MathCastle

Linear Algebra

Complete course booklet • 11 topics

Contents

1 Matrices & Systems **3**

1.1 Matrix Multiplication 3

1.2 Proofs & Why It Matters 4

1.3 Going Deeper: Explanations & Worked Problems 4

2 Row Reduction & Rank **5**

2.1 Gaussian Elimination 5

2.2 Proofs & Why It Matters 6

2.3 Going Deeper: Explanations & Worked Problems 7

3 Determinants & Inverses **8**

3.1 The Determinant 8

3.2 Proofs & Why It Matters 9

3.3 Going Deeper: Explanations & Worked Problems 10

4 Vector Spaces & Bases **11**

4.1 The Language of Vector Spaces 11

4.2 Proofs & Why It Matters 12

4.3 Going Deeper: Explanations & Worked Problems 12

5 Linear Transformations **13**

5.1 Maps That Respect Addition 13

5.2 Proofs & Why It Matters 14

5.3 Going Deeper: Explanations & Worked Problems 15

6 Orthogonality & Least Squares **16**

6.1 Projection and Best Approximation 16

6.2 Proofs & Why It Matters 17

6.3 Going Deeper: Explanations & Worked Problems 18

7 Eigenvalues **19**

7.1 Eigenvalues & Eigenvectors 19

7.2 Proofs & Why It Matters 20

7.3 Going Deeper: Explanations & Worked Problems 20

8 Diagonalization & Matrix Powers **22**

8.1 Changing to the Eigenbasis 22

8.2 Proofs & Why It Matters 23

8.3 Going Deeper: Worked Problems 23

9 Symmetric Matrices & Quadratic Forms **24**

9.1 The Nicest Matrices in Mathematics 24

9.2 Proofs & Why It Matters 25

9.3 Going Deeper: Worked Problems 26

10 Singular Values & the SVD **26**

10.1 The Master Decomposition 26

10.2 Proofs & Why It Matters 27

10.3 Going Deeper: Worked Problems 28

11 Markov Chains & Long-Run Behavior **29**

11.1 Probability Meets Eigenvectors 29

11.2 Proofs & Why It Matters 30

11.3 Going Deeper: Worked Problems 30

1 Matrices & Systems

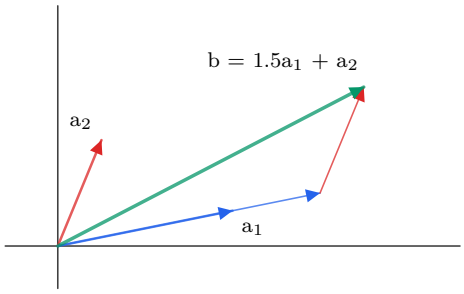
Arithmetic with grids of numbers

1.1 Matrix Multiplication

Row times column

The (i, j) entry of AB is row i of A dotted with column j of B. Sizes must chain: $(m \times n)(n \times p) = m \times p$. Order MATTERS: $AB \neq BA$ in general — the single biggest habit change from ordinary algebra.

Ax is a combination of the columns of A.



A matrix is a machine for systems

The system $2x + y = 7$, $x - y = 2$ is one equation $A\mathbf{x} = \mathbf{b}$ about the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}$. Solving the system, inverting A, and asking whether columns are independent are all the SAME question in different clothes.

Linear dependence

$(1, 2, 3)$ and $(2, 4, t)$ are dependent exactly when one is a multiple of the other; the first two coordinates force the multiplier to be 2, so $t = 6$ — dependence is a rigid, all-coordinates condition.

Significance: matrices are how computers see everything

An image is a matrix of pixels; a social network is an adjacency matrix; a dataset is a rows-by-features matrix; a 3D scene is transformed by 4×4 matrices sixty times a second. Matrix multiplication being composition is why graphics pipelines can collapse a chain of transformations into ONE matrix and apply it to millions of points.

Try it: predict before you multiply

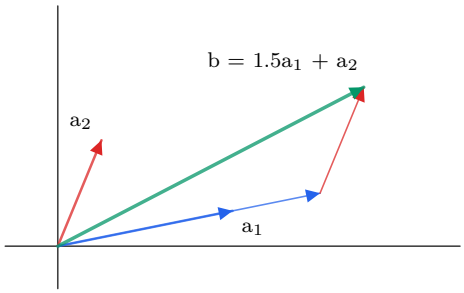
Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ (rotation by 90°). Predict A^2, then compute. Work: two quarter-turns are a half-turn, so A^2 should be $-I$ — and indeed $A^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ ✓. Thinking in maps first turns matrix arithmetic into geometry.

1.2 Proofs & Why It Matters

Proof: matrix multiplication is composition

Let T_A and T_B be the maps $\mathbf{x} \mapsto A\mathbf{x}$ and $\mathbf{x} \mapsto B\mathbf{x}$. Column j of AB is $A(B\mathbf{e}_j)$ — feed the jth basis vector through B, then through A.

Ax is a combination of the columns of A.



But a linear map is determined by where it sends basis vectors, so the matrix whose columns are $A(B\mathbf{e}_j)$ IS the matrix of $T_A \circ T_B$. The row-times-column rule is just this computed entry by entry — and composition of maps is famously order-sensitive, which is WHY $AB \neq BA$. ■

Proof: row operations preserve solutions

Each operation is reversible: swapping rows swaps back; scaling by $c \neq 0$ undoes by $\frac{1}{c}$; adding c(row i) to row j undoes by subtracting. A solution of the old system satisfies every new equation (each new row is a combination of old rows), and reversibility gives the converse. So the solution set never changes on the way to echelon form. ■

1.3 Going Deeper: Explanations & Worked Problems

What a matrix does: track the basis vectors

Everything about $A\mathbf{x}$ is visible in two special inputs. Feed in $\mathbf{e}_1 = (1, 0)$: out comes the FIRST COLUMN of A. Feed in $\mathbf{e}_2 = (0, 1)$: the second column.

Any other input is a combination $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$, and linearity forces $A\mathbf{x} = x_1(\text{col}_1) + x_2(\text{col}_2)$ —

matrix-times-vector is a WEIGHTED SUM OF COLUMNS. Example: $A = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$ sends the unit square spanned by $\mathbf{e}_1, \mathbf{e}_2$ to the parallelogram spanned by $(2,0)$ and $(1,3)$: the plane is stretched, sheared, and every area is multiplied by $|\det A| = 6$. This picture explains the multiplication rule too: the columns of AB are A applied to the columns of B — do B first, then A — which is why AB and BA genuinely differ: shear-then-rotate is not rotate-then-shear.

Worked: a full 2×2 product, entry by entry

Compute AB for $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ 5 & 2 \end{pmatrix}$. Entry $(1,1)$: row 1 of A dot column 1 of B: $1 \cdot 0 + 2 \cdot 5 = 10$. Entry $(1,2)$: row 1 dot column 2: $1 \cdot 1 + 2 \cdot 2 = 5$. Entry $(2,1)$: row 2 dot column 1: $3 \cdot 0 + 4 \cdot 5 = 20$.

Entry $(2,2)$: $3{\cdot}1{+}4{\cdot}2 = 11$. So $AB = \begin{pmatrix} 10 & 5 \\ 20 & 11 \end{pmatrix}$. Now the other order: $BA = \begin{pmatrix} 0 \cdot 1 + 1 \cdot 3 & 0 \cdot 2 + 1 \cdot 4 \\ 5 \cdot 1 + 2 \cdot 3 & 5 \cdot 2 + 2 \cdot 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 11 & 18 \end{pmatrix}$ — completely different, as composition order predicts. One determinant check seals both: $\det A \cdot \det B = (-2)(-5) = 10$, and $\det(AB) = 10 \cdot 11 - 5 \cdot 20 = 10$ ✓.

Worked: solving a system as a matrix question

Solve $2x + y = 7$, $x - y = 2$ three compatible ways.

Way 1 — elimination: add the equations to kill y: $3x = 9$, so $x = 3$, then $y = 7 - 6 = 1$. Way 2 — inverse matrix: $A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}$ has $\det A = -2 - 1 = -3$, so $A^{-1} = \frac{1}{-3}\begin{pmatrix} -1 & -1 \\ -1 & 2 \end{pmatrix}$, and $\mathbf{x} = A^{-1}\begin{pmatrix} 7 \\ 2 \end{pmatrix} = \frac{1}{-3}\begin{pmatrix} -9 \\ -3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$. Way 3 — column picture: find weights making $x\begin{pmatrix} 2 \\ 1 \end{pmatrix} + y\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 7 \\ 2 \end{pmatrix}$; the weights 3 and 1 work. Three languages, one answer $(3,1)$ — and fluency means moving between them at will.

2 Row Reduction & Rank

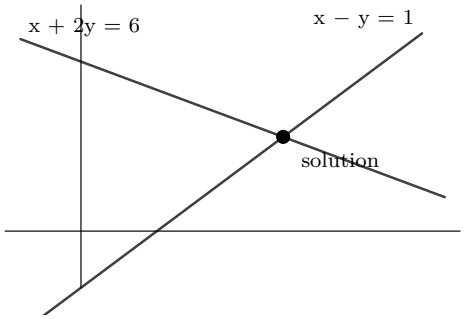
Gaussian elimination and the shape of solution sets

2.1 Gaussian Elimination

Three moves solve every system

Swap rows, scale a row, add a multiple of one row to another — none changes the solution set. March to echelon form; each PIVOT pins a variable, each pivotless column leaves a FREE variable. Unique solution, infinitely many, or none: the pivots decide.

Each equation is a line; the solution is where they meet.



Rank and the rank–nullity theorem

The rank is the number of pivots — the true dimension of what the matrix can reach. For an $m \times n$ matrix, (free variables) $= n -$ rank: a 3×5 homogeneous system of rank 3 always has a 2-parameter family of solutions.

Reading consistency off proportions

$x + 2y = 3$ and $2x + 4y = h$: the left sides are proportional, so either $h = 6$ (same line — infinitely many solutions) or $h \neq 6$ (parallel lines — none). Elimination makes the dichotomy mechanical.

Significance: the most-executed algorithm on Earth

Every engineering simulation, circuit solver, weather model, and structural analysis ends in a huge linear system, and Gaussian elimination (with clever pivoting) is how they are all solved. Its cost — about $\frac{n^3}{3}$ operations — is a number every computational scientist knows by heart.

Try it: read the verdict without finishing

The system reduces to rows $(1, 2 \,|\, 5)$, $(0, 3 \,|\, 6)$, $(0, 0 \,|\, c)$. For which c is it consistent? Work: the last row reads $0 = c$, so consistent only when $c = 0$ — and then two pivots, two variables: a UNIQUE solution ($y = 2$, $x = 1$). One glance at the echelon form settles existence, uniqueness, and the answer.

2.2 Proofs & Why It Matters

Proof: rank–nullity

Row reduce A ($m \times n$, rank r): r pivot columns and $n - r$ free columns. Each free variable can be set to 1 with the others 0, and back-substitution fills in the pivot variables — producing $n - r$ solutions of $A\mathbf{x} = \mathbf{0}$ that are independent (each has a 1 where the others have 0).

Conversely every homogeneous solution is determined by its free-variable values, so these span: the null space has dimension exactly $n - r$. ■

Proof: pivots decide the trichotomy

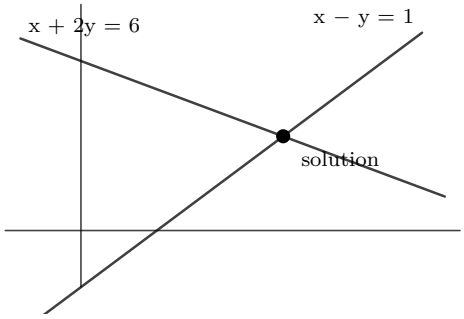
In echelon form, a row $(0; \cdots; 0 \,|\, b)$ with $b \neq 0$ reads $0 = b$: no solution. Otherwise, if every column has a pivot, back-substitution determines each variable uniquely: exactly one solution. Otherwise some column is pivot-free: its variable is free, and each of its infinitely many values extends to a solution. No fourth case exists. ■

2.3 Going Deeper: Explanations & Worked Problems

Elimination as an algorithm, not an art

The method is fully mechanical. (1) Find the leftmost column with a nonzero entry; swap that entry's row to the top — it is the first PIVOT. (2) Subtract multiples of the pivot row from every row below, zeroing that column beneath the pivot. (3) Ignore the pivot row and repeat on the remaining rows.

Each equation is a line; the solution is where they meet.



The result is echelon form: pivots marching down and right, zeros below each. Then read off the verdict: a row $(0 \cdots 0 \,|\, b)$, $b \neq 0$, means INCONSISTENT; otherwise count pivots — every column with a pivot pins a variable, every column without one donates a free parameter. Back-substitution finishes: solve the last pivot equation, substitute upward. The whole of solvability theory — unique/infinite/none, rank, dimension of the solution space — is a bookkeeping layer on this one procedure.

Worked: a full 3×3 elimination

Solve $x + y + z = 6$, $2x + y + 3z = 13$, $x + 2y - z = 2$.

Step 1 — clear column 1 below the pivot: $R_2 \leftarrow R_2 - 2R_1$: $(0, -1, 1 \,|\, 1)$; $R_3 \leftarrow R_3 - R_1$: $(0, 1, -2 \,|\, -4)$. Step 2 — clear column 2 below the new pivot: $R_3 \leftarrow R_3 + R_2$: $(0, 0, -1 \,|\, -3)$. Step 3 — echelon form reached with three pivots $(1, -1, -1)$: a unique solution exists. Step 4 — back-substitute: from row 3, $z = 3$; from row 2, $-y + 3 = 1$ so $y = 2$; from row 1, $x = 6 - 2 - 3 = 1$. Step 5 — verify in the ORIGINAL equations: $1 + 2 + 3 = 6$ ✓, $2 + 2 + 9 = 13$ ✓, $1 + 4 - 3 = 2$ ✓. The verification step costs ten seconds and catches nearly every slip.

Worked: reading rank and the solution space

Reduce $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 1 \end{pmatrix}$.

Step 1: $R_2 \leftarrow R_2 - 2R_1 = (0, 0, 0)$ — row 2 was twice row 1, and elimination EXPOSES the dependence as a zero row. Step 2: $R_3 \leftarrow R_3 - R_1 = (0, -1, -2)$. Echelon form: pivots in columns 1 and 2, none in column 3: rank 2. Step 3 — the homogeneous solutions: z is free; row 2 gives $y = -2z$; row 1 gives $x = -2y - 3z = 4z - 3z = z$. So the null space is the LINE of multiples of $(1, -2, 1)$, dimension 1 — and rank–nullity checks: 3 columns = 2 (rank) +1 (nullity) ✓. Rank counts surviving information; nullity counts what the matrix destroys.

3 Determinants & Inverses

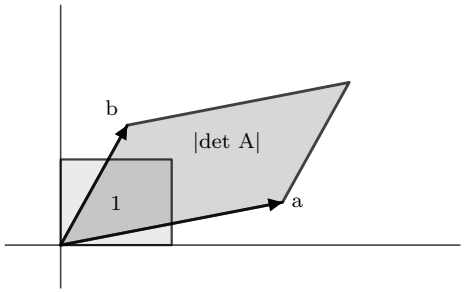
One number that decides invertibility

3.1 The Determinant

2×2 and 3×3 by hand

$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$; for 3×3, expand along a row: each entry times the 2×2 determinant of what remains, with alternating signs. Geometrically, $|\det|$ is the area (or volume) scale factor of the map.

The determinant is the area scale factor.



Invertible nonzero determinant

A^{-1} exists exactly when $\det A \neq 0$, and for 2×2: $A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ — swap the diagonal, negate the off-diagonal, divide by the determinant. Also $\det(AB) = \det A \det B$, so $\det(A^n) = (\det A)^n$.

Singular tuning

$\begin{pmatrix} 2 & 4 \\ 3 & k \end{pmatrix}$ fails to invert when $2k = 12$: at $k = 6$ the rows become parallel and the matrix flattens the plane onto a line.

Significance: one number, many jobs

Jacobians in calculus are determinants (area scaling); Cramer's rule solves small systems by determinants; the characteristic polynomial that finds eigenvalues IS a determinant; and orientation (does a transformation flip space?) is its sign. It is the only scalar that summarizes an entire matrix's invertibility and volume behavior.

Try it: zero without computing

What is det $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{pmatrix}$? Work: row 3 = row 1 + row 2 — a dependence, so the determinant is 0: no expansion needed. Spotting structure beats grinding cofactors; check by expansion if unconvinced.

3.2 Proofs & Why It Matters

Proof: ad bc is the area factor

The matrix sends the unit square (spanned by $\mathbf{e}_1, \mathbf{e}_2$) to the parallelogram spanned by the columns (a, c) and (b, d).

Embed in 3D and take the cross product: $(a, c, 0) \times (b, d, 0) = (0, 0, ad - bc)$, so the parallelogram's area is $|ad - bc|$ — the factor by which EVERY area scales, since any region is a limit of little squares. The sign records whether orientation flips. ■

Proof: the 2×2 inverse formula

Multiply directly: $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = (\det A)\, I.$

When $\det A \neq 0$, dividing by it gives A^{-1}; when $\det A = 0$ the product is the zero matrix, so no inverse can exist (an invertible product would force $\det A \cdot \det A^{-1} = 1$). ■

Proof: det(AB) = det A · det B (2×2, by areas)

The map $\mathbf{x} \mapsto AB\mathbf{x}$ first scales areas by $\det B$ (applying B), then by $\det A$ (applying A) — so

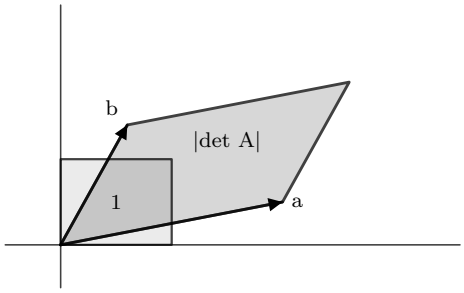
the composite scales areas by the product. Since the determinant IS the (signed) area factor, $\det(AB) = \det A \det B$. Setting $B = A^{n-1}$ repeatedly gives $\det(A^n) = (\det A)^n$. ■

3.3 Going Deeper: Explanations & Worked Problems

Three faces of one number

The determinant answers three seemingly different questions with one number. ALGEBRA: $A\mathbf{x} = \mathbf{b}$ has a unique solution iff $\det A \neq 0$ — the determinant is the product of the pivots (up to sign from row swaps), so "no zero pivot" and "nonzero determinant" are the same statement.

The determinant is the area scale factor.



GEOMETRY: $|\det A|$ is the factor by which A scales all areas (volumes in 3D), and its sign records orientation — negative means the plane got flipped. COMPUTATION: expand along any row or column (each entry times its minor, alternating signs $+, -, +, \ldots$ in a checkerboard), or better, row reduce and multiply pivots — for large matrices, elimination is exponentially faster than cofactors. Useful consequences worth having at your fingertips: a matrix with two equal rows has determinant 0 (swap them: the sign flips yet nothing changed); triangular matrices have determinant = product of the diagonal; and $\det(A^\mathsf{T}) = \det A$.

Worked: a 3×3 determinant by cofactors, then again by elimination

Let $M = \begin{pmatrix} 2 & 1 & 3 \\ 0 & 4 & 1 \\ 1 & 2 & 0 \end{pmatrix}$. COFACTORS along row 1: $\det M = 2\det\begin{pmatrix} 4 & 1 \\ 2 & 0 \end{pmatrix} - 1\det\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + 3\det\begin{pmatrix} 0 & 4 \\ 1 & 2 \end{pmatrix}$. Evaluate each 2×2: $(0 - 2) = -2$; $(0 - 1) = -1$; $(0 - 4) = -4$.

Assemble with the $+, -, +$ signs: $2(-2) - 1(-1) + 3(-4) = -4 + 1 - 12 = -15$. ELIMINATION: swap $R_1 \leftrightarrow R_3$ (determinant flips sign), then clear below: the pivots come out $1, 4, \frac{15}{4}\ldots$ product 15, times the swap sign gives -15 ✓. Same number, two engines — cofactors for small or sparse matrices, elimination for everything else.

Worked: using the inverse formula end to end

Solve $\begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 7 \\ 12 \end{pmatrix}$ by the inverse.

Step 1 — determinant: $3 \cdot 2 - 1 \cdot 5 = 1$. Step 2 — the 2×2 inverse recipe (swap diagonal, negate off-diagonal, divide): $A^{-1} = \frac{1}{1} \begin{pmatrix} 2 & -1 \\ -5 & 3 \end{pmatrix}$. Step 3 — multiply: $\mathbf{x} = \begin{pmatrix} 2 & -1 \\ -5 & 3 \end{pmatrix} \begin{pmatrix} 7 \\ 12 \end{pmatrix} = \begin{pmatrix} 14 - 12 \\ -35 + 36 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Step 4 — check in the original system: $3 \cdot 2 + 1 = 7$ ✓, $5 \cdot 2 + 2 = 12$ ✓. Because $\det = 1$, this matrix even preserves areas exactly — an example of the geometry and the algebra agreeing to the digit.

4 Vector Spaces & Bases

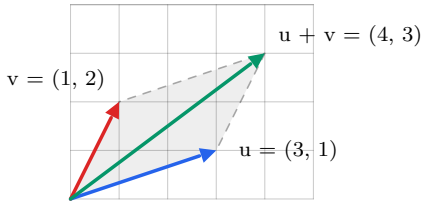
Span, independence, basis, dimension

4.1 The Language of Vector Spaces

Span, independence, basis

The SPAN of a set is everything reachable by linear combinations. A set is INDEPENDENT when no member is a combination of the others. A BASIS is both: independent AND spanning — every vector then has exactly ONE coordinate representation, and every basis of a space has the same size: its DIMENSION.

A subspace is closed under addition and scaling.



Subspaces cut dimension by conditions

Each independent linear condition on $\mathbb{R}^n$ removes one dimension: the hyperplane $x_1 + x_2 + x_3 + x_4 = 0$ in $\mathbb{R}^4$ has dimension 3. Solution sets of homogeneous systems ARE subspaces — this is where geometry and equations meet.

Coordinates in a new basis

To write $(7, 3)$ in the basis $\{(1, 1), (1, -1)\}$: solve $a + b = 7$, $a - b = 3$, giving $a = 5$, $b = 2$. Changing basis is solving a linear system — and orthogonal bases make it a pair of one-line dot products.

Significance: the same axioms, everywhere

Polynomials, audio signals, quantum states, and solutions of linear ODEs all form vector spaces — the span/basis/dimension language transfers wholesale. That is why "the solution space of $y'' + y = 0$ is 2-dimensional with basis $\{\cos, \sin\}$" is a sentence, not an analogy: the differential equations course runs on this chapter.

Try it: dimensions by conditions

What is the dimension of the space of 2×2 matrices with trace 0? Work: the space of all 2×2 matrices has dimension 4; trace zero is ONE linear condition; dimension $4 - 1 = 3$. A basis: $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$.

4.2 Proofs & Why It Matters

Proof: coordinates in a basis are unique

Suppose $\mathbf{v} = \sum a_i \mathbf{b}_i = \sum c_i \mathbf{b}_i$ in a basis $\{\mathbf{b}_i\}$. Subtract: $\sum (a_i - c_i)\mathbf{b}_i = \mathbf{0}$. Independence forces every coefficient to vanish: $a_i = c_i$. Existence comes from spanning, uniqueness from independence — a basis delivers both at once. ■

Proof sketch: every basis has the same size

The exchange argument: if $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is independent and $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ spans, insert the $\mathbf{u}$s one at a time into the spanning set, each time ejecting some $\mathbf{b}$ (writing the new $\mathbf{u}$ in the current spanning set must use a surviving $\mathbf{b}$, which can then be solved for). The process never runs out of $\mathbf{b}$s to eject, so $m \leq n$. Applying this both ways to two bases gives $m \leq n$ and $n \leq m$: dimension is well-defined. ■

4.3 Going Deeper: Explanations & Worked Problems

Independence testing is elimination in disguise

To decide whether vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are independent, ask whether $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$ forces all $c_i = 0$ — but that is a homogeneous linear SYSTEM in the c_i, with the vectors as columns.

Row reduce: if every column earns a pivot, only the zero combination works — independent; a pivotless column means a free variable, hence a nonzero combination summing to zero — dependent,

and back-substitution hands you the actual dependence. The same computation delivers dimension: the pivot columns form a basis of the span, so dim(span) = rank. This is the great economy of the subject — span, independence, basis, dimension, and solvability are all read off ONE echelon form.

Worked: dependence exposed, coefficients and all

Test $(1,0,1),(0,1,1),(1,1,2)$. Step 1 — put them as columns and reduce: $\begin{pmatrix}1&0&1\\0&1&1\\1&1&2\end{pmatrix} \to$ $R_3 - R_1 : \begin{pmatrix}1&0&1\\0&1&1\\0&1&1\end{pmatrix} \to R_3 - R_2 : \begin{pmatrix}1&0&1\\0&1&1\\0&0&0\end{pmatrix}$.

Step 2 — read: pivots in columns 1, 2 only — DEPENDENT, dimension of the span is 2. Step 3 — recover the dependence: the free column says c_3 is free; rows give $c_1 = -c_3$ and $c_2 = -c_3$; take $c_3 = 1$: $-(1,0,1) - (0,1,1) + (1,1,2) = (0,0,0)$ ✓ — i.e. $\mathbf{v}_3 = \mathbf{v}_1 + \mathbf{v}_2$, the hidden relation, now explicit. Step 4 — basis: keep the pivot columns $\{(1,0,1),(0,1,1)\}$; they span the same plane with nothing wasted.

Worked: change of basis with an orthogonal shortcut

Write $(7,3)$ in the basis $\mathbf{b}_1 = (1,1)$, $\mathbf{b}_2 = (1,-1)$. THE SYSTEM ROUTE: $a(1,1)+b(1,-1) = (7,3)$ means $a+b=7$ and $a-b=3$; adding, $2a=10$, $a=5$; subtracting, $b=2$.

THE ORTHOGONAL SHORTCUT: because $\mathbf{b}_1 \cdot \mathbf{b}_2 = 1-1=0$, each coefficient is an independent projection: $a = \dfrac{(7,3)\cdot(1,1)}{(1,1)\cdot(1,1)} = \dfrac{10}{2} = 5$ and $b = \dfrac{(7,3)\cdot(1,-1)}{2} = \dfrac{4}{2} = 2$ — no system solved, no interaction between coordinates. Check: $5(1,1) + 2(1,-1) = (7,3)$ ✓. This is precisely why orthogonal bases (and eventually Fourier series) are prized: coordinates decouple.

5 Linear Transformations

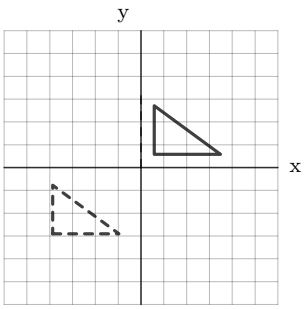
Kernel, image, rotations and reflections

5.1 Maps That Respect Addition

The matrix IS the transformation

A linear map is determined by where it sends the basis vectors — those images are the COLUMNS of its matrix. Rotation by 90°: $\begin{pmatrix}0&-1\\1&0\end{pmatrix}$; reflection across $y=x$: $\begin{pmatrix}0&1\\1&0\end{pmatrix}$; composition of maps is a matrix product.

A linear map is determined by where the basis goes.



Kernel and image

The KERNEL is everything sent to zero (the solution space of $A\mathbf{x} = \mathbf{0}$); the IMAGE is everything hit (the span of the columns). Their dimensions add to the number of input variables — rank–nullity, now with geometric meaning.

Determinants classify the geometry

A reflection has determinant -1 (flips orientation, preserves area); rotations have determinant $+1$; a projection has determinant 0 (it flattens). Two quarter-turns compose to $R_{90}^2 = -I$: every point goes to its antipode.

Significance: geometry becomes computation

Every rotation of a 3D game character, every camera projection, every robot-arm pose is a linear (or affine) transformation composed from simpler ones by matrix multiplication. Kernel and image answer engineering questions directly: the kernel of a measurement matrix is what your sensors CANNOT see.

Try it: build a matrix from geometry

Write the matrix that stretches x by 3 and reflects y. Work: $\mathbf{e}_1 \mapsto (3, 0)$ and $\mathbf{e}_2 \mapsto (0, -1)$: columns give $\begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$. Determinant -3: areas scale by 3, orientation flips — both facts visible before multiplying anything.

5.2 Proofs & Why It Matters

Proof: the columns determine the map

Any $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, so by linearity $T(\mathbf{x}) = x_1T(\mathbf{e}_1) + \cdots + x_nT(\mathbf{e}_n)$ — a combination of the images of the basis vectors with weights given by the coordinates. Assembling $T(\mathbf{e}_j)$ as columns of a matrix A makes this exactly $A\mathbf{x}$. Linearity leaves no freedom beyond the

columns. ■

Proof: reflections have determinant 1

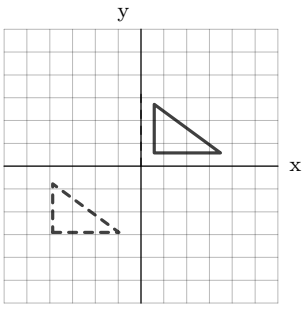
A reflection fixes a line and flips the perpendicular direction. In a basis adapted to those two directions its matrix is $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, with determinant -1; changing basis conjugates the matrix, and $\det(P^{-1}MP) = \det M$, so the determinant is -1 in EVERY basis. Areas are preserved ($|-1| = 1$) but orientation reverses. ■

5.3 Going Deeper: Explanations & Worked Problems

A gallery of maps, with their matrices derived

Derive, never memorize: a map's matrix has columns $T(\mathbf{e}_1), T(\mathbf{e}_2)$. ROTATION by θ: $\mathbf{e}_1 = (1,0)$ lands at $(\cos\theta, \sin\theta)$; $\mathbf{e}_2 = (0,1)$ at $(-\sin\theta, \cos\theta)$; hence $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, $\det = \cos^2 + \sin^2 = 1$. REFLECTION across $y = x$: the basis vectors swap, so $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\det = -1$. PROJECTION onto the x-axis: $\mathbf{e}_1$ stays, $\mathbf{e}_2$ dies: $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\det = 0$ — information is destroyed, and indeed the kernel is the whole y-axis. SHEAR: $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ slides horizontals by their height, $\det = 1$ (areas survive shearing). The determinant column of this gallery — $1, -1, 0, 1$ — is a complete orientation-and-area story at a glance.

A linear map is determined by where the basis goes.



Worked: kernel and image of one map, completely

Let $T(x,y,z) = (x+y, ; y+z)$, matrix $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$. KERNEL: solve $x + y = 0$, $y + z = 0$: $x = -y$, $z = -y$, y free — the line spanned by $(-1,1,-1)$ (check: $T(-1,1,-1) = (0,0)$ ✓). Dimension 1.

IMAGE: the span of the columns $(1,0), (1,1), (0,1)$; already the first two are independent, so the

image is ALL of $\mathbb{R}^2$, dimension 2. RANK–NULLITY audit: $3 = 2+1$ ✓. Interpretation: T squashes three-dimensional space onto the plane, and the crushing happens exactly along the kernel line — every fiber $T^{-1}(\mathbf{b})$ is a translate of that line, which is the geometric meaning of "general solution = particular + homogeneous."

Worked: composing transformations by multiplying matrices

What single map is "reflect across $y = x$, THEN rotate 90° counterclockwise"?

Step 1 — matrices: reflection $F = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, rotation $R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Step 2 — compose in the right ORDER: the reflection acts first, so the product is RF (closest to the input acts first): $RF = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ — reflection across the y-axis. Step 3 — the other order: $FR = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, reflection across the x-axis: DIFFERENT, as expected. Step 4 — determinant ledger: $\det R \cdot \det F = (1)(-1) = -1 = \det(RF)$ ✓ — a rotation composed with a reflection is always another reflection.

6 Orthogonality & Least Squares

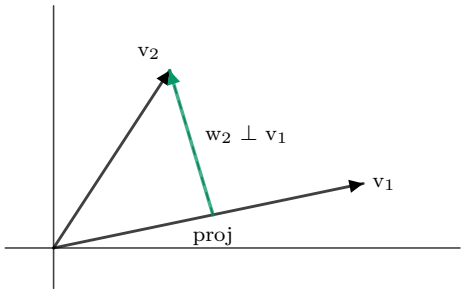
Projections, distance, and best fit

6.1 Projection and Best Approximation

The projection formula

The projection of $\mathbf{u}$ onto $\mathbf{v}$ is $\dfrac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\,\mathbf{v}$ — the shadow $\mathbf{u}$ casts on the line through $\mathbf{v}$. What is LEFT OVER, $\mathbf{u}$ – proj, is perpendicular to $\mathbf{v}$; its length is the distance to the line.

Subtracting the projection leaves the orthogonal part.



Least squares: closest, not exact

When $A\mathbf{x} = \mathbf{b}$ has no solution, solve $A^{\mathsf{T}}A\,\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$ instead: $A\hat{\mathbf{x}}$ is the PROJECTION of

b onto the column space — the best possible fit. Every regression line in statistics is this computation.

The simplest regression

Fitting a constant c to data $1, 3, 5$ minimizes $(c-1)^2 + (c-3)^2 + (c-5)^2$; the derivative vanishes at the MEAN, $c = 3$. Least squares generalizes the average — and orthogonality is why it works.

Significance: least squares built the modern world

Gauss invented least squares to find a lost asteroid (Ceres, 1801) — and today every regression, every GPS position fix (overdetermined satellite equations), and every curve fit in every lab runs the normal equations from this topic. Orthogonality is also why JPEG and MP3 work: signals are projected onto orthogonal cosine bases and the small components discarded.

Try it: a distance via projection

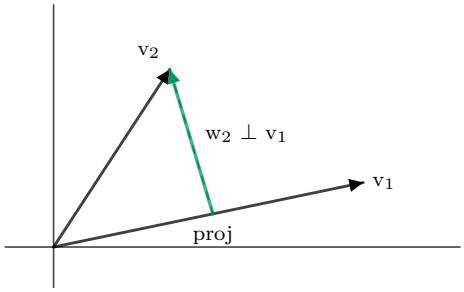
Find the distance from $(5, 1)$ to the line through the origin along $(2, 1)$. Work: coefficient $c = \frac{10+1}{4+1} = \frac{11}{5}$; projection $\frac{11}{5}(2, 1)$; remainder $(5, 1) - \left(\frac{22}{5}, \frac{11}{5}\right) = \left(\frac{3}{5}, -\frac{6}{5}\right)$; distance $= \sqrt{\frac{9+36}{25}} = \frac{3\sqrt{5}}{5}$. Verify: remainder $\cdot(2, 1) = \frac{6-6}{5} = 0$ ✓.

6.2 Proofs & Why It Matters

Proof: the projection formula

Seek the multiple $c\mathbf{v}$ of $\mathbf{v}$ making the remainder $\mathbf{u} - c\mathbf{v}$ orthogonal to $\mathbf{v}$: $(\mathbf{u} - c\mathbf{v}) \cdot \mathbf{v} = 0$ gives $c = \dfrac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}$ in one step.

Subtracting the projection leaves the orthogonal part.



This $c\mathbf{v}$ is also the CLOSEST point on the line: for any other multiple $t\mathbf{v}$, Pythagoras gives $|\mathbf{u} - t\mathbf{v}|^2 = |\mathbf{u} - c\mathbf{v}|^2 + |c\mathbf{v} - t\mathbf{v}|^2 \geq |\mathbf{u} - c\mathbf{v}|^2$. ■

Proof: the normal equations of least squares

The best $\hat{\mathbf{x}}$ minimizes $|A\mathbf{x} - \mathbf{b}|$, i.e. makes $A\hat{\mathbf{x}}$ the closest point to $\mathbf{b}$ in the column space. Closest means the error $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (else sliding along that column would shrink it — the projection argument again).

Orthogonality to all columns at once is $A^{\mathsf{T}}(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$, i.e. $A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$. Fitting a constant to data makes this one equation whose solution is the mean. ■

6.3 Going Deeper: Explanations & Worked Problems

Decomposing a vector: the along-and-across split

The projection formula does more than find shadows — it SPLITS any $\mathbf{u}$ into two orthogonal pieces relative to a direction $\mathbf{v}$: the part along, $\text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}$, and the part across, $\mathbf{u} - \text{proj}_{\mathbf{v}}\mathbf{u}$, which is orthogonal to $\mathbf{v}$ by construction. Pythagoras then bookkeeps the lengths: $|\mathbf{u}|^2 = |\text{along}|^2 + |\text{across}|^2$. The "across" length IS the distance from the point to the line — that is where the point-to-line distance formula comes from. Iterating the split against several directions in turn is Gram–Schmidt: peel off the component along each previously built direction, keep what remains, normalize — manufacturing an orthogonal basis from any basis, one subtraction at a time.

Worked: project, split, and measure — one example, all the way

Take $\mathbf{u} = (3, 4)$ and direction $\mathbf{v} = (1, 1)$.

Step 1 — the coefficient: $c = \dfrac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}} = \dfrac{3+4}{1+1} = \dfrac{7}{2}$. Step 2 — the along part: $\frac{7}{2}(1,1) = \left(\frac{7}{2}, \frac{7}{2}\right)$. Step 3 — the across part: $(3,4) - \left(\frac{7}{2}, \frac{7}{2}\right) = \left(-\frac{1}{2}, \frac{1}{2}\right)$; check orthogonality: $\left(-\frac{1}{2}\right)(1) + \frac{1}{2}(1) = 0$ ✓. Step 4 — distance from $(3,4)$ to the line $y = x$: the across length, $\sqrt{\frac{1}{4} + \frac{1}{4}} = \frac{\sqrt{2}}{2}$ — matching the point-to-line formula $\frac{|3-4|}{\sqrt{2}}$ ✓. Step 5 — Pythagoras audit: $\left(\frac{7\sqrt{2}}{2}\right)^2 + \left(\frac{\sqrt{2}}{2}\right)^2 = \frac{49}{2} + \frac{1}{2} = 25 = |\mathbf{u}|^2$ ✓. Five steps, three self-checks — the decomposition verifies itself.

Worked: an actual least-squares line, normal equations and all

Fit $y = mx + c$ to the points $(0,1), (1,3), (2,5)$... which happen to be collinear — so fit $(0,1), (1,2), (2,4)$ instead.

Step 1 — the (unsolvable) system: $c = 1$, $m + c = 2$, $2m + c = 4$, i.e. $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$, $\mathbf{b} = (1, 2, 4)$. Step 2 — normal equations $A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$: compute $A^{\mathsf{T}}A = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}$ and $A^{\mathsf{T}}\mathbf{b} = \begin{pmatrix} 10 \\ 7 \end{pmatrix}$. Step 3 — solve: $5m + 3c = 10$ and $3m + 3c = 7$; subtracting, $2m = 3$, $m = \frac{3}{2}$, then $c = \frac{7 - 4.5}{3} = \frac{5}{6}$. Step 4 — the fitted line $y = \frac{3}{2}x + \frac{5}{6}$ predicts $\frac{5}{6}, \frac{7}{3}, \frac{23}{6}$; residuals $\frac{1}{6}, -\frac{1}{3}, \frac{1}{6}$ sum to 0 — as they must, since orthogonality of the error to the all-ones column is literally one of the normal equations.

7 Eigenvalues

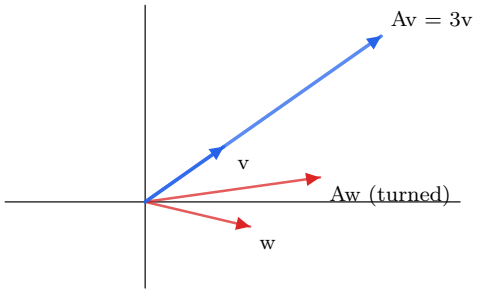
The directions a matrix merely stretches

7.1 Eigenvalues & Eigenvectors

Stretch directions

An eigenvector $\mathbf{v}$ satisfies $A\mathbf{v} = \lambda\mathbf{v}$: the matrix only STRETCHES it by the eigenvalue λ. Find eigenvalues from $\det(A - \lambda I) = 0$; for 2×2 this is $\lambda^2 - (\operatorname{tr} A)\lambda + \det A = 0$, so eigenvalues SUM to the trace and MULTIPLY to the determinant.

An eigenvector keeps its direction; A only scales it.



Powers become easy

If $A\mathbf{v} = \lambda\mathbf{v}$ then $A^k\mathbf{v} = \lambda^k\mathbf{v}$, and $\operatorname{tr}(A^k) = \sum \lambda_i^k$. This is why eigenvalues run the long-term behavior of everything iterative — Fibonacci (Q^n holds F_n, and $\det Q = -1$ gives Cassini's identity in one line), Markov chains, and populations alike.

A symmetric 2×2

$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$: trace 4, determinant 3, so $\lambda^2 - 4\lambda + 3 = 0$ gives $\lambda = 1, 3$ — with perpendicular eigenvectors $(1, -1)$ and $(1, 1)$, as symmetry guarantees.

Significance: the axes the world actually uses

Google ranks pages by an eigenvector; bridges have resonant eigenfrequencies (soldiers break step crossing them); quantum energy levels ARE eigenvalues; population models stabilize along their dominant eigenvector. Whenever a system is applied repeatedly, its eigenvalues write its future.

Try it: eigenvalues at sight

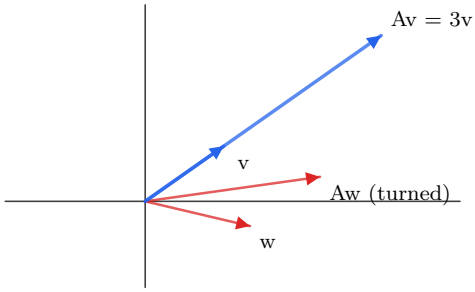
Find the eigenvalues of the triangular matrix $\begin{pmatrix} 4 & 7 \\ 0 & 2 \end{pmatrix}$. Work: for triangular matrices the eigenvalues ARE the diagonal entries: 4 and 2 (the characteristic determinant is $(4-\lambda)(2-\lambda)$ regardless of the 7). Check the trace: $4+2=6$ ✓ and det: 8 ✓.

7.2 Proofs & Why It Matters

Proof: eigenvalues sum to the trace, multiply to the determinant (2×2)

Expand the characteristic polynomial: $\det \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} = \lambda^2 - (a+d)\lambda + (ad-bc)$. By Vieta, its roots λ_1, λ_2 satisfy $\lambda_1 + \lambda_2 = a + d = \operatorname{tr} A$ and $\lambda_1 \lambda_2 = ad - bc = \det A$.

An eigenvector keeps its direction; A only scales it.



The trace and determinant are the eigenvalues' fingerprints, readable without solving anything. ■

Proof: Av = v

Induction on k. Base: $A\mathbf{v} = \lambda\mathbf{v}$ is given. Step: if $A^k\mathbf{v} = \lambda^k\mathbf{v}$, apply A once more: $A^{k+1}\mathbf{v} = A(\lambda^k\mathbf{v}) = \lambda^k(A\mathbf{v}) = \lambda^{k+1}\mathbf{v}$.

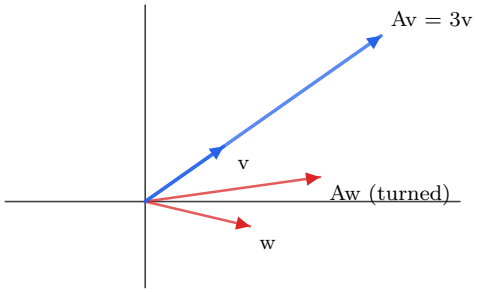
So along an eigenvector, iterating the matrix is just repeated scalar multiplication — the engine behind diagonalization, Fibonacci formulas, and Markov-chain limits. ■

7.3 Going Deeper: Explanations & Worked Problems

Finding eigenvalues and eigenvectors: the complete procedure

Step one: eigenvalues are the roots of $\det(A - \lambda I) = 0$ — the values of λ making $A - \lambda I$ singular, because only a singular matrix can send a NONZERO vector to zero. For a 2×2, skip straight to $\lambda^2 - (\operatorname{tr} A)\lambda + \det A = 0$.

An eigenvector keeps its direction; A only scales it.



Step two: for EACH root, solve $(A-\lambda I)\mathbf{v}=\mathbf{0}$ — the matrix is now singular by design, so its rows are dependent and one row determines $\mathbf{v}$ up to scale. Step three: use them — along eigenvectors the matrix acts as a number, so A^k acts as λ^k, systems decouple, and $A^k\mathbf{x}_0$ for large k tilts toward the eigenvector of the LARGEST $|\lambda|$ (the power method, and the reason Markov chains forget their starting point). Two structural gifts: symmetric matrices always have real eigenvalues with orthogonal eigenvectors, and distinct eigenvalues always give independent eigenvectors.

Worked: full eigen-analysis of a symmetric matrix

Analyze $A=\begin{pmatrix}2&1\\1&2\end{pmatrix}$.

Step 1 — characteristic equation via trace and determinant: tr $=4$, det $=3$, so $\lambda^2-4\lambda+3=0$, factoring as $(\lambda-1)(\lambda-3)=0$: eigenvalues 1 and 3. Step 2 — eigenvector for $\lambda=3$: $(A-3I)\mathbf{v}=\begin{pmatrix}-1&1\\1&-1\end{pmatrix}\mathbf{v}=\mathbf{0}$ gives $v_1=v_2$: take $(1,1)$. Step 3 — for $\lambda=1$: $\begin{pmatrix}1&1\\1&1\end{pmatrix}\mathbf{v}=\mathbf{0}$ gives $v_1=-v_2$: take $(1,-1)$. Step 4 — checks: $A(1,1)=(3,3)=3(1,1)$ ✓; $A(1,-1)=(1,-1)$ ✓; and $(1,1)\cdot(1,-1)=0$ — orthogonal, as symmetry promised. Step 5 — payoff: $A^{10}(1,0)=A^{10}\cdot\frac{1}{2}\left[(1,1)+(1,-1)\right]=\frac{1}{2}\left[3^{10}(1,1)+(1,-1)\right]$ — a closed form with no repeated multiplication anywhere.

Worked: eigenvalues running a dynamical system

A population splits between two sites, redistributing each year by $\mathbf{p}_{k+1}=M\mathbf{p}_k$ with $M=\begin{pmatrix}0.8&0.3\\0.2&0.7\end{pmatrix}$ (columns sum to 1: nobody is lost).

Step 1 — eigenvalues: tr $=1.5$, det $=0.56-0.06=0.5$: $\lambda^2-1.5\lambda+0.5=0$ factors as $(\lambda-1)(\lambda-0.5)=0$. Step 2 — the $\lambda=1$ eigenvector: $(M-I)\mathbf{v}=\mathbf{0}$ gives $-0.2v_1+0.3v_2=0$, so $\mathbf{v}\propto(3,2)$. Step 3 — conclusion: the $\lambda=0.5$ component HALVES every year and dies; every starting split converges to the ratio $3:2$ — sixty percent at site one, forever, regardless of history. Step 4 — check one iteration from $(1,0)$: $M(1,0)=(0.8,0.2)$, then $(0.7,0.3)$, then $(0.65,0.35)$ — marching monotonically toward $(0.6,0.4)$ ✓. The eigenvalue 1 holds the destination; the second eigenvalue sets the speed of forgetting.

8 Diagonalization & Matrix Powers

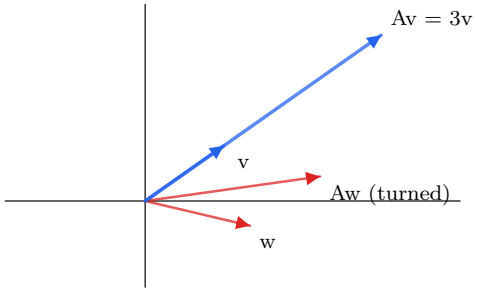
$A = PDP$ and everything it unlocks

8.1 Changing to the Eigenbasis

Diagonalization is a change of glasses

If a matrix A has n independent eigenvectors, pack them as the columns of P and their eigenvalues into the diagonal of D: then $A = PDP^{-1}$. Reading right to left: P^{-1} translates a vector into eigen-coordinates, D stretches each coordinate by its eigenvalue (the easy part), and P translates back.

In the eigenbasis the map is just scaling.



The matrix was never complicated — we were looking at it in the wrong basis. Consequences cascade: $A^k = PD^kP^{-1}$ (powers act on eigenvalues only), $\text{tr}(A^k) = \sum \lambda_i^k$, and functions of matrices ($\sqrt{A}$, e^{At}) are defined by applying them to the diagonal. Computing A^{100} costs a diagonalization, not a hundred multiplications.

When diagonalization fails

A repeated eigenvalue may come with too few eigenvectors: the shear $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has the double eigenvalue 1 but only ONE eigenvector direction — it is DEFECTIVE, and no basis makes it diagonal. The fix (Jordan form) adds an off-diagonal 1 and, in differential equations, produces exactly the $te^{\lambda t}$ solutions you met with repeated characteristic roots. The two "repeated root" phenomena are one phenomenon.

Recurrences are matrix powers

Fibonacci's $F_{n+1} = F_n + F_{n-1}$ is the matrix iteration $\binom{F_{n+1}}{F_n} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \binom{F_n}{F_{n-1}}$, so F_n lives inside Q^n.

Diagonalizing Q (eigenvalues $\varphi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$) yields BINET'S FORMULA $F_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}$ — a closed form for a recursive sequence, and since $|\psi| < 1$, the growth rate of Fibonacci is exactly

the golden ratio. Every linear recurrence yields to the same treatment: its characteristic equation is the characteristic polynomial of its companion matrix.

Try it: trace of a big power in ten seconds

Compute $\operatorname{tr}(A^5)$ for $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Work: eigenvalues from trace 4, det 3: $\lambda = 1, 3$. Then $\operatorname{tr}(A^5) = 1^5 + 3^5 = 244$. No matrix was multiplied. (For contrast, try computing A^5 by hand — five multiplications, sixteen entries each — then take the trace and confirm.)

8.2 Proofs & Why It Matters

Proof: A = PDP, and why traces are power sums

$A^2 = (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1} = PD^2P^{-1}$ — the inner pair cancels; induction extends to all k. For the trace, use $\operatorname{tr}(XY) = \operatorname{tr}(YX)$: $\operatorname{tr}(A^k) = \operatorname{tr}(PD^kP^{-1}) = \operatorname{tr}(D^kP^{-1}P) = \operatorname{tr}(D^k) = \sum \lambda_i^k$.

■ SIGNIFICANCE: this identity is how one counts closed walks in graphs (adjacency-matrix powers), computes partition functions in physics, and analyzes mixing of Markov chains — three fields, one cancellation.

Significance: exponentials of matrices

Define $e^{At} = Pe^{Dt}P^{-1}$, i.e. exponentiate each eigenvalue. Then $\mathbf{u}(t) = e^{At}\mathbf{u}_0$ solves $\mathbf{u}' = A\mathbf{u}$ — the systems topic of Differential Equations is literally this formula. Diagonalization is the load-bearing wall between the two courses.

8.3 Going Deeper: Worked Problems

Worked: a full diagonalization, then a closed-form power

Diagonalize $A = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$ and use it to find a formula for A^n.

Step 1 — eigenvalues: trace 7, det 10: $\lambda^2 - 7\lambda + 10 = (\lambda - 2)(\lambda - 5) = 0$. Step 2 — eigenvectors: for $\lambda = 5$, $(A - 5I)\mathbf{v} = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}\mathbf{v} = \mathbf{0}$ gives $(1, 1)$; for $\lambda = 2$, $\begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}\mathbf{v} = \mathbf{0}$ gives $(1, -2)$. Step 3 — $P = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$, $D = \operatorname{diag}(5, 2)$, and $P^{-1} = \frac{1}{-3}\begin{pmatrix} -2 & -1 \\ -1 & 1 \end{pmatrix}$. Step 4 — $A^n = PD^nP^{-1} = \frac{1}{3}\begin{pmatrix} 2 \cdot 5^n + 2^n & 5^n - 2^n \\ 2 \cdot 5^n - 2 \cdot 2^n & 5^n + 2 \cdot 2^n \end{pmatrix}$. Step 5 — check $n = 1$: $\frac{1}{3}\begin{pmatrix} 12 & 3 \\ 6 & 9 \end{pmatrix} = A$ ✓. Every entry is now a formula in n; the 5^n terms dominate, along the eigenvector $(1, 1)$.

Worked: solving a recurrence by diagonalization

A sequence satisfies $a_{n+1} = 5a_n - 6a_{n-1}$ with $a_0 = 1$, $a_1 = 4$. Find a closed form.

Step 1 — companion matrix: $\binom{a_{n+1}}{a_n} = \begin{pmatrix} 5 & -6 \\ 1 & 0 \end{pmatrix} \binom{a_n}{a_{n-1}}$. Step 2 — its characteristic polynomial is $\lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3)$ — exactly the recurrence with λ in place of the shift. Step 3 — so $a_n = c_1 \cdot 2^n + c_2 \cdot 3^n$; the data give $c_1 + c_2 = 1$ and $2c_1 + 3c_2 = 4$, hence $c_2 = 2$, $c_1 = -1$: $a_n = 2 \cdot 3^n - 2^n$. Step 4 — verify a_2: recurrence gives $5 \cdot 4 - 6 \cdot 1 = 14$; formula gives $18 - 4 = 14$ ✓. Diagonalizing the companion matrix IS the "characteristic equation" method — now you know why it works.

9 Symmetric Matrices & Quadratic Forms

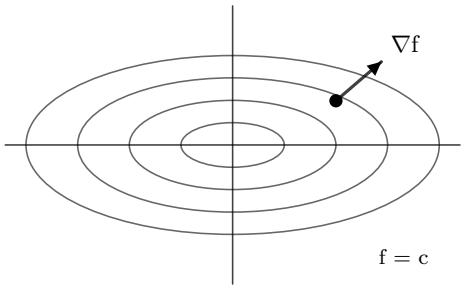
The spectral theorem and definiteness

9.1 The Nicest Matrices in Mathematics

The spectral theorem

A SYMMETRIC matrix ($A = A^\mathsf{T}$) is as good as it gets: all eigenvalues real, eigenvectors of distinct eigenvalues automatically ORTHOGONAL, and a full orthonormal eigenbasis always exists — $A = Q\Lambda Q^\mathsf{T}$ with Q a rotation.

Level sets of a positive definite form are ellipses.



Geometrically: every symmetric matrix is a pure stretch along some perpendicular set of axes, no shearing, no rotation mixed in. This is the spectral theorem, and it is why symmetric matrices rule applied mathematics: covariance matrices, Hessians, stiffness matrices, and graph Laplacians are all symmetric by construction.

Quadratic forms and definiteness

A quadratic form $q(\mathbf{x}) = \mathbf{x}^\mathsf{T} A\mathbf{x}$ (write $ax^2 + bxy + cy^2$ with the symmetric matrix $\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$ — HALVE the cross term) is classified by eigenvalue signs: all positive — positive definite, a

bowl with a genuine minimum; all negative — a dome; mixed — a saddle. On the unit circle, q ranges exactly over $[\lambda_{\min}, \lambda_{\max}]$, attained along the eigenvectors (the PRINCIPAL AXES). The quick test without eigenvalues: positive definite iff $a > 0$ and $\det A > 0$. Recognize this as the honest form of the second-derivative D-test from multivariable calculus — the Hessian is a symmetric matrix, and $D > 0, f_{xx} > 0$ is exactly the 2×2 definiteness test.

Ellipses are eigen-pictures

The level curve $q(\mathbf{x}) = 1$ of a positive-definite form is an ELLIPSE whose axes point along the eigenvectors with half-lengths $1/\sqrt{\lambda_i}$. Rotate to the eigenbasis and the cross term vanishes: $\lambda_1 u^2 + \lambda_2 v^2 = 1$. Every tilted ellipse you have ever completed the square on was a symmetric matrix asking to be diagonalized.

Try it: extremes on the unit circle

Maximize and minimize $q = 5x^2 + 4xy + 5y^2$ on $x^2 + y^2 = 1$. Work: matrix $\begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix}$; trace 10, det 21; eigenvalues 3 and 7. Max $= 7$ at $\pm\frac{1}{\sqrt{2}}(1,1)$, min $= 3$ at $\pm\frac{1}{\sqrt{2}}(1,-1)$.

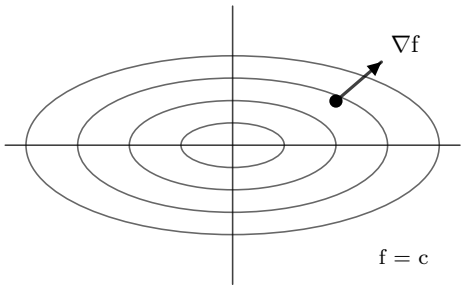
Answer. Verify: $q\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{5}{2} + 2 + \frac{5}{2} = 7$ ✓. Lagrange multipliers on a quadratic ARE the eigenvalue problem.

9.2 Proofs & Why It Matters

Proof: eigenvectors of a symmetric matrix are orthogonal

Let $A\mathbf{u} = \lambda\mathbf{u}$ and $A\mathbf{v} = \mu\mathbf{v}$ with $\lambda \neq \mu$. Compute $\lambda(\mathbf{u} \cdot \mathbf{v}) = (A\mathbf{u}) \cdot \mathbf{v} = \mathbf{u}^\mathsf{T} A^\mathsf{T} \mathbf{v} = \mathbf{u}^\mathsf{T} A\mathbf{v} = \mu(\mathbf{u} \cdot \mathbf{v})$, using symmetry in the middle.

Level sets of a positive definite form are ellipses.



So $(\lambda - \mu)(\mathbf{u} \cdot \mathbf{v}) = 0$, and since $\lambda \neq \mu$: $\mathbf{u} \cdot \mathbf{v} = 0$. ■ SIGNIFICANCE: orthogonal eigenvectors mean the world decomposes into independent, non-interacting directions — normal modes of vibration, principal components of data, energy levels in quantum mechanics. The proof is three lines; the consequences fill physics departments.

Proof: the range of q on the unit circle

Expand $\mathbf{x}$ in the orthonormal eigenbasis: $\mathbf{x} = \sum c_i \mathbf{q}_i$ with $\sum c_i^2 = 1$ on the unit circle. Then $q(\mathbf{x}) = \sum \lambda_i c_i^2$ — a weighted AVERAGE of the eigenvalues with weights c_i^2.

An average lies between the extremes, hitting $\lambda_{\min}$ and $\lambda_{\max}$ when all weight sits on one eigenvector. ■ This one-line argument (the Rayleigh quotient) is how Google-scale eigenvalue problems are actually solved: maximize the quotient numerically.

9.3 Going Deeper: Worked Problems

Worked: rotating away a cross term

Identify the conic $5x^2 - 4xy + 8y^2 = 36$ by diagonalizing its quadratic form.

Step 1 — the symmetric matrix is $\begin{pmatrix} 5 & -2 \\ -2 & 8 \end{pmatrix}$ (half the cross term). Step 2 — eigenvalues: trace 13, det 36: $\lambda^2 - 13\lambda + 36 = (\lambda - 4)(\lambda - 9) = 0$. Step 3 — in the eigenbasis the equation becomes $4u^2 + 9v^2 = 36$, i.e. $\frac{u^2}{9} + \frac{v^2}{4} = 1$: an ELLIPSE with semi-axes 3 and 2. Step 4 — the axes' directions are the eigenvectors: for $\lambda = 4$, $(A - 4I)\mathbf{v} = 0$ gives $(2, 1)$ — the long axis points along $(2, 1)$. Both eigenvalues positive confirmed the ellipse before any drawing; a negative one would have meant a hyperbola.

Worked: the Hessian test as a definiteness check

Classify the critical point of $f(x, y) = x^2 + xy + y^2 - 3x$ at $(2, -1)$.

Step 1 — confirm it is critical: $f_x = 2x + y - 3 = 0$ and $f_y = x + 2y = 0$ at $(2, -1)$ ✓. Step 2 — the Hessian is the constant symmetric matrix $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Step 3 — leading minors: $2 > 0$ and $\det = 3 > 0$: POSITIVE DEFINITE, so the quadratic form near the point is a bowl and $(2, -1)$ is a local (in fact global) minimum, with $f = 4 - 2 + 1 - 6 = -3$. Step 4 — connect: the calculus "$D = f_{xx}f_{yy} - f_{xy}^2 = 3 > 0$, $f_{xx} > 0$" test is literally the leading-minor test for positive definiteness. Two courses, one criterion.

10 Singular Values & the SVD

Every matrix is rank-one layers, sorted by importance

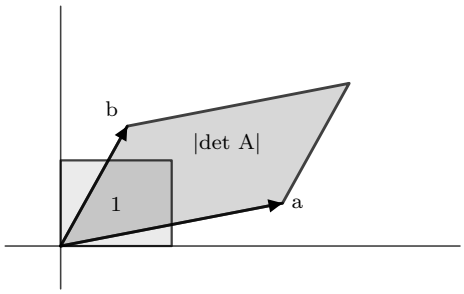
10.1 The Master Decomposition

Singular values: stretch factors for ANY matrix

Eigenvalues need square matrices and can be complex; SINGULAR VALUES exist for every

matrix and are always real and nonnegative: $\sigma_i = \sqrt{\lambda_i(A^\mathsf{T}A)}$ ($A^\mathsf{T}A$ is symmetric positive semidefinite, so this is legal).

Rotate, stretch along the axes, rotate again.



Meaning: σ_1 is the largest stretch A applies to any unit vector; the number of nonzero σ's is the RANK; and $\sum \sigma_i^2$ equals the sum of squares of all entries. The unit circle maps to an ellipse with semi-axes σ_1, σ_2 — every matrix, however ugly, is geometrically just rotate–stretch–rotate.

The SVD and best low-rank approximation

The singular value decomposition writes $A = \sum_i \sigma_i \, \mathbf{u}_i \mathbf{v}_i^\mathsf{T}$: a sum of RANK-ONE layers, ordered by importance $\sigma_1 \geq \sigma_2 \geq \cdots$. The Eckart–Young theorem: keeping just the first k layers gives the BEST possible rank-k approximation, with error exactly σ_{k+1}.

This is data compression as a theorem: an image is a matrix, most of its singular values are tiny, and storing a few layers reproduces it almost perfectly. Principal component analysis is the same truncation applied to centered data — "find the directions that matter, drop the rest."

Eigen vs. singular: when they agree

For symmetric positive-definite matrices, singular values ARE the eigenvalues; for the diagonal matrix $\mathrm{diag}(3, -4)$ they are the absolute values $3, 4$ — stretch has no sign. For a genuinely non-symmetric matrix the two sets differ, and the singular values are the honest geometry: a matrix with all eigenvalues zero (nilpotent) can still stretch vectors enormously, and its σ_1 says by how much.

Try it: a rank-one matrix bare-handed

Find the singular values of $A = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix}$. Work: rows are proportional — rank 1, so $\sigma_2 = 0$ and $\sigma_1^2 = \mathrm{tr}(A^\mathsf{T}A) = 4 + 16 + 1 + 4 = 25$: $\sigma_1 = 5$.

Answer. Structure check: $A = \binom{2}{1}(1;;2)$, and $\left|\binom{2}{1}\right| \cdot |(1,2)| = \sqrt{5} \cdot \sqrt{5} = 5$ ✓ — a rank-one matrix's only singular value is the product of the lengths of its two factors.

10.2 Proofs & Why It Matters

Proof: is the maximum stretch

Maximize $|A\mathbf{x}|^2 = \mathbf{x}^\mathsf{T}(A^\mathsf{T}A)\mathbf{x}$ over unit vectors $\mathbf{x}$. That is a quadratic form in the symmetric matrix $A^\mathsf{T}A$, and the previous topic proved its maximum on the unit sphere is the largest eigenvalue $\lambda_1(A^\mathsf{T}A)$.

Take square roots: $\max|A\mathbf{x}| = \sqrt{\lambda_1} = \sigma_1$. ■ SIGNIFICANCE: σ_1 is the operator norm — it bounds how much a matrix can amplify errors, which is why σ_1/σ_n (the condition number) decides whether a linear system is numerically trustworthy. Every numerical analyst's first question about a matrix is about its singular values.

Where you have already met the SVD

Recommendation engines factor the user-movie ratings matrix into a few rank-one "taste" layers; latent semantic analysis does it to word-document counts; noise reduction truncates small singular values because noise spreads evenly across all of them while signal concentrates in the top few. The SVD is the closest thing linear algebra has to a universal tool.

10.3 Going Deeper: Worked Problems

Worked: singular values of a non-square matrix

Find the singular values of $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ (3×2).

Step 1 — $A^\mathsf{T}A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ (a 2×2, so exactly two singular values). Step 2 — its eigenvalues: trace 4, det 3: $\lambda = 3, 1$. Step 3 — singular values $\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$. Step 4 — interpret: A maps the plane into 3-space, stretching by $\sqrt{3}$ along $(1,1)/\sqrt{2}$ and by 1 along $(1,-1)/\sqrt{2}$; the check $\sigma_1^2 + \sigma_2^2 = 4 =$ sum of squared entries $(1+0+0+1+1+1)$ ✓ confirms the arithmetic.

Worked: how much a rank-one approximation loses

For $A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$, find σ_1, σ_2 and the error of the best rank-one approximation.

Step 1 — $A^\mathsf{T}A = \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix}$. Step 2 — eigenvalues: trace 50, det 225: $\lambda^2 - 50\lambda + 225 = (\lambda - 45)(\lambda - 5)$. Step 3 — $\sigma_1 = \sqrt{45} = 3\sqrt{5} \approx 6.71$, $\sigma_2 = \sqrt{5} \approx 2.24$. Step 4 — by Eckart–Young the best rank-one approximation errs by exactly $\sigma_2 = \sqrt{5}$ in the spectral norm; the fraction of "energy" captured is $\dfrac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \dfrac{45}{50} = 90\%$. Sanity: $\sigma_1\sigma_2 = 15 = |\det A|$ ✓ — the product of singular values is always the absolute determinant.

11 Markov Chains & Long-Run Behavior

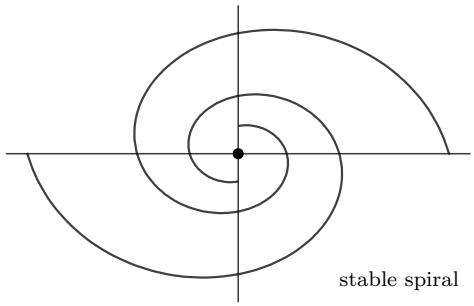
Steady states, mixing, and absorbing states

11.1 Probability Meets Eigenvectors

Transition matrices and the eigenvalue 1

A Markov chain hops among states with fixed probabilities; its transition matrix M has non-negative entries and columns summing to 1.

Every start converges to the steady state.



That column condition forces $\lambda = 1$ to be an eigenvalue (the all-ones vector is a LEFT eigenvector), and its right eigenvector — normalized to sum to 1 — is the STEADY STATE: the distribution the chain settles into regardless of where it started. Finding it is one homogeneous solve: $(M - I)\mathbf{v} = \mathbf{0}$. All other eigenvalues have $|\lambda| \leq 1$, and the second-largest, $|\lambda_2|$, sets the MIXING RATE: the memory of the starting state decays like $|\lambda_2|^n$.

PageRank is a steady state

Google's original algorithm models a random web surfer: from each page, follow a random link (with occasional random jumps). The importance of a page is its share of the STEADY-STATE distribution of that Markov chain — the $\lambda = 1$ eigenvector of a matrix with billions of rows, computed by simply iterating $\mathbf{v} \mapsto M\mathbf{v}$ (the power method) until it stops changing. A trillion-dollar company, built on this chapter.

Absorbing chains and expected times

Some states trap you (absorbing); the natural question becomes "how long until absorption?" First-step analysis: from state i, one step happens, then you are somewhere new — so $E_i = 1 + \sum_j p_{ij} E_j$, a small LINEAR SYSTEM in the expected times.

A state that advances with probability p and stalls otherwise contributes $\frac{1}{p}$ expected steps (the geometric distribution). Gambler's ruin, board-game durations, drug absorption, and queue lengths are all this one linear system with different numbers.

Try it: steady state and mixing in one go

For $M = \begin{pmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{pmatrix}$: (a) steady state: $-0.1v_1 + 0.2v_2 = 0$ gives $v_1 : v_2 = 2 : 1$, so $\left(\frac{2}{3}, \frac{1}{3}\right)$; (b) mixing rate: trace 1.7 minus the known eigenvalue 1 leaves $\lambda_2 = 0.7$ — deviations shrink 30% per step, so after 10 steps only $(0.7)^{10} \approx 3\%$ of any initial imbalance survives. Verify (a) by one multiplication: $M\binom{2/3}{1/3} = \binom{2/3}{1/3}$ ✓.

11.2 Proofs & Why It Matters

Proof: $= 1$ is always an eigenvalue

Let $\mathbf{1} = (1, 1, \ldots, 1)$. Column sums equal 1 means $\mathbf{1}^\mathsf{T} M = \mathbf{1}^\mathsf{T}$ — so 1 is an eigenvalue of M^T, and a matrix and its transpose share a characteristic polynomial, hence eigenvalues: 1 is an eigenvalue of M too.

Moreover no eigenvalue can exceed 1 in magnitude: if $M\mathbf{v} = \lambda\mathbf{v}$, comparing the largest-magnitude entry of both sides shows $|\lambda| \leq 1$ (each entry of $M\mathbf{v}$ is a weighted average of entries of $\mathbf{v}$). ■ SIGNIFICANCE: existence of a steady state is thus STRUCTURAL — probability is conserved, so equilibrium is guaranteed; the only question is how fast you get there, which is $|\lambda_2|$'s job.

The bridge back

Markov chains tie the whole course together: the steady state is an eigenvector (eigenvalues topic), computing M^n is diagonalization (powers topic), the entries of M^n converge because $|\lambda_2| < 1$ (spectral thinking), and expected absorption times are a linear system (row reduction). One applied object exercising every tool in the box — which is exactly why it closes the course.

11.3 Going Deeper: Worked Problems

Worked: a three-state weather chain, start to steady state

Sunny goes to sunny with probability 0.7, else rainy; rainy goes to rainy 0.5, sunny 0.3, cloudy 0.2; cloudy goes to sunny 0.4, cloudy 0.6. Find the long-run fraction of sunny days.

Step 1 — transition matrix (columns from-state, rows to-state): $M = \begin{pmatrix} 0.7 & 0.3 & 0.4 \\ 0.3 & 0.5 & 0 \\ 0 & 0.2 & 0.6 \end{pmatrix}$; each column sums to 1 ✓. Step 2 — steady state solves $M\mathbf{v} = \mathbf{v}$: from row 2, $0.3s + 0.5r = r$ gives $r = 0.6s$; from row 3, $0.2r + 0.6c = c$ gives $c = 0.5r = 0.3s$. Step 3 — normalize: $s + 0.6s + 0.3s = 1$, so $s = \dfrac{1}{1.9} = \dfrac{10}{19} \approx 52.6\%$ sunny, $r = \frac{6}{19}$, $c = \frac{3}{19}$. Step 4 — verify row 1: $0.7 \cdot \frac{10}{19} + 0.3 \cdot \frac{6}{19} + 0.4 \cdot \frac{3}{19} = \frac{7+1.8+1.2}{19} = \frac{10}{19}$ ✓. The forecast for a year from now needs no calendar — only this eigenvector.

Worked: expected time to absorb, by first-step equations

A token starts at 0 and each second moves +1 with probability $\frac{2}{3}$ or stays with probability $\frac{1}{3}$; it stops on reaching 3. Expected time to stop?

Step 1 — let E_k be the expected remaining time from position k, with $E_3 = 0$. Step 2 — first-step analysis at each state: $E_k = 1 + \frac{2}{3}E_{k+1} + \frac{1}{3}E_k$, so $\frac{2}{3}E_k = 1 + \frac{2}{3}E_{k+1}$, i.e. $E_k = \frac{3}{2} + E_{k+1}$. Step 3 — unwind from the end: $E_2 = 1.5$, $E_1 = 3$, $E_0 = 4.5$. Step 4 — interpret: each step forward takes $\frac{1}{2/3} = 1.5$ expected seconds (geometric distribution), and three steps are needed — linearity of expectation gives 4.5 directly, and the linear system confirms it.