

Linear Algebra

Aayush Bajaj

July 10, 2026

Contents

1	vector spaces	1
2	independence, basis, dimension	2
3	linear maps and matrices	2
4	the four fundamental subspaces	2
5	inner products and orthogonality	3
6	determinants	3
7	eigentheory	4
7.1	eigenvalues and diagonalisation	4
7.2	the spectral theorem	4
7.3	power iteration	4
8	the singular value decomposition	5
8.1	existence and geometry	5
8.2	low-rank approximation	5
9	positive definiteness and cholesky	6
10	numerics: eigendecomposition vs svd	6
11	see also	7

linear algebra is the study of vector spaces and the structure-preserving maps between them. it is the one branch of mathematics that computers execute natively — every model fit, every graphics frame, every pagerank iteration is matrix arithmetic — and the local model that calculus reduces every smooth problem to. the plot of this page: spaces, then maps, then the four subspaces every matrix carries, then the two great factorisations — the spectral theorem and the svd (Anton, Howard, 2010).

the jacobian on the multivariable page is exactly this reduction: differentiate, then hand the problem to linear algebra.

1 vector spaces

a **vector space** over $\mathbb{R}$ is a set V with addition and scalar multiplication satisfying the usual eight axioms, compactly: $(V, +)$ is a commutative group (associative, commutative, zero vector, negatives), and scalar multiplication is associative ($a(b\mathbf{v}) = (ab)\mathbf{v}$), unital ($1\mathbf{v} = \mathbf{v}$), and distributes over both additions.

a subset $W \subseteq V$ is a **subspace** iff it is nonempty and closed under linear combinations: $\mathbf{u}, \mathbf{v} \in W \implies a\mathbf{u} + b\mathbf{v} \in W$. the examples that matter:

- $\mathbb{R}^n$, the prototype; every n -dimensional real vector space is isomorphic to it.
- polynomials of degree $\leq n$; all polynomials (infinite-dimensional).
- continuous functions on $[a, b]$ — the setting of **functional analysis**, where dimension becomes infinite and topology starts to matter.
- solution sets of homogeneous linear systems $A\mathbf{x} = \mathbf{0}$ — never of inhomogeneous ones (no zero vector).

2 independence, basis, dimension

- vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are **linearly independent** if $c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k = \mathbf{0}$ forces every $c_i = 0$ — no vector is redundant.
- a **basis** is an independent spanning set; every vector then has *unique* coordinates.
- any two bases of a space have the same size (steinitz exchange), and that common size is the **dimension**.

the working test in $\mathbb{R}^n$: stack the vectors as columns and row-reduce; independence means a pivot in every column, spanning means a pivot in every row.

3 linear maps and matrices

a map $T : V \rightarrow W$ is **linear** if $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$.

representation theorem. fix bases $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ of V and $(\mathbf{w}_1, \dots, \mathbf{w}_m)$ of W . then every linear T is realised by exactly one $m \times n$ matrix A — its j -th column holds the coordinates of $T(\mathbf{v}_j)$ — and conversely every matrix defines a linear map. composition of maps corresponds to matrix multiplication: $[S \circ T] = [S][T]$.

this is why matrix multiplication is defined the odd way it is: it was reverse-engineered from composition, not invented and then interpreted.

the dictionary runs deep: a linear map is the *object*, a matrix is its *coordinates*. change of basis replaces A by $P^{-1}AP$ — same map, new coordinates — and the properties that survive this similarity (rank, determinant, eigenvalues, trace) are exactly the properties of the map itself.

4 the four fundamental subspaces

an $m \times n$ matrix A of rank r carries four subspaces:

subspace	lives in	dimension
column space $C(A)$	$\mathbb{R}^m$	r
null space $N(A)$	$\mathbb{R}^n$	$n - r$
row space $C(A^\top)$	$\mathbb{R}^n$	r
left null space $N(A^\top)$	$\mathbb{R}^m$	$m - r$

rank–nullity theorem. $\dim C(A) + \dim N(A) = n$.

proof sketch: take a basis $\mathbf{u}_1, \dots, \mathbf{u}_k$ of $N(A)$ and extend it to a basis $\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{w}_1, \dots, \mathbf{w}_{n-k}$ of $\mathbb{R}^n$. the images $A\mathbf{w}_1, \dots, A\mathbf{w}_{n-k}$ span $C(A)$ (the $\mathbf{u}$'s die), and they are independent: if $\sum c_i A\mathbf{w}_i = \mathbf{0}$ then $\sum c_i \mathbf{w}_i \in N(A)$, so it is a combination of $\mathbf{u}$'s — impossible for basis vectors unless all $c_i = 0$. hence $\dim C(A) = n - k$. ■

the deeper geometry is the orthogonality: the row space and null space are orthogonal complements in $\mathbb{R}^n$ (each null vector is orthogonal to each row by definition of $A\mathbf{x} = \mathbf{0}$), and likewise $C(A) \perp N(A^\top)$ in $\mathbb{R}^m$. every $\mathbf{x} \in \mathbb{R}^n$ splits uniquely as $\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$, and A acts invertibly from the row space onto the column space while annihilating the null component:

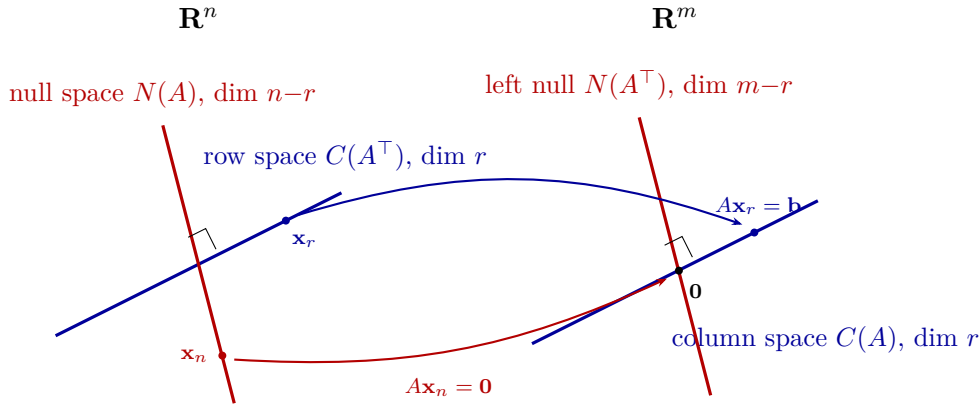


Figure 1: the four fundamental subspaces: A maps the row space bijectively onto the column space and crushes the null space to $\mathbf{0}$. orthogonal pairs meet at right angles.

5 inner products and orthogonality

an **inner product** on a real space is a symmetric, bilinear, positive-definite form $\langle \cdot, \cdot \rangle$; it buys lengths $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, angles via $\cos \theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|}$ (legal by cauchy-schwarz), and the notion of orthogonality $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. orthonormal bases make coordinates trivial: $\mathbf{v} = \sum_i \langle \mathbf{v}, \mathbf{q}_i \rangle \mathbf{q}_i$, no system to solve.

gram-schmidt manufactures an orthonormal basis from any independent list $\mathbf{a}_1, \dots, \mathbf{a}_n$:

- **input:** independent vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$.
- **loop:** for $j = 1, \dots, n$: subtract the components along everything already built,

$$\tilde{\mathbf{q}}_j = \mathbf{a}_j - \sum_{i < j} \langle \mathbf{a}_j, \mathbf{q}_i \rangle \mathbf{q}_i, \quad (1)$$

- **normalise:** $\mathbf{q}_j = \tilde{\mathbf{q}}_j / \|\tilde{\mathbf{q}}_j\|$.
- **output:** orthonormal $\mathbf{q}_1, \dots, \mathbf{q}_n$ spanning the same space, stage by stage.

collecting the bookkeeping as a matrix identity gives the **qr factorisation** $A = QR$ with Q orthonormal columns and R upper triangular — the workhorse of least squares, where $A\mathbf{x} \approx \mathbf{b}$ collapses to the triangular solve $R\mathbf{x} = Q^\top \mathbf{b}$.

in floating point one uses modified gram-schmidt or householder reflections — classical gram-schmidt loses orthogonality to rounding.

6 determinants

take the properties as the definition (Hill, David, 2001): $\det$ is the unique function of the rows of an $n \times n$ matrix that is (i) linear in each row separately, (ii) alternating (equal rows kill it, swapping rows flips its sign), and (iii) normalised by $\det I = 1$. everything else is a corollary:

- expansion formulas (cofactor, leibniz sum over permutations) — *consequences*, not definitions.

- $\det(AB) = \det A \det B$; $\det A^\top = \det A$; triangular matrices multiply their diagonals.
- $\det A \neq 0$ iff A is invertible.
- geometrically, $|\det A|$ is the volume-scaling factor of the map — which is exactly why it appears in the [change-of-variables formula](#).

invertible matrix theorem (the grand equivalence, for square A): invertible $\iff \det A \neq 0 \iff \text{rank } n \iff \text{columns independent} \iff \text{columns span} \iff N(A) = \{\mathbf{0}\} \iff A\mathbf{x} = \mathbf{b}$ uniquely solvable for every $\mathbf{b} \iff 0$ is not an eigenvalue. one condition fails, all fail.

7 eigentheory

7.1 eigenvalues and diagonalisation

$A\mathbf{v} = \lambda\mathbf{v}$ with $\mathbf{v} \neq \mathbf{0}$: the directions the map merely stretches. eigenvalues are roots of the **characteristic polynomial** $p(\lambda) = \det(A - \lambda I)$, degree n , so there are n of them over $\mathbb{C}$ with multiplicity.

A is **diagonalisable** iff it has n linearly independent eigenvectors: $A = PDP^{-1}$ with $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. eigenvectors of distinct eigenvalues are automatically independent, so distinct roots suffice; repeated roots may or may not (the jordan story). diagonalisation is the licence to compute functions of matrices: $A^k = PD^kP^{-1}$, $e^A = Pe^DP^{-1}$ — which is how linear [differential equations](#) get solved.

cayley–hamilton. every square matrix satisfies its own characteristic polynomial: $p(A) = 0$. (proved honestly via the adjugate identity — substituting $\lambda = A$ into $\det(A - \lambda I)$ is not a proof.)

7.2 the spectral theorem

theorem. every real symmetric matrix S has an orthonormal basis of eigenvectors with real eigenvalues:

$$S = Q\Lambda Q^\top, \quad Q^\top Q = I, \quad \Lambda \text{ real diagonal.} \quad (2)$$

symmetry buys three miracles at once: real eigenvalues, orthogonal eigenvectors (across distinct eigenvalues), and no defectiveness — symmetric matrices are always diagonalisable. this is the theorem behind pca, behind the hessian test in [optimisation](#), behind every quadratic form being a sum of squares in the right coordinates (Deisenroth, Marc Peter and Faisal, A. Aldo and Ong, Cheng Soon, 2020).

7.3 power iteration

the simplest eigenvalue algorithm, and the ancestor of pagerank:

- **input:** A with a strictly dominant eigenvalue $|\lambda_1| > |\lambda_2|$, random $\mathbf{x}_0$ not orthogonal to the top eigenvector.
- **loop:** $\mathbf{x}_{k+1} = A\mathbf{x}_k / \|A\mathbf{x}_k\|$.
- **read off:** $\mathbf{x}_k$ converges (up to sign) to the dominant eigenvector at geometric rate $|\lambda_2/\lambda_1|^k$; the rayleigh quotient $\mathbf{x}_k^\top A\mathbf{x}_k$ converges to λ_1 .

writing $\mathbf{x}_0$ in the eigenbasis shows why: each iteration multiplies the i -th coefficient by λ_i , and the dominant one wins the race.

perron–frobenius (why pagerank is well-posed): a matrix with strictly positive entries has a real, simple, dominant eigenvalue with a strictly positive eigenvector — a *column*-stochastic such matrix has it at $\lambda = 1$, and power iteration on it converges to the unique stationary distribution (for a row-stochastic matrix the stationary distribution is the *left* eigenvector — pagerank’s transition matrix is written column-stochastic for exactly this reason).

8 the singular value decomposition

8.1 existence and geometry

theorem. every real $m \times n$ matrix factors as

$$A = U\Sigma V^\top, \quad (3)$$

with U ($m \times m$) and V ($n \times n$) orthogonal and Σ diagonal with entries $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$ — the **singular values**. existence rides on the spectral theorem: $A^\top A$ is symmetric positive semi-definite, its orthonormal eigenvectors are V , its eigenvalues are σ_i^2 , and $\mathbf{u}_i = A\mathbf{v}_i/\sigma_i$.

geometrically: *every* matrix is a rotation, then an axis-aligned scaling, then another rotation. the unit sphere always maps to an ellipsoid (possibly squashed flat); the σ_i are its semi-axes.

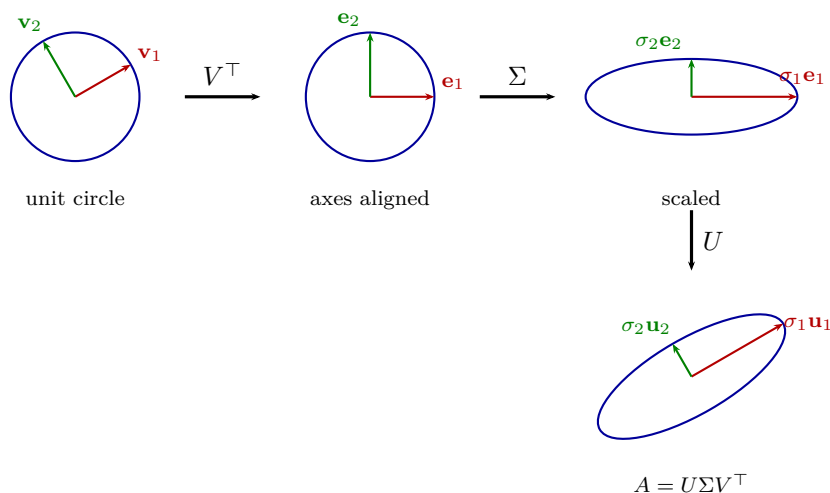


Figure 2: the svd as geometry: $V^\top$ rotates the unit circle’s preferred axes onto the coordinate axes, Σ scales along them, U rotates the result into place. every matrix does exactly this.

8.2 low-rank approximation

truncating the svd after k terms gives $A_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$, and the **eckart–young theorem** says this is the *best* rank- k approximation:

$$\min_{\text{rank}(B) \leq k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}, \quad (4)$$

with the analogous statement in frobenius norm (error $\sqrt{\sigma_{k+1}^2 + \dots}$). this single theorem is the licence behind pca, latent semantic analysis, matrix-completion recommenders and lossy image compression: keep the big singular values, discard the rest, and no rank- k object can do better. the svd also hands over the pseudoinverse $A^+ = V\Sigma^+U^\top$ (least-squares in one formula) and the condition number $\kappa = \sigma_{\max}/\sigma_{\min}$ — how much a linear solve amplifies input error.

9 positive definiteness and cholesky

a symmetric S is **positive definite** (pd) when any of these equivalent conditions holds:

- $\mathbf{x}^\top S \mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$;
- all eigenvalues > 0 ;
- all leading principal minors > 0 (sylvester);
- $S = R^\top R$ for some invertible R .

pd matrices are the geometry-carriers of applied mathematics: covariance matrices, kernel gram matrices, hessians at strict minima, the metric in every quadratic norm $\|\mathbf{x}\|_S^2 = \mathbf{x}^\top S \mathbf{x}$.

the constructive certificate is the **cholesky factorisation** $S = LL^\top$ with L lower triangular and positive diagonal — it exists iff S is pd, costs $\frac{n^3}{3}$ flops (half of lu, symmetry exploited), and either completes (proof of definiteness) or hits a negative pivot (counterexample found). solving $S\mathbf{x} = \mathbf{b}$, sampling correlated gaussians, and kalman filters all run on it.

10 numerics: eigendecomposition vs svd

```

1  import numpy as np
2
3  rng = np.random.default_rng(42)
4
5  # --- symmetric psd matrix: eigendecomposition and svd agree ---
6  B = rng.normal(size=(4, 4))
7  S = B @ B.T                                # symmetric positive semi-definite
8  evals, evecs = np.linalg.eigh(S)           # ascending
9  U, sig, Vt = np.linalg.svd(S)              # descending
10
11 print("eigenvalues (desc):", np.round(evals[::-1], 6))
12 print("singular values  :", np.round(sig, 6))
13 print("max |eig - sv|    :", f"{np.abs(evals[::-1] - sig).max():.2e}")
14
15 # --- power iteration finds the top eigenpair ---
16 x = rng.normal(size=4)
17 for _ in range(200):
18     x = S @ x
19     x /= np.linalg.norm(x)
20 rayleigh = x @ S @ x
21 print("power iteration  :", f"lambda_max = {rayleigh:.6f}",
22       f"(eigh: {evals[-1]:.6f})")
23
24 # --- eckart-young: best rank-1 error (spectral norm) is sigma_2 ---
25 A = rng.normal(size=(5, 3))
26 U, sig, Vt = np.linalg.svd(A, full_matrices=False)
27 A1 = sig[0] * np.outer(U[:, 0], Vt[0])      # truncated svd, rank 1

```

```

28 err = np.linalg.norm(A - A1, ord=2)
29 print("rank-1 error      :", f"{err:.6f} (sigma_2 = {sig[1]:.6f})")
30
31 # --- nonsymmetric: eigenvalues may be complex, singular values never ---
32 C = np.array([[0., -1.], [1., 0.]]) # rotation by 90 degrees
33 print("rotation eigvals :", np.linalg.eigvals(C))
34 print("rotation sigmas  :", np.linalg.svd(C, compute_uv=False))

```

```

eigenvalues (desc): [7.307998 4.208078 1.052883 0.010111]
singular values   : [7.307998 4.208078 1.052883 0.010111]
max |eig - sv|    : 3.55e-15
power iteration    : lambda_max = 7.307998 (eigh: 7.307998)
rank-1 error       : 1.256959 (sigma_2 = 1.256959)
rotation eigvals   : [0.+1.j 0.-1.j]
rotation sigmas    : [1. 1.]

```

three lessons in ten lines: for symmetric psd matrices the two decompositions coincide (to machine precision); power iteration recovers the dominant pair; the truncated svd's error is *exactly* σ_{k+1} , as eckart–young promises. and the rotation matrix shows why the svd is the safer tool: its eigenvalues flee to $\mathbb{C}$ while its singular values sit calmly at 1.

11 see also

- [multivariable calculus](#) — jacobians and hessians consume this page
- [optimisation](#) — definiteness decides convexity
- [statistics](#) — covariance, pca, least squares
- [functional analysis](#) — vector spaces grown infinite-dimensional
- [discrete mathematics](#) — graphs as adjacency matrices

references

Anton, Howard (2010). *Elementary Linear Algebra, with Applications*, Wiley.

Deisenroth, Marc Peter and Faisal, A. Aldo and Ong, Cheng Soon (2020). *Mathematics for Machine Learning*, Cambridge University Press.

Hill, David (2001). *Elementary Linear Algebra*, Cengage Learning.