F} = \mathbf{C}$. The obstruction is that over $\mathbf{R}$ the minimal polynomial can have irreducible quadratic factors with no real zeros, so "the only zero of the minimal polynomial is 0" does not force the minimal polynomial to be z^m.
Remark: the dimension 3 used here is minimal. Suppose V is a real vector space with $\dim V \le 2$ and 0 is the only eigenvalue of $T \in \mathcal{L}(V)$. The minimal polynomial p of T has degree at most $\dim V \le 2$ (by 5.22), and by 5.27(a) the zeros of p in $\mathbf{R}$ are exactly the eigenvalues of T, namely just 0. A monic real polynomial of degree 1 or 2 having 0 as a zero is z, or z^2, or $z(z-a)$ with $a \neq 0$; the last is excluded because a would then be an eigenvalue. Hence $p(z) = z$ or $p(z) = z^2$, so $T^2 = 0$ and T is nilpotent.

Problem 8A.24. For each item in Example 8.15, find a basis of the domain vector space such that the matrix of the nilpotent operator with respect to that basis has the upper-triangular form promised by 8.18(c).

Solution. Recall that 8.18(c) asks for a basis $v_1, \dots, v_n$ of the domain with respect to which the matrix of the operator has all entries on and below the diagonal equal to 0. Since the k-th column of the matrix of T lists the coefficients of Tv_k in the basis $v_1, \dots, v_n$ (see the definition 5.35), this requirement says exactly that

$$Tv_k \in \operatorname{span}(v_1, \dots, v_{k-1}) \quad \text{for each } k = 1, \dots, n,$$

with the convention that $\operatorname{span}(\,) = \{0\}$, so in particular $Tv_1 = 0$.
Part (a). Here $T \in \mathcal{L}(\mathbf{F}^4)$ is defined by $T(z_1, z_2, z_3, z_4) = (0, 0, z_1, z_2)$. Writing e_1, e_2, e_3, e_4 for the standard basis, we have

$$Te_1 = e_3, \qquad Te_2 = e_4, \qquad Te_3 = 0, \qquad Te_4 = 0.$$

The standard basis in its usual order does not work, since $Te_1 = e_3 \notin \operatorname{span}(e_1)$. But simply reordering it does work: take

$$v_1 = e_3, \qquad v_2 = e_1, \qquad v_3 = e_4, \qquad v_4 = e_2.$$

This is a basis of $\mathbf{F}^4$ (it is a permutation of the standard basis), and

$$Tv_1 = 0, \qquad Tv_2 = e_3 = v_1, \qquad Tv_3 = 0, \qquad Tv_4 = e_4 = v_3.$$

Hence the matrix of T with respect to v_1, v_2, v_3, v_4 is

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

whose entries on and below the diagonal are all 0, as required. (This matrix consists of two 2-by-2 Jordan-type blocks, reflecting $T^2 = 0$ with $\dim \operatorname{null} T = 2$.)
Part (b). Here $T \in \mathcal{L}(\mathbf{F}^3)$ has matrix

$$A = \begin{pmatrix} -3 & 9 & 0 \\ -7 & 9 & 6 \\ 4 & 0 & -6 \end{pmatrix}$$

with respect to the standard basis, and $T^3 = 0$. Computing,

$$A^2 = \begin{pmatrix} -54 & 54 & 54 \\ -18 & 18 & 18 \\ -36 & 36 & 36 \end{pmatrix}, \qquad A^3 = 0.$$

Since $A^2 \neq 0$, there is a vector v with $T^2 v \neq 0$; the first column of A^2 shows that $v = (1,0,0)$ works, because $T^2 v = (-54,-18,-36) \neq 0$. Now put

$$v_1 = T^2 v = (-54,-18,-36), \qquad v_2 = Tv = (-3,-7,4), \qquad v_3 = v = (1,0,0).$$

Since $T^2 v \neq 0$ and $T^3 v = 0$, Exercise 2 in Section 8A (with $m = 3$) shows that $v, Tv, T^2 v$ is linearly independent; being a linearly independent list of length $3 = \dim \mathbf{F}^3$, it is a basis of $\mathbf{F}^3$ (by 2.38). Moreover

$$Tv_1 = T^3 v = 0, \qquad Tv_2 = T^2 v = v_1, \qquad Tv_3 = Tv = v_2.$$

Hence the matrix of T with respect to v_1, v_2, v_3 is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

which has the required form. (One may of course rescale the whole chain: replacing v by cv for any $c \neq 0$ replaces v_1, v_2, v_3 by cv_1, cv_2, cv_3 and gives the same matrix.)
Part (c). Here $V = \mathcal{P}_m(\mathbf{R})$ and $T = D$ is differentiation, $Dp = p'$. The standard basis, in its usual order, already works: take

$$v_1 = 1, \quad v_2 = x, \quad v_3 = x^2, \quad \ldots, \quad v_{m+1} = x^m.$$

Then $Dv_1 = 0$ and, for $k = 1, \ldots, m$,

$$Dv_{k+1} = D(x^k) = kx^{k-1} = kv_k \in \operatorname{span}(v_1, \ldots, v_k),$$

so the criterion displayed at the start of this solution is met (for the basis vector with index $k+1$ the image lies in the span of the strictly earlier basis vectors, and for index 1 the image is 0). Concretely, the matrix of D with respect to $v_1, \ldots, v_{m+1}$ is the $(m+1)$-by-$(m+1)$ matrix whose entry in row k, column $k+1$ equals k for $k = 1, \ldots, m$ and whose every other entry is 0:

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 2 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & & m-1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & m \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

All entries on and below the diagonal are 0, as promised by 8.18(c). Note that the superdiagonal entries $1, 2, \ldots, m$ are all nonzero, matching the fact that $D^m \neq 0$ on $\mathcal{P}_m(\mathbf{R})$ while $D^{m+1} = 0$. (The display above is written for $m \geq 1$; in the degenerate case $m = 0$ the space $\mathcal{P}_0(\mathbf{R})$ has dimension 1, the basis is the single vector $v_1 = 1$, and the matrix of D is the 1-by-1 zero matrix, which again has all entries on and below the diagonal equal to 0.)

Problem 8A.25. Suppose that V is an inner product space and $T \in \mathcal{L}(V)$ is nilpotent. Show that there is an orthonormal basis of V with respect to which the matrix of T has the upper-triangular form promised by 8.18(c).

Solution. As always in this chapter, V is finite-dimensional (and, being an inner product space, $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$). Let $n = \dim V$; we may assume $n \geq 1$, since for $V = \{0\}$ there is nothing to prove.
First record what the desired conclusion says. If $u_1, \ldots, u_n$ is a basis of V, then by the definition 5.35 of the matrix of an operator, the k-th column of $\mathcal{M}(T)$ with respect to $u_1, \ldots, u_n$ consists of the

coefficients in the expansion of Tu_k in that basis. Hence

$$\begin{aligned}&\mathcal{M}(T) \text{ has all entries on and below the diagonal equal to } 0\\ &\quad\iff Tu_k \in \operatorname{span}(u_1,\dots,u_{k-1}) \text{ for each } k = 1,\dots,n,\end{aligned}$$

where $\operatorname{span}(u_1,\dots,u_{k-1})$ is interpreted as $\{0\}$ when $k = 1$.
Because T is nilpotent, 8.18 (the implication (a) $\Rightarrow$ (c)) provides a basis $v_1,\dots,v_n$ of V with respect to which the matrix of T has all entries on and below the diagonal equal to 0; by the paragraph above this means

$$(\star)\qquad Tv_j \in \operatorname{span}(v_1,\dots,v_{j-1}) \quad \text{for each } j = 1,\dots,n.$$

This basis need not be orthonormal, so we improve it with the Gram–Schmidt procedure.
Since $v_1,\dots,v_n$ is linearly independent, applying the Gram–Schmidt procedure (6.32) to it produces an orthonormal list $e_1,\dots,e_n$ in V with

$$\operatorname{span}(e_1,\dots,e_k) = \operatorname{span}(v_1,\dots,v_k) \quad \text{for each } k = 1,\dots,n.$$

This is an orthonormal list of length $n = \dim V$, so it is an orthonormal basis of V (by 6.28).
Now fix $k \in \{1,\dots,n\}$. Because $e_k \in \operatorname{span}(e_1,\dots,e_k) = \operatorname{span}(v_1,\dots,v_k)$, there exist $a_1,\dots,a_k \in \mathbf{F}$ with

$$e_k = a_1v_1 + \cdots + a_kv_k.$$

Applying T and using $(\star)$ together with the inclusions $\operatorname{span}(v_1,\dots,v_{j-1}) \subseteq \operatorname{span}(v_1,\dots,v_{k-1})$ for $j \le k$, we get

$$Te_k = a_1Tv_1 + \cdots + a_kTv_k \in \operatorname{span}(v_1,\dots,v_{k-1}) = \operatorname{span}(e_1,\dots,e_{k-1}),$$

where for $k = 1$ both of the last two spans are $\{0\}$, so the statement reads $Te_1 = 0$.
Thus $Te_k \in \operatorname{span}(e_1,\dots,e_{k-1})$ for every $k = 1,\dots,n$. By the criterion recorded at the start, the matrix of T with respect to the orthonormal basis $e_1,\dots,e_n$ has all entries on and below the diagonal equal to 0, that is, it has the form

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix}$$

promised by 8.18(c). This completes the proof.
Remark. There is a shorter route through 6.37, and it works over $\mathbf{R}$ just as well as over $\mathbf{C}$. Because T is nilpotent, 8.18 (the implication (a) $\Rightarrow$ (b)) says the minimal polynomial of T equals z^m for some positive integer m, which is of the form $(z-\lambda_1)\cdots(z-\lambda_m)$ with every $\lambda_j = 0$. Hence 6.37 gives an orthonormal basis $e_1,\dots,e_n$ of V with respect to which the matrix of T is upper triangular. By 5.41 the eigenvalues of T are precisely the diagonal entries of that matrix, and by 8.17(a) the only eigenvalue of T is 0; so every diagonal entry equals 0, and the matrix has all entries on and below the diagonal equal to 0. Note that 6.37 is stated for an arbitrary $\mathbf{F}$, with the splitting of the minimal polynomial as its hypothesis, so no appeal to $\mathbf{F} = \mathbf{C}$ is needed here; the proof of 6.37 is itself the Gram–Schmidt argument given above, applied to a basis supplied by 5.44 rather than by 8.18(c).

8.2 Exercises 8B

Problem 8B.1. Define $T \in \mathcal{L}(\mathbf{C}^2)$ by $T(w,z) = (-z,w)$. Find the generalized eigenspaces corresponding to the distinct eigenvalues of T.

Solution. The answer is

$$G(i,T) = \operatorname{span}\big((1,-i)\big), \qquad G(-i,T) = \operatorname{span}\big((1,i)\big),$$

and $\mathbf{C}^2 = G(i,T) \oplus G(-i,T)$, as promised by 8.22.

First we find the eigenvalues. If $T(w,z) = \lambda(w,z)$, then $-z = \lambda w$ and $w = \lambda z$. Substituting the second equation into the first gives $-z = \lambda^2 z$; if $z = 0$ then also $w = \lambda z = 0$, so for an eigenvector we must have $z \neq 0$ and hence $\lambda^2 = -1$. Thus the only possible eigenvalues are i and $-i$. Both occur: $T(1,-i) = (i,1) = i(1,-i)$ and $T(1,i) = (-i,1) = -i(1,i)$. Hence the distinct eigenvalues of T are i and $-i$, with

$$E(i,T) = \operatorname{span}\big((1,-i)\big), \qquad E(-i,T) = \operatorname{span}\big((1,i)\big),$$

because in each case the eigenvector equation $w = \lambda z$ determines the eigenvector up to a scalar.
Now we compute the generalized eigenspaces. Since $\dim \mathbf{C}^2 = 2$, we have $G(\lambda,T) = \operatorname{null}(T-\lambda I)^2$ by 8.20. Note that

$$T^2(w,z) = T(-z,w) = (-w,-z),$$

so $T^2 = -I$. Therefore

$$(T - iI)^2 = T^2 - 2iT - I = -I - 2iT - I = -2(I + iT).$$

Because $-2 \neq 0$, we get $G(i,T) = \operatorname{null}(I + iT)$. Now

$$(I + iT)(w,z) = (w,z) + i(-z,w) = (w - iz,\ z + iw),$$

which equals $(0,0)$ exactly when $w = iz$ (the second coordinate gives $z = -iw$, which is the same condition). Hence

$$G(i,T) = \{(iz,z) : z \in \mathbf{C}\} = \operatorname{span}\big((i,1)\big) = \operatorname{span}\big((1,-i)\big).$$

Similarly,

$$(T + iI)^2 = T^2 + 2iT - I = -2(I - iT),$$

and $(I - iT)(w,z) = (w + iz,\ z - iw)$, which equals $(0,0)$ exactly when $w = -iz$. Hence

$$G(-i,T) = \{(-iz,z) : z \in \mathbf{C}\} = \operatorname{span}\big((-i,1)\big) = \operatorname{span}\big((1,i)\big).$$

Thus each generalized eigenspace equals the corresponding eigenspace, and each eigenvalue has multiplicity 1; the multiplicities sum to $2 = \dim \mathbf{C}^2$, in agreement with 8.25. Equivalently: the minimal polynomial of T is $z^2 + 1 = (z - i)(z + i)$, which has no repeated zeros, so T is diagonalizable and generalized eigenspaces coincide with eigenspaces.

Problem 8B.2. Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that $G(\lambda,T) = G\big(\frac{1}{\lambda}, T^{-1}\big)$ for every $\lambda \in \mathbf{F}$ with $\lambda \neq 0$.

Solution. Fix $\lambda \in \mathbf{F}$ with $\lambda \neq 0$. The key is the algebraic identity

$$T^{-1} - \tfrac{1}{\lambda} I = \tfrac{1}{\lambda} T^{-1}(\lambda I - T) = -\tfrac{1}{\lambda} T^{-1}(T - \lambda I),$$

which is verified by expanding the middle expression: $\frac{1}{\lambda}T^{-1}(\lambda I - T) = T^{-1} - \frac{1}{\lambda} I$.
Because T^{-1} commutes with T, it also commutes with $T - \lambda I$. Hence for every positive integer k the factors on the right side above commute, and so

$$\Big(T^{-1} - \tfrac{1}{\lambda} I\Big)^k = \Big(-\tfrac{1}{\lambda}\Big)^k (T^{-1})^k (T - \lambda I)^k.$$

Now fix a positive integer k and let $v \in V$. Since $(T^{-1})^k$ is invertible (it is the inverse of T^k) and $\big(-\frac{1}{\lambda}\big)^k \neq 0$, the operator $\big(-\frac{1}{\lambda}\big)^k (T^{-1})^k$ is injective. Therefore

$$\Big(T^{-1} - \tfrac{1}{\lambda} I\Big)^k v = 0 \iff (T - \lambda I)^k v = 0.$$

In other words,

$$\operatorname{null}\Big(T^{-1} - \tfrac{1}{\lambda} I\Big)^k = \operatorname{null}(T - \lambda I)^k$$

for every positive integer k.
By the definition 8.19 of a generalized eigenspace, $G(\lambda, T)$ is the union over all positive integers k of $\operatorname{null}(T-\lambda I)^k$, and $G\left(\frac{1}{\lambda}, T^{-1}\right)$ is the union over all positive integers k of $\operatorname{null}\left(T^{-1}-\frac{1}{\lambda}I\right)^k$. These unions are taken over equal sets, so

$$G(\lambda, T) = G\left(\tfrac{1}{\lambda}, T^{-1}\right),$$

as desired. (Alternatively, take $k = \dim V$ in the displayed equality of null spaces and use 8.20 twice.)
Two remarks. First, no assumption that λ is an eigenvalue is needed: if λ is not an eigenvalue of T, then both sides equal $\{0\}$, and the equality of the null spaces above shows that λ is an eigenvalue of T if and only if $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}. Second, taking dimensions in the case $k = \dim V$ shows that corresponding eigenvalues have equal multiplicities.

Problem 8B.3. Suppose $T \in \mathcal{L}(V)$. Suppose $S \in \mathcal{L}(V)$ is invertible. Prove that T and $S^{-1}TS$ have the same eigenvalues with the same multiplicities.

Solution. Let $n = \dim V$ and fix $\lambda \in \mathbf{F}$. Because $S^{-1}(\lambda I)S = \lambda I$, we have

$$S^{-1}TS - \lambda I = S^{-1}(T - \lambda I)S.$$

Hence for every positive integer k, the inner factors of S and S^{-1} cancel in the product and we get

$$(S^{-1}TS - \lambda I)^k = \left(S^{-1}(T - \lambda I)S\right)^k = S^{-1}(T - \lambda I)^k S.$$

Now fix a positive integer k and consider the restriction of S to $\operatorname{null}(S^{-1}TS - \lambda I)^k$. If $v \in \operatorname{null}(S^{-1}TS - \lambda I)^k$, then

$$0 = (S^{-1}TS - \lambda I)^k v = S^{-1}(T - \lambda I)^k (Sv),$$

and applying the injective map S to both sides gives $(T - \lambda I)^k (Sv) = 0$; thus $Sv \in \operatorname{null}(T - \lambda I)^k$. Conversely, if $u \in \operatorname{null}(T - \lambda I)^k$ and $v = S^{-1}u$, then

$$(S^{-1}TS - \lambda I)^k v = S^{-1}(T - \lambda I)^k S S^{-1} u = S^{-1}(T - \lambda I)^k u = 0,$$

so $v \in \operatorname{null}(S^{-1}TS - \lambda I)^k$ and $Sv = u$. Therefore S restricted to $\operatorname{null}(S^{-1}TS - \lambda I)^k$ is a linear map onto $\operatorname{null}(T - \lambda I)^k$, and it is injective because S is injective. Hence these two subspaces are isomorphic, and in particular

$$\dim \operatorname{null}(S^{-1}TS - \lambda I)^k = \dim \operatorname{null}(T - \lambda I)^k$$

for every positive integer k and every $\lambda \in \mathbf{F}$.
Taking $k = 1$: $\operatorname{null}(S^{-1}TS - \lambda I) \neq \{0\}$ if and only if $\operatorname{null}(T - \lambda I) \neq \{0\}$. Thus λ is an eigenvalue of $S^{-1}TS$ if and only if λ is an eigenvalue of T. So T and $S^{-1}TS$ have the same eigenvalues.
Taking $k = n$: by 8.20 the displayed equation says

$$\dim G(\lambda, S^{-1}TS) = \dim G(\lambda, T)$$

(both operators act on the same space V, so the same exponent $n = \dim V$ is used for both). By the definition 8.23 of multiplicity, this says exactly that each eigenvalue λ has the same multiplicity for $S^{-1}TS$ as for T.
Thus T and $S^{-1}TS$ have the same eigenvalues with the same multiplicities. In fact the argument above proves more: S maps $G(\lambda, S^{-1}TS)$ isomorphically onto $G(\lambda, T)$ for every $\lambda \in \mathbf{F}$.

Problem 8B.4. Suppose $\dim V \geq 2$ and $T \in \mathcal{L}(V)$ is such that $\operatorname{null} T^{\dim V - 2} \neq \operatorname{null} T^{\dim V - 1}$. Prove that T has at most two distinct eigenvalues.

Solution. Let $n = \dim V$, so $n \geq 2$ and $\operatorname{null} T^{n-2} \neq \operatorname{null} T^{n-1}$.

Step 1: the null spaces of $T^0, T^1, \ldots, T^{n-1}$ strictly increase. By 8.1 we have the chain

$$\{0\} = \operatorname{null} T^0 \subseteq \operatorname{null} T^1 \subseteq \cdots \subseteq \operatorname{null} T^{n-1}.$$

The inclusions in this chain are $\operatorname{null} T^j \subseteq \operatorname{null} T^{j+1}$ for $j = 0, 1, \ldots, n-2$. The last of them, the one with $j = n-2$, is strict by hypothesis. Suppose one of the others were an equality, say $\operatorname{null} T^j = \operatorname{null} T^{j+1}$ with $0 \leq j \leq n-3$. Then 8.2 would give $\operatorname{null} T^j = \operatorname{null} T^k$ for all $k \geq j$; since $j \leq n-3$, both $n-2$ and $n-1$ are at least j, so this yields $\operatorname{null} T^{n-2} = \operatorname{null} T^j = \operatorname{null} T^{n-1}$, contradicting the hypothesis. Hence every inclusion in the chain above is strict:

$$\{0\} = \operatorname{null} T^0 \subsetneq \operatorname{null} T^1 \subsetneq \cdots \subsetneq \operatorname{null} T^{n-1}.$$

At each of the $n-1$ strict inclusions the dimension increases by at least 1, so

$$\dim \operatorname{null} T^{n-1} \geq n-1.$$

Step 2: 0 is an eigenvalue of T with multiplicity at least $n-1$. Since $n \geq 2$, the chain in Step 1 includes the strict inclusion $\operatorname{null} T^0 \subsetneq \operatorname{null} T^1$, so $\operatorname{null} T \neq \{0\}$ and therefore 0 is an eigenvalue of T. By 8.1, $\operatorname{null} T^{n-1} \subseteq \operatorname{null} T^n$, so by 8.20 (applied with $\lambda = 0$),

$$\dim G(0, T) = \dim \operatorname{null} T^n \geq \dim \operatorname{null} T^{n-1} \geq n-1.$$

Thus the multiplicity of the eigenvalue 0 is at least $n-1$.
Step 3: the generalized eigenspaces sum directly. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, and suppose $v_1 + \cdots + v_m = 0$ with $v_k \in G(\lambda_k, T)$ for each k. Discard the terms v_k that equal 0; the remaining v_k form a list of generalized eigenvectors of T corresponding to distinct eigenvalues, hence a linearly independent list by 8.12. Since that list sums to 0 with all coefficients equal to 1, the list must be empty; that is, every v_k equals 0. By 1.45 the sum $G(\lambda_1, T) + \cdots + G(\lambda_m, T)$ is a direct sum. Therefore, by 3.94,

$$\sum_{k=1}^{m} \dim G(\lambda_k, T) = \dim \big(G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)\big) \leq \dim V = n.$$

(Note that this step does not require $\mathbf{F} = \mathbf{C}$; only 8.12 is used, not the generalized eigenspace decomposition 8.22.)
Step 4: conclusion. By Step 2, one of the eigenvalues, say $\lambda_1 = 0$, satisfies $\dim G(\lambda_1, T) \geq n-1$. Each remaining generalized eigenspace has dimension at least 1, since $G(\lambda_k, T) \supseteq E(\lambda_k, T) \neq \{0\}$ for an eigenvalue λ_k. Hence Step 3 gives

$$n \geq (n-1) + (m-1),$$

which simplifies to $m \leq 2$. Thus T has at most two distinct eigenvalues.

Problem 8B.5. Suppose $T \in \mathcal{L}(V)$ and 3 and 8 are eigenvalues of T. Let $n = \dim V$. Prove that $V = (\operatorname{null} T^{n-2}) \oplus (\operatorname{range} T^{n-2})$.

Solution. Because T has the two distinct eigenvalues 3 and 8, eigenvectors corresponding to them are linearly independent (by 5.11), so $n \geq 2$ and the exponent $n-2$ is a nonnegative integer.
Step 1: $\dim \operatorname{null} T^n \leq n-2$. By 8.20, $\operatorname{null} T^n = G(0, T)$. As in Exercise 4 above, the generalized eigenspaces of T corresponding to distinct eigenvalues form a direct sum: if $v_1 + \cdots + v_m = 0$ with each v_k in the generalized eigenspace of a distinct eigenvalue, then after discarding the zero terms the rest would be a linearly independent list (by 8.12) summing to 0 with coefficients 1, which forces the list to be empty; hence all $v_k = 0$, and the sum is direct by 1.45.
If 0 is an eigenvalue of T, then 0, 3, 8 are three distinct eigenvalues, and the direct sum $G(0, T) \oplus G(3, T) \oplus G(8, T)$ is a subspace of V, so by 3.94

$$\dim G(0, T) + \dim G(3, T) + \dim G(8, T) \leq n.$$

Since $G(3,T) \supseteq E(3,T) \neq \{0\}$ and $G(8,T) \supseteq E(8,T) \neq \{0\}$, each of the last two dimensions is at least 1, so $\dim G(0,T) \leq n-2$.
If 0 is not an eigenvalue of T, then T is injective, so T^n is injective and $\dim G(0,T) = \dim \operatorname{null} T^n = 0 \leq n-2$ (using $n \geq 2$).
In both cases $\dim \operatorname{null} T^n \leq n-2$.
Step 2: $\operatorname{null} T^{n-2} = \operatorname{null} T^n$. Let m be the smallest nonnegative integer such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$; such an m exists and satisfies $m \leq n$ by 8.3. By minimality of m, all the inclusions

$$\{0\} = \operatorname{null} T^0 \subsetneq \operatorname{null} T^1 \subsetneq \cdots \subsetneq \operatorname{null} T^m$$

are strict (an equality earlier in the chain would, by 8.2, force $\operatorname{null} T^j = \operatorname{null} T^{j+1}$ for a smaller j, contradicting minimality). Each strict inclusion raises the dimension by at least 1, so

$$m \leq \dim \operatorname{null} T^m.$$

By 8.2, $\operatorname{null} T^m = \operatorname{null} T^k$ for all $k \geq m$; in particular $\operatorname{null} T^m = \operatorname{null} T^n$. Combining this with Step 1 gives

$$m \leq \dim \operatorname{null} T^m = \dim \operatorname{null} T^n \leq n-2.$$

Hence $n-2 \geq m$, and so, again by 8.2,

$$\operatorname{null} T^{n-2} = \operatorname{null} T^m = \operatorname{null} T^n.$$

Step 3: $\operatorname{range} T^{n-2} = \operatorname{range} T^n$. Since $T^n = T^2 T^{n-2}$, we have $\operatorname{range} T^n \subseteq \operatorname{range} T^{n-2}$. By the fundamental theorem of linear maps (3.21) applied to T^{n-2} and to T^n, together with Step 2,

$$\dim \operatorname{range} T^{n-2} = n - \dim \operatorname{null} T^{n-2} = n - \dim \operatorname{null} T^n = \dim \operatorname{range} T^n.$$

A subspace of the same dimension as the space containing it equals that space (2.39), so $\operatorname{range} T^{n-2} = \operatorname{range} T^n$.
Step 4: conclusion. By 8.4,

$$V = (\operatorname{null} T^n) \oplus (\operatorname{range} T^n).$$

Substituting the equalities from Steps 2 and 3 gives

$$V = (\operatorname{null} T^{n-2}) \oplus (\operatorname{range} T^{n-2}),$$

as desired.

Problem 8B.6. Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. Explain why the exponent of $z-\lambda$ in the factorization of the minimal polynomial of T is the smallest positive integer m such that $(T-\lambda I)^m|_{G(\lambda,T)} = 0$.

Solution. Let $n = \dim V$ and write $G = G(\lambda, T)$. Let p be the minimal polynomial of T. Since λ is an eigenvalue of T, we know $p(\lambda) = 0$ by 5.27(a), so we may factor

$$p(z) = (z-\lambda)^k q(z),$$

where k is a positive integer, $q \in \mathcal{P}(\mathbf{F})$, and $q(\lambda) \neq 0$ (take k to be the largest power of $z-\lambda$ dividing p; this k is the exponent referred to in the problem).
First we check that the integer m in the statement exists. By 8.20, $G = \operatorname{null}(T-\lambda I)^n$, so $(T-\lambda I)^n v = 0$ for every $v \in G$. Also G is invariant under T by 5.18 (with the polynomial $(z-\lambda)^n$), hence invariant under $T-\lambda I$, so $N := (T-\lambda I)|_G \in \mathcal{L}(G)$ is defined and $N^n = 0$. Thus there is a smallest positive integer m with $N^m = (T-\lambda I)^m|_G = 0$, and $m \leq n$. Note also that G is invariant under every polynomial in T, and that for any polynomial r we have $r(T)|_G = r(T|_G)$.
Claim 1: $m \leq k$. Expand q in powers of $z-\lambda$:

$$q(z) = c_0 + c_1(z-\lambda) + \cdots + c_d(z-\lambda)^d,$$

where $c_0 = q(\lambda) \neq 0$. Restricting to G and using $N = (T - \lambda I)|_G$,

$$q(T)|_G = c_0 I|_G + A, \quad \text{where } A = c_1 N + \cdots + c_d N^d.$$

Writing $A = NB$ with $B = c_1 I|_G + c_2 N + \cdots + c_d N^{d-1}$, which commutes with N, gives $A^m = N^m B^m = 0$, so A is nilpotent. An operator of the form $c_0 I + A$ with $c_0 \neq 0$ and A nilpotent is invertible: one checks directly that its inverse is $c_0^{-1} \sum_{i=0}^{m-1} (-c_0^{-1} A)^i$, since the telescoping product $(c_0 I + A) \sum_{i=0}^{m-1} (-c_0^{-1} A)^i = c_0 \big(I - (-c_0^{-1} A)^m\big) = c_0 I$. Hence $q(T)|_G$ is invertible.
Now $p(T) = 0$, so restricting to the invariant subspace G and using that $(T - \lambda I)^k$ and $q(T)$ commute,

$$0 = p(T)|_G = \big((T - \lambda I)^k|_G\big)\big(q(T)|_G\big).$$

Composing on the right with the inverse of $q(T)|_G$ gives $(T - \lambda I)^k|_G = 0$. By the minimality of m, we conclude $m \leq k$.
Claim 2: $k \leq m$. Let $R = \operatorname{range}(T - \lambda I)^n$. Applying 8.4 to the operator $T - \lambda I$ gives

$$V = G \oplus R,$$

and R is invariant under T by 5.18. We show $q(T)|_R = 0$. First, $(T - \lambda I)|_R$ is injective: if $u \in R$ and $(T - \lambda I)u = 0$, then $u \in \operatorname{null}(T - \lambda I) \subseteq G$, so $u \in G \cap R = \{0\}$. Hence $(T - \lambda I)^k|_R$ is injective as well. Now let $u \in R$. Since R is invariant under T, we have $q(T)u \in R$, and

$$(T - \lambda I)^k \big(q(T)u\big) = p(T)u = 0.$$

Injectivity of $(T - \lambda I)^k$ on R forces $q(T)u = 0$. Thus $q(T)|_R = 0$.
Set $r(z) = (z - \lambda)^m q(z)$. On G we have $r(T)|_G = \big((T - \lambda I)^m|_G\big)\big(q(T)|_G\big) = 0$ by the definition of m. On R we have $r(T)|_R = \big((T - \lambda I)^m|_R\big)\big(q(T)|_R\big) = 0$ by the previous paragraph. Since $V = G \oplus R$, every $v \in V$ is a sum of a vector in G and a vector in R, so $r(T) = 0$. By 5.29, r is a polynomial multiple of p; since $r \neq 0$ (because $q \neq 0$, as q is a factor of the nonzero polynomial p), that multiple is nonzero and hence

$$\deg r \geq \deg p, \quad \text{i.e.} \quad m + \deg q \geq k + \deg q,$$

which gives $k \leq m$.
Combining Claims 1 and 2 yields $k = m$. That is, the exponent of $z - \lambda$ in the factorization of the minimal polynomial of T equals the smallest positive integer m such that $(T - \lambda I)^m|_{G(\lambda,T)} = 0$, as claimed. (This is the equality of items (a) and (b) in Exercise 18 of this section. Note that no assumption $\mathbf{F} = \mathbf{C}$ is needed: only 8.4, 8.20, 5.18, 5.27(a), and 5.29 were used.)

Problem 8B.7. Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T with multiplicity d. Prove that $G(\lambda, T) = \operatorname{null}(T - \lambda I)^d$.
[If $d < \dim V$, then this exercise improves 8.20.]

Solution. Let $n = \dim V$ and write $G = G(\lambda, T)$. By the definition 8.23 of multiplicity, $d = \dim G$, and $d \geq 1$ because $G \supseteq E(\lambda, T) \neq \{0\}$.
The inclusion $\operatorname{null}(T - \lambda I)^d \subseteq G$ is immediate from the definition 8.19 of the generalized eigenspace: if $(T - \lambda I)^d v = 0$, then v satisfies $(T - \lambda I)^k v = 0$ for the positive integer $k = d$, so $v \in G$.
For the reverse inclusion we show $(T - \lambda I)^d|_G = 0$.
By 8.20 we have $G = \operatorname{null}(T - \lambda I)^n$, so 5.18 (applied with the polynomial $(z - \lambda)^n$) shows that G is invariant under T, hence also under $T - \lambda I$. Therefore

$$N := (T - \lambda I)\big|_G \in \mathcal{L}(G)$$

is an operator on the d-dimensional vector space G.
The operator N is nilpotent: if $v \in G$, then $(T - \lambda I)^n v = 0$ by 8.20, and since G is invariant under $T - \lambda I$ we have $N^n v = (T - \lambda I)^n v = 0$; thus $N^n = 0$.
Now apply 8.16 to the nilpotent operator N on the vector space G: a nilpotent operator raised to the

dimension of its domain is 0. Since $\dim G = d$, this gives

$$N^d = 0.$$

That is, $(T-\lambda I)^d v = N^d v = 0$ for every $v \in G$, so $G \subseteq \operatorname{null}(T-\lambda I)^d$.
Combining the two inclusions,

$$G(\lambda, T) = \operatorname{null}(T-\lambda I)^d,$$

as desired.
Remark. Since $d = \dim G(\lambda, T) \le \dim V = n$, this improves 8.20, which asserts $G(\lambda, T) = \operatorname{null}(T-\lambda I)^n$. What improves is the exponent, not the subspace: the two null spaces $\operatorname{null}(T-\lambda I)^d$ and $\operatorname{null}(T-\lambda I)^n$ are equal as sets in every case, but when $d < n$ the result above reaches $G(\lambda, T)$ using the smaller exponent d. Exercise 6 sharpens this further, replacing d by the exponent of $z-\lambda$ in the factorization of the minimal polynomial of T. When $\mathbf{F} = \mathbf{C}$, that exponent is at most d: the exponent of $z-\lambda$ in the characteristic polynomial of T is exactly d by the definition 8.26, and the characteristic polynomial is a polynomial multiple of the minimal polynomial by 8.30.
Note finally that nothing above assumed $\mathbf{F} = \mathbf{C}$. The results used, namely 8.16, 8.19, 8.20, 8.23, and 5.18, all hold for $\mathbf{F} = \mathbf{R}$ as well.

Problem 8B.8. Suppose $T \in \mathcal{L}(V)$ and $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T. Prove that

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$$

if and only if the minimal polynomial of T equals $(z-\lambda_1)^{k_1} \cdots (z-\lambda_m)^{k_m}$ for some positive integers $k_1, \dots, k_m$.
The case $\mathbf{F} = \mathbf{C}$ follows immediately from 5.27(b) and the generalized eigenspace decomposition (8.22); thus this exercise is interesting only when $\mathbf{F} = \mathbf{R}$.

Solution. Throughout, $n = \dim V$ and p denotes the minimal polynomial of T. Recall (8.20) that $G(\lambda, T) = \operatorname{null}(T-\lambda I)^n$ for every $\lambda \in \mathbf{F}$, and that this subspace is invariant under T by 5.18 applied to the polynomial $(z-\lambda)^n$.
Two preliminary observations will be used repeatedly.
Observation 1. The sum $G(\lambda_1, T) + \cdots + G(\lambda_m, T)$ is always a direct sum. Indeed, suppose $v_1 + \cdots + v_m = 0$ with $v_k \in G(\lambda_k, T)$ for each k. If some v_k were nonzero, then the nonzero vectors among $v_1, \dots, v_m$ would be a linearly dependent list of generalized eigenvectors of T corresponding to distinct eigenvalues, contradicting 8.12. Hence every v_k equals 0, and 1.45 gives that the sum is direct. Thus the displayed equation in the exercise is equivalent to the single statement $V = G(\lambda_1, T) + \cdots + G(\lambda_m, T)$.
Observation 2. For each k, the operator $(T-\lambda_k I)|_{G(\lambda_k, T)}$ is nilpotent, and if $d_k = \dim G(\lambda_k, T)$ then $(T-\lambda_k I)^{d_k}|_{G(\lambda_k, T)} = 0$. Every $v \in G(\lambda_k, T)$ satisfies $(T-\lambda_k I)^n v = 0$ by 8.20, so the n-th power of the restriction is 0; hence the restriction is nilpotent, and 8.16 applied on the space $G(\lambda_k, T)$ (of dimension d_k) gives $(T-\lambda_k I)^{d_k}|_{G(\lambda_k, T)} = 0$.
The forward implication. Suppose $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$. Because $V \neq \{0\}$, we have $m \ge 1$. Put

$$q(z) = (z-\lambda_1)^{d_1} \cdots (z-\lambda_m)^{d_m}, \qquad d_k = \dim G(\lambda_k, T).$$

Fix j and let $v \in G(\lambda_j, T)$. The factors of $q(T)$ commute with one another, so we may write

$$q(T) = \Big[\prod_{k \neq j} (T-\lambda_k I)^{d_k} \Big] (T-\lambda_j I)^{d_j},$$

and Observation 2 gives $(T-\lambda_j I)^{d_j} v = 0$; hence $q(T)v = 0$. Since every vector of V is a sum of vectors from the subspaces $G(\lambda_1, T), \dots, G(\lambda_m, T)$, we conclude $q(T) = 0$. By 5.29, p divides q, so

$$p(z) = (z-\lambda_1)^{k_1} \cdots (z-\lambda_m)^{k_m}$$

for some integers $k_1, \dots, k_m$ with $0 \le k_j \le d_j$ for each j. Finally, each λ_j is an eigenvalue of T, so by 5.27(a) each λ_j is a zero of p, forcing $k_j \ge 1$. Thus p has the required form with positive exponents.

The reverse implication. By Observation 1 it suffices to prove that V is the sum of the generalized eigenspaces. We prove the following statement by induction on the number m of distinct eigenvalues: if $W \neq \{0\}$ is a finite-dimensional vector space over $\mathbf{F}$ and $S \in \mathcal{L}(W)$ has a minimal polynomial that factors into linear factors over $\mathbf{F}$, then W is the sum of the generalized eigenspaces of S corresponding to the distinct eigenvalues of S.
This suffices: if the minimal polynomial of T equals $(z-\lambda_1)^{k_1}\cdots(z-\lambda_m)^{k_m}$ with all $k_j \geq 1$, then it factors into linear factors over $\mathbf{F}$, and by 5.27(a) its distinct zeros $\lambda_1, \ldots, \lambda_m$ are exactly the distinct eigenvalues of T.
Note first that by 5.27(a) the distinct zeros of the minimal polynomial of such an S are precisely the distinct eigenvalues of S; in particular S has at least one eigenvalue, because the minimal polynomial of an operator on a nonzero space has degree at least 1 and, factoring into linear factors, therefore has a zero.
Base case $m = 1$. Here the minimal polynomial of S is $(z-\mu)^k$ for a single $\mu \in \mathbf{F}$, so $(S-\mu I)^k = 0$. Hence every vector of W lies in $\operatorname{null}(S-\mu I)^k \subseteq G(\mu, S)$, so $W = G(\mu, S)$.
Inductive step. Let $m \geq 2$ and assume the statement holds for all operators (on all nonzero finite-dimensional $\mathbf{F}$-vector spaces) having fewer than m distinct eigenvalues. Suppose $S \in \mathcal{L}(W)$ has minimal polynomial splitting into linear factors and has exactly m distinct eigenvalues $\mu_1, \ldots, \mu_m$. Let $N = S - \mu_m I$ and $\ell = \dim W$. By 8.4 applied to N,

$$W = \operatorname{null} N^{\ell} \oplus \operatorname{range} N^{\ell} = G(\mu_m, S) \oplus U, \qquad U := \operatorname{range}(S-\mu_m I)^{\ell},$$

where the identification $\operatorname{null} N^{\ell} = G(\mu_m, S)$ is 8.20. The subspace U is invariant under S by 5.18 (applied to the polynomial $(z-\mu_m)^{\ell}$).
If $U = \{0\}$ then $W = G(\mu_m, S)$; but then any eigenvector u of S with eigenvalue μ_j, $j \neq m$, would satisfy $0 = (S-\mu_m I)^{\ell} u = (\mu_j - \mu_m)^{\ell} u$, forcing $u = 0$, a contradiction since $m \geq 2$. So $U \neq \{0\}$.
Let $R = S|_U \in \mathcal{L}(U)$. Since the minimal polynomial r of S satisfies $r(S) = 0$ and U is invariant under S, we get $r(R) = r(S)|_U = 0$, so by 5.29 the minimal polynomial of R divides r. In particular the minimal polynomial of R also factors into linear factors over $\mathbf{F}$, and by 5.27(a) every eigenvalue of R is a zero of r, hence lies in $\{\mu_1, \ldots, \mu_m\}$. Moreover μ_m is not an eigenvalue of R: if $u \in U$ is nonzero with $Su = \mu_m u$, then $Nu = 0$, so $u \in \operatorname{null} N^{\ell} \cap U = \{0\}$, a contradiction. Therefore R has at most $m-1$ distinct eigenvalues, all belonging to $\{\mu_1, \ldots, \mu_{m-1}\}$.
By the induction hypothesis, U is the sum of the generalized eigenspaces $G(\mu, R)$ over the distinct eigenvalues μ of R. For each such μ we have $G(\mu, R) \subseteq G(\mu, S)$, because a vector $u \in U$ annihilated by some power of $R - \mu I_U$ is annihilated by the same power of $S - \mu I$, as R is the restriction of S to U. Hence

$$U \subseteq G(\mu_1, S) + \cdots + G(\mu_{m-1}, S),$$

and therefore

$$W = G(\mu_m, S) + U \subseteq G(\mu_1, S) + \cdots + G(\mu_m, S) \subseteq W.$$

So W is the sum of the generalized eigenspaces of S, completing the induction and the proof.
Remark. When $\mathbf{F} = \mathbf{C}$ both sides always hold: 5.27(b) says the minimal polynomial factors as $(z-\lambda_1)^{k_1}\cdots(z-\lambda_m)^{k_m}$ over the distinct eigenvalues, and 8.22(c) gives the generalized eigenspace decomposition. The content of the exercise is therefore the real case, where the minimal polynomial need not split.

Problem 8B.9. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that there exist $D, N \in \mathcal{L}(V)$ such that $T = D + N$, the operator D is diagonalizable, N is nilpotent, and $DN = ND$.

Solution. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T and let $n = \dim V$. Since $\mathbf{F} = \mathbf{C}$, the generalized eigenspace decomposition (8.22) gives

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

each $G(\lambda_k, T)$ is invariant under T, and each $(T - \lambda_k I)|_{G(\lambda_k, T)}$ is nilpotent.

Definition of D. Because the sum above is direct, every $v \in V$ has a unique representation $v = v_1 + \cdots + v_m$ with $v_k \in G(\lambda_k, T)$. Define

$$Dv = \lambda_1 v_1 + \cdots + \lambda_m v_m.$$

Uniqueness of the representation makes D a well-defined function, and it is linear: if $v = v_1 + \cdots + v_m$ and $w = w_1 + \cdots + w_m$ are the representations of v and w, and $\alpha \in \mathbf{C}$, then $(v_1 + \alpha w_1) + \cdots + (v_m + \alpha w_m)$ is the representation of $v + \alpha w$ (each summand lies in the correct subspace, since each $G(\lambda_k, T)$ is a subspace), so $D(v + \alpha w) = Dv + \alpha Dw$. Thus $D \in \mathcal{L}(V)$, and by construction

$$Dv = \lambda_k v \quad \text{for every } v \in G(\lambda_k, T).$$

D is diagonalizable. Choose a basis of each $G(\lambda_k, T)$ and concatenate; by the generalized eigenspace decomposition this list is a basis of V, and by the displayed formula every vector in it is an eigenvector of D. Hence V has a basis consisting of eigenvectors of D, so D is diagonalizable by 5.55.
$N := T - D$ is nilpotent. Fix k and let $v \in G(\lambda_k, T)$. Then $Nv = Tv - \lambda_k v = (T - \lambda_k I)v$, and $(T - \lambda_k I)v \in G(\lambda_k, T)$ because $G(\lambda_k, T)$ is invariant under T (8.22(a)) and hence under $T - \lambda_k I$. So $G(\lambda_k, T)$ is invariant under N and

$$N|_{G(\lambda_k, T)} = (T - \lambda_k I)|_{G(\lambda_k, T)},$$

which is nilpotent by 8.22(b). Applying 8.16 on the space $G(\lambda_k, T)$, whose dimension is at most n, gives $\big(N|_{G(\lambda_k, T)}\big)^n = 0$, i.e. $N^n v = 0$ for all $v \in G(\lambda_k, T)$. Since V is the sum of the subspaces $G(\lambda_k, T)$, we get $N^n = 0$, so N is nilpotent.
$DN = ND$. Fix k and let $v \in G(\lambda_k, T)$. As just shown, $Nv \in G(\lambda_k, T)$, so $D(Nv) = \lambda_k Nv$. On the other hand $Dv = \lambda_k v$, so $N(Dv) = \lambda_k Nv$. Hence $DNv = NDv$ for every $v \in G(\lambda_k, T)$ and every k. Since V is the sum of these subspaces and DN, ND are linear, $DN = ND$.
Thus $T = D + N$ with D diagonalizable, N nilpotent, and $DN = ND$, as required.

Problem 8B.10. Suppose V is a complex inner product space, $e_1, \ldots, e_n$ is an orthonormal basis of V, and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, each included as many times as its multiplicity. Prove that

$$|\lambda_1|^2 + \cdots + |\lambda_n|^2 \le \|Te_1\|^2 + \cdots + \|Te_n\|^2.$$

See the comment after Exercise 5 in Section 7A.

Solution. First note that the list $\lambda_1, \ldots, \lambda_n$ really does have $n = \dim V$ entries: by 8.25 the sum of the multiplicities of all eigenvalues of T equals $\dim V$.
Step 1: the right-hand side does not depend on the orthonormal basis. This is the content of Exercise 5 in Section 7A and the comment printed after it; here is the short proof. Fix an orthonormal basis $f_1, \ldots, f_n$ of V once and for all, and let $e_1, \ldots, e_n$ be any orthonormal basis of V. Applying Parseval's identity 6.30(b) to the vector Te_k with respect to $f_1, \ldots, f_n$, and then the definition of the adjoint (7.1),

$$\sum_{k=1}^{n} \|Te_k\|^2 = \sum_{k=1}^{n}\sum_{j=1}^{n} \big|\langle Te_k, f_j\rangle\big|^2 = \sum_{j=1}^{n}\sum_{k=1}^{n} \big|\langle e_k, T^* f_j\rangle\big|^2 = \sum_{j=1}^{n} \|T^* f_j\|^2,$$

the last equality being Parseval's identity applied to $T^* f_j$ with respect to $e_1, \ldots, e_n$. The far right side does not involve $e_1, \ldots, e_n$ at all, so the quantity $\sum_{k=1}^{n} \|Te_k\|^2$ is the same for every orthonormal basis $e_1, \ldots, e_n$ of V.
Step 2: compute the sum using a Schur basis. Since V is a finite-dimensional complex inner product space, Schur's theorem (6.38) provides an orthonormal basis $u_1, \ldots, u_n$ of V with respect to which T has an upper-triangular matrix

$$A = \mathcal{M}\big(T, (u_1, \ldots, u_n)\big) = (a_{j,k}),$$

so $a_{j,k} = 0$ whenever $j > k$. By 5.41 the eigenvalues of T are exactly the diagonal entries $a_{1,1}, \dots, a_{n,n}$, and by 8.31 each eigenvalue λ of T appears on this diagonal exactly as many times as its multiplicity. Hence the list $a_{1,1}, \dots, a_{n,n}$ is a rearrangement of the list $\lambda_1, \dots, \lambda_n$, and in particular

$$\sum_{k=1}^{n} |a_{k,k}|^2 = \sum_{k=1}^{n} |\lambda_k|^2.$$

For each k we have $Tu_k = \sum_{j=1}^{n} a_{j,k} u_j$ by the definition of the matrix of an operator, so by the formula for the norm of an orthonormal linear combination (6.24),

$$\|Tu_k\|^2 = \sum_{j=1}^{n} |a_{j,k}|^2.$$

Summing over k,

$$\sum_{k=1}^{n} \|Tu_k\|^2 = \sum_{k=1}^{n} \sum_{j=1}^{n} |a_{j,k}|^2 \geq \sum_{k=1}^{n} |a_{k,k}|^2 = \sum_{k=1}^{n} |\lambda_k|^2,$$

where the inequality holds because we have discarded the nonnegative terms $|a_{j,k}|^2$ with $j \neq k$.
Step 3: conclude. By Step 1, $\sum_{k=1}^{n} \|Te_k\|^2 = \sum_{k=1}^{n} \|Tu_k\|^2$. Combining with Step 2,

$$|\lambda_1|^2 + \cdots + |\lambda_n|^2 \leq \|Te_1\|^2 + \cdots + \|Te_n\|^2,$$

as desired.
Remark. The proof shows exactly when equality holds: equality forces $a_{j,k} = 0$ for all $j \neq k$, i.e. T has a diagonal matrix with respect to some orthonormal basis, which by the complex spectral theorem (7.31) happens precisely when T is normal.

Problem 8B.11. Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $(z-7)^2(z-8)^2$.

Solution. Define $T \in \mathcal{L}(\mathbf{C}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (7z_1,\ 7z_2,\ 8z_3,\ 8z_4).$$

With respect to the standard basis e_1, e_2, e_3, e_4 of $\mathbf{C}^4$, the matrix of T is the diagonal matrix with diagonal entries $7, 7, 8, 8$.
By 5.41 the eigenvalues of T are 7 and 8. To compute the multiplicities, note that $T - 7I$ is the operator $(z_1, z_2, z_3, z_4) \mapsto (0, 0, z_3, z_4)$, so $(T-7I)^4$ is the same operator, whose null space is $\operatorname{span}(e_1, e_2)$. Hence by 8.20 and 8.23,

$$G(7, T) = \operatorname{null}(T - 7I)^4 = \operatorname{span}(e_1, e_2),$$

so 7 has multiplicity 2. Similarly $T - 8I$ is $(z_1, z_2, z_3, z_4) \mapsto (-z_1, -z_2, 0, 0)$, so $(T - 8I)^4$ is $(z_1, z_2, z_3, z_4) \mapsto (z_1, z_2, 0, 0)$ and

$$G(8, T) = \operatorname{null}(T - 8I)^4 = \operatorname{span}(e_3, e_4),$$

so 8 has multiplicity 2. By the definition of the characteristic polynomial (8.26), the characteristic polynomial of T is $(z-7)^2(z-8)^2$.
(The minimal polynomial of this particular T is $(z-7)(z-8)$, since T is diagonalizable; see 5.62. The exercise imposes no condition on the minimal polynomial, so any operator whose matrix is upper triangular with diagonal $7, 7, 8, 8$ would also work, by 8.31.)

Problem 8B.12. Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $(z-1)(z-5)^3$ and whose minimal polynomial equals $(z-1)(z-5)^2$.

Solution. Define $T \in \mathcal{L}(\mathbf{C}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (z_1,\ 5z_2 + z_3,\ 5z_3,\ 5z_4),$$

so that with respect to the standard basis e_1, e_2, e_3, e_4 the matrix of T is the block diagonal matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

Explicitly, $Te_1 = e_1$, $Te_2 = 5e_2$, $Te_3 = e_2 + 5e_3$, $Te_4 = 5e_4$.
Characteristic polynomial. The matrix is upper triangular with diagonal entries $1, 5, 5, 5$, so by 5.41 the eigenvalues of T are 1 and 5, and by 8.31 their multiplicities are 1 and 3 respectively. Hence by 8.26 the characteristic polynomial of T is $(z-1)(z-5)^3$.
(One can also see the multiplicities directly. Let $U = \operatorname{span}(e_1)$ and $W = \operatorname{span}(e_2, e_3, e_4)$; both are invariant under T, and $\mathbf{C}^4 = U \oplus W$. On U the operator $T - 5I$ acts as multiplication by -4, hence is injective, while on W the operator $T - 5I$ satisfies $e_2 \mapsto 0$, $e_3 \mapsto e_2$, $e_4 \mapsto 0$ and so is nilpotent. Therefore $G(5, T) = \operatorname{null}(T - 5I)^4 = W$, of dimension 3. Symmetrically $T - I$ is injective on W, since its matrix on W is upper triangular with all diagonal entries equal to 4, and $(T - I)e_1 = 0$; so $G(1, T) = U$, of dimension 1.)
Minimal polynomial. Let $q(z) = (z-1)(z-5)^2$. We check $q(T) = 0$ on the basis. Since $(T - I)e_1 = 0$ and the factors $(T - I)$ and $(T - 5I)^2$ commute,

$$q(T)e_1 = (T - 5I)^2(T - I)e_1 = 0.$$

For e_2, e_3, e_4 it suffices to note that $(T - 5I)^2$ kills each of them: $(T - 5I)e_2 = 0$, $(T - 5I)e_4 = 0$, and $(T - 5I)e_3 = e_2$ so that $(T - 5I)^2 e_3 = (T - 5I)e_2 = 0$. Hence $q(T)e_k = (T - I)(T - 5I)^2 e_k = 0$ for $k = 2, 3, 4$. Thus $q(T) = 0$, so by 5.29 the minimal polynomial p of T divides q.
By 5.27(a) the zeros of p are exactly the eigenvalues 1 and 5 of T, so p is either $(z-1)(z-5)$ or $(z-1)(z-5)^2$. The first is ruled out:

$$(T - I)(T - 5I)e_3 = (T - I)e_2 = 5e_2 - e_2 = 4e_2 \neq 0.$$

Therefore the minimal polynomial of T equals $(z-1)(z-5)^2$, as required.

Problem 8B.13. Give an example of an operator on $\mathbf{C}^4$ whose characteristic and minimal polynomials both equal $z(z-1)^2(z-3)$.

Solution. Define $T \in \mathcal{L}(\mathbf{C}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (0,\ z_2 + z_3,\ z_3,\ 3z_4).$$

On the standard basis this reads $Te_1 = 0$, $Te_2 = e_2$, $Te_3 = e_2 + e_3$, $Te_4 = 3e_4$, so the matrix of T with respect to e_1, e_2, e_3, e_4 is the block diagonal matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Characteristic polynomial. This matrix is upper triangular with diagonal entries $0, 1, 1, 3$. By 5.41 the eigenvalues of T are 0, 1, and 3, and by 8.31 their multiplicities are the numbers of times they occur

on the diagonal, namely 1, 2, and 1. Hence by 8.26 the characteristic polynomial of T is

$$z(z-1)^2(z-3).$$

Equivalently, one checks directly that $G(0,T) = \operatorname{span}(e_1)$, $G(1,T) = \operatorname{span}(e_2, e_3)$, and $G(3,T) = \operatorname{span}(e_4)$: each of these three subspaces is invariant under T, they give a direct sum decomposition of $\mathbf{C}^4$, and on each of them the corresponding operator $T - \lambda I$ is nilpotent while on the other two summands $T - \lambda I$ is upper triangular with nonzero diagonal entries, hence injective.
Minimal polynomial. Let $q(z) = z(z-1)^2(z-3)$. The factors of $q(T)$ commute with one another, so to check $q(T) = 0$ it suffices to check that each basis vector is killed by one factor:

- $Te_1 = 0$;
- $(T-I)^2 e_2 = 0$ and $(T-I)^2 e_3 = (T-I)e_2 = 0$, since $(T-I)e_3 = e_2$ and $(T-I)e_2 = 0$;
- $(T-3I)e_4 = 0$.

Hence $q(T) = 0$, so by 5.29 the minimal polynomial p of T divides q. By 5.27(a) the zeros of p are exactly 0, 1, 3, so p equals $z(z-1)(z-3)$ or $z(z-1)^2(z-3)$. The first possibility fails:

$$(T-3I)e_3 = e_2 - 2e_3, \qquad (T-I)(e_2 - 2e_3) = 0 - 2e_2 = -2e_2, \qquad T(-2e_2) = -2e_2 \neq 0.$$

Thus $T(T-I)(T-3I) \neq 0$, and therefore the minimal polynomial of T equals $z(z-1)^2(z-3)$, which is also its characteristic polynomial.

Problem 8B.14. Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $z(z-1)^2(z-3)$ and whose minimal polynomial equals $z(z-1)(z-3)$.

Solution. Define $T \in \mathcal{L}(\mathbf{C}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (0,\ z_2,\ z_3,\ 3z_4),$$

so that the matrix of T with respect to the standard basis e_1, e_2, e_3, e_4 is the diagonal matrix with diagonal entries $0, 1, 1, 3$:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

This is the operator of the previous exercise with the off-diagonal 1 deleted; deleting it does not change the diagonal, hence does not change the characteristic polynomial, but it does shrink the minimal polynomial.
Characteristic polynomial. The matrix is (upper) triangular with diagonal entries $0, 1, 1, 3$, so by 5.41 the eigenvalues of T are 0, 1, 3, and by 8.31 their multiplicities are 1, 2, 1. Concretely, $T - \lambda I$ is diagonal with entries $-\lambda, 1-\lambda, 1-\lambda, 3-\lambda$, so

$$G(0,T) = \operatorname{null} T^4 = \operatorname{span}(e_1), \quad G(1,T) = \operatorname{null}(T-I)^4 = \operatorname{span}(e_2, e_3), \quad G(3,T) = \operatorname{span}(e_4),$$

of dimensions 1, 2, 1. By 8.26 the characteristic polynomial of T is therefore

$$z(z-1)^2(z-3).$$

Minimal polynomial. The standard basis consists of eigenvectors of T, so T is diagonalizable (5.55). By 5.62 the minimal polynomial of a diagonalizable operator is $(z-\lambda_1)\cdots(z-\lambda_m)$, where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues; here that gives

$$z(z-1)(z-3).$$

Directly: writing $q(z) = z(z-1)(z-3)$, the commuting factors of $q(T)$ satisfy $Te_1 = 0$, $(T-I)e_2 = (T-I)e_3 = 0$, and $(T-3I)e_4 = 0$, so $q(T) = 0$ on a basis and hence $q(T) = 0$. By 5.29 the minimal polynomial divides q, and by 5.27(a) it must have 0, 1, 3 among its zeros, so it equals q. Thus the minimal polynomial is $z(z-1)(z-3)$ while the characteristic polynomial is $z(z-1)^2(z-3)$, as required.

Problem 8B.15. Let T be the operator on $\mathbf{C}^4$ defined by $T(z_1, z_2, z_3, z_4) = (0, z_1, z_2, z_3)$. Find the characteristic polynomial and the minimal polynomial of T.

Solution. Write e_1, e_2, e_3, e_4 for the standard basis of $\mathbf{C}^4$. The formula for T says

$$Te_1 = e_2, \qquad Te_2 = e_3, \qquad Te_3 = e_4, \qquad Te_4 = 0.$$

Iterating the formula gives

$$\begin{aligned} T^2(z_1, z_2, z_3, z_4) &= T(0, z_1, z_2, z_3) = (0, 0, z_1, z_2), \\ T^3(z_1, z_2, z_3, z_4) &= T(0, 0, z_1, z_2) = (0, 0, 0, z_1), \\ T^4(z_1, z_2, z_3, z_4) &= T(0, 0, 0, z_1) = (0, 0, 0, 0). \end{aligned}$$

Thus $T^4 = 0$, so T is nilpotent, while $T^3 \neq 0$ because $T^3 e_1 = e_4 \neq 0$.
Minimal polynomial. Because T is nilpotent, its minimal polynomial is z^m for some positive integer m (by 8.18, or directly by 5.29 applied to the polynomial z^4, which annihilates T: the minimal polynomial must divide z^4 and hence has the form z^m with $m \leq 4$). Since $T^3 \neq 0$, we cannot have $m \leq 3$. Hence $m = 4$ and the minimal polynomial of T is

$$z^4.$$

Characteristic polynomial. Since T is nilpotent, 0 is the only eigenvalue of T (by 8.17). Hence the characteristic polynomial of T is z^d, where d is the multiplicity of the eigenvalue 0, that is, $d = \dim G(0, T)$. By 8.25 the multiplicities of the eigenvalues of T sum to $\dim \mathbf{C}^4 = 4$, and 0 is the only eigenvalue, so $d = 4$. (Equivalently: $G(0, T) = \operatorname{null} T^4 = \mathbf{C}^4$ by 8.20.) Therefore the characteristic polynomial of T is

$$z^4.$$

So for this operator the characteristic polynomial and the minimal polynomial both equal z^4. This is consistent with 8.30, which requires the characteristic polynomial to be a polynomial multiple of the minimal polynomial.

Problem 8B.16. Let T be the operator on $\mathbf{C}^6$ defined by

$$T(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0).$$

Find the characteristic polynomial and the minimal polynomial of T.

Solution. Let $e_1, \ldots, e_6$ be the standard basis of $\mathbf{C}^6$. Reading off the formula for T (the entry z_1 sits in slot 2, the entry z_2 in slot 3, the entry z_4 in slot 5, and no other z_j appears), we get

$$Te_1 = e_2, \quad Te_2 = e_3, \quad Te_3 = 0, \quad Te_4 = e_5, \quad Te_5 = 0, \quad Te_6 = 0.$$

So T splits into three independent chains: $e_1 \mapsto e_2 \mapsto e_3 \mapsto 0$, then $e_4 \mapsto e_5 \mapsto 0$, then $e_6 \mapsto 0$.
Computing powers directly from the formula,

$$\begin{aligned} T^2(z_1, \ldots, z_6) &= T(0, z_1, z_2, 0, z_4, 0) = (0, 0, z_1, 0, 0, 0), \\ T^3(z_1, \ldots, z_6) &= T(0, 0, z_1, 0, 0, 0) = (0, 0, 0, 0, 0, 0). \end{aligned}$$

Thus $T^3 = 0$, so T is nilpotent, while $T^2 \neq 0$ because $T^2 e_1 = e_3 \neq 0$.
Minimal polynomial. Because $T^3 = 0$, the minimal polynomial of T divides z^3 (by 5.29), so it equals z^m for some $m \in \{1, 2, 3\}$. Since $T^2 \neq 0$, we must have $m = 3$. Hence the minimal polynomial of T is

$$z^3.$$

Characteristic polynomial. Since T is nilpotent, 0 is the only eigenvalue of T (by 8.17). By 8.20,

$$G(0, T) = \operatorname{null} T^6 = \mathbf{C}^6$$

because $T^6 = 0$. So the multiplicity of the eigenvalue 0 is 6, and the characteristic polynomial of T is

$$z^6.$$

Note that z^6 is indeed a polynomial multiple of z^3, as 8.30 demands, and that the characteristic polynomial has degree $6 = \dim \mathbf{C}^6$, as 8.28(a) demands.

Problem 8B.17. Suppose $\mathbf{F} = \mathbf{C}$ and $P \in \mathcal{L}(V)$ is such that $P^2 = P$. Prove that the characteristic polynomial of P is $z^m(z-1)^n$, where $m = \dim \operatorname{null} P$ and $n = \dim \operatorname{range} P$.

Solution. Throughout, $m = \dim \operatorname{null} P$ and $n = \dim \operatorname{range} P$; by the fundamental theorem of linear maps (3.21),

$$m + n = \dim V.$$

The characteristic polynomial of P is by definition (8.26) the product of $(z-\lambda)^{\dim G(\lambda,P)}$ over the distinct eigenvalues λ of P. So we must show two things: every eigenvalue of P lies in $\{0,1\}$, and the generalized eigenspaces satisfy $\dim G(0,P) = m$ and $\dim G(1,P) = n$. (When $m = 0$ the factor z^m is 1 and 0 is not an eigenvalue, and similarly when $n = 0$; so the claimed formula is correct in those degenerate cases too, and we will see this below.)

If $\dim V = 0$ then P has no eigenvalues, its characteristic polynomial is the empty product 1, and $m = n = 0$, so $z^m(z-1)^n = 1$ and the result holds. Assume from now on that $\dim V \geq 1$.

Step 1: the possible eigenvalues. Suppose $\lambda \in \mathbf{C}$ and $v \in V$ satisfy $v \neq 0$ and $Pv = \lambda v$. Applying P again and using $P^2 = P$ gives

$$\lambda v = Pv = P^2 v = P(\lambda v) = \lambda^2 v.$$

Since $v \neq 0$ this forces $\lambda^2 = \lambda$, that is, $\lambda \in \{0,1\}$.

Step 2: $G(0,P) = \operatorname{null} P$. Since $P^2 = P$, induction gives $P^k = P$ for every positive integer k. Hence taking $k = \dim V \geq 1$ we get, using 8.20,

$$G(0,P) = \operatorname{null} P^{\dim V} = \operatorname{null} P,$$

so $\dim G(0,P) = m$. Consequently 0 is an eigenvalue of P exactly when $m > 0$, and in that case its multiplicity is m.

Step 3: $G(1,P) = \operatorname{range} P$. First note that P restricted to $\operatorname{range} P$ is the identity: if $v \in \operatorname{range} P$, say $v = Pu$, then

$$Pv = P^2 u = Pu = v.$$

Thus $\operatorname{range} P \subseteq \operatorname{null}(P - I)$. Conversely if $Pv = v$ then $v \in \operatorname{range} P$. Hence

$$\operatorname{null}(P - I) = \operatorname{range} P.$$

Next, $(P-I)^2 = P^2 - 2P + I = P - 2P + I = I - P$, and more generally, since $I - P$ is also idempotent (because $(I-P)^2 = I - 2P + P^2 = I - P$), induction gives $(P-I)^k = (-1)^k(I-P)$ for every positive integer k. Therefore $\operatorname{null}(P-I)^k = \operatorname{null}(I-P) = \operatorname{null}(P-I) = \operatorname{range} P$ for every positive integer k. Taking $k = \dim V$ and using 8.20,

$$G(1,P) = \operatorname{null}(P-I)^{\dim V} = \operatorname{range} P,$$

so $\dim G(1,P) = n$. Consequently 1 is an eigenvalue of P exactly when $n > 0$, and in that case its multiplicity is n.

Step 4: assembling the characteristic polynomial. By Step 1 the distinct eigenvalues of P form a subset of $\{0,1\}$, and by Steps 2 and 3 the eigenvalues are precisely those elements λ of $\{0,1\}$ for which the corresponding number (m for $\lambda = 0$, n for $\lambda = 1$) is positive, with multiplicity equal to that number. Hence the characteristic polynomial of P is

$$z^m(z-1)^n,$$

where a factor with exponent 0 contributes nothing, exactly matching the definition 8.26. (As a check, the degree of this polynomial is $m + n = \dim V$, in agreement with 8.28(a).)

Remark. Steps 2 and 3 also reprove the direct sum decomposition $V = \operatorname{null} P \oplus \operatorname{range} P$ for an idempotent P: it is the generalized eigenspace decomposition 8.22(c) in this case.

Problem 8B.18. Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. Explain why the following four numbers equal each other.

- (a) The exponent of $z - \lambda$ in the factorization of the minimal polynomial of T.
- (b) The smallest positive integer m such that $(T - \lambda I)^m|_{G(\lambda,T)} = 0$.
- (c) The smallest positive integer m such that $\operatorname{null}(T - \lambda I)^m = \operatorname{null}(T - \lambda I)^{m+1}$.
- (d) The smallest positive integer m such that $\operatorname{range}(T - \lambda I)^m = \operatorname{range}(T - \lambda I)^{m+1}$.

Solution. Write $N = T - \lambda I$ and $n = \dim V$, and let p denote the minimal polynomial of T. Call the four numbers a, b, c, d respectively. We show c is well defined, then prove $d = c$, then $b = c$, then $a = c$.

Preliminaries on the number c. By 8.1 the null spaces of the powers of N increase,

$$\{0\} = \operatorname{null} N^0 \subseteq \operatorname{null} N^1 \subseteq \operatorname{null} N^2 \subseteq \cdots,$$

and by 8.3 we have $\operatorname{null} N^n = \operatorname{null} N^{n+1}$. Hence the set of positive integers m with $\operatorname{null} N^m = \operatorname{null} N^{m+1}$ is nonempty (it contains n), so c exists and $c \leq n$. Moreover, by 8.2, once two consecutive null spaces agree all later ones agree:

$$\operatorname{null} N^c = \operatorname{null} N^{c+1} = \operatorname{null} N^{c+2} = \cdots.$$

In particular, since $c \leq n$,

$$\operatorname{null} N^c = \operatorname{null} N^n = G(\lambda, T), \tag{1}$$

the last equality by 8.20. Also, by minimality of c and 8.2 again, the inclusions

$$\{0\} = \operatorname{null} N^0 \subsetneq \operatorname{null} N^1 \subsetneq \cdots \subsetneq \operatorname{null} N^c$$

are all strict. Finally, $c \geq 1$ trivially, and $\operatorname{null} N^1 \neq \{0\}$ because λ is an eigenvalue of T.

Proof that $d = c$. Fix a positive integer m. The ranges decrease, $\operatorname{range} N^{m+1} \subseteq \operatorname{range} N^m$, because $N^{m+1}v = N^m(Nv)$. Hence $\operatorname{range} N^m = \operatorname{range} N^{m+1}$ if and only if $\dim \operatorname{range} N^m = \dim \operatorname{range} N^{m+1}$ (by 2.39). By the fundamental theorem of linear maps (3.21),

$$\dim \operatorname{range} N^k = n - \dim \operatorname{null} N^k$$

for every k. So $\dim \operatorname{range} N^m = \dim \operatorname{range} N^{m+1}$ if and only if $\dim \operatorname{null} N^m = \dim \operatorname{null} N^{m+1}$, which (since $\operatorname{null} N^m \subseteq \operatorname{null} N^{m+1}$) holds if and only if $\operatorname{null} N^m = \operatorname{null} N^{m+1}$. Thus a positive integer m satisfies the condition in (d) precisely when it satisfies the condition in (c); the smallest such integers therefore agree, so $d = c$.

Proof that $b = c$. For a positive integer m we have

$$N^m|_{G(\lambda,T)} = 0 \iff G(\lambda, T) \subseteq \operatorname{null} N^m.$$

By (1), $G(\lambda, T) = \operatorname{null} N^c$. If $m \geq c$ then $\operatorname{null} N^c \subseteq \operatorname{null} N^m$, so the condition holds. If $m < c$ then $\operatorname{null} N^m \subsetneq \operatorname{null} N^c = G(\lambda, T)$ by the strict inclusions noted above, so $G(\lambda, T) \not\subseteq \operatorname{null} N^m$ and the condition fails. Hence the smallest positive integer m with $N^m|_{G(\lambda,T)} = 0$ is c; that is, $b = c$.

Proof that $a = c$. Since λ is an eigenvalue of T, λ is a zero of p (by 5.27(a)), so $a \geq 1$. Write

$$p(z) = (z - \lambda)^a q(z),$$

where $q \in \mathcal{P}(\mathbf{F})$ and $q(\lambda) \neq 0$ (this is what it means for a to be the exponent of $z - \lambda$ in the factorization of p: divide out the factor $z - \lambda$ as often as possible). Note q is monic because p is monic.

First we show $c \leq a$, by showing $\operatorname{null} N^a = \operatorname{null} N^{a+1}$. The inclusion $\subseteq$ is 8.1. For the reverse, suppose $v \in \operatorname{null} N^{a+1}$ and set $w = N^a v$. Then $Nw = N^{a+1}v = 0$, so $Tw = \lambda w$, and therefore $q(T)w = q(\lambda)w$. On the other hand, because the operators $N^a = (T - \lambda I)^a$ and $q(T)$ are polynomials in T they commute, so

$$q(\lambda)w = q(T)w = q(T)N^a v = p(T)v = 0,$$

using $p(T) = 0$. Since $q(\lambda) \neq 0$, we get $w = 0$, that is, $N^a v = 0$. Hence $\operatorname{null} N^{a+1} \subseteq \operatorname{null} N^a$, so the two null spaces are equal and therefore $c \leq a$ by minimality of c.

Now we show $a \le c$. Suppose instead $a > c$, so $a - 1 \ge c$ and $a - 1 \ge 1$. Set

$$r(z) = (z - \lambda)^{a-1} q(z),$$

a monic polynomial with $\deg r = \deg p - 1$. We claim $r(T) = 0$. Let $v \in V$. Then

$$N^a\big(q(T)v\big) = p(T)v = 0,$$

so $q(T)v \in \operatorname{null} N^a$. Since $a-1 \ge c$, the stabilization statement following (1) gives $\operatorname{null} N^a = \operatorname{null} N^{a-1}$. Hence $q(T)v \in \operatorname{null} N^{a-1}$, that is,

$$r(T)v = N^{a-1} q(T)v = 0.$$

As $v \in V$ was arbitrary, $r(T) = 0$. But r is a monic polynomial with $\deg r < \deg p$, contradicting the fact that the minimal polynomial p is the monic polynomial of smallest degree annihilating T (5.22). This contradiction shows $a \le c$.
Combining, $a = c$. Together with $b = c$ and $d = c$, all four numbers are equal.

Problem 8B.19. Suppose $\mathbf{F} = \mathbf{C}$ and $S \in \mathcal{L}(V)$ is a unitary operator. Prove that the constant term in the characteristic polynomial of S has absolute value 1.

Solution. Here V is a finite-dimensional complex inner product space and S is unitary, meaning S is an isometry of V onto itself, so $\|Sv\| = \|v\|$ for all $v \in V$.
Let $\lambda_1, \dots, \lambda_M$ denote the distinct eigenvalues of S, with multiplicities $d_1, \dots, d_M$, and let q denote the characteristic polynomial of S, so by 8.26

$$q(z) = (z - \lambda_1)^{d_1} \cdots (z - \lambda_M)^{d_M}.$$

The constant term of a polynomial q is $q(0)$. Evaluating,

$$q(0) = (-\lambda_1)^{d_1} \cdots (-\lambda_M)^{d_M} = (-1)^{d_1 + \cdots + d_M} \lambda_1^{d_1} \cdots \lambda_M^{d_M}.$$

Taking absolute values and using multiplicativity of the absolute value on $\mathbf{C}$,

$$|q(0)| = |\lambda_1|^{d_1} \cdots |\lambda_M|^{d_M}.$$

Every eigenvalue of a unitary operator has absolute value 1 (by 7.54; indeed, if $Sv = \lambda v$ with $v \ne 0$, then $|\lambda| \, \|v\| = \|\lambda v\| = \|Sv\| = \|v\|$, and dividing by $\|v\| \ne 0$ gives $|\lambda| = 1$). Hence $|\lambda_k| = 1$ for each k, and therefore

$$|q(0)| = 1^{d_1} \cdots 1^{d_M} = 1.$$

So the constant term in the characteristic polynomial of S has absolute value 1, as desired. (If $\dim V = 0$ there are no eigenvalues, the characteristic polynomial is the empty product 1, and its constant term is 1, which also has absolute value 1.)
Remark. The computation above shows more: writing $n = \dim V$, we have $d_1 + \cdots + d_M = n$ by 8.25, so the constant term equals $(-1)^n \lambda_1^{d_1} \cdots \lambda_M^{d_M}$, which is $(-1)^n$ times a product of complex numbers of absolute value 1. In determinant language this is the statement that a unitary operator has determinant of absolute value 1.

Problem 8B.20. Suppose that $\mathbf{F} = \mathbf{C}$ and $V_1, \dots, V_m$ are nonzero subspaces of V such that

$$V = V_1 \oplus \cdots \oplus V_m.$$

Suppose $T \in \mathcal{L}(V)$ and each V_k is invariant under T. For each k, let p_k denote the characteristic polynomial of $T|_{V_k}$. Prove that the characteristic polynomial of T equals $p_1 \cdots p_m$.

Solution. Let $n = \dim V$ and, for $\lambda \in \mathbf{C}$, write $N = T - \lambda I$. For a subspace U of V invariant under T we write $G(\lambda, T|_U)$ for the generalized eigenspace of the operator $T|_U \in \mathcal{L}(U)$; recall from 8.20 that

$$G(\lambda, T|_U) = \operatorname{null}\big((T|_U - \lambda I_U)^{\dim U}\big) = \{u \in U : N^{\dim U} u = 0\}.$$

It is convenient to extend the notion of multiplicity to every $\lambda \in \mathbf{C}$ by setting the multiplicity of λ (for an operator R) equal to $\dim G(\lambda, R)$; this agrees with 8.23 when λ is an eigenvalue, and it equals 0 when λ is not an eigenvalue, because then $R - \lambda I$ is injective, hence so is every power of it, and thus $G(\lambda, R) = \operatorname{null}(R - \lambda I)^{\dim W} = \{0\}$, where W is the space on which R acts. With this convention, the definition 8.26 of the characteristic polynomial of an operator R on a complex vector space reads

$$\text{char poly of } R = \prod_{\lambda \in \mathbf{C}} (z - \lambda)^{\dim G(\lambda, R)},$$

where all but finitely many factors equal 1.
Key claim. For every $\lambda \in \mathbf{C}$,

$$G(\lambda, T) = G(\lambda, T|_{V_1}) \oplus \cdots \oplus G(\lambda, T|_{V_m}).$$

Proof of the claim. First, each V_k is invariant under T and hence under $N = T - \lambda I$, and hence under every power N^j.
Inclusion $\supseteq$. Suppose $v_k \in G(\lambda, T|_{V_k})$. Then $N^{\dim V_k} v_k = 0$. Since $\dim V_k \le n$, we get $N^n v_k = N^{n - \dim V_k}\big(N^{\dim V_k} v_k\big) = 0$, so $v_k \in \operatorname{null} N^n = G(\lambda, T)$ by 8.20. As $G(\lambda, T)$ is a subspace, it contains the sum of the $G(\lambda, T|_{V_k})$.
Inclusion $\subseteq$. Suppose $v \in G(\lambda, T) = \operatorname{null} N^n$. Because $V = V_1 \oplus \cdots \oplus V_m$, we may write $v = v_1 + \cdots + v_m$ with $v_k \in V_k$. Applying N^n and using the invariance of each V_k under N^n,

$$0 = N^n v = N^n v_1 + \cdots + N^n v_m, \qquad N^n v_k \in V_k.$$

By the uniqueness of the decomposition of 0 in a direct sum (1.45), $N^n v_k = 0$ for each k. Thus v_k lies in $\operatorname{null}\big((T|_{V_k} - \lambda I_{V_k})^n\big)$. Since $n \ge \dim V_k$, the null spaces of the powers of $T|_{V_k} - \lambda I_{V_k}$ have stopped growing by exponent $\dim V_k$ (by 8.3 applied on the space V_k), so

$$\operatorname{null}\big((T|_{V_k} - \lambda I_{V_k})^n\big) = \operatorname{null}\big((T|_{V_k} - \lambda I_{V_k})^{\dim V_k}\big) = G(\lambda, T|_{V_k}),$$

the last equality by 8.20. Hence $v_k \in G(\lambda, T|_{V_k})$ for each k, and v lies in the sum of these subspaces.
Finally, the sum is direct: the subspaces $G(\lambda, T|_{V_k})$ satisfy $G(\lambda, T|_{V_k}) \subseteq V_k$, and $V_1 + \cdots + V_m$ is a direct sum, so any relation $u_1 + \cdots + u_m = 0$ with $u_k \in G(\lambda, T|_{V_k}) \subseteq V_k$ forces every $u_k = 0$. This proves the claim.
Conclusion. Taking dimensions in the claim and using the formula for the dimension of a direct sum (3.94),

$$\dim G(\lambda, T) = \sum_{k=1}^{m} \dim G(\lambda, T|_{V_k})$$

for every $\lambda \in \mathbf{C}$. Therefore

$$\begin{aligned} p_1 \cdots p_m &= \prod_{k=1}^{m} \prod_{\lambda \in \mathbf{C}} (z - \lambda)^{\dim G(\lambda, T|_{V_k})} \\ &= \prod_{\lambda \in \mathbf{C}} (z - \lambda)^{\sum_{k=1}^{m} \dim G(\lambda, T|_{V_k})} \\ &= \prod_{\lambda \in \mathbf{C}} (z - \lambda)^{\dim G(\lambda, T)}, \end{aligned}$$

which is the characteristic polynomial of T. (All products above have only finitely many factors different from 1, so the rearrangement is legitimate.)
As a consistency check on degrees: $\deg p_k = \dim V_k$ by 8.28(a), and $\sum_k \dim V_k = \dim V$ by 3.94, so $\deg(p_1 \cdots p_m) = \dim V$, which is indeed the degree of the characteristic polynomial of T.

Problem 8B.21. Suppose $p, q \in \mathcal{P}(\mathbf{C})$ are monic polynomials with the same zeros and q is a polynomial multiple of p. Prove that there exists $T \in \mathcal{L}(\mathbf{C}^{\deg q})$ such that the characteristic polynomial of T is q and the minimal polynomial of T is p.
[This exercise implies that every monic polynomial is the characteristic polynomial of some operator.]

Solution. Degenerate case. If p and q have no zeros, then by the fundamental theorem of algebra (4.12) both are constant, and being monic they both equal 1. Then $\deg q = 0$, and the only operator on $\mathbf{C}^0 = \{0\}$ is $T = 0$; its characteristic polynomial is the empty product $1 = q$ and its minimal polynomial is the constant polynomial $1 = p$ (see 5.22). So assume from now on that p and q have at least one zero.
Setting up the factorizations. Let $\lambda_1, \dots, \lambda_M$ be the distinct zeros of p, which by hypothesis are also exactly the distinct zeros of q. By the fundamental theorem of algebra, second version (4.13), and because p and q are monic,

$$p(z) = (z - \lambda_1)^{a_1} \cdots (z - \lambda_M)^{a_M}, \qquad q(z) = (z - \lambda_1)^{b_1} \cdots (z - \lambda_M)^{b_M},$$

where a_k, b_k are positive integers. Since q is a polynomial multiple of p, uniqueness of the factorization in 4.13 forces

$$a_k \le b_k \qquad \text{for each } k \in \{1, \dots, M\}.$$

Set $b = b_1 + \cdots + b_M = \deg q$.
Building the blocks. Fix k. On $\mathbf{C}^{b_k}$ define N_k to be the linear map determined on the standard basis $f_1, \dots, f_{b_k}$ of $\mathbf{C}^{b_k}$ by

$$N_k f_j = f_{j+1} \ \text{ for } j < a_k, \qquad N_k f_j = 0 \ \text{ for } j \ge a_k.$$

This makes sense because $a_k \le b_k$, so the index $j+1$ appearing above never exceeds b_k. Thus N_k moves f_1 along the chain $f_1 \mapsto f_2 \mapsto \cdots \mapsto f_{a_k} \mapsto 0$ and kills f_j for $j \ge a_k$. Hence for $j \in \{1, \dots, b_k\}$ we have $N_k^{a_k - j + 1} f_j = 0$ if $j \le a_k$ and $N_k f_j = 0$ if $j \ge a_k$; in either case $N_k^{a_k} f_j = 0$, because $a_k - j + 1 \le a_k$ when $j \ge 1$. Therefore $N_k^{a_k} = 0$. On the other hand $N_k^{a_k - 1} f_1 = f_{a_k} \ne 0$, so $N_k^{a_k - 1} \ne 0$. Thus N_k is nilpotent, and a_k is the smallest exponent killing it.
Now define

$$T_k = \lambda_k I + N_k \in \mathcal{L}(\mathbf{C}^{b_k}).$$

Characteristic and minimal polynomials of T_k. Since $T_k - \lambda_k I = N_k$ is nilpotent, 0 is the only eigenvalue of N_k (by 8.17(a)); consequently λ_k is the only eigenvalue of T_k, because μ is an eigenvalue of T_k if and only if $\mu - \lambda_k$ is an eigenvalue of N_k. Moreover, by 8.20,

$$G(\lambda_k, T_k) = \operatorname{null} N_k^{b_k} = \mathbf{C}^{b_k}$$

since $N_k^{b_k} = 0$ (as $b_k \ge a_k$). Hence the multiplicity of λ_k is b_k, and by 8.26 the characteristic polynomial of T_k is $(z - \lambda_k)^{b_k}$.
For the minimal polynomial of T_k: since $(T_k - \lambda_k I)^{a_k} = N_k^{a_k} = 0$, the minimal polynomial of T_k divides $(z - \lambda_k)^{a_k}$ (by 5.29), so it equals $(z - \lambda_k)^c$ for some integer c with $0 \le c \le a_k$ (the monic divisors of $(z - \lambda_k)^{a_k}$ are exactly these, by 4.13). Here $c \ne 0$, since the constant polynomial 1 does not annihilate the identity operator on the nonzero space $\mathbf{C}^{b_k}$. If $c < a_k$ then $c \le a_k - 1$, so $N_k^{a_k - 1} = N_k^{a_k - 1 - c} N_k^c = 0$, where $N_k^c = (T_k - \lambda_k I)^c = 0$; this contradicts $N_k^{a_k - 1} \ne 0$. Hence $c = a_k$ and the minimal polynomial of T_k is $(z - \lambda_k)^{a_k}$.
Assembling T. Since $b_1 + \cdots + b_M = b$, we may identify

$$\mathbf{C}^b = V_1 \oplus \cdots \oplus V_M,$$

where V_k is the coordinate subspace of $\mathbf{C}^b$ spanned by the standard basis vectors in positions $b_1 + \cdots + b_{k-1} + 1$ through $b_1 + \cdots + b_k$; thus V_k is naturally identified with $\mathbf{C}^{b_k}$. Define $T \in \mathcal{L}(\mathbf{C}^b)$ by

$$T(v_1 + \cdots + v_M) = T_1 v_1 + \cdots + T_M v_M \qquad (v_k \in V_k),$$

using the identification $V_k = \mathbf{C}^{b_k}$. This is well defined and linear because the decomposition of each vector as $v_1 + \cdots + v_M$ is unique. Each V_k is invariant under T, and $T|_{V_k} = T_k$. (In matrix language, T has a block diagonal matrix whose k-th block is the b_k-by-b_k matrix with λ_k on the diagonal, 1 in the first $a_k - 1$ entries just below the diagonal, and 0 elsewhere.)
Characteristic polynomial of T. Each V_k is nonzero (because $b_k \geq 1$) and invariant under T, so Exercise 20 in this section applies: the characteristic polynomial of T is the product of the characteristic polynomials of the $T|_{V_k} = T_k$, namely

$$(z - \lambda_1)^{b_1} \cdots (z - \lambda_M)^{b_M} = q.$$

Minimal polynomial of T. Let $r \in \mathcal{P}(\mathbf{C})$. Since T acts blockwise and each V_k is invariant under T, we have $r(T)|_{V_k} = r(T_k)$ for each k. Because $\mathbf{C}^b$ is the sum of the V_k, it follows that

$$r(T) = 0 \iff r(T_k) = 0 \text{ for every } k \in \{1, \ldots, M\}.$$

By 5.29 applied to each T_k, whose minimal polynomial we computed to be $(z - \lambda_k)^{a_k}$, this is equivalent to

$$(z - \lambda_k)^{a_k} \text{ divides } r \text{ for every } k.$$

Since $\lambda_1, \ldots, \lambda_M$ are distinct, this last condition is equivalent to divisibility by the product. Indeed, if $r = 0$ then both conditions hold, since every polynomial divides 0. And if $r \neq 0$, factor r into linear factors as in 4.13; the condition says that λ_k occurs at least a_k times among those factors, for each k, and as the λ_k are distinct these occurrences are disjoint, so $(z - \lambda_1)^{a_1} \cdots (z - \lambda_M)^{a_M} = p$ divides r. (Conversely if p divides r then certainly each $(z - \lambda_k)^{a_k}$ divides r.) Hence

$$r(T) = 0 \iff p \text{ divides } r.$$

In particular $p(T) = 0$, and no monic polynomial of degree smaller than $\deg p$ annihilates T (a nonzero polynomial divisible by p has degree at least $\deg p$). Since p is monic, p is the minimal polynomial of T (by 5.22).
Therefore $T \in \mathcal{L}(\mathbf{C}^{\deg q})$ has characteristic polynomial q and minimal polynomial p, as required.
Remark on the bracketed comment. Given any monic $q \in \mathcal{P}(\mathbf{C})$, let p be the product of $(z - \lambda)$ over the distinct zeros λ of q. Then p and q are monic with the same zeros and q is a multiple of p (by 4.13), so the result produces an operator on $\mathbf{C}^{\deg q}$ whose characteristic polynomial is q. Thus every monic polynomial in $\mathcal{P}(\mathbf{C})$ is the characteristic polynomial of some operator.

Problem 8B.22. Suppose A and B are block diagonal matrices of the form

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}, \qquad B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{pmatrix},$$

where A_k and B_k are square matrices of the same size for each $k = 1, \ldots, m$. Show that AB is a block diagonal matrix of the form

$$AB = \begin{pmatrix} A_1B_1 & & 0 \\ & \ddots & \\ 0 & & A_mB_m \end{pmatrix}.$$

Solution. Notation. For each $k \in \{1, \ldots, m\}$ let n_k denote the common size of the square matrices A_k and B_k, and put

$$n = n_1 + \cdots + n_m, \qquad N_0 = 0, \qquad N_k = n_1 + \cdots + n_k.$$

Thus A and B are both n-by-n matrices. For each k set

$$I_k = \{N_{k-1} + 1,\ N_{k-1} + 2,\ \ldots,\ N_k\},$$

a set of n_k consecutive integers. The sets $I_1, \dots, I_m$ are pairwise disjoint and their union is $\{1, \dots, n\}$; hence each index in $\{1, \dots, n\}$ lies in exactly one of them.
Unwinding the definition of a block diagonal matrix (8.35), the hypothesis on A says exactly this. If $j \in I_k$ and $l \in I_{k'}$ with $k \neq k'$, then

$$A_{j,l} = 0;$$

if instead $j, l \in I_k$ for one and the same k, then

$$A_{j,l} = (A_k)_{\,j-N_{k-1},\, l-N_{k-1}}.$$

Here the second formula makes sense because $j, l \in I_k$ forces $j - N_{k-1}$ and $l - N_{k-1}$ to lie in $\{1, \dots, n_k\}$, which is the index range for the n_k-by-n_k matrix A_k. The matrix B satisfies the two corresponding statements with A replaced by B throughout.
The computation. Fix $j, l \in \{1, \dots, n\}$, and let k and k' be the unique indices with $j \in I_k$ and $l \in I_{k'}$. By the definition of matrix multiplication,

$$(AB)_{j,l} = \sum_{r=1}^{n} A_{j,r} B_{r,l}$$

(this is 3.47, the formula for an entry of a matrix product).
Case 1: $k \neq k'$. If $r \notin I_k$, then j and r lie in different sets of the partition, so $A_{j,r} = 0$ and the r^{th} summand is 0. If $r \in I_k$, then $r \notin I_{k'}$ (the sets I_k and $I_{k'}$ are disjoint), so r and l lie in different sets of the partition, giving $B_{r,l} = 0$ and again a summand equal to 0. Every summand vanishes, so

$$(AB)_{j,l} = 0.$$

Thus every entry of AB whose row index and column index lie in different sets of the partition $I_1, \dots, I_m$ equals 0. In other words, AB is a block diagonal matrix whose diagonal blocks have sizes $n_1, \dots, n_m$, exactly the sizes of the diagonal blocks of A and of B.
Case 2: $k = k'$, so that $j, l \in I_k$. Again $A_{j,r} = 0$ for every $r \notin I_k$, so only the terms with $r \in I_k$ survive. Writing $r = N_{k-1} + s$ with s running over $\{1, \dots, n_k\}$, we get

$$\begin{aligned}
(AB)_{j,l} &= \sum_{r \in I_k} A_{j,r} B_{r,l} \\
&= \sum_{s=1}^{n_k} A_{\,j,\, N_{k-1}+s}\; B_{\,N_{k-1}+s,\, l} \\
&= \sum_{s=1}^{n_k} (A_k)_{\,j-N_{k-1},\, s}\; (B_k)_{\,s,\, l-N_{k-1}} \\
&= (A_k B_k)_{\,j-N_{k-1},\, l-N_{k-1}},
\end{aligned}$$

where the third equality used the description of the entries of A and of B (legitimate because $j, N_{k-1} + s \in I_k$ and $N_{k-1} + s, l \in I_k$), and the last equality is the definition of matrix multiplication applied to the n_k-by-n_k matrices A_k and B_k. Note that the product $A_k B_k$ is defined and is again n_k-by-n_k precisely because A_k and B_k have the same size, which is the hypothesis.
Conclusion. Case 1 shows that AB is block diagonal with blocks of sizes $n_1, \dots, n_m$, and Case 2 shows that its k^{th} diagonal block, whose entries are the numbers $(AB)_{j,l}$ for $j, l \in I_k$ read in order, is precisely $A_k B_k$. Hence

$$AB = \begin{pmatrix} A_1 B_1 & & 0 \\ & \ddots & \\ 0 & & A_m B_m \end{pmatrix},$$

as desired.
Remark. There is an equivalent operator-theoretic way to see this. Suppose $v_1, \dots, v_n$ is a basis of V and $S, T \in \mathcal{L}(V)$ have matrices A and B with respect to this basis. Let $V_k = \operatorname{span}(v_j : j \in I_k)$, so $V = V_1 \oplus \cdots \oplus V_m$. Saying A and B are block diagonal with these block sizes says exactly that each

V_k is invariant under both S and T, with A_k and B_k the matrices of $S|_{V_k}$ and $T|_{V_k}$ with respect to $v_{N_{k-1}+1}, \dots, v_{N_k}$. Then each V_k is invariant under ST and $(ST)|_{V_k} = S|_{V_k} T|_{V_k}$, so the matrix of ST is block diagonal with k^{th} block equal to the matrix of $S|_{V_k} T|_{V_k}$, which is $A_k B_k$ because the matrix of a product of linear maps is the product of their matrices (3.43). The index computation above is nothing more than this argument written out in coordinates.

Problem 8B.23. Suppose $\mathbf{F} = \mathbf{R}$, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{C}$.
(a) Show that $u + iv \in G(\lambda, T_{\mathbf{C}})$ if and only if $u - iv \in G(\overline{\lambda}, T_{\mathbf{C}})$.
(b) Show that the multiplicity of λ as an eigenvalue of $T_{\mathbf{C}}$ equals the multiplicity of $\overline{\lambda}$ as an eigenvalue of $T_{\mathbf{C}}$.
(c) Use (b) and the result about the sum of the multiplicities (8.25) to show that if $\dim V$ is an odd number, then $T_{\mathbf{C}}$ has a real eigenvalue.
(d) Use (c) and the result about real eigenvalues of $T_{\mathbf{C}}$ (Exercise 17 in Section 5A) to show that if $\dim V$ is an odd number, then T has an eigenvalue (thus giving an alternative proof of 5.34).
[See Exercise 33 in Section 3B for the definition of the complexification $T_{\mathbf{C}}$.]

Solution. Throughout, V is a finite-dimensional real vector space, $V_{\mathbf{C}} = \{u + iv : u, v \in V\}$ is its complexification (Exercise 8 in Section 1B), and $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$ is defined by $T_{\mathbf{C}}(u + iv) = Tu + iTv$ (Exercise 33 in Section 3B). Recall that if $v_1, \dots, v_n$ is a basis of V over $\mathbf{R}$, then $v_1, \dots, v_n$ is also a basis of $V_{\mathbf{C}}$ over $\mathbf{C}$ (Exercise 11 in Section 2B); hence

$$\dim_{\mathbf{C}} V_{\mathbf{C}} = \dim_{\mathbf{R}} V.$$

The conjugation map. Define $\sigma : V_{\mathbf{C}} \to V_{\mathbf{C}}$ by

$$\sigma(u + iv) = u - iv \qquad \text{for all } u, v \in V.$$

(Here $u - iv$ means $u + i(-v)$, so σ is well defined: an element of $V_{\mathbf{C}}$ is by definition the ordered pair (u, v), and $\sigma(u, v) = (u, -v)$.) Three properties of σ drive everything below.
Property 1. σ is additive, and $\sigma \circ \sigma$ is the identity on $V_{\mathbf{C}}$. Indeed

$$\sigma\big((u_1 + iv_1) + (u_2 + iv_2)\big) = \sigma\big((u_1 + u_2) + i(v_1 + v_2)\big) = (u_1 + u_2) - i(v_1 + v_2),$$

which equals $\sigma(u_1 + iv_1) + \sigma(u_2 + iv_2)$; and $\sigma(\sigma(u + iv)) = \sigma(u - iv) = u + iv$. In particular σ is a bijection and $\sigma(0) = 0$.
Property 2. σ is conjugate-homogeneous: $\sigma(\alpha x) = \overline{\alpha}\,\sigma(x)$ for all $\alpha \in \mathbf{C}$ and $x \in V_{\mathbf{C}}$. To verify this, write $\alpha = a + bi$ with $a, b \in \mathbf{R}$ and $x = u + iv$. By the definition of complex scalar multiplication on $V_{\mathbf{C}}$,

$$\alpha x = (au - bv) + i(av + bu), \qquad \text{so} \qquad \sigma(\alpha x) = (au - bv) + i\big(-av - bu\big).$$

On the other hand $\overline{\alpha} = a + (-b)i$ and $\sigma(x) = u + i(-v)$, so

$$\overline{\alpha}\,\sigma(x) = \big(au - (-b)(-v)\big) + i\big(a(-v) + (-b)u\big) = (au - bv) + i\big(-av - bu\big).$$

The two right sides agree, proving the claim.
Property 3. σ commutes with $T_{\mathbf{C}}$: for all $u, v \in V$,

$$\sigma\big(T_{\mathbf{C}}(u + iv)\big) = \sigma(Tu + iTv) = Tu - iTv = T_{\mathbf{C}}(u - iv) = T_{\mathbf{C}}\big(\sigma(u + iv)\big).$$

Combining Properties 1, 2, and 3 gives the key identity: for every $x \in V_{\mathbf{C}}$ and every $\lambda \in \mathbf{C}$,

$$\sigma\big((T_{\mathbf{C}} - \lambda I)x\big) = \sigma(T_{\mathbf{C}} x) - \sigma(\lambda x) = T_{\mathbf{C}}\sigma(x) - \overline{\lambda}\,\sigma(x) = (T_{\mathbf{C}} - \overline{\lambda} I)\sigma(x),$$

where we used additivity of σ together with $\sigma(-y) = -\sigma(y)$ (which follows from additivity and $\sigma(0) = 0$). Iterating this identity, an easy induction on k gives

$$\sigma\big((T_{\mathbf{C}} - \lambda I)^k x\big) = (T_{\mathbf{C}} - \overline{\lambda} I)^k \sigma(x) \qquad \text{for every positive integer } k \text{ and every } x \in V_{\mathbf{C}}.$$

Indeed, the case $k = 1$ is the identity just proved, and if it holds for k then

$$\sigma\big((T_{\mathbf{C}} - \lambda I)^{k+1}x\big) = \sigma\big((T_{\mathbf{C}} - \lambda I)^k\,(T_{\mathbf{C}} - \lambda I)x\big) = (T_{\mathbf{C}} - \overline{\lambda} I)^k \sigma\big((T_{\mathbf{C}} - \lambda I)x\big) = (T_{\mathbf{C}} - \overline{\lambda} I)^{k+1}\sigma(x).$$

Solution to (a). Suppose $u + iv \in G(\lambda, T_{\mathbf{C}})$. By the definition of generalized eigenspace (8.19), there is a positive integer k with $(T_{\mathbf{C}} - \lambda I)^k(u + iv) = 0$. Applying σ and using the displayed identity above together with $\sigma(0) = 0$,

$$(T_{\mathbf{C}} - \overline{\lambda} I)^k(u - iv) = (T_{\mathbf{C}} - \overline{\lambda} I)^k \sigma(u + iv) = \sigma\big((T_{\mathbf{C}} - \lambda I)^k(u + iv)\big) = \sigma(0) = 0.$$

Hence $u - iv \in G(\overline{\lambda}, T_{\mathbf{C}})$.
Conversely, suppose $u - iv \in G(\overline{\lambda}, T_{\mathbf{C}})$. Apply the implication just proved, with λ replaced by $\overline{\lambda}$ and with the pair (u, v) replaced by $(u, -v)$: it gives $u + iv \in G(\overline{\overline{\lambda}}, T_{\mathbf{C}}) = G(\lambda, T_{\mathbf{C}})$. This completes (a).
Equivalently: σ maps $G(\lambda, T_{\mathbf{C}})$ into $G(\overline{\lambda}, T_{\mathbf{C}})$ and maps $G(\overline{\lambda}, T_{\mathbf{C}})$ into $G(\lambda, T_{\mathbf{C}})$; since $\sigma \circ \sigma$ is the identity, σ restricts to a bijection

$$\sigma|_{G(\lambda, T_{\mathbf{C}})} : G(\lambda, T_{\mathbf{C}}) \to G(\overline{\lambda}, T_{\mathbf{C}}).$$

Solution to (b). First note that for any $\mu \in \mathbf{C}$ we have $G(\mu, T_{\mathbf{C}}) \neq \{0\}$ if and only if μ is an eigenvalue of $T_{\mathbf{C}}$: an eigenvector for μ lies in $G(\mu, T_{\mathbf{C}})$, and conversely if $G(\mu, T_{\mathbf{C}}) \neq \{0\}$ then $(T_{\mathbf{C}} - \mu I)^k$ is not injective for some k, which forces $T_{\mathbf{C}} - \mu I$ itself to be non-injective (a composition of injective maps is injective), i.e. μ is an eigenvalue. Combining this with (a), which says σ carries $G(\lambda, T_{\mathbf{C}})$ bijectively onto $G(\overline{\lambda}, T_{\mathbf{C}})$, we see that λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if $\overline{\lambda}$ is. So the two multiplicities in (b) are simultaneously defined (or, adopting the harmless convention that a non-eigenvalue has multiplicity $0 = \dim\{0\}$, both are 0).
Now assume λ is an eigenvalue and let $d = \dim G(\lambda, T_{\mathbf{C}})$ be its multiplicity (8.23). Let $x_1, \ldots, x_d$ be a basis of $G(\lambda, T_{\mathbf{C}})$. We claim $\sigma(x_1), \ldots, \sigma(x_d)$ is a basis of $G(\overline{\lambda}, T_{\mathbf{C}})$; by (a) these vectors do lie in $G(\overline{\lambda}, T_{\mathbf{C}})$.
Linear independence: suppose $c_1, \ldots, c_d \in \mathbf{C}$ and $c_1\sigma(x_1) + \cdots + c_d\sigma(x_d) = 0$. By additivity and conjugate-homogeneity of σ,

$$\sigma\big(\overline{c_1}x_1 + \cdots + \overline{c_d}x_d\big) = c_1\sigma(x_1) + \cdots + c_d\sigma(x_d) = 0.$$

Since σ is injective and $\sigma(0) = 0$, we get $\overline{c_1}x_1 + \cdots + \overline{c_d}x_d = 0$, so each $\overline{c_j} = 0$ by linear independence of $x_1, \ldots, x_d$, so each $c_j = 0$.
Spanning: suppose $y \in G(\overline{\lambda}, T_{\mathbf{C}})$. By (a) we have $\sigma(y) \in G(\lambda, T_{\mathbf{C}})$, so $\sigma(y) = a_1x_1 + \cdots + a_dx_d$ for some $a_1, \ldots, a_d \in \mathbf{C}$. Applying σ and using $\sigma \circ \sigma = I$ again,

$$y = \sigma(\sigma(y)) = \overline{a_1}\,\sigma(x_1) + \cdots + \overline{a_d}\,\sigma(x_d).$$

Hence $\sigma(x_1), \ldots, \sigma(x_d)$ is a basis of $G(\overline{\lambda}, T_{\mathbf{C}})$, and therefore

$$\text{multiplicity of } \overline{\lambda} = \dim G(\overline{\lambda}, T_{\mathbf{C}}) = d = \dim G(\lambda, T_{\mathbf{C}}) = \text{multiplicity of } \lambda.$$

Solution to (c). Let $n = \dim V$ and suppose n is odd; in particular $n \geq 1$, so $V_{\mathbf{C}} \neq \{0\}$, and $\dim_{\mathbf{C}} V_{\mathbf{C}} = n$ as noted at the start. Since $V_{\mathbf{C}}$ is a nonzero finite-dimensional complex vector space and $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$, the result on the sum of the multiplicities (8.25) applies: if $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of $T_{\mathbf{C}}$ and $d_1, \ldots, d_m$ are their multiplicities, then

$$d_1 + \cdots + d_m = n.$$

Split $\{\lambda_1, \ldots, \lambda_m\}$ into the set R of real eigenvalues and the set N of non-real eigenvalues, and write S_R and S_N for the sums of the corresponding multiplicities, so $S_R + S_N = n$.
By (b), complex conjugation maps the set of eigenvalues of $T_{\mathbf{C}}$ into itself, and it clearly maps non-real numbers to non-real numbers; since conjugation is its own inverse, it is a bijection of N onto N with no fixed points (a fixed point of conjugation is a real number). Consequently N is partitioned into

two-element subsets $\{\mu, \overline{\mu}\}$. By (b) the two eigenvalues in each such pair have equal multiplicity, say d, so each pair contributes $2d$ to S_N. Summing over the pairs shows S_N is even.
Therefore

$$S_R = n - S_N$$

is odd, since n is odd and S_N is even. In particular $S_R \neq 0$, so the set R is nonempty (an empty sum is 0). That is, $T_{\mathbf{C}}$ has a real eigenvalue.
Solution to (d). Suppose $\dim V$ is odd. By (c), there exists $\lambda \in \mathbf{R}$ that is an eigenvalue of $T_{\mathbf{C}}$. Exercise 17 in Section 5A states that for $\mathbf{F} = \mathbf{R}$, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{R}$, the number λ is an eigenvalue of T if and only if λ is an eigenvalue of $T_{\mathbf{C}}$. Applying that equivalence in the direction from $T_{\mathbf{C}}$ to T, we conclude that λ is an eigenvalue of T.
Thus every operator on an odd-dimensional real vector space has an eigenvalue. Together with the fact that every operator on a nonzero finite-dimensional complex vector space has an eigenvalue (5.19), this gives an alternative proof of 5.34.

8.3 Exercises 8C

Problem 8C.1. Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is the operator defined by $T(z_1, z_2, z_3) = (z_2, z_3, 0)$. Prove that T does not have a square root.

Solution. First we record two facts about T.
Computing directly,

$$T^2(z_1, z_2, z_3) = T(z_2, z_3, 0) = (z_3, 0, 0), \qquad T^3(z_1, z_2, z_3) = T(z_3, 0, 0) = (0, 0, 0).$$

Thus $T^3 = 0$, so T is nilpotent; and $T^2 \neq 0$, because $T^2(0, 0, 1) = (1, 0, 0) \neq 0$.
Now suppose, seeking a contradiction, that $R \in \mathcal{L}(\mathbf{C}^3)$ satisfies $R^2 = T$. Then

$$R^6 = (R^2)^3 = T^3 = 0,$$

so R is nilpotent. Because $\dim \mathbf{C}^3 = 3$, the result on nilpotent operators raised to the dimension of the domain (8.16) gives $R^3 = 0$. Hence

$$T^2 = (R^2)^2 = R^4 = R^3 R = 0 \cdot R = 0,$$

contradicting $T^2 \neq 0$.
Therefore no such R exists: T does not have a square root.
[Remark: the same argument shows more generally that if $N \in \mathcal{L}(V)$ is nilpotent with $N^{\dim V - 1} \neq 0$ and $\dim V \geq 2$, then N has no square root. Indeed a square root R of N would be nilpotent, hence satisfy $R^{\dim V} = 0$, so $N^{\lceil \dim V/2 \rceil} = R^{2\lceil \dim V/2 \rceil} = 0$, and $\lceil \dim V/2 \rceil \leq \dim V - 1$ when $\dim V \geq 2$.]

Problem 8C.2. Define $T \in \mathcal{L}(\mathbf{F}^5)$ by $T(x_1, x_2, x_3, x_4, x_5) = (2x_2, 3x_3, -x_4, 4x_5, 0)$.
(a) Show that T is nilpotent.
(b) Find a square root of $I + T$.

Solution. (a) Let $e_1, \ldots, e_5$ denote the standard basis of $\mathbf{F}^5$. From the formula for T we read off

$$Te_1 = 0, \quad Te_2 = 2e_1, \quad Te_3 = 3e_2, \quad Te_4 = -e_3, \quad Te_5 = 4e_4.$$

Thus T maps $\operatorname{span}(e_1, \ldots, e_k)$ into $\operatorname{span}(e_1, \ldots, e_{k-1})$ for each k, where span of the empty list is $\{0\}$. Applying this five times, T^5 maps $\mathbf{F}^5 = \operatorname{span}(e_1, \ldots, e_5)$ into $\{0\}$. Hence $T^5 = 0$ and T is nilpotent.

Explicitly, iterating the formula for T gives

$$\begin{aligned}
T^2(x_1,\dots,x_5) &= (6x_3,-3x_4,-4x_5,0,0),\\
T^3(x_1,\dots,x_5) &= (-6x_4,-12x_5,0,0,0),\\
T^4(x_1,\dots,x_5) &= (-24x_5,0,0,0,0),\\
T^5(x_1,\dots,x_5) &= (0,0,0,0,0).
\end{aligned}$$

(b) Following the proof of 8.39, we look for a square root of the form

$$R = I + a_1T + a_2T^2 + a_3T^3 + a_4T^4,$$

where the a_k are the coefficients of the Taylor series of $\sqrt{1+x}$. Because $T^5 = 0$, only the terms of degree at most 4 can survive when we square, so it suffices to choose a_1, a_2, a_3, a_4 making the coefficients of T, T^2, T^3, T^4 in R^2 equal to $1, 0, 0, 0$ respectively. Squaring and collecting terms (using $T^k = 0$ for $k \geq 5$),

$$R^2 = I + 2a_1T + (2a_2 + a_1^2)T^2 + (2a_3 + 2a_1a_2)T^3 + (2a_4 + 2a_1a_3 + a_2^2)T^4.$$

Solving successively:

$$2a_1 = 1 \implies a_1 = \tfrac{1}{2}; \qquad 2a_2 + \tfrac{1}{4} = 0 \implies a_2 = -\tfrac{1}{8};$$

$$2a_3 + 2\cdot\tfrac{1}{2}\cdot\left(-\tfrac{1}{8}\right) = 0 \implies a_3 = \tfrac{1}{16}; \qquad 2a_4 + 2\cdot\tfrac{1}{2}\cdot\tfrac{1}{16} + \tfrac{1}{64} = 0 \implies a_4 = -\tfrac{5}{128}.$$

Hence

$$R = I + \tfrac{1}{2}T - \tfrac{1}{8}T^2 + \tfrac{1}{16}T^3 - \tfrac{5}{128}T^4$$

satisfies $R^2 = I + T$, by the displayed computation of R^2 together with the choices above.
Using the explicit formulas for the powers of T found in part (a), this operator is

$$R(x_1,x_2,x_3,x_4,x_5) = \Big(x_1 + x_2 - \tfrac{3}{4}x_3 - \tfrac{3}{8}x_4 + \tfrac{15}{16}x_5,\ x_2 + \tfrac{3}{2}x_3 + \tfrac{3}{8}x_4 - \tfrac{3}{4}x_5,\ x_3 - \tfrac{1}{2}x_4 + \tfrac{1}{2}x_5,\ x_4 + 2x_5,\ x_5\Big).$$

Equivalently, the matrix of R with respect to the standard basis is

$$\begin{pmatrix}
1 & 1 & -\frac{3}{4} & -\frac{3}{8} & \frac{15}{16}\\
0 & 1 & \frac{3}{2} & \frac{3}{8} & -\frac{3}{4}\\
0 & 0 & 1 & -\frac{1}{2} & \frac{1}{2}\\
0 & 0 & 0 & 1 & 2\\
0 & 0 & 0 & 0 & 1
\end{pmatrix},$$

whose square is the matrix of $I + T$, namely the matrix with 1's on the diagonal and $2, 3, -1, 4$ on the line directly above the diagonal.
Note that this works over $\mathbf{F} = \mathbf{R}$ as well as over $\mathbf{F} = \mathbf{C}$, since all the coefficients a_k are real.

Problem 8C.3. Suppose V is a complex vector space. Prove that every invertible operator on V has a cube root.

Solution. As always in this chapter, V is finite-dimensional and nonzero. We imitate the proofs of 8.39 and 8.41, replacing squares by cubes.
Step 1: if $N \in \mathcal{L}(V)$ is nilpotent, then $I + N$ has a cube root.
Let s be a positive integer with $N^s = 0$. If $s = 1$ then $N = 0$ and I is a cube root of $I + N$, so assume $s \geq 2$. Set $b_0 = 1$ and consider a polynomial

$$p(x) = b_0 + b_1x + b_2x^2 + \cdots + b_{s-1}x^{s-1}$$

with coefficients yet to be chosen. For $0 \le k \le s-1$, the coefficient of x^k in $p(x)^3$ equals

$$c_k = \sum_{\substack{i+j+\ell=k \\ i,j,\ell \ge 0}} b_i b_j b_\ell.$$

In this sum, the triples in which some index equals k are exactly the three triples $(k,0,0)$, $(0,k,0)$, $(0,0,k)$ (when $k \ge 1$), each contributing $b_k b_0^2 = b_k$; every other triple has all three indices at most $k-1$. Hence for $k \ge 1$ we may write

$$c_k = 3b_k + q_k(b_1, \dots, b_{k-1}),$$

where q_k is a specific polynomial expression in $b_1, \dots, b_{k-1}$ (with $q_1 = 0$). Also $c_0 = b_0^3 = 1$.
Now choose the coefficients recursively: let $b_1 = \frac{1}{3}$, so that $c_1 = 3b_1 = 1$; and for $k = 2, \dots, s-1$ let

$$b_k = -\tfrac{1}{3}\, q_k(b_1, \dots, b_{k-1}),$$

so that $c_k = 0$. (These are of course the Taylor coefficients of $(1+x)^{1/3}$, but we need only that they exist, and the recursion above defines them unambiguously.) With this choice,

$$p(x)^3 = 1 + x + (\text{terms of degree at least } s).$$

Let $R = p(N) = I + b_1 N + \cdots + b_{s-1} N^{s-1}$. Substituting N for x in the identity above and using $N^k = 0$ for all $k \ge s$, we get $R^3 = I + N$. So $I + N$ has a cube root.
Step 2: every complex number has a cube root.
Given $\lambda \in \mathbf{C}$, write $\lambda = r(\cos\theta + i\sin\theta)$ with $r \ge 0$. Then, by the addition formulas for sine and cosine (applied twice),

$$\Big(r^{1/3}\big(\cos\tfrac{\theta}{3} + i\sin\tfrac{\theta}{3}\big)\Big)^3 = r(\cos\theta + i\sin\theta) = \lambda.$$

Step 3: the general case.
Suppose $T \in \mathcal{L}(V)$ is invertible. Because V is a nonzero finite-dimensional complex vector space, T has at least one eigenvalue (5.19); let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. By the generalized eigenspace decomposition (8.22),

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

each $G(\lambda_k, T)$ is invariant under T, and there is a nilpotent $N_k \in \mathcal{L}\big(G(\lambda_k, T)\big)$ with

$$T|_{G(\lambda_k, T)} = \lambda_k I + N_k.$$

Because T is invertible, 0 is not an eigenvalue of T, so $\lambda_k \ne 0$ for each k. Thus

$$T|_{G(\lambda_k, T)} = \lambda_k\Big(I + \frac{N_k}{\lambda_k}\Big).$$

The operator N_k/λ_k is nilpotent, so by Step 1 there exists $S_k \in \mathcal{L}\big(G(\lambda_k, T)\big)$ with $S_k^3 = I + N_k/\lambda_k$. By Step 2 choose $c_k \in \mathbf{C}$ with $c_k^3 = \lambda_k$, and set $R_k = c_k S_k \in \mathcal{L}\big(G(\lambda_k, T)\big)$. Then

$$R_k^3 = c_k^3 S_k^3 = \lambda_k\Big(I + \frac{N_k}{\lambda_k}\Big) = T|_{G(\lambda_k, T)}.$$

Finally, each $v \in V$ can be written uniquely as $v = u_1 + \cdots + u_m$ with $u_k \in G(\lambda_k, T)$; define $R \in \mathcal{L}(V)$ by

$$Rv = R_1 u_1 + \cdots + R_m u_m.$$

This map is linear, because each of the maps $v \mapsto u_k$ is linear (uniqueness of the decomposition in a direct sum) and each R_k is linear. Moreover $R_k u_k \in G(\lambda_k, T)$, so $R_1 u_1 + \cdots + R_m u_m$ is exactly the decomposition of Rv into its generalized-eigenspace components. Applying R three times therefore gives

$$R^3 v = R_1^3 u_1 + \cdots + R_m^3 u_m = T u_1 + \cdots + T u_m = T(u_1 + \cdots + u_m) = Tv.$$

Hence $R^3 = T$, so T has a cube root.
[The same argument, with 3 replaced by any positive integer n in Steps 1 and 2, shows that every invertible operator on a complex vector space has an nth root, as remarked after 8.41.]

Problem 8C.4. Suppose V is a real vector space. Prove that the operator $-I$ on V has a square root if and only if $\dim V$ is an even number.

Solution. First suppose $-I$ has a square root, so there exists $R \in \mathcal{L}(V)$ with $R^2 = -I$.
We claim R has no eigenvalues. Indeed, suppose $\lambda \in \mathbf{R}$ and $v \in V$ with $v \neq 0$ and $Rv = \lambda v$. Applying R again gives

$$R^2 v = \lambda^2 v.$$

But $R^2 = -I$, so $R^2 v = -v$. Hence $\lambda^2 v = -v$, and since $v \neq 0$ this forces $\lambda^2 = -1$. No real number has square -1, so no such λ exists; thus R has no eigenvalues.
Now if $\dim V$ were an odd number, then R would have an eigenvalue, because every operator on an odd-dimensional real vector space has an eigenvalue (5.34). This contradicts the previous paragraph. Hence $\dim V$ is an even number.
Conversely, suppose $\dim V$ is an even number, say $\dim V = 2n$. Because V is nonzero (a standing assumption of this chapter), $n \geq 1$. Let $v_1, \ldots, v_{2n}$ be a basis of V. Define $R \in \mathcal{L}(V)$ by specifying its values on this basis (3.4):

$$Rv_{2k-1} = v_{2k}, \qquad Rv_{2k} = -v_{2k-1} \qquad \text{for } k = 1, \ldots, n.$$

Then for each $k \in \{1, \ldots, n\}$,

$$R^2 v_{2k-1} = Rv_{2k} = -v_{2k-1}, \qquad R^2 v_{2k} = R(-v_{2k-1}) = -v_{2k}.$$

Thus R^2 agrees with $-I$ on a basis of V, and hence $R^2 = -I$. So $-I$ has a square root.
[If one drops the convention that $V \neq \{0\}$, the case $\dim V = 0$ also causes no trouble: the only operator on $\{0\}$ is 0, and it equals $-I$, hence is its own square root.]
[In matrix language, the second half says that $-I$ has the square root whose matrix with respect to $v_1, \ldots, v_{2n}$ is block diagonal with n blocks, each equal to $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$: rotation by 90° in each coordinate plane, squaring to rotation by 180°.]

Problem 8C.5. Suppose $T \in \mathcal{L}(\mathbf{C}^2)$ is the operator defined by $T(w, z) = (-w - z,\ 9w + 5z)$. Find a Jordan basis for T.

Solution. First we find the eigenvalues. Suppose $\lambda \in \mathbf{C}$ and $(w, z) \neq (0, 0)$ satisfy $T(w, z) = \lambda(w, z)$, that is,

$$-w - z = \lambda w \quad \text{and} \quad 9w + 5z = \lambda z.$$

If $w = 0$ then the first equation gives $z = 0$, contradiction; so $w \neq 0$, and the first equation gives $z = -(\lambda + 1)w$. Substituting into the second equation and dividing by w:

$$9 - 5(\lambda + 1) = -\lambda(\lambda + 1),$$

which simplifies to $\lambda^2 - 4\lambda + 4 = 0$, that is, $(\lambda - 2)^2 = 0$. Hence 2 is the only eigenvalue of T, and the corresponding eigenvectors are the nonzero multiples of $(1, -3)$ (take $\lambda = 2$, $w = 1$, $z = -3$).
A Jordan matrix is upper triangular, so its diagonal entries are the eigenvalues of T (5.41); hence every diagonal entry of a Jordan matrix for T equals 2. Moreover the eigenspace $E(2, T) = \operatorname{span}\big((1, -3)\big)$ has dimension $1 < 2 = \dim \mathbf{C}^2$, so there is no basis of $\mathbf{C}^2$ consisting of eigenvectors of T; thus the Jordan matrix cannot be diagonal, ruling out two blocks of size 1. So a Jordan basis for T must give a single 2-by-2 block with 2 on the diagonal; that is, we need vectors v_1, v_2 forming a basis with

$$Tv_1 = 2v_1 \quad \text{and} \quad Tv_2 = v_1 + 2v_2.$$

Note that

$$(T - 2I)(w, z) = (-3w - z,\ 9w + 3z).$$

Take $v_2 = (1, 0)$; then

$$v_1 := (T - 2I)v_2 = (-3, 9) \neq 0.$$

Then v_1, v_2 is the desired list. Let us verify everything directly:

$$Tv_1 = T(-3, 9) = \big(3 - 9,\ -27 + 45\big) = (-6, 18) = 2(-3, 9) = 2v_1,$$

$$Tv_2 = T(1, 0) = (-1, 9) = (-3, 9) + 2(1, 0) = v_1 + 2v_2.$$

Also v_1, v_2 is linearly independent: if $a(-3, 9) + b(1, 0) = (0, 0)$, the second coordinate gives $9a = 0$, so $a = 0$, and then $b = 0$. Being a linearly independent list of length 2 in the 2-dimensional space $\mathbf{C}^2$, the list v_1, v_2 is a basis (2.38).
With respect to the basis

$$v_1 = (-3, 9), \qquad v_2 = (1, 0),$$

the matrix of T is

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix},$$

which has the required Jordan form (a single block, with the eigenvalue 2 on the diagonal and 1 directly above the diagonal). Hence $(-3, 9), (1, 0)$ is a Jordan basis for T.
[Any $v_2 \notin E(2, T)$ works, with $v_1 = (T - 2I)v_2$; for instance $v_2 = \big(-\frac{1}{3}, 0\big)$ gives the Jordan basis $(1, -3), \big(-\frac{1}{3}, 0\big)$.]

Problem 8C.6. Find a basis of $\mathcal{P}_4(\mathbf{R})$ that is a Jordan basis for the differentiation operator D on $\mathcal{P}_4(\mathbf{R})$ defined by $Dp = p'$.

Solution. The operator D is nilpotent: the fifth derivative of every polynomial of degree at most 4 is 0, so $D^5 = 0$ [see 8.15(c)]. Because D is nilpotent, its only eigenvalue is 0 (8.17); a Jordan matrix is upper triangular, so its diagonal entries are the eigenvalues (5.41), and hence every diagonal entry in a Jordan matrix for D must be 0. Also $D^4 \neq 0$, since $D^4(x^4) = 24 \neq 0$. If every Jordan block had size at most 4, then D^4 would kill every vector of the Jordan basis (a nilpotent Jordan block of size s shifts the corresponding basis vectors down by one and so is killed by the sth power of the operator), forcing $D^4 = 0$. Hence some block has size at least 5; since $\dim \mathcal{P}_4(\mathbf{R}) = 5$, there is exactly one block, of size 5.
Concretely, following the pattern of 8.42, we take a vector u with $D^4u \neq 0$ and use the list D^4u, D^3u, D^2u, Du, u; we then verify directly that this list is a basis and that the matrix of D with respect to it is a single Jordan block. Taking $u = x^4/24$ gives the list

$$p_1 = 1, \qquad p_2 = x, \qquad p_3 = \frac{x^2}{2}, \qquad p_4 = \frac{x^3}{6}, \qquad p_5 = \frac{x^4}{24},$$

that is, $p_k = x^{k-1}/(k-1)!$ for $k = 1, \ldots, 5$.
This list is a basis of $\mathcal{P}_4(\mathbf{R})$: its entries are nonzero scalar multiples of $1, x, x^2, x^3, x^4$, which is the standard basis of $\mathcal{P}_4(\mathbf{R})$; scaling the vectors of a basis by nonzero scalars again yields a basis (or: the list has 5 polynomials of distinct degrees $0, 1, 2, 3, 4$, hence is linearly independent, and a linearly independent list of length $5 = \dim \mathcal{P}_4(\mathbf{R})$ is a basis by 2.38).
Now compute the action of D on this list:

$$Dp_1 = 0, \qquad Dp_2 = 1 = p_1, \qquad Dp_3 = x = p_2, \qquad Dp_4 = \frac{x^2}{2} = p_3, \qquad Dp_5 = \frac{x^3}{6} = p_4.$$

Therefore the matrix of D with respect to $p_1, \ldots, p_5$ is

$$\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

which is a single Jordan block with eigenvalue 0 on the diagonal and 1's on the line directly above the diagonal. Hence

$$1,\ x,\ \frac{x^2}{2},\ \frac{x^3}{6},\ \frac{x^4}{24}$$

is a Jordan basis for D.
[Any list of the form D^4u, D^3u, D^2u, Du, u with $\deg u = 4$ works; for example $24, 24x, 12x^2, 4x^3, x^4$, obtained from $u = x^4$.]

Problem 8C.7. Suppose $T \in \mathcal{L}(V)$ is nilpotent and $v_1, \dots, v_n$ is a Jordan basis for T. Prove that the minimal polynomial of T is z^{m+1}, where m is the length of the longest consecutive string of 1's that appears on the line directly above the diagonal in the matrix of T with respect to $v_1, \dots, v_n$.

Solution. Setting up notation. By the definition of a Jordan basis (8.44), the matrix of T with respect to $v_1, \dots, v_n$ is block diagonal with blocks $A_1, \dots, A_p$, where A_k is a d_k-by-d_k matrix having some scalar λ_k on its diagonal, 1's on the line directly above its diagonal, and 0's elsewhere; here $d_1 + \cdots + d_p = n$. This matrix is upper triangular, so by 5.41 the eigenvalues of T are precisely the diagonal entries $\lambda_1, \dots, \lambda_p$. Since T is nilpotent, 0 is the only eigenvalue of T (8.17), so

$$\lambda_1 = \cdots = \lambda_p = 0.$$

It is convenient to rename the basis by blocks: for $k = 1, \dots, p$ and $i = 1, \dots, d_k$, put

$$u_{k,i} = v_{d_1 + \cdots + d_{k-1} + i}$$

(so $u_{k,1}, \dots, u_{k,d_k}$ are the basis vectors belonging to the kth block). Reading off the columns of the matrix, and using $\lambda_k = 0$, we get

$$Tu_{k,1} = 0 \qquad \text{and} \qquad Tu_{k,i} = u_{k,i-1} \ \text{ for } 2 \le i \le d_k.$$

An easy induction on j then gives, for all k and i,

$$T^j u_{k,i} = \begin{cases} u_{k,\,i-j} & \text{if } 0 \le j \le i-1, \\ 0 & \text{if } j \ge i. \end{cases}$$

Step 1: identifying m. Let $d = \max\{d_1, \dots, d_p\}$. We claim $m = d - 1$.
The entries directly above the diagonal of the full n-by-n matrix are the entries in positions $(i, i+1)$ for $i = 1, \dots, n-1$. If i and $i+1$ belong to the same block A_k, then this entry lies directly above the diagonal of A_k, so it equals 1; there are exactly $d_k - 1$ such i for each k, and they are consecutive. If i is the last index of block A_k and $i+1$ is the first index of block A_{k+1} (this happens once for each $k < p$), then the entry in position $(i, i+1)$ lies outside all the diagonal blocks, so it equals 0.
Hence, reading the line directly above the diagonal from left to right, we see

$$\underbrace{1, \dots, 1}_{d_1 - 1},\ 0,\ \underbrace{1, \dots, 1}_{d_2 - 1},\ 0,\ \dots,\ 0,\ \underbrace{1, \dots, 1}_{d_p - 1}.$$

The consecutive strings of 1's are therefore exactly the strings coming from the individual blocks, of lengths $d_1 - 1, \dots, d_p - 1$. Thus

$$m = \max\{d_1 - 1, \dots, d_p - 1\} = d - 1, \qquad \text{so } d = m + 1.$$

(If every d_k equals 1, then no 1's appear at all and $m = 0 = d - 1$, consistent with the formula.)
Step 2: $T^d = 0$ and $T^{d-1} \ne 0$. Since $d \ge d_k \ge i$ for every basis vector $u_{k,i}$, the displayed formula for $T^j u_{k,i}$ gives $T^d u_{k,i} = 0$ for every k and i. As T^d vanishes on a basis of V, we conclude $T^d = 0$.
Choose k with $d_k = d$. If $d \ge 2$, then $T^{d-1} u_{k,d} = u_{k,1} \ne 0$ (basis vectors are nonzero), so $T^{d-1} \ne 0$. If $d = 1$, then $T^{d-1} = T^0 = I \ne 0$ because $V \ne \{0\}$. Either way $T^{d-1} \ne 0$.

Step 3: the minimal polynomial. Let q denote the minimal polynomial of T, that is, the unique monic polynomial of smallest degree with $q(T) = 0$ (5.24). Write $e = \deg q$ and

$$q(z) = c_0 + c_1 z + \cdots + c_{e-1} z^{e-1} + z^e.$$

Since z^d is a monic polynomial with $T^d = 0$ by Step 2, minimality of the degree gives $e \le d$.
Suppose toward a contradiction that $e < d$. Again choose k with $d_k = d$ and apply $q(T)$ to $u_{k,d}$. Setting $c_e = 1$, and using $d - i \ge d - e \ge 1$ for each $i \in \{0, \ldots, e\}$, the formula for $T^j u_{k,i}$ gives

$$0 = q(T)u_{k,d} = \sum_{i=0}^{e} c_i \, T^i u_{k,d} = \sum_{i=0}^{e} c_i \, u_{k,\, d-i}.$$

The vectors $u_{k,d}, u_{k,d-1}, \ldots, u_{k,d-e}$ are distinct members of the basis $v_1, \ldots, v_n$, hence are linearly independent. Therefore $c_i = 0$ for every $i \in \{0, \ldots, e\}$; in particular $c_e = 0$, contradicting $c_e = 1$.
Hence $e = d$. Now z^d is a monic polynomial of degree $d = e$ that annihilates T, and by 5.24 the minimal polynomial is the unique monic polynomial of that (smallest) degree annihilating T. Therefore

$$q(z) = z^d = z^{m+1},$$

where the last equality uses $d = m + 1$ from Step 1. This completes the proof.

Problem 8C.8. Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V that is a Jordan basis for T. Describe the matrix of T^2 with respect to this basis.

Solution. By the definition of a Jordan basis (8.44), the matrix of T with respect to $v_1, \ldots, v_n$ is block diagonal,

$$\mathcal{M}(T) = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

where A_k is a d_k-by-d_k matrix with some scalar λ_k on the diagonal, 1 on the line directly above the diagonal, and 0 everywhere else; here $d_1 + \cdots + d_p = n$.
By 3.81 (the matrix of a product of linear maps), $\mathcal{M}(T^2) = \mathcal{M}(T)\,\mathcal{M}(T) = \mathcal{M}(T)^2$. Squaring a block diagonal matrix squares each block, so

$$\mathcal{M}(T^2) = \begin{pmatrix} A_1^2 & & 0 \\ & \ddots & \\ 0 & & A_p^2 \end{pmatrix}.$$

It remains to compute A_k^2. Write $A_k = \lambda_k I + N_k$, where I is the d_k-by-d_k identity matrix and N_k has 1 on the line directly above the diagonal and 0 elsewhere. A direct computation shows that for each positive integer j, the matrix N_k^j has 1 in every entry on the j-th line above the diagonal and 0 in all other entries (and $N_k^j = 0$ if $j \ge d_k$). Because I commutes with N_k,

$$A_k^2 = (\lambda_k I + N_k)^2 = \lambda_k^2 I + 2\lambda_k N_k + N_k^2.$$

Thus A_k^2 is the upper-triangular matrix with λ_k^2 on the diagonal, $2\lambda_k$ on the line directly above the diagonal, 1 on the second line above the diagonal, and 0 in every other entry. For example, if $d_k = 5$ then

$$A_k^2 = \begin{pmatrix} \lambda_k^2 & 2\lambda_k & 1 & 0 & 0 \\ 0 & \lambda_k^2 & 2\lambda_k & 1 & 0 \\ 0 & 0 & \lambda_k^2 & 2\lambda_k & 1 \\ 0 & 0 & 0 & \lambda_k^2 & 2\lambda_k \\ 0 & 0 & 0 & 0 & \lambda_k^2 \end{pmatrix}.$$

(If $d_k = 1$ the block is just (λ_k^2); if $d_k = 2$ it is the 2-by-2 matrix with λ_k^2 on the diagonal and $2\lambda_k$ in the upper-right corner.)
Equivalently, in terms of the basis vectors themselves: if $u_1, \dots, u_{d_k}$ are the basis vectors belonging to the k-th block, so that $Tu_j = \lambda_k u_j + u_{j-1}$ with the convention $u_0 = 0$, then

$$T^2 u_j = \lambda_k^2 u_j + 2\lambda_k u_{j-1} + u_{j-2},$$

with the convention $u_{-1} = u_0 = 0$. This is exactly the column description of A_k^2 given above.
So the matrix of T^2 is block diagonal with the same block sizes as the matrix of T, and each block is upper triangular with λ_k^2 on the diagonal, $2\lambda_k$ on the first superdiagonal, and 1 on the second superdiagonal.
Two remarks. First, $\mathcal{M}(T^2)$ is upper triangular with diagonal entries λ_k^2, so by 5.41 the eigenvalues of T^2 are exactly the squares of the eigenvalues of T, as expected. Second, $v_1, \dots, v_n$ is generally not a Jordan basis for T^2: for instance, if $\lambda_k = 0$ and $d_k = 3$, then A_k^2 has 1 in its upper-right corner and 0 everywhere else, which is not of the form required in 8.44.

Problem 8C.9. Suppose $T \in \mathcal{L}(V)$ is nilpotent. Explain why there exist $v_1, \dots, v_n \in V$ and nonnegative integers $m_1, \dots, m_n$ such that (a) and (b) below both hold.
- (a) $T^{m_1} v_1, \dots, Tv_1, v_1, \dots, T^{m_n} v_n, \dots, Tv_n, v_n$ is a basis of V.
- (b) $T^{m_1+1} v_1 = \cdots = T^{m_n+1} v_n = 0$.

Solution. Because T is nilpotent, 8.45 tells us that there is a basis of V that is a Jordan basis for T. Thus by 8.44 the matrix of T with respect to this basis is block diagonal,

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

where A_k is d_k-by-d_k with some scalar λ_k on the diagonal, 1 on the line directly above the diagonal, and 0 elsewhere.
First we show that every λ_k equals 0. The block diagonal matrix above is upper triangular, so by 5.41 its diagonal entries are precisely the eigenvalues of T; in particular each λ_k is an eigenvalue of T. But the only eigenvalue of a nilpotent operator is 0: if $Tv = \lambda v$ with $v \neq 0$ and $T^M = 0$, then $0 = T^M v = \lambda^M v$, forcing $\lambda^M = 0$ and hence $\lambda = 0$. Thus $\lambda_1 = \cdots = \lambda_p = 0$, and each A_k has 0 on its diagonal and 1 on the line directly above its diagonal.
Now name the basis vectors block by block: write the Jordan basis as

$$u_1^1, \dots, u_{d_1}^1,\ u_1^2, \dots, u_{d_2}^2,\ \dots,\ u_1^p, \dots, u_{d_p}^p,$$

where $u_1^k, \dots, u_{d_k}^k$ are the d_k consecutive basis vectors corresponding to the block A_k. The column of the matrix belonging to u_j^k lists the coefficients of Tu_j^k in terms of the basis. Block diagonality says that all entries of this column outside the rows of the k-th block are 0, and inside the block the column is the j-th column of A_k. Since A_k has 0 on the diagonal and 1 directly above the diagonal, the first column of A_k is 0 and, for $j \geq 2$, the j-th column of A_k has a 1 in row $j-1$ and 0 elsewhere. Hence

$$Tu_1^k = 0 \qquad \text{and} \qquad Tu_j^k = u_{j-1}^k \quad \text{for } j = 2, \dots, d_k.$$

Now set $n = p$ and, for $k = 1, \dots, n$, define

$$v_k = u_{d_k}^k, \qquad m_k = d_k - 1.$$

Each m_k is a nonnegative integer because $d_k \geq 1$.
Applying T repeatedly to v_k and using the displayed formulas gives, by induction on i,

$$T^i v_k = u_{d_k - i}^k \qquad \text{for } i = 0, 1, \dots, d_k - 1,$$

and then

$$T^{m_k+1}v_k = T^{d_k}v_k = T\big(T^{d_k-1}v_k\big) = Tu_1^k = 0.$$

This proves (b).
For (a), the formula $T^i v_k = u^k_{d_k-i}$ shows that

$$T^{m_k}v_k, \dots, Tv_k, v_k \;=\; u_1^k, u_2^k, \dots, u_{d_k}^k,$$

which is exactly the k-th block of basis vectors, in its original order. Concatenating over $k = 1, \dots, n$, the list

$$T^{m_1}v_1, \dots, Tv_1, v_1, \dots, T^{m_n}v_n, \dots, Tv_n, v_n$$

is precisely the Jordan basis $u_1^1, \dots, u_{d_p}^p$ written in its original order. Hence it is a basis of V, proving (a).

Problem 8C.10. Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of V that is a Jordan basis for T. Describe the matrix of T with respect to the basis $v_n, \dots, v_1$ obtained by reversing the order of the v's.

Solution. Write $A = \mathcal{M}\big(T, (v_1, \dots, v_n)\big)$. By 8.44 this is block diagonal with blocks $A_1, \dots, A_p$ of sizes $d_1, \dots, d_p$, where A_k has λ_k on the diagonal, 1 on the line directly above the diagonal, and 0 elsewhere.
First, a general observation about reversing a basis. Put $w_j = v_{n+1-j}$, so $w_1, \dots, w_n$ is the reversed basis, and let $B = \mathcal{M}\big(T, (w_1, \dots, w_n)\big)$. For each k,

$$Tw_k = Tv_{n+1-k} = \sum_{i=1}^{n} A_{i,\,n+1-k}\, v_i = \sum_{i=1}^{n} A_{i,\,n+1-k}\, w_{n+1-i} = \sum_{j=1}^{n} A_{n+1-j,\,n+1-k}\, w_j,$$

where the last equality comes from the substitution $j = n + 1 - i$. Hence

$$B_{j,k} = A_{n+1-j,\,n+1-k}.$$

In other words, B is obtained from A by rotating it through 180 degrees: reverse the order of the rows and reverse the order of the columns.
Applying this to our block diagonal A: rotating a block diagonal matrix through 180 degrees produces a block diagonal matrix whose blocks are the original blocks in reverse order, each individually rotated through 180 degrees. Rotating A_k through 180 degrees sends the diagonal to the diagonal and sends the line directly above the diagonal to the line directly below it, so the rotated block is the d_k-by-d_k matrix with λ_k on the diagonal, 1 on the line directly below the diagonal, and 0 elsewhere. That matrix is exactly A_k^{t}, the transpose of A_k.
Therefore

$$\mathcal{M}\big(T, (v_n, \dots, v_1)\big) = \begin{pmatrix} A_p^{\mathrm{t}} & & 0 \\ & \ddots & \\ 0 & & A_1^{\mathrm{t}} \end{pmatrix}, \qquad A_k^{\mathrm{t}} = \begin{pmatrix} \lambda_k & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda_k \end{pmatrix}.$$

It is worth confirming this directly with vectors. Suppose $u_1, \dots, u_d$ are the consecutive basis vectors belonging to the block A_k, so that

$$Tu_j = \lambda_k u_j + u_{j-1} \quad (j = 1, \dots, d), \qquad u_0 = 0.$$

Reversing the whole basis reverses the order of the blocks and also reverses the order of the vectors within each block, so within this block the new ordering is $\tilde{u}_1, \dots, \tilde{u}_d$ where $\tilde{u}_i = u_{d+1-i}$. Then for $i = 1, \dots, d$,

$$T\tilde{u}_i = Tu_{d+1-i} = \lambda_k u_{d+1-i} + u_{d-i} = \lambda_k \tilde{u}_i + \tilde{u}_{i+1},$$

with the convention $\tilde{u}_{d+1} = 0$. Thus the column of $T\tilde{u}_i$ has λ_k in row i and 1 in row $i+1$, which is precisely a lower-triangular block with λ_k on the diagonal and 1's directly below the diagonal, as claimed.

So the matrix of T with respect to the reversed basis is block diagonal with the blocks in reversed order, each block being lower triangular with an eigenvalue of T repeated on its diagonal, 1's on the line directly below the diagonal, and 0's elsewhere. In particular it is a lower-triangular matrix, and it is a Jordan matrix in the sense of 8.44 exactly when every block has size 1, that is, exactly when $\mathcal{M}(T)$ was already diagonal.

Problem 8C.11. Suppose $T \in \mathcal{L}(V)$. Explain why every vector in each Jordan basis for T is a generalized eigenvector of T.

Solution. Let $v_1, \dots, v_n$ be a Jordan basis for T. By 8.44, the matrix of T with respect to this basis is block diagonal with blocks $A_1, \dots, A_p$, where A_k is d_k-by-d_k with a scalar λ_k on the diagonal, 1 on the line directly above the diagonal, and 0 elsewhere.
Fix k and let $u_1, \dots, u_{d_k}$ denote the consecutive basis vectors belonging to the block A_k; thus every v_j is one of these vectors for exactly one k. Reading off the columns of the block diagonal matrix (all entries outside the rows of the k-th block are 0, and inside the block the j-th column is the j-th column of A_k) gives

$$Tu_j = \lambda_k u_j + u_{j-1} \quad \text{for } j = 1, \dots, d_k, \qquad \text{where } u_0 = 0.$$

Equivalently,

$$(T - \lambda_k I)u_j = u_{j-1} \quad \text{for } j = 1, \dots, d_k, \qquad u_0 = 0.$$

Iterating this identity, for each $j \in \{1, \dots, d_k\}$ and each $i \in \{0, 1, \dots, j\}$ we get $(T - \lambda_k I)^i u_j = u_{j-i}$. Taking $i = j$ gives

$$(T - \lambda_k I)^j u_j = u_0 = 0.$$

Two things remain to be checked so that the definition 8.8 of a generalized eigenvector applies.
First, λ_k must be an eigenvalue of T. The block diagonal matrix above is upper triangular, and its diagonal entries are $\lambda_1, \dots, \lambda_1, \lambda_2, \dots, \lambda_p$ (each λ_k repeated d_k times). By 5.41, the eigenvalues of T are exactly the diagonal entries of an upper-triangular matrix of T. Hence each λ_k is an eigenvalue of T. (Alternatively and directly: $(T - \lambda_k I)u_1 = 0$ with $u_1 \neq 0$, so u_1 is an eigenvector of T with eigenvalue λ_k.)
Second, the vector in question must be nonzero. Each u_j is a member of a basis of V, and no vector in a linearly independent list is 0; hence $u_j \neq 0$.
Therefore each u_j is a nonzero vector satisfying $(T - \lambda_k I)^j u_j = 0$ for the eigenvalue λ_k of T, so by 8.8 it is a generalized eigenvector of T corresponding to λ_k. Since every vector of the Jordan basis is such a u_j for some block k, every vector in the Jordan basis is a generalized eigenvector of T.
Note also that this is consistent with 8.11: each of these basis vectors corresponds to the unique eigenvalue λ_k attached to the block it lies in.

Problem 8C.12. Suppose $T \in \mathcal{L}(V)$ is diagonalizable. Show that $\mathcal{M}(T)$ is a diagonal matrix with respect to every Jordan basis for T.

Solution. Let $v_1, \dots, v_n$ be a Jordan basis for T. By 8.44 the matrix of T with respect to it is block diagonal with blocks $A_1, \dots, A_p$ of sizes $d_1, \dots, d_p$, where A_k has λ_k on the diagonal, 1 directly above the diagonal, and 0 elsewhere. Since $d_1 + \cdots + d_p = n$, the matrix $\mathcal{M}(T)$ is diagonal if and only if $d_k = 1$ for every k (a block of size 1 contributes only the diagonal entry λ_k, while a block of size at least 2 contributes a nonzero off-diagonal entry, namely a 1 directly above the diagonal). So it suffices to prove that every block has size 1.
We first record the key property of diagonalizable operators that we need.
Claim. If T is diagonalizable, $\lambda \in \mathbf{F}$, and $v \in V$ satisfies $(T - \lambda I)^2 v = 0$, then $(T - \lambda I)v = 0$.
Proof of claim. Let $\lambda'_1, \dots, \lambda'_m$ be the distinct eigenvalues of T. Because T is diagonalizable, 5.55 gives

$$V = E(\lambda'_1, T) \oplus \cdots \oplus E(\lambda'_m, T).$$

Write $v = x_1 + \cdots + x_m$ with $x_i \in E(\lambda'_i, T)$. Each $E(\lambda'_i, T)$ is invariant under T and hence under

$T-\lambda I$, and on $E(\lambda_i', T)$ the operator $T-\lambda I$ is multiplication by $\lambda_i'-\lambda$. Thus

$$0=(T-\lambda I)^2 v=(\lambda_1'-\lambda)^2 x_1+\cdots+(\lambda_m'-\lambda)^2 x_m,$$

with the i-th summand lying in $E(\lambda_i', T)$. Because the sum of the eigenspaces is direct, each summand must be 0 (see 1.45): $(\lambda_i'-\lambda)^2 x_i=0$ for every i. Hence $x_i=0$ whenever $\lambda_i'\neq\lambda$. Consequently

$$(T-\lambda I)v=(\lambda_1'-\lambda)x_1+\cdots+(\lambda_m'-\lambda)x_m=0,$$

since each summand has either $x_i=0$ or $\lambda_i'-\lambda=0$. This proves the claim.
Now suppose, for contradiction, that some block has $d_k\geq 2$. Let $u_1,\ldots,u_{d_k}$ be the consecutive basis vectors belonging to that block. Reading off the first two columns of A_k inside the block diagonal matrix gives

$$Tu_1=\lambda_k u_1, \qquad Tu_2=\lambda_k u_2+u_1,$$

that is,

$$(T-\lambda_k I)u_2=u_1 \qquad \text{and} \qquad (T-\lambda_k I)^2 u_2=(T-\lambda_k I)u_1=0.$$

By the claim (applied with $\lambda=\lambda_k$ and $v=u_2$), we get $(T-\lambda_k I)u_2=0$, that is, $u_1=0$. But u_1 belongs to a basis of V, so $u_1\neq 0$. This contradiction shows that $d_k\geq 2$ is impossible.
Hence every block has size 1, so $p=n$ and

$$\mathcal{M}(T)=\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix},$$

a diagonal matrix. Since the Jordan basis was arbitrary, $\mathcal{M}(T)$ is diagonal with respect to every Jordan basis for T.
Remark. The claim above is exactly the statement that for a diagonalizable operator every generalized eigenvector is an eigenvector, so the argument can be summarized as: a Jordan block of size at least 2 would produce a generalized eigenvector that is not an eigenvector.

Problem 8C.13. Suppose $T\in\mathcal{L}(V)$ is nilpotent. Prove that if $v_1,\ldots,v_n$ are vectors in V and $m_1,\ldots,m_n$ are nonnegative integers such that

$$T^{m_1}v_1,\ldots,Tv_1,v_1,\ldots,T^{m_n}v_n,\ldots,Tv_n,v_n \quad \text{is a basis of } V$$

and

$$T^{m_1+1}v_1=\cdots=T^{m_n+1}v_n=0,$$

then $T^{m_1}v_1,\ldots,T^{m_n}v_n$ is a basis of $\operatorname{null}T$.
[This exercise shows that $n=\dim\operatorname{null}T$. Thus the positive integer n that appears above depends only on T and not on the specific Jordan basis chosen for T.]

Solution. Write $\mathcal{B}$ for the given basis of V; its members are the vectors

$$T^i v_j \qquad \text{for } j=1,\ldots,n \text{ and } i=0,1,\ldots,m_j,$$

and these are pairwise distinct members of a basis, so in particular they form a linearly independent list of length $(m_1+1)+\cdots+(m_n+1)=\dim V$.
We show that $T^{m_1}v_1,\ldots,T^{m_n}v_n$ lies in $\operatorname{null}T$, is linearly independent, and spans $\operatorname{null}T$.
Step 1: each $T^{m_j}v_j$ is in $\operatorname{null}T$. Indeed, by hypothesis

$$T\big(T^{m_j}v_j\big)=T^{m_j+1}v_j=0.$$

Step 2: $T^{m_1}v_1,\ldots,T^{m_n}v_n$ is linearly independent. This list is a sublist of the basis $\mathcal{B}$ (it consists of the vectors $T^i v_j$ with $i=m_j$, one for each j, and these are n distinct members of $\mathcal{B}$). Every sublist of a linearly independent list is linearly independent.

Step 3: $T^{m_1}v_1, \dots, T^{m_n}v_n$ spans $\operatorname{null} T$. Suppose $v \in \operatorname{null} T$. Because $\mathcal{B}$ is a basis of V, there exist scalars $a_{j,i} \in \mathbf{F}$ (for $j = 1, \dots, n$ and $i = 0, \dots, m_j$) such that

$$v = \sum_{j=1}^{n} \sum_{i=0}^{m_j} a_{j,i} T^i v_j.$$

Apply T and use $Tv = 0$ together with $T^{m_j+1}v_j = 0$:

$$0 = Tv = \sum_{j=1}^{n} \sum_{i=0}^{m_j} a_{j,i} T^{i+1} v_j = \sum_{j=1}^{n} \sum_{i=0}^{m_j-1} a_{j,i} T^{i+1} v_j,$$

where the $i = m_j$ terms dropped out because $T^{m_j+1}v_j = 0$. (If $m_j = 0$ the inner sum on the right is empty.)
Now every vector $T^{i+1}v_j$ appearing on the right satisfies $1 \le i+1 \le m_j$, so it is a member of $\mathcal{B}$; moreover distinct pairs (j, i) give distinct members of $\mathcal{B}$. Since $\mathcal{B}$ is linearly independent, all the coefficients vanish:

$$a_{j,i} = 0 \qquad \text{for } j = 1, \dots, n \text{ and } i = 0, \dots, m_j - 1.$$

Substituting back into the expression for v leaves only the terms with $i = m_j$:

$$v = \sum_{j=1}^{n} a_{j,m_j} T^{m_j} v_j.$$

Hence $v \in \operatorname{span}\big(T^{m_1}v_1, \dots, T^{m_n}v_n\big)$.
Combining Step 1 (the list lies in $\operatorname{null} T$), Step 3 (it spans $\operatorname{null} T$), and Step 2 (it is linearly independent), we conclude that $T^{m_1}v_1, \dots, T^{m_n}v_n$ is a basis of $\operatorname{null} T$.
In particular $n = \dim \operatorname{null} T$. Since 8C.9 produces such a list from any Jordan basis for the nilpotent operator T, with n equal to the number of blocks and $m_j + 1$ equal to the size of the j-th block, this shows that every Jordan basis for a nilpotent T has exactly $\dim \operatorname{null} T$ blocks: the number of blocks does not depend on the choice of Jordan basis.

Problem 8C.14. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that there does not exist a direct sum decomposition of V into two nonzero subspaces invariant under T if and only if the minimal polynomial of T is of the form $(z - \lambda)^{\dim V}$ for some $\lambda \in \mathbf{C}$.

Solution. Throughout, V is a nonzero finite-dimensional complex vector space; write $n = \dim V \ge 1$.
We first record a small fact used in both directions.
Fact. If q is a monic polynomial in $\mathcal{P}(\mathbf{C})$ that divides $(z - \lambda)^N$ for some $\lambda \in \mathbf{C}$ and some nonnegative integer N, then $q = (z - \lambda)^r$ for some integer r with $0 \le r \le N$. Indeed, if q is constant then $q = 1 = (z - \lambda)^0$. Otherwise 4.13 (the fundamental theorem of algebra, second version) factors q as $(z - \mu_1) \cdots (z - \mu_r)$ with $r = \deg q \ge 1$; each μ_i is then a zero of q, hence a zero of $(z - \lambda)^N$ (because q divides $(z - \lambda)^N$), so $\mu_i = \lambda$. Thus $q = (z - \lambda)^r$, and comparing degrees gives $r \le N$.
First direction. Suppose there does not exist a direct sum decomposition of V into two nonzero subspaces invariant under T. We must produce λ with minimal polynomial $(z - \lambda)^n$.
Since $\mathbf{F} = \mathbf{C}$, by 8.46 there is a basis $v_1, \dots, v_n$ of V that is a Jordan basis for T. By 8.44 the matrix of T with respect to it is block diagonal with blocks $A_1, \dots, A_p$ of sizes $d_1, \dots, d_p$, the block A_k having a scalar λ_k on its diagonal, 1 directly above the diagonal, and 0 elsewhere. For $k = 1, \dots, p$ let U_k be the span of the d_k consecutive basis vectors belonging to the block A_k. If $u_1, \dots, u_{d_k}$ are those vectors, then reading off the columns of the block diagonal matrix gives

$$Tu_j = \lambda_k u_j + u_{j-1} \in U_k \qquad (u_0 = 0),$$

so each U_k is invariant under T. Also each $U_k \ne \{0\}$, and the concatenation of bases of $U_1, \dots, U_p$ is

the basis $v_1, \dots, v_n$ of V, so

$$V = U_1 \oplus \cdots \oplus U_p.$$

Suppose $p \geq 2$. Put $U = U_1$ and $W = U_2 + \cdots + U_p$. Then U and W are nonzero (each $U_k \neq \{0\}$), both are invariant under T (a sum of subspaces invariant under T is invariant under T), $U + W = V$, and $\dim U + \dim W = d_1 + (d_2 + \cdots + d_p) = n = \dim V$, so $V = U \oplus W$ (by 2.39 together with 3.94, or directly because the union of the corresponding bases is the basis $v_1, \dots, v_n$). This is a direct sum decomposition of V into two nonzero subspaces invariant under T, contradicting our hypothesis.
Hence $p = 1$: there is a single Jordan block, of size n, with some $\lambda \in \mathbf{C}$ on its diagonal. Thus

$$(T - \lambda I)v_j = v_{j-1} \quad \text{for } j = 1, \dots, n, \qquad v_0 = 0.$$

Iterating, $(T - \lambda I)^i v_j = v_{j-i}$ whenever $0 \leq i \leq j$, with v_{j-i} interpreted as 0 when $i = j$. Consequently $(T - \lambda I)^n v_j = 0$ for every j, so $(T - \lambda I)^n = 0$; and

$$(T - \lambda I)^{n-1} v_n = v_1 \neq 0,$$

so $(T - \lambda I)^{n-1} \neq 0$.
Let q be the minimal polynomial of T. By 5.29, $(z - \lambda)^n$ is a polynomial multiple of q, so q divides $(z - \lambda)^n$, and by the Fact $q = (z - \lambda)^r$ with $0 \leq r \leq n$. Since $q(T) = 0$ but $(T - \lambda I)^{n-1} \neq 0$, we cannot have $r \leq n - 1$. Therefore $r = n$ and the minimal polynomial of T is $(z - \lambda)^{\dim V}$.
Second direction. Conversely, suppose the minimal polynomial of T is $(z - \lambda)^n$ for some $\lambda \in \mathbf{C}$, where $n = \dim V$. Suppose toward a contradiction that

$$V = U \oplus W$$

with U and W nonzero subspaces of V invariant under T.
Let p_U be the minimal polynomial of $T|_U \in \mathcal{L}(U)$. By 5.31 the minimal polynomial of T is a polynomial multiple of p_U; that is, p_U divides $(z - \lambda)^n$. By the Fact, $p_U = (z - \lambda)^a$ for some nonnegative integer a, and by 5.22 (which bounds the degree of a minimal polynomial by the dimension) $a = \deg p_U \leq \dim U$.
Symmetrically, the minimal polynomial of $T|_W$ is $(z - \lambda)^b$ with $b \leq \dim W$.
Let $c = \max\{a, b\}$. Then $(z - \lambda)^c$ is a polynomial multiple of p_U, so by 5.29 applied to $T|_U$ we get $\big((T - \lambda I)|_U\big)^c = 0$, i.e. $(T - \lambda I)^c u = 0$ for all $u \in U$; likewise $(T - \lambda I)^c w = 0$ for all $w \in W$. Every $v \in V$ can be written as $v = u + w$ with $u \in U$ and $w \in W$, so

$$(T - \lambda I)^c v = (T - \lambda I)^c u + (T - \lambda I)^c w = 0.$$

Thus $(T - \lambda I)^c = 0$, so by 5.29 the polynomial $(z - \lambda)^c$ is a polynomial multiple of the minimal polynomial $(z - \lambda)^n$ of T; comparing degrees gives $n \leq c$.
On the other hand, U and W are both nonzero and $\dim U + \dim W = \dim V = n$ (by 3.94), so $\dim U \leq n - 1$ and $\dim W \leq n - 1$. Hence

$$c = \max\{a, b\} \leq \max\{\dim U, \dim W\} \leq n - 1,$$

contradicting $n \leq c$.
Therefore no such decomposition exists, completing the proof.
Remark. The first direction shows more: over $\mathbf{C}$, the operator T admits no nontrivial decomposition into invariant subspaces exactly when its Jordan form consists of a single block. Note also that the second direction used neither $\mathbf{F} = \mathbf{C}$ nor Jordan form; only the first direction needs the complex hypothesis, which is essential there, since for instance the minimal polynomial of a rotation by 90 degrees on $\mathbf{R}^2$ is $z^2 + 1$, which is not of the form $(z - \lambda)^2$ with $\lambda \in \mathbf{R}$, yet that operator admits no decomposition into two nonzero invariant subspaces (it has no eigenvalues, hence no invariant subspace of dimension 1).

8.4 Exercises 8D

Problem 8D.1. Suppose V is an inner product space and $v, w \in V$. Define an operator $T \in \mathcal{L}(V)$ by $Tu = \langle u, v\rangle w$. Find a formula for $\operatorname{tr} T$.

Solution. The answer is

$$\operatorname{tr} T = \langle w, v\rangle.$$

To prove this, first note that T really is an operator on V: the map $u \mapsto \langle u, v\rangle w$ is linear because the inner product is linear in its first slot.
By the standing assumptions, V is finite-dimensional, so V has an orthonormal basis $e_1, \dots, e_n$ (by 6.35). Because the trace of an operator on an inner product space can be computed from any orthonormal basis (by 8.55),

$$\operatorname{tr} T = \sum_{k=1}^{n} \langle Te_k, e_k\rangle = \sum_{k=1}^{n} \big\langle \langle e_k, v\rangle w,\, e_k\big\rangle = \sum_{k=1}^{n} \langle e_k, v\rangle\langle w, e_k\rangle,$$

where the last equality holds because scalars come out of the first slot of an inner product.
On the other hand, 6.30(a) tells us that

$$w = \langle w, e_1\rangle e_1 + \cdots + \langle w, e_n\rangle e_n.$$

Taking the inner product of both sides with v and using linearity in the first slot gives

$$\langle w, v\rangle = \sum_{k=1}^{n} \langle w, e_k\rangle\langle e_k, v\rangle.$$

The right side of this equation is exactly the sum computed above for $\operatorname{tr} T$. Hence

$$\operatorname{tr} T = \langle w, v\rangle,$$

as claimed.
As a check, take $v = w = e_1$ for some unit vector e_1. Then T is the orthogonal projection of V onto $\operatorname{span}(e_1)$, whose trace is 1; and indeed $\langle w, v\rangle = \|e_1\|^2 = 1$.

Problem 8D.2. Suppose $P \in \mathcal{L}(V)$ satisfies $P^2 = P$. Prove that

$$\operatorname{tr} P = \dim \operatorname{range} P.$$

Solution. The idea is to produce a basis of V with respect to which the matrix of P is diagonal with only 0's and 1's on the diagonal, the number of 1's being $\dim \operatorname{range} P$.
Step 1: $\operatorname{range} P = \{v \in V : Pv = v\}$.
If $Pv = v$, then v is in the range of P. Conversely, suppose $v \in \operatorname{range} P$, say $v = Pu$ for some $u \in V$. Then

$$Pv = P(Pu) = P^2u = Pu = v.$$

Thus P acts as the identity on $\operatorname{range} P$.
Step 2: $\operatorname{null} P \cap \operatorname{range} P = \{0\}$, and every vector of V is the sum of a vector in $\operatorname{range} P$ and a vector in $\operatorname{null} P$.
If $v \in \operatorname{null} P \cap \operatorname{range} P$, then Step 1 gives $v = Pv = 0$. For the second assertion, let $v \in V$ and write

$$v = Pv + (v - Pv).$$

Here $Pv \in \operatorname{range} P$, and $P(v - Pv) = Pv - P^2v = Pv - Pv = 0$, so $v - Pv \in \operatorname{null} P$.
Step 3: a good basis.

Let $m = \dim \operatorname{range} P$ and $k = \dim \operatorname{null} P$. Choose a basis $u_1, \dots, u_m$ of $\operatorname{range} P$ and a basis $w_1, \dots, w_k$ of $\operatorname{null} P$ (either list may be empty if the corresponding subspace is $\{0\}$). Consider the list

$$u_1, \dots, u_m, w_1, \dots, w_k.$$

By Step 2 every $v \in V$ is a sum of an element of $\operatorname{range} P$ and an element of $\operatorname{null} P$, hence a linear combination of the list above; so this list spans V. Its length is $m + k$, which equals $\dim V$ by the fundamental theorem of linear maps (3.21). Thus the list is a basis of V (by 2.42).
Step 4: computing the trace.
With respect to the basis of Step 3 we have, using Step 1,

$$Pu_j = u_j \quad \text{for } j = 1, \dots, m, \qquad Pw_j = 0 \quad \text{for } j = 1, \dots, k.$$

Hence $\mathcal{M}\big(P, (u_1, \dots, u_m, w_1, \dots, w_k)\big)$ is the diagonal matrix whose first m diagonal entries equal 1 and whose remaining k diagonal entries equal 0. The trace of an operator may be computed from the matrix with respect to any basis (by 8.50 and the definition 8.51), so

$$\operatorname{tr} P = \underbrace{1 + \cdots + 1}_{m \text{ times}} + \underbrace{0 + \cdots + 0}_{k \text{ times}} = m = \dim \operatorname{range} P,$$

as desired.
Note that no inner product and no assumption on $\mathbf{F}$ was used; the result holds for $\mathbf{F} = \mathbf{R}$ and $\mathbf{F} = \mathbf{C}$ alike.

Problem 8D.3. Suppose $T \in \mathcal{L}(V)$ and $T^5 = T$. Prove that the real and imaginary parts of $\operatorname{tr} T$ are both integers.

Solution. We first do the case $\mathbf{F} = \mathbf{C}$ and then reduce the case $\mathbf{F} = \mathbf{R}$ to it.
Case 1: $\mathbf{F} = \mathbf{C}$.
Suppose $\lambda \in \mathbf{C}$ is an eigenvalue of T, with eigenvector $v \neq 0$. Then $Tv = \lambda v$, and applying T repeatedly gives $T^5 v = \lambda^5 v$. Because $T^5 = T$, we get

$$\lambda^5 v = T^5 v = Tv = \lambda v,$$

so $(\lambda^5 - \lambda)v = 0$. Since $v \neq 0$, this forces

$$\lambda^5 - \lambda = 0, \qquad \text{that is,} \qquad \lambda(\lambda - 1)(\lambda + 1)(\lambda^2 + 1) = 0.$$

Hence

$$\lambda \in \{0,\, 1,\, -1,\, i,\, -i\}.$$

Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \dots, d_m$; each d_j is a positive integer (see 8.23). By 8.52,

$$\operatorname{tr} T = d_1\lambda_1 + \cdots + d_m\lambda_m.$$

Each λ_j lies in $\{0, 1, -1, i, -i\}$. Group the terms: let a be the sum of those d_j for which $\lambda_j = 1$ minus the sum of those d_j for which $\lambda_j = -1$, and let b be the sum of those d_j for which $\lambda_j = i$ minus the sum of those d_j for which $\lambda_j = -i$ (the terms with $\lambda_j = 0$ contribute nothing). Then a and b are integers and

$$\operatorname{tr} T = a + bi.$$

Thus $\operatorname{Re}(\operatorname{tr} T) = a$ and $\operatorname{Im}(\operatorname{tr} T) = b$ are both integers.
Case 2: $\mathbf{F} = \mathbf{R}$.
Let $n = \dim V$, choose a basis of V, and let $A = \mathcal{M}(T)$ with respect to that basis; thus A is an n-by-n matrix with real entries. Because the matrix of a product of operators is the product of their matrices (3.43),

$$A^5 = \mathcal{M}(T)^5 = \mathcal{M}(T^5) = \mathcal{M}(T) = A.$$

Now regard A as a matrix with complex entries and let $S \in \mathcal{L}(\mathbf{C}^n)$ be the operator whose matrix with respect to the standard basis of $\mathbf{C}^n$ is A. Then, using 3.43 once more, $\mathcal{M}(S^5) = A^5 = A = \mathcal{M}(S)$; because an operator is determined by its matrix with respect to a fixed basis, this gives $S^5 = S$. By Case 1, $\operatorname{tr} S$ has integer real and imaginary parts. But

$$\operatorname{tr} S = \operatorname{tr} A = \operatorname{tr} T,$$

since both traces are the sum of the diagonal entries of the same matrix A. As $\operatorname{tr} T$ is a real number, its imaginary part is 0, an integer, and its real part equals $\operatorname{tr} T$ itself, which by the previous sentence is an integer.
In both cases the real and imaginary parts of $\operatorname{tr} T$ are integers.
Remark. Case 1 in fact shows more: $\operatorname{tr} T$ is a Gaussian integer $a + bi$ with $|a| + |b| \le \dim V$, because the multiplicities of the eigenvalues of T sum to $\dim V$ (by 8.25).

Problem 8D.4. Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Prove that

$$\operatorname{tr} T^* = \overline{\operatorname{tr} T}.$$

Solution. Because V is a finite-dimensional inner product space, it has an orthonormal basis $e_1, \dots, e_n$ (by 6.35). We give two proofs.
First proof (via 8.55 and the definition of the adjoint).
By 8.55 applied to T^*,

$$\operatorname{tr} T^* = \sum_{k=1}^{n} \langle T^* e_k, e_k \rangle.$$

Fix k. By conjugate symmetry of the inner product and the defining property of the adjoint (7.1), which says $\langle Tv, w \rangle = \langle v, T^* w \rangle$ for all $v, w \in V$,

$$\langle T^* e_k, e_k \rangle = \overline{\langle e_k, T^* e_k \rangle} = \overline{\langle T e_k, e_k \rangle}.$$

Summing over k and using that complex conjugation is additive,

$$\operatorname{tr} T^* = \sum_{k=1}^{n} \overline{\langle T e_k, e_k \rangle} = \overline{\sum_{k=1}^{n} \langle T e_k, e_k \rangle} = \overline{\operatorname{tr} T},$$

where the last equality is 8.55 applied to T.
Second proof (via matrices).
Let $A = \mathcal{M}\big(T, (e_1, \dots, e_n)\big)$. Since $e_1, \dots, e_n$ is orthonormal, 7.9 gives

$$\mathcal{M}\big(T^*, (e_1, \dots, e_n)\big) = A^*,$$

the conjugate transpose of A. By the definition 7.7 of conjugate transpose, $(A^*)_{k,k} = \overline{A_{k,k}}$ for each k; that is, the diagonal entries of A^* are the complex conjugates of the diagonal entries of A. Therefore

$$\operatorname{tr} T^* = \sum_{k=1}^{n} (A^*)_{k,k} = \sum_{k=1}^{n} \overline{A_{k,k}} = \overline{\sum_{k=1}^{n} A_{k,k}} = \overline{\operatorname{tr} T},$$

where the outer equalities use the definition 8.51 of the trace of an operator (computing with the orthonormal basis $e_1, \dots, e_n$, which is legitimate by 8.50).
If $\mathbf{F} = \mathbf{R}$, complex conjugation is the identity and the result simply says $\operatorname{tr} T^* = \operatorname{tr} T$.

Problem 8D.5. Suppose V is an inner product space. Suppose $T \in \mathcal{L}(V)$ is a positive operator and $\operatorname{tr} T = 0$. Prove that $T = 0$.

Solution. Because T is a positive operator, the implication (a) $\Rightarrow$ (c) of 7.38 provides an orthonormal basis $e_1, \dots, e_n$ of V with respect to which the matrix of T is diagonal with only nonnegative numbers on the diagonal. Write

$$\mathcal{M}\big(T, (e_1, \dots, e_n)\big) = \operatorname{diag}(\lambda_1, \dots, \lambda_n), \qquad \lambda_k \geq 0 \text{ for each } k.$$

The trace of an operator may be computed from its matrix with respect to any basis (8.50 and 8.51), so

$$0 = \operatorname{tr} T = \lambda_1 + \cdots + \lambda_n.$$

A sum of nonnegative real numbers equals 0 only if every term equals 0: indeed, if some $\lambda_j > 0$, then $\lambda_1 + \cdots + \lambda_n \geq \lambda_j > 0$. Hence $\lambda_1 = \cdots = \lambda_n = 0$, so $\mathcal{M}\big(T, (e_1, \dots, e_n)\big)$ is the zero matrix. Therefore $Te_k = 0$ for each k, and since $e_1, \dots, e_n$ is a basis of V, linearity gives $Tv = 0$ for every $v \in V$. Thus $T = 0$.

Alternative proof avoiding the spectral theorem. Since T is positive, 7.38 (a) $\Rightarrow$ (f) gives $R \in \mathcal{L}(V)$ with $T = R^*R$ (one may take $R = \sqrt{T}$, using 7.39 and the self-adjointness of $\sqrt{T}$). Let $e_1, \dots, e_n$ be any orthonormal basis of V (6.35). For each k we have $\langle R^*(Re_k), e_k\rangle = \langle Re_k, (R^*)^* e_k\rangle = \langle Re_k, Re_k\rangle$, using the defining property 7.1 of the adjoint applied to R^* together with $(R^*)^* = R$ (7.5(c)). Hence, by 8.55,

$$0 = \operatorname{tr} T = \sum_{k=1}^{n} \langle R^*Re_k, e_k\rangle = \sum_{k=1}^{n} \langle Re_k, Re_k\rangle = \sum_{k=1}^{n} \|Re_k\|^2.$$

Each summand is nonnegative, so $\|Re_k\| = 0$, that is, $Re_k = 0$, for every k. Hence $R = 0$ and therefore $T = R^*R = 0$.

Problem 8D.6. Suppose V is an inner product space and $P, Q \in \mathcal{L}(V)$ are orthogonal projections. Prove that $\operatorname{tr}(PQ) \geq 0$.

Solution. The key point is that $\operatorname{tr}(PQ)$ can be rewritten as $\operatorname{tr}(S^*S)$ for a suitable S, and traces of operators of that form are nonnegative.

Step 1: an orthogonal projection is idempotent and self-adjoint.

Say $P = P_U$ and $Q = P_W$ for subspaces U, W of V; these subspaces are finite-dimensional because V is. By 6.57(g),

$$P^2 = P \qquad \text{and} \qquad Q^2 = Q.$$

Next we check that $P^* = P$. Let $v, w \in V$. By 6.49 we may write

$$v = u_1 + x_1, \qquad w = u_2 + x_2$$

with $u_1, u_2 \in U$ and $x_1, x_2 \in U^\perp$; then $Pv = u_1$ and $Pw = u_2$ by the definition 6.55 of P_U. Since $\langle u_1, x_2\rangle = 0$ and $\langle x_1, u_2\rangle = 0$,

$$\langle Pv, w\rangle = \langle u_1, u_2 + x_2\rangle = \langle u_1, u_2\rangle = \langle u_1 + x_1, u_2\rangle = \langle v, Pw\rangle.$$

As this holds for all $v, w \in V$, the uniqueness of the adjoint gives $P^* = P$. The same argument gives $Q^* = Q$.

Step 2: $\operatorname{tr}(PQ) = \operatorname{tr}(S^*S)$ where $S = QP$.

Using 7.5(d) and Step 1,

$$S^*S = (QP)^*(QP) = P^*Q^*QP = PQQP = PQ^2P = PQP.$$

By 8.56, $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for all $A, B \in \mathcal{L}(V)$; taking $A = PQ$ and $B = P$ gives

$$\operatorname{tr}(S^*S) = \operatorname{tr}\big((PQ)P\big) = \operatorname{tr}\big(P(PQ)\big) = \operatorname{tr}(P^2Q) = \operatorname{tr}(PQ).$$

Step 3: $\operatorname{tr}(S^*S) \geq 0$.

Let $e_1, \ldots, e_n$ be an orthonormal basis of V (6.35). By 8.55 and the defining property of the adjoint (together with $S^{**} = S$, which is 7.5(c)),

$$\operatorname{tr}(S^*S) = \sum_{k=1}^{n} \langle S^*Se_k, e_k \rangle = \sum_{k=1}^{n} \langle Se_k, Se_k \rangle = \sum_{k=1}^{n} \|Se_k\|^2 \geq 0.$$

Combining Steps 2 and 3,

$$\operatorname{tr}(PQ) = \sum_{k=1}^{n} \|QPe_k\|^2 \geq 0,$$

as desired.
Remark. The displayed formula shows moreover that $\operatorname{tr}(PQ) = 0$ if and only if $QP = 0$, which happens exactly when $\operatorname{range} P \subseteq (\operatorname{range} Q)^{\perp}$, that is, when the two subspaces onto which P and Q project are orthogonal to each other.

Problem 8D.7. Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is the operator whose matrix is

$$\begin{pmatrix} 51 & -12 & -21 \\ 60 & -40 & -28 \\ 57 & -68 & 1 \end{pmatrix}.$$

Someone tells you (accurately) that -48 and 24 are eigenvalues of T. Without using a computer or writing anything down, find the third eigenvalue of T.

Solution. The third eigenvalue is 36.
Here is the reasoning, all of which can be done in one's head. The trace of T is the sum of the diagonal entries of its matrix (8.47 and 8.51):

$$\operatorname{tr} T = 51 + (-40) + 1 = 12.$$

Because $\mathbf{F} = \mathbf{C}$, the trace equals the sum of the eigenvalues of T, each included as many times as its multiplicity (8.52). Also, the multiplicities of the eigenvalues of T add up to $\dim \mathbf{C}^3 = 3$ (8.25).
We are told that -48 and 24 are eigenvalues, so each has multiplicity at least 1, accounting for at least 2 of the 3 available. There are two possibilities.
Possibility A: -48 and 24 are the only eigenvalues of T. Then their multiplicities are 2 and 1, or 1 and 2, and the sum of the eigenvalues with multiplicity would be

$$2(-48) + 24 = -72 \qquad \text{or} \qquad (-48) + 2(24) = 0.$$

Neither equals 12, so by 8.52 this possibility is ruled out.
Possibility B: T has a third eigenvalue λ, distinct from -48 and 24. Then T has three distinct eigenvalues whose multiplicities are positive integers summing to 3, so each multiplicity equals 1. Hence 8.52 gives

$$12 = \operatorname{tr} T = (-48) + 24 + \lambda = -24 + \lambda,$$

so

$$\lambda = 12 + 24 = 36.$$

Thus the third eigenvalue of T is 36, and the eigenvalues of T are -48, 24, 36, each with multiplicity 1.
In practice the whole computation is the mental one-liner: add the diagonal to get 12, subtract the two known eigenvalues, $12 - (-48) - 24 = 36$.

Problem 8D.8. Prove or give a counterexample: If $S, T \in \mathcal{L}(V)$, then $\operatorname{tr}(ST) = (\operatorname{tr} S)(\operatorname{tr} T)$.

Solution. The statement is false. In fact it fails for every V with $\dim V \geq 2$.
Counterexample. Let $V = \mathbf{F}^2$ and take $S = T = I$. Then $ST = I$, and the matrix of I with respect to the standard basis is the 2-by-2 identity matrix, whose diagonal entries are 1 and 1. Hence, by the definition of the trace of an operator (8.51) together with the definition of the trace of a matrix (8.47),

$$\operatorname{tr}(ST) = \operatorname{tr} I = 2,$$

whereas

$$(\operatorname{tr} S)(\operatorname{tr} T) = (\operatorname{tr} I)(\operatorname{tr} I) = 2 \cdot 2 = 4.$$

Because $2 \neq 4$, the proposed formula fails.
The same example works on any V with $\dim V \geq 2$: if $n = \dim V$ and $S = T = I$, then $\operatorname{tr}(ST) = n$ while $(\operatorname{tr} S)(\operatorname{tr} T) = n^2$, and $n \neq n^2$ whenever $n \geq 2$. So the proposed identity fails for every such V, which settles the exercise: as a claim about an arbitrary finite-dimensional nonzero vector space it is false.
For completeness, the one degenerate case is genuinely different. If $\dim V = 1$ then every operator on V has the form λI, whose matrix with respect to any basis is the 1-by-1 matrix (λ); so for $S = \lambda I$ and $T = \mu I$ both $\operatorname{tr}(ST)$ and $(\operatorname{tr} S)(\operatorname{tr} T)$ equal $\lambda\mu$, and the formula happens to hold. This is why the counterexample must be taken in dimension at least 2.
It is worth noting what is true instead: the trace is linear (8.56), so $\operatorname{tr}(S + T) = \operatorname{tr} S + \operatorname{tr} T$ and $\operatorname{tr}(\lambda T) = \lambda \operatorname{tr} T$; and $\operatorname{tr}(ST) = \operatorname{tr}(TS)$. But the trace is not multiplicative.

Problem 8D.9. Suppose $T \in \mathcal{L}(V)$ is such that $\operatorname{tr}(ST) = 0$ for all $S \in \mathcal{L}(V)$. Prove that $T = 0$.

Solution. Let $n = \dim V$ and let $v_1, \ldots, v_n$ be a basis of V. All matrices below are with respect to this basis; write $A = \mathcal{M}(T)$, so that

$$Tv_k = \sum_{r=1}^{n} A_{r,k} v_r \quad \text{for } k = 1, \ldots, n.$$

For each pair $j, k \in \{1, \ldots, n\}$ define $P_{j,k} \in \mathcal{L}(V)$ by

$$P_{j,k}(a_1 v_1 + \cdots + a_n v_n) = a_k v_j.$$

Each $P_{j,k}$ is linear because the coefficient functionals $a_1 v_1 + \cdots + a_n v_n \mapsto a_k$ are linear. Since $P_{j,k} v_k = v_j$ and $P_{j,k} v_i = 0$ for $i \neq k$, the matrix $\mathcal{M}(P_{j,k})$ is the matrix $E_{j,k}$ whose entry in row j, column k equals 1 and whose other entries all equal 0.
Now fix j, k and take $S = P_{j,k}$. By 3.43 the matrix of a product of operators is the product of their matrices, so

$$\mathcal{M}(P_{j,k} T) = E_{j,k} A.$$

The entry of $E_{j,k} A$ in row r, column s equals

$$\sum_{t=1}^{n} (E_{j,k})_{r,t} A_{t,s} = \begin{cases} A_{k,s} & \text{if } r = j, \\ 0 & \text{if } r \neq j, \end{cases}$$

because the only nonzero entry of row r of $E_{j,k}$ occurs when $r = j$, in column k. Hence the only possibly nonzero diagonal entry of $E_{j,k} A$ is the one in row j, column j, which equals $A_{k,j}$. Therefore, by 8.47 and 8.51,

$$\operatorname{tr}(P_{j,k} T) = \operatorname{tr}(E_{j,k} A) = A_{k,j}.$$

By hypothesis $\operatorname{tr}(ST) = 0$ for every $S \in \mathcal{L}(V)$, and in particular for $S = P_{j,k}$. Thus $A_{k,j} = 0$. As j and k range over $\{1, \ldots, n\}$, this shows that every entry of A equals 0, so $A = \mathcal{M}(T)$ is the zero matrix. Because $Tv_k = \sum_r A_{r,k} v_r = 0$ for each k and $v_1, \ldots, v_n$ spans V, we conclude that $T = 0$.

Remark. If V is an inner product space, there is a slicker proof: take $S = T^*$ and use 8.55 together with an orthonormal basis $e_1, \dots, e_n$ to get

$$0 = \operatorname{tr}(T^*T) = \sum_{k=1}^{n} \langle T^*Te_k, e_k \rangle = \sum_{k=1}^{n} \|Te_k\|^2,$$

which forces $Te_k = 0$ for each k and hence $T = 0$. The matrix argument above has the advantage of requiring no inner product, which matters here since V is only assumed to be a finite-dimensional vector space.

Problem 8D.10. Prove that the trace is the only linear functional $\tau\colon \mathcal{L}(V) \to \mathbf{F}$ such that

$$\tau(ST) = \tau(TS)$$

for all $S, T \in \mathcal{L}(V)$ and $\tau(I) = \dim V$.
Hint: Suppose that $v_1, \dots, v_n$ is a basis of V. For $j, k \in \{1, \dots, n\}$, define $P_{j,k} \in \mathcal{L}(V)$ by $P_{j,k}(a_1v_1 + \cdots + a_nv_n) = a_kv_j$. Prove that

$$\tau(P_{j,k}) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Then for $T \in \mathcal{L}(V)$, use the equation $T = \sum_{k=1}^{n} \sum_{j=1}^{n} \mathcal{M}(T)_{j,k} P_{j,k}$ to show that $\tau(T) = \operatorname{tr} T$.

Solution. Two things must be shown: that the trace has the two stated properties, and that no other linear functional does.
Existence. The function $\operatorname{tr}\colon \mathcal{L}(V) \to \mathbf{F}$ is a linear functional satisfying $\operatorname{tr}(ST) = \operatorname{tr}(TS)$ for all $S, T \in \mathcal{L}(V)$; this is 8.56. Moreover the matrix of I with respect to any basis of V is the n-by-n identity matrix, whose n diagonal entries all equal 1, so $\operatorname{tr} I = n = \dim V$ by 8.47 and 8.51. Thus the trace is such a functional.
Uniqueness. Suppose $\tau\colon \mathcal{L}(V) \to \mathbf{F}$ is linear, $\tau(ST) = \tau(TS)$ for all $S, T \in \mathcal{L}(V)$, and $\tau(I) = n$, where $n = \dim V$. Fix a basis $v_1, \dots, v_n$ of V and define $P_{j,k} \in \mathcal{L}(V)$ for $j, k \in \{1, \dots, n\}$ by

$$P_{j,k}(a_1v_1 + \cdots + a_nv_n) = a_kv_j.$$

Each $P_{j,k}$ is linear (the map sending a vector to its k^{th} coordinate with respect to the basis is linear). Equivalently, $P_{j,k}$ is determined on the basis by

$$P_{j,k}v_i = \begin{cases} v_j & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

Step 1: the multiplication rule. For all $j, k, l, m \in \{1, \dots, n\}$,

$$P_{j,k}P_{l,m} = \begin{cases} P_{j,m} & \text{if } k = l, \\ 0 & \text{if } k \neq l. \end{cases}$$

Indeed, apply both sides to a basis vector v_i. If $i \neq m$ then $P_{l,m}v_i = 0$, so $(P_{j,k}P_{l,m})v_i = 0$; and $P_{j,m}v_i = 0$ as well. If $i = m$ then $(P_{j,k}P_{l,m})v_m = P_{j,k}v_l$, which equals v_j when $l = k$ and equals 0 when $l \neq k$; this agrees with $P_{j,m}v_m = v_j$ in the first case and with 0 in the second. Since two linear maps agreeing on a basis are equal, the rule holds.
Step 2: $\tau(P_{j,k}) = 0$ when $j \neq k$. Suppose $j \neq k$. Apply Step 1 to $P_{j,j}P_{j,k}$: the second index of the left factor is j and the first index of the right factor is j, and these agree, so $P_{j,j}P_{j,k} = P_{j,k}$. Apply Step 1 to $P_{j,k}P_{j,j}$: now the second index of the left factor is k while the first index of the right factor is j, and $k \neq j$, so $P_{j,k}P_{j,j} = 0$. Hence

$$\tau(P_{j,k}) = \tau(P_{j,j}P_{j,k}) = \tau(P_{j,k}P_{j,j}) = \tau(0) = 0,$$

where the second equality uses the hypothesis $\tau(SU) = \tau(US)$ with $S = P_{j,j}$ and $U = P_{j,k}$, and the last equality holds because τ is linear.
Step 3: $\tau(P_{j,j})$ is the same for all j. Fix j, k. By Step 1,

$$P_{j,k}P_{k,j} = P_{j,j} \quad \text{and} \quad P_{k,j}P_{j,k} = P_{k,k}.$$

Therefore

$$\tau(P_{j,j}) = \tau(P_{j,k}P_{k,j}) = \tau(P_{k,j}P_{j,k}) = \tau(P_{k,k}).$$

Thus there is $c \in \mathbf{F}$ with $\tau(P_{j,j}) = c$ for every j.
Step 4: $c = 1$. For every $v = a_1v_1 + \cdots + a_nv_n \in V$ we have

$$\Big(\sum_{j=1}^{n} P_{j,j}\Big)v = a_1v_1 + \cdots + a_nv_n = v,$$

so $\sum_{j=1}^{n} P_{j,j} = I$. Applying the linear functional τ and using Step 3,

$$n = \tau(I) = \sum_{j=1}^{n} \tau(P_{j,j}) = nc.$$

Since $n = \dim V \geq 1$, we have $n \neq 0$ in $\mathbf{F}$ (here $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$, which have characteristic 0), so $c = 1$. Combining Steps 2, 3, and 4:

$$\tau(P_{j,k}) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Step 5: $\tau = \operatorname{tr}$. Let $T \in \mathcal{L}(V)$ and let $A = \mathcal{M}(T, (v_1, \ldots, v_n))$, so $Tv_k = \sum_{j=1}^{n} A_{j,k}v_j$ for each k. I claim

$$T = \sum_{k=1}^{n}\sum_{j=1}^{n} A_{j,k}P_{j,k}.$$

To verify this, apply the right side to the basis vector v_i: since $P_{j,k}v_i = 0$ unless $k = i$, in which case it equals v_j, the right side sends v_i to $\sum_{j=1}^{n} A_{j,i}v_j = Tv_i$. Two linear maps agreeing on a basis are equal, which proves the claim.
Now apply τ and use its linearity together with Step 4:

$$\tau(T) = \sum_{k=1}^{n}\sum_{j=1}^{n} A_{j,k}\,\tau(P_{j,k}) = \sum_{j=1}^{n} A_{j,j} = \operatorname{tr} A = \operatorname{tr} T,$$

the last equality being the definition 8.51 of the trace of an operator. Since $T \in \mathcal{L}(V)$ was arbitrary, $\tau = \operatorname{tr}$, completing the proof of uniqueness.

Problem 8D.11. Suppose V and W are inner product spaces and $T \in \mathcal{L}(V, W)$. Prove that if $e_1, \ldots, e_n$ is an orthonormal basis of V and $f_1, \ldots, f_m$ is an orthonormal basis of W, then

$$\operatorname{tr}(T^*T) = \sum_{k=1}^{n}\sum_{j=1}^{m} |\langle Te_k, f_j\rangle|^2.$$

The numbers $\langle Te_k, f_j\rangle$ are the entries of the matrix of T with respect to the orthonormal bases $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$. These numbers depend on the bases, but $\operatorname{tr}(T^*T)$ does not depend on a choice of bases. Thus this exercise shows that the sum of the squares of the absolute values of the matrix entries does not depend on which orthonormal bases are used.

Solution. First note that the hypotheses make everything in the asserted formula meaningful. Because V has a basis $e_1, \ldots, e_n$ and W has a basis $f_1, \ldots, f_m$, both V and W are finite-dimensional; hence

the adjoint T^* is defined (7.1) and $T^* \in \mathcal{L}(W,V)$ by 7.4. Thus $T^*T \in \mathcal{L}(V)$, an operator on a finite-dimensional space, so $\operatorname{tr}(T^*T)$ is defined.
Since $e_1,\dots,e_n$ is an orthonormal basis of the inner product space V and T^*T is an operator on V, the formula 8.55 for the trace on an inner product space gives

$$\operatorname{tr}(T^*T) = \sum_{k=1}^{n} \langle (T^*T)e_k, e_k\rangle.$$

For each k, the definition 7.1 of the adjoint (applied with $v = e_k \in V$ and $w = Te_k \in W$, in the form $\langle Tv, w\rangle = \langle v, T^*w\rangle$, and using conjugate symmetry of the inner product) yields

$$\langle T^*(Te_k), e_k\rangle = \overline{\langle e_k, T^*(Te_k)\rangle} = \overline{\langle Te_k, Te_k\rangle} = \|Te_k\|^2,$$

where the last step uses that $\langle Te_k, Te_k\rangle = \|Te_k\|^2$ is a nonnegative real number, hence equal to its own complex conjugate. Therefore

$$\operatorname{tr}(T^*T) = \sum_{k=1}^{n} \|Te_k\|^2.$$

Now fix k. The vector Te_k lies in W, and $f_1,\dots,f_m$ is an orthonormal basis of W, so Parseval's identity 6.30(b) gives

$$\|Te_k\|^2 = \sum_{j=1}^{m} |\langle Te_k, f_j\rangle|^2.$$

Substituting this into the previous display produces

$$\operatorname{tr}(T^*T) = \sum_{k=1}^{n}\sum_{j=1}^{m} |\langle Te_k, f_j\rangle|^2,$$

as desired.
Interpretation. By 6.30(a) applied to $Te_k \in W$, we have $Te_k = \sum_{j=1}^{m}\langle Te_k, f_j\rangle f_j$, so $\langle Te_k, f_j\rangle$ is exactly the entry in row j, column k of $\mathcal{M}\big(T,(e_1,\dots,e_n),(f_1,\dots,f_m)\big)$. The left side $\operatorname{tr}(T^*T)$ is defined without reference to any basis. Hence the sum of the squares of the absolute values of the entries of the matrix of T is the same for every choice of orthonormal bases of V and of W, even though the individual entries change.

Problem 8D.12. Suppose V and W are finite-dimensional inner product spaces.
(a) Prove that $\langle S,T\rangle = \operatorname{tr}(T^*S)$ defines an inner product on $\mathcal{L}(V,W)$.
(b) Suppose $e_1,\dots,e_n$ is an orthonormal basis of V and $f_1,\dots,f_m$ is an orthonormal basis of W. Show that the inner product on $\mathcal{L}(V,W)$ from (a) is the same as the standard inner product on $\mathbf{F}^{mn}$, where we identify each element of $\mathcal{L}(V,W)$ with its matrix (with respect to the bases just mentioned) and then with an element of $\mathbf{F}^{mn}$.
Caution: The norm of a linear map $T \in \mathcal{L}(V,W)$ as defined by 7.86 is not the same as the norm that comes from the inner product in (a) above. Unless explicitly stated otherwise, always assume that $\|T\|$ refers to the norm as defined by 7.86. The norm that comes from the inner product in (a) is called the Frobenius norm or the Hilbert–Schmidt norm.

Solution. Throughout, fix an orthonormal basis $e_1,\dots,e_n$ of V and an orthonormal basis $f_1,\dots,f_m$ of W; such bases exist because every finite-dimensional inner product space has an orthonormal basis (6.35, via Gram–Schmidt). For $S,T \in \mathcal{L}(V,W)$ the adjoint T^* is defined (7.1) and $T^* \in \mathcal{L}(W,V)$ by 7.4, so $T^*S \in \mathcal{L}(V)$ and the proposed expression $\operatorname{tr}(T^*S)$ is a well-defined scalar in $\mathbf{F}$.
We record one identity used twice below. For $R \in \mathcal{L}(V)$,

$$\operatorname{tr}(R^*) = \overline{\operatorname{tr} R}.$$

Indeed, by 8.55 and the definition 7.1 of the adjoint,

$$\operatorname{tr}(R^*) = \sum_{k=1}^{n} \langle R^* e_k, e_k \rangle = \sum_{k=1}^{n} \overline{\langle e_k, R^* e_k \rangle} = \sum_{k=1}^{n} \overline{\langle R e_k, e_k \rangle} = \overline{\operatorname{tr} R},$$

using conjugate symmetry of the inner product in the second and last steps. (This is Exercise 8D.4.)
(a) We verify the axioms of an inner product (6.2) for $\langle S, T \rangle = \operatorname{tr}(T^* S)$ on $\mathcal{L}(V, W)$.
Positivity and definiteness. Let $S \in \mathcal{L}(V, W)$. By exactly the computation carried out in Exercise 8D.11,

$$\langle S, S \rangle = \operatorname{tr}(S^* S) = \sum_{k=1}^{n} \langle S^* S e_k, e_k \rangle = \sum_{k=1}^{n} \langle S e_k, S e_k \rangle = \sum_{k=1}^{n} \|S e_k\|^2,$$

which is a sum of nonnegative real numbers, hence $\langle S, S \rangle \geq 0$. Moreover $\langle S, S \rangle = 0$ forces $\|S e_k\| = 0$, that is $S e_k = 0$, for every k; since $e_1, \ldots, e_n$ is a basis of V, this gives $S = 0$. Conversely $\langle 0, 0 \rangle = \operatorname{tr} 0 = 0$.
Additivity in the first slot. For $S_1, S_2, T \in \mathcal{L}(V, W)$, distributivity of composition over addition gives $T^*(S_1 + S_2) = T^* S_1 + T^* S_2$, so by linearity of the trace (8.56),

$$\langle S_1 + S_2, T \rangle = \operatorname{tr}(T^* S_1 + T^* S_2) = \operatorname{tr}(T^* S_1) + \operatorname{tr}(T^* S_2) = \langle S_1, T \rangle + \langle S_2, T \rangle.$$

Homogeneity in the first slot. For $\lambda \in \mathbf{F}$, we have $T^*(\lambda S) = \lambda (T^* S)$, so again by 8.56,

$$\langle \lambda S, T \rangle = \operatorname{tr}(\lambda\, T^* S) = \lambda \operatorname{tr}(T^* S) = \lambda \langle S, T \rangle.$$

Conjugate symmetry. By 7.5 we have $(T^* S)^* = S^* (T^*)^* = S^* T$. Hence, using the identity recorded above with $R = T^* S \in \mathcal{L}(V)$,

$$\langle T, S \rangle = \operatorname{tr}(S^* T) = \operatorname{tr}\big((T^* S)^*\big) = \overline{\operatorname{tr}(T^* S)} = \overline{\langle S, T \rangle}.$$

All the axioms hold, so $\langle \cdot, \cdot \rangle$ is an inner product on $\mathcal{L}(V, W)$.
(b) For $T \in \mathcal{L}(V, W)$ write $A = \mathcal{M}\big(T, (e_1, \ldots, e_n), (f_1, \ldots, f_m)\big)$, an m-by-n matrix. Since $T e_k \in W$ and $f_1, \ldots, f_m$ is an orthonormal basis of W, 6.30(a) gives $T e_k = \sum_{j=1}^{m} \langle T e_k, f_j \rangle f_j$, so

$$A_{j,k} = \langle T e_k, f_j \rangle.$$

Similarly write $B_{j,k} = \langle S e_k, f_j \rangle$ for the entries of the matrix B of S with respect to the same bases. Now compute, using 8.55 and then 7.1 as in part (a),

$$\langle S, T \rangle = \operatorname{tr}(T^* S) = \sum_{k=1}^{n} \langle T^* S e_k, e_k \rangle = \sum_{k=1}^{n} \langle S e_k, T e_k \rangle,$$

where the last equality holds because $\langle T^*(S e_k), e_k \rangle = \overline{\langle e_k, T^*(S e_k) \rangle} = \overline{\langle T e_k, S e_k \rangle} = \langle S e_k, T e_k \rangle$. Applying 6.30(c) in W to the pair $S e_k$, $T e_k$ and the orthonormal basis $f_1, \ldots, f_m$,

$$\langle S e_k, T e_k \rangle = \sum_{j=1}^{m} \langle S e_k, f_j \rangle \overline{\langle T e_k, f_j \rangle} = \sum_{j=1}^{m} B_{j,k} \overline{A_{j,k}}.$$

Summing over k,

$$\langle S, T \rangle = \sum_{k=1}^{n} \sum_{j=1}^{m} B_{j,k} \overline{A_{j,k}}.$$

Fix any bijective listing of the mn index pairs (j, k) and use it to identify an m-by-n matrix with an element of $\mathbf{F}^{mn}$; under the composite identification $T \mapsto A \mapsto (A_{j,k})_{(j,k)} \in \mathbf{F}^{mn}$, the displayed formula reads

$$\langle S, T \rangle = \sum_{(j,k)} B_{j,k} \overline{A_{j,k}},$$

which is precisely the standard inner product on $\mathbf{F}^{mn}$ of the vectors corresponding to S and to T. Thus the map $\mathcal{L}(V,W) \to \mathbf{F}^{mn}$ sending each linear map to its matrix (with respect to the chosen orthonormal bases), read as a vector, is a linear bijection (by 3.71, $\mathcal{M}$ is an isomorphism from $\mathcal{L}(V,W)$ onto $\mathbf{F}^{m,n}$) that preserves inner products; the two inner products agree under this identification.
In particular the associated norm is

$$\sqrt{\langle T, T\rangle} = \Big(\sum_{k=1}^{n}\sum_{j=1}^{m} |A_{j,k}|^2\Big)^{1/2},$$

the Frobenius (Hilbert–Schmidt) norm; consistently with Exercise 8D.11, its square equals $\operatorname{tr}(T^*T)$ and therefore does not depend on which orthonormal bases were chosen.

Problem 8D.13. Find $S, T \in \mathcal{L}(\mathcal{P}(\mathbf{F}))$ such that $ST - TS = I$.
Hint: Make an appropriate modification of the operators in Example 3.9.
This exercise shows that additional hypotheses are needed on S and T to extend 8.57 to the setting of infinite-dimensional vector spaces.

Solution. Let $\mathcal{P}(\mathbf{F})$ denote the vector space of all polynomials with coefficients in $\mathbf{F}$. Define $S, T \in \mathcal{L}(\mathcal{P}(\mathbf{F}))$ by

$$(Sp)(x) = p'(x) \qquad \text{and} \qquad (Tp)(x) = x\,p(x)$$

for $p \in \mathcal{P}(\mathbf{F})$. In other words, S is differentiation and T is multiplication by x; these are the operators of Example 3.9 with multiplication by x^2 replaced by multiplication by x. Both are linear maps of $\mathcal{P}(\mathbf{F})$ into itself: differentiation of polynomials is linear and takes polynomials to polynomials, and $p \mapsto xp$ is clearly linear and takes polynomials to polynomials. (When $\mathbf{F} = \mathbf{C}$, interpret p' as the formal derivative $\sum_k k a_k x^{k-1}$ of $p(x) = \sum_k a_k x^k$; this is linear for the same reason.)
I claim $ST - TS = I$.
Let $p \in \mathcal{P}(\mathbf{F})$. Then $(Tp)(x) = xp(x)$, and applying the product rule for differentiation (equivalently, computing with the formal derivative, as done below),

$$\big((ST)p\big)(x) = \big(x\,p(x)\big)' = p(x) + x\,p'(x).$$

On the other hand,

$$\big((TS)p\big)(x) = x\,p'(x).$$

Subtracting,

$$\big((ST - TS)p\big)(x) = \big(p(x) + x\,p'(x)\big) - x\,p'(x) = p(x)$$

for all x. Hence $(ST - TS)p = p$ for every $p \in \mathcal{P}(\mathbf{F})$, that is, $ST - TS = I$.
For a computation valid over $\mathbf{F} = \mathbf{R}$ and $\mathbf{F} = \mathbf{C}$ alike that uses only the formal derivative, write $p(x) = a_0 + a_1 x + \cdots + a_d x^d$. Then $(Tp)(x) = a_0 x + a_1 x^2 + \cdots + a_d x^{d+1}$, so

$$\big((ST)p\big)(x) = \sum_{k=0}^{d} (k+1) a_k x^k,$$

while $(Sp)(x) = \sum_{k=1}^{d} k a_k x^{k-1}$, so

$$\big((TS)p\big)(x) = \sum_{k=1}^{d} k a_k x^k = \sum_{k=0}^{d} k a_k x^k.$$

Therefore

$$\big((ST - TS)p\big)(x) = \sum_{k=0}^{d} \big((k+1) - k\big) a_k x^k = \sum_{k=0}^{d} a_k x^k = p(x),$$

confirming that $ST - TS = I$.

Remark. This does not contradict 8.57, which says that no operators S, T on V satisfy $ST - TS = I$: that result carries the standing assumption that V is finite-dimensional and nonzero. Its proof computes $\operatorname{tr}(ST - TS) = \operatorname{tr}(ST) - \operatorname{tr}(TS) = 0$ by 8.56 and compares this with $\operatorname{tr} I = \dim V \neq 0$, so it needs both the existence of the trace, which is defined only for operators on finite-dimensional spaces, and $\dim V \neq 0$. Here $\mathcal{P}(\mathbf{F})$ is infinite-dimensional (the polynomials $1, x, x^2, \ldots$ form a linearly independent list of arbitrary length), so there is no trace available and the argument of 8.57 collapses. Note also that neither S nor T above has any eigenvalue-based obstruction: T has no eigenvalues at all, while S has eigenvalue 0 with eigenvectors the nonzero constant polynomials.

9 Multilinear Algebra and Determinants

Contents

9.1 Exercises 9A 552
9.2 Exercises 9B 560
9.3 Exercises 9C 567
9.4 Exercises 9D 586

9.1 Exercises 9A

Problem 9A.1. Prove that if β is a bilinear form on $\mathbf{F}$, then there exists $c \in \mathbf{F}$ such that

$$\beta(x, y) = c\, x\, y$$

for all $x, y \in \mathbf{F}$.

Solution. Here $\mathbf{F}$ is regarded as a one-dimensional vector space over itself, so the list consisting of the single vector 1 is a basis of $\mathbf{F}$.
Set

$$c = \beta(1, 1).$$

Let $x, y \in \mathbf{F}$. Because β is a bilinear form, the function $u \mapsto \beta(u, y)$ is a linear functional on $\mathbf{F}$ for each fixed y, and the function $w \mapsto \beta(x, w)$ is a linear functional on $\mathbf{F}$ for each fixed x. Writing $x = x \cdot 1$ and $y = y \cdot 1$ and using homogeneity in the first slot and then in the second slot, we get

$$\beta(x, y) = \beta(x \cdot 1,\ y \cdot 1) = x\, \beta(1,\ y \cdot 1) = x\, y\, \beta(1, 1) = c\, x\, y,$$

as desired.
Note that c is uniquely determined by β: taking $x = y = 1$ in the displayed formula forces $c = \beta(1, 1)$. Thus the map $\beta \mapsto \beta(1, 1)$ is an isomorphism of $\mathbf{F}^{(2)}$ onto $\mathbf{F}$, which is the case $n = 1$ of 9.5 (the assertion $\dim V^{(2)} = (\dim V)^2$).

Problem 9A.2. Let $n = \dim V$. Suppose β is a bilinear form on V. Prove that there exist $\varphi_1, \ldots, \varphi_n, \tau_1, \ldots, \tau_n \in V'$ such that

$$\beta(u, v) = \varphi_1(u) \cdot \tau_1(v) + \cdots + \varphi_n(u) \cdot \tau_n(v)$$

for all $u, v \in V$.
[This exercise shows that if $n = \dim V$, then every bilinear form on V is of the form given by the last bullet point of Example 9.2.]

Solution. Let $e_1, \ldots, e_n$ be a basis of V and let $\varphi_1, \ldots, \varphi_n \in V'$ be the dual basis of $e_1, \ldots, e_n$ (see 3.112), so that $\varphi_j(e_k)$ equals 1 if $j = k$ and equals 0 if $j \neq k$.
For each $k \in \{1, \ldots, n\}$ define $\tau_k \colon V \to \mathbf{F}$ by

$$\tau_k(v) = \beta(e_k, v).$$

Because β is a bilinear form, the function $v \mapsto \beta(e_k, v)$ is a linear functional on V; thus $\tau_k \in V'$ for each k.
Now suppose $u, v \in V$. By 3.114 (the dual basis gives the coefficients in a basis expansion),

$$u = \varphi_1(u)\, e_1 + \cdots + \varphi_n(u)\, e_n.$$

Because β is linear in its first slot, applying $\beta(\cdot\,, v)$ to both sides gives

$$\begin{aligned} \beta(u, v) &= \beta\big(\varphi_1(u)\, e_1 + \cdots + \varphi_n(u)\, e_n,\ v\big) \\ &= \varphi_1(u)\, \beta(e_1, v) + \cdots + \varphi_n(u)\, \beta(e_n, v) \\ &= \varphi_1(u) \cdot \tau_1(v) + \cdots + \varphi_n(u) \cdot \tau_n(v), \end{aligned}$$

as desired.
Two remarks. First, the scalars $\varphi_k(u)$ pulled out of β legitimately, since homogeneity in the first slot is part of bilinearity. Second, the construction makes the meaning of the τ_k transparent: if $A = \mathcal{M}(\beta, (e_1, \ldots, e_n))$ is the matrix of β with respect to $e_1, \ldots, e_n$ (see 9.4), then $\tau_k = A_{k,1}\varphi_1 + \cdots + A_{k,n}\varphi_n$,

because $\tau_k(e_m) = \beta(e_k, e_m) = A_{k,m}$. Thus

$$\beta(u, v) = \sum_{k=1}^{n} \sum_{m=1}^{n} A_{k,m}\, \varphi_k(u)\, \varphi_m(v)$$

for all $u, v \in V$, which exhibits β explicitly in terms of the dual basis and recovers the formula used in the proof of 9.5.

Problem 9A.3. Suppose $\beta\colon V \times V \to \mathbf{F}$ is a bilinear form on V and also is a linear functional on $V \times V$. Prove that $\beta = 0$.

Solution. Recall that $V \times V$ is a vector space under coordinatewise addition and scalar multiplication, so it makes sense to ask whether the function β is linear on $V \times V$. The two hypotheses are genuinely different: bilinearity says β is linear in each slot separately when the other slot is held fixed, while linearity on $V \times V$ says β is linear in the pair.
First note a consequence of bilinearity alone: for every $u \in V$,

$$\beta(u, 0) = 0 \qquad \text{and} \qquad \beta(0, u) = 0.$$

Indeed, for fixed u the function $w \mapsto \beta(u, w)$ is a linear map from V to $\mathbf{F}$, and every linear map takes 0 to 0; this gives $\beta(u, 0) = 0$. Similarly, for fixed u the function $w \mapsto \beta(w, u)$ is linear, so $\beta(0, u) = 0$.
Now suppose $u, v \in V$. In the vector space $V \times V$ we have

$$(u, v) = (u, 0) + (0, v).$$

Because β is a linear functional on $V \times V$, it is additive on $V \times V$, so

$$\beta(u, v) = \beta\big((u, 0) + (0, v)\big) = \beta(u, 0) + \beta(0, v) = 0 + 0 = 0.$$

Since $u, v \in V$ were arbitrary, $\beta = 0$, as desired.
Conversely, the zero function is both a bilinear form on V and a linear functional on $V \times V$. Hence the zero form is the only function satisfying both hypotheses, which is exactly the assertion made in the text preceding 9.3.

Problem 9A.4. Suppose V is a real inner product space and β is a bilinear form on V. Show that there exists a unique operator $T \in \mathcal{L}(V)$ such that

$$\beta(u, v) = \langle u, Tv \rangle$$

for all $u, v \in V$.
[This exercise states that if V is a real inner product space, then every bilinear form on V is of the form given by the third bullet point in 9.2.]

Solution. Throughout, V is finite-dimensional (as is assumed throughout this chapter) and $\mathbf{F} = \mathbf{R}$.
Existence. Fix $v \in V$. Because β is a bilinear form, the function

$$u \mapsto \beta(u, v)$$

is a linear functional on V. By the Riesz representation theorem (6.42), there is a unique vector, which we call Tv, such that

$$\beta(u, v) = \langle u, Tv \rangle \quad \text{for all } u \in V.$$

Since Tv is uniquely determined by v, this defines a function $T\colon V \to V$, and by construction $\beta(u, v) = \langle u, Tv \rangle$ for all $u, v \in V$.
We must check that T is linear. Before doing so, record the standard fact that if $w \in V$ satisfies $\langle u, w \rangle = 0$ for all $u \in V$, then $w = 0$: take $u = w$ to get $\langle w, w \rangle = 0$, which forces $w = 0$ by the

definiteness requirement in the definition of inner product (6.2).
Suppose $v_1, v_2 \in V$. For every $u \in V$,

$$\begin{aligned}\langle u, T(v_1 + v_2)\rangle &= \beta(u, v_1 + v_2)\\ &= \beta(u, v_1) + \beta(u, v_2)\\ &= \langle u, Tv_1\rangle + \langle u, Tv_2\rangle\\ &= \langle u, Tv_1 + Tv_2\rangle,\end{aligned}$$

where the second line uses additivity of β in its second slot and the last line uses additivity of the real inner product in its second slot. Hence $\langle u,\ T(v_1 + v_2) - Tv_1 - Tv_2\rangle = 0$ for all $u \in V$, and therefore $T(v_1 + v_2) = Tv_1 + Tv_2$.
Now suppose $\lambda \in \mathbf{R}$ and $v \in V$. For every $u \in V$,

$$\langle u, T(\lambda v)\rangle = \beta(u, \lambda v) = \lambda\, \beta(u, v) = \lambda\langle u, Tv\rangle = \langle u, \lambda Tv\rangle,$$

where the last equality holds because the inner product is real (so it is homogeneous, not conjugate-homogeneous, in its second slot). Hence $T(\lambda v) = \lambda Tv$.
Thus $T \in \mathcal{L}(V)$.
Uniqueness. Suppose $S, T \in \mathcal{L}(V)$ both satisfy $\beta(u, v) = \langle u, Sv\rangle$ and $\beta(u, v) = \langle u, Tv\rangle$ for all $u, v \in V$. Fix $v \in V$. Then $\langle u, Sv - Tv\rangle = 0$ for all $u \in V$, so $Sv = Tv$ by the fact recorded above. Because $v \in V$ was arbitrary, $S = T$.
Hence there is exactly one such operator T, completing the proof.
Remark: the same argument works verbatim on a complex inner product space if one asks instead for $\beta(u, v) = \langle u, Tv\rangle$ with β linear in the first slot and conjugate-linear in the second; for a genuine bilinear form over $\mathbf{C}$ the homogeneity computation above would produce $T(\lambda v) = \overline{\lambda}\, Tv$, so T need not be linear. This is why the hypothesis $\mathbf{F} = \mathbf{R}$ appears.

Problem 9A.5. Suppose β is a bilinear form on a real inner product space V and T is the unique operator on V such that $\beta(u, v) = \langle u, Tv\rangle$ for all $u, v \in V$ (see Exercise 4). Show that β is an inner product on V if and only if T is an invertible positive operator on V.

Solution. Throughout, V is a finite-dimensional real inner product space with inner product $\langle \cdot, \cdot \rangle$, and $\beta(u, v) = \langle u, Tv\rangle$ for all $u, v \in V$.
Recall the two definitions in play. By 6.2, a function $\beta\colon V \times V \to \mathbf{R}$ is an inner product on V if it satisfies: positivity ($\beta(v, v) \geq 0$ for all $v \in V$); definiteness ($\beta(v, v) = 0$ if and only if $v = 0$); additivity and homogeneity in the first slot; and symmetry ($\beta(u, v) = \beta(v, u)$, since we are over $\mathbf{R}$). By 7.34, an operator $T \in \mathcal{L}(V)$ is positive if T is self-adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$.
Note at the outset that β is bilinear by hypothesis, so additivity and homogeneity in the first slot hold automatically. Thus β is an inner product if and only if it is symmetric, positive, and definite.
Suppose first that T is an invertible positive operator.
Symmetry: because T is self-adjoint and the inner product on the real space V is symmetric,

$$\beta(u, v) = \langle u, Tv\rangle = \langle Tu, v\rangle = \langle v, Tu\rangle = \beta(v, u)$$

for all $u, v \in V$, where the second equality is the definition of T being self-adjoint ($T^* = T$).
Positivity: for $v \in V$,

$$\beta(v, v) = \langle v, Tv\rangle = \langle Tv, v\rangle \geq 0,$$

the inequality holding because T is a positive operator.
Definiteness: certainly $\beta(0, 0) = 0$. Conversely, suppose $\beta(v, v) = 0$. Because T is positive, by 7.38 (the equivalence of (a) and (d)) there exists a positive operator $R \in \mathcal{L}(V)$ with $R^2 = T$; in particular R is self-adjoint. Then

$$0 = \beta(v, v) = \langle Tv, v\rangle = \langle R^2 v, v\rangle = \langle Rv, R^* v\rangle = \langle Rv, Rv\rangle = \|Rv\|^2.$$

Hence $Rv = 0$, and therefore $Tv = R(Rv) = 0$. Since T is invertible, T is injective, so $v = 0$.

Thus β is symmetric, positive, and definite, so β is an inner product on V.
Now suppose conversely that β is an inner product on V.
T is self-adjoint: for all $u, v \in V$, symmetry of β and symmetry of the real inner product give

$$\langle u, Tv\rangle = \beta(u,v) = \beta(v,u) = \langle v, Tu\rangle = \langle Tu, v\rangle = \langle u, T^*v\rangle.$$

Hence $\langle u,\ Tv - T^*v\rangle = 0$ for all $u \in V$; taking $u = Tv - T^*v$ shows $Tv = T^*v$. As $v \in V$ was arbitrary, $T^* = T$.
T is positive: for $v \in V$,

$$\langle Tv, v\rangle = \langle v, Tv\rangle = \beta(v,v) \geq 0$$

by positivity of the inner product β. Combined with self-adjointness, T is a positive operator by 7.34.
T is invertible: suppose $v \in \operatorname{null} T$. Then

$$\beta(v,v) = \langle v, Tv\rangle = \langle v, 0\rangle = 0,$$

so $v = 0$ by definiteness of the inner product β. Thus $\operatorname{null} T = \{0\}$, so T is injective; because V is finite-dimensional, T is therefore invertible (by 3.65, injectivity is equivalent to surjectivity for operators on a finite-dimensional space).
Hence T is an invertible positive operator, completing the proof of both directions.
Remark: since a positive operator has only nonnegative eigenvalues (7.38), the condition "invertible and positive" is exactly the condition that all eigenvalues of the self-adjoint operator T are strictly positive, i.e. that T is what is usually called positive definite.

Problem 9A.6. Prove or give a counterexample: If ρ is a symmetric bilinear form on V, then

$$\{v \in V : \rho(v,v) = 0\}$$

is a subspace of V.

Solution. The statement is false. Here is a counterexample.
Take $\mathbf{F} = \mathbf{R}$ and $V = \mathbf{R}^2$, and define $\rho\colon \mathbf{R}^2 \times \mathbf{R}^2 \to \mathbf{R}$ by

$$\rho\big((x_1,x_2),(y_1,y_2)\big) = x_1y_1 - x_2y_2.$$

First, ρ is a bilinear form: for fixed (y_1, y_2) the map $(x_1, x_2) \mapsto x_1y_1 - x_2y_2$ is a linear functional on $\mathbf{R}^2$, and similarly with the roles of the two slots reversed. It is symmetric because

$$\rho\big((y_1,y_2),(x_1,x_2)\big) = y_1x_1 - y_2x_2 = x_1y_1 - x_2y_2 = \rho\big((x_1,x_2),(y_1,y_2)\big).$$

Equivalently, the matrix of ρ with respect to the standard basis is the symmetric matrix

$$\mathcal{M}(\rho) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

so ρ is symmetric by 9.12.
Now

$$\{v \in \mathbf{R}^2 : \rho(v,v) = 0\} = \{(x_1,x_2) \in \mathbf{R}^2 : x_1^2 = x_2^2\} = \{(x_1,x_2) : x_2 = x_1\} \cup \{(x_1,x_2) : x_2 = -x_1\},$$

the union of the two diagonal lines in the plane.
This set is not a subspace of $\mathbf{R}^2$. Indeed, $(1,1)$ and $(1,-1)$ both belong to it, since

$$\rho\big((1,1),(1,1)\big) = 1 - 1 = 0 \qquad \text{and} \qquad \rho\big((1,-1),(1,-1)\big) = 1 - 1 = 0,$$

but their sum $(1,1) + (1,-1) = (2,0)$ does not, since

$$\rho\big((2,0),(2,0)\big) = 4 - 0 = 4 \neq 0.$$

So the set is not closed under addition, hence is not a subspace.
Two comments on what does survive. The set $\{v \in V : \rho(v,v) = 0\}$ always contains 0 and is always closed under scalar multiplication, because $\rho(\lambda v, \lambda v) = \lambda^2 \rho(v,v)$ by bilinearity; closure under addition is the only property that can fail. And it can fail only through the cross term: by symmetry,

$$\rho(u+w, u+w) = \rho(u,u) + 2\rho(u,w) + \rho(w,w),$$

so if u and w lie in the set, then $u+w$ lies in the set precisely when $\rho(u,w) = 0$. In the counterexample above, $\rho\big((1,1),(1,-1)\big) = 1 + 1 = 2 \neq 0$.
Finally, note the contrast with alternating bilinear forms: if α is alternating, then by definition (9.14) the corresponding set is all of V, which is of course a subspace. So the set can fail to be a subspace only when ρ is not alternating, meaning only when $\rho(v,v) \neq 0$ for some $v \in V$.

Problem 9A.7. Explain why the proof of 9.13 (diagonalization of a symmetric bilinear form by an orthonormal basis on a real inner product space) fails if the hypothesis that $\mathbf{F} = \mathbf{R}$ is dropped.

Solution. Recall the proof of 9.13. One starts with an orthonormal basis $f_1, \dots, f_n$ of V and sets $B = \mathcal{M}(\rho, (f_1, \dots, f_n))$, which is a symmetric matrix by 9.12. One then lets $T \in \mathcal{L}(V)$ be the operator with $\mathcal{M}(T, (f_1, \dots, f_n)) = B$ and asserts that T is self-adjoint. The real spectral theorem (7.29) then produces an orthonormal basis $e_1, \dots, e_n$ with respect to which T has a diagonal matrix. With $C = \mathcal{M}(I, (e_1, \dots, e_n), (f_1, \dots, f_n))$, one has $C^{-1}BC = \mathcal{M}(T, (e_1, \dots, e_n))$ diagonal by 3.84, and one finishes with

$$\mathcal{M}(\rho, (e_1, \dots, e_n)) = C^t B C = C^{-1} B C,$$

the first equality by 9.7 and the second because C is a unitary matrix with real entries, so $C^{-1} = C^t$ (7.57).
Three separate steps of this argument use $\mathbf{F} = \mathbf{R}$, and each of them breaks over $\mathbf{C}$.
(1) The passage from "B is symmetric" to "T is self-adjoint" fails. With respect to an orthonormal basis, the matrix of T^* is the conjugate transpose of the matrix of T (7.9), not the transpose. So T is self-adjoint if and only if $B = \overline{B^t}$, i.e. if and only if B is Hermitian. A symmetric complex matrix satisfies $B = B^t$, and this coincides with $B = \overline{B^t}$ only when all entries of B are real. For a concrete instance, let $V = \mathbf{C}$ with its usual inner product, let $f_1 = 1$, and let $\rho(u,w) = i\,uw$. Then ρ is a symmetric bilinear form on V and $B = \big(i\big)$, which is symmetric; but the associated operator T is multiplication by i, whose adjoint is multiplication by $-i$. Thus T is not self-adjoint, and the proof stops here.
(2) The real spectral theorem (7.29) is stated only for $\mathbf{F} = \mathbf{R}$, so it is unavailable in any case. One might hope to substitute the complex spectral theorem (7.31), but that requires T to be normal, and the operator T built from a complex symmetric matrix B need not be normal. For an explicit instance, let $V = \mathbf{C}^2$ with its usual inner product and standard (orthonormal) basis, and let ρ be the symmetric bilinear form whose matrix is

$$B = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix},$$

that is, $\rho\big((u_1,u_2),(w_1,w_2)\big) = u_1w_1 + iu_1w_2 + iu_2w_1 - u_2w_2$. The operator T with $\mathcal{M}(T) = B$ has

$$B^*B = \begin{pmatrix} 2 & 2i \\ -2i & 2 \end{pmatrix} \qquad \text{and} \qquad BB^* = \begin{pmatrix} 2 & -2i \\ 2i & 2 \end{pmatrix},$$

so $T^*T \neq TT^*$ and T is not normal. Hence neither spectral theorem applies. The structural obstruction is that symmetry of a matrix is a statement about the transpose, whereas both spectral theorems are statements about the conjugate transpose.
(3) Even if one somehow obtained an orthonormal basis $e_1, \dots, e_n$ diagonalizing the operator T, the final step would still collapse. Changing the basis for a bilinear form conjugates its matrix by C^t on the left (9.7), while changing the basis for an operator conjugates its matrix by C^{-1} on the left (3.84). The whole point of the real proof is that these two operations agree, because a change of basis between two orthonormal bases of a real inner product space is given by a unitary matrix with real entries, for

which $C^{-1} = C^t$ (7.57). Over $\mathbf{C}$, a unitary C satisfies $C^{-1} = C^* = \overline{C^t}$, so C^tBC and $C^{-1}BC$ are in general different matrices, and diagonalizing one says nothing about the other. In the language of matrices: over $\mathbf{R}$ with orthogonal C, congruence and similarity coincide; over $\mathbf{C}$ with unitary C they do not.

A word of caution about what this does and does not establish. The exercise asks only why the proof fails, and the three points above answer that. Failure of this particular proof does not by itself show that the conclusion of 9.13 is false over $\mathbf{C}$, and in fact the conclusion remains true there, although its proof requires a tool not developed in this book, namely Takagi's factorization: every complex symmetric matrix B can be written $B = U\Sigma U^t$ with U unitary and Σ diagonal with nonnegative real entries. Granting that, put $C = \overline{U}$, which is again unitary. Then

$$C^tBC = U^*\big(U\Sigma U^t\big)\overline{U} = \Sigma\, U^t\overline{U} = \Sigma,$$

because $U^t\overline{U} = \overline{U^*U} = \overline{I} = I$. Defining $e_k = \sum_{j=1}^{n} C_{j,k} f_j$ makes $e_1, \dots, e_n$ an orthonormal basis of V (the columns of a unitary matrix are orthonormal, and $f_1, \dots, f_n$ is orthonormal) with $C = \mathcal{M}(I, (e_1, \dots, e_n), (f_1, \dots, f_n))$, so $\mathcal{M}(\rho, (e_1, \dots, e_n)) = C^tBC = \Sigma$ is diagonal by 9.7. So the correct summary is: the argument given for 9.13 is genuinely a real-inner-product-space argument at each of its three key steps, and a complex proof must proceed along entirely different lines.

Problem 9A.8. Find formulas for $\dim V_{\text{sym}}^{(2)}$ and $\dim V_{\text{alt}}^{(2)}$ in terms of $\dim V$.

Solution. Let $n = \dim V$. We show

$$\dim V_{\text{sym}}^{(2)} = \frac{n(n+1)}{2}, \qquad \dim V_{\text{alt}}^{(2)} = \frac{n(n-1)}{2}.$$

Fix a basis $e_1, \dots, e_n$ of V and let $\beta \mapsto \mathcal{M}(\beta)$ denote the matrix map with respect to this basis, which by 9.5 is an isomorphism of $V^{(2)}$ onto $\mathbf{F}^{n,n}$. Since an isomorphism carries a subspace onto a subspace of the same dimension, it suffices to identify the images of $V_{\text{sym}}^{(2)}$ and $V_{\text{alt}}^{(2)}$ under this map and to compute the dimensions of those matrix spaces.

Step 1: the images are the symmetric and the skew-symmetric matrices.

By 9.12, a bilinear form ρ on V is symmetric if and only if $\mathcal{M}(\rho)$ is a symmetric matrix: if ρ is symmetric then its matrix with respect to every basis is symmetric (implication (a) $\Rightarrow$ (b) there), and if its matrix with respect to our fixed basis is symmetric then ρ is symmetric (implication (c) $\Rightarrow$ (a) there). Hence $\beta \mapsto \mathcal{M}(\beta)$ maps $V_{\text{sym}}^{(2)}$ onto

$$\mathcal{S} = \{A \in \mathbf{F}^{n,n} : A_{j,k} = A_{k,j} \text{ for all } j, k\}.$$

Now consider the alternating forms. If $\alpha \in V_{\text{alt}}^{(2)}$, then 9.16 gives $\alpha(u, w) = -\alpha(w, u)$ for all $u, w \in V$; taking $u = e_j$ and $w = e_k$ shows $\mathcal{M}(\alpha)_{j,k} = -\mathcal{M}(\alpha)_{k,j}$, so $\mathcal{M}(\alpha)$ is skew-symmetric. Conversely, suppose $\alpha \in V^{(2)}$ and $A = \mathcal{M}(\alpha)$ satisfies $A_{j,k} = -A_{k,j}$ for all j, k. If $v = x_1e_1 + \cdots + x_ne_n \in V$, then bilinearity gives

$$\alpha(v, v) = \sum_{k=1}^{n}\sum_{j=1}^{n} x_j x_k A_{j,k}.$$

Interchanging the names of the summation indices j and k in this double sum and then using $A_{k,j} = -A_{j,k}$ shows that the sum equals its own negative; because $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$ we have $2 \neq 0$, and hence $\alpha(v, v) = 0$. Thus α is alternating. Hence $\beta \mapsto \mathcal{M}(\beta)$ maps $V_{\text{alt}}^{(2)}$ onto

$$\mathcal{A} = \{A \in \mathbf{F}^{n,n} : A_{j,k} = -A_{k,j} \text{ for all } j, k\}.$$

Step 2: dimensions of $\mathcal{S}$ and $\mathcal{A}$.

For $j, k \in \{1, \dots, n\}$ let $E_{j,k} \in \mathbf{F}^{n,n}$ be the matrix with 1 in row j, column k, and 0 elsewhere; the list of all $E_{j,k}$ is the standard basis of $\mathbf{F}^{n,n}$.

Consider the list consisting of $E_{j,j}$ for $j \in \{1, \dots, n\}$ together with $E_{j,k} + E_{k,j}$ for all pairs $j < k$. Each of these matrices lies in $\mathcal{S}$. They span $\mathcal{S}$: if $A \in \mathcal{S}$, then

$$A = \sum_{j=1}^{n} A_{j,j} E_{j,j} + \sum_{j<k} A_{j,k}(E_{j,k} + E_{k,j}),$$

because the right side has the same entry as A in every position (using $A_{k,j} = A_{j,k}$). They are linearly independent: if a linear combination with coefficients c_j (for the diagonal matrices) and $c_{j,k}$ (for $j < k$) equals 0, then looking at the entry in row j, column j gives $c_j = 0$, and looking at the entry in row j, column k with $j < k$ gives $c_{j,k} = 0$. The number of matrices in this basis is n (the diagonal ones) plus the number of pairs $j < k$, which is $\binom{n}{2} = n(n-1)/2$. Hence

$$\dim \mathcal{S} = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}.$$

Now consider the list $E_{j,k} - E_{k,j}$ for all pairs $j < k$; each such matrix lies in $\mathcal{A}$. If $A \in \mathcal{A}$, then $A_{j,j} = -A_{j,j}$ forces $A_{j,j} = 0$ for each j (here we use that $\mathbf{F}$ is $\mathbf{R}$ or $\mathbf{C}$, so $2 \neq 0$), and therefore

$$A = \sum_{j<k} A_{j,k}(E_{j,k} - E_{k,j}),$$

since both sides agree in every entry (using $A_{k,j} = -A_{j,k}$ and $A_{j,j} = 0$). Linear independence follows by looking at the entry in row j, column k for $j < k$, exactly as above. Hence

$$\dim \mathcal{A} = \frac{n(n-1)}{2}.$$

Step 3: conclusion.
Because $\beta \mapsto \mathcal{M}(\beta)$ is an isomorphism of $V^{(2)}$ onto $\mathbf{F}^{n,n}$ carrying $V^{(2)}_{\text{sym}}$ onto $\mathcal{S}$ and $V^{(2)}_{\text{alt}}$ onto $\mathcal{A}$, its restrictions to those subspaces are isomorphisms onto $\mathcal{S}$ and $\mathcal{A}$. Therefore

$$\dim V^{(2)}_{\text{sym}} = \frac{(\dim V)(\dim V + 1)}{2}, \qquad \dim V^{(2)}_{\text{alt}} = \frac{(\dim V)(\dim V - 1)}{2}.$$

As a check, these two numbers add up to n^2, which agrees with 9.17 (stating that $V^{(2)} = V^{(2)}_{\text{sym}} \oplus V^{(2)}_{\text{alt}}$) combined with 9.5 (stating that $\dim V^{(2)} = (\dim V)^2$). Indeed, the direct sum decomposition together with the value of $\dim V^{(2)}_{\text{sym}}$ computed above gives a second derivation of $\dim V^{(2)}_{\text{alt}} = n^2 - n(n+1)/2 = n(n-1)/2$.

Problem 9A.9. Suppose that n is a positive integer and $V = \{p \in \mathcal{P}_n(\mathbf{R}) : p(0) = p(1)\}$. Define $\alpha : V \times V \to \mathbf{R}$ by

$$\alpha(p, q) = \int_0^1 pq'.$$

Show that α is an alternating bilinear form on V.

Solution. First note that V is indeed a vector space: the map $\varphi : \mathcal{P}_n(\mathbf{R}) \to \mathbf{R}$ defined by $\varphi(p) = p(1) - p(0)$ is a linear functional on $\mathcal{P}_n(\mathbf{R})$, and $V = \operatorname{null} \varphi$, so V is a subspace of $\mathcal{P}_n(\mathbf{R})$ and hence a finite-dimensional real vector space. Also, α is well defined: if $p, q \in V$, then pq' is a polynomial, so it is continuous on $[0, 1]$ and $\int_0^1 pq'$ is a real number.
Bilinearity. Fix $q \in V$. If $p_1, p_2 \in V$ and $\lambda \in \mathbf{R}$, then $(p_1 + \lambda p_2)q' = p_1 q' + \lambda p_2 q'$ pointwise, so linearity of the integral gives

$$\alpha(p_1 + \lambda p_2, q) = \int_0^1 (p_1 + \lambda p_2)q' = \int_0^1 p_1 q' + \lambda \int_0^1 p_2 q' = \alpha(p_1, q) + \lambda \alpha(p_2, q).$$

Thus $p \mapsto \alpha(p, q)$ is linear on V for each fixed $q \in V$. Now fix $p \in V$. Differentiation is linear, so

$(q_1 + \lambda q_2)' = q_1' + \lambda q_2'$ for $q_1, q_2 \in V$ and $\lambda \in \mathbf{R}$; hence, again by linearity of the integral,

$$\alpha(p, q_1 + \lambda q_2) = \int_0^1 p(q_1' + \lambda q_2') = \int_0^1 pq_1' + \lambda \int_0^1 pq_2' = \alpha(p, q_1) + \lambda\alpha(p, q_2).$$

Thus $q \mapsto \alpha(p, q)$ is linear on V for each fixed $p \in V$. Hence α is a bilinear form on V (as defined in 9.1).
Alternating. Suppose $p \in V$. The product rule gives $(p^2)' = 2pp'$, so $pp' = \frac{1}{2}(p^2)'$. Because p^2 is a polynomial, it is differentiable with continuous derivative on $[0, 1]$, so the fundamental theorem of calculus applies:

$$\alpha(p, p) = \int_0^1 pp' = \frac{1}{2}\int_0^1 (p^2)' = \frac{p(1)^2 - p(0)^2}{2} = 0,$$

where the last equality holds because $p \in V$ means $p(0) = p(1)$.
Thus $\alpha(p, p) = 0$ for every $p \in V$, which is exactly the definition (9.14) of an alternating bilinear form. Hence α is an alternating bilinear form on V.
Remark. By 9.16, this is equivalent to the statement that $\alpha(p, q) = -\alpha(q, p)$ for all $p, q \in V$, which can also be seen directly by integration by parts: since $(pq)' = p'q + pq'$, the fundamental theorem of calculus gives

$$\alpha(p, q) + \alpha(q, p) = \int_0^1 (pq' + p'q) = \int_0^1 (pq)' = p(1)q(1) - p(0)q(0) = 0,$$

where the last equality uses $p(0) = p(1)$ and $q(0) = q(1)$.

Problem 9A.10. Suppose that n is a positive integer and

$$V = \{p \in \mathcal{P}_n(\mathbf{R}) : p(0) = p(1) \text{ and } p'(0) = p'(1)\}.$$

Define $\rho : V \times V \to \mathbf{R}$ by

$$\rho(p, q) = \int_0^1 pq''.$$

Show that ρ is a symmetric bilinear form on V.

Solution. First note that V is a vector space: the map $\Phi : \mathcal{P}_n(\mathbf{R}) \to \mathbf{R}^2$ defined by

$$\Phi(p) = \big(p(1) - p(0),\ p'(1) - p'(0)\big)$$

is linear (each coordinate is a composition of the linear map $p \mapsto p$ or $p \mapsto p'$ with evaluation at a point, and differences of linear functionals are linear), and $V = \operatorname{null}\Phi$. Hence V is a subspace of $\mathcal{P}_n(\mathbf{R})$ and therefore a finite-dimensional real vector space. Also ρ is well defined: for $p, q \in V$ the function pq'' is a polynomial, hence continuous on $[0, 1]$, so $\int_0^1 pq''$ is a real number.
Bilinearity. Fix $q \in V$. For $p_1, p_2 \in V$ and $\lambda \in \mathbf{R}$ we have $(p_1 + \lambda p_2)q'' = p_1q'' + \lambda p_2q''$ pointwise, so linearity of the integral gives

$$\rho(p_1 + \lambda p_2, q) = \int_0^1 (p_1q'' + \lambda p_2q'') = \rho(p_1, q) + \lambda\rho(p_2, q).$$

Thus $p \mapsto \rho(p, q)$ is linear on V. Now fix $p \in V$. Because differentiation is linear, so is the map $q \mapsto q''$; hence for $q_1, q_2 \in V$ and $\lambda \in \mathbf{R}$ we have $(q_1 + \lambda q_2)'' = q_1'' + \lambda q_2''$, and therefore

$$\rho(p, q_1 + \lambda q_2) = \int_0^1 p(q_1'' + \lambda q_2'') = \rho(p, q_1) + \lambda\rho(p, q_2).$$

Thus $q \mapsto \rho(p, q)$ is linear on V. Hence ρ is a bilinear form on V (as defined in 9.1).
Symmetry. Suppose $p, q \in V$. The product rule gives

$$(pq')' = p'q' + pq''.$$

The function pq' is a polynomial, so it is differentiable with continuous derivative on $[0,1]$, and the fundamental theorem of calculus yields

$$\int_0^1 (p'q' + pq'') = \int_0^1 (pq')' = p(1)q'(1) - p(0)q'(0).$$

Because $p \in V$ we have $p(0) = p(1)$, and because $q \in V$ we have $q'(0) = q'(1)$; hence

$$p(1)q'(1) - p(0)q'(0) = p(1)q'(1) - p(1)q'(1) = 0.$$

Therefore, using linearity of the integral,

$$\rho(p,q) = \int_0^1 pq'' = -\int_0^1 p'q'.$$

The right side of this last equation is unchanged if p and q are interchanged, since $\int_0^1 p'q' = \int_0^1 q'p'$. Applying the displayed formula with the roles of p and q reversed (which is legitimate because $q \in V$ and $p \in V$, so $q(0) = q(1)$ and $p'(0) = p'(1)$ as needed) gives

$$\rho(q,p) = -\int_0^1 q'p' = -\int_0^1 p'q' = \rho(p,q).$$

Thus $\rho(p,q) = \rho(q,p)$ for all $p, q \in V$, which is the definition (9.9) of a symmetric bilinear form. Hence ρ is a symmetric bilinear form on V.
Remark. The computation above shows more: $\rho(p,q) = -\int_0^1 p'q'$ for all $p, q \in V$, and hence the associated quadratic form (see 9.18) satisfies $\rho(p,p) = -\int_0^1 (p')^2 \le 0$, with equality if and only if $p' = 0$ on $[0,1]$, that is, if and only if p is a constant polynomial. So $-\rho$ is a positive semidefinite symmetric bilinear form whose null directions are exactly the constant polynomials.

9.2 Exercises 9B

Problem 9B.1. Suppose m is a positive integer. Show that $\dim V^{(m)} = (\dim V)^m$.

Solution. Let $n = \dim V$ and fix a basis $e_1, \dots, e_n$ of V. Let $\varphi_1, \dots, \varphi_n \in V'$ be the dual basis, so that

$$v = \sum_{j=1}^{n} \varphi_j(v)\, e_j \quad \text{for every } v \in V$$

(see 3.114) and $\varphi_j(e_k)$ equals 1 if $j = k$ and equals 0 if $j \ne k$.
For each list $(j_1, \dots, j_m)$ of elements of $\{1, \dots, n\}$, define $\beta_{j_1,\dots,j_m} : V^m \to \mathbf{F}$ by

$$\beta_{j_1,\dots,j_m}(v_1, \dots, v_m) = \varphi_{j_1}(v_1) \cdots \varphi_{j_m}(v_m).$$

First, each $\beta_{j_1,\dots,j_m}$ is an m-linear form on V. Indeed, fix $k \in \{1, \dots, m\}$ and fix $u_1, \dots, u_m \in V$. Then

$$v \mapsto \beta_{j_1,\dots,j_m}(u_1, \dots, u_{k-1}, v, u_{k+1}, \dots, u_m) = c\, \varphi_{j_k}(v),$$

where $c = \prod_{p \ne k} \varphi_{j_p}(u_p) \in \mathbf{F}$ does not depend on v. A scalar multiple of the linear functional φ_{j_k} is linear, so $\beta_{j_1,\dots,j_m}$ is linear in each slot when the other slots are held fixed. Thus $\beta_{j_1,\dots,j_m} \in V^{(m)}$.
We now show that the list of the n^m forms $\beta_{j_1,\dots,j_m}$, as $(j_1, \dots, j_m)$ ranges over all lists of length m of elements of $\{1, \dots, n\}$, is a basis of the vector space $V^{(m)}$.
Spanning. Suppose $\beta \in V^{(m)}$ and $v_1, \dots, v_m \in V$. Writing each $v_k = \sum_{j=1}^{n} \varphi_j(v_k) e_j$ and using linearity

of β in the first slot, then in the second slot, and so on through the m-th slot, we get

$$\begin{aligned}\beta(v_1,\dots,v_m) &= \beta\Big(\sum_{j_1=1}^{n}\varphi_{j_1}(v_1)e_{j_1},\ \dots,\ \sum_{j_m=1}^{n}\varphi_{j_m}(v_m)e_{j_m}\Big)\\ &= \sum_{j_1=1}^{n}\cdots\sum_{j_m=1}^{n}\varphi_{j_1}(v_1)\cdots\varphi_{j_m}(v_m)\,\beta(e_{j_1},\dots,e_{j_m})\\ &= \sum_{j_1=1}^{n}\cdots\sum_{j_m=1}^{n}\beta(e_{j_1},\dots,e_{j_m})\,\beta_{j_1,\dots,j_m}(v_1,\dots,v_m).\end{aligned}$$

Because $v_1,\dots,v_m\in V$ were arbitrary, this says

$$\beta=\sum_{j_1=1}^{n}\cdots\sum_{j_m=1}^{n}\beta(e_{j_1},\dots,e_{j_m})\,\beta_{j_1,\dots,j_m}.$$

Hence our list of forms spans $V^{(m)}$.
Linear independence. Suppose scalars $c_{j_1,\dots,j_m}\in\mathbf{F}$ satisfy

$$\sum_{j_1=1}^{n}\cdots\sum_{j_m=1}^{n}c_{j_1,\dots,j_m}\,\beta_{j_1,\dots,j_m}=0.$$

Fix a list $(k_1,\dots,k_m)$ of elements of $\{1,\dots,n\}$ and evaluate the left side at $(e_{k_1},\dots,e_{k_m})$. Because

$$\beta_{j_1,\dots,j_m}(e_{k_1},\dots,e_{k_m})=\varphi_{j_1}(e_{k_1})\cdots\varphi_{j_m}(e_{k_m})$$

equals 1 if $j_p=k_p$ for every $p\in\{1,\dots,m\}$ and equals 0 otherwise, only the single term indexed by $(k_1,\dots,k_m)$ survives. Thus $c_{k_1,\dots,k_m}=0$. As $(k_1,\dots,k_m)$ was arbitrary, all the coefficients are 0, proving linear independence.
Therefore the forms $\beta_{j_1,\dots,j_m}$ constitute a basis of $V^{(m)}$. The number of lists $(j_1,\dots,j_m)$ of length m with entries in the n-element set $\{1,\dots,n\}$ is n^m, and distinct index lists give distinct basis vectors (they are linearly independent, hence pairwise distinct). Hence

$$\dim V^{(m)}=n^m=(\dim V)^m,$$

as desired.
The degenerate case $V=\{0\}$ is covered by the same argument: there the index set is empty, the empty spanning sum shows that every m-linear form on V is 0, and $\dim V^{(m)}=0=0^m$.

Problem 9B.2. Suppose $n\ge 3$ and $\alpha\colon\mathbf{F}^n\times\mathbf{F}^n\times\mathbf{F}^n\to\mathbf{F}$ is defined by

$$\begin{aligned}&\alpha\big((x_1,\dots,x_n),(y_1,\dots,y_n),(z_1,\dots,z_n)\big)\\ &\qquad=x_1y_2z_3-x_2y_1z_3-x_3y_2z_1-x_1y_3z_2+x_3y_1z_2+x_2y_3z_1.\end{aligned}$$

Show that α is an alternating 3-linear form on $\mathbf{F}^n$.

Solution. For $j\in\{1,\dots,n\}$ let $\varphi_j\colon\mathbf{F}^n\to\mathbf{F}$ be the j-th coordinate functional, defined by $\varphi_j(x_1,\dots,x_n)=x_j$. Each φ_j is a linear functional on $\mathbf{F}^n$.
The key observation is that the six terms defining α are exactly the six terms indexed by perm 3, each weighted by the sign of its permutation. Indeed, listing the elements of perm 3 together with their signs as computed from 9.32:

$$\begin{aligned}&\operatorname{sign}(1,2,3)=1, &&\operatorname{sign}(2,1,3)=-1, &&\operatorname{sign}(3,2,1)=-1,\\ &\operatorname{sign}(1,3,2)=-1, &&\operatorname{sign}(3,1,2)=1, &&\operatorname{sign}(2,3,1)=1.\end{aligned}$$

(For example, in $(3,1,2)$ the pairs (k,ℓ) with $k<\ell$ such that k appears after ℓ are $(1,3)$ and $(2,3)$, giving $(-1)^2=1$; in $(3,2,1)$ the three pairs $(1,2)$, $(1,3)$, $(2,3)$ all appear reversed, giving $(-1)^3=-1$.) Matching these against the displayed formula term by term, we obtain

$$\alpha(u,v,w)=\sum_{(j_1,j_2,j_3)\in\operatorname{perm}3}\big(\operatorname{sign}(j_1,j_2,j_3)\big)\varphi_{j_1}(u)\,\varphi_{j_2}(v)\,\varphi_{j_3}(w)$$

for all $u,v,w\in\mathbf{F}^n$: the permutation $(1,2,3)$ contributes $x_1y_2z_3$, the permutation $(2,1,3)$ contributes $-x_2y_1z_3$, the permutation $(3,2,1)$ contributes $-x_3y_2z_1$, the permutation $(1,3,2)$ contributes $-x_1y_3z_2$, the permutation $(3,1,2)$ contributes $+x_3y_1z_2$, and the permutation $(2,3,1)$ contributes $+x_2y_3z_1$, where $u=(x_1,\dots,x_n)$, $v=(y_1,\dots,y_n)$, $w=(z_1,\dots,z_n)$.
3-linearity. Fix $v,w\in\mathbf{F}^n$. By the formula above, the function

$$u\mapsto\alpha(u,v,w)=\sum_{(j_1,j_2,j_3)\in\operatorname{perm}3}\Big(\big(\operatorname{sign}(j_1,j_2,j_3)\big)\varphi_{j_2}(v)\varphi_{j_3}(w)\Big)\varphi_{j_1}(u)$$

is a linear combination of the linear functionals $\varphi_1,\varphi_2,\varphi_3$ with coefficients not depending on u, hence is a linear map from $\mathbf{F}^n$ to $\mathbf{F}$. The same argument applies with the second or third slot free instead of the first. Thus α is linear in each slot when the other two slots are held fixed, so α is a 3-linear form on $\mathbf{F}^n$.
Alternating. By 9.27 we must show $\alpha(u,v,w)=0$ whenever two of the three vectors u,v,w are equal. Suppose first that $u=v$. Define a map on $\operatorname{perm}3$ by

$$(j_1,j_2,j_3)\mapsto(j_2,j_1,j_3),$$

that is, swap the entries in the first two slots of the permutation. This map sends $\operatorname{perm}3$ into itself, it is its own inverse, and it has no fixed point (since $j_1\neq j_2$ always). Hence it partitions $\operatorname{perm}3$ into three two-element pairs. By 9.34, $\operatorname{sign}(j_2,j_1,j_3)=-\operatorname{sign}(j_1,j_2,j_3)$. Because $u=v$,

$$\varphi_{j_2}(u)\,\varphi_{j_1}(v)\,\varphi_{j_3}(w)=\varphi_{j_2}(u)\,\varphi_{j_1}(u)\,\varphi_{j_3}(w)=\varphi_{j_1}(u)\,\varphi_{j_2}(v)\,\varphi_{j_3}(w).$$

Thus the two terms of the displayed sum coming from a paired couple (j_1,j_2,j_3) and (j_2,j_1,j_3) have equal magnitude and opposite sign, so they cancel. Summing over the three pairs gives $\alpha(u,v,w)=0$. If $v=w$, use instead the fixed-point-free sign-reversing involution $(j_1,j_2,j_3)\mapsto(j_1,j_3,j_2)$; if $u=w$, use $(j_1,j_2,j_3)\mapsto(j_3,j_2,j_1)$. In each case the same computation shows that the paired terms cancel, giving $\alpha(u,v,w)=0$.
Hence α is an alternating 3-linear form on $\mathbf{F}^n$.
For a concrete check of the first case, take $u=v=(x_1,\dots,x_n)$ directly in the defining formula:

$$x_1x_2z_3-x_2x_1z_3-x_3x_2z_1-x_1x_3z_2+x_3x_1z_2+x_2x_3z_1=0,$$

since the first two terms cancel, the third cancels the sixth, and the fourth cancels the fifth.

Problem 9B.3. Suppose m is a positive integer and α is an m-linear form on V such that $\alpha(v_1,\dots,v_m)=0$ whenever $v_1,\dots,v_m$ is a list of vectors in V with $v_j=v_{j+1}$ for some $j\in\{1,\dots,m-1\}$. Prove that α is an alternating m-linear form on V.

Solution. If $m=1$ there is nothing to prove, because the alternating condition in 9.27 is vacuous when $m=1$ (there are no two distinct indices in $\{1\}$). So assume $m\geq 2$.
Step 1: swapping the vectors in two adjacent slots reverses the sign of α.
Fix $j\in\{1,\dots,m-1\}$, fix vectors $v_i\in V$ for $i\notin\{j,j+1\}$, and let $u,w\in V$. Write

$$\alpha(\dots,a,b,\dots)$$

for the value of α on the list whose j-th entry is a, whose $(j+1)$-st entry is b, and whose other entries

are the fixed vectors v_i. Putting $u+w$ in both slots j and $j+1$, the hypothesis gives

$$0 = \alpha(\dots, u+w, u+w, \dots).$$

Expanding by linearity in slot j and then in slot $j+1$ (which is legitimate because α is m-linear) yields

$$0 = \alpha(\dots, u, u, \dots) + \alpha(\dots, u, w, \dots) + \alpha(\dots, w, u, \dots) + \alpha(\dots, w, w, \dots).$$

The first and last terms on the right are 0 by the hypothesis, because in each of them slots j and $j+1$ hold equal vectors. Hence

$$\alpha(\dots, w, u, \dots) = -\alpha(\dots, u, w, \dots).$$

In other words, interchanging the entries in slots j and $j+1$ multiplies the value of α by -1, for every $j \in \{1, \dots, m-1\}$ and every choice of the entries in the remaining slots.
Step 2: $\alpha(v_1, \dots, v_m) = 0$ whenever $v_j = v_k$ for some $j < k$.
Suppose $v_1, \dots, v_m$ is a list in V with $v_j = v_k$, where $1 \le j < k \le m$. If $k = j+1$, then $\alpha(v_1, \dots, v_m) = 0$ directly by hypothesis.
Now suppose $k > j+1$. Starting from the list $v_1, \dots, v_m$, perform the following $k-j-1$ adjacent interchanges: first swap the entries in slots $k-1$ and k, then swap the entries in slots $k-2$ and $k-1$, and continue until finally swapping the entries in slots $j+1$ and $j+2$. Each of these swaps involves only slots with indices in $\{j+1, \dots, k\}$, so the entry in slot j is never moved, and the effect of the whole sequence is to move the vector v_k leftward from slot k into slot $j+1$, shifting each of $v_{j+1}, \dots, v_{k-1}$ one slot to the right. The resulting list is

$$v_1, \dots, v_{j-1}, v_j, v_k, v_{j+1}, \dots, v_{k-1}, v_{k+1}, \dots, v_m,$$

whose entries in slots j and $j+1$ are v_j and v_k, which are equal by assumption. Hence by hypothesis α of this new list equals 0.
By Step 1, each of the $k-j-1$ swaps multiplied the value of α by -1, so

$$\alpha(v_1, \dots, v_m) = (-1)^{k-j-1}\, \alpha\big(v_1, \dots, v_{j-1}, v_j, v_k, v_{j+1}, \dots, v_{k-1}, v_{k+1}, \dots, v_m\big) = (-1)^{k-j-1} \cdot 0 = 0.$$

Thus $\alpha(v_1, \dots, v_m) = 0$ whenever two entries of the list $v_1, \dots, v_m$ in distinct slots are equal. By 9.27, α is an alternating m-linear form on V.

Problem 9B.4. Prove or give a counterexample: If $\alpha \in V_{\text{alt}}^{(4)}$, then

$$\{(v_1, v_2, v_3, v_4) \in V^4 : \alpha(v_1, v_2, v_3, v_4) = 0\}$$

is a subspace of V^4.

Solution. The statement is false. Below is a counterexample.
Let $V = \mathbf{F}^4$ and let e_1, e_2, e_3, e_4 be the standard basis of $\mathbf{F}^4$. By 9.37 the vector space $V_{\text{alt}}^{(4)}$ has dimension one, so there exists a nonzero $\alpha \in V_{\text{alt}}^{(4)}$; moreover the explicit construction 9.38 in the proof of 9.37, applied with the basis e_1, e_2, e_3, e_4, produces such an α with

$$\alpha(e_1, e_2, e_3, e_4) = 1.$$

(Concretely, $\alpha(v_1, v_2, v_3, v_4) = \sum_{(j_1, j_2, j_3, j_4) \in \operatorname{perm} 4} (\operatorname{sign}(j_1, j_2, j_3, j_4)) \varphi_{j_1}(v_1) \varphi_{j_2}(v_2) \varphi_{j_3}(v_3) \varphi_{j_4}(v_4)$, where φ_j is the j-th coordinate functional on $\mathbf{F}^4$.)
Let

$$E = \{(v_1, v_2, v_3, v_4) \in V^4 : \alpha(v_1, v_2, v_3, v_4) = 0\}$$

and consider the two elements of V^4

$$u = (e_1, e_2, e_3, e_3) \qquad \text{and} \qquad w = (0, 0, 0, e_4).$$

First, $u \in E$: the third and fourth entries of u are equal, so $\alpha(e_1, e_2, e_3, e_3) = 0$ because α is alternating (9.27).
Second, $w \in E$: because α is linear in its first slot, and a linear map sends 0 to 0, we have $\alpha(0, 0, 0, e_4) = 0$.
However $u + w \notin E$. Addition in V^4 is defined coordinatewise, so

$$u + w = (e_1, e_2, e_3, e_3 + e_4),$$

and by linearity of α in the fourth slot,

$$\alpha(e_1, e_2, e_3, e_3 + e_4) = \alpha(e_1, e_2, e_3, e_3) + \alpha(e_1, e_2, e_3, e_4) = 0 + 1 = 1 \neq 0.$$

Thus E is not closed under addition, so E is not a subspace of V^4. This gives the desired counterexample.
Two remarks clarify what is really going on. First, E never fails the other two subspace conditions: it contains the additive identity $(0, 0, 0, 0)$ of V^4, and it is closed under scalar multiplication, since for $c \in \mathbf{F}$,

$$\alpha(cv_1, cv_2, cv_3, cv_4) = c^4 \alpha(v_1, v_2, v_3, v_4)$$

by linearity in each of the four slots, so $\alpha(v_1, v_2, v_3, v_4) = 0$ implies $\alpha(cv_1, cv_2, cv_3, cv_4) = 0$. Only closure under addition can fail.
Second, some hypothesis on $\dim V$ is unavoidable for a counterexample: if $\dim V \leq 3$, then $4 > \dim V$, so 9.29 forces $\alpha = 0$, in which case $E = V^4$, which is indeed a subspace. The construction above works for any V with $\dim V \geq 4$ and any nonzero $\alpha \in V_{\text{alt}}^{(4)}$: choose $e_1, e_2, e_3, e_4 \in V$ with $\alpha(e_1, e_2, e_3, e_4) \neq 0$ (possible since $\alpha \neq 0$) and run the same computation.

Problem 9B.5. Suppose m is a positive integer and β is an m-linear form on V. Define an m-linear form α on V by

$$\alpha(v_1, \ldots, v_m) = \sum_{(j_1, \ldots, j_m) \in \operatorname{perm} m} \big(\operatorname{sign}(j_1, \ldots, j_m)\big) \beta(v_{j_1}, \ldots, v_{j_m})$$

for $v_1, \ldots, v_m \in V$. Explain why $\alpha \in V_{\text{alt}}^{(m)}$.

Solution. Throughout, we identify a permutation $(j_1, \ldots, j_m) \in \operatorname{perm} m$ with the function j from $\{1, \ldots, m\}$ to $\{1, \ldots, m\}$ defined by $j(p) = j_p$. Because the list $(j_1, \ldots, j_m)$ contains each of $1, \ldots, m$ exactly once, this function is a bijection.
First we check that α really is an m-linear form on V (the exercise asserts this, but it is worth verifying). Fix $(j_1, \ldots, j_m) \in \operatorname{perm} m$ and define $\gamma \colon V^m \to \mathbf{F}$ by $\gamma(v_1, \ldots, v_m) = \beta(v_{j_1}, \ldots, v_{j_m})$. Fix $k \in \{1, \ldots, m\}$ and fix $u_1, \ldots, u_m \in V$. Let p be the unique index with $j_p = k$ (unique because j is a bijection). Then

$$v \mapsto \gamma(u_1, \ldots, u_{k-1}, v, u_{k+1}, \ldots, u_m)$$

is the map obtained from β by placing v in slot p of β and holding the entries in the other slots of β fixed (they are among $u_1, \ldots, u_m$, and none of them equals the free vector v because j takes the value k only at p). Since β is m-linear, this map is linear from V to $\mathbf{F}$. Hence $\gamma \in V^{(m)}$. As α is a finite linear combination of such γ's, and $V^{(m)}$ is a vector space under the usual operations on functions, we conclude $\alpha \in V^{(m)}$.
Now we show that α is alternating. If $m = 1$ this is vacuous, so assume $m \geq 2$. Suppose $v_1, \ldots, v_m \in V$ and $v_r = v_s$ where r and s are distinct elements of $\{1, \ldots, m\}$. We must show $\alpha(v_1, \ldots, v_m) = 0$.
Define a map $\Phi \colon \operatorname{perm} m \to \operatorname{perm} m$ as follows. Given $(j_1, \ldots, j_m) \in \operatorname{perm} m$, let p and q be the unique indices with $j_p = r$ and $j_q = s$; note $p \neq q$ because $r \neq s$. Let $\Phi(j_1, \ldots, j_m)$ be the list obtained from $(j_1, \ldots, j_m)$ by interchanging the entries in slots p and q. Thus $\Phi(j_1, \ldots, j_m)$ still contains each of $1, \ldots, m$ exactly once, so it is indeed an element of $\operatorname{perm} m$; it is the list $(j'_1, \ldots, j'_m)$ with $j'_p = s$, $j'_q = r$, and $j'_i = j_i$ for $i \notin \{p, q\}$.

The map Φ has three properties we need.
Φ is an involution with no fixed points. Applying Φ to $(j'_1, \dots, j'_m)$, the entry equal to r sits in slot q and the entry equal to s sits in slot p, so Φ interchanges those same two slots and returns $(j_1, \dots, j_m)$. Hence $\Phi \circ \Phi$ is the identity. Also $\Phi(j_1, \dots, j_m) \neq (j_1, \dots, j_m)$, since $j'_p = s \neq r = j_p$. Therefore Φ partitions $\operatorname{perm} m$ into two-element subsets $\{(j_1, \dots, j_m), \Phi(j_1, \dots, j_m)\}$.
Φ reverses signs. Because $\Phi(j_1, \dots, j_m)$ is obtained from $(j_1, \dots, j_m)$ by swapping two entries, 9.34 gives

$$\operatorname{sign}\big(\Phi(j_1, \dots, j_m)\big) = -\operatorname{sign}(j_1, \dots, j_m).$$

Φ leaves the value of β unchanged. With notation as above,

$$v_{j'_p} = v_s = v_r = v_{j_p}, \qquad v_{j'_q} = v_r = v_s = v_{j_q},$$

using the hypothesis $v_r = v_s$, and $v_{j'_i} = v_{j_i}$ for all other i. Hence the lists $v_{j'_1}, \dots, v_{j'_m}$ and $v_{j_1}, \dots, v_{j_m}$ are identical, so

$$\beta(v_{j'_1}, \dots, v_{j'_m}) = \beta(v_{j_1}, \dots, v_{j_m}).$$

Combining the last two properties, for each $(j_1, \dots, j_m) \in \operatorname{perm} m$ the two summands of $\alpha(v_1, \dots, v_m)$ indexed by $(j_1, \dots, j_m)$ and by $\Phi(j_1, \dots, j_m)$ add to

$$\big(\operatorname{sign}(j_1, \dots, j_m)\big)\beta(v_{j_1}, \dots, v_{j_m}) - \big(\operatorname{sign}(j_1, \dots, j_m)\big)\beta(v_{j_1}, \dots, v_{j_m}) = 0.$$

Because Φ partitions $\operatorname{perm} m$ into such two-element subsets, the entire sum defining $\alpha(v_1, \dots, v_m)$ splits into these cancelling pairs. Hence $\alpha(v_1, \dots, v_m) = 0$.
Since $v_1, \dots, v_m$ was an arbitrary list with two equal entries in distinct slots, 9.27 gives $\alpha \in V_{\text{alt}}^{(m)}$.

Problem 9B.6. Suppose m is a positive integer and β is an m-linear form on V. Define an m-linear form α on V by

$$\alpha(v_1, \dots, v_m) = \sum_{(j_1, \dots, j_m) \in \operatorname{perm} m} \beta(v_{j_1}, \dots, v_{j_m})$$

for $v_1, \dots, v_m \in V$. Explain why

$$\alpha(v_{k_1}, \dots, v_{k_m}) = \alpha(v_1, \dots, v_m)$$

for all $v_1, \dots, v_m \in V$ and all $(k_1, \dots, k_m) \in \operatorname{perm} m$.

Solution. As in the previous exercise, identify a permutation $(j_1, \dots, j_m) \in \operatorname{perm} m$ with the function j from $\{1, \dots, m\}$ to itself given by $j(p) = j_p$. Saying that the list $(j_1, \dots, j_m)$ contains each of $1, \dots, m$ exactly once says precisely that this function is a bijection of $\{1, \dots, m\}$ onto itself. Conversely, every bijection of $\{1, \dots, m\}$ onto itself arises this way from exactly one element of $\operatorname{perm} m$.
Suppose $v_1, \dots, v_m \in V$ and $(k_1, \dots, k_m) \in \operatorname{perm} m$. Set

$$w_p = v_{k_p} \quad \text{for } p \in \{1, \dots, m\}.$$

Then $w_{j_p} = v_{k_{j_p}}$ for all p, and so by the definition of α,

$$\alpha(v_{k_1}, \dots, v_{k_m}) = \alpha(w_1, \dots, w_m) = \sum_{(j_1, \dots, j_m) \in \operatorname{perm} m} \beta(w_{j_1}, \dots, w_{j_m}) = \sum_{(j_1, \dots, j_m) \in \operatorname{perm} m} \beta(v_{k_{j_1}}, \dots, v_{k_{j_m}}).$$

Now define, for each $(j_1, \dots, j_m) \in \operatorname{perm} m$,

$$\Psi(j_1, \dots, j_m) = (k_{j_1}, \dots, k_{j_m}),$$

which is the list corresponding to the composition $k \circ j$ of the two bijections k and j. Two observations finish the proof.
Ψ maps $\operatorname{perm} m$ into $\operatorname{perm} m$. The composition of two bijections of $\{1, \dots, m\}$ onto itself is again a bijection of $\{1, \dots, m\}$ onto itself, so the list $(k_{j_1}, \dots, k_{j_m})$ contains each of $1, \dots, m$ exactly once. Hence $\Psi(j_1, \dots, j_m) \in \operatorname{perm} m$.

Ψ is a bijection of $\operatorname{perm} m$ onto $\operatorname{perm} m$. Suppose $\Psi(j_1,\dots,j_m)=\Psi(j_1',\dots,j_m')$, so $k_{j_p}=k_{j_p'}$ for every p. Because the function k is injective, this forces $j_p=j_p'$ for every p, so $(j_1,\dots,j_m)=(j_1',\dots,j_m')$. Thus Ψ is injective. An injective function from the finite set $\operatorname{perm} m$ to itself is surjective, so Ψ is a bijection of $\operatorname{perm} m$ onto itself. (Alternatively, Ψ is invertible because composing with k^{-1} undoes it.) Therefore, as $(j_1,\dots,j_m)$ runs over $\operatorname{perm} m$, the list $(k_{j_1},\dots,k_{j_m})$ runs over each element of $\operatorname{perm} m$ exactly once. A finite sum does not depend on the order in which its terms are added, and reindexing a finite sum by a bijection of the index set leaves the sum unchanged. Hence

$$\sum_{(j_1,\dots,j_m)\in\operatorname{perm} m}\beta(v_{k_{j_1}},\dots,v_{k_{j_m}})=\sum_{(i_1,\dots,i_m)\in\operatorname{perm} m}\beta(v_{i_1},\dots,v_{i_m})=\alpha(v_1,\dots,v_m).$$

Combining the two displayed chains of equalities gives

$$\alpha(v_{k_1},\dots,v_{k_m})=\alpha(v_1,\dots,v_m),$$

as desired. In words: α is the symmetrization of β, obtained by summing β over all rearrangements of its inputs, and rearranging the inputs of α merely permutes the terms of that sum.

Problem 9B.7. Give an example of a nonzero alternating 2-linear form α on $\mathbf{R}^3$ and a linearly independent list v_1,v_2 in $\mathbf{R}^3$ such that $\alpha(v_1,v_2)=0$.
This exercise shows that 9.39 can fail if the hypothesis that $n=\dim V$ is deleted.

Solution. Define $\alpha\colon \mathbf{R}^3\times\mathbf{R}^3\to\mathbf{R}$ by

$$\alpha\big((x_1,x_2,x_3),(y_1,y_2,y_3)\big)=x_1y_2-x_2y_1,$$

and take

$$v_1=(1,0,0),\qquad v_2=(0,0,1).$$

α is a 2-linear form. Fix $v=(y_1,y_2,y_3)\in\mathbf{R}^3$. Then the function

$$(x_1,x_2,x_3)\mapsto\alpha\big((x_1,x_2,x_3),v\big)=y_2x_1-y_1x_2$$

is a linear functional on $\mathbf{R}^3$, being a linear combination of the first two coordinate functionals with coefficients not depending on (x_1,x_2,x_3). Similarly, for fixed $u=(x_1,x_2,x_3)\in\mathbf{R}^3$, the function

$$(y_1,y_2,y_3)\mapsto\alpha\big(u,(y_1,y_2,y_3)\big)=-x_2y_1+x_1y_2$$

is a linear functional on $\mathbf{R}^3$. So α is linear in each slot when the other slot is held fixed, which by 9.25 means α is a 2-linear form on $\mathbf{R}^3$.
α is alternating. For $m=2$ the condition in 9.27 says exactly that $\alpha(u,u)=0$ for all $u\in\mathbf{R}^3$. Indeed, for $u=(x_1,x_2,x_3)$,

$$\alpha(u,u)=x_1x_2-x_2x_1=0.$$

Hence $\alpha\in(\mathbf{R}^3)^{(2)}_{\text{alt}}$.
α is nonzero. Taking $e_1=(1,0,0)$ and $e_2=(0,1,0)$ gives

$$\alpha(e_1,e_2)=1\cdot1-0\cdot0=1\neq0,$$

so α is not the zero function on $(\mathbf{R}^3)^2$.
The list v_1,v_2 is linearly independent. Suppose $a,b\in\mathbf{R}$ and $av_1+bv_2=0$. Then $(a,0,b)=(0,0,0)$, so $a=b=0$. Hence v_1,v_2 is linearly independent.
Finally,

$$\alpha(v_1,v_2)=\alpha\big((1,0,0),(0,0,1)\big)=1\cdot0-0\cdot0=0.$$

Thus α is a nonzero alternating 2-linear form on $\mathbf{R}^3$ and v_1,v_2 is a linearly independent list in $\mathbf{R}^3$ with $\alpha(v_1,v_2)=0$, as required.

This does not contradict 9.39, whose hypothesis is that the number of slots equals $\dim V$. Here $\dim \mathbf{R}^3 = 3$ but α has only 2 slots, so 9.39 does not apply. Geometrically, $\alpha(u, v)$ is the determinant of the 2×2 matrix formed by the first two coordinates of u and v, that is, the signed area of the parallelogram spanned by the projections of u and v onto the x_1x_2-plane. The vectors v_1 and v_2 are independent in $\mathbf{R}^3$, but v_2 projects to 0 in that plane, so the projected parallelogram is degenerate and $\alpha(v_1, v_2) = 0$.

9.3 Exercises 9C

Problem 9C.1. Prove or give a counterexample: $S, T \in \mathcal{L}(V) \implies \det(S + T) = \det S + \det T$.

Solution. The statement is false whenever $\dim V \geq 2$; here is a counterexample.
Take $V = \mathbf{F}^2$ and $S = T = I$, the identity operator on V. Then $S + T = 2I$, and the second bullet point of 9.42 gives

$$\det(S + T) = \det(2I) = 2^{\dim V} = 2^2 = 4.$$

On the other hand, the first bullet point of 9.42 gives $\det S = \det T = \det I = 1$, so

$$\det S + \det T = 1 + 1 = 2.$$

Since $4 \neq 2$, we have $\det(S + T) \neq \det S + \det T$ for this choice of S and T. Hence the asserted implication fails.
Two remarks help to explain what is going on.
First, the counterexample is not special to dimension two: if $\dim V = n$ and $S = T = I$, then $\det(S + T) = 2^n$ while $\det S + \det T = 2$, and $2^n \neq 2$ for every integer $n \geq 2$.
Second, the hypothesis $\dim V \geq 2$ is genuinely needed. If $\dim V = 1$, then every operator on V is of the form λI for some $\lambda \in \mathbf{F}$, and the second bullet point of 9.42 gives $\det(\lambda I) = \lambda^1 = \lambda$; thus in dimension one the determinant is a linear functional on $\mathcal{L}(V)$ and the displayed identity does hold. The failure in higher dimensions reflects the fact that, by the formula 9.46, the determinant of an n-by-n matrix is a homogeneous polynomial of degree n in the entries: replacing A by $2A$ multiplies $\det A$ by 2^n rather than by 2, which is exactly what the computation above exhibits.

Problem 9C.2. Suppose the first column of a square matrix A consists of all zeros except possibly the first entry $A_{1,1}$. Let B be the matrix obtained from A by deleting the first row and the first column of A. Show that $\det A = A_{1,1} \det B$.

Solution. Let n be the size of A. Since the book defines determinants only for m-by-m matrices with m a positive integer (see 9.43), the hypothesis that B is a matrix forces $n \geq 2$, and we assume this throughout. (If one adopts the usual convention that the determinant of the empty 0-by-0 matrix is 1, then the case $n = 1$ reads $\det A = A_{1,1} \cdot 1$, which is correct.) The hypothesis says

$$A_{j,1} = 0 \quad \text{for all } j \in \{2, \dots, n\},$$

and B is the $(n-1)$-by-$(n-1)$ matrix given by

$$B_{r,s} = A_{r+1,\, s+1} \quad \text{for } r, s \in \{1, \dots, n-1\}.$$

Start from the formula 9.46 for the determinant of a matrix:

$$\det A = \sum_{(j_1, \dots, j_n) \in \operatorname{perm} n} \big(\operatorname{sign}(j_1, \dots, j_n)\big) A_{j_1,1} \cdots A_{j_n,n}.$$

If $(j_1, \dots, j_n) \in \operatorname{perm} n$ has $j_1 \neq 1$, then $A_{j_1,1} = 0$ and that term of the sum vanishes. Hence only the permutations whose first entry equals 1 can contribute, and

$$\det A = \sum_{\substack{(j_1, \dots, j_n) \in \operatorname{perm} n \\ j_1 = 1}} \big(\operatorname{sign}(j_1, \dots, j_n)\big) A_{1,1} A_{j_2,2} \cdots A_{j_n,n}.$$

Now set up a bijection between the index set of this sum and perm$(n-1)$. If $(1, j_2, \dots, j_n) \in \operatorname{perm} n$, then $j_2, \dots, j_n$ is a list containing each of $2, \dots, n$ exactly once, so

$$(k_1, \dots, k_{n-1}) := (j_2 - 1, \dots, j_n - 1)$$

lies in perm$(n-1)$. Conversely, every element of perm$(n-1)$ arises from exactly one such permutation, via $j_{r+1} = k_r + 1$. Thus the correspondence is a bijection.
Next, the signs agree:

$$\operatorname{sign}(1, j_2, \dots, j_n) = \operatorname{sign}(k_1, \dots, k_{n-1}).$$

To see this, recall 9.32: the sign of a permutation is $(-1)^N$, where N counts the pairs of integers (k, ℓ) with $k < \ell$ such that k appears after ℓ in the list. In the list $(1, j_2, \dots, j_n)$, no pair of the form $(1, \ell)$ is counted, because 1 occupies the first slot and hence appears before everything else. So N counts exactly the pairs (k, ℓ) with $2 \le k < \ell \le n$ such that k appears after ℓ in $(j_2, \dots, j_n)$. Subtracting 1 from every entry is an order-preserving relabeling, so such a pair (k, ℓ) is counted for $(1, j_2, \dots, j_n)$ if and only if the pair $(k-1, \ell-1)$ is counted for $(k_1, \dots, k_{n-1})$. The two counts therefore agree, and so do the signs.
Finally, the products agree: for $r \in \{1, \dots, n-1\}$ we have $j_{r+1} = k_r + 1$, so

$$A_{j_{r+1}, r+1} = A_{k_r+1, r+1} = B_{k_r, r}.$$

Hence

$$A_{1,1} A_{j_2,2} \cdots A_{j_n,n} = A_{1,1} B_{k_1,1} \cdots B_{k_{n-1}, n-1}.$$

Substituting these three facts into the displayed sum gives

$$\det A = A_{1,1} \sum_{(k_1, \dots, k_{n-1}) \in \operatorname{perm}(n-1)} \big(\operatorname{sign}(k_1, \dots, k_{n-1})\big) B_{k_1,1} \cdots B_{k_{n-1}, n-1} = A_{1,1} \det B,$$

where the last equality is 9.46 applied to B. This is the desired conclusion.

Problem 9C.3. Suppose $T \in \mathcal{L}(V)$ is nilpotent. Prove that $\det(I + T) = 1$.

Solution. Let $n = \dim V$. Because T is nilpotent, the equivalence of (a) and (c) in 8.18 shows that there is a basis $e_1, \dots, e_n$ of V with respect to which the matrix of T is upper triangular with all entries on and below the diagonal equal to 0; write $N = \mathcal{M}(T, (e_1, \dots, e_n))$ for this matrix, so

$$N_{j,k} = 0 \quad \text{whenever } j \ge k.$$

(Note that 8.18 is valid over $\mathbf{R}$ as well as over $\mathbf{C}$, so no assumption $\mathbf{F} = \mathbf{C}$ is needed here. Indeed, its proof of the implication from (a) to (c) goes through the minimal polynomial of T, which for a nilpotent operator is z^m for some positive integer m; since z^m is already a product of factors of the form $z - \lambda$ with $\lambda = 0 \in \mathbf{F}$, the criterion 5.44 applies over either field and yields an upper-triangular matrix, whose diagonal entries are then all 0 by 5.41.)
With respect to the same basis, the matrix of $I + T$ is $\mathcal{M}(I) + \mathcal{M}(T) = I + N$, where I now denotes the n-by-n identity matrix. Since N has zeros on and below the diagonal, the matrix $I + N$ is upper triangular and each of its diagonal entries equals 1. Therefore 9.48 gives

$$\det(I + N) = 1 \cdots 1 = 1.$$

Finally, 9.53 says that the determinant of an operator equals the determinant of its matrix with respect to any basis, so

$$\det(I + T) = \det \mathcal{M}\big(I + T, (e_1, \dots, e_n)\big) = \det(I + N) = 1,$$

as desired.
Here is a second proof in the case $\mathbf{F} = \mathbf{C}$, which explains the result conceptually. If λ is an eigenvalue of $I + T$ with eigenvector v, then $Tv = (\lambda - 1)v$, so $\lambda - 1$ is an eigenvalue of T; by 8.17(a) the only

eigenvalue of a nilpotent operator is 0, so $\lambda = 1$. Thus 1 is the only eigenvalue of $I + T$, and its multiplicity must be n because the multiplicities of the eigenvalues of an operator on a complex vector space sum to $\dim V$ (by 8.25). By 9.55 the determinant of $I + T$ is the product of its eigenvalues counted with multiplicity, which is $1^n = 1$.

Problem 9C.4. Suppose V is an inner product space and $S \in \mathcal{L}(V)$. Prove that S is unitary if and only if $|\det S| = \|S\| = 1$.

Solution. Here $\|S\|$ denotes the norm of the linear map S, defined in 7.86 by

$$\|S\| = \max\{\|Sv\| : v \in V \text{ and } \|v\| \le 1\}.$$

Recall the standing assumption of Chapter 9 that V is a finite-dimensional nonzero vector space; the hypothesis here adds an inner product on V. The assumption $V \neq \{0\}$ is used below (and is needed: on the zero vector space every operator has norm 0).
Let $n = \dim V$ and let $s_1, \dots, s_n$ be the list of singular values of S. By the definition 7.65 these are listed in decreasing order, so s_1 is the largest singular value, and there are exactly n of them because they are the nonnegative square roots of the eigenvalues of the positive operator S^*S, each repeated according to the dimension of the corresponding eigenspace, and those dimensions sum to $\dim V$ by the spectral theorem. Two facts from the text drive the whole proof:

$$\|S\| = s_1 \qquad \text{and} \qquad |\det S| = s_1 \cdots s_n.$$

The first is the content of 7.85 together with the definition 7.86 (and it holds trivially in the remaining case $S = 0$, when both sides are 0); the second is 9.60.
First suppose S is unitary. By the definition 7.51, a unitary operator is an invertible isometry, so S is an isometry; by the definition 7.44 of isometry this means $\|Sv\| = \|v\|$ for every $v \in V$. Hence $\|Sv\| \le 1$ for every v with $\|v\| \le 1$, and taking any v with $\|v\| = 1$ (such v exists because $V \neq \{0\}$) gives $\|Sv\| = 1$. Thus the maximum defining $\|S\|$ equals 1. Also $|\det S| = 1$ by 9.58. So $|\det S| = \|S\| = 1$.
Now suppose conversely that $|\det S| = \|S\| = 1$. From $\|S\| = 1$ and the identity $\|S\| = s_1$ we get $s_1 = 1$, and since s_1 is the largest of the singular values and singular values are nonnegative by definition,

$$0 \le s_k \le 1 \quad \text{for each } k \in \{1, \dots, n\}.$$

From $|\det S| = 1$ and the identity $|\det S| = s_1 \cdots s_n$ we get $s_1 \cdots s_n = 1$. If some s_k were strictly less than 1, then, because every factor lies in $[0, 1]$, the product $s_1 \cdots s_n$ would be at most $s_k < 1$, a contradiction. Hence

$$s_1 = \cdots = s_n = 1.$$

In particular all n singular values of S are positive, so the singular value decomposition 7.70 provides orthonormal lists $e_1, \dots, e_n$ in V and $f_1, \dots, f_n$ in V such that

$$Sv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n = \langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n$$

for every $v \in V$. Each of these orthonormal lists has length $n = \dim V$, so each is an orthonormal basis of V (an orthonormal list is linearly independent by 6.25, and a linearly independent list of length $\dim V$ is a basis by 2.38).
Now let $v \in V$. Because $f_1, \dots, f_n$ is orthonormal, 6.24 applied to the displayed sum gives

$$\|Sv\|^2 = \sum_{k=1}^{n} \big|\langle v, e_k\rangle\big|^2.$$

Because $e_1, \dots, e_n$ is an orthonormal basis of V, Parseval's identity (6.30(b)) gives

$$\sum_{k=1}^{n} \big|\langle v, e_k\rangle\big|^2 = \|v\|^2.$$

Combining the two displays yields $\|Sv\| = \|v\|$ for all $v \in V$. Thus S is an isometry. Every isometry is injective (as noted in the text just after 7.44: if $Sv = 0$ then $\|v\| = \|Sv\| = 0$, so $v = 0$), and an injective operator on a finite-dimensional vector space is invertible (3.65). Hence S is an invertible isometry, which is exactly a unitary operator by the definition 7.51.
This completes both directions.
One comment on why both hypotheses are needed; assume $\dim V \geq 2$ for these examples. The condition $\|S\| = 1$ alone does not force S to be unitary: the orthogonal projection of V onto a proper nonzero subspace has norm 1 but determinant 0. The condition $|\det S| = 1$ alone does not suffice either: the operator on $\mathbf{F}^2$ whose matrix with respect to the standard basis is $\begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$ has determinant 1 but norm 2. It is the two conditions together, which force every singular value to equal 1, that characterize unitarity.

Problem 9C.5. Suppose A is a block upper-triangular matrix

$$A = \begin{pmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each A_k along the diagonal is a square matrix. Prove that

$$\det A = (\det A_1) \cdots (\det A_m).$$

Solution. First we make the hypothesis precise. Say A_k is n_k-by-n_k and put $n = n_1 + \cdots + n_m$, so A is n-by-n. Partition $\{1, \ldots, n\}$ into the consecutive blocks

$$I_k = \{n_1 + \cdots + n_{k-1} + 1, \ldots, n_1 + \cdots + n_k\} \quad \text{for } k \in \{1, \ldots, m\},$$

so that I_k has n_k elements. The picture in the statement says exactly two things: for row and column indices j, ℓ lying in the same block I_k, the entries of A are the entries of A_k, that is,

$$A_{j,\ell} = (A_k)_{j-c_k,\, \ell-c_k} \quad \text{for } j, \ell \in I_k, \text{ where } c_k = n_1 + \cdots + n_{k-1},$$

and everything strictly below the diagonal blocks vanishes, that is,

$$A_{j,\ell} = 0 \quad \text{whenever } j \in I_a \text{ and } \ell \in I_b \text{ with } a > b.$$

The entries in positions with $a < b$ (the region marked $*$) are arbitrary.
The heart of the matter is the case of two diagonal blocks; the general case then follows by induction on m.
Step 1: the case $m = 2$.
Suppose

$$A = \begin{pmatrix} A_1 & C \\ 0 & A_2 \end{pmatrix},$$

where A_1 is p-by-p, A_2 is q-by-q, and C is p-by-q. Put $n = p + q$. In terms of entries, the hypothesis is

$$A_{j,k} = 0 \quad \text{whenever } j > p \text{ and } k \leq p,$$

while

$$A_{j,k} = (A_1)_{j,k} \text{ for } j, k \leq p, \qquad A_{j,k} = (A_2)_{j-p,\, k-p} \text{ for } j, k > p.$$

Use the formula 9.46:

$$\det A = \sum_{(j_1, \ldots, j_n) \in \operatorname{perm} n} \big(\operatorname{sign}(j_1, \ldots, j_n)\big) A_{j_1,1} \cdots A_{j_n,n}.$$

Fix $(j_1, \dots, j_n) \in \operatorname{perm} n$ and suppose its term is nonzero. Then $A_{j_k,k} \neq 0$ for every k, so by the vanishing condition above we must have $j_k \leq p$ for each $k \in \{1, \dots, p\}$. Thus $j_1, \dots, j_p$ are p distinct elements of $\{1, \dots, p\}$, which forces $(j_1, \dots, j_p) \in \operatorname{perm} p$; and then $j_{p+1}, \dots, j_n$ are exactly the remaining values $p+1, \dots, n$ in some order, so

$$(j_{p+1} - p, \dots, j_n - p) \in \operatorname{perm} q.$$

Conversely, every pair $\sigma = (\sigma_1, \dots, \sigma_p) \in \operatorname{perm} p$ and $\tau = (\tau_1, \dots, \tau_q) \in \operatorname{perm} q$ determines exactly one such permutation of $(1, \dots, n)$, namely

$$(\sigma_1, \dots, \sigma_p,\ \tau_1 + p, \dots, \tau_q + p).$$

So the permutations that can contribute are in bijection with $\operatorname{perm} p \times \operatorname{perm} q$.
Next compare signs. Recall 9.32: the sign is $(-1)^N$, where N counts the pairs (k, ℓ) with $k < \ell$ such that k appears after ℓ in the list. Split the pairs of the list $(\sigma_1, \dots, \sigma_p, \tau_1 + p, \dots, \tau_q + p)$ into three kinds.

- Both members in $\{1, \dots, p\}$. These occupy the first p slots, and the count of such inverted pairs is exactly the count for σ.
- Both members in $\{p+1, \dots, n\}$. These occupy the last q slots, and since subtracting p from each entry is an order-preserving relabeling, the count of such inverted pairs is exactly the count for τ.
- One member $k \leq p$ and the other $\ell > p$. Every value at most p occupies one of the first p slots and every value greater than p occupies one of the last q slots, so k always appears before ℓ. No such pair is inverted.

Hence $N = N_\sigma + N_\tau$, and therefore

$$\operatorname{sign}(\sigma_1, \dots, \sigma_p, \tau_1 + p, \dots, \tau_q + p) = (\operatorname{sign} \sigma)(\operatorname{sign} \tau).$$

Finally compare the products. For $k \leq p$ we have $A_{\sigma_k,k} = (A_1)_{\sigma_k,k}$, and for $k = p + r$ with $1 \leq r \leq q$ we have $A_{\tau_r+p,\, p+r} = (A_2)_{\tau_r,r}$. Hence

$$A_{j_1,1} \cdots A_{j_n,n} = \Big(\prod_{k=1}^{p} (A_1)_{\sigma_k,k}\Big)\Big(\prod_{r=1}^{q} (A_2)_{\tau_r,r}\Big).$$

Putting the three observations together and factoring the double sum,

$$\begin{aligned}
\det A &= \sum_{\sigma \in \operatorname{perm} p} \sum_{\tau \in \operatorname{perm} q} (\operatorname{sign} \sigma)(\operatorname{sign} \tau)\Big(\prod_{k=1}^{p} (A_1)_{\sigma_k,k}\Big)\Big(\prod_{r=1}^{q} (A_2)_{\tau_r,r}\Big) \\
&= \Big(\sum_{\sigma \in \operatorname{perm} p} (\operatorname{sign} \sigma) \prod_{k=1}^{p} (A_1)_{\sigma_k,k}\Big)\Big(\sum_{\tau \in \operatorname{perm} q} (\operatorname{sign} \tau) \prod_{r=1}^{q} (A_2)_{\tau_r,r}\Big) \\
&= (\det A_1)(\det A_2),
\end{aligned}$$

the last line by 9.46 applied to A_1 and to A_2. Note that the entries of the block C never appear: they sit in positions (j, k) with $j \leq p < k$, and no contributing permutation ever selects such a position, since it sends $\{1, \dots, p\}$ onto itself and hence $\{p+1, \dots, n\}$ onto itself.
Step 2: induction on m.
The case $m = 1$ is trivial, and $m = 2$ is Step 1. Suppose $m > 2$ and the result holds for block upper-triangular matrices with fewer than m diagonal blocks. Write A_1 as a p-by-p block and let A' be the square matrix obtained from A by deleting the first p rows and the first p columns. Then

$$A = \begin{pmatrix} A_1 & C \\ 0 & A' \end{pmatrix}$$

for some matrix C: indeed, taking $p = n_1$, so that $I_1 = \{1, \dots, p\}$, the entries of A lying below the block A_1 are precisely the entries $A_{j,\ell}$ with $j \in I_a$ for some $a > 1$ and $\ell \in I_1$, and all of these are 0 by the vanishing condition above. Moreover A' is itself block upper triangular with diagonal blocks $A_2, \dots, A_m$: its index set is $\{1, \dots, n-p\}$ partitioned into the consecutive blocks $I_2 - p, \dots, I_m - p$,

and both conditions in the description of A transfer to A' verbatim because subtracting p from every index is a translation that preserves the block structure and the order of the blocks. By Step 1 and then the induction hypothesis,

$$\det A = (\det A_1)(\det A') = (\det A_1)(\det A_2)\cdots(\det A_m),$$

which completes the induction and the proof.
Observe that taking each A_k to be 1-by-1 recovers 9.48, the formula for the determinant of an upper-triangular matrix; and taking $m = 2$ with A_1 the 1-by-1 block $(A_{1,1})$ recovers Exercise 9C.2.

Problem 9C.6. Suppose $A = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ is an n-by-n matrix, with v_k denoting the kth column of A. Show that if $(m_1, \ldots, m_n) \in \operatorname{perm} n$, then

$$\det \begin{pmatrix} v_{m_1} & \cdots & v_{m_n} \end{pmatrix} = \big(\operatorname{sign}(m_1, \ldots, m_n)\big) \det A.$$

Solution. Define $\alpha\colon (\mathbf{F}^n)^n \to \mathbf{F}$ by

$$\alpha(u_1, \ldots, u_n) = \det \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix}$$

for $u_1, \ldots, u_n \in \mathbf{F}^n$, where as usual $\begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix}$ denotes the n-by-n matrix whose kth column is u_k. By 9.45, α is an alternating n-linear form on $\mathbf{F}^n$.
Now apply 9.35, which states that if α is an alternating m-linear form on a vector space and $(m_1, \ldots, m_m)$ is a permutation of $(1, \ldots, m)$, then

$$\alpha(u_{m_1}, \ldots, u_{m_m}) = \big(\operatorname{sign}(m_1, \ldots, m_m)\big)\, \alpha(u_1, \ldots, u_m)$$

for every list $u_1, \ldots, u_m$ of vectors. Taking $m = n$ and $u_k = v_k$ gives

$$\det \begin{pmatrix} v_{m_1} & \cdots & v_{m_n} \end{pmatrix} = \alpha(v_{m_1}, \ldots, v_{m_n}) = \big(\operatorname{sign}(m_1, \ldots, m_n)\big)\, \alpha(v_1, \ldots, v_n) = \big(\operatorname{sign}(m_1, \ldots, m_n)\big) \det A,$$

as desired.
For completeness, here is why 9.35 holds in this setting, spelled out. The permutation $(m_1, \ldots, m_n)$ can be transformed into $(1, \ldots, n)$ by a finite sequence of swaps of two entries. Each swap changes the value of α by a factor of -1, since α is alternating (9.30, whose proof uses only that α vanishes when two of its arguments are equal, together with multilinearity), and each swap also changes the sign of the permutation by a factor of -1 (9.34). Hence the quantity

$$\big(\operatorname{sign}(m_1, \ldots, m_n)\big)\, \alpha(v_{m_1}, \ldots, v_{m_n})$$

is unchanged by each swap. After the full sequence of swaps we reach the identity permutation, whose sign is 1 (9.33), and the quantity above becomes $\alpha(v_1, \ldots, v_n)$. Therefore

$$\big(\operatorname{sign}(m_1, \ldots, m_n)\big)\, \alpha(v_{m_1}, \ldots, v_{m_n}) = \alpha(v_1, \ldots, v_n).$$

Multiplying both sides by $\operatorname{sign}(m_1, \ldots, m_n)$ and using that this sign is ± 1, hence squares to 1, gives the stated identity.

Problem 9C.7. Suppose $T \in \mathcal{L}(V)$ is invertible. Let p denote the characteristic polynomial of T and let q denote the characteristic polynomial of T^{-1}. Prove that

$$q(z) = \frac{1}{p(0)} z^{\dim V} p\left(\frac{1}{z}\right)$$

for all nonzero $z \in \mathbf{F}$.

Solution. Let $n = \dim V$. By the definition 9.63 of the characteristic polynomial,

$$p(z) = \det(zI - T) \qquad \text{and} \qquad q(z) = \det(zI - T^{-1}) \quad \text{for } z \in \mathbf{F}.$$

First we evaluate $p(0)$ and check that it is nonzero, so that the right side of the asserted formula makes sense. Using the third bullet point of 9.42 with $\lambda = -1$,

$$p(0) = \det(0 \cdot I - T) = \det(-T) = (-1)^n \det T.$$

Because T is invertible, $\det T \neq 0$ by 9.50; hence $p(0) \neq 0$. Note also that $(-1)^n$ is its own reciprocal, so

$$\frac{1}{p(0)} = \frac{1}{(-1)^n \det T} = \frac{(-1)^n}{\det T}.$$

Now fix a nonzero $z \in \mathbf{F}$. The key algebraic identity is

$$zI - T^{-1} = -T^{-1}(I - zT),$$

which one checks by expanding the right side: $-T^{-1}(I - zT) = -T^{-1} + zT^{-1}T = -T^{-1} + zI$.
Apply the multiplicativity of the determinant, 9.49(a), to this factorization:

$$q(z) = \det(zI - T^{-1}) = \big(\det(-T^{-1})\big)\big(\det(I - zT)\big).$$

We compute the two factors separately.
For the first factor, the third bullet point of 9.42 (with $\lambda = -1$ applied to the operator T^{-1}) together with 9.50 gives

$$\det(-T^{-1}) = (-1)^n \det(T^{-1}) = \frac{(-1)^n}{\det T}.$$

For the second factor, use $z \neq 0$ to write

$$I - zT = z\left(\frac{1}{z}I - T\right),$$

and then the third bullet point of 9.42 again, this time with the scalar z:

$$\det(I - zT) = z^n \det\left(\frac{1}{z}I - T\right) = z^n\, p\left(\frac{1}{z}\right),$$

the last equality by the definition of p (note $1/z \in \mathbf{F}$ since z is a nonzero element of the field $\mathbf{F}$).
Combining,

$$q(z) = \frac{(-1)^n}{\det T} \cdot z^n\, p\left(\frac{1}{z}\right) = \frac{1}{p(0)} z^n p\left(\frac{1}{z}\right),$$

where the last equality uses the computation of $1/p(0)$ above. Since $n = \dim V$, this is exactly the asserted formula, valid for every nonzero $z \in \mathbf{F}$.
A remark on what the formula says. Both p and q are monic of degree n (as noted after 9.63). Writing $p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ with $a_0 = p(0) \neq 0$, the identity just proved gives

$$q(z) = \frac{1}{a_0}\left(1 + a_{n-1}z + \cdots + a_1 z^{n-1} + a_0 z^n\right) = z^n + \frac{a_1}{a_0}z^{n-1} + \cdots + \frac{a_{n-1}}{a_0}z + \frac{1}{a_0}.$$

In other words, the coefficient list of the characteristic polynomial of T^{-1} is the reversed coefficient list of that of T, rescaled to be monic. Consistently with this, λ is a zero of q if and only if $\lambda \neq 0$ and $1/\lambda$ is a zero of p, which recovers the familiar fact that the eigenvalues of T^{-1} are precisely the reciprocals of the eigenvalues of T.

Problem 9C.8. Suppose $T \in \mathcal{L}(V)$ is an operator with no eigenvalues (which implies that $\mathbf{F} = \mathbf{R}$). Prove that $\det T > 0$.

Solution. Let $n = \dim V$; recall that in this chapter V is a nonzero finite-dimensional vector space, so $n \geq 1$. Because T has no eigenvalues and every operator on a finite-dimensional complex vector space has an eigenvalue (5.19), we have $\mathbf{F} = \mathbf{R}$, as the exercise notes.
Let p be the characteristic polynomial of T, so

$$p(x) = \det(xI - T)$$

for all $x \in \mathbf{R}$ (9.63). Fix a basis of V and let $A = \mathcal{M}(T)$ with respect to that basis, so that $p(x) = \det(xI - A)$ by 9.53. In the formula 9.46 for $\det(xI - A)$, the identity permutation contributes $(x - A_{1,1}) \cdots (x - A_{n,n})$, and every other permutation contributes a product containing at most $n-2$ diagonal factors $x - A_{k,k}$; hence p is a monic polynomial of degree n with real coefficients (this is the observation made just after 9.63 and used in the proof of 9.65).
By 9.51, a number $\lambda \in \mathbf{R}$ is an eigenvalue of T if and only if $p(\lambda) = \det(\lambda I - T) = 0$. Since T has no eigenvalues, p has no zeros in $\mathbf{R}$.
Because $n \geq 1$, the polynomial p is nonconstant, so the factorization of a polynomial over $\mathbf{R}$ (4.16) gives

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M),$$

where $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbf{R}$ and $b_k^2 < 4c_k$ for each k. Each λ_j appearing above would be a real zero of p; since p has no real zeros, we must have $m = 0$. Thus

$$p(x) = c(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M).$$

Comparing leading coefficients and degrees with the fact that p is monic of degree n, we get $c = 1$ and $n = 2M$. In particular n is even (and $M \geq 1$, because $n \geq 1$).
Now evaluate at 0. For each k we have $c_k > b_k^2/4 \geq 0$, so $c_k > 0$. Hence

$$p(0) = c_1 \cdots c_M > 0.$$

On the other hand, $p(0) = \det(0 \cdot I - T) = \det(-T) = (-1)^n \det T$ by the third bullet point of 9.42, and $(-1)^n = 1$ because n is even. Therefore

$$\det T = p(0) > 0,$$

as desired.

Problem 9C.9. Suppose that V is a real vector space of even dimension, $T \in \mathcal{L}(V)$, and $\det T < 0$. Prove that T has at least two distinct eigenvalues.

Solution. Let $n = \dim V$, which is even, and let $p(x) = \det(xI - T)$ be the characteristic polynomial of T (9.63). As recalled in the solution to Exercise 8, p is a monic polynomial of degree n with real coefficients, and by 9.51 the zeros of p in $\mathbf{R}$ are exactly the eigenvalues of T.
First note that

$$p(0) = \det(-T) = (-1)^n \det T = \det T < 0,$$

using the third bullet point of 9.42 and the evenness of n. Also, since $\det T \neq 0$, the operator T is invertible (9.50), so 0 is not an eigenvalue of T.
Suppose, for the sake of contradiction, that T has at most one distinct eigenvalue. Since $n \geq 1$, the polynomial p is nonconstant, so by the factorization of a polynomial over $\mathbf{R}$ (4.16) we may write

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M),$$

with all coefficients real and $b_k^2 < 4c_k$ for each k; comparing leading coefficients shows $c = 1$, and comparing degrees shows $m + 2M = n$. Each quadratic factor $x^2 + b_k x + c_k$ has no real zeros, because

completing the square gives $x^2 + b_k x + c_k = (x + b_k/2)^2 + (c_k - b_k^2/4)$ with $c_k - b_k^2/4 > 0$. Hence the real zeros of p are precisely $\lambda_1, \ldots, \lambda_m$, which by 9.51 are precisely the eigenvalues of T. By our contradiction hypothesis there is at most one distinct such number; call it λ when $m > 0$ (so every λ_j equals λ). Thus

$$p(x) = (x - \lambda)^m q(x), \qquad q(x) = (x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M),$$

where the first factor is interpreted as 1 when $m = 0$.
For each k we have $c_k > b_k^2/4 \geq 0$, so $c_k > 0$, and therefore

$$q(0) = c_1 \cdots c_M > 0.$$

Moreover $m = n - 2M$ is even, because n is even.
If $m = 0$, then $p(0) = q(0) > 0$. If $m > 0$, then λ is an eigenvalue of T, so $\lambda \neq 0$ (as T is invertible), and

$$p(0) = (-\lambda)^m q(0) = \lambda^m q(0) > 0,$$

since m is even and $\lambda \neq 0$. In both cases $p(0) > 0$, contradicting $p(0) = \det T < 0$.
This contradiction shows that T has at least two distinct eigenvalues.

Problem 9C.10. Suppose V is a real vector space of odd dimension and $T \in \mathcal{L}(V)$. Without using the minimal polynomial, prove that T has an eigenvalue.
[This result was previously proved without using determinants or the characteristic polynomial; see 5.34.]

Solution. Let $n = \dim V$, an odd positive integer, and let

$$p(x) = \det(xI - T)$$

be the characteristic polynomial of T (9.63). Fix a basis of V and let $A = \mathcal{M}(T)$ with respect to that basis; then $p(x) = \det(xI - A)$ by 9.53. Applying the formula 9.46 to the matrix $xI - A$, the identity permutation contributes the term $(x - A_{1,1}) \cdots (x - A_{n,n})$, while every other permutation $(j_1, \ldots, j_n)$ satisfies $j_k \neq k$ for at least two indices k and hence contributes a product with at most $n - 2$ factors of the form $x - A_{k,k}$, so it contributes a polynomial of degree at most $n - 2$. Consequently p is a monic polynomial of degree n with real coefficients. (No use of the minimal polynomial has been made, nor will any be made below.)
Since $n \geq 1$, the polynomial p is nonconstant. By the factorization of a polynomial over $\mathbf{R}$ (4.16) there exist real numbers $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M$ with $b_k^2 < 4c_k$ for each k such that

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M)$$

for all $x \in \mathbf{R}$. Comparing degrees of the two sides gives

$$n = m + 2M.$$

Because n is odd, m cannot be 0; indeed m is odd, so in particular $m \geq 1$.
Thus there is a real number λ_1 with

$$\det(\lambda_1 I - T) = p(\lambda_1) = 0.$$

By 9.51, this says precisely that λ_1 is an eigenvalue of T. Hence T has an eigenvalue.

Problem 9C.11. Prove or give a counterexample: If $\mathbf{F} = \mathbf{R}$, $T \in \mathcal{L}(V)$, and $\det T > 0$, then T has a square root.
[If $\mathbf{F} = \mathbf{C}$, $T \in \mathcal{L}(V)$, and $\det T \neq 0$, then T has a square root (see 8.41).]

Solution. The statement is false. Here is a counterexample.
Take $\mathbf{F} = \mathbf{R}$, $V = \mathbf{R}^2$, and define $T \in \mathcal{L}(\mathbf{R}^2)$ by

$$T(x, y) = (-x, -2y).$$

With respect to the standard basis e_1, e_2 of $\mathbf{R}^2$, the matrix of T is the diagonal matrix with $-1, -2$ on the diagonal. Hence 9.53 and 9.48 give

$$\det T = (-1)(-2) = 2 > 0.$$

We claim that T has no square root, meaning that there is no $S \in \mathcal{L}(\mathbf{R}^2)$ with $S^2 = T$.
Suppose such an S exists. First observe that S commutes with T, since

$$ST = SS^2 = S^2S = TS.$$

Next, identify the eigenspace of T corresponding to the eigenvalue -1. For $(x, y) \in \mathbf{R}^2$ we have

$$(T + I)(x, y) = (0, -y),$$

so $(T + I)(x, y) = 0$ if and only if $y = 0$. Thus

$$E(-1, T) = \operatorname{null}(T + I) = \operatorname{span}(e_1).$$

Because S commutes with T, the subspace $E(-1, T)$ is invariant under S: if $Tv = -v$, then

$$T(Sv) = S(Tv) = S(-v) = -(Sv),$$

so $Sv \in E(-1, T)$. Since $E(-1, T) = \operatorname{span}(e_1)$, there exists $a \in \mathbf{R}$ with

$$Se_1 = ae_1.$$

Applying S again gives

$$-e_1 = Te_1 = S^2e_1 = a^2e_1,$$

and hence $a^2 = -1$. No real number satisfies this equation, a contradiction.
Therefore no $S \in \mathcal{L}(\mathbf{R}^2)$ satisfies $S^2 = T$, even though $\det T = 2 > 0$. This disproves the statement.
Remark: the obstruction here is that a real operator can have negative eigenvalues while its determinant remains positive; the analogous complex statement holds because every nonzero complex number has a square root, which is what makes 8.41 work.

Problem 9C.12. Suppose $S, T \in \mathcal{L}(V)$ and S is invertible. Define $p\colon \mathbf{F} \to \mathbf{F}$ by

$$p(z) = \det(zS - T).$$

Prove that p is a polynomial of degree $\dim V$ and that the coefficient of $z^{\dim V}$ in this polynomial is $\det S$.

Solution. Let $n = \dim V$.
Because S is invertible, for every $z \in \mathbf{F}$ we can factor

$$zS - T = S\left(zI - S^{-1}T\right),$$

as is verified by expanding the right side: $S(zI - S^{-1}T) = zS - SS^{-1}T = zS - T$.
Since the determinant is multiplicative (9.49(a)), applying det to the displayed factorization gives

$$p(z) = \det(zS - T) = (\det S)\det\left(zI - S^{-1}T\right)$$

for all $z \in \mathbf{F}$.

Let q denote the characteristic polynomial of the operator $S^{-1}T \in \mathcal{L}(V)$, so that $q(z) = \det(zI - S^{-1}T)$ by 9.63. We claim q is monic of degree n. To see this, fix a basis of V and let $A = \mathcal{M}(S^{-1}T)$ with respect to that basis; then $q(z) = \det(zI - A)$ by 9.53. Apply the formula 9.46 to the n-by-n matrix $zI - A$, whose entry in row j, column k equals $z - A_{k,k}$ if $j = k$ and equals $-A_{j,k}$ if $j \neq k$. The identity permutation $(1, \dots, n)$ contributes

$$(z - A_{1,1}) \cdots (z - A_{n,n}),$$

a monic polynomial in z of degree n. Any other permutation $(j_1, \dots, j_n) \in \operatorname{perm} n$ satisfies $j_k \neq k$ for at least two values of k, so its contribution is a product containing at most $n - 2$ factors of the form $z - A_{k,k}$ and otherwise only constants; such a contribution is a polynomial of degree at most $n - 2$. Summing, q is a polynomial of degree n whose coefficient of z^n equals 1. (This is exactly the observation recorded just after 9.63.)
Therefore

$$p(z) = (\det S)\, q(z)$$

for all $z \in \mathbf{F}$, and p is the polynomial $(\det S)q$. Because S is invertible, $\det S \neq 0$ by 9.50. Multiplying the monic degree-n polynomial q by the nonzero scalar $\det S$ produces a polynomial of degree exactly $n = \dim V$ whose coefficient of z^n equals $\det S \cdot 1 = \det S$.
Thus p is a polynomial of degree $\dim V$ and the coefficient of $z^{\dim V}$ in p is $\det S$, as desired.

Problem 9C.13. Suppose $\mathbf{F} = \mathbf{C}$, $T \in \mathcal{L}(V)$, and $n = \dim V > 2$. Let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of T, with each eigenvalue included as many times as its multiplicity.
(a) Find a formula for the coefficient of z^{n-2} in the characteristic polynomial of T in terms of $\lambda_1, \dots, \lambda_n$.
(b) Find a formula for the coefficient of z in the characteristic polynomial of T in terms of $\lambda_1, \dots, \lambda_n$.

Solution. Let q denote the characteristic polynomial of T, so $q(z) = \det(zI - T)$ (9.63). Let $\mu_1, \dots, \mu_m$ be the distinct eigenvalues of T with multiplicities $d_1, \dots, d_m$. By 9.62,

$$q(z) = (z - \mu_1)^{d_1} \cdots (z - \mu_m)^{d_m}.$$

Since $\mathbf{F} = \mathbf{C}$, the sum of the multiplicities equals $\dim V = n$ (8.25), and the list $\lambda_1, \dots, \lambda_n$ is by definition the list in which each μ_j is repeated d_j times. Hence the displayed product is exactly $\prod_{k=1}^{n}(z - \lambda_k)$, so

$$q(z) = (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n).$$

Now expand this product. Multiplying out means choosing, from each factor $z - \lambda_k$, either the term z or the term $-\lambda_k$, and summing over all such choices. Encoding a choice by the set $A \subseteq \{1, \dots, n\}$ of indices from which $-\lambda_k$ is chosen, we obtain

$$q(z) = \sum_{A \subseteq \{1,\dots,n\}} \left(\prod_{k \in A} (-\lambda_k) \right) z^{n - |A|} = \sum_{r=0}^{n} (-1)^r \left(\sum_{\substack{A \subseteq \{1,\dots,n\} \\ |A| = r}} \prod_{k \in A} \lambda_k \right) z^{n-r}.$$

Thus for each $r \in \{0, 1, \dots, n\}$, the coefficient of z^{n-r} in q is $(-1)^r e_r$, where

$$e_r = \sum_{1 \leq k_1 < k_2 < \cdots < k_r \leq n} \lambda_{k_1} \lambda_{k_2} \cdots \lambda_{k_r}$$

is the r-th elementary symmetric function of $\lambda_1, \dots, \lambda_n$ (with $e_0 = 1$).
(a) Take $r = 2$, which is legitimate since $n > 2$ guarantees $0 \leq n - 2 < n$. The coefficient of z^{n-2} in the characteristic polynomial of T is

$$(-1)^2 e_2 = \sum_{1 \leq j < k \leq n} \lambda_j \lambda_k.$$

(b) The coefficient of $z = z^1$ corresponds to $n - r = 1$, that is, $r = n - 1$. A subset A of $\{1, \dots, n\}$ with $|A| = n - 1$ is exactly the complement of a single index k, so

$$e_{n-1} = \sum_{k=1}^{n} \prod_{\substack{j=1 \\ j \neq k}}^{n} \lambda_j.$$

Hence the coefficient of z in the characteristic polynomial of T is

$$(-1)^{n-1} e_{n-1} = (-1)^{n-1} \sum_{k=1}^{n} \lambda_1 \cdots \lambda_{k-1} \lambda_{k+1} \cdots \lambda_n.$$

Two consistency checks. Taking $r = 1$ in the general formula gives $-(\lambda_1 + \cdots + \lambda_n)$ as the coefficient of z^{n-1}, and taking $r = n$ gives $(-1)^n \lambda_1 \cdots \lambda_n$ as the constant term; these agree with 9.65 together with 8.52 and 9.55, which say that the trace of T is the sum of the eigenvalues and the determinant of T is the product of the eigenvalues (each with multiplicity). Also, when $n = 3$ the answers in (a) and (b) must coincide, since then $z^{n-2} = z$: indeed (a) gives $\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3$ and (b) gives $(-1)^2(\lambda_2\lambda_3 + \lambda_1\lambda_3 + \lambda_1\lambda_2)$, the same number.

Problem 9C.14. Suppose V is an inner product space and T is a positive operator on V. Prove that

$$\det \sqrt{T} = \sqrt{\det T}.$$

Solution. Because T is a positive operator, it has a unique positive square root (7.39), denoted $\sqrt{T}$ (7.40). Thus $\sqrt{T}$ is a positive operator on V with

$$\sqrt{T}\,\sqrt{T} = T.$$

Two applications of 9.59 (every positive operator has nonnegative determinant), one to T and one to $\sqrt{T}$, show that both $\det T$ and $\det \sqrt{T}$ are nonnegative real numbers. In particular $\sqrt{\det T}$ makes sense as the unique nonnegative real square root of the nonnegative real number $\det T$.
Because the determinant is multiplicative (9.49(a)),

$$\det T = \det\Big(\sqrt{T}\,\sqrt{T}\Big) = \Big(\det \sqrt{T}\Big)\Big(\det \sqrt{T}\Big) = \Big(\det \sqrt{T}\Big)^2.$$

So $\det \sqrt{T}$ is a nonnegative real number whose square equals $\det T$. Since a nonnegative real number has exactly one nonnegative real square root, we conclude that

$$\det \sqrt{T} = \sqrt{\det T},$$

as desired.
Note that the nonnegativity of $\det \sqrt{T}$ is essential: from $(\det \sqrt{T})^2 = \det T$ alone one could only conclude $\det \sqrt{T} = \pm\sqrt{\det T}$, and 9.59 is exactly what rules out the minus sign.

Problem 9C.15. Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Use the polar decomposition to give a proof that

$$|\det T| = \sqrt{\det(T^*T)}$$

that is different from the proof given earlier (see 9.60).

Solution. First note that T^*T is a positive operator: it is self-adjoint because $(T^*T)^* = T^*(T^*)^* = T^*T$, and for every $v \in V$ we have

$$\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2 \geq 0.$$

Hence T^*T has a unique positive square root $R = \sqrt{T^*T}$ (by 7.39 and 7.40), and $R^2 = T^*T$.

By the polar decomposition (7.93) there exists a unitary operator $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T} = SR.$$

Because the determinant is multiplicative (9.49(a)),

$$\det T = (\det S)(\det R).$$

Taking absolute values and using $|\det S| = 1$ (9.58, every unitary operator has determinant of absolute value 1), we get

$$|\det T| = |\det S|\,|\det R| = |\det R|.$$

Since R is a positive operator, $\det R \geq 0$ (9.59), so $|\det R| = \det R$ and therefore

$$|\det T| = \det R.$$

Finally, applying 9.49(a) again to $R^2 = T^*T$ gives

$$(\det R)^2 = \det(R^2) = \det(T^*T).$$

Because $\det R \geq 0$, taking nonnegative square roots of both sides of the equation above yields $\det R = \sqrt{\det(T^*T)}$ (the number $\det(T^*T)$ is nonnegative, again by 9.59 applied to the positive operator T^*T). Combining this with $|\det T| = \det R$ gives

$$|\det T| = \sqrt{\det(T^*T)},$$

as desired.
This proof differs from the proof of 9.60, which multiplied $\det T$ by its complex conjugate and applied $\det(T^*) = \overline{\det T}$ (9.56(c)) directly to T. The argument above instead factors T into a unitary operator times a positive operator and then uses only the multiplicativity of the determinant together with two structural facts: unitary operators have determinant of absolute value 1 (9.58), and positive operators have nonnegative determinant (9.59).

Problem 9C.16. Suppose $T \in \mathcal{L}(V)$. Define $g\colon \mathbf{F} \to \mathbf{F}$ by $g(x) = \det(I + xT)$. Show that $g'(0) = \operatorname{tr} T$.
[Look for a clean solution to this exercise, without using the explicit but complicated formula for the determinant of a matrix.]

Solution. Let $n = \dim V$, fix a basis $u_1, \ldots, u_n$ of V, and let $A = \mathcal{M}(T, (u_1, \ldots, u_n))$. Then $\mathcal{M}(I + xT, (u_1, \ldots, u_n)) = I + xA$ for every $x \in \mathbf{F}$, so 9.53 gives

$$g(x) = \det(I + xT) = \det(I + xA).$$

Also $\operatorname{tr} T = A_{1,1} + \cdots + A_{n,n}$, by the definition of the trace of an operator.
Let $e_1, \ldots, e_n$ be the standard basis of $\mathbf{F}^n$ and let $A_{\cdot,k} \in \mathbf{F}^n$ denote the k^{th} column of A. Let

$$D(v_1, \ldots, v_n) = \det \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$$

for $v_1, \ldots, v_n \in \mathbf{F}^n$. By 9.45, D is an alternating n-linear form on $\mathbf{F}^n$. The k^{th} column of $I + xA$ equals $e_k + xA_{\cdot,k}$, and thus

$$g(x) = D(e_1 + xA_{\cdot,1},\ \ldots,\ e_n + xA_{\cdot,n}).$$

Now expand using linearity in each of the n slots, one slot at a time. Each slot contributes either e_k or $xA_{\cdot,k}$, so the expansion is indexed by the subsets of $\{1, \ldots, n\}$: for $S \subseteq \{1, \ldots, n\}$ put $u_k^S = A_{\cdot,k}$ if $k \in S$ and $u_k^S = e_k$ if $k \notin S$. Pulling the scalars x out of the slots indexed by S gives

$$g(x) = \sum_{S \subseteq \{1,\ldots,n\}} x^{|S|}\, D(u_1^S, \ldots, u_n^S) = \sum_{m=0}^{n} c_m x^m, \qquad \text{where } c_m = \sum_{|S|=m} D(u_1^S, \ldots, u_n^S).$$

In particular g is a polynomial function of x of degree at most n, so $g'(0) = c_1$ (for $\mathbf{F} = \mathbf{C}$ this is the complex derivative, and in either case the derivative at 0 of a polynomial function is its coefficient of x).
It remains to compute c_1. The only subsets with $|S| = 1$ are the singletons $\{k\}$, so

$$c_1 = \sum_{k=1}^{n} D(e_1, \ldots, e_{k-1}, A_{\cdot,k}, e_{k+1}, \ldots, e_n).$$

Fix k and write $A_{\cdot,k} = \sum_{j=1}^{n} A_{j,k} e_j$. Using linearity of D in the k^{th} slot,

$$D(e_1, \ldots, e_{k-1}, A_{\cdot,k}, e_{k+1}, \ldots, e_n) = \sum_{j=1}^{n} A_{j,k}\, D(e_1, \ldots, e_{k-1}, e_j, e_{k+1}, \ldots, e_n).$$

If $j \neq k$, then the list $e_1, \ldots, e_{k-1}, e_j, e_{k+1}, \ldots, e_n$ has the vector e_j in two different slots, so the corresponding term is 0 because D is alternating. Hence only the term $j = k$ survives, and it equals

$$A_{k,k}\, D(e_1, \ldots, e_n) = A_{k,k} \det I = A_{k,k}.$$

Therefore

$$g'(0) = c_1 = \sum_{k=1}^{n} A_{k,k} = \operatorname{tr} A = \operatorname{tr} T,$$

as desired.
Incidentally, the computation above shows more: writing out the coefficients c_m gives $g(x) = 1 + (\operatorname{tr} T)x + c_2 x^2 + \cdots + c_n x^n$ with $c_n = \det T$ (the subset $S = \{1, \ldots, n\}$ contributes $x^n D(A_{\cdot,1}, \ldots, A_{\cdot,n}) = x^n \det A$).

Problem 9C.17. Suppose a, b, c are positive numbers. Find the volume of the ellipsoid

$$\left\{(x, y, z) \in \mathbf{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} < 1\right\}$$

by finding a set $\Omega \subseteq \mathbf{R}^3$ whose volume you know and an operator T on $\mathbf{R}^3$ such that $T(\Omega)$ equals the ellipsoid above.

Solution. Let E denote the ellipsoid above. Let

$$\Omega = \{(x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2 < 1\}$$

be the open unit ball of $\mathbf{R}^3$, whose volume is the familiar number $\frac{4\pi}{3}$, and define $T \in \mathcal{L}(\mathbf{R}^3)$ by

$$T(x, y, z) = (ax, by, cz).$$

First we check that $T(\Omega) = E$.
Suppose $(x, y, z) \in \Omega$. Then $T(x, y, z) = (ax, by, cz)$ and

$$\frac{(ax)^2}{a^2} + \frac{(by)^2}{b^2} + \frac{(cz)^2}{c^2} = x^2 + y^2 + z^2 < 1,$$

so $T(x, y, z) \in E$. Hence $T(\Omega) \subseteq E$.
Conversely, suppose $(u, v, w) \in E$. Because a, b, c are nonzero we may set $(x, y, z) = \left(\frac{u}{a}, \frac{v}{b}, \frac{w}{c}\right)$. Then

$$x^2 + y^2 + z^2 = \frac{u^2}{a^2} + \frac{v^2}{b^2} + \frac{w^2}{c^2} < 1,$$

so $(x, y, z) \in \Omega$, and $T(x, y, z) = (u, v, w)$. Hence $E \subseteq T(\Omega)$. Thus $T(\Omega) = E$.

Next we compute $\det T$. The matrix of T with respect to the standard basis e_1, e_2, e_3 of $\mathbf{R}^3$ is the diagonal matrix

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix},$$

which is upper triangular with a, b, c on its diagonal. Hence $\det T = abc$ by 9.53 and 9.48. (Equivalently: e_1, e_2, e_3 is a basis of eigenvectors of T with corresponding eigenvalues a, b, c, so the last bullet point of 9.42 gives $\det T = abc$.)
Now 9.61 (an operator changes volume by a factor of the absolute value of its determinant) gives

$$\text{volume } E = \text{volume } T(\Omega) = |\det T| \,(\text{volume } \Omega) = abc \cdot \frac{4\pi}{3} = \frac{4\pi abc}{3},$$

where we used $|\det T| = |abc| = abc$ because a, b, c are positive.
Thus the ellipsoid has volume $\frac{4}{3}\pi abc$. As a sanity check, taking $a = b = c = 1$ recovers the volume $\frac{4\pi}{3}$ of the unit ball.

Problem 9C.18. Suppose that A is an invertible square matrix. Prove that Hadamard's inequality (9.66) is an equality if and only if each column of A is orthogonal to the other columns.

Solution. Let A be an invertible n-by-n matrix and let $v_1, \ldots, v_n \in \mathbf{F}^n$ denote its columns. Hadamard's inequality (9.66) states that $|\det A| \le \prod_{k=1}^n \|v_k\|$; we must show that equality holds if and only if $\langle v_j, v_k \rangle = 0$ whenever $j \ne k$.
Because A is invertible, its columns are linearly independent, so the QR factorization (7.58) provides a unitary matrix Q and an upper-triangular matrix R whose diagonal entries are positive numbers, with $A = QR$. As in the proof of 9.66, write $R_{\cdot,k}$ for the k^{th} column of R. Then $v_k = QR_{\cdot,k}$, and because Q is an isometry,

$$\|v_k\| = \|QR_{\cdot,k}\| = \|R_{\cdot,k}\| \qquad \text{for each } k = 1, \ldots, n.$$

Furthermore $|\det A| = |\det Q|\,|\det R| = |\det R| = \prod_{k=1}^n R_{k,k}$, using 9.49(b), then 9.58, then 9.48.
Thus Hadamard's inequality for A reads

$$\prod_{k=1}^n R_{k,k} \le \prod_{k=1}^n \|R_{\cdot,k}\|,$$

and it is proved factor by factor: because R is upper triangular, $R_{j,k} = 0$ for $j > k$, so

$$\|R_{\cdot,k}\|^2 = \sum_{j=1}^k |R_{j,k}|^2 = R_{k,k}^2 + \sum_{j=1}^{k-1} |R_{j,k}|^2 \ge R_{k,k}^2,$$

which gives $0 < R_{k,k} \le \|R_{\cdot,k}\|$, with equality if and only if $R_{j,k} = 0$ for all $j < k$.
Suppose now that some factor is strict, say $R_{m,m} < \|R_{\cdot,m}\|$. Since every factor on each side is a positive number and $R_{k,k} \le \|R_{\cdot,k}\|$ for all k, multiplying the inequalities gives $\prod_k R_{k,k} < \prod_k \|R_{\cdot,k}\|$. Hence

$$|\det A| = \prod_{k=1}^n \|v_k\| \iff R_{k,k} = \|R_{\cdot,k}\| \text{ for each } k \iff R_{j,k} = 0 \text{ whenever } j < k.$$

Combined with the upper-triangularity of R, the last condition says exactly that R is a diagonal matrix. So it suffices to prove:

$$R \text{ is diagonal} \iff \text{the columns of } A \text{ are pairwise orthogonal.}$$

First suppose R is diagonal. Then $R_{\cdot,k} = R_{k,k} e_k$, so $v_k = QR_{\cdot,k} = R_{k,k} Q_{\cdot,k}$, where $Q_{\cdot,k} = Qe_k$ is the k^{th} column of Q. Because Q is unitary, it maps the orthonormal basis $e_1, \ldots, e_n$ to an orthonormal list, so $Q_{\cdot,1}, \ldots, Q_{\cdot,n}$ is orthonormal. Hence for $j \ne k$,

$$\langle v_j, v_k \rangle = R_{j,j}\, \overline{R_{k,k}}\, \langle Q_{\cdot,j}, Q_{\cdot,k} \rangle = 0.$$

Thus each column of A is orthogonal to the other columns.
Conversely, suppose the columns of A are pairwise orthogonal. From $A = QR$ we get $R = Q^{-1}A = Q^*A$, so $R_{\cdot,k} = Q^*v_k$. Because Q^* is unitary it preserves inner products, so for $j \neq k$,

$$\langle R_{\cdot,j}, R_{\cdot,k}\rangle = \langle Q^*v_j, Q^*v_k\rangle = \langle v_j, v_k\rangle = 0;$$

that is, the columns of R are pairwise orthogonal. We now show by induction on k that $R_{\cdot,k} = R_{k,k}e_k$, which says R is diagonal. For $k = 1$ this holds because R is upper triangular. Suppose $1 < k \leq n$ and $R_{\cdot,j} = R_{j,j}e_j$ for every $j < k$. Then for each $j < k$,

$$0 = \langle R_{\cdot,k}, R_{\cdot,j}\rangle = \langle R_{\cdot,k}, R_{j,j}e_j\rangle = \overline{R_{j,j}}\, R_{j,k},$$

and $R_{j,j} > 0$, so $R_{j,k} = 0$. Since also $R_{j,k} = 0$ for $j > k$ (upper-triangularity), we conclude $R_{\cdot,k} = R_{k,k}e_k$, completing the induction.
Therefore Hadamard's inequality is an equality for the invertible matrix A if and only if each column of A is orthogonal to the other columns.
Two remarks. First, the direction "pairwise orthogonal columns imply equality" also has a one-line proof that does not need A to be invertible: if the columns of A are pairwise orthogonal, then $(A^*A)_{j,k} = \langle v_k, v_j\rangle$ shows that A^*A is the diagonal matrix with $\|v_1\|^2, \ldots, \|v_n\|^2$ on the diagonal, so $|\det A|^2 = \det(A^*A) = \prod_{k=1}^n \|v_k\|^2$, by 9.60 applied to the operator on $\mathbf{F}^n$ whose matrix with respect to the standard basis is A (the matrix of the adjoint of that operator is A^*, by 7.9). Second, invertibility really is needed for the direction proved via the QR factorization: the 3-by-3 matrix with columns $v_1 = (1, 0, 0)$, $v_2 = (1, 1, 0)$, $v_3 = (0, 0, 0)$ satisfies $|\det A| = 0 = 1 \cdot \sqrt{2} \cdot 0 = \prod_{k=1}^3 \|v_k\|$, so Hadamard's inequality is an equality even though v_1 and v_2 are not orthogonal.

Problem 9C.19. Suppose V is an inner product space, $e_1, \ldots, e_n$ is an orthonormal basis of V, and $T \in \mathcal{L}(V)$ is a positive operator.
(a) Prove that $\det T \leq \prod_{k=1}^n \langle Te_k, e_k\rangle$.
(b) Prove that if T is invertible, then the inequality in (a) is an equality if and only if e_k is an eigenvector of T for each $k = 1, \ldots, n$.

Solution. Let $R = \sqrt{T}$ be the unique positive square root of T (7.39 and 7.40), so R is self-adjoint and $R^*R = R^2 = T$. Let

$$B = \mathcal{M}(R, (e_1, \ldots, e_n)),$$

and write $B_{\cdot,k} \in \mathbf{F}^n$ for the k^{th} column of B. Two computations set everything up.
First, by 9.49(a) and 9.56(c),

$$\det T = \det(R^*R) = (\det R^*)(\det R) = \overline{\det R}\, \det R = |\det R|^2 = |\det B|^2,$$

where the last equality uses $\det R = \det B$ (9.53).
Second, for each k,

$$\langle Te_k, e_k\rangle = \langle R^*Re_k, e_k\rangle = \langle Re_k, Re_k\rangle = \|Re_k\|^2.$$

Because the k^{th} column of B consists of the coefficients in $Re_k = \sum_{j=1}^n B_{j,k}e_j$, and because $e_1, \ldots, e_n$ is orthonormal, Parseval's identity 6.30(b) gives

$$\|Re_k\|^2 = \sum_{j=1}^n |B_{j,k}|^2 = \|B_{\cdot,k}\|^2,$$

the last norm being the one coming from the standard inner product on $\mathbf{F}^n$. Hence

$$\langle Te_k, e_k\rangle = \|B_{\cdot,k}\|^2 \qquad \text{for each } k = 1, \ldots, n.$$

(a) Apply Hadamard's inequality (9.66) to the matrix B, whose columns are $B_{\cdot,1}, \ldots, B_{\cdot,n}$:

$$|\det B| \leq \prod_{k=1}^n \|B_{\cdot,k}\|.$$

Both sides are nonnegative, so squaring preserves the inequality:

$$\det T = |\det B|^2 \le \prod_{k=1}^{n} \|B_{\cdot,k}\|^2 = \prod_{k=1}^{n} \langle Te_k, e_k\rangle.$$

This is the desired inequality.
(b) Now suppose T is invertible. Then $\det T \ne 0$ (9.50), so $\det B \ne 0$ and hence B is an invertible matrix.
Because $x \mapsto x^2$ is injective on $[0,\infty)$ and both $|\det B|$ and $\prod_k \|B_{\cdot,k}\|$ are nonnegative, the inequality in (a) is an equality if and only if Hadamard's inequality for B is an equality. Since B is invertible, Exercise 18 says that this happens if and only if the columns of B are pairwise orthogonal.
Finally, translate that condition back to T. Since $e_1, \dots, e_n$ is orthonormal, 6.30(c) says that taking coordinates with respect to $e_1, \dots, e_n$ preserves inner products, so for $j \ne k$,

$$\langle B_{\cdot,j}, B_{\cdot,k}\rangle = \langle Re_j, Re_k\rangle = \langle R^* Re_j, e_k\rangle = \langle Te_j, e_k\rangle.$$

Thus the columns of B are pairwise orthogonal if and only if

$$\langle Te_j, e_k\rangle = 0 \qquad \text{whenever } j \ne k.$$

By 6.30(a) we have $Te_j = \sum_{k=1}^{n} \langle Te_j, e_k\rangle e_k$ for each j. Hence the displayed condition holds if and only if $Te_j = \langle Te_j, e_j\rangle e_j$ for each j, which (because each e_j is a nonzero vector) says exactly that each e_j is an eigenvector of T, with corresponding eigenvalue $\langle Te_j, e_j\rangle$.
Putting the equivalences together: if T is invertible, then $\det T = \prod_{k=1}^{n} \langle Te_k, e_k\rangle$ if and only if e_k is an eigenvector of T for each $k = 1, \dots, n$.
Two remarks. The implication "each e_k is an eigenvector $\implies$ equality" does not need invertibility: if $Te_k = \lambda_k e_k$ for each k, then $\langle Te_k, e_k\rangle = \lambda_k \|e_k\|^2 = \lambda_k$, while the last bullet point of 9.42 gives $\det T = \lambda_1 \cdots \lambda_n$; hence both sides of (a) equal $\lambda_1 \cdots \lambda_n$. Invertibility is used only for the converse, and it cannot be dropped there. For a counterexample, work in $\mathbf{R}^3$ with the standard basis e_1, e_2, e_3 and let T be the orthogonal projection onto the one-dimensional subspace spanned by $v = \frac{1}{\sqrt{2}}(1,1,0)$, so $Tu = \langle u, v\rangle v$ for $u \in \mathbf{R}^3$. Every orthogonal projection is positive, since T is self-adjoint and $\langle Tu, u\rangle = |\langle u, v\rangle|^2 \ge 0$. Here $Te_3 = 0$, so $\langle Te_3, e_3\rangle = 0$ and hence $\prod_{k=1}^{3} \langle Te_k, e_k\rangle = 0$; also $\det T = 0$ because T is not injective. So the inequality in (a) is an equality. But $Te_1 = \frac{1}{\sqrt{2}} v = \left(\frac{1}{2}, \frac{1}{2}, 0\right)$, which is not a scalar multiple of e_1, so e_1 is not an eigenvector of T.

Problem 9C.20. Suppose A is an n-by-n matrix, and suppose c is such that $|A_{j,k}| \le c$ for all $j, k \in \{1, \dots, n\}$. Prove that

$$|\det A| \le c^n n^{n/2}.$$

[The formula for the determinant of a matrix (9.46) shows that $|\det A| \le c^n n!$. However, the estimate given by this exercise is much better. For example, if $c = 1$ and $n = 100$, then $c^n n! \approx 10^{158}$, but the estimate given by this exercise is the much smaller number 10^{100}. If n is an integer power of 2, then the inequality above is sharp and cannot be improved.]

Solution. Let $v_1, \dots, v_n \in \mathbf{F}^n$ denote the columns of A, so that $v_k = (A_{1,k}, \dots, A_{n,k})$. Note first that $c \ge 0$, because $0 \le |A_{1,1}| \le c$.
Fix $k \in \{1, \dots, n\}$. Using the hypothesis $|A_{j,k}| \le c$ for each j,

$$\|v_k\|^2 = \sum_{j=1}^{n} |A_{j,k}|^2 \le \sum_{j=1}^{n} c^2 = nc^2.$$

Taking nonnegative square roots gives

$$\|v_k\| \le c\sqrt{n} \qquad \text{for each } k = 1, \dots, n.$$

Now Hadamard's inequality (9.66) applied to A, followed by the bound above (all the quantities involved are nonnegative, so the factors may be multiplied), gives

$$|\det A| \le \prod_{k=1}^{n} \|v_k\| \le \prod_{k=1}^{n} \left(c\sqrt{n}\right) = \left(c\sqrt{n}\right)^n = c^n n^{n/2},$$

as desired.
Regarding the sharpness mentioned in the exercise, suppose $c > 0$ and equality holds. Then $|\det A| = c^n n^{n/2} > 0$, so A is invertible, and the chain of inequalities above collapses: $|\det A| = \prod_{k=1}^{n} \|v_k\| = c^n n^{n/2}$. The second of these equalities, together with $\|v_k\| \le c\sqrt{n}$ for every k and the positivity of all the factors, forces $\|v_k\| = c\sqrt{n}$ for each k; then $\sum_{j=1}^{n} |A_{j,k}|^2 = nc^2$ with each $|A_{j,k}| \le c$ forces $|A_{j,k}| = c$ for all j, k. The first equality, by Exercise 18, forces the columns of A to be pairwise orthogonal. For $c = 1$ and $\mathbf{F} = \mathbf{R}$ such matrices are exactly the Hadamard matrices, and Sylvester's doubling construction $H \mapsto \begin{pmatrix} H & H \\ H & -H \end{pmatrix}$ (which starts from the 1-by-1 matrix (1) and takes a Hadamard matrix of size m to one of size $2m$, since the doubled matrix again has entries ± 1 and pairwise orthogonal columns) produces one for every size n that is an integer power of 2.

Problem 9C.21. Suppose n is a positive integer and $\delta\colon \mathbf{C}^{n,n} \to \mathbf{C}$ is a function such that

$$\delta(AB) = \delta(A) \cdot \delta(B)$$

for all $A, B \in \mathbf{C}^{n,n}$ and $\delta(A)$ equals the product of the diagonal entries of A for each diagonal matrix $A \in \mathbf{C}^{n,n}$. Prove that

$$\delta(A) = \det A$$

for all $A \in \mathbf{C}^{n,n}$.
[Recall that $\mathbf{C}^{n,n}$ denotes the set of n-by-n matrices with entries in $\mathbf{C}$. This exercise shows that the determinant is the unique function defined on square matrices that is multiplicative and has the desired behavior on diagonal matrices. This result is analogous to Exercise 10 in Section 8D, which shows that the trace is uniquely determined by its algebraic properties.]

Solution. Throughout, $e_1, \dots, e_n$ is the standard basis of $\mathbf{C}^n$, and for $j, k \in \{1, \dots, n\}$ we write $E^{(j,k)}$ for the matrix whose entry in row j, column k equals 1 and whose other entries are 0. Recall the multiplication rule $E^{(j,k)}E^{(p,q)} = 0$ if $k \ne p$ and $E^{(j,k)}E^{(k,q)} = E^{(j,q)}$. We freely identify an n-by-n matrix with the operator on $\mathbf{C}^n$ whose matrix with respect to $e_1, \dots, e_n$ it is; under this identification a matrix is invertible if and only if the corresponding operator is invertible, and det of the matrix equals det of the operator.
Step 1: $\delta(I) = 1$. Indeed, I is a diagonal matrix whose diagonal entries are all 1, so by hypothesis $\delta(I)$ is the product of those entries, namely 1.
Step 2: if A is invertible, then $\delta(A) \ne 0$; and $\delta(SAS^{-1}) = \delta(A)$ for every invertible $S \in \mathbf{C}^{n,n}$. For the first assertion, multiplicativity and Step 1 give

$$\delta(A)\,\delta(A^{-1}) = \delta(AA^{-1}) = \delta(I) = 1,$$

so $\delta(A) \ne 0$ (and $\delta(A^{-1}) = 1/\delta(A)$). For the second, using multiplicativity and the fact that complex numbers commute,

$$\delta(SAS^{-1}) = \delta(S)\,\delta(A)\,\delta(S^{-1}) = \delta(A)\,\delta(S)\,\delta(S^{-1}) = \delta(A) \cdot 1 = \delta(A).$$

Step 3: if A is not invertible, then $\delta(A) = 0 = \det A$. That $\det A = 0$ follows from 9.50. For δ: because A is not invertible, the corresponding operator is not injective, so there exists $v \in \mathbf{C}^n$ with $v \ne 0$ and $Av = 0$. Extend v to a basis, writing the basis in the order $v_1, \dots, v_{n-1}, v$ (possible by 2.32). Let Q be the matrix whose columns are $v_1, \dots, v_{n-1}, v$; then Q is invertible because its columns form a basis of $\mathbf{C}^n$. The k^{th} column of AQ is A applied to the k^{th} column of Q, so the columns of AQ are $Av_1, \dots, Av_{n-1}, 0$. In particular the last column of AQ is 0.

Let D_0 be the diagonal matrix with diagonal entries $1, \dots, 1, 0$. For any matrix M and any diagonal matrix D we have $(MD)_{j,k} = \sum_m M_{j,m} D_{m,k} = D_{k,k} M_{j,k}$, so right multiplication by D multiplies the k^{th} column of M by $D_{k,k}$. Applying this with $M = AQ$ and $D = D_0$: the first $n-1$ columns of $(AQ)D_0$ agree with those of AQ and the last column of $(AQ)D_0$ is 0, which is also the last column of AQ. Hence

$$(AQ)D_0 = AQ.$$

Applying δ and using $\delta(D_0) = 1 \cdots 1 \cdot 0 = 0$ (the hypothesis on diagonal matrices),

$$\delta(A)\,\delta(Q) = \delta(AQ) = \delta(AQ)\,\delta(D_0) = 0.$$

Since Q is invertible, $\delta(Q) \neq 0$ by Step 2, and therefore $\delta(A) = 0$.
Step 4: $\delta(I + aE^{(j,k)}) = 1$ whenever $a \in \mathbf{C}$ and $j \neq k$. Write $T_{j,k}(a) = I + aE^{(j,k)}$. Because $j \neq k$ we have $E^{(j,k)}E^{(j,k)} = 0$, so for all $a, b \in \mathbf{C}$,

$$T_{j,k}(a)\,T_{j,k}(b) = I + (a+b)E^{(j,k)} + ab\,E^{(j,k)}E^{(j,k)} = T_{j,k}(a+b).$$

In particular $T_{j,k}(a)T_{j,k}(-a) = T_{j,k}(0) = I$, so $T_{j,k}(a)$ is invertible, and $T_{j,k}(a)^2 = T_{j,k}(2a)$.
Now let D be the diagonal matrix with $D_{j,j} = 2$ and all other diagonal entries equal to 1; note D is invertible and, because $j \neq k$, $D_{k,k} = 1$. For a diagonal invertible D one computes $(DE^{(j,k)}D^{-1})_{p,q} = D_{p,p}(E^{(j,k)})_{p,q}(D^{-1})_{q,q}$, which vanishes unless $(p,q) = (j,k)$ and equals $D_{j,j}/D_{k,k}$ there. Hence

$$D\,T_{j,k}(a)\,D^{-1} = I + a\frac{D_{j,j}}{D_{k,k}}E^{(j,k)} = T_{j,k}(2a) = T_{j,k}(a)^2.$$

Applying δ and using Step 2 (conjugation invariance) and multiplicativity,

$$\delta\big(T_{j,k}(a)\big) = \delta\big(D\,T_{j,k}(a)\,D^{-1}\big) = \delta\big(T_{j,k}(a)^2\big) = \delta\big(T_{j,k}(a)\big)^2.$$

Thus $\delta(T_{j,k}(a))$ equals its own square, so it equals 0 or 1; it is not 0 because $T_{j,k}(a)$ is invertible (Step 2). Hence $\delta(T_{j,k}(a)) = 1$.
Step 5 (a lemma about matrices, with no reference to δ): for every positive integer m, every m-by-m upper-triangular matrix all of whose diagonal entries equal 1 is a product of matrices of the form $I + aE^{(j,k)}$ with $1 \leq j < k \leq m$ and $a \in \mathbf{C}$.
We induct on m. If $m = 1$, then the only such matrix is $(1) = I$, which is the empty product. Suppose $m > 1$ and the claim holds for $m-1$. Let N be m-by-m, upper triangular, with 1's on the diagonal. Write N in block form

$$N = \begin{pmatrix} 1 & r \\ 0 & N' \end{pmatrix},$$

where r is the row of length $m-1$ whose entries are $r_k = N_{1,k}$ for $k = 2, \dots, m$, and N' is the $(m-1)$-by-$(m-1)$ upper-triangular matrix with 1's on the diagonal obtained from N by deleting its first row and first column. Block multiplication gives

$$\begin{pmatrix} 1 & 0 \\ 0 & N' \end{pmatrix}\begin{pmatrix} 1 & r \\ 0 & I \end{pmatrix} = \begin{pmatrix} 1 & r \\ 0 & N' \end{pmatrix} = N.$$

For the second factor, note that $\begin{pmatrix} 1 & r \\ 0 & I \end{pmatrix} = I + \sum_{k=2}^{m} r_k E^{(1,k)}$ and that $E^{(1,j)}E^{(1,k)} = 0$ for all $j, k \geq 2$; hence expanding the product below leaves only the terms shown:

$$\prod_{k=2}^{m}\left(I + r_k E^{(1,k)}\right) = I + \sum_{k=2}^{m} r_k E^{(1,k)} = \begin{pmatrix} 1 & r \\ 0 & I \end{pmatrix},$$

so the second factor is a product of matrices of the required form (each with $j = 1 < k$). For the first factor, let ι send an $(m-1)$-by-$(m-1)$ matrix M to $\begin{pmatrix} 1 & 0 \\ 0 & M \end{pmatrix}$; block multiplication shows $\iota(M)\iota(M'') = \iota(MM'')$, and $\iota(I + aE^{(j,k)}) = I + aE^{(j+1,k+1)}$. By the induction hypothesis N' is a

product of matrices $I+aE^{(j,k)}$ with $1 \le j < k \le m-1$, so $\iota(N')$ is a product of matrices $I+aE^{(j+1,k+1)}$ with $1 < j+1 < k+1 \le m$, again of the required form. This completes the induction.
Combining Step 5 with Step 4 and multiplicativity: if $N \in \mathbf{C}^{n,n}$ is upper triangular with all diagonal entries equal to 1, then $\delta(N)$ is a product of numbers each equal to 1, so

$$\delta(N) = 1.$$

Step 6: $\delta(A) = \det A$ for invertible A. Let S be the operator on $\mathbf{C}^n$ whose matrix with respect to $e_1, \dots, e_n$ is A. By 5.47 there is a basis $u_1, \dots, u_n$ of $\mathbf{C}^n$ with respect to which S has an upper-triangular matrix U; thus $Su_k = \sum_{j=1}^{n} U_{j,k} u_j$ for each k. Let P be the invertible matrix whose columns are $u_1, \dots, u_n$, so $Pe_k = u_k$. Then for each k,

$$APe_k = Su_k = \sum_{j=1}^{n} U_{j,k} u_j = \sum_{j=1}^{n} U_{j,k} Pe_j = P\left(\sum_{j=1}^{n} U_{j,k} e_j\right) = PUe_k,$$

so $AP = PU$ and hence $A = PUP^{-1}$.
By Step 2, $\delta(A) = \delta(U)$. Because A is invertible and $U = P^{-1}AP$, the matrix U is invertible; equivalently, U is the matrix of the invertible operator S with respect to the basis $u_1, \dots, u_n$, so 9.53 and 9.50 give $\det U = \det S \neq 0$. Since $\det U = U_{1,1} \cdots U_{n,n}$ by 9.48, every diagonal entry of U is nonzero. Let D be the diagonal matrix with diagonal entries $U_{1,1}, \dots, U_{n,n}$, which is invertible, and set $N = D^{-1}U$. Then $N_{j,k} = U_{j,k}/U_{j,j}$, so N is upper triangular with all diagonal entries equal to 1. Hence $\delta(N) = 1$ by the conclusion of Step 5, and therefore

$$\delta(U) = \delta(DN) = \delta(D)\,\delta(N) = U_{1,1} \cdots U_{n,n} = \det U,$$

using the hypothesis on diagonal matrices for $\delta(D)$ and 9.48 for $\det U$. Finally, 9.49(b) applied twice gives $\det A = \det(PUP^{-1}) = (\det P)(\det U)\det(P^{-1})$, while 9.49(b) applied to $PP^{-1} = I$ gives $(\det P)\det(P^{-1}) = \det I = 1$. Since complex numbers commute,

$$\det A = (\det U)(\det P)\det(P^{-1}) = \det U.$$

Combining the last three displays yields $\delta(A) = \delta(U) = \det U = \det A$.
Steps 3 and 6 together cover all $A \in \mathbf{C}^{n,n}$, so $\delta(A) = \det A$ for every $A \in \mathbf{C}^{n,n}$, as desired.
Note that both hypotheses were needed, and only through a few consequences: multiplicativity gave conjugation invariance and the reduction to upper-triangular matrices, while the value of δ on diagonal matrices was used exactly three times, for I, for the singular matrix D_0, and for the diagonal part D of U.

9.4 Exercises 9D

Problem 9D.1. Suppose $v \in V$ and $w \in W$. Prove that $v \otimes w = 0$ if and only if $v = 0$ or $w = 0$.

Solution. Recall from 9.71 that $V \otimes W = \mathcal{B}(V', W')$ and that $v \otimes w$ is the bilinear functional on $V' \times W'$ defined by

$$(v \otimes w)(\varphi, \tau) = \varphi(v)\,\tau(w)$$

for all $(\varphi, \tau) \in V' \times W'$. Thus $v \otimes w = 0$ means precisely that $\varphi(v)\tau(w) = 0$ for every $\varphi \in V'$ and every $\tau \in W'$.
First suppose $v = 0$ or $w = 0$. If $v = 0$, then $\varphi(v) = 0$ for every $\varphi \in V'$ (every linear functional maps 0 to 0), and hence $(v \otimes w)(\varphi, \tau) = \varphi(v)\tau(w) = 0$ for all $(\varphi, \tau) \in V' \times W'$. Thus $v \otimes w = 0$. The case $w = 0$ is handled the same way, using $\tau(w) = 0$ for every $\tau \in W'$.
Now suppose $v \neq 0$ and $w \neq 0$; we show $v \otimes w \neq 0$. Because $v \neq 0$, the list v (of length one) is linearly independent in V, so by 2.32 it extends to a basis $v, u_2, \dots, u_n$ of V. By the linear map lemma 3.4

there is a linear map $\varphi\colon V \to \mathbb{F}$ with

$$\varphi(v) = 1 \quad \text{and} \quad \varphi(u_j) = 0 \text{ for } j = 2, \dots, n.$$

Then $\varphi \in V'$ and $\varphi(v) = 1$. Similarly, because $w \neq 0$, the same argument applied to W produces $\tau \in W'$ with $\tau(w) = 1$. Hence

$$(v \otimes w)(\varphi, \tau) = \varphi(v)\,\tau(w) = 1 \cdot 1 = 1 \neq 0,$$

so the bilinear functional $v \otimes w$ is not the zero element of $\mathcal{B}(V', W') = V \otimes W$. That is, $v \otimes w \neq 0$. Taking the contrapositive of this last paragraph: if $v \otimes w = 0$, then $v = 0$ or $w = 0$. Combined with the first paragraph, this proves that $v \otimes w = 0$ if and only if $v = 0$ or $w = 0$.

Problem 9D.2. Give an example of six distinct vectors $v_1, v_2, v_3, w_1, w_2, w_3$ in $\mathbb{R}^3$ such that

$$v_1 \otimes w_1 + v_2 \otimes w_2 + v_3 \otimes w_3 = 0$$

but none of $v_1 \otimes w_1$, $v_2 \otimes w_2$, $v_3 \otimes w_3$ is a scalar multiple of another element of this list.

Solution. Let e_1, e_2, e_3 denote the standard basis of $\mathbb{R}^3$ and take

$$\begin{aligned} v_1 &= (1,0,0) = e_1, & w_1 &= (0,1,0) = e_2, \\ v_2 &= (2,0,0) = 2e_1, & w_2 &= (0,0,1) = e_3, \\ v_3 &= (-1,0,0) = -e_1, & w_3 &= (0,1,2) = e_2 + 2e_3. \end{aligned}$$

These six vectors are distinct, as one sees by inspecting the six displayed triples: $(1,0,0)$, $(2,0,0)$, $(-1,0,0)$, $(0,1,0)$, $(0,0,1)$, $(0,1,2)$ are pairwise different.
To verify the required equation, put $A = e_1 \otimes e_2$ and $B = e_1 \otimes e_3$. Using the bilinearity of the tensor product (9.73) we get

$$v_1 \otimes w_1 = A, \qquad v_2 \otimes w_2 = (2e_1) \otimes e_3 = 2B,$$

$$v_3 \otimes w_3 = (-e_1) \otimes (e_2 + 2e_3) = -(e_1 \otimes e_2) - 2(e_1 \otimes e_3) = -A - 2B.$$

Hence

$$v_1 \otimes w_1 + v_2 \otimes w_2 + v_3 \otimes w_3 = A + 2B + (-A - 2B) = 0,$$

as desired.
It remains to show that none of the three elements A, $2B$, $-A-2B$ is a scalar multiple of another one. The list e_1 is linearly independent in $\mathbb{R}^3$ and the list e_2, e_3 is linearly independent in $\mathbb{R}^3$, so 9.74(a) tells us that the list A, B (that is, $e_1 \otimes e_2$, $e_1 \otimes e_3$) is linearly independent in $\mathbb{R}^3 \otimes \mathbb{R}^3$. Now suppose one of our three elements equals λ times another one, for some $\lambda \in \mathbb{R}$. Each of the six possible equations, after moving everything to one side, has the form $\alpha A + \beta B = 0$, and linear independence of A, B forces $\alpha = \beta = 0$:

- $A = \lambda(2B)$ gives $A - 2\lambda B = 0$, so the coefficient of A satisfies $1 = 0$, a contradiction.
- $2B = \lambda A$ gives $-\lambda A + 2B = 0$, so $2 = 0$, a contradiction.
- $A = \lambda(-A - 2B)$ gives $(1+\lambda)A + 2\lambda B = 0$, so $\lambda = 0$ and $1 + \lambda = 0$, a contradiction.
- $-A - 2B = \lambda A$ gives $(-1-\lambda)A - 2B = 0$, so $2 = 0$, a contradiction.
- $2B = \lambda(-A - 2B)$ gives $\lambda A + (2 + 2\lambda)B = 0$, so $\lambda = 0$ and $2 + 2\lambda = 0$, a contradiction.
- $-A - 2B = \lambda(2B)$ gives $-A - (2 + 2\lambda)B = 0$, so $-1 = 0$, a contradiction.

Thus no element of the list $v_1 \otimes w_1$, $v_2 \otimes w_2$, $v_3 \otimes w_3$ is a scalar multiple of another element of the list, completing the example.
As a check, one can also see all of this through the matrix identification of 9.76: $v \otimes w$ corresponds to the 3-by-3 matrix whose entry in row j, column k is $v_j w_k$. Since v_1, v_2, v_3 each have only a first coordinate, the three matrices here have all rows equal to 0 except for row 1, and their first rows are $(0,1,0)$, $(0,0,2)$, and $(0,-1,-2)$. These three rows sum to $(0,0,0)$ and are visibly pairwise non-proportional.

Problem 9D.3. Suppose that $v_1, \dots, v_m$ is a linearly independent list in V. Suppose also that $w_1, \dots, w_m$ is a list in W such that

$$v_1 \otimes w_1 + \cdots + v_m \otimes w_m = 0.$$

Prove that $w_1 = \cdots = w_m = 0$.

Solution. Fix $k \in \{1, \dots, m\}$; we show $w_k = 0$.
Because $v_1, \dots, v_m$ is linearly independent, 2.32 tells us that this list extends to a basis $v_1, \dots, v_m, u_{m+1}, \dots, u_n$ of V. By the linear map lemma 3.4, there is a (unique) linear functional $\varphi \in V'$ with

$$\varphi(v_j) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k, \end{cases} \qquad \text{and} \qquad \varphi(u_i) = 0 \text{ for } i = m+1, \dots, n.$$

(This is the same construction used in the proof of 9.74.)
Now let $\tau \in W'$ be arbitrary. Recall from 9.71 that elements of $V \otimes W$ are bilinear functionals on $V' \times W'$, with $(v \otimes w)(\varphi, \tau) = \varphi(v)\tau(w)$. Applying the zero functional $v_1 \otimes w_1 + \cdots + v_m \otimes w_m$ to the pair (φ, τ) gives

$$0 = \sum_{j=1}^{m} (v_j \otimes w_j)(\varphi, \tau) = \sum_{j=1}^{m} \varphi(v_j)\, \tau(w_j) = \tau(w_k),$$

where the last equality holds because $\varphi(v_j) = 0$ for $j \neq k$ and $\varphi(v_k) = 1$.
Thus $\tau(w_k) = 0$ for every $\tau \in W'$. If we had $w_k \neq 0$, then (exactly as in the solution to Exercise 9D.1) the linearly independent list w_k could be extended to a basis of W by 2.32, and the linear map lemma 3.4 would produce $\tau \in W'$ with $\tau(w_k) = 1 \neq 0$, contradicting what we just proved. Hence $w_k = 0$.
Because $k \in \{1, \dots, m\}$ was arbitrary, we conclude $w_1 = \cdots = w_m = 0$.

Problem 9D.4. Suppose $\dim V > 1$ and $\dim W > 1$. Prove that

$$\{v \otimes w : (v, w) \in V \times W\}$$

is not a subspace of $V \otimes W$.
This exercise implies that if $\dim V > 1$ and $\dim W > 1$, then

$$\{v \otimes w : (v, w) \in V \times W\} \neq V \otimes W.$$

Solution. Write $E = \{v \otimes w : (v, w) \in V \times W\}$ for the set of so-called pure tensors. We show that E is not closed under addition, which implies that E is not a subspace of $V \otimes W$.
Because $\dim V > 1$, a basis of V has length at least 2; taking its first two vectors gives a linearly independent list v_1, v_2 in V. Similarly, because $\dim W > 1$ there is a linearly independent list w_1, w_2 in W. Both $v_1 \otimes w_1$ and $v_2 \otimes w_2$ belong to E. We claim that their sum does not.
Suppose, for the sake of contradiction, that there exist $v \in V$ and $w \in W$ with

$$v_1 \otimes w_1 + v_2 \otimes w_2 = v \otimes w.$$

By 2.32, extend the linearly independent list v_1, v_2 to a basis $v_1, \dots, v_n$ of V (here $n = \dim V \geq 2$). Write

$$v = a_1 v_1 + \cdots + a_n v_n$$

with $a_1, \dots, a_n \in \mathbb{F}$. Then by the bilinearity of the tensor product (9.73),

$$v \otimes w = \sum_{j=1}^{n} (a_j v_j) \otimes w = \sum_{j=1}^{n} v_j \otimes (a_j w).$$

Substituting this into the displayed equation and moving everything to one side gives

$$v_1 \otimes (a_1 w - w_1) + v_2 \otimes (a_2 w - w_2) + \sum_{j=3}^{n} v_j \otimes (a_j w) = 0.$$

The list $v_1, \ldots, v_n$ is linearly independent (it is a basis), so Exercise 9D.3 applies to this equation and tells us that every vector appearing in the second slot is 0. In particular

$$w_1 = a_1 w \qquad \text{and} \qquad w_2 = a_2 w.$$

Thus w_1 and w_2 both lie in $\operatorname{span}(w)$. Now w_1, w_2 is a linearly independent list of length 2 in the vector space $\operatorname{span}(w)$, which is spanned by the list w of length 1. This contradicts 2.22, which says that the length of a linearly independent list is at most the length of a spanning list.
Therefore no such v, w exist: the element $v_1 \otimes w_1 + v_2 \otimes w_2$ of $V \otimes W$ is not in E, even though $v_1 \otimes w_1 \in E$ and $v_2 \otimes w_2 \in E$. So E is not closed under addition, and hence E is not a subspace of $V \otimes W$.
Finally, note the consequence stated in the exercise. The set $V \otimes W$ is of course a subspace of itself. Since E is not a subspace of $V \otimes W$, we must have $E \neq V \otimes W$ whenever $\dim V > 1$ and $\dim W > 1$. (Concretely, the element $v_1 \otimes w_1 + v_2 \otimes w_2$ produced above lies in $V \otimes W$ but not in E.)

Problem 9D.5. Suppose m and n are positive integers. For $v \in \mathbb{F}^m$ and $w \in \mathbb{F}^n$, identify $v \otimes w$ with an m-by-n matrix as in Example 9.76. With that identification, show that the set

$$\{v \otimes w : v \in \mathbb{F}^m \text{ and } w \in \mathbb{F}^n\}$$

is the set of m-by-n matrices (with entries in $\mathbb{F}$) that have rank at most one.

Solution. Recall the identification of 9.76: if $v = (v_1, \ldots, v_m) \in \mathbb{F}^m$ and $w = (w_1, \ldots, w_n) \in \mathbb{F}^n$, then $v \otimes w$ is identified with the m-by-n matrix whose entry in row j, column k equals $v_j w_k$. Denote this matrix by $M(v, w)$. Recall also that the rank of a matrix is by definition (3.58) its column rank, which by 3.52 is the dimension of the span of the columns of the matrix, viewed as vectors in $\mathbb{F}^{m,1}$.
Let $S = \{M(v, w) : v \in \mathbb{F}^m,\ w \in \mathbb{F}^n\}$ and let $R = \{A \in \mathbb{F}^{m,n} : \operatorname{rank} A \leq 1\}$. We prove $S = R$ by proving both inclusions.
First we show $S \subseteq R$. Suppose $v \in \mathbb{F}^m$ and $w \in \mathbb{F}^n$, and let $A = M(v, w)$, so $A_{j,k} = v_j w_k$. Let $\hat{v} \in \mathbb{F}^{m,1}$ denote the column vector whose j-th entry is v_j. For each $k \in \{1, \ldots, n\}$, column k of A is the column vector whose j-th entry is $v_j w_k$, which is exactly $w_k \hat{v}$. Hence every column of A lies in $\operatorname{span}(\hat{v})$, and therefore

$$\operatorname{span}(\text{columns of } A) \subseteq \operatorname{span}(\hat{v}).$$

Because $\dim \operatorname{span}(\hat{v}) \leq 1$, we get $\operatorname{rank} A \leq 1$ (using 2.37, which says that a subspace of a finite-dimensional vector space has dimension at most that of the whole space). Thus $A \in R$.
Now we show $R \subseteq S$. Suppose $A \in \mathbb{F}^{m,n}$ with $\operatorname{rank} A \leq 1$.
Case 1: $\operatorname{rank} A = 0$. Then the span of the columns of A is $\{0\}$, so every column of A is 0, meaning $A = 0$. Taking $v = 0 \in \mathbb{F}^m$ and any $w \in \mathbb{F}^n$ gives $M(v, w) = 0 = A$, so $A \in S$.
Case 2: $\operatorname{rank} A = 1$. Then the span of the columns of A is a one-dimensional subspace of $\mathbb{F}^{m,1}$; let $\hat{v}$ be a basis of it, and let $v = (v_1, \ldots, v_m) \in \mathbb{F}^m$ be the corresponding element of $\mathbb{F}^m$ (so the j-th entry of $\hat{v}$ is v_j). For each $k \in \{1, \ldots, n\}$, column k of A lies in $\operatorname{span}(\hat{v})$, so there is a scalar $w_k \in \mathbb{F}$ with

$$\text{column } k \text{ of } A = w_k \hat{v}.$$

Comparing entries in row j gives $A_{j,k} = v_j w_k$ for all $j \in \{1, \ldots, m\}$ and all $k \in \{1, \ldots, n\}$. Setting $w = (w_1, \ldots, w_n) \in \mathbb{F}^n$, we conclude $A = M(v, w)$, so $A \in S$.
Both inclusions hold, so

$$\{v \otimes w : v \in \mathbb{F}^m \text{ and } w \in \mathbb{F}^n\} = \{A \in \mathbb{F}^{m,n} : \operatorname{rank} A \leq 1\},$$

as desired.
Note that this gives another way to see Exercise 9D.4 for the concrete spaces $V = \mathbb{F}^m$ and $W = \mathbb{F}^n$ with $m, n > 1$: the set of matrices of rank at most one is not closed under addition, since for instance the two rank-one matrices with 1 in position $(1,1)$ (respectively $(2,2)$) and zeros elsewhere sum to a matrix of rank 2.

Problem 9D.6. Suppose m and n are positive integers. Give a description, analogous to Exercise 5, of the set of m-by-n matrices (with entries in $\mathbb{F}$) that have rank at most two.

Solution. The description is: with the identification of Example 9.76, the set of m-by-n matrices with entries in $\mathbb{F}$ that have rank at most two is

$$\{v_1 \otimes w_1 + v_2 \otimes w_2 : v_1, v_2 \in \mathbb{F}^m \text{ and } w_1, w_2 \in \mathbb{F}^n\}.$$

In other words, the matrices of rank at most two are exactly the sums of two matrices of rank at most one.
As in the solution to Exercise 9D.5, for $v = (v_1, \dots, v_m) \in \mathbb{F}^m$ and $w = (w_1, \dots, w_n) \in \mathbb{F}^n$ let $M(v, w)$ denote the m-by-n matrix whose entry in row j, column k is $v_j w_k$; this is the matrix identified with $v \otimes w$. For $u \in \mathbb{F}^m$ let $\hat{u} \in \mathbb{F}^{m,1}$ denote the corresponding column vector. Recall (3.58 and 3.52) that the rank of a matrix is the dimension of the span of its columns in $\mathbb{F}^{m,1}$.
Let $S = \{M(v_1, w_1) + M(v_2, w_2) : v_1, v_2 \in \mathbb{F}^m,\ w_1, w_2 \in \mathbb{F}^n\}$ and let $R = \{A \in \mathbb{F}^{m,n} : \operatorname{rank} A \le 2\}$.
First we show $S \subseteq R$. Suppose $v_1, v_2 \in \mathbb{F}^m$ and $w_1, w_2 \in \mathbb{F}^n$, and let $A = M(v_1, w_1) + M(v_2, w_2)$, so that

$$A_{j,k} = (v_1)_j (w_1)_k + (v_2)_j (w_2)_k.$$

For each k, column k of A is therefore

$$(w_1)_k \, \hat{v_1} + (w_2)_k \, \hat{v_2} \in \operatorname{span}(\hat{v_1}, \hat{v_2}).$$

Hence the span of the columns of A is contained in $\operatorname{span}(\hat{v_1}, \hat{v_2})$, a subspace of dimension at most 2 (a basis of $\operatorname{span}(\hat{v_1}, \hat{v_2})$ is a linearly independent list in that space, while $\hat{v_1}, \hat{v_2}$ spans it, so 2.22 gives the bound). By 2.37, $\operatorname{rank} A \le 2$, so $A \in R$.
Now we show $R \subseteq S$. Suppose $A \in \mathbb{F}^{m,n}$ with $r := \operatorname{rank} A \le 2$. The span C of the columns of A is a subspace of $\mathbb{F}^{m,1}$ with $\dim C = r \le 2$. Choose a basis of C and, if $r < 2$, pad it with zero vectors to obtain a list $\hat{v_1}, \hat{v_2}$ of length 2 in $\mathbb{F}^{m,1}$ whose span contains (indeed, equals) C. Let $v_1, v_2 \in \mathbb{F}^m$ be the corresponding elements of $\mathbb{F}^m$.
For each $k \in \{1, \dots, n\}$, column k of A lies in $C \subseteq \operatorname{span}(\hat{v_1}, \hat{v_2})$, so there are scalars $c_k, d_k \in \mathbb{F}$ with

$$\text{column } k \text{ of } A = c_k \, \hat{v_1} + d_k \, \hat{v_2}.$$

Comparing entries in row j gives $A_{j,k} = (v_1)_j c_k + (v_2)_j d_k$. Setting $w_1 = (c_1, \dots, c_n) \in \mathbb{F}^n$ and $w_2 = (d_1, \dots, d_n) \in \mathbb{F}^n$, we conclude

$$A = M(v_1, w_1) + M(v_2, w_2),$$

so $A \in S$.
Both inclusions hold, which proves the description stated at the beginning.
The same argument works verbatim with 2 replaced by any positive integer p: with the identification of 9.76, the set of m-by-n matrices of rank at most p is

$$\{v_1 \otimes w_1 + \cdots + v_p \otimes w_p : v_1, \dots, v_p \in \mathbb{F}^m \text{ and } w_1, \dots, w_p \in \mathbb{F}^n\}.$$

Exercise 9D.5 is the case $p = 1$.

Problem 9D.7. Suppose $\dim V > 2$ and $\dim W > 2$. Prove that

$$\{v_1 \otimes w_1 + v_2 \otimes w_2 : v_1, v_2 \in V \text{ and } w_1, w_2 \in W\} \neq V \otimes W.$$

Solution. Let T denote the set on the left side. We exhibit an element of $V \otimes W$ that is not in T.
Because $\dim V > 2$, a basis of V has length at least 3; taking its first three vectors gives a linearly independent list x_1, x_2, x_3 in V. Similarly, because $\dim W > 2$ there is a linearly independent list y_1, y_2, y_3 in W. Set

$$z = x_1 \otimes y_1 + x_2 \otimes y_2 + x_3 \otimes y_3 \in V \otimes W.$$

We claim $z \notin T$.
Suppose, for the sake of contradiction, that there exist $v_1, v_2 \in V$ and $w_1, w_2 \in W$ with

$$z = v_1 \otimes w_1 + v_2 \otimes w_2.$$

Let $U = \operatorname{span}(v_1, v_2)$ and let $r = \dim U$, so $r \leq 2$ (a basis of U is linearly independent in U while v_1, v_2 spans U, so 2.22 gives $r \leq 2$). Let $u_1, \ldots, u_r$ be a basis of U and extend it, using 2.32, to a basis $u_1, \ldots, u_n$ of V, where $n = \dim V$. (If $r = 0$, so that $v_1 = v_2 = 0$, then this list is empty and every sum below indexed by $p \leq r$ is the empty sum, equal to 0; the argument that follows still runs, and indeed 2.4 tells us that the span of the empty list is $\{0\} = U$.)
Since $v_1, v_2 \in U = \operatorname{span}(u_1, \ldots, u_r)$, we may write $v_i = \sum_{p=1}^{r} b_{i,p} u_p$ for $i = 1, 2$, with scalars $b_{i,p} \in \mathbb{F}$. By the bilinearity of the tensor product (9.73),

$$v_1 \otimes w_1 + v_2 \otimes w_2 = \sum_{p=1}^{r} u_p \otimes (b_{1,p} w_1 + b_{2,p} w_2) = \sum_{p=1}^{r} u_p \otimes c_p,$$

where $c_p = b_{1,p} w_1 + b_{2,p} w_2 \in W$ for $p = 1, \ldots, r$. Note the key point: this expression involves only $u_1, \ldots, u_r$, and no u_p with $p > r$.
On the other hand, write each x_l in the basis $u_1, \ldots, u_n$:

$$x_l = \sum_{p=1}^{n} \lambda_{l,p} u_p \qquad (l = 1, 2, 3)$$

with $\lambda_{l,p} \in \mathbb{F}$. Again by bilinearity (9.73),

$$z = \sum_{l=1}^{3} x_l \otimes y_l = \sum_{l=1}^{3} \sum_{p=1}^{n} u_p \otimes (\lambda_{l,p} y_l) = \sum_{p=1}^{n} u_p \otimes d_p, \qquad \text{where } d_p = \sum_{l=1}^{3} \lambda_{l,p}\, y_l.$$

Now subtract the two expressions for z. Setting $c_p = 0$ for $p > r$ (a harmless convention), we obtain

$$\sum_{p=1}^{n} u_p \otimes (d_p - c_p) = 0.$$

The list $u_1, \ldots, u_n$ is a basis of V and hence linearly independent, so Exercise 9D.3 applies and gives $d_p - c_p = 0$ for every $p \in \{1, \ldots, n\}$. In particular, for each p with $r < p \leq n$ we have $c_p = 0$ and therefore

$$d_p = \lambda_{1,p}\, y_1 + \lambda_{2,p}\, y_2 + \lambda_{3,p}\, y_3 = 0.$$

Because y_1, y_2, y_3 is linearly independent in W, this forces

$$\lambda_{1,p} = \lambda_{2,p} = \lambda_{3,p} = 0 \qquad \text{for all } p \text{ with } r < p \leq n.$$

Hence for each $l \in \{1, 2, 3\}$ we have $x_l = \sum_{p=1}^{r} \lambda_{l,p} u_p \in \operatorname{span}(u_1, \ldots, u_r) = U$. So x_1, x_2, x_3 is a linearly independent list of length 3 in U, while $u_1, \ldots, u_r$ is a spanning list of U of length $r \leq 2$. This contradicts 2.22.

Therefore no such v_1, v_2, w_1, w_2 exist, so $z \notin T$. Since $z \in V \otimes W$, we conclude

$$\{v_1 \otimes w_1 + v_2 \otimes w_2 : v_1, v_2 \in V \text{ and } w_1, w_2 \in W\} \neq V \otimes W,$$

as desired.
The argument above proves more generally: if $\dim V > k$ and $\dim W > k$ for a positive integer k, then the set of elements of $V \otimes W$ expressible as a sum of k elements of the form $v \otimes w$ is not all of $V \otimes W$. The case $k = 1$ is the last assertion of Exercise 9D.4.

Problem 9D.8. Suppose $v_1, \dots, v_m \in V$ and $w_1, \dots, w_m \in W$ are such that

$$v_1 \otimes w_1 + \cdots + v_m \otimes w_m = 0.$$

Suppose that U is a vector space and $\Gamma\colon V \times W \to U$ is a bilinear map. Show that

$$\Gamma(v_1, w_1) + \cdots + \Gamma(v_m, w_m) = 0.$$

Solution. The point of this exercise is that a relation among elementary tensors is automatically respected by every bilinear map; this is exactly the content of the universal property 9.79(a).
Because $\Gamma\colon V \times W \to U$ is a bilinear map, 9.79(a) provides a (unique) linear map

$$\hat{\Gamma}\colon V \otimes W \to U$$

such that $\hat{\Gamma}(v \otimes w) = \Gamma(v, w)$ for all $(v, w) \in V \times W$.
Now apply $\hat{\Gamma}$ to both sides of the hypothesis $v_1 \otimes w_1 + \cdots + v_m \otimes w_m = 0$. Since $\hat{\Gamma}$ is linear, it maps the additive identity of $V \otimes W$ to the additive identity of U (by 3.10), and it preserves sums. Hence

$$\begin{aligned} 0 &= \hat{\Gamma}(0) \\ &= \hat{\Gamma}(v_1 \otimes w_1 + \cdots + v_m \otimes w_m) \\ &= \hat{\Gamma}(v_1 \otimes w_1) + \cdots + \hat{\Gamma}(v_m \otimes w_m) \\ &= \Gamma(v_1, w_1) + \cdots + \Gamma(v_m, w_m), \end{aligned}$$

where the last line holds by the defining property of $\hat{\Gamma}$. This is the desired conclusion.
Remark. The hypothesis really is needed in the form given: a bilinear map Γ is not linear on $V \times W$ (for instance $\Gamma(2v, 2w) = 4\Gamma(v, w)$ in general), so one cannot argue directly on $V \times W$. It is the passage through $V \otimes W$, where bilinearity has been converted into linearity, that makes the argument work. Note also that no assumption of linear independence is placed on $v_1, \dots, v_m$ or on $w_1, \dots, w_m$.

Problem 9D.9. Suppose $S \in \mathcal{L}(V)$ and $T \in \mathcal{L}(W)$. Prove that there exists a unique operator on $V \otimes W$ that takes $v \otimes w$ to $Sv \otimes Tw$ for all $v \in V$ and $w \in W$.
[In an abuse of notation, the operator on $V \otimes W$ given by this exercise is often called $S \otimes T$.]

Solution. Define $\Gamma\colon V \times W \to V \otimes W$ by

$$\Gamma(v, w) = Sv \otimes Tw.$$

First we check that Γ is a bilinear map from $V \times W$ to the vector space $U = V \otimes W$, in the sense of 9.77. Fix $w \in W$. If $v_1, v_2 \in V$ and $\lambda \in \mathbf{F}$, then

$$\Gamma(v_1 + v_2, w) = S(v_1 + v_2) \otimes Tw = (Sv_1 + Sv_2) \otimes Tw = Sv_1 \otimes Tw + Sv_2 \otimes Tw,$$

where the second equality holds because S is linear and the third holds by the bilinearity of the tensor product (9.73); and

$$\Gamma(\lambda v_1, w) = S(\lambda v_1) \otimes Tw = (\lambda Sv_1) \otimes Tw = \lambda(Sv_1 \otimes Tw) = \lambda\Gamma(v_1, w),$$

again using the linearity of S and 9.73. Thus $v \mapsto \Gamma(v,w)$ is a linear map from V to $V \otimes W$ for each $w \in W$. The same computation with the roles of the two slots interchanged, using the linearity of T and 9.73, shows that $w \mapsto \Gamma(v,w)$ is a linear map from W to $V \otimes W$ for each $v \in V$. Hence Γ is a bilinear map.
Now 9.79(a), applied with $U = V \otimes W$, produces a linear map

$$\hat{\Gamma}\colon V \otimes W \to V \otimes W$$

such that $\hat{\Gamma}(v \otimes w) = \Gamma(v,w) = Sv \otimes Tw$ for all $(v,w) \in V \times W$. Because $\hat{\Gamma}$ maps $V \otimes W$ into itself, it is an operator on $V \otimes W$. This proves existence.
For uniqueness, suppose A and B are operators on $V \otimes W$ with

$$A(v \otimes w) = Sv \otimes Tw = B(v \otimes w)$$

for all $v \in V$ and $w \in W$. Let $e_1, \dots, e_m$ be a basis of V and $f_1, \dots, f_n$ a basis of W. By 9.74(b), the list $\{e_j \otimes f_k\}_{j=1,\dots,m;\, k=1,\dots,n}$ is a basis of $V \otimes W$. The linear maps A and B agree on each vector $e_j \otimes f_k$ of this basis (each such vector is of the form $v \otimes w$), so $A = B$: indeed, every $\alpha \in V \otimes W$ can be written as $\alpha = \sum_{k=1}^{n} \sum_{j=1}^{m} a_{j,k}\, e_j \otimes f_k$, and then linearity gives

$$A\alpha = \sum_{k=1}^{n} \sum_{j=1}^{m} a_{j,k}\, A(e_j \otimes f_k) = \sum_{k=1}^{n} \sum_{j=1}^{m} a_{j,k}\, B(e_j \otimes f_k) = B\alpha.$$

Hence the operator is unique, and we may unambiguously denote it by $S \otimes T$.
Remark. The uniqueness argument shows more generally that a linear map on $V \otimes W$ is determined by its values on the elementary tensors $v \otimes w$, since those span $V \otimes W$ by 9.74(b). We use this repeatedly in the following exercises. Note also that existence could not be established by simply declaring the value on each $v \otimes w$ and extending linearly, because an element of $V \otimes W$ has no unique representation as a sum of elementary tensors; 9.79(a) is precisely what removes that obstacle.

Problem 9D.10. Suppose $S \in \mathcal{L}(V)$ and $T \in \mathcal{L}(W)$. Prove that $S \otimes T$ is an invertible operator on $V \otimes W$ if and only if both S and T are invertible operators. Also, prove that if both S and T are invertible operators, then $(S \otimes T)^{-1} = S^{-1} \otimes T^{-1}$, where we are using the notation from the comment after Exercise 9.

Solution. Throughout, $S \otimes T$ denotes the operator on $V \otimes W$ produced by Exercise 9D.9, so that $(S \otimes T)(v \otimes w) = Sv \otimes Tw$ for all $v \in V$ and $w \in W$. As in Exercise 9D.9, we use freely the fact that a linear map on $V \otimes W$ is determined by its values on the elementary tensors $v \otimes w$, because the basis $\{e_j \otimes f_k\}$ of 9.74(b) consists of such vectors.
Recall the standing assumptions of Chapter 9: V and W are finite-dimensional nonzero vector spaces over $\mathbf{F}$. The hypothesis $V \neq \{0\} \neq W$ is genuinely needed for one of the two implications: if, say, $V = \{0\}$, then $\dim(V \otimes W) = (\dim V)(\dim W) = 0$ by 9.72, so $V \otimes W = \{0\}$ and every operator on it is invertible, while T could be any noninvertible operator on W.
Step 1: two multiplicative identities.
First, if $S, S' \in \mathcal{L}(V)$ and $T, T' \in \mathcal{L}(W)$, then

$$(S \otimes T)(S' \otimes T') = (SS') \otimes (TT').$$

Indeed, both sides are operators on $V \otimes W$ (products of linear maps are linear), and for all $v \in V$, $w \in W$,

$$(S \otimes T)\big((S' \otimes T')(v \otimes w)\big) = (S \otimes T)(S'v \otimes T'w) = SS'v \otimes TT'w = \big((SS') \otimes (TT')\big)(v \otimes w).$$

Since the two operators agree on all elementary tensors, they agree on the basis $\{e_j \otimes f_k\}$ and hence are equal.
Second,

$$I_V \otimes I_W = I_{V \otimes W},$$

because $(I_V \otimes I_W)(v \otimes w) = v \otimes w$ for all $v \in V$ and $w \in W$, so the operator $I_V \otimes I_W$ agrees with the identity on the basis $\{e_j \otimes f_k\}$.
Step 2: if S and T are invertible, then so is $S \otimes T$, with $(S \otimes T)^{-1} = S^{-1} \otimes T^{-1}$.
Applying Step 1 twice,

$$(S \otimes T)(S^{-1} \otimes T^{-1}) = (SS^{-1}) \otimes (TT^{-1}) = I_V \otimes I_W = I_{V \otimes W}$$

and

$$(S^{-1} \otimes T^{-1})(S \otimes T) = (S^{-1}S) \otimes (T^{-1}T) = I_V \otimes I_W = I_{V \otimes W}.$$

Thus $S \otimes T$ is invertible and its inverse is $S^{-1} \otimes T^{-1}$ (by the definition of invertible and inverse, 3.59, together with the uniqueness of the inverse, 3.60).
Step 3: if $S \otimes T$ is invertible, then S and T are invertible.
We prove the contrapositive. Suppose S is not invertible. Because V is finite-dimensional, an operator on V is invertible if and only if it is injective (by 3.65); hence S is not injective, so there exists $v \in V$ with $v \neq 0$ and $Sv = 0$. Because $W \neq \{0\}$, we may choose $w \in W$ with $w \neq 0$. Then

$$(S \otimes T)(v \otimes w) = Sv \otimes Tw = 0 \otimes Tw = 0,$$

where the last equality holds because $0 \otimes u = (0 \cdot 0) \otimes u = 0(0 \otimes u) = 0$ by 9.73. On the other hand $v \otimes w \neq 0$, since $v \neq 0$ and $w \neq 0$ (Exercise 9D.1; alternatively, extend v to a basis of V and w to a basis of W and apply 9.74). Hence $S \otimes T$ is not injective, so by 3.65 it is not invertible.
The argument when T is not invertible is identical with the roles of the two factors interchanged: choose $w \neq 0$ with $Tw = 0$ and any $v \neq 0$ in V, and note $(S \otimes T)(v \otimes w) = Sv \otimes 0 = 0$ while $v \otimes w \neq 0$.
Combining Steps 2 and 3: $S \otimes T$ is invertible if and only if both S and T are invertible, and in that case $(S \otimes T)^{-1} = S^{-1} \otimes T^{-1}$.

Problem 9D.11. Suppose V and W are inner product spaces. Prove that if $S \in \mathcal{L}(V)$ and $T \in \mathcal{L}(W)$, then $(S \otimes T)^* = S^* \otimes T^*$, where we are using the notation from the comment after Exercise 9.

Solution. By the standing assumptions of Chapter 9, V and W are finite-dimensional nonzero vector spaces; here they are also equipped with inner products. Then $\dim(V \otimes W) = (\dim V)(\dim W)$ is finite by 9.72, so $V \otimes W$ is a finite-dimensional inner product space once it carries the inner product of 9.82, and hence every operator on $V \otimes W$ has an adjoint. That inner product is characterized by

$$\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle \langle w, x \rangle$$

for all $v, u \in V$ and $w, x \in W$. As in Exercise 9D.9, $S \otimes T$ denotes the unique operator on $V \otimes W$ with $(S \otimes T)(v \otimes w) = Sv \otimes Tw$, and similarly $S^* \otimes T^*$ denotes the unique operator on $V \otimes W$ with $(S^* \otimes T^*)(u \otimes x) = S^*u \otimes T^*x$.
Step 1: the identity on elementary tensors.
Let $v \in V$, $w \in W$, $u \in V$, $x \in W$. Then

$$\begin{aligned}
\langle (S \otimes T)(v \otimes w), u \otimes x \rangle &= \langle Sv \otimes Tw, u \otimes x \rangle \\
&= \langle Sv, u \rangle \langle Tw, x \rangle \\
&= \langle v, S^*u \rangle \langle w, T^*x \rangle \\
&= \langle v \otimes w, S^*u \otimes T^*x \rangle \\
&= \langle v \otimes w, (S^* \otimes T^*)(u \otimes x) \rangle,
\end{aligned}$$

where the second and fourth equalities use 9.82 and the third uses the defining property of the adjoints S^* and T^* (7.1).
Step 2: passing from elementary tensors to all of $V \otimes W$.
Fix an orthonormal basis $e_1, \ldots, e_m$ of V and an orthonormal basis $f_1, \ldots, f_n$ of W; by 9.83 the doubly indexed list $\{e_j \otimes f_k\}_{j=1,\ldots,m;\, k=1,\ldots,n}$ is a basis of $V \otimes W$ (only its being a basis matters here, which is 9.74(b)). In particular every element of $V \otimes W$ is a linear combination of elementary tensors.

Fix $\beta \in V \otimes W$ of the form $\beta = u \otimes x$ and consider the two functions

$$\alpha \mapsto \langle (S \otimes T)\alpha, \beta \rangle \quad \text{and} \quad \alpha \mapsto \langle \alpha, (S^* \otimes T^*)\beta \rangle$$

from $V \otimes W$ to $\mathbf{F}$. Both are linear in α: the first because $S \otimes T$ is linear and the inner product is linear in its first slot, the second because the inner product is linear in its first slot. By Step 1 they agree at every elementary tensor $\alpha = v \otimes w$, hence at each basis vector $e_j \otimes f_k$, hence (being linear) on all of $V \otimes W$. So

$$\langle (S \otimes T)\alpha, u \otimes x \rangle = \langle \alpha, (S^* \otimes T^*)(u \otimes x) \rangle$$

for all $\alpha \in V \otimes W$ and all $u \in V$, $x \in W$.
Now fix $\alpha \in V \otimes W$ and consider the two functions

$$\beta \mapsto \langle (S \otimes T)\alpha, \beta \rangle \quad \text{and} \quad \beta \mapsto \langle \alpha, (S^* \otimes T^*)\beta \rangle.$$

Both are additive in β and satisfy $g(\lambda\beta) = \overline{\lambda}\, g(\beta)$ for $\lambda \in \mathbf{F}$: for the first this is additivity and conjugate homogeneity of the inner product in its second slot (6.6(d) and 6.6(e)), and for the second it is those same properties combined with the linearity of $S^* \otimes T^*$. Consequently each of them is determined by its values on any spanning list, since it takes the linear combination $\beta = \sum c_{j,k}\, e_j \otimes f_k$ to $\sum \overline{c_{j,k}}\, g(e_j \otimes f_k)$. By the previous paragraph the two functions agree at every $\beta = e_j \otimes f_k$, so they agree for all $\beta \in V \otimes W$.
Step 3: conclusion.
We have shown

$$\langle (S \otimes T)\alpha, \beta \rangle = \langle \alpha, (S^* \otimes T^*)\beta \rangle$$

for all $\alpha, \beta \in V \otimes W$. On the other hand, by the definition of the adjoint (7.1),

$$\langle (S \otimes T)\alpha, \beta \rangle = \langle \alpha, (S \otimes T)^*\beta \rangle$$

for all $\alpha, \beta \in V \otimes W$. Subtracting and using additivity and conjugate homogeneity of the inner product in its second slot (6.6), we get $\langle \alpha, (S \otimes T)^*\beta - (S^* \otimes T^*)\beta \rangle = 0$ for all α, β. Taking $\alpha = (S \otimes T)^*\beta - (S^* \otimes T^*)\beta$ shows that this vector has norm 0, hence is 0, for every β. Therefore

$$(S \otimes T)^* = S^* \otimes T^*,$$

as desired.

Problem 9D.12. Suppose that $V_1, \dots, V_m$ are finite-dimensional inner product spaces. Prove that there is a unique inner product on $V_1 \otimes \cdots \otimes V_m$ such that

$$\langle v_1 \otimes \cdots \otimes v_m,\ u_1 \otimes \cdots \otimes u_m \rangle = \langle v_1, u_1 \rangle \cdots \langle v_m, u_m \rangle$$

for all $(v_1, \dots, v_m)$ and $(u_1, \dots, u_m)$ in $V_1 \times \cdots \times V_m$.
[Note that the equation above implies that

$$\|v_1 \otimes \cdots \otimes v_m\| = \|v_1\| \times \cdots \times \|v_m\|$$

for all $(v_1, \dots, v_m) \in V_1 \times \cdots \times V_m$.]

Solution. This is the m-fold analogue of 9.80, and the proof follows that pattern.
Notation. For each $k \in \{1, \dots, m\}$ let $n_k = \dim V_k$ and let $e_1^k, \dots, e_{n_k}^k$ be an orthonormal basis of V_k (such a basis exists by 6.35, since V_k is a finite-dimensional inner product space). Write $J = (j_1, \dots, j_m)$ for a multi-index with $j_k \in \{1, \dots, n_k\}$, let $\mathcal{J}$ denote the set of all such multi-indices, and set

$$E_J = e_{j_1}^1 \otimes \cdots \otimes e_{j_m}^m.$$

By 9.90 the list $\{E_J\}_{J \in \mathcal{J}}$ is a basis of $V_1 \otimes \cdots \otimes V_m$.
Step 0: multilinearity of the tensor product of vectors.

The map $(v_1, \dots, v_m) \mapsto v_1 \otimes \cdots \otimes v_m$ is an m-linear map from $V_1 \times \cdots \times V_m$ to $V_1 \otimes \cdots \otimes V_m$. To see this, fix k and fix the entries in the slots other than the k^{th}. For $\varphi_1 \in V_1', \dots, \varphi_m \in V_m'$, definition 9.88 gives

$$(v_1 \otimes \cdots \otimes v_m)(\varphi_1, \dots, \varphi_m) = \varphi_1(v_1) \cdots \varphi_m(v_m),$$

and the right side is, as a function of v_k with the other vectors fixed, a scalar multiple of the linear functional φ_k evaluated at v_k. Hence for $v_k, v_k' \in V_k$ and $\lambda \in \mathbf{F}$ the elements $v_1 \otimes \cdots \otimes (v_k + v_k') \otimes \cdots \otimes v_m$ and $v_1 \otimes \cdots \otimes v_k \otimes \cdots \otimes v_m + v_1 \otimes \cdots \otimes v_k' \otimes \cdots \otimes v_m$ take the same value at every $(\varphi_1, \dots, \varphi_m)$, so they are equal; similarly $v_1 \otimes \cdots \otimes (\lambda v_k) \otimes \cdots \otimes v_m = \lambda(v_1 \otimes \cdots \otimes v_m)$. This is the generalization of 9.73 to m factors.

Consequently, if $v_k = \sum_{j=1}^{n_k} a_j^k e_j^k$ for each k, then expanding each slot in turn gives

$$v_1 \otimes \cdots \otimes v_m = \sum_{J \in \mathcal{J}} a_{j_1}^1 a_{j_2}^2 \cdots a_{j_m}^m \, E_J.$$

Step 1: existence.

Define, for scalars $\{b_J\}_{J \in \mathcal{J}}$ and $\{c_J\}_{J \in \mathcal{J}}$,

$$\Big\langle \sum_{J \in \mathcal{J}} b_J E_J, \ \sum_{J \in \mathcal{J}} c_J E_J \Big\rangle = \sum_{J \in \mathcal{J}} b_J \overline{c_J}.$$

This is well defined because $\{E_J\}$ is a basis of $V_1 \otimes \cdots \otimes V_m$ (9.90), so the coefficients b_J and c_J are uniquely determined by the vectors. It is an inner product: positivity holds because $\langle \sum b_J E_J, \sum b_J E_J \rangle = \sum_J |b_J|^2 \geq 0$; definiteness holds because this sum equals 0 exactly when every $b_J = 0$, that is, exactly when the vector is 0; linearity in the first slot and conjugate symmetry are immediate from the formula. (This is just the transport of the standard inner product on $\mathbf{F}^{n_1 \cdots n_m}$ through the coordinate isomorphism determined by the basis $\{E_J\}$; the verification is the same as the one left to the reader in the proof of 9.80.)

Now we check the required identity. Suppose $(v_1, \dots, v_m)$ and $(u_1, \dots, u_m)$ lie in $V_1 \times \cdots \times V_m$, and write $v_k = \sum_{j=1}^{n_k} a_j^k e_j^k$ and $u_k = \sum_{j=1}^{n_k} d_j^k e_j^k$. By the expansion displayed at the end of Step 0 and the definition above,

$$\begin{aligned}
\langle v_1 \otimes \cdots \otimes v_m, \ u_1 \otimes \cdots \otimes u_m \rangle &= \sum_{J \in \mathcal{J}} \big(a_{j_1}^1 \cdots a_{j_m}^m\big)\overline{\big(d_{j_1}^1 \cdots d_{j_m}^m\big)} \\
&= \sum_{j_1=1}^{n_1} \cdots \sum_{j_m=1}^{n_m} \prod_{k=1}^{m} a_{j_k}^k \, \overline{d_{j_k}^k} \\
&= \prod_{k=1}^{m} \Big(\sum_{j=1}^{n_k} a_j^k \overline{d_j^k} \Big) \\
&= \langle v_1, u_1 \rangle \cdots \langle v_m, u_m \rangle,
\end{aligned}$$

where the third equality is the distributive law (expanding the product of the m sums produces exactly one term for each multi-index J), and the last equality holds because $e_1^k, \dots, e_{n_k}^k$ is an orthonormal basis of V_k, so that $\langle v_k, u_k \rangle = \sum_{j=1}^{n_k} a_j^k \overline{d_j^k}$; the last identity follows by expanding both vectors in the orthonormal basis and using additivity and homogeneity in the first slot together with 6.6(d) and 6.6(e) in the second slot, so that only the diagonal terms $\langle e_j^k, e_j^k \rangle = 1$ survive. This proves existence.

Step 2: uniqueness.

Suppose $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ are two inner products on $V_1 \otimes \cdots \otimes V_m$ satisfying the stated identity. Then they agree on every pair of elementary tensors, in particular

$$\langle E_J, E_K \rangle_1 = \langle E_J, E_K \rangle_2$$

for all $J, K \in \mathcal{J}$. Given arbitrary $\alpha, \beta \in V_1 \otimes \cdots \otimes V_m$, write $\alpha = \sum_J b_J E_J$ and $\beta = \sum_K c_K E_K$ using the basis $\{E_J\}$. Since every inner product is linear in its first slot and is additive and conjugate homogeneous in its second slot (6.6), both inner products satisfy

$$\langle \alpha, \beta \rangle_i = \sum_J \sum_K b_J \overline{c_K} \, \langle E_J, E_K \rangle_i$$

for $i = 1, 2$, and the two right-hand sides are equal. Hence $\langle \alpha, \beta \rangle_1 = \langle \alpha, \beta \rangle_2$ for all α, β, so the two inner products coincide. This proves uniqueness.
Step 3: the norm identity.
Taking $u_k = v_k$ for each k in the defining identity gives

$$\|v_1 \otimes \cdots \otimes v_m\|^2 = \langle v_1, v_1 \rangle \cdots \langle v_m, v_m \rangle = \|v_1\|^2 \cdots \|v_m\|^2.$$

Taking (nonnegative) square roots yields $\|v_1 \otimes \cdots \otimes v_m\| = \|v_1\| \times \cdots \times \|v_m\|$, as noted in the statement.

Problem 9D.13. Suppose that $V_1, \ldots, V_m$ are finite-dimensional inner product spaces and $V_1 \otimes \cdots \otimes V_m$ is made into an inner product space using the inner product from Exercise 12. Suppose $e_1^k, \ldots, e_{n_k}^k$ is an orthonormal basis of V_k for each $k = 1, \ldots, m$. Show that the list

$$\{e_{j_1}^1 \otimes \cdots \otimes e_{j_m}^m\}_{j_1=1,\ldots,n_1;\ \cdots;\ j_m=1,\ldots,n_m}$$

is an orthonormal basis of $V_1 \otimes \cdots \otimes V_m$.

Solution. This is the m-fold analogue of 9.83. The essential point is that the orthonormal bases $e_1^k, \ldots, e_{n_k}^k$ used here may be different from the orthonormal bases used to construct the inner product in Exercise 9D.12; the characterizing identity of that inner product is basis free, so the conclusion holds for every choice of orthonormal bases.
Write $J = (j_1, \ldots, j_m)$ and $K = (l_1, \ldots, l_m)$ for multi-indices with $j_k, l_k \in \{1, \ldots, n_k\}$, and put

$$E_J = e_{j_1}^1 \otimes \cdots \otimes e_{j_m}^m.$$

Note that $n_k = \dim V_k$, since $e_1^k, \ldots, e_{n_k}^k$ is a basis of V_k.
Step 1: orthonormality.
Let J and K be multi-indices. By the defining property of the inner product from Exercise 9D.12,

$$\langle E_J, E_K \rangle = \langle e_{j_1}^1, e_{l_1}^1 \rangle \, \langle e_{j_2}^2, e_{l_2}^2 \rangle \cdots \langle e_{j_m}^m, e_{l_m}^m \rangle.$$

Because $e_1^k, \ldots, e_{n_k}^k$ is an orthonormal list in V_k, the k^{th} factor equals 1 if $j_k = l_k$ and equals 0 if $j_k \neq l_k$. Hence the product equals 1 if $j_k = l_k$ for every k, that is if $J = K$, and equals 0 otherwise, because in that case at least one factor is 0. Thus

$$\langle E_J, E_K \rangle = \begin{cases} 1 & \text{if } J = K, \\ 0 & \text{if } J \neq K, \end{cases}$$

which says precisely that the doubly (indeed m-tuply) indexed list $\{E_J\}$ is orthonormal. In particular $\|E_J\| = 1$ for each J.
Step 2: the list is a basis.
The list $\{E_J\}$ is a basis of $V_1 \otimes \cdots \otimes V_m$ by 9.90, since $e_1^k, \ldots, e_{n_k}^k$ is a basis of V_k for each k.
Alternatively, one can obtain this from Step 1 without invoking 9.90: an orthonormal list is linearly independent (by 6.25, applied to the list $\{E_J\}$ in some order), and the length of the list is $n_1 n_2 \cdots n_m$, which equals

$$(\dim V_1)(\dim V_2) \cdots (\dim V_m) = \dim(V_1 \otimes \cdots \otimes V_m)$$

by 9.89. A linearly independent list whose length equals the dimension of the space is a basis (by 2.38). Hence $\{E_J\}$ is a basis.
Combining Steps 1 and 2, the list

$$\{e_{j_1}^1 \otimes \cdots \otimes e_{j_m}^m\}_{j_1=1,\ldots,n_1;\ \cdots;\ j_m=1,\ldots,n_m}$$

is an orthonormal basis of $V_1 \otimes \cdots \otimes V_m$, as desired. (As usual for these multiply indexed lists, the ordering is immaterial; choose any convenient one.)