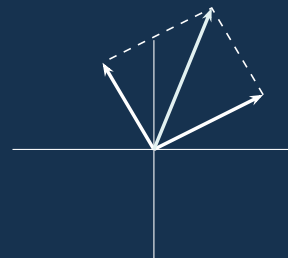


Linear Algebra

Comprehensive Reference Notes



SCS 1306 | University of Colombo School of Computing

Course	SCS 1306 — Linear Algebra (2L + 1P, 3 credits)
Lecturer	Dr. Randil Pushpananda
Primary text	G. Strang, <i>Introduction to Linear Algebra</i> , 6th edition (2022)
Also referenced	Friedberg, Insel & Spence, <i>Linear Algebra</i> (2018); J. Hefferon, <i>Linear Algebra</i>
Assessment	80% final paper, 20% in-class assignments

How to use these notes. Each chapter follows the same rhythm: an overview, the definitions and theorems in coloured boxes, geometric intuition supported by diagrams, fully worked examples drawn from the tutorial sheets, and an *Exam Preparation* panel summarising formulas, pitfalls and self-check questions. Green side-boxes point to the corresponding sections of Strang's 6th edition so you can read further at any point.

Compiled as a study reference from the SCS 1306 lecture slides, tutorial sheets 01–06, Assignment 01, and the recommended textbook.

Preface and Course Map

These notes cover the whole of SCS 1306 in a single reference document. They were assembled from the lecture slides, the six tutorial sheets, the in-class assignment, and Gilbert Strang's *Introduction to Linear Algebra* (6th edition), which is the recommended text for the course.

The one idea the course is built on

Almost everything in a first linear algebra course is an answer to one question:

**What can I build out of a fixed set of vectors,
and which targets are out of reach?**

Written algebraically, “what can I build” is the set of all linear combinations $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$; written in matrix form it is $A\mathbf{x}$; written as a question it is $A\mathbf{x} = \mathbf{b}$. Every major topic is a different lens on that same equation:

Topic	What it says about $A\mathbf{x} = \mathbf{b}$
Elimination	A reliable procedure for <i>solving</i> it.
Inverses	When A can be undone, $\mathbf{x} = A^{-1}\mathbf{b}$.
Determinants	A single number that detects when it breaks.
Vector spaces	The exact set of reachable $\mathbf{b}$, and the ambiguity in $\mathbf{x}$.
Orthogonality	What to do when there is no exact solution.
Eigenvalues	Directions in which A acts as mere scaling.
SVD / PCA	The best possible coordinates for <i>any</i> A .
Transformations	The geometric meaning of multiplying by A .

Course outline and where it lives in these notes

	Syllabus topic	Chapters here	Strang 6th ed.
1	Linear systems	1–5	1.1–1.4, 2.1–2.5
2	Vector spaces and subspaces	8–9	3.1–3.5
3	Orthogonality	10	4.1–4.4
4	Determinants	6–7	5.1–5.3
5	Eigenvalues and eigenvectors	11–12	6.1–6.5
6	Singular Value Decomposition	13–14	7.1–7.4
7	Linear transformations	15	8.1–8.3
8	Applications	16	Ch. 9–10

Remark • on the missing Lecture 2

Lecture 2 was not available when these notes were compiled. Judging from Tutorial 01 and Tutorial 02 — both of which are drawn almost verbatim from Strang §1.1–§1.3 — it covered vectors, linear combinations, lengths and dot products, and the beginnings of matrix notation. That material is reconstructed in full in Chapters 1 and 3 from the textbook, so nothing in the syllabus is left uncovered.

How the boxes are colour-coded

Definition 0.1 Sample

Blue boxes state definitions. These are the things you must be able to recite precisely; exam questions frequently begin “State the definition of...”.

Theorem 0.1 Sample

Teal boxes hold theorems, propositions, lemmas and corollaries — the results you are allowed to *use*, usually with a proof following.

Worked Example 0.1 • Sample

Gold boxes are fully worked examples. Most are taken directly from the tutorial sheets so the arithmetic matches what you have already seen.

Intuition

Purple boxes explain *why* a result is true or what it looks like geometrically, without the formalism.

Common Pitfall

Rust boxes flag the specific mistakes that cost marks: sign errors in cofactors, forgetting $\det(A + B) \neq \det A + \det B$, and so on.

Method Sample

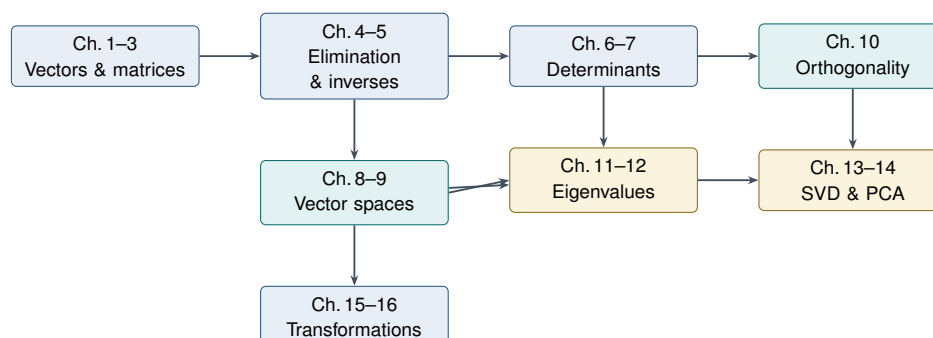
Green boxes are step-by-step procedures — the algorithm to follow when you are asked to *compute* something.

White boxes with a green edge point to the matching section of the textbook, e.g. §2.2, pp. 45–52. STRANG | 6th ed.

Each chapter closes with an **Exam Preparation** panel: the formulas worth memorising, the pitfalls specific to that chapter, and a handful of self-check questions with answers.

A suggested study order

If you are revising under time pressure, the dependency structure is:



Chapters 1–5 are prerequisites for everything. Determinants (6–7) and vector spaces (8–9) can be read in either order. Eigenvalues need both. SVD needs eigenvalues *and* orthogonality.

Contents

Preface and Course Map	i
Notation	x
I Vectors, Matrices and Linear Systems	1
1 Vectors and Linear Combinations	2
1.1 What a vector is	2
1.1.1 The two operations	3
1.2 Linear combinations	4
1.2.1 The geometry of “all combinations”	4
1.2.2 Independence, informally	6
1.3 Lengths, dot products and angles	6
1.3.1 Unit vectors and the standard basis	8
1.4 Vectors in three dimensions and beyond	9
1.5 Exam preparation	10
2 Systems of Linear Equations	11
2.1 Linear equations	11
2.2 The row picture	12
2.2.1 Two equations in two unknowns	12
2.2.2 Three equations in three unknowns	13
2.3 The column picture	14
2.4 Matrix form	16
2.5 Square systems and what can go wrong	18
2.6 Exam preparation	19
3 Matrix Algebra	20
3.1 Matrices, sizes and equality	20
3.2 Addition and scalar multiplication	20
3.3 Matrix multiplication	21
3.3.1 Four ways to see the same product	22
3.3.2 The rules — and the one that fails	23
3.4 Special matrices	24
3.5 The trace	25

3.6	Block multiplication	27
3.7	Exam preparation	27
4	Elimination and Back Substitution	29
4.1	The idea	29
4.2	Gaussian elimination	30
4.3	Elimination as matrix multiplication	31
4.4	When elimination breaks down	33
4.5	Echelon forms	34
4.6	Gaussian versus Gauss–Jordan	35
4.7	Systems with parameters	36
4.8	Exam preparation	38
5	Inverse Matrices	40
5.1	Definition and uniqueness	40
5.2	The 2×2 formula	41
5.3	Gauss–Jordan: computing A^{-1}	41
5.4	Rules for inverses	43
5.5	Triangular and diagonal inverses	44
5.6	The Invertible Matrix Theorem	45
5.7	Why not to use A^{-1}	46
5.8	Exam preparation	46
II	Determinants	48
6	The Transpose and Symmetric Matrices	49
6.1	The transpose	49
6.1.1	The transpose and the dot product	50
6.2	Symmetric matrices	51
6.3	Permutation matrices	53
6.4	Exam preparation	54
7	Determinants	56
7.1	What the determinant is for	56
7.2	Small determinants	57
7.3	Minors, cofactors and expansion	58
7.4	Rules for determinants	60
7.4.1	Determinants by elimination	61
7.5	The adjugate and a formula for A^{-1}	62

7.6	Cramer's rule	64
7.7	Determinants and singularity with a parameter	65
7.8	Exam preparation	67
III Vector Spaces and the Four Subspaces		69
8	Vector Spaces and Subspaces	70
8.1	The axioms	70
8.1.1	Examples — and non-examples	71
8.2	Subspaces	71
8.3	Span	73
8.4	Linear independence	73
8.5	Basis and dimension	74
8.6	Exam preparation	76
9	Column Space, Null Space and the Four Subspaces	78
9.1	The column space	78
9.2	The null space	79
9.2.1	Computing the null space	79
9.2.2	The complete solution of $Ax = b$	80
9.3	The rank	81
9.4	The four fundamental subspaces	83
9.5	Rank, solvability and the shape of the answer	85
9.6	Exam preparation	86
IV Orthogonality		87
10	Orthogonality, Projections and Least Squares	88
10.1	Orthogonal vectors	88
10.2	Orthogonal complements	89
10.3	Orthonormal sets and orthogonal matrices	90
10.4	Projections	92
10.5	Least squares	93
10.6	Gram–Schmidt and QR	94
10.7	Exam preparation	96
V Eigenvalues, Eigenvectors and Decompositions		97
11	Eigenvalues and Eigenvectors	98
11.1	The definition	98

11.2	Finding eigenvalues	99
11.3	Properties of eigenvalues	102
11.4	Eigenspaces are subspaces	104
11.5	Special matrices and their eigenvalues	105
11.6	Similar matrices	106
11.7	Exam preparation	107
12	Diagonalization and the Spectral Theorem	108
12.1	The eigen-decomposition	108
12.2	When is a matrix diagonalizable?	110
12.3	Symmetric matrices: the spectral theorem	111
12.4	Positive definite matrices	112
12.5	An application: Markov chains	113
12.6	Exam preparation	114
13	The Singular Value Decomposition	116
13.1	The statement	116
13.2	Why it works	117
13.3	Computing an SVD	118
13.4	The SVD and the four subspaces	120
13.5	Low-rank approximation	121
13.6	The pseudoinverse	122
13.7	PCA versus SVD	122
13.8	Exam preparation	123
14	Principal Component Analysis	124
14.1	The problem PCA solves	124
14.2	Centring and the covariance matrix	125
14.3	The PCA procedure	126
14.4	Choosing how many components	128
14.5	What PCA is used for	128
14.6	Exam preparation	130
VI	Linear Transformations and Applications	131
15	Linear Transformations, Translations and Affine Maps	132
15.1	Linear transformations	132
15.2	Geometric transformations in 2D	134
15.2.1	Rotation	134

15.2.2	Reflection in an arbitrary line	135
15.2.3	Shear	136
15.3	Geometric transformations in 3D	137
15.4	Translation and why it breaks linearity	138
15.5	Homogeneous coordinates	138
15.6	Composing transformations	140
15.7	Change of basis	141
15.8	Exam preparation	142
16	Applications of Linear Algebra	144
16.1	Computer graphics	144
16.2	Image compression	145
16.3	Machine learning and data science	146
16.4	Networks and graphs	147
16.5	Cryptography	148
16.6	Differential equations	149
16.7	Numerical computing with Octave	149
16.8	Exam preparation	150
VII	Appendices	152
A	Worked Tutorial Solutions	153
A.1	Tutorial 01 — Matrix arithmetic	153
A.2	Tutorial 02 — Vectors and combinations (Strang §1.1–1.2)	155
A.3	Tutorial 04 — Inverses, rank and elimination	156
A.4	Tutorial 05 — Determinants	159
A.5	Tutorial 06 — Eigenvalues (Part A)	161
B	Assignment 01 — Full Solutions	163
B.1	Question 1 (40 marks)	163
B.2	Question 2 (30 marks)	166
B.3	Question 3 (30 marks)	167
B.4	Marks summary and common errors	170
C	Formula Reference Sheet	171
C.1	Vectors	171
C.2	Matrix algebra	171
C.3	Elimination and factorization	172
C.4	Determinants	172

C.5	Vector spaces and the four subspaces	172
C.6	Orthogonality and least squares	173
C.7	Eigenvalues and diagonalization	173
C.8	SVD and PCA	174
C.9	Transformations	174
C.10	The Invertible Matrix Theorem	174
C.11	Decision table: which method?	175
C.12	Twelve mistakes that cost marks	175
C.13	Greek letters used in this course	176
Index		177

Notation

Notation in linear algebra is not fully standardised. The conventions below are used consistently throughout these notes and follow Strang where possible.

Objects

$\mathbb{R}, \mathbb{C}$	the real numbers, the complex numbers
$\mathbb{R}^n$	the set of column vectors with n real entries
$\mathbb{R}^{m \times n}$	the set of real matrices with m rows and n columns
v, x, b	vectors — bold lower case , always <i>column</i> vectors
A, B, Q, Λ	matrices — upper case
c, d, λ, α	scalars — lower case italic or Greek
$\mathbf{0}$	the zero vector; O or 0 the zero matrix
I or I_n	the $n \times n$ identity matrix
$e_1, \dots, e_n$	the standard basis vectors (columns of I)

Entries and pieces

a_{ij}	the entry of A in row i , column j
a_j	the j th <i>column</i> of A
$A = [a_1 \ a_2 \ \cdots \ a_n]$	A displayed by columns
v_i	the i th component of the vector v

Operations

A^T	transpose of A ; $(A^T)_{ij} = a_{ji}$
A^{-1}	inverse of A (exists only if A is square and nonsingular)
A^{-T}	shorthand for $(A^{-1})^T = (A^T)^{-1}$
$x \cdot y$ or $x^T y$	dot (inner) product
$\ x\ $	length (Euclidean norm) $\sqrt{x^T x}$
$\det A$ or $ A $	determinant of A
$\text{tr } A$	trace of A , the sum of the diagonal entries
$\text{rank } A$	rank of A
$\text{adj } A$	adjugate (classical adjoint) of A
M_{ij}	the (i, j) minor of A
C_{ij}	the (i, j) cofactor, $C_{ij} = (-1)^{i+j} M_{ij}$

Subspaces

$\mathbf{C}(A)$	column space of A — a subspace of $\mathbb{R}^m$
$\mathbf{N}(A)$	null space of A — a subspace of $\mathbb{R}^n$
$\mathbf{C}(A^T)$	row space of A — a subspace of $\mathbb{R}^n$
$\mathbf{N}(A^T)$	left null space of A — a subspace of $\mathbb{R}^m$
$\text{span}\{v_1, \dots, v_k\}$	the set of all linear combinations of those vectors
$V^\perp$	the orthogonal complement of the subspace V
$\dim V$	dimension of V (number of vectors in a basis)

Row operations

$R_i \leftrightarrow R_j$	exchange rows i and j
$R_i \rightarrow cR_i$	multiply row i by the nonzero scalar c
$R_i \rightarrow R_i - \ell R_j$	subtract ℓ times row j from row i
ℓ_{ij}	the multiplier used to clear the (i, j) entry

Decompositions

$A = LU$	lower triangular times upper triangular
$PA = LU$	the same with a row-exchange (permutation) matrix P
$A = QR$	orthonormal Q times upper triangular R
$A = X\Lambda X^{-1}$	eigen-decomposition (diagonalization)
$S = Q\Lambda Q^T$	spectral decomposition of a symmetric matrix
$A = U\Sigma V^T$	singular value decomposition

Remark • Columns, not rows

Every vector in these notes is a column. When a row is needed it is written as a transpose, e.g. $v^T = [1 \ 2 \ 3]$. Writing $v = (1, 2, 3)$ in running text is an abbreviation for the column $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ — this is Strang’s convention too, and it saves a great deal of vertical space.

Common Pitfall • with the word “adjoint”

In this course $\text{adj } A$ means the *adjugate*: the transpose of the cofactor matrix, used in the formula $A^{-1} = \frac{1}{\det A} \text{adj } A$. In more advanced courses “adjoint” means the conjugate transpose A^* , which is a completely different object. Tutorial 05 uses the first sense.

PART I

Vectors, Matrices and Linear Systems

Vectors and Linear Combinations

This chapter builds the vocabulary the rest of the course speaks in. A vector is a list of numbers with two operations attached — addition and scalar multiplication — and *every* construction in linear algebra is assembled from just those two. Combining them gives the linear combination $cv + dw$, and the set of all such combinations is the single most important object in the subject. We finish with lengths, angles and the dot product, which is what lets us measure inside a vector space.

Strang, *Introduction to Linear Algebra*, 6th ed., §1.1 (“Vectors and Linear Combinations”) and §1.2 (“Lengths and Angles from Dot Products”). ^{STRANG | 6th ed.} Tutorial 02 is essentially the problem set of §1.1.

1.1 What a vector is

Definition 1.1 Column vector

A **vector** in $\mathbb{R}^n$ is an ordered list of n real numbers written as a column:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n.$$

The numbers $v_1, \dots, v_n$ are the **components** of v . Two vectors are equal exactly when all corresponding components are equal — so equality in $\mathbb{R}^n$ is really n separate equations.

A vector carries three simultaneous readings, and fluency means switching between them without thinking:

Reading	What $v = (3, 2)$ means	When it helps
Algebraic	the pair of numbers 3 and 2	computation, code
Point	the location $(3, 2)$ in the plane	describing solution sets
Arrow	displacement 3 right, 2 up	addition, geometry, physics

1.1.1 The two operations

Definition 1.2 Vector addition and scalar multiplication

For $v, w \in \mathbb{R}^n$ and a scalar $c \in \mathbb{R}$:

$$v + w = \begin{bmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{bmatrix}, \quad cv = \begin{bmatrix} cv_1 \\ \vdots \\ cv_n \end{bmatrix}.$$

Both operations act *componentwise*. Note that $v + w$ is only defined when v and w have the *same* number of components.

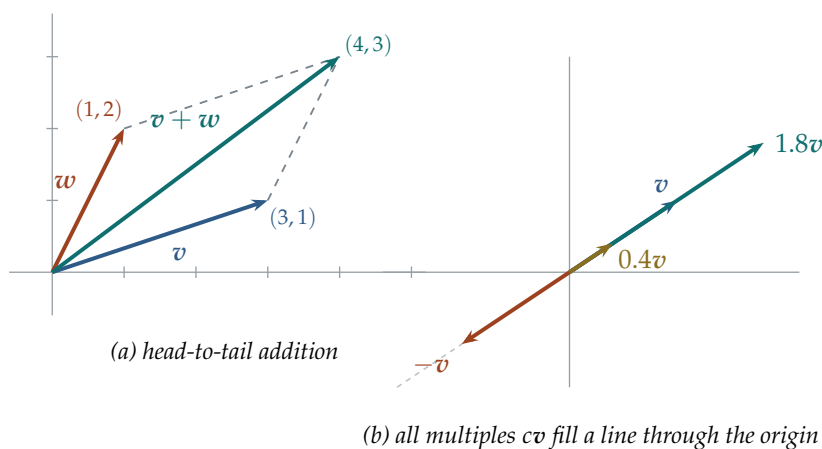


Figure 1.1: The two operations. Addition completes a parallelogram; scalar multiplication slides a vector along a fixed line through the origin.

Intuition • why the parallelogram

Adding w to v means “start at the tip of v and apply the displacement w ”. Doing it in the other order — apply w first, then v — lands in the same place, which is exactly why $v + w = w + v$ and why the two routes bound a parallelogram.

Proposition 1.1 Algebraic rules

For all $u, v, w \in \mathbb{R}^n$ and all scalars $c, d \in \mathbb{R}$:

- | | |
|----------------------------------|---------------------------|
| (V1) $v + w = w + v$ | (V5) $c(v + w) = cv + cw$ |
| (V2) $(u + v) + w = u + (v + w)$ | (V6) $(c + d)v = cv + dv$ |
| (V3) $v + \mathbf{0} = v$ | (V7) $(cd)v = c(dv)$ |
| (V4) $v + (-v) = \mathbf{0}$ | (V8) $1v = v$ |

Proof. Each rule is the corresponding rule for real numbers applied in every component simultaneously. For (V6), the i th component of the left side is $(c + d)v_i$ and of the right side is $cv_i + dv_i$; these agree by the distributive law in $\mathbb{R}$. Since i was arbitrary, the vectors are equal. ■

Remark • these eight rules are the whole story

Chapter 8 turns this proposition on its head: instead of *deriving* the eight rules for $\mathbb{R}^n$, we will *demand* them of any set whatever, and call such a set a vector space. Everything proved using only (V1)–(V8) then applies to polynomials, functions and matrices without a single extra line of work.

1.2 Linear combinations

Definition 1.3 Linear combination

Given vectors $v_1, \dots, v_k \in \mathbb{R}^n$ and scalars $c_1, \dots, c_k \in \mathbb{R}$, the vector

$$c_1 v_1 + c_2 v_2 + \dots + c_k v_k$$

is a **linear combination** of $v_1, \dots, v_k$. The scalars c_i are the **coefficients** or **weights**.

This is the central construction of the entire subject. Every question you will meet is a question about linear combinations wearing a disguise:

Question	Really asking
Does $Ax = b$ have a solution?	Is b a linear combination of the columns of A ?
Is this set of vectors independent?	Is 0 a linear combination in a non-trivial way?
What is the column space?	What is the set of <i>all</i> linear combinations?
What is a basis?	Which minimal set has all others as combinations?

1.2.1 The geometry of “all combinations”

Theorem 1.1 Shape of the set of all combinations

Let S be the set of all linear combinations of given vectors in $\mathbb{R}^3$. Then S is always one of exactly four things, and it always contains 0 :

S	Occurs when	Example
$\{0\}$	all the vectors are 0	$c0$
a line through 0	all nonzero vectors are parallel	$c(1, 2, 3) + d(3, 6, 9)$
a plane through 0	exactly two independent directions	$c(1, 0, 0) + d(0, 2, 3)$
all of $\mathbb{R}^3$	three independent directions	$ce_1 + de_2 + ee_3$

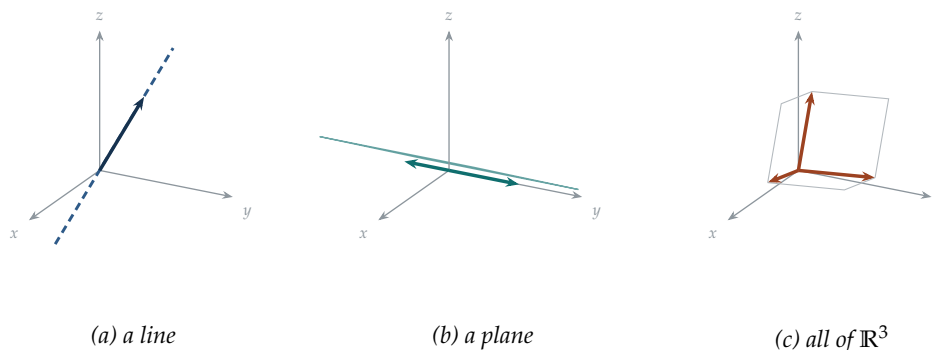


Figure 1.2: The three nontrivial possibilities for the set of all linear combinations in $\mathbb{R}^3$. Which one occurs is decided entirely by how many independent directions the vectors supply — this count is the rank, introduced in Chapter 9.

Worked Example 1.1 • Tutorial 02, Q3 — classifying spans

Describe geometrically all linear combinations of:

(a) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$

(b) $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$

(c) $\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$

(a) A line. The second vector is 3 times the first: $(3, 6, 9) = 3(1, 2, 3)$. So every combination $c(1, 2, 3) + d(3, 6, 9) = (c + 3d)(1, 2, 3)$ is just a multiple of the single vector $(1, 2, 3)$. The set is the line through the origin in the direction $(1, 2, 3)$.

(b) A plane. Neither vector is a multiple of the other (the first has a nonzero first component, the second has zero there). Two independent directions sweep out a plane through the origin. To name that plane, find a normal $\mathbf{n}$ with $\mathbf{n} \cdot (1, 0, 0) = 0$ and $\mathbf{n} \cdot (0, 2, 3) = 0$: the first gives $n_1 = 0$, the second gives $2n_2 + 3n_3 = 0$, so $\mathbf{n} = (0, 3, -2)$ works. The plane is $3y - 2z = 0$.

(c) All of $\mathbb{R}^3$. Check whether the third vector is a combination of the first two. Solve $c(2, 0, 0) + d(0, 2, 2) = (2, 2, 3)$: from the second component $2d = 2$ so $d = 1$; from the third component $2d = 3$ so $d = \frac{3}{2}$. Contradiction — the third vector is *not* in the plane of the first two. Three independent directions in $\mathbb{R}^3$ produce all of $\mathbb{R}^3$.

Common Pitfall • the set of combinations always contains the origin

Taking $c_1 = \cdots = c_k = 0$ gives $\mathbf{0}$, so the set of all linear combinations can never be a line or plane that misses the origin. This is exactly why translation is *not* a linear operation (Chapter 15) and why the solution set of $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} \neq \mathbf{0}$ is not a subspace.

1.2.2 Independence, informally

Definition 1.4 Linear independence — first pass

Vectors $v_1, \dots, v_k$ are **linearly independent** if none of them is a linear combination of the others. Equivalently,

$$c_1 v_1 + \dots + c_k v_k = \mathbf{0} \implies c_1 = c_2 = \dots = c_k = 0.$$

If some non-trivial choice of coefficients does give $\mathbf{0}$, the vectors are **linearly dependent**.

The second form is the one to use in proofs: it converts a statement about “none of them is” (which would need k separate checks) into a single homogeneous system. Chapter 8 develops this properly.

Worked Example 1.2 • Tutorial 02, Q10 — an independence test with a parameter

Decide whether $u = (1, 1, 0)$, $v = (0, 1, 1)$ and $w = (a, b, c)$ are independent or dependent.

Set $xu + yv + zw = \mathbf{0}$ and read off the components:

$$x + az = 0, \quad x + y + bz = 0, \quad y + cz = 0.$$

From the first and third, $x = -az$ and $y = -cz$. Substituting into the second gives $-az - cz + bz = 0$, i.e. $z(b - a - c) = 0$.

- If $b - a - c \neq 0$ the only possibility is $z = 0$, which forces $x = y = 0$: the vectors are **independent**.
- If $b = a + c$ we may take $z = 1$, giving $x = -a$, $y = -c$ — a non-trivial combination equal to $\mathbf{0}$. The vectors are **dependent**, and indeed $w = au + cv$.

So: independent precisely when $b \neq a + c$. Geometrically, u and v span the plane $x - y + z = 0$; w lies in that plane exactly when $a - b + c = 0$.

1.3 Lengths, dot products and angles

Addition and scalar multiplication cannot by themselves tell you that one vector is longer than another, or that two vectors are perpendicular. For that we need one more operation.

Definition 1.5 Dot product

For $v, w \in \mathbb{R}^n$,

$$v \cdot w = v^T w = v_1 w_1 + v_2 w_2 + \dots + v_n w_n = \sum_{i=1}^n v_i w_i.$$

The result is a *scalar*, not a vector.

Definition 1.6 Length / Euclidean norm

The **length** (or norm, or magnitude) of $v \in \mathbb{R}^n$ is

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + \cdots + v_n^2}, \quad \text{so} \quad \|v\|^2 = v \cdot v.$$

A vector with $\|v\| = 1$ is a **unit vector**. For any $v \neq 0$, the vector $u = v / \|v\|$ is the unit vector in the direction of v ; forming it is called **normalizing**.

Proposition 1.2 Properties of the dot product

For all $u, v, w \in \mathbb{R}^n$ and $c \in \mathbb{R}$:

$$\begin{aligned} \text{(symmetry)} \quad & v \cdot w = w \cdot v \\ \text{(linearity)} \quad & (u + v) \cdot w = u \cdot w + v \cdot w, \quad (cv) \cdot w = c(v \cdot w) \\ \text{(positivity)} \quad & v \cdot v \geq 0, \text{ with equality iff } v = 0 \end{aligned}$$

Theorem 1.2 The angle formula

For nonzero $v, w \in \mathbb{R}^n$, the angle $\theta \in [0, \pi]$ between them satisfies

$$\cos \theta = \frac{v \cdot w}{\|v\| \|w\|} \quad \text{equivalently} \quad v \cdot w = \|v\| \|w\| \cos \theta.$$

In particular $v \cdot w = 0$ exactly when $\theta = 90^\circ$: the vectors are **orthogonal** (perpendicular).

Sketch, via the law of cosines. The triangle with sides v, w and $v - w$ gives $\|v - w\|^2 = \|v\|^2 + \|w\|^2 - 2 \|v\| \|w\| \cos \theta$. Expanding the left side algebraically,

$$\|v - w\|^2 = (v - w) \cdot (v - w) = \|v\|^2 - 2 v \cdot w + \|w\|^2.$$

Comparing the two expressions and cancelling $\|v\|^2 + \|w\|^2$ leaves $-2 v \cdot w = -2 \|v\| \|w\| \cos \theta$. ■

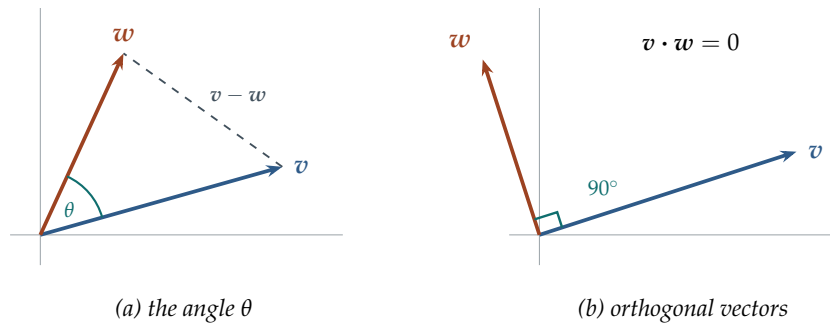


Figure 1.3: The dot product measures alignment. It is positive for an acute angle, zero at 90° , and negative for an obtuse angle.

Theorem 1.3 Cauchy–Schwarz and the triangle inequality

For all $v, w \in \mathbb{R}^n$:

$$|v \cdot w| \leq \|v\| \|w\| \quad (\text{Cauchy–Schwarz})$$

$$\|v + w\| \leq \|v\| + \|w\| \quad (\text{triangle inequality})$$

Equality holds in Cauchy–Schwarz exactly when v and w are parallel.

Proof. Cauchy–Schwarz is immediate from the angle formula since $|\cos \theta| \leq 1$. For the triangle inequality, square the left side:

$$\|v + w\|^2 = \|v\|^2 + 2v \cdot w + \|w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2,$$

using Cauchy–Schwarz in the middle. Taking square roots (both sides are non-negative) finishes it. ■

Worked Example 1.3 • Tutorial 02, Q2 — a triangle of vectors

Going around a triangle from $(0,0)$ to $(5,0)$ to $(0,12)$ and back to $(0,0)$, what are the three vectors u, v, w ? What is $u + v + w$? What are their lengths?

Each side is (head) – (tail):

$$u = \begin{bmatrix} 5 \\ 0 \end{bmatrix}, \quad v = \begin{bmatrix} 0 - 5 \\ 12 - 0 \end{bmatrix} = \begin{bmatrix} -5 \\ 12 \end{bmatrix}, \quad w = \begin{bmatrix} 0 - 0 \\ 0 - 12 \end{bmatrix} = \begin{bmatrix} 0 \\ -12 \end{bmatrix}.$$

Their sum is $u + v + w = \begin{bmatrix} 5-5+0 \\ 0+12-12 \end{bmatrix} = \mathbf{0}$, which must happen: a closed path returns you to where you started, so the total displacement is zero.

Lengths: $\|u\| = \sqrt{25 + 0} = 5$, $\|v\| = \sqrt{25 + 144} = \sqrt{169} = 13$, $\|w\| = \sqrt{0 + 144} = 12$. The familiar 5–12–13 right triangle. Consistently, $u \cdot w = 5 \cdot 0 + 0 \cdot (-12) = 0$, so the sides u and w meet at a right angle.

Worked Example 1.4 • Assignment 01, Q2(c) — the angle between two vectors

With $a = (1, 0, -1)$ and $b = (0, 2, 1)$, find the angle between them.

$$a \cdot b = 1 \cdot 0 + 0 \cdot 2 + (-1) \cdot 1 = -1, \quad \|a\| = \sqrt{2}, \quad \|b\| = \sqrt{5}.$$

Hence

$$\cos \theta = \frac{-1}{\sqrt{2}\sqrt{5}} = \frac{-1}{\sqrt{10}} \approx -0.3162, \quad \theta = \arccos(-0.3162) \approx 108.4^\circ.$$

The negative dot product tells you immediately, before any arithmetic, that the angle is obtuse.

1.3.1 Unit vectors and the standard basis

Definition 1.7 Standard basis vectors

In $\mathbb{R}^n$ the vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

are the **standard basis** of $\mathbb{R}^n$. In $\mathbb{R}^3$ they are often written i, j, k . They are mutually orthogonal unit vectors, and every $v \in \mathbb{R}^n$ has the unique expression $v = v_1 e_1 + \cdots + v_n e_n$.

That last sentence is worth pausing on: the components of a vector *are* its coefficients in the standard basis. Most of the second half of this course is about choosing a *better* basis — eigenvectors in Chapter 11, singular vectors in Chapter 13 — in which the coefficients say something more useful.

Worked Example 1.5 • Tutorial 02, Q14 — the clock vectors

Twelve vectors point from the centre of a clock to the hours 1:00, 2:00, ..., 12:00, each of length 1.

- (a) What is their sum V ?
- (b) If the 2:00 vector is removed, why do the remaining 11 add to the 8:00 vector?
- (c) The 2:00 vector is $(\cos \theta, \sin \theta)$. What is θ ?

(a) $V = \mathbf{0}$. The twelve vectors come in six opposite pairs: 12:00 with 6:00, 1:00 with 7:00, and so on. Each pair sums to $\mathbf{0}$, so the total is $\mathbf{0}$.
(b) If the sum of all twelve is $\mathbf{0}$, then removing one vector c_2 leaves $\mathbf{0} - c_2 = -c_2$. And $-c_2$ is the vector pointing in the opposite direction to 2:00, which on a clock face is 8:00.
(c) Measuring θ counter-clockwise from the positive x -axis (the 3:00 direction), each hour is 30° and the hands move clockwise. The 2:00 position is one hour *before* 3:00, so $\theta = 30^\circ = \pi/6$, giving $c_2 = (\cos 30^\circ, \sin 30^\circ) = (\sqrt{3}/2, 1/2)$.

1.4 Vectors in three dimensions and beyond

Everything above was written for $\mathbb{R}^n$, and none of the algebra changes when $n > 3$. What is lost is the picture; what is gained is that the same tools describe a 1000-dimensional data record as easily as a point in space.

Worked Example 1.6 • Tutorial 02, Q24 — a cube in four dimensions

How many corners $(\pm 1, \pm 1, \pm 1, \pm 1)$ does a cube of side 2 have in four dimensions? What is its volume? How many 3D faces? How many edges?

Corners. Each of the 4 coordinates independently takes one of 2 values, so there are $2^4 = 16$ corners.

Volume. Side length 2 in each of 4 dimensions: $2^4 = 16$.

3D faces. A face is obtained by freezing one coordinate at $+1$ or -1 : 4 coordinates $\times$ 2 choices = 8 three-dimensional faces (each of which is an ordinary cube).

Edges. An edge freezes 3 coordinates and lets 1 vary. Choose which coordinate varies (4 ways) and fix the other three ($2^3 = 8$ ways): $4 \times 8 = 32$ edges. One edge runs from $(-1, 1, 1, 1)$ to $(1, 1, 1, 1)$.

The general pattern for the n -cube: 2^n corners, $2n$ faces of dimension $n - 1$, and $n 2^{n-1}$ edges.

Remark • why dimension counting matters

Notice that the number of corners grows like 2^n while the number of edges grows only like $n2^{n-1}$. High-dimensional geometry is full of such surprises — most of the volume of a high-dimensional cube sits near its corners, far from the centre. This “curse of dimensionality” is precisely what PCA (Chapter 14) is designed to fight.

1.5 Exam preparation

Exam Preparation • Chapter 1

Formulas to know cold

$$\begin{aligned} \mathbf{v} \cdot \mathbf{w} &= \sum_i v_i w_i = \mathbf{v}^T \mathbf{w} && \text{(a scalar)} \\ \|\mathbf{v}\| &= \sqrt{\mathbf{v} \cdot \mathbf{v}} && \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v} \\ \cos \theta &= \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} && \mathbf{v} \perp \mathbf{w} \iff \mathbf{v} \cdot \mathbf{w} = 0 \\ \|\mathbf{v} \pm \mathbf{w}\|^2 &= \|\mathbf{v}\|^2 \pm 2\mathbf{v} \cdot \mathbf{w} + \|\mathbf{w}\|^2 && \text{(expand, don't guess)} \end{aligned}$$

Pitfalls

- $\mathbf{v} \cdot \mathbf{w}$ is a *number*. Writing “ $\mathbf{v} \cdot \mathbf{w} = \mathbf{0}$ ” (bold zero) loses a mark.
- $\|\mathbf{v} + \mathbf{w}\| \neq \|\mathbf{v}\| + \|\mathbf{w}\|$ in general — it is $\leq$, with equality only for parallel same-direction vectors.
- The span of any set always contains $\mathbf{0}$. A line or plane that misses the origin is never a span.
- “None is a combination of the others” must be checked for *all* the vectors, not just the last one. Use the $\sum c_i \mathbf{v}_i = \mathbf{0}$ form instead.

Self-check

1. If $\mathbf{v} + \mathbf{w} = (5, 1)$ and $\mathbf{v} - \mathbf{w} = (1, 5)$, find $\mathbf{v}$ and $\mathbf{w}$. [Tutorial 02, Q5]
2. Every combination of $\mathbf{v} = (1, -2, 1)$ and $\mathbf{w} = (0, 1, -1)$ has components adding to what? Why is $(3, 3, 6)$ impossible as $c\mathbf{v} + d\mathbf{w}$? [Tutorial 02, Q8]
3. Under what condition on a, b, c, d (all nonzero) is $\begin{bmatrix} c \\ d \end{bmatrix}$ a multiple of $\begin{bmatrix} a \\ b \end{bmatrix}$? [Tutorial 02, Q1]
4. What combination $c(1, 2) + d(3, 1)$ gives $(14, 8)$? [Tutorial 02, Q17]

Answers

1. Add: $2\mathbf{v} = (6, 6)$ so $\mathbf{v} = (3, 3)$; subtract: $2\mathbf{w} = (4, -4)$ so $\mathbf{w} = (2, -2)$.
2. Components of $\mathbf{v}$ add to 0 and of $\mathbf{w}$ add to 0, so *every* combination has components adding to 0. Since $3 + 3 + 6 = 12 \neq 0$, $(3, 3, 6)$ is unreachable. For $(3, 3, -6)$: $c = 3, d = 9$.
3. Eliminating m from $c = ma, d = mb$ gives $c/a = d/b$, i.e. $ad = bc$ (equivalently $ad - bc = 0$ — this is the 2×2 determinant, Chapter 7).
4. $c + 3d = 14$ and $2c + d = 8$. Solving: $d = 8 - 2c$, so $c + 24 - 6c = 14, -5c = -10, c = 2, d = 4$.

Systems of Linear Equations

Strang calls solving $Ax = b$ the *fundamental problem of linear algebra*, and the course is organised around it. This chapter sets up the problem three ways — as intersecting rows, as combining columns, and as a single matrix equation — and shows why the column picture is the one that generalises. Understanding the difference between the row and column pictures is the single highest-yield idea in the first half of the course.

Strang §1.3 (“Matrices and Column Spaces”) and §1.4; the row/column picture discussion in Lecture 1. Assignment 01 Q1 and Q3 are direct applications of §1.3. STRANG | 6th ed.

2.1 Linear equations

Definition 2.1 Linear equation

An equation in the unknowns $x_1, \dots, x_n$ is **linear** if it can be written

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b$$

with constants $a_1, \dots, a_n$ (the **coefficients**) and b (the **constant term**). Every unknown appears to the first power only, multiplied by a constant, and never multiplied by another unknown.

Linear	Not linear
$x + 2y + z = 5$	$xy = 1$ (product of unknowns)
$x = 3$	$x^2 + y = 0$ (power $\neq 1$)
$3x + y + z + w = -8$	$\sqrt{x} + y = 2$
$0x_1 + 0x_2 = 0$	$\sin x + y = 0$
$\pi x_1 - \sqrt{2}x_2 = e$	$1/x + y = 4$

Remark

The coefficients may be any real numbers — including irrational ones. “Linear” constrains how the *unknowns* appear, not the constants.

Definition 2.2 System of linear equations

A **system** of m linear equations in n unknowns is

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + \cdots + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \cdots + & a_{mn}x_n & = & b_m \end{array}$$

A **solution** is a list $(x_1, \dots, x_n)$ satisfying *all* m equations at once. The **solution set** is the set of all solutions. The system is **consistent** if the solution set is non-empty and **inconsistent** otherwise. If every $b_i = 0$ the system is **homogeneous**.

Worked Example 2.1 • Lecture 1 — the ice-cream problem

Saman buys two ice creams and two drinks for Rs. 300. Kamala buys two ice creams and one drink for Rs. 200. Find the price of each.

Let x be the price of an ice cream and y the price of a drink:

$$2x + 2y = 300$$

$$2x + y = 200$$

Subtracting the second equation from the first eliminates x and gives $y = 100$; back-substituting, $2x + 100 = 200$, so $x = 50$.

Notice what happened: *one* equation ($2x + 2y = 300$) was not enough to determine two unknowns — it has infinitely many solutions. Adding a second, genuinely new equation pinned the answer down. This is the entire subject in miniature: how many independent equations do you need, and what happens when you have too few or too many?

2.2 The row picture

Definition 2.3 Row picture

In the **row picture**, each equation is read as one geometric object — a line in $\mathbb{R}^2$, a plane in $\mathbb{R}^3$, a **hyperplane** in $\mathbb{R}^n$ — and the solution set is the *intersection* of all of them.

2.2.1 Two equations in two unknowns

Take the system

$$x - 2y = 1$$

$$2x + y = 7$$

Each equation is a line; the solution $(3, 1)$ is where they cross.

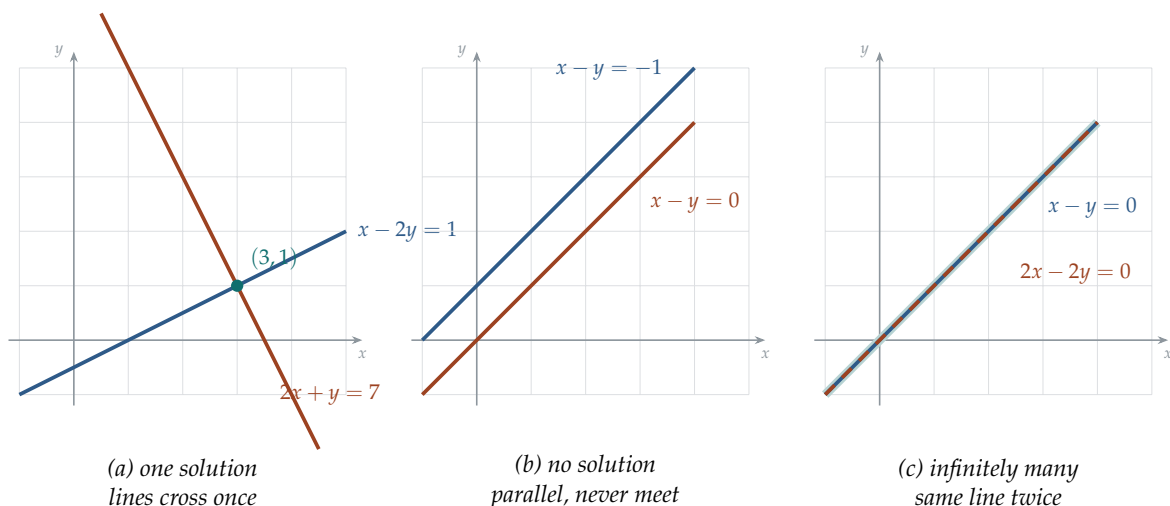


Figure 2.1: The row picture in $\mathbb{R}^2$. Two lines in a plane must either cross once, be parallel and distinct, or coincide. There is no fourth possibility — which is why a linear system has 0, 1, or infinitely many solutions and never exactly two.

Theorem 2.1 Trichotomy for linear systems

Every system of linear equations has exactly one of:

- (i) **no** solution (inconsistent),
- (ii) **exactly one** solution,
- (iii) **infinitely many** solutions.

No linear system has exactly 2, or exactly 17, solutions.

Proof. Suppose $x_1 \neq x_2$ are both solutions of $Ax = b$. For any $t \in \mathbb{R}$ put $x_t = x_1 + t(x_2 - x_1)$. Then

$$Ax_t = Ax_1 + t(Ax_2 - Ax_1) = b + t(b - b) = b,$$

so every x_t is a solution. Since $x_2 - x_1 \neq 0$, distinct values of t give distinct solutions — infinitely many. Hence two solutions force infinitely many. ■

Intuition

The proof is worth remembering as a picture: if you have two solutions, the whole *line through them* consists of solutions. Linearity means solution sets are flat — points, lines, planes — and a flat set with two points in it has infinitely many.

2.2.2 Three equations in three unknowns

Lecture 1 works through

$$\begin{aligned} 2x - y &= 0 \\ -x + 2y - z &= -1 \\ -3y + 4z &= 4 \end{aligned}$$

Each equation is now a *plane* in $\mathbb{R}^3$. Three planes typically meet in a single point.

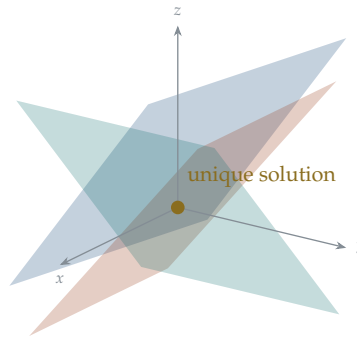


Figure 2.2: Row picture in $\mathbb{R}^3$: three planes meeting at one point. Special cases: the three planes can share a whole line (infinitely many solutions), or form a triangular “prism” with no common point (inconsistent), or two can be parallel.

Common Pitfall • the row picture stops being useful fast

In $\mathbb{R}^3$ the row picture is already hard to draw reliably — “do these three planes meet in a point?” is not something the eye judges well. In $\mathbb{R}^4$ and beyond it is unusable. The column picture, next, stays manageable because it asks a question about *one* vector rather than about the intersection of many surfaces.

2.3 The column picture

Definition 2.4 Column picture

Group the same system by columns instead of rows:

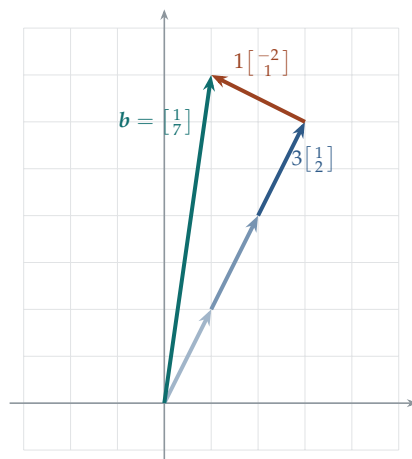
$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

In the **column picture** the question is not “where do these surfaces meet?” but:

Which combination of the column vectors produces $\mathbf{b}$?

For the 2×2 system above this reads

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}, \quad \text{and indeed} \quad 3 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}.$$



The column picture: $3\mathbf{a}_1 + 1\mathbf{a}_2 = \mathbf{b}$

Figure 2.3: Column picture for $x - 2y = 1$, $2x + y = 7$. We walk three steps along the first column, then one step along the second, and arrive at $\mathbf{b}$. The solution $(3, 1)$ is the recipe.

Intuition • why the column picture wins

The row picture asks a question about m different surfaces in $\mathbb{R}^n$; its difficulty grows with the size of the system. The column picture asks one question — “is $\mathbf{b}$ in the set of combinations of these columns?” — and that question has the *same shape* whether $n = 2$ or $n = 2000$. Everything that follows (column space, rank, independence, basis) is machinery for answering it.

Worked Example 2.2 • Lecture 1 — when the columns fail

Consider

$$\begin{aligned} 2x - y &= 1 \\ -x + 2y - z &= 1 \\ -3y + 4z &= -3 \end{aligned}$$

Look at the columns:

$$\mathbf{a}_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}.$$

First check whether the columns are independent. Try $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3 = \mathbf{0}$:

$$2c_1 - c_2 = 0, \quad -c_1 + 2c_2 - c_3 = 0, \quad -3c_2 + 4c_3 = 0.$$

From the first, $c_2 = 2c_1$; from the third, $c_3 = \frac{3}{4}c_2 = \frac{3}{2}c_1$. Substituting into the second: $-c_1 + 4c_1 - \frac{3}{2}c_1 = \frac{3}{2}c_1 = 0$, so $c_1 = 0$ and then $c_2 = c_3 = 0$. The columns are **independent**, so their combinations fill all of $\mathbb{R}^3$ and *every* $\mathbf{b}$ is reachable — there is exactly one solution. Solving (by elimination, Chapter 4) gives $x = 1$, $y = 1$, $z = 0$. Check: $1\mathbf{a}_1 + 1\mathbf{a}_2 + 0\mathbf{a}_3 = (2 - 1, -1 + 2, 0 - 3) = (1, 1, -3) = \mathbf{b}$. ✓

Worked Example 2.3 • Assignment 01, Q3 — when all columns lie on one line

Let

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix}.$$

Show all columns lie on the same line, describe $C(A)$, and decide whether $Ax = (12, 6, -6)$ and $Ax = (1, 0, 0)$ are solvable.

Columns on one line. Write the columns:

$$a_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} = 2a_1, \quad a_3 = \begin{bmatrix} 6 \\ 3 \\ -3 \end{bmatrix} = 3a_1.$$

All three are multiples of a_1 , so they all lie on the line through the origin in the direction $(2, 1, -1)$. Only *one* independent direction is present, so $\text{rank } A = 1$.

Column space. $C(A)$ is the set of all combinations $x_1 a_1 + x_2 (2a_1) + x_3 (3a_1) = (x_1 + 2x_2 + 3x_3)a_1$, which is just *all multiples of a_1* — the line $\{t(2, 1, -1) : t \in \mathbb{R}\}$ in $\mathbb{R}^3$. It is a one-dimensional subspace, not a plane and certainly not all of $\mathbb{R}^3$.

Is $(12, 6, -6)$ reachable? Yes: $(12, 6, -6) = 6(2, 1, -1) = 6a_1$, so it lies on the line. The system is consistent. Solutions satisfy $x_1 + 2x_2 + 3x_3 = 6$ — a plane's worth of solutions, i.e. infinitely many. Taking $x_2 = s, x_3 = t$ free:

$$x = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}.$$

Is $(1, 0, 0)$ reachable? No. Every vector in $C(A)$ has the form $(2t, t, -t)$; matching the second component forces $t = 0$, which gives $(0, 0, 0) \neq (1, 0, 0)$. So $b \notin C(A)$ and the system is **inconsistent**.

2.4 Matrix form

Definition 2.5 Coefficient matrix and matrix form

Collecting the coefficients into an $m \times n$ matrix A , the unknowns into $x \in \mathbb{R}^n$ and the right-hand sides into $b \in \mathbb{R}^m$, the system becomes the single equation

$$\boxed{Ax = b} \quad \text{where} \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

A is the **coefficient matrix**. Appending b as an extra column gives the **augmented matrix** $[A \mid b]$, of size $m \times (n + 1)$.

Definition 2.6 Matrix–vector product, two readings

For $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$:

column reading: $A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ (a combination of columns)

row reading: $(A\mathbf{x})_i = \sum_{j=1}^n a_{ij}x_j = (\text{row } i \text{ of } A) \cdot \mathbf{x}$ (a list of dot products)

Both give the same answer. The column reading is the conceptual one; the row reading is how you compute by hand.

Column reading

$$\begin{array}{|c|c|} \hline 1 & -2 \\ \hline 2 & 1 \\ \hline \end{array} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} (1)(3) + (-2)(1) \\ (2)(3) + (1)(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

Figure 2.4: The same product $A\mathbf{x}$ computed two ways. Boxes mark what is being grouped.

Worked Example 2.4 • Tutorial 02, Q27 — writing a combination as $A\mathbf{x} = \mathbf{b}$

With $\mathbf{u} = (2, -1, 0)$, $\mathbf{v} = (-1, 2, -1)$, $\mathbf{w} = (0, -1, 2)$ and $\mathbf{b} = (1, 0, 0)$, write three equations for c, d, e so that $c\mathbf{u} + d\mathbf{v} + e\mathbf{w} = \mathbf{b}$, express this as $A\mathbf{x} = \mathbf{b}$, and solve.

Matching components gives

$$\begin{array}{rcl} 2c - d & = & 1 \\ -c + 2d - e & = & 0 \\ -d + 2e & = & 0 \end{array} \iff \underbrace{\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}}_A \begin{bmatrix} c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

From the third equation $d = 2e$. Substituting into the second: $-c + 4e - e = 0$, so $c = 3e$. The first then gives $6e - 2e = 1$, i.e. $4e = 1$, $e = \frac{1}{4}$. Hence

$$c = \frac{3}{4}, \quad d = \frac{1}{2}, \quad e = \frac{1}{4}.$$

Check the first equation: $2(\frac{3}{4}) - \frac{1}{2} = \frac{3}{2} - \frac{1}{2} = 1$. ✓

This A is Strang's second-difference matrix. It is symmetric, its columns are independent, and it reappears throughout the book.

2.5 Square systems and what can go wrong

Definition 2.7 Square, singular, nonsingular

A matrix is **square** if $m = n$. A square matrix A is **nonsingular** (equivalently **invertible**) if its columns are linearly independent; otherwise it is **singular**.

Theorem 2.2 The square case

Let A be $n \times n$. Exactly one of the following holds.

A nonsingular	A singular
columns independent	columns dependent
columns span all of $\mathbb{R}^n$	columns span a lower-dimensional subspace
$Ax = b$ has exactly one solution for every b	$Ax = b$ has no solution for some b , and infinitely many for others
$Ax = 0$ only when $x = 0$	$Ax = 0$ has nonzero solutions
$\det A \neq 0$	$\det A = 0$
A^{-1} exists	A^{-1} does not exist
n nonzero pivots	fewer than n pivots
$\text{rank } A = n$	$\text{rank } A < n$

This table — often called the *Invertible Matrix Theorem* — is assembled piece by piece over the next several chapters. Each column is a list of statements that are all equivalent to one another. Learn it as a single fact with many faces, not as eight separate facts.

Worked Example 2.5 • Assignment 01, Q1 — a parameter that changes everything

Let

$$A(t) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & t & 2 \\ 1 & 2 & t \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

For which t does $A(t)x = b$ have a unique solution for every $b \in \mathbb{R}^3$?

Unique solution for every $b \iff A(t)$ is nonsingular $\iff \det A(t) \neq 0$. Expanding along the first row:

$$\det A(t) = 1 \begin{vmatrix} t & 2 \\ 2 & t \end{vmatrix} - 1 \begin{vmatrix} 1 & 2 \\ 1 & t \end{vmatrix} + 1 \begin{vmatrix} 1 & t \\ 1 & 2 \end{vmatrix} = (t^2 - 4) - (t - 2) + (2 - t).$$

Simplify:

$$\det A(t) = t^2 - 4 - t + 2 + 2 - t = t^2 - 2t = t(t - 2).$$

So $\det A(t) = 0$ exactly when $t = 0$ or $t = 2$.

$t \notin \{0, 2\}$	nonsingular: unique solution for every b
$t = 0$ or $t = 2$	singular: solvable only for special b

The full analysis of the two singular cases — including the exact conditions on b_1, b_2, b_3 — is in Appendix B, Question 1.

Common Pitfall • “more equations than unknowns” does not mean “no solution”

A 5×3 system can be perfectly consistent if the five right-hand sides happen to be compatible; and a 3×5 system (more unknowns than equations) is *usually* but not always solvable — if the rows are dependent in a bad way it can still be inconsistent. Count *pivots*, not equations. See Chapter 9.

2.6 Exam preparation

Exam Preparation • Chapter 2

The three views of one system

View	Object	Question
Row	m hyperplanes in $\mathbb{R}^n$	Where do they intersect?
Column	n vectors in $\mathbb{R}^m$	Which combination gives $\mathbf{b}$?
Matrix	one equation $A\mathbf{x} = \mathbf{b}$	Can A be undone?

Facts to state without hesitation

- A linear system has 0, 1, or infinitely many solutions — never finitely many others.
- $A\mathbf{x}$ is a linear combination of the *columns* of A , with the entries of $\mathbf{x}$ as weights.
- $A\mathbf{x} = \mathbf{b}$ is solvable $\iff \mathbf{b} \in \mathbf{C}(A)$.
- A square A is nonsingular $\iff$ its columns are independent $\iff \det A \neq 0 \iff A\mathbf{x} = \mathbf{0}$ forces $\mathbf{x} = \mathbf{0}$.

Pitfalls

- Forgetting that $A\mathbf{x}$ requires #columns of $A =$ #entries of $\mathbf{x}$.
- Confusing the augmented matrix $[A \mid \mathbf{b}]$ with the coefficient matrix A when computing rank — they can differ, and that difference *is* inconsistency.
- Drawing the row picture in $\mathbb{R}^3$ and trusting the drawing.

Self-check

1. Write $x + 2y + 3z = 5$, $2x + 5y + 7z = 8$ in matrix form. What are the sizes of A , $\mathbf{x}$, $\mathbf{b}$?
2. The columns of A are $(1, 1)$, $(2, 2)$, $(3, 3)$. Describe $\mathbf{C}(A)$. For which $\mathbf{b}$ is $A\mathbf{x} = \mathbf{b}$ solvable?
3. If A is 4×6 , can $A\mathbf{x} = \mathbf{b}$ have exactly one solution? Why?
4. Give a system of three equations in three unknowns that is inconsistent even though no two of the planes are parallel.

Answers

1. $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \end{bmatrix}$ is 2×3 , $\mathbf{x} \in \mathbb{R}^3$, $\mathbf{b} = (5, 8) \in \mathbb{R}^2$.
2. All three columns are multiples of $(1, 1)$, so $\mathbf{C}(A)$ is the line $\{t(1, 1)\}$ in $\mathbb{R}^2$. Solvable exactly when $b_1 = b_2$.
3. No. With 6 unknowns and at most 4 pivots there are always at least 2 free variables, so any solution comes with infinitely many others. It is 0 or ∞ , never 1.
4. $x + y = 0$, $y + z = 0$, $x - z = 1$. Adding the first two and subtracting the third gives $0 = -1$. No two normals $(1, 1, 0)$, $(0, 1, 1)$, $(1, 0, -1)$ are parallel, so no two planes are parallel — they form a prism.

Matrix Algebra

Matrices are not merely rectangular arrays of numbers — they are *operations*, and matrix multiplication is composition of operations. This chapter defines the arithmetic, catalogues the special shapes that recur throughout the course, and explains the one property that trips everyone up: multiplication is not commutative. The block view of multiplication introduced at the end is the tool that makes *LU*, *QR* and the SVD readable.

Strang §1.3–§1.4 and §2.4 (“Rules for Matrix Operations”). Lecture 3 opens with matrix multiplication “the row way” and “the column way”; Tutorial 01 Q1–Q4 is the drill. STRANG | 6th ed.

3.1 Matrices, sizes and equality

Definition 3.1 Matrix

An $m \times n$ **matrix** is a rectangular array

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n},$$

with m rows and n columns. **Rows first, columns second** — in the size, in the subscripts, everywhere. Two matrices are equal iff they have the same size and $a_{ij} = b_{ij}$ for all i, j .

Common Pitfall • row/column order

a_{23} is row 2, column 3. A 3×5 matrix is tall-ish and wide: 3 rows, 5 columns. Getting this backwards makes every subsequent dimension check fail, and it is the most common source of lost marks in the first half of the course.

3.2 Addition and scalar multiplication

Definition 3.2 Addition, scalar multiple, subtraction

For $A, B \in \mathbb{R}^{m \times n}$ of the same size and $c \in \mathbb{R}$:

$$(A + B)_{ij} = a_{ij} + b_{ij}, \quad (cA)_{ij} = c a_{ij}, \quad A - B = A + (-1)B.$$

Addition of differently-sized matrices is undefined.

Proposition 3.1 Rules

For matrices of matching sizes and scalars c, d :

$$\begin{aligned} A + B &= B + A, & (A + B) + C &= A + (B + C), & A + O &= A, & A + (-A) &= O, \\ c(A + B) &= cA + cB, & (c + d)A &= cA + dA, & c(dA) &= (cd)A, & 1A &= A. \end{aligned}$$

These are exactly the eight vector-space axioms of Proposition 1.1 — which tells you that $\mathbb{R}^{m \times n}$ is itself a vector space, of dimension mn (Chapter 8).

Worked Example 3.1 • Tutorial 01, Q1 — basic combinations

With

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 6 & 7 \\ 3 & 4 & 5 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & 5 & 7 \\ 5 & -6 & 5 \end{bmatrix},$$

compute $A + B$, $A - B$, $A + B - C$ and $2A - B - 3C$.

$$A + B = \begin{bmatrix} 6 & 8 & 10 \\ 10 & 12 & 14 \end{bmatrix}, \quad A - B = \begin{bmatrix} -4 & -4 & -4 \\ 4 & 4 & 4 \end{bmatrix},$$

$$A + B - C = \begin{bmatrix} 6 - (-1) & 8 - 5 & 10 - 7 \\ 10 - 5 & 12 - (-6) & 14 - 5 \end{bmatrix} = \begin{bmatrix} 7 & 3 & 3 \\ 5 & 18 & 9 \end{bmatrix},$$

$$2A = \begin{bmatrix} 2 & 4 & 6 \\ 14 & 16 & 18 \end{bmatrix}, \quad 3C = \begin{bmatrix} -3 & 15 & 21 \\ 15 & -18 & 15 \end{bmatrix},$$

$$2A - B - 3C = \begin{bmatrix} 2 - 5 - (-3) & 4 - 6 - 15 & 6 - 7 - 21 \\ 14 - 3 - 15 & 16 - 4 - (-18) & 18 - 5 - 15 \end{bmatrix} = \begin{bmatrix} 0 & -17 & -22 \\ -4 & 30 & -2 \end{bmatrix}.$$

3.3 Matrix multiplication

Definition 3.3 Matrix product

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. The product $AB \in \mathbb{R}^{m \times p}$ has entries

$$(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj} = (\text{row } i \text{ of } A) \cdot (\text{column } j \text{ of } B).$$

The product is defined *only* when the inner dimensions agree:

$$\underbrace{(m \times n)}_A \underbrace{(n \times p)}_B = (m \times p).$$

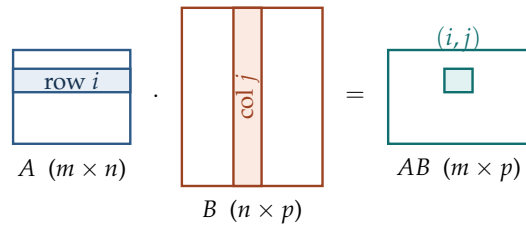


Figure 3.1: Entry (i, j) of AB is the dot product of row i of A with column j of B . The shared dimension n is consumed.

3.3.1 Four ways to see the same product

Lecture 3 emphasises the “row way” and the “column way”. There are in fact four useful decompositions, and different ones make different theorems obvious.

View	Formula	Makes obvious
Entry (dot)	$(AB)_{ij} = (\text{row}_i A) \cdot (\text{col}_j B)$	how to compute by hand
Columns	$AB = [Ab_1 \ Ab_2 \ \cdots \ Ab_p]$	every column of AB is in $\mathbf{C}(A)$
Rows	$\text{row}_i(AB) = (\text{row}_i A) B$	every row of AB is in the row space of B
Outer products	$AB = \sum_{k=1}^n a_k (\text{row}_k B)$	rank, and the SVD (Ch. 13)

Intuition • the column view is the one to internalise

$AB = [Ab_1 \ \cdots \ Ab_p]$ says: to multiply by B , apply A to each column of B separately. Two consequences follow with no further work — $\mathbf{C}(AB) \subseteq \mathbf{C}(A)$, and $\text{rank}(AB) \leq \text{rank}(A)$. Both are used constantly later.

Worked Example 3.2 • Tutorial 01, Q3 — verifying $A(B + C) = AB + AC$

Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & -1 \\ 1 & 0 & -1 \\ -2 & 1 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 1 & 4 \\ 2 & 1 & 3 \\ 2 & 2 & 1 \end{bmatrix}.$$

Left side. First $B + C = \begin{bmatrix} 5 & 0 & 3 \\ 3 & 1 & 2 \\ 0 & 3 & 3 \end{bmatrix}$, then

$$A(B + C) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 & 3 \\ 3 & 1 & 2 \\ 0 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 8 & 4 & 8 \\ 6 & 5 & 7 \\ 5 & 3 & 6 \end{bmatrix}.$$

For instance the $(1, 1)$ entry is $1(5) + 1(3) + 1(0) = 8$ and the $(2, 3)$ entry is $0(3) + 2(2) + 1(3) = 7$.

Right side.

$$AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \text{— compute directly: } AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

since row 1 of A dotted with column 1 of B is $2 + 1 - 2 = 1$, with column 2 is $-1 + 0 + 1 = 0$, with column 3 is $-1 - 1 + 2 = 0$; and similarly for the other rows. (So $B = A^{-1}$ — a pleasant accident of this exercise.) Next

$$AC = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ 2 & 1 & 3 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 4 & 8 \\ 6 & 4 & 7 \\ 5 & 3 & 5 \end{bmatrix}.$$

Adding,

$$AB + AC = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 7 & 4 & 8 \\ 6 & 4 & 7 \\ 5 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 8 & 4 & 8 \\ 6 & 5 & 7 \\ 5 & 3 & 6 \end{bmatrix} = A(B + C). \checkmark$$

3.3.2 The rules — and the one that fails

Theorem 3.1 Algebra of matrix multiplication

Whenever the sizes make the products defined:

$(AB)C = A(BC)$	associative
$A(B + C) = AB + AC, \quad (A + B)C = AC + BC$	distributive
$c(AB) = (cA)B = A(cB)$	scalars slide
$I_m A = A = A I_n$	identity
$(AB)^T = B^T A^T$	transpose reverses

But in general

$$AB \neq BA$$

Matrix multiplication is **not commutative**.

Proof of associativity. Compare entry (i, ℓ) of each side:

$$((AB)C)_{i\ell} = \sum_j (AB)_{ij} c_{j\ell} = \sum_j \left(\sum_k a_{ik} b_{kj} \right) c_{j\ell} = \sum_k \sum_j a_{ik} b_{kj} c_{j\ell},$$

$$(A(BC))_{i\ell} = \sum_k a_{ik} (BC)_{k\ell} = \sum_k a_{ik} \left(\sum_j b_{kj} c_{j\ell} \right) = \sum_k \sum_j a_{ik} b_{kj} c_{j\ell}.$$

The two double sums are identical, so the matrices are equal. ■

Worked Example 3.3 • Lecture 3 — $AB \neq BA$

Take $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (a row swapper).

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}.$$

AB swapped the *columns* of A ; BA swapped the *rows*. These are genuinely different operations, so of course the answers differ.

Worse, the product can fail to exist in one order. If A is 2×3 and B is 3×4 then AB is 2×4 but BA is undefined.

Common Pitfall • three consequences of non-commutativity

- $(A + B)^2 = A^2 + AB + BA + B^2$, **not** $A^2 + 2AB + B^2$.
- $(AB)^{-1} = B^{-1}A^{-1}$ and $(AB)^T = B^T A^T$ — the order *reverses*. Think of putting on socks then shoes: to undo, take off shoes then socks.
- $AB = O$ does **not** imply $A = O$ or $B = O$. Example: $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Likewise $AB = AC$ does not let you cancel A unless A is invertible.

Worked Example 3.4 • Tutorial 01, Q4 — solving matrix equations

(a) Solve $\begin{bmatrix} x & 3 \\ z & 2 \end{bmatrix} + \begin{bmatrix} 2-y & z \\ 2-z & -x \end{bmatrix} = \begin{bmatrix} 3 & 7 \\ 2 & 0 \end{bmatrix}$ for x, y, z .

(b) With $A = \begin{bmatrix} 3 & 4 \\ -1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 2 \\ -1 & 2 \end{bmatrix}$, find X with $2X + B = 3A$.

(a) Equating entries one at a time:

$$\begin{aligned} (1,1) : x + 2 - y &= 3 & (1,2) : 3 + z &= 7 \\ (2,1) : z + 2 - z &= 2 & (2,2) : 2 - x &= 0 \end{aligned}$$

The (1,2) entry gives $z = 4$. The (2,2) entry gives $x = 2$. Then (1,1) gives $2 + 2 - y = 3$, so $y = 1$. The (2,1) entry reads $2 = 2$ — automatically satisfied, carrying no information. Answer: $x = 2$, $y = 1$, $z = 4$.

(b) Rearranging, $X = \frac{1}{2}(3A - B)$. Now

$$3A = \begin{bmatrix} 9 & 12 \\ -3 & 6 \end{bmatrix}, \quad 3A - B = \begin{bmatrix} 9-3 & 12-2 \\ -3+1 & 6-2 \end{bmatrix} = \begin{bmatrix} 6 & 10 \\ -2 & 4 \end{bmatrix},$$

$$X = \frac{1}{2} \begin{bmatrix} 6 & 10 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ -1 & 2 \end{bmatrix}.$$

Check: $2X + B = \begin{bmatrix} 6 & 10 \\ -2 & 4 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 9 & 12 \\ -3 & 6 \end{bmatrix} = 3A. \checkmark$

3.4 Special matrices

These shapes recur constantly; recognising one immediately tells you something.

Name	Condition	Why it matters
Zero matrix O	all $a_{ij} = 0$	additive identity
Identity I_n	$a_{ii} = 1$, else 0	$IA = AI = A$
Diagonal D	$a_{ij} = 0$ for $i \neq j$	powers and inverses are trivial
Upper triangular U	$a_{ij} = 0$ for $i > j$	what elimination produces
Lower triangular L	$a_{ij} = 0$ for $i < j$	records the elimination multipliers
Symmetric S	$S^T = S$	real eigenvalues, orthogonal eigenvectors
Skew-symmetric	$A^T = -A$	zero diagonal; rotations' generators
Permutation P	rows of I reordered	$P^{-1} = P^T$
Orthogonal Q	$Q^T Q = I$	preserves lengths and angles

Proposition 3.2 Diagonal matrices are easy

If $D = \text{diag}(d_1, \dots, d_n)$ then

$$D^k = \text{diag}(d_1^k, \dots, d_n^k), \quad D^{-1} = \text{diag}(d_1^{-1}, \dots, d_n^{-1}) \text{ if all } d_i \neq 0.$$

Also, DA scales the *rows* of A (row i by d_i), while AD scales the *columns* (column j by d_j).

Worked Example 3.5 • Tutorial 04, Q4 — powers of a diagonal matrix

Let $A = \text{diag}(1, -3, 2)$. Find A^{-1} , A^5 and A^{-5} .

$$A^{-1} = \text{diag}\left(1, -\frac{1}{3}, \frac{1}{2}\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix},$$

$$A^5 = \text{diag}(1^5, (-3)^5, 2^5) = \text{diag}(1, -243, 32),$$

$$A^{-5} = (A^5)^{-1} = \text{diag}\left(1, -\frac{1}{243}, \frac{1}{32}\right).$$

Note how much work diagonality saves: computing A^5 by repeated multiplication would take four full matrix products. This is exactly why *diagonalization* (Chapter 12) is worth the effort — it turns any well-behaved matrix into this easy case.

3.5 The trace

Definition 3.4 Trace

For a *square* matrix $A \in \mathbb{R}^{n \times n}$, the **trace** is the sum of the diagonal entries:

$$\text{tr } A = a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^n a_{ii}.$$

Proposition 3.3 Properties of the trace

For square A, B of the same size and $c \in \mathbb{R}$:

$$\text{tr}(A + B) = \text{tr } A + \text{tr } B, \quad \text{tr}(cA) = c \text{tr } A, \quad \text{tr}(A^T) = \text{tr } A,$$

and, remarkably,

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

even though $AB \neq BA$ in general. (This also holds when A is $m \times n$ and B is $n \times m$, so both products are square.)

Proof of the cyclic property.

$$\operatorname{tr}(AB) = \sum_i (AB)_{ii} = \sum_i \sum_k a_{ik} b_{ki} = \sum_k \sum_i b_{ki} a_{ik} = \sum_k (BA)_{kk} = \operatorname{tr}(BA).$$

Only the order of a finite double sum was changed. ■

Remark • the trace resurfaces

In Chapter 11 we prove $\operatorname{tr} A = \lambda_1 + \cdots + \lambda_n$: the trace is the sum of the eigenvalues. Together with $\det A = \lambda_1 \cdots \lambda_n$ this gives a two-second sanity check on any eigenvalue computation, and Tutorial 06 Q5 asks you to prove it.

Worked Example 3.6 • Tutorial 01, Q2 — traces and transposes

With

$$A = \begin{bmatrix} -5 & 2 & 0 \\ 7 & -3 & 4 \\ -1 & 3 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & -1 & 8 \\ 6 & -14 & 2 \\ 9 & 5 & 1 \end{bmatrix},$$

evaluate $\operatorname{tr}(A + B)$ and verify $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

$\operatorname{tr} A = -5 - 3 + 2 = -6$ and $\operatorname{tr} B = 5 - 14 + 1 = -8$, so $\operatorname{tr}(A + B) = -6 - 8 = -14$.

Directly: $A + B = \begin{bmatrix} 0 & 1 & 8 \\ 13 & -17 & 6 \\ 8 & 8 & 3 \end{bmatrix}$ has diagonal sum $0 - 17 + 3 = -14$. ✓

For $\operatorname{tr}(AB)$ we only need the three diagonal entries of AB :

$$(AB)_{11} = (-5)(5) + (2)(6) + (0)(9) = -25 + 12 + 0 = -13,$$

$$(AB)_{22} = (7)(-1) + (-3)(-14) + (4)(5) = -7 + 42 + 20 = 55,$$

$$(AB)_{33} = (-1)(8) + (3)(2) + (2)(1) = -8 + 6 + 2 = 0.$$

So $\operatorname{tr}(AB) = -13 + 55 + 0 = 42$. Similarly for BA :

$$(BA)_{11} = (5)(-5) + (-1)(7) + (8)(-1) = -25 - 7 - 8 = -40,$$

$$(BA)_{22} = (6)(2) + (-14)(-3) + (2)(3) = 12 + 42 + 6 = 60,$$

$$(BA)_{33} = (9)(0) + (5)(4) + (1)(2) = 0 + 20 + 2 = 22.$$

$\operatorname{tr}(BA) = -40 + 60 + 22 = 42 = \operatorname{tr}(AB)$. ✓

3.6 Block multiplication

Definition 3.5 Block (partitioned) matrices

A matrix may be cut into rectangular **blocks**. If the cuts are compatible — the column cuts of A match the row cuts of B — then blocks multiply by the same rule as entries:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

Order matters inside each product, since the blocks are matrices.

Special cases you have already used without naming them:

Partition	What it gives
B split into single columns	$AB = [Ab_1 \cdots Ab_p]$
A split into single rows	rows of AB are rows of A times B
A by columns, B by rows	$AB = \sum_k a_k(\text{row}_k B)$ — outer products
$[A \mid b]$	the augmented matrix of Chapter 4
$[A \mid I]$	the Gauss–Jordan inverse tableau (Chapter 5)

Worked Example 3.7 • Outer-product form and rank

Compute $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ as a sum of outer products, and read off its rank.

Write $A = a_1 r_1 + a_2 r_2$ where a_k are the columns of the left factor and r_k the rows of the right. Taking $A = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix}$ directly:

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = A.$$

A is a *single* outer product — a column times a row — so $\text{rank } A = 1$. Every rank-one matrix has exactly this form uv^T , and the SVD (Chapter 13) writes *any* matrix as a sum of r such pieces, ordered by importance.

3.7 Exam preparation

Exam Preparation • Chapter 3

Dimension bookkeeping

$$(m \times n)(n \times p) = (m \times p).$$

Before every product, write the two sizes down and check the inner pair matches. Most computational errors are caught here for free.

Rules that hold

$$(AB)C = A(BC), \quad A(B+C) = AB+AC, \quad (AB)^T = B^T A^T, \quad \text{tr}(AB) = \text{tr}(BA).$$

Rules that fail

$$AB \neq BA, \quad (A+B)^2 \neq A^2 + 2AB + B^2, \quad AB = O \not\Rightarrow A = O \text{ or } B = O.$$

Pitfalls

- Cancelling: $AB = AC$ gives $B = C$ *only* if A is invertible.
- Reversing order in $(AB)^T$ and $(AB)^{-1}$ — forget the reversal and every subsequent line is wrong.
- Trace is defined only for square matrices, and $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$ but *not* $\text{tr}(ACB)$.

Self-check

1. A is 3×4 and B is 4×2 . Which of $AB, BA, A^T B^T, B^T A^T$ are defined, and what are their sizes?
2. Find a 2×2 matrix $A \neq O$ with $A^2 = O$.
3. If A is $m \times n$, what size is $A^T A$? What size is AA^T ? Are they equal?
4. Find a solution $X \in \mathbb{R}^{2 \times 2}$ of $X^2 - 2X = \begin{bmatrix} -1 & 0 \\ 6 & 3 \end{bmatrix}$. [Tutorial 01, Q4(c)]

Answers

1. AB is 3×2 . BA is undefined ($2 \neq 3$). $A^T B^T$ is undefined (4×3 times 2×4). $B^T A^T$ is 2×3 — and equals $(AB)^T$.
2. $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Then $A^2 = O$ although $A \neq O$: matrices can be *nilpotent*.
3. $A^T A$ is $n \times n$; AA^T is $m \times m$. They are different sizes unless $m = n$, and even then usually unequal — but both are symmetric, and (Ch. 13) they share the same nonzero eigenvalues.
4. Try $X = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix}$. Then $X^2 - 2X = \begin{bmatrix} a^2 - 2a & 0 \\ c(a+d-2) & d^2 - 2d \end{bmatrix}$. Matching: $a^2 - 2a = -1$ gives $a = 1$; $d^2 - 2d = 3$ gives $d = 3$ or $d = -1$; take $d = 3$, then $c(1+3-2) = 6$ gives $c = 3$. So $X = \begin{bmatrix} 1 & 0 \\ 3 & 3 \end{bmatrix}$. (Solutions are not unique — matrix “square roots” rarely are.)

Elimination and Back Substitution

Elimination is the algorithm that turns a system into a triangular one you can solve by inspection. It is completely mechanical, it works for any size, and — crucially — the record it leaves behind (the multipliers, the pivots, the row exchanges) contains almost everything you will later want to know about the matrix: its rank, its determinant, its factorization $A = LU$, its four subspaces. This chapter covers Gaussian elimination, back substitution, echelon forms, Gauss–Jordan, and the two ways elimination can break down.

Strang §2.1–§2.3 (“Elimination”, “Elimination Using Matrices”, “Rules for Matrix Operations”) and §2.6 for $A = LU$. This is Lecture 3 and the second half of Tutorial 04. STRANG | 6th.ed.

4.1 The idea

Intuition

Solving $2x + 3y = 8$, $4x + 7y = 18$ is hard because both unknowns appear in both equations. Subtract $2 \times (\text{first})$ from (second) and x vanishes from equation two, leaving $y = 2$ immediately. Elimination is nothing more than doing that systematically, one variable at a time, until the system is triangular — at which point the last equation has one unknown, the second-to-last has two (one already known), and so on backwards.

Definition 4.1 Elementary row operations

There are exactly three, and each is reversible:

Type	Notation	Inverse operation
Replacement	$R_i \rightarrow R_i - \ell R_j \ (i \neq j)$	$R_i \rightarrow R_i + \ell R_j$
Interchange	$R_i \leftrightarrow R_j$	itself
Scaling	$R_i \rightarrow c R_i \ (c \neq 0)$	$R_i \rightarrow \frac{1}{c} R_i$

Theorem 4.1 Row operations preserve the solution set

If $[A' \mid \mathbf{b}']$ is obtained from $[A \mid \mathbf{b}]$ by a sequence of elementary row operations, then $A\mathbf{x} = \mathbf{b}$ and $A'\mathbf{x} = \mathbf{b}'$ have exactly the same solutions. Such systems are called **row equivalent**.

Proof. It suffices to check one operation, since a sequence is a composition. If $\mathbf{x}$ satisfies every equation of the original system, it satisfies any linear combination of them — in particular the combination (eqn i) $-\ell$ (eqn j). So every solution of the old system solves the new one. Because

each operation is reversible (right-hand column of the table), the same argument run backwards shows every solution of the new system solves the old. Hence the solution sets coincide. ■

Common Pitfall • the operation must be reversible

$R_i \rightarrow 0 \cdot R_i$ is *not* allowed: it destroys information and can turn an inconsistent system into a consistent one. Similarly $R_i \rightarrow R_i - \ell R_i$ (using the same row) is forbidden. Always subtract a multiple of a *different* row.

4.2 Gaussian elimination

Definition 4.2 Pivot, upper triangular form

Working left to right, the **pivot** in each step is the first nonzero entry of the current row, in the column being cleared. A matrix U is **upper triangular** if $u_{ij} = 0$ whenever $i > j$ — all entries below the main diagonal are zero.

Gaussian elimination uses row replacements (and row exchanges when needed) to convert A into an upper triangular U , carrying $\mathbf{b}$ along to become $\mathbf{c}$:

$$[A \mid \mathbf{b}] \rightarrow [U \mid \mathbf{c}], \quad Ax = \mathbf{b} \iff Ux = \mathbf{c}.$$

Method Gaussian elimination

For $j = 1, 2, \dots$ (columns):

1. Find the pivot: the entry in column j at or below the current pivot row. If it is zero, exchange with a lower row that has a nonzero entry there. If the whole column below is zero, move to the next column (this column has no pivot).
2. For each row i below the pivot row, compute the **multiplier**

$$\ell_{ij} = \frac{\text{entry to eliminate}}{\text{pivot}}$$

and apply $R_i \rightarrow R_i - \ell_{ij}R_{\text{pivot}}$.

3. Move down one row and right one column; repeat.

Never divide by a pivot of zero. Record every ℓ_{ij} — they are the entries of L in the factorization $A = LU$.

Definition 4.3 Back substitution

Once $Ux = \mathbf{c}$ is triangular, solve *upwards*: the last equation gives x_n directly; substitute into the second-to-last to get x_{n-1} ; and so on. Explicitly,

$$x_i = \frac{1}{u_{ii}} \left(c_i - \sum_{j>i} u_{ij}x_j \right), \quad i = n, n-1, \dots, 1.$$

Solve

$$\begin{aligned}x + 2y + z &= 0 \\2x + 3y + 3z &= 10 \\3x - y + 2z &= 14\end{aligned}$$

Augmented matrix.

$$[A | \mathbf{b}] = \left[\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 2 & 3 & 3 & 10 \\ 3 & -1 & 2 & 14 \end{array} \right]$$

Step 1 — clear column 1. Pivot = 1 in position (1,1). Multipliers $\ell_{21} = 2/1 = 2$ and $\ell_{31} = 3/1 = 3$:

$$\begin{aligned}R_2 &\rightarrow R_2 - 2R_1 \\R_3 &\rightarrow R_3 - 3R_1\end{aligned} \implies \left[\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & -1 & 1 & 10 \\ 0 & -7 & -1 & 14 \end{array} \right]$$

Step 2 — clear column 2. Pivot = -1 in position (2,2). Multiplier $\ell_{32} = (-7)/(-1) = 7$:

$$R_3 \rightarrow R_3 - 7R_2 \implies \left[\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & -1 & 1 & 10 \\ 0 & 0 & -8 & -56 \end{array} \right] = [U | \mathbf{c}]$$

The pivots are 1, -1, -8 — all nonzero, so the solution is unique.

Back substitution.

$$-8z = -56 \implies z = 7.$$

$$-y + z = 10 \implies -y + 7 = 10 \implies y = -3.$$

$$x + 2y + z = 0 \implies x - 6 + 7 = 0 \implies x = -1.$$

Answer: $(x, y, z) = (-1, -3, 7)$. Check in the third original equation: $3(-1) - (-3) + 2(7) = -3 + 3 + 14 = 14$. ✓

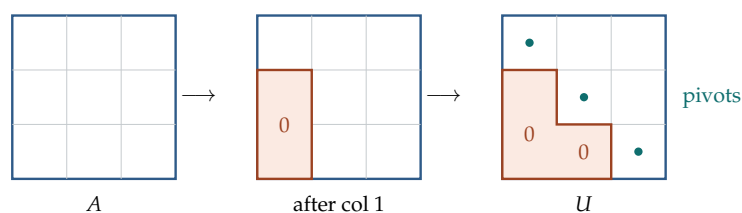


Figure 4.1: Elimination clears one column at a time, below the diagonal. What survives on the diagonal are the pivots; their number is the rank and their product (up to sign) is the determinant.

4.3 Elimination as matrix multiplication

Every row operation is itself a matrix. This is the observation that turns elimination from an algorithm into a *factorization*.

Definition 4.4 Elementary matrices

$E_{ij}(\ell)$ denotes the identity with an extra $-\ell$ in position (i, j) . Multiplying on the left performs $R_i \rightarrow R_i - \ell R_j$. For 3×3 :

$$E_{21}(\ell) = \begin{bmatrix} 1 & 0 & 0 \\ -\ell & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_{21}(\ell)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \ell & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = E_{21}(-\ell).$$

Similarly P_{ij} (the identity with rows i, j swapped) performs a row exchange, and $D_i(c)$ scales row i .

Common Pitfall • left multiplies rows, right multiplies columns

EA performs a *row* operation on A . AE performs a *column* operation. Elimination is always left multiplication.

Theorem 4.2 The LU factorization

If Gaussian elimination on a square A requires no row exchanges, then

$$A = LU$$

where U is the upper triangular result and L is *lower triangular with 1s on the diagonal*, whose (i, j) entry for $i > j$ is exactly the multiplier ℓ_{ij} used to clear position (i, j) .

Why L holds the multipliers. Elimination is $E_k \cdots E_2 E_1 A = U$, so $A = (E_k \cdots E_1)^{-1} U = E_1^{-1} E_2^{-1} \cdots E_k^{-1} U$. Each $E_{ij}(\ell)^{-1}$ is the identity with $+\ell$ in position (i, j) , and — this is the small miracle — when these are multiplied together in this particular order the ℓ 's do not interact: they simply drop into their own slots. So $L = E_1^{-1} \cdots E_k^{-1}$ has ℓ_{ij} in position (i, j) and nothing else off the diagonal. ■

Worked Example 4.2 • LU for the example above

In Worked Example 4.1 the multipliers were $\ell_{21} = 2$, $\ell_{31} = 3$, $\ell_{32} = 7$. Hence

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 3 \\ 3 & -1 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 7 & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & -8 \end{bmatrix}}_U.$$

Verify the $(3, 2)$ entry of LU : row 3 of L is $(3, 7, 1)$, column 2 of U is $(2, -1, 0)$, and $3(2) + 7(-1) + 1(0) = 6 - 7 = -1$. ✓ And the $(3, 3)$ entry: $3(1) + 7(1) + 1(-8) = 3 + 7 - 8 = 2$. ✓

Why bother? Because once you have L and U , solving $Ax = b$ for a *new* b costs two triangular solves instead of a fresh elimination:

$$Lc = b \text{ (forward substitution),} \quad Ux = c \text{ (back substitution).}$$

Elimination costs about $\frac{1}{3}n^3$ operations; each triangular solve costs about n^2 . For repeated right-hand sides the saving is enormous.

4.4 When elimination breaks down

Definition 4.5 Temporary and permanent breakdown

A zero appears in the pivot position. Two cases:

- **Temporary breakdown:** some row *below* has a nonzero entry in that column. Exchange rows and continue. The matrix may still be nonsingular.
- **Permanent breakdown:** the whole column below is zero. No exchange helps; this column has no pivot. If A is square it is *singular*.

Worked Example 4.3 • Both kinds of breakdown

(a) Temporary.

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 3 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix} \xrightarrow{R_3 \rightarrow R_3 - 2R_1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & -1 & 1 \end{bmatrix}$$

$$\xrightarrow{R_3 \rightarrow R_3 + \frac{1}{2}R_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}.$$

Three pivots 1, 2, $\frac{3}{2}$: nonsingular after all.

(b) Permanent.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 8 \\ 1 & 2 & 5 \end{bmatrix} \xrightarrow{R_2 - 2R_1, R_3 - R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 2 \end{bmatrix} \xrightarrow{R_3 - R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Column 2 has nothing to pivot on — the whole column below the first row is zero, and no exchange can fix it. Only 2 pivots, so $\text{rank } A = 2 < 3$: singular. Column 2 is $2 \times$ column 1, which is the dependency causing it.

Theorem 4.3 $PA = LU$

For *any* square matrix A there is a permutation matrix P (a reordering of the rows of I) such that

$$PA = LU$$

with L unit lower triangular and U upper triangular. P records the row exchanges made during elimination. Permutation matrices satisfy $P^{-1} = P^T$.

Remark • partial pivoting

Numerically one exchanges rows even when not strictly necessary, always choosing the *largest* available entry as pivot. This is **partial pivoting**; it keeps the multipliers $|\ell_{ij}| \leq 1$ and stops rounding errors from being amplified. Every serious linear-algebra library does this, and it is why `A\b` in Octave/MATLAB is reliable where naive elimination is not.

4.5 Echelon forms

Definition 4.6 Row echelon form (REF)

A matrix is in **row echelon form** if:

- (i) all zero rows are at the bottom;
- (ii) the leading (leftmost nonzero) entry of each nonzero row — the **pivot** — is strictly to the right of the pivot in the row above.

The columns containing pivots are the **pivot columns** (Tutorial 04 calls them *basic columns*); the others are **free columns**.

Definition 4.7 Reduced row echelon form (RREF)

R is in **reduced row echelon form** if it is in REF and moreover:

- (iii) every pivot equals 1;
- (iv) each pivot is the *only* nonzero entry in its column — cleared above as well as below.

The RREF of a matrix is **unique**; the REF is not.

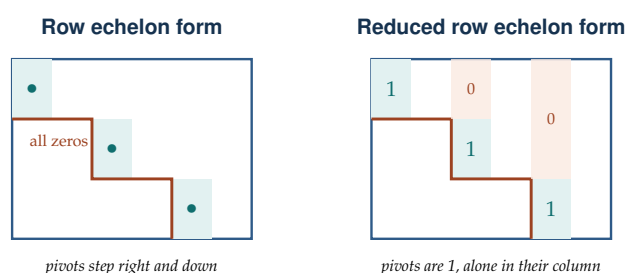


Figure 4.2: Echelon forms. Gaussian elimination stops at REF; Gauss–Jordan continues to RREF.

Worked Example 4.4 • Tutorial 04, Q10(a) — REF and RREF

Reduce $A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 4 & 2 & 2 \\ 3 & 6 & 3 & 4 \end{bmatrix}$ to REF, then to RREF.

To REF. Pivot 1 at $(1, 1)$; $\ell_{21} = 2$, $\ell_{31} = 3$:

$$\begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 4 & 2 & 2 \\ 3 & 6 & 3 & 4 \end{bmatrix} \xrightarrow{R_2 - 2R_1, R_3 - 3R_1} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Pivots sit in columns 1 and 4, so $\text{rank } A = 2$; columns 2 and 3 are free.

To RREF. The pivots are already 1. Clear *above* the second pivot: $R_1 \rightarrow R_1 - R_2$:

$$R = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Reading off the free columns: column 2 is $2 \times$ column 1 and column 3 equals column 1. Those relations hold in the *original* A too — row operations never change the dependencies among columns. Indeed in A : column 2 = $2 \cdot$ column 1 ✓ and column 3 = column 1 ✓.

Intuition • what RREF tells you at a glance

From R alone you can read: the rank (number of pivots), which columns are independent (the pivot columns), how each free column depends on the pivot columns (its entries *are* the coefficients), and the complete solution of $Ax = 0$. Chapter 9 makes each of these precise.

4.6 Gaussian versus Gauss–Jordan

	Gaussian elimination	Gauss–Jordan elimination
Stops at	row echelon form U	reduced row echelon form R
Clears	below each pivot only	below <i>and above</i> each pivot
Pivots	left as they are	scaled to 1
Finish with	back substitution	nothing — read the answer off
Cost ($n \times n$)	$\approx \frac{1}{3}n^3$	$\approx \frac{1}{2}n^3$
Best for	solving one system; LU	finding A^{-1} ; RREF theory

Worked Example 4.5 • Tutorial 04, Q13(b) — Gauss–Jordan on a dependent system

Solve

$$\begin{aligned}x + 2y - 3z &= 2 \\6x + 3y - 9z &= 6 \\7x + 14y - 21z &= 13\end{aligned}$$

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & 2 \\ 6 & 3 & -9 & 6 \\ 7 & 14 & -21 & 13 \end{array} \right] \xrightarrow{R_2 - 6R_1, R_3 - 7R_1} \left[\begin{array}{ccc|c} 1 & 2 & -3 & 2 \\ 0 & -9 & 9 & -6 \\ 0 & 0 & 0 & -1 \end{array} \right]$$

The last row reads $0x + 0y + 0z = -1$, i.e. $0 = -1$. **The system is inconsistent** — there is no solution.

Note how elimination detects this automatically: a pivot appears in the *augmented* column. In rank language, $\text{rank } A = 2$ but $\text{rank}[A \mid \mathbf{b}] = 3$, and

$$\text{rank}[A \mid \mathbf{b}] > \text{rank } A \iff \text{inconsistent}.$$

Geometrically, the third plane is parallel to (but offset from) the plane formed by the first two, since $7 \cdot \text{row}_1 = (7, 14, -21)$ matches the left-hand side of row 3 but $7 \cdot 2 = 14 \neq 13$.

Worked Example 4.6 • Tutorial 04, Q13(a) — Gauss–Jordan with a free variable

Solve

$$\begin{aligned}2x_2 + 3x_3 - 4x_4 &= 1 \\2x_3 + 3x_4 &= 4 \\2x_1 + 2x_2 - 5x_3 + 2x_4 &= 4 \\2x_1 &- 6x_3 + 9x_4 = 7\end{aligned}$$

Order the rows so a pivot appears in column 1:

$$\begin{aligned} & \left[\begin{array}{cccc|c} 2 & 2 & -5 & 2 & 4 \\ 2 & 0 & -6 & 9 & 7 \\ 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \end{array} \right] \xrightarrow{R_2 - R_1} \left[\begin{array}{cccc|c} 2 & 2 & -5 & 2 & 4 \\ 0 & -2 & -1 & 7 & 3 \\ 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \end{array} \right] \\ & \xrightarrow{R_3 + R_2} \left[\begin{array}{cccc|c} 2 & 2 & -5 & 2 & 4 \\ 0 & -2 & -1 & 7 & 3 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 0 & 2 & 3 & 4 \end{array} \right] \xrightarrow{R_4 - R_3} \left[\begin{array}{cccc|c} 2 & 2 & -5 & 2 & 4 \\ 0 & -2 & -1 & 7 & 3 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \end{aligned}$$

Three pivots (columns 1, 2, 3), one free variable (x_4). Continue to RREF: scale to make pivots 1 and clear upward.

$$\begin{aligned} & \xrightarrow{R_1/2, R_2/(-2), R_3/2} \left[\begin{array}{cccc|c} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} & -\frac{3}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ & \xrightarrow{R_2 - \frac{1}{2}R_3, R_1 + \frac{5}{2}R_3} \left[\begin{array}{cccc|c} 1 & 1 & 0 & \frac{19}{4} & 7 \\ 0 & 1 & 0 & -\frac{17}{4} & -\frac{5}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \xrightarrow{R_1 - R_2} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 9 & \frac{19}{2} \\ 0 & 1 & 0 & -\frac{17}{4} & -\frac{5}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \end{aligned}$$

Setting the free variable $x_4 = t$, the complete solution is

$$\mathbf{x} = \begin{bmatrix} \frac{19}{2} \\ -\frac{5}{2} \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} -9 \\ \frac{17}{4} \\ -\frac{3}{2} \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}.$$

The first vector is a *particular solution* ($t = 0$); the second spans the *null space*. This “particular + null” structure is the subject of Chapter 9.

4.7 Systems with parameters

Exam questions very often attach a parameter to one entry and ask for which values the system has 0, 1, or ∞ solutions. The method is always the same: eliminate as usual, then look at the last pivot.

Method Parameter problems

1. Eliminate normally, treating the parameter as a symbol. *Never divide by an expression that could be zero without splitting into cases.*
2. You will reach a last row of the form $(0 \ 0 \ \alpha \mid \beta)$.

3. Then:

- $\alpha \neq 0$: unique solution.
- $\alpha = 0$ and $\beta \neq 0$: no solution.
- $\alpha = 0$ and $\beta = 0$: infinitely many solutions.

Worked Example 4.7 • Tutorial 04, Q11(b) — full case analysis

Find α and β for which

$$3x + y - 6z = -10$$

$$2x + y - 5z = -8$$

$$6x - 3y + \alpha z = \beta$$

has (i) no solution, (ii) a unique solution, (iii) infinitely many.

$$\left[\begin{array}{ccc|c} 3 & 1 & -6 & -10 \\ 2 & 1 & -5 & -8 \\ 6 & -3 & \alpha & \beta \end{array} \right] \xrightarrow{R_2 - \frac{2}{3}R_1, R_3 - 2R_1} \left[\begin{array}{ccc|c} 3 & 1 & -6 & -10 \\ 0 & \frac{1}{3} & -1 & -\frac{4}{3} \\ 0 & -5 & \alpha + 12 & \beta + 20 \end{array} \right]$$

Now $\ell_{32} = (-5)/(\frac{1}{3}) = -15$, so $R_3 \rightarrow R_3 + 15R_2$:

$$\left[\begin{array}{ccc|c} 3 & 1 & -6 & -10 \\ 0 & \frac{1}{3} & -1 & -\frac{4}{3} \\ 0 & 0 & \alpha + 12 - 15 & \beta + 20 - 20 \end{array} \right] = \left[\begin{array}{ccc|c} 3 & 1 & -6 & -10 \\ 0 & \frac{1}{3} & -1 & -\frac{4}{3} \\ 0 & 0 & \alpha - 3 & \beta \end{array} \right]$$

Applying the recipe to the last row $(0 \ 0 \ \alpha - 3 \mid \beta)$:

Case	Condition	Outcome
(ii) unique	$\alpha \neq 3$	one solution, any β
(i) none	$\alpha = 3$ and $\beta \neq 0$	inconsistent
(iii) ∞	$\alpha = 3$ and $\beta = 0$	one free variable

In case (iii), back-substituting with $z = t$ free gives $\frac{1}{3}y = -\frac{4}{3} + t$, so $y = 3t - 4$, and $3x = -10 - y + 6z = -10 - (3t - 4) + 6t = 3t - 6$, so $x = t - 2$:

$$\mathbf{x} = \begin{bmatrix} -2 \\ -4 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}.$$

Worked Example 4.8 • Tutorial 04, Q11(a) — a consistency condition

Show that $2x + y + 3z = a$, $2x + 6y + 8z = b$, $6x + 8y + 14z = c$ has no solutions unless $c = 2a + b$.

$$\left[\begin{array}{ccc|c} 2 & 1 & 3 & a \\ 2 & 6 & 8 & b \\ 6 & 8 & 14 & c \end{array} \right] \xrightarrow{R_2 - R_1, R_3 - 3R_1} \left[\begin{array}{ccc|c} 2 & 1 & 3 & a \\ 0 & 5 & 5 & b - a \\ 0 & 5 & 5 & c - 3a \end{array} \right]$$

$$\xrightarrow{R_3 - R_2} \left[\begin{array}{ccc|c} 2 & 1 & 3 & a \\ 0 & 5 & 5 & b-a \\ 0 & 0 & 0 & c-3a-(b-a) \end{array} \right] = \left[\begin{array}{ccc|c} 2 & 1 & 3 & a \\ 0 & 5 & 5 & b-a \\ 0 & 0 & 0 & c-2a-b \end{array} \right]$$

The last row says $0 = c - 2a - b$. Unless $c = 2a + b$ this is a contradiction and there is no solution. When $c = 2a + b$ the row is $0 = 0$ and, with only two pivots, there is one free variable — infinitely many solutions.

Column-space reading. $\mathbf{C}(A)$ is the plane $\{(a, b, c) : c = 2a + b\}$ in $\mathbb{R}^3$, i.e. $2a + b - c = 0$ with normal $(2, 1, -1)$. The system is solvable exactly for $\mathbf{b} \in \mathbf{C}(A)$ — which is what the algebra just told us.

4.8 Exam preparation

Exam Preparation • Chapter 4

The algorithm in one line

$$[A | \mathbf{b}] \xrightarrow[\text{clear below}]{\text{Gaussian}} [U | \mathbf{c}] \xrightarrow[\text{substitute up}]{\text{back}} \mathbf{x} \quad \text{or} \quad [A | \mathbf{b}] \xrightarrow{\text{Gauss-Jordan}} [R | \mathbf{d}] \text{ read off.}$$

Diagnosing the outcome from the echelon form

Seen in $[U \mathbf{c}]$	Rank condition	Conclusion
row $(0 \cdots 0 \text{nonzero})$	$\text{rank}[A \mathbf{b}] > \text{rank } A$	no solution
n pivots, no bad row	$\text{rank } A = n$	unique solution
$< n$ pivots, no bad row	$\text{rank } A < n$	∞ many; $n - \text{rank } A$ free variables

Facts

- Number of pivots = rank A . Number of free variables = $n - \text{rank } A$.
- RREF is unique; REF and L, U depend on the order of operations.
- $\det A = \pm(\text{product of pivots})$, $\text{sign} = (-1)^{\#\text{row swaps}}$.
- Row operations change the column *space* but never the dependencies *among* columns.

Pitfalls

- Dividing by a parameter without checking it is nonzero. Split into cases *first*.
- Forgetting to carry $\mathbf{b}$ through the operations.
- Treating a $0 = 0$ row as a contradiction (it is not) or a $0 = 5$ row as a free variable (it is not).
- Writing L with the multipliers negated. L stores $+\ell_{ij}$, even though the operation subtracted.

Self-check

1. Solve $x - 2y + z = 0$, $2x - 7y - 3z = 5$, $4x - 7y + z = -1$ by Gaussian elimination. [Tutorial 04, Q12(b)]
2. Factor $A = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix}$ as LU .
3. A 4×6 matrix has 3 pivots. How many free variables does $A\mathbf{x} = \mathbf{0}$ have? Can $A\mathbf{x} = \mathbf{b}$ be inconsistent?
4. Show $A = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$ and $B = \begin{bmatrix} 1 & \tan \theta \\ 0 & 1 \end{bmatrix}$ are row equivalent. [Tutorial 04, Q3]

Answers

1. Eliminating: $R_2 - 2R_1$ gives $(0, -3, -5 \mid 5)$; $R_3 - 4R_1$ gives $(0, 1, -3 \mid -1)$; then $R_3 + \frac{1}{3}R_2$ gives $(0, 0, -\frac{14}{3} \mid \frac{2}{3})$. So $z = -\frac{1}{7}$, $y = \frac{1}{3}(-5 - 5 \cdot \frac{1}{7}) \cdot (-1) \dots$ carefully: $-3y - 5z = 5 \Rightarrow -3y + \frac{5}{7} = 5 \Rightarrow y = -\frac{10}{7}$; then $x = 2y - z = -\frac{20}{7} + \frac{1}{7} = -\frac{19}{7}$. Solution $(-\frac{19}{7}, -\frac{10}{7}, -\frac{1}{7})$.
2. $\ell_{21} = 6/2 = 3$, so $U = \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix}$ and $L = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$. Check: $LU = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix}$. ✓
3. $6 - 3 = 3$ free variables. Yes it can be inconsistent: with 4 rows and only 3 pivots, one row of U is zero, and if the corresponding entry of c is nonzero the system fails.
4. A is invertible ($\det = \cos^2 + \sin^2 = 1$) and so is B ($\det = 1$). Any two invertible matrices of the same size are row equivalent, since both reduce to I . Concretely, $R_1 \rightarrow R_1 / \cos \theta$ then $R_2 \rightarrow R_2 + \sin \theta R_1$ turns A into $\begin{bmatrix} 1 & \tan \theta \\ 0 & \cos \theta + \sin \theta \tan \theta \end{bmatrix} = \begin{bmatrix} 1 & \tan \theta \\ 0 & 1 / \cos \theta \end{bmatrix}$, and scaling R_2 by $\cos \theta$ gives B . (Requires $\cos \theta \neq 0$.)

Inverse Matrices

If A describes an operation, A^{-1} undoes it. This chapter defines the inverse, proves it is unique when it exists, gives the Gauss–Jordan algorithm for computing it, and — most importantly — catalogues the many equivalent ways of saying “ A is invertible”. A warning up front: A^{-1} is a wonderful theoretical object and a poor computational one. Lecture 3 says it bluntly: *do not compute A^{-1} when all you want is to solve $Ax = b$.*

Strang §2.5 (“Inverse Matrices”) and §2.6. Lectures 3 and 5; Tutorial 04 Q1, Q4, Q5, Q8 and Tutorial 05 Q2 are the drill. STRANG, 6th ed.

5.1 Definition and uniqueness

Definition 5.1 Inverse matrix

A square matrix $A \in \mathbb{R}^{n \times n}$ is **invertible** (or **nonsingular**) if there exists a matrix $A^{-1} \in \mathbb{R}^{n \times n}$ with

$$A A^{-1} = I \quad \text{and} \quad A^{-1} A = I.$$

A^{-1} is the **inverse** of A . If no such matrix exists, A is **singular**.

Proposition 5.1 The inverse is unique

If A has an inverse, it has only one.

Proof. Suppose B and C both satisfy $AB = BA = I$ and $AC = CA = I$. Then

$$B = BI = B(AC) = (BA)C = IC = C.$$

Only associativity was used. ■

Remark • one-sided inverses suffice — for square matrices

For square A it is in fact enough to check *one* of the two equations: if $AB = I$ for square A, B then automatically $BA = I$. This fails for rectangular matrices, where a matrix can have a left inverse but no right inverse (or vice versa) — see Chapter 10 on pseudoinverses.

5.2 The 2×2 formula

Theorem 5.1 Inverse of a 2×2 matrix

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then A is invertible if and only if $\det A = ad - bc \neq 0$, and in that case

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

— swap the diagonal entries, negate the off-diagonal entries, divide by the determinant.

Proof. Direct multiplication:

$$\frac{1}{ad - bc} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} ad - bc & -ab + ab \\ cd - cd & -cb + ad \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Conversely if $ad - bc = 0$ then the columns (a, c) and (b, d) are proportional (Tutorial 02 Q1), so they are dependent and A is singular. ■

Worked Example 5.1 • Tutorial 05, Q2(a) — verifying an inverse

Verify that $A = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} -4 & 3 \\ 3 & -2 \end{bmatrix}$ are inverses.

$$AB = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} -8 + 9 & 6 - 6 \\ -12 + 12 & 9 - 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

$$BA = \begin{bmatrix} -4 & 3 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} -8 + 9 & -12 + 12 \\ 6 - 6 & 9 - 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Both products give I , so $B = A^{-1}$. ✓

Consistency check with the formula: $\det A = 2(4) - 3(3) = 8 - 9 = -1$, so

$$A^{-1} = \frac{1}{-1} \begin{bmatrix} 4 & -3 \\ -3 & 2 \end{bmatrix} = \begin{bmatrix} -4 & 3 \\ 3 & -2 \end{bmatrix} = B. \checkmark$$

5.3 Gauss–Jordan: computing A^{-1}

Intuition

The columns of A^{-1} are the solutions of n systems at once: $Ax_j = e_j$ for $j = 1, \dots, n$. Rather than eliminating n separate times, augment A with the whole identity and eliminate once.

Method Gauss–Jordan inversion

1. Form the $n \times 2n$ block $[A \mid I]$.

2. Row reduce all the way to reduced row echelon form.

3. If the left block becomes I , the right block is A^{-1} :

$$[A \mid I] \longrightarrow [I \mid A^{-1}].$$

4. If at any stage a row of the left block becomes entirely zero, stop: A is **singular** and has no inverse.

Why this works. Row reduction is left multiplication by some invertible M (the product of the elementary matrices used). Applying it to the block gives $[MA \mid MI] = [MA \mid M]$. If the left block reaches I then $MA = I$, so $M = A^{-1}$, and the right block — which holds M — is exactly A^{-1} . ■

Worked Example 5.2 • Tutorial 04, Q5(a) — inversion by elimination

Find A^{-1} for $A = \begin{bmatrix} 3 & -3 & 4 \\ 2 & -3 & 4 \\ 0 & -1 & 1 \end{bmatrix}$.

$$[A \mid I] = \left[\begin{array}{ccc|ccc} 3 & -3 & 4 & 1 & 0 & 0 \\ 2 & -3 & 4 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array} \right]$$

Step 1. $R_1 \rightarrow R_1 - R_2$ (a convenient trick: it makes the first pivot 1 without fractions):

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 2 & -3 & 4 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array} \right]$$

Step 2. $R_2 \rightarrow R_2 - 2R_1$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & -3 & 4 & -2 & 3 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array} \right]$$

Step 3. $R_2 \leftrightarrow R_3$, then $R_3 \rightarrow R_3 - 3R_2$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \\ 0 & -3 & 4 & -2 & 3 & 0 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -2 & 3 & -3 \end{array} \right]$$

Step 4. $R_2 \rightarrow R_2 - R_3$, then $R_2 \rightarrow -R_2$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & -1 & 0 & 2 & -3 & 4 \\ 0 & 0 & 1 & -2 & 3 & -3 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -2 & 3 & -4 \\ 0 & 0 & 1 & -2 & 3 & -3 \end{array} \right]$$

Hence

$$A^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 3 & -4 \\ -2 & 3 & -3 \end{bmatrix}.$$

Check. Row 2 of A times column 1 of A^{-1} : $2(1) + (-3)(-2) + 4(-2) = 2 + 6 - 8 = 0$. ✓ Row 3 times column 3: $0(0) + (-1)(-4) + 1(-3) = 4 - 3 = 1$. ✓

Worked Example 5.3 • Lecture 5 — detecting a singular matrix

Attempt to invert $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$.

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array} \right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array} \right]$$

The left block now has a zero row and can never become I . A is **singular**. The reason is visible in the original: column 2 is $2 \times$ column 1, so the columns are dependent, and $\det A = 1(4) - 2(2) = 0$.

The bottom row also reads $0x_1 + 0x_2 = -2$ for the first system — an outright contradiction, confirming $Ax = e_1$ has no solution.

Worked Example 5.4 • Tutorial 04, Q1 — recovering unknown entries

The inverse of $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$ is $A^{-1} = \begin{bmatrix} a & 1 & -1 \\ 1 & b & 1 \\ -1 & 1 & 0 \end{bmatrix}$. Find a and b .

Use $AA^{-1} = I$ and pick the two cheapest entries.

The $(1,1)$ entry of AA^{-1} : row 1 of A is $(1,1,1)$, column 1 of A^{-1} is $(a,1,-1)$:

$$1 \cdot a + 1 \cdot 1 + 1 \cdot (-1) = a = 1 \implies a = 1.$$

The $(2,2)$ entry: row 2 of A is $(1,1,2)$, column 2 of A^{-1} is $(1,b,1)$:

$$1(1) + 1(b) + 2(1) = b + 3 = 1 \implies b = -2.$$

$$\text{So } A^{-1} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 1 \\ -1 & 1 & 0 \end{bmatrix}.$$

Verify with a third entry. $(3,3)$: row 3 of A is $(1,2,3)$, column 3 of A^{-1} is $(-1,1,0)$:
 $1(-1) + 2(1) + 3(0) = 1$. ✓

5.4 Rules for inverses

Theorem 5.2 Algebra of inverses

Let A, B be invertible $n \times n$ matrices and $c \neq 0$ a scalar. Then all of the following are invertible, with

$$\begin{aligned} (A^{-1})^{-1} &= A \\ (cA)^{-1} &= \frac{1}{c}A^{-1} \\ (AB)^{-1} &= B^{-1}A^{-1} && \text{order reverses} \\ (A^T)^{-1} &= (A^{-1})^T && \text{transpose and inverse commute} \\ (A^k)^{-1} &= (A^{-1})^k \end{aligned}$$

Proof. For the product rule, multiply out:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I,$$

and similarly $(B^{-1}A^{-1})(AB) = I$. By uniqueness, $(AB)^{-1} = B^{-1}A^{-1}$.

For the transpose rule, take transposes in $AA^{-1} = I$. Using $(XY)^T = Y^T X^T$ and $I^T = I$:

$$(A^{-1})^T A^T = I,$$

so $(A^{-1})^T$ is a left inverse of A^T ; the same argument on $A^{-1}A = I$ gives it as a right inverse. Hence $(A^T)^{-1} = (A^{-1})^T$. ■

Common Pitfall • the order reversal

$(AB)^{-1} = B^{-1}A^{-1}$, *not* $A^{-1}B^{-1}$. The socks-and-shoes image: to undo “socks then shoes” you must do “shoes off, then socks off”. The same reversal appears in $(AB)^T = B^T A^T$. Marks are routinely lost here.

Also note that $A + B$ invertible does **not** follow from A, B invertible: I and $-I$ are both invertible but $I + (-I) = O$ is not. There is no useful formula for $(A + B)^{-1}$.

Worked Example 5.5 • Powers and negative powers

For $A = \text{diag}(1, -3, 2)$ (Tutorial 04 Q4), the rules give

$$A^{-5} = (A^5)^{-1} = (A^{-1})^5 = \text{diag}(1^{-5}, (-3)^{-5}, 2^{-5}) = \text{diag}(1, -\frac{1}{243}, \frac{1}{32}),$$

matching the direct computation in Chapter 3. Both routes agree because $(A^k)^{-1} = (A^{-1})^k$.

5.5 Triangular and diagonal inverses

Proposition 5.2 Structure is preserved

- $D = \text{diag}(d_1, \dots, d_n)$ is invertible iff every $d_i \neq 0$, and then $D^{-1} = \text{diag}(1/d_1, \dots, 1/d_n)$.
- A triangular matrix is invertible iff every diagonal entry is nonzero, and its inverse is triangular of the same kind (lower stays lower, upper stays upper).
- An orthogonal matrix Q (with $Q^T Q = I$) has $Q^{-1} = Q^T$ — no computation required at all.

Worked Example 5.6 • Inverse of a unit lower triangular matrix

Invert $L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 7 & 1 \end{bmatrix}$ (the L from Chapter 4).

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 3 & 7 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{R_2 - 2R_1, R_3 - 3R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 7 & 1 & -3 & 0 & 1 \end{array} \right]$$

$$\xrightarrow{R_3 - 7R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & 11 & -7 & 1 \end{array} \right]$$

So $L^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 11 & -7 & 1 \end{bmatrix}$ — still unit lower triangular, as promised. But notice the entries

are *not* simply the negatives of those of L : the $(2,1)$ and $(3,2)$ entries are $-\ell_{21} = -2$ and $-\ell_{32} = -7$ as one might guess, yet

$$(L^{-1})_{31} = -\ell_{31} + \ell_{32}\ell_{21} = -3 + 7 \cdot 2 = 11,$$

picking up a product term from the two-step path $3 \rightarrow 2 \rightarrow 1$. Fill-in of this kind is why L^{-1} is rarely formed in practice — forward substitution with L itself is cheaper and exact.

5.6 The Invertible Matrix Theorem

Everything proved so far about square matrices collapses into a single statement.

Theorem 5.3 Invertible Matrix Theorem

For a square matrix $A \in \mathbb{R}^{n \times n}$, the following are **equivalent** — if any one holds, all hold; if any one fails, all fail.

- | | |
|---|---|
| (1) A is invertible. | (9) $\text{rank } A = n$ (full rank). |
| (2) A^{-1} exists and is unique. | (10) A has n nonzero pivots. |
| (3) $Ax = b$ has a unique solution for every $b \in \mathbb{R}^n$. | (11) The RREF of A is I . |
| (4) $Ax = 0$ has only the solution $x = 0$. | (12) $\det A \neq 0$. |
| (5) The columns of A are linearly independent. | (13) $N(A) = \{0\}$, i.e. nullity $A = 0$. |
| (6) The columns of A span $\mathbb{R}^n$. | (14) $C(A) = \mathbb{R}^n$. |
| (7) The columns of A form a basis of $\mathbb{R}^n$. | (15) 0 is <i>not</i> an eigenvalue of A . |
| (8) The rows of A are linearly independent. | (16) A is a product of elementary matrices. |

Intuition • how to use this theorem

In an exam you are typically given one of these sixteen and asked for another. The theorem says you may jump freely between them. In practice:

- To *prove* invertibility of a small explicit matrix, compute $\det A$ (7) or count pivots.
- To *prove* singularity, exhibit a nonzero x with $Ax = 0$, or a dependency among the columns.
- To prove a *theoretical* statement, item (4) — “ $Ax = 0$ forces $x = 0$ ” — is almost always the most convenient.

Worked Example 5.7 • Using item (4)

Prove: if A and B are $n \times n$ and AB is invertible, then both A and B are invertible.

B is invertible. Suppose $Bx = 0$. Then $(AB)x = A(Bx) = A0 = 0$. Since AB is invertible, item (4) gives $x = 0$. So $Bx = 0$ only for $x = 0$, and B is invertible by item (4) again.

A is invertible. Now that B^{-1} exists, write $A = (AB)B^{-1}$, a product of two invertible matrices, hence invertible.

5.7 Why not to use A^{-1}

Common Pitfall • “Do not compute A^{-1} ”

Lecture 3 states this in capital letters, and it is worth understanding why. To solve $Ax = b$:

Method	Cost (flops)	Accuracy
Elimination (LU) then substitute	$\approx \frac{1}{3}n^3$	good
Form A^{-1} , then $A^{-1}b$	$\approx n^3$ (3× more)	worse

The inverse costs about three times as much, and multiplying by it introduces extra rounding error. It also destroys sparsity: the inverse of a sparse matrix is usually dense, so for large problems A^{-1} may not even fit in memory.

A^{-1} remains indispensable *in proofs and formulas* — just not in code. In Octave, write `x = A\b`, not `x = inv(A)*b`.

5.8 Exam preparation

Exam Preparation • Chapter 5

Formulas

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad (ad - bc \neq 0)$$

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A \quad (\text{Chapter 7 — for } 3 \times 3 \text{ and up})$$

$$(AB)^{-1} = B^{-1}A^{-1}, \quad (A^T)^{-1} = (A^{-1})^T, \quad (cA)^{-1} = \frac{1}{c}A^{-1}$$

Algorithm

$$[A \mid I] \xrightarrow{\text{Gauss-Jordan}} [I \mid A^{-1}] \quad \text{or a zero row} \Rightarrow \text{singular.}$$

Pitfalls

- $(AB)^{-1} \neq A^{-1}B^{-1}$. The order reverses.

- $(A + B)^{-1} \neq A^{-1} + B^{-1}$. There is no such rule at all.
- Only *square* matrices have inverses in this sense.
- Row operations on $[A \mid I]$ must be applied to the *whole* row — both blocks — every time.
- Dividing by $\det A$ before checking it is nonzero.

Self-check

1. Find A^{-1} for $A = \begin{bmatrix} 1 & 0 & -3 \\ 2 & -2 & 1 \\ 0 & -1 & 3 \end{bmatrix}$. [Tutorial 05, Q5(a)(i)]
2. If A is invertible and $AB = AC$, show $B = C$. Where exactly does the argument fail if A is singular?
3. Show that if $A^2 = A$ and $A \neq I$, then A is singular.
4. A is 3×3 with $A^3 = O$. Is A invertible? Is $I - A$ invertible?

Answers

1. $\det A = 1(-6 + 1) - 0 + (-3)(-2 - 0) = -5 + 6 = 1$, so $A^{-1} = \begin{bmatrix} -5 & 3 & -6 \\ -6 & 3 & -7 \\ -2 & 1 & -2 \end{bmatrix}$. (Check row 1 of A times column 1 of A^{-1} : $1(-5) + 0(-6) + (-3)(-2) = -5 + 6 = 1$. ✓)
2. Multiply on the left by A^{-1} : $A^{-1}AB = A^{-1}AC$ gives $IB = IC$, so $B = C$. If A is singular no A^{-1} exists and the first step is illegal — and the conclusion genuinely fails, e.g. $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ give $AB = AC = O$ but $B \neq C$.
3. If A were invertible, multiplying $A^2 = A$ by A^{-1} gives $A = I$, contradicting $A \neq I$. So A is singular. (Such A is called *idempotent* — a projection; see Chapter 10.)
4. A is **not** invertible: if it were, $A^3 = O$ would give $A = A^{-2}A^3 = O$, and O is not invertible — contradiction. But $I - A$ **is** invertible, with $(I - A)^{-1} = I + A + A^2$, since $(I - A)(I + A + A^2) = I - A^3 = I$.

PART II

Determinants

The Transpose and Symmetric Matrices

Transposing flips a matrix across its main diagonal. That sounds trivial, and computationally it is — but the transpose is what connects the row world to the column world, and matrices satisfying $S^T = S$ turn out to be the best-behaved objects in the whole subject. Symmetric matrices have real eigenvalues and orthogonal eigenvectors (Chapter 11), and every matrix produces two of them, $A^T A$ and $A A^T$, which is exactly how the SVD is built.

Strang §2.7 (“Transposes and Permutations”), §4.1 for the geometric role of A^T , and §6.3 for symmetric matrices. Lectures 4 and 5; Tutorial 01 Q2 and Tutorial 05 Q1. STRANG | 6th ed.

6.1 The transpose

Definition 6.1 Transpose

The **transpose** of $A \in \mathbb{R}^{m \times n}$ is the matrix $A^T \in \mathbb{R}^{n \times m}$ with

$$(A^T)_{ij} = a_{ji}.$$

Row i of A becomes column i of A^T . The size flips from $m \times n$ to $n \times m$.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \implies A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

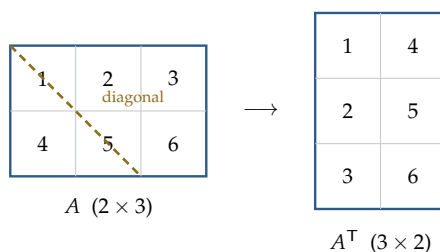


Figure 6.1: Transposing reflects the entries across the main diagonal. Rows become columns.

Theorem 6.1 Properties of the transpose

Whenever the operations are defined:

$$\begin{aligned}
 (A^T)^T &= A \\
 (A + B)^T &= A^T + B^T \\
 (cA)^T &= c A^T \\
 (AB)^T &= B^T A^T && \text{order reverses} \\
 (A^{-1})^T &= (A^T)^{-1} && \text{for invertible } A \\
 \text{tr}(A^T) &= \text{tr } A \\
 \det(A^T) &= \det A && \text{(Chapter 7)}
 \end{aligned}$$

Proof of $(AB)^T = B^T A^T$. Compare entries. Writing $C = AB$,

$$(C^T)_{ij} = c_{ji} = \sum_k a_{jk} b_{ki}.$$

On the other side,

$$(B^T A^T)_{ij} = \sum_k (B^T)_{ik} (A^T)_{kj} = \sum_k b_{ki} a_{jk} = \sum_k a_{jk} b_{ki}.$$

The two agree for all i, j . (Note also that the sizes work out: if A is $m \times n$ and B is $n \times p$, then $(AB)^T$ is $p \times m$, and $B^T A^T$ is $(p \times n)(n \times m) = p \times m$, whereas $A^T B^T$ would be $(n \times m)(p \times n)$ — undefined unless $m = p$.) ■

Proof of $(A^{-1})^T = (A^T)^{-1}$. Transpose both sides of $AA^{-1} = I$:

$$(AA^{-1})^T = I^T \implies (A^{-1})^T A^T = I.$$

Similarly, transposing $A^{-1}A = I$ gives $A^T(A^{-1})^T = I$. So $(A^{-1})^T$ is a two-sided inverse of A^T , and by uniqueness it is $(A^T)^{-1}$. ■

Remark • Lecture 4's phrasing

Lecture 4 states this as: *for an invertible matrix you can change the order of the inverse and transpose*. The common shorthand is A^{-T} , unambiguous precisely because the two orders agree.

Common Pitfall

$(AB)^T = B^T A^T$, not $A^T B^T$. As shown in the proof, $A^T B^T$ is usually not even defined. The same reversal governs $(ABC)^T = C^T B^T A^T$.

6.1.1 The transpose and the dot product

Proposition 6.1 The defining property of A^T

For all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$,

$$(Ax) \cdot y = x \cdot (A^T y)$$

Proof. $(Ax) \cdot y = (Ax)^T y = x^T A^T y = x^T (A^T y) = x \cdot (A^T y)$. ■

Intuition • this is what the transpose is

The formula above is not a property of the transpose — it is the definition, in disguise. A^T is the unique matrix that lets you move A from one side of a dot product to the other. Everything about orthogonality in Chapter 10, and the entire theory of the four fundamental subspaces, follows from this one identity. It is also why $\mathbf{N}(A^T)$ (the left null space) is exactly the set of vectors orthogonal to every column of A .

Worked Example 6.1 • Tutorial 01, Q2 — transposes of sums and products

With

$$A = \begin{bmatrix} -5 & 2 & 0 \\ 7 & -3 & 4 \\ -1 & 3 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & -1 & 8 \\ 6 & -14 & 2 \\ 9 & 5 & 1 \end{bmatrix},$$

verify $(A + B)^T = A^T + B^T$ and compute $(-A)^T$.

$$A^T = \begin{bmatrix} -5 & 7 & -1 \\ 2 & -3 & 3 \\ 0 & 4 & 2 \end{bmatrix}, \quad B^T = \begin{bmatrix} 5 & 6 & 9 \\ -1 & -14 & 5 \\ 8 & 2 & 1 \end{bmatrix}.$$

Then

$$A^T + B^T = \begin{bmatrix} 0 & 13 & 8 \\ 1 & -17 & 8 \\ 8 & 6 & 3 \end{bmatrix}.$$

Directly, $A + B = \begin{bmatrix} 0 & 1 & 8 \\ 13 & -17 & 6 \\ 8 & 8 & 3 \end{bmatrix}$ and transposing gives the same matrix. ✓

Finally $(-A)^T = -A^T = \begin{bmatrix} 5 & -7 & 1 \\ -2 & 3 & -3 \\ 0 & -4 & -2 \end{bmatrix}$: negation and transposition commute, being special cases of $(cA)^T = cA^T$.

6.2 Symmetric matrices

Definition 6.2 Symmetric and skew-symmetric

A square matrix S is **symmetric** if

$$S^T = S, \quad \text{i.e.} \quad s_{ij} = s_{ji} \text{ for all } i, j,$$

and **skew-symmetric** (antisymmetric) if $A^T = -A$, i.e. $a_{ij} = -a_{ji}$. A skew-symmetric matrix necessarily has zeros on the diagonal, since $a_{ii} = -a_{ii}$ forces $a_{ii} = 0$.

$$\underbrace{\begin{bmatrix} 2 & 1 & 5 \\ 1 & 3 & 0 \\ 5 & 0 & -4 \end{bmatrix}}_{\text{symmetric}} \quad \underbrace{\begin{bmatrix} 0 & 1 & -5 \\ -1 & 0 & 2 \\ 5 & -2 & 0 \end{bmatrix}}_{\text{skew-symmetric}}$$

Proposition 6.2 Every matrix splits

Any square A decomposes uniquely as

$$A = \underbrace{\frac{1}{2}(A + A^T)}_{\text{symmetric}} + \underbrace{\frac{1}{2}(A - A^T)}_{\text{skew-symmetric}}.$$

Proof. The two pieces sum to A , and $(\frac{1}{2}(A + A^T))^T = \frac{1}{2}(A^T + A)$ is the same matrix, while $(\frac{1}{2}(A - A^T))^T = \frac{1}{2}(A^T - A)$ is its negative. For uniqueness, if $A = S + K$ with $S^T = S$ and $K^T = -K$, then $A^T = S - K$; adding and subtracting recovers $S = \frac{1}{2}(A + A^T)$ and $K = \frac{1}{2}(A - A^T)$. ■

Theorem 6.2 $A^T A$ is always symmetric

For any $A \in \mathbb{R}^{m \times n}$ (square or not), both

$$A^T A \in \mathbb{R}^{n \times n} \quad \text{and} \quad A A^T \in \mathbb{R}^{m \times m}$$

are symmetric. Moreover both are **positive semidefinite**: $x^T(A^T A)x = \|Ax\|^2 \geq 0$ for every x .

Proof. $(A^T A)^T = A^T(A^T)^T = A^T A$, using the reversal rule and $(A^T)^T = A$. The same argument gives $(A A^T)^T = A A^T$. For semidefiniteness,

$$x^T A^T A x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0.$$

■

Remark • why this theorem matters so much

A itself may be rectangular, complex-eigenvalued and thoroughly badly behaved. But $A^T A$ is always square, symmetric and positive semidefinite — so it has real non-negative eigenvalues and a full set of orthonormal eigenvectors. The SVD (Chapter 13) is built by applying the spectral theorem to $A^T A$ and $A A^T$, then transferring the result back to A . Lecture notes on SVD open with exactly this observation: “We need a square matrix. Matrices $A^T A$ and $A A^T$ will be square. They will also be symmetric.”

Worked Example 6.2 • Tutorial 05, Q1 — solving for entries via $A A^T$

Let $A = \begin{bmatrix} a & 0 & -1 \\ 2 & 3 & b \end{bmatrix}$. Find a and b so that $A A^T = \begin{bmatrix} 26 & -11 \\ -11 & 14 \end{bmatrix}$.

Compute AA^T symbolically. With $A^T = \begin{bmatrix} a & 2 \\ 0 & 3 \\ -1 & b \end{bmatrix}$:

$$AA^T = \begin{bmatrix} a & 0 & -1 \\ 2 & 3 & b \end{bmatrix} \begin{bmatrix} a & 2 \\ 0 & 3 \\ -1 & b \end{bmatrix} = \begin{bmatrix} a^2 + 0 + 1 & 2a + 0 - b \\ 2a + 0 - b & 4 + 9 + b^2 \end{bmatrix} = \begin{bmatrix} a^2 + 1 & 2a - b \\ 2a - b & 13 + b^2 \end{bmatrix}.$$

(Note the off-diagonal entries agree automatically — AA^T must be symmetric.)

Matching entries:

$$a^2 + 1 = 26 \Rightarrow a^2 = 25 \Rightarrow a = \pm 5,$$

$$13 + b^2 = 14 \Rightarrow b^2 = 1 \Rightarrow b = \pm 1,$$

$$2a - b = -11.$$

The last equation selects the signs. Testing: $a = 5, b = 1$ gives $10 - 1 = 9$ ✗; $a = 5, b = -1$ gives $10 + 1 = 11$ ✗; $a = -5, b = 1$ gives $-10 - 1 = -11$ ✓; $a = -5, b = -1$ gives $-10 + 1 = -9$ ✗.

Answer: $a = -5, b = 1$. Verify: $A = \begin{bmatrix} -5 & 0 & -1 \\ 2 & 3 & 1 \end{bmatrix}$ and $AA^T = \begin{bmatrix} 25 + 0 + 1 & -10 + 0 - 1 \\ -10 + 0 - 1 & 4 + 9 + 1 \end{bmatrix} = \begin{bmatrix} 26 & -11 \\ -11 & 14 \end{bmatrix}$. ✓

6.3 Permutation matrices

Definition 6.3 Permutation matrix

P is a **permutation matrix** if it is obtained by reordering the rows of the identity. Equivalently, every row and every column contains exactly one 1 and zeros elsewhere. There are $n!$ permutation matrices of size n .

Proposition 6.3 Permutations are orthogonal

For any permutation matrix P :

$$P^T P = I, \quad \text{so} \quad P^{-1} = P^T, \quad \text{and} \quad \det P = \pm 1.$$

The sign is $+1$ for an even number of row swaps and -1 for an odd number.

Proof. $(P^T P)_{ij}$ is the dot product of columns i and j of P . Distinct columns have their single 1 in different rows, so that dot product is 0; and each column dotted with itself gives 1. Hence $P^T P = I$. ■

Worked Example 6.3 • A 3×3 permutation

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad P^T = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

PA moves row 2 of A to row 1, row 3 to row 2, and row 1 to row 3 — a cyclic shift. And

$$P^T P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I. \checkmark$$

This P is a 3-cycle, achievable by two swaps, so $\det P = (-1)^2 = +1$.

6.4 Exam preparation

Exam Preparation • Chapter 6

Rules

$$(A^T)^T = A, \quad (A+B)^T = A^T + B^T, \quad (cA)^T = cA^T, \\ (AB)^T = B^T A^T, \quad (A^{-1})^T = (A^T)^{-1}, \quad \text{tr}(A^T) = \text{tr } A, \quad \det(A^T) = \det A.$$

Key facts

- $(Ax) \cdot y = x \cdot (A^T y)$ — the transpose moves across a dot product.
- $A^T A$ and AA^T are always symmetric and positive semidefinite, for *any* A .
- S symmetric $\Rightarrow$ real eigenvalues, orthogonal eigenvectors, $S = Q\Lambda Q^T$ (Chapter 12).
- P a permutation $\Rightarrow P^{-1} = P^T$, $\det P = \pm 1$.
- Every square A is uniquely (symmetric) + (skew-symmetric).

Pitfalls

- Writing $(AB)^T = A^T B^T$. Usually not even defined.
- Assuming AB symmetric when A, B are. It is symmetric *only if* $AB = BA$, since $(AB)^T = B^T A^T = BA$.
- Assuming a symmetric matrix must be invertible. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is symmetric and singular.

Self-check

1. If A is 4×7 , what are the sizes of $A^T A$ and AA^T ? Which one can possibly be invertible?
2. Prove that $A^T A = O$ implies $A = O$.
3. If S is symmetric and invertible, show S^{-1} is symmetric.
4. Show that $x^T A y$ is a scalar and equals $y^T A^T x$.

Answers

1. $A^T A$ is 7×7 , AA^T is 4×4 . Since $\text{rank } A \leq \min(4, 7) = 4$, the 7×7 matrix $A^T A$ has rank at most $4 < 7$ and is *always* singular. Only AA^T can be invertible (and is, exactly when A has rank 4).
2. If $A^T A = O$ then for every x , $\|Ax\|^2 = x^T A^T A x = 0$, so $Ax = \mathbf{0}$ for all x . Taking $x = e_j$ shows every column of A is zero, i.e. $A = O$.
3. $(S^{-1})^T = (S^T)^{-1} = S^{-1}$.

4. $\mathbf{x}^T A \mathbf{y}$ is $(1 \times m)(m \times n)(n \times 1) = 1 \times 1$, a scalar. A scalar equals its own transpose, so $\mathbf{x}^T A \mathbf{y} = (\mathbf{x}^T A \mathbf{y})^T = \mathbf{y}^T A^T \mathbf{x}$.

Determinants

The determinant compresses an entire square matrix into a single number that answers one question: *is this matrix invertible?* It answers it by measuring volume. $\det A$ is the signed volume of the box spanned by the columns of A ; the box collapses to zero volume exactly when the columns are dependent. This chapter gives the definition by cofactors, the computational rules, the adjugate formula for A^{-1} , and Cramer's rule — everything Lectures 4–5 and Tutorial 05 cover.

Strang §5.1 (“3 by 3 Determinants and Cofactors”), §5.2 (“Computing and Using Determinants”), §5.3 (“Areas and Volumes”). Lectures 4 and 5; Tutorial 05 is entirely on this chapter. STRANG | 6th ed.

7.1 What the determinant is for

Definition 7.1 Determinant — informally

The **determinant** of a square matrix A , written $\det A$ or $|A|$, is a single real number satisfying

$$A \text{ is invertible} \iff \det A \neq 0$$

It is defined only for square matrices.

Intuition • determinant = signed volume

Place the columns of A as edges of a box from the origin.

- In $\mathbb{R}^2$: $|\det A|$ is the *area* of the parallelogram with edges a_1, a_2 .
- In $\mathbb{R}^3$: $|\det A|$ is the *volume* of the parallelepiped with edges a_1, a_2, a_3 .
- The *sign* records orientation — whether the columns form a right-handed or left-handed frame.

$\det A = 0$ means the box is flat: the columns lie in a lower-dimensional subspace, so they are dependent and A is singular. Every algebraic rule below is a statement about volume.

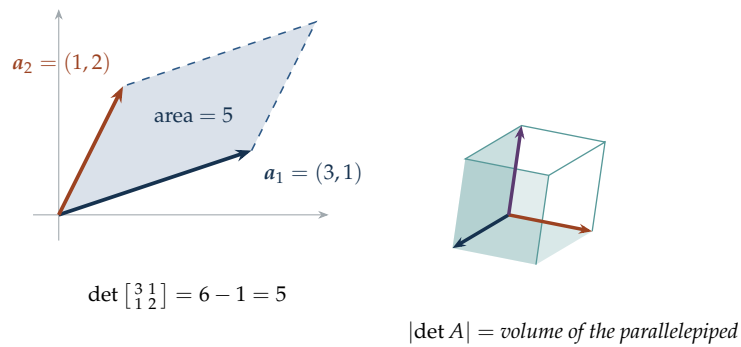


Figure 7.1: The determinant as signed area (2D) and signed volume (3D) spanned by the columns.

7.2 Small determinants

Definition 7.2 1×1 and 2×2

$$\det[a] = a, \quad \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Definition 7.3 3×3 by the rule of Sarrus

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi.$$

Mnemonic: copy the first two columns to the right, add the three “down-right” diagonals, subtract the three “down-left” diagonals.

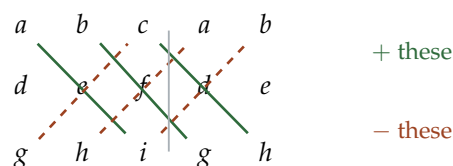


Figure 7.2: Sarrus' rule for 3×3 determinants. *It works only for 3×3 — there is no 4×4 version.*

Common Pitfall • Sarrus does not generalise

The diagonal trick is valid for 2×2 and 3×3 only. For 4×4 and larger you *must* use cofactor expansion or row reduction. Attempting a “ 4×4 Sarrus” produces 8 terms where the correct answer has 24.

7.3 Minors, cofactors and expansion

Definition 7.4 Minor

For an $n \times n$ matrix A , the **minor** M_{ij} is the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting row i and column j of A .

Definition 7.5 Cofactor

The **cofactor** of the entry a_{ij} is the signed minor

$$C_{ij} = (-1)^{i+j} M_{ij}.$$

The sign pattern is a checkerboard beginning with $+$ at the top left:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Theorem 7.1 Cofactor (Laplace) expansion

For any $n \times n$ matrix A and *any* fixed row i or column j :

$$\det A = \sum_{k=1}^n a_{ik} C_{ik} \quad (\text{expansion along row } i),$$

$$\det A = \sum_{k=1}^n a_{kj} C_{kj} \quad (\text{expansion along column } j).$$

Every choice of row or column gives the same answer.

Method Choosing where to expand

Expand along the row or column containing the *most zeros*. Each zero entry kills an entire minor computation. If there are no zeros, use row operations first to create some — adding a multiple of one row to another does not change the determinant (Section 7.4).

Worked Example 7.1 • Lecture 4 — expansion along the best row

Compute $\det A$ for

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

By Sarrus:

$$\begin{aligned} \det A &= 1(5)(9) + 2(6)(7) + 3(4)(8) - 3(5)(7) - 1(6)(8) - 2(4)(9) \\ &= 45 + 84 + 96 - 105 - 48 - 72 = 225 - 225 = 0. \end{aligned}$$

By cofactors along row 1:

$$\det A = 1 \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$$

$$= 1(45 - 48) - 2(36 - 42) + 3(32 - 35) = -3 + 12 - 9 = 0.$$

Same answer. ✓

Why zero? Column 2 is the average of columns 1 and 3: $a_2 = \frac{1}{2}(a_1 + a_3)$. The columns are dependent, so the box is flat and A is singular.

Worked Example 7.2 • Tutorial 05, Q3 — three determinants

Compute

$$A = \begin{bmatrix} -2 & 3 & -2 \\ -4 & -2 & 1 \\ 2 & 4 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -1 & 4 \\ 2 & 5 & 1 \\ 2 & 0 & 6 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 2 & -1 & 1 \\ 2 & 1 & 3 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}.$$

det A, expanding along row 1:

$$\det A = -2 \begin{vmatrix} -2 & 1 \\ 4 & 2 \end{vmatrix} - 3 \begin{vmatrix} -4 & 1 \\ 2 & 2 \end{vmatrix} + (-2) \begin{vmatrix} -4 & -2 \\ 2 & 4 \end{vmatrix}$$

$$= -2(-4 - 4) - 3(-8 - 2) - 2(-16 + 4) = -2(-8) - 3(-10) - 2(-12) = 16 + 30 + 24 = 70.$$

det B, expanding along row 3 (it contains a zero):

$$\det B = 2 \begin{vmatrix} -1 & 4 \\ 5 & 1 \end{vmatrix} - 0 \begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix} + 6 \begin{vmatrix} 3 & -1 \\ 2 & 5 \end{vmatrix}$$

$$= 2(-1 - 20) - 0 + 6(15 + 2) = 2(-21) + 6(17) = -42 + 102 = 60.$$

det C (4 × 4) — reduce first. Use row operations to create zeros in column 1 (these do not change the determinant):

$$\begin{array}{l} R_2 \rightarrow R_2 - 2R_1 \\ R_3 \rightarrow R_3 - 2R_1 \\ R_4 \rightarrow R_4 - R_1 \end{array} \implies \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 1 & 1 & -2 \\ 0 & 1 & -1 & 0 \end{bmatrix}.$$

Expanding along column 1 leaves the single term $1 \cdot C_{11}$:

$$\det C = \begin{vmatrix} 2 & -3 & -1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{vmatrix}.$$

Expand this along row 3 (it has a zero):

$$= 1 \begin{vmatrix} -3 & -1 \\ 1 & -2 \end{vmatrix} - (-1) \begin{vmatrix} 2 & -1 \\ 1 & -2 \end{vmatrix} + 0 = (6 + 1) + (-4 + 1) = 7 - 3 = 4.$$

So $\det C = 4$.

Summary: $\det A = 70$, $\det B = 60$, $\det C = 4$. All three matrices are invertible.

7.4 Rules for determinants

Theorem 7.2 The rules

Let A, B be $n \times n$ and $c \in \mathbb{R}$.

	Rule	Reason (volume)
D1	$\det I = 1$	the unit cube has volume 1
D2	Exchanging two rows <i>reverses the sign</i>	orientation flips
D3	$\det$ is linear in each row separately	scaling one edge scales the volume
D4	$R_i \rightarrow R_i + cR_j$ leaves $\det$ unchanged	shearing preserves volume
D5	Two equal rows $\Rightarrow \det A = 0$	the box is flat
D6	A zero row $\Rightarrow \det A = 0$	an edge of length 0
D7	Triangular $\Rightarrow \det A = \prod a_{ii}$	a box with axis-aligned edges
D8	$\det(AB) = (\det A)(\det B)$	volumes multiply
D9	$\det(A^T) = \det A$	rows and columns play symmetric roles
D10	$\det(cA) = c^n \det A$	all n edges scaled by c
D11	$\det(A^{-1}) = 1/\det A$	from D8 with $B = A^{-1}$

Common Pitfall • the two everyone gets wrong

$$\det(A + B) \neq \det A + \det B \quad \text{and} \quad \det(cA) = c^n \det A \neq c \det A.$$

For the first: $A = B = I_2$ gives $\det(A + B) = \det(2I) = 4$ but $\det A + \det B = 2$. For the second, the exponent n is the *size* of the matrix, and forgetting it is the single most common error in Tutorial 05 Q4.

Worked Example 7.3 • Tutorial 05, Q4(a) — determinant gymnastics

Given $\det A = 4$ where $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ (so $n = 3$), evaluate:

(i) $\det(3A)$

(ii) $\det(2A^{-1})$

(iii) $\det((2A)^{-1})$

(iv) $\begin{vmatrix} b & a & c \\ 2e & 2d & 2f \\ h & g & i \end{vmatrix}$

(v) $\begin{vmatrix} 2a+3d & 2b+3e & 2c+3f \\ d & e & f \\ 5g-7d & 5h-7e & 5i-7f \end{vmatrix}$

(i) By D10 with $c = 3, n = 3$: $\det(3A) = 3^3 \det A = 27 \cdot 4 = \boxed{108}$.

(ii) $\det(2A^{-1}) = 2^3 \det(A^{-1}) = 8 \cdot \frac{1}{4} = \boxed{2}$.

(iii) $\det((2A)^{-1}) = \frac{1}{\det(2A)} = \frac{1}{2^3 \cdot 4} = \frac{1}{32}$. So $\boxed{1/32}$. (Note (ii) $\neq$ (iii): $2A^{-1}$ and $(2A)^{-1} = \frac{1}{2}A^{-1}$ are different matrices.)

(iv) Two changes from A : columns 1 and 2 are swapped (factor -1 by D2 applied to columns, valid by D9), and row 2 is multiplied by 2 (factor 2 by D3). So

$$\det = (-1)(2)(4) = \boxed{-8}.$$

(v) Work row by row. Row 2 is unchanged. Row 1 is $2(\text{row}_1) + 3(\text{row}_2)$: by D3 this contributes a factor 2, and by D4 the added multiple of row 2 changes nothing. Row 3 is $5(\text{row}_3) - 7(\text{row}_2)$: a factor 5, and again the row 2 term is free. Hence

$$\det = 2 \cdot 5 \cdot \det A = 10 \cdot 4 = \boxed{40}.$$

Worked Example 7.4 • Tutorial 05, Q4(b) — a proof-style question

A, B are 2×2 with $\det A = 3$ and $\det B = 2$. Show that $\det((3A^{-1})(B^T)) = 6$.

Apply the rules in order, remembering $n = 2$:

$$\det((3A^{-1})(B^T)) \stackrel{D8}{=} \det(3A^{-1}) \cdot \det(B^T) \stackrel{D9}{=} \det(3A^{-1}) \cdot \det B.$$

Now

$$\det(3A^{-1}) \stackrel{D10}{=} 3^2 \det(A^{-1}) \stackrel{D11}{=} 9 \cdot \frac{1}{3} = 3.$$

Therefore $\det((3A^{-1})(B^T)) = 3 \cdot 2 = 6$. ✓

7.4.1 Determinants by elimination

Method The efficient method for large n

Cofactor expansion on an $n \times n$ matrix costs about $n!$ multiplications — for $n = 10$ that is over three million. Instead:

1. Row reduce A to upper triangular U using only $R_i \rightarrow R_i - \ell R_j$ (which by D4 changes nothing) and row exchanges.
2. Count the row exchanges, say s of them.
3. Then $\det A = (-1)^s \times (\text{product of the pivots})$.

This costs about $\frac{1}{3}n^3$ operations — and it is exactly the elimination you already did in Chapter 4, so the determinant comes for free.

Worked Example 7.5 • Determinant from the pivots

In Chapter 4 we found $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 3 \\ 3 & -1 & 2 \end{bmatrix}$ reduces to $U = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & -8 \end{bmatrix}$ with no row exchanges.

Hence $s = 0$ and

$$\det A = (+1) \cdot 1 \cdot (-1) \cdot (-8) = 8.$$

Cross-check by cofactors along row 1:

$$1 \begin{vmatrix} 3 & 3 \\ -1 & 2 \end{vmatrix} - 2 \begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix} + 1 \begin{vmatrix} 2 & 3 \\ 3 & -1 \end{vmatrix} = (6 + 3) - 2(4 - 9) + (-2 - 9) = 9 + 10 - 11 = 8. \checkmark$$

7.5 The adjugate and a formula for A^{-1}

Definition 7.6 Cofactor matrix and adjugate

The **cofactor matrix** of A is $C = [C_{ij}]$, the matrix of all cofactors. Its transpose is the **adjugate** (Lecture 4 and Tutorial 05 call it the *adjoint*):

$$\text{adj } A = C^T, \quad (\text{adj } A)_{ij} = C_{ji}.$$

Do not forget the transpose — it is the step most often skipped.

Theorem 7.3 Adjugate formula for the inverse

For any square A ,

$$A (\text{adj } A) = (\text{adj } A) A = (\det A) I.$$

Consequently, if $\det A \neq 0$,

$$A^{-1} = \frac{1}{\det A} \text{adj } A$$

Sketch. Entry (i, i) of $A \text{adj } A$ is $\sum_k a_{ik} C_{ik} = \det A$ — exactly the cofactor expansion along row i . Entry (i, j) with $i \neq j$ is $\sum_k a_{ik} C_{jk}$, which is the cofactor expansion of the matrix obtained from A by replacing row j with a copy of row i . That matrix has two equal rows, so by D5 its determinant is 0. Hence $A \text{adj } A = (\det A) I$. ■

Method Finding A^{-1} via the adjugate

1. Compute $\det A$. If it is 0, stop — no inverse.
2. Compute all n^2 minors M_{ij} .
3. Attach signs: $C_{ij} = (-1)^{i+j} M_{ij}$.
4. **Transpose** the cofactor matrix to get $\text{adj } A$.
5. Divide by $\det A$.

Practical for 2×2 and 3×3 ; hopeless beyond that — use Gauss–Jordan (Chapter 5) instead.

Worked Example 7.6 • Tutorial 05, Q5(a)(i) — adjugate and inverse

Find $\text{adj } A$ and A^{-1} for $A = \begin{bmatrix} 1 & 0 & -3 \\ 2 & -2 & 1 \\ 0 & -1 & 3 \end{bmatrix}$.

Step 1 — determinant (expand along row 1):

$$\det A = 1 \begin{vmatrix} -2 & 1 \\ -1 & 3 \end{vmatrix} - 0 + (-3) \begin{vmatrix} 2 & -2 \\ 0 & -1 \end{vmatrix} = 1(-6 + 1) - 3(-2 - 0) = -5 + 6 = 1.$$

Step 2–3 — cofactors.

$$\begin{aligned} C_{11} &= + \begin{vmatrix} -2 & 1 \\ -1 & 3 \end{vmatrix} = -5 & C_{12} &= - \begin{vmatrix} 2 & 1 \\ 0 & 3 \end{vmatrix} = -6 & C_{13} &= + \begin{vmatrix} 2 & -2 \\ 0 & -1 \end{vmatrix} = -2 \\ C_{21} &= - \begin{vmatrix} 0 & -3 \\ -1 & 3 \end{vmatrix} = 3 & C_{22} &= + \begin{vmatrix} 1 & -3 \\ 0 & 3 \end{vmatrix} = 3 & C_{23} &= - \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} = 1 \\ C_{31} &= + \begin{vmatrix} 0 & -3 \\ -2 & 1 \end{vmatrix} = -6 & C_{32} &= - \begin{vmatrix} 1 & -3 \\ 2 & 1 \end{vmatrix} = -7 & C_{33} &= + \begin{vmatrix} 1 & 0 \\ 2 & -2 \end{vmatrix} = -2 \end{aligned}$$

So the cofactor matrix is $C = \begin{bmatrix} -5 & -6 & -2 \\ 3 & 3 & 1 \\ -6 & -7 & -2 \end{bmatrix}$.

Step 4 — transpose.

$$\text{adj } A = C^T = \begin{bmatrix} -5 & 3 & -6 \\ -6 & 3 & -7 \\ -2 & 1 & -2 \end{bmatrix}.$$

Step 5 — divide. Since $\det A = 1$,

$$A^{-1} = \frac{1}{1} \text{adj } A = \begin{bmatrix} -5 & 3 & -6 \\ -6 & 3 & -7 \\ -2 & 1 & -2 \end{bmatrix}.$$

Check. Row 1 of A times column 1 of A^{-1} : $1(-5) + 0(-6) + (-3)(-2) = -5 + 6 = 1$. ✓ Row 2 times column 2: $2(3) + (-2)(3) + 1(1) = 6 - 6 + 1 = 1$. ✓ Row 3 times column 1: $0(-5) + (-1)(-6) + 3(-2) = 6 - 6 = 0$. ✓

Worked Example 7.7 • Tutorial 05, Q5(a)(ii)

Find the inverse of $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 4 \\ 3 & 0 & 1 \end{bmatrix}$.

Determinant (expand along column 2, which has two zeros):

$$\det A = 0 \cdot C_{12} + 1 \cdot C_{22} + 0 \cdot C_{32} = + \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} = 1 - 6 = -5.$$

Cofactors.

$$\begin{aligned} C_{11} &= + \begin{vmatrix} 1 & 4 \\ 0 & 1 \end{vmatrix} = 1 & C_{12} &= - \begin{vmatrix} 0 & 4 \\ 3 & 1 \end{vmatrix} = 12 & C_{13} &= + \begin{vmatrix} 0 & 1 \\ 3 & 0 \end{vmatrix} = -3 \\ C_{21} &= - \begin{vmatrix} 0 & 2 \\ 0 & 1 \end{vmatrix} = 0 & C_{22} &= + \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} = -5 & C_{23} &= - \begin{vmatrix} 1 & 0 \\ 3 & 0 \end{vmatrix} = 0 \\ C_{31} &= + \begin{vmatrix} 0 & 2 \\ 1 & 4 \end{vmatrix} = -2 & C_{32} &= - \begin{vmatrix} 1 & 2 \\ 0 & 4 \end{vmatrix} = -4 & C_{33} &= + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \end{aligned}$$

$$\text{adj } A = C^T = \begin{bmatrix} 1 & 0 & -2 \\ 12 & -5 & -4 \\ -3 & 0 & 1 \end{bmatrix}, \quad A^{-1} = \frac{1}{-5} \begin{bmatrix} 1 & 0 & -2 \\ 12 & -5 & -4 \\ -3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} & 0 & \frac{2}{5} \\ -\frac{12}{5} & 1 & \frac{4}{5} \\ \frac{3}{5} & 0 & -\frac{1}{5} \end{bmatrix}.$$

Check the $(1, 1)$ entry of AA^{-1} : $1(-\frac{1}{5}) + 0(-\frac{12}{5}) + 2(\frac{3}{5}) = -\frac{1}{5} + \frac{6}{5} = 1$. ✓

7.6 Cramer's rule

Theorem 7.4 Cramer's rule

Let A be $n \times n$ with $\det A \neq 0$, and let A_j denote the matrix A with its j th column replaced by $\mathbf{b}$. Then the unique solution of $A\mathbf{x} = \mathbf{b}$ has components

$$x_j = \frac{\det A_j}{\det A}, \quad j = 1, \dots, n.$$

Proof. From $A^{-1} = \frac{1}{\det A} \text{adj } A$ we get $\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det A} (\text{adj } A)\mathbf{b}$, so

$$x_j = \frac{1}{\det A} \sum_{k=1}^n (\text{adj } A)_{jk} b_k = \frac{1}{\det A} \sum_{k=1}^n C_{kj} b_k.$$

The last sum is precisely the cofactor expansion of $\det A_j$ along its j th column (which holds $\mathbf{b}$), because deleting that column and row k from A_j gives the same minor as deleting column j and row k from A . ■

Worked Example 7.8 • Tutorial 05, Q6(b)(i) — Cramer's rule

Solve

$$4x + 8y + z = -6$$

$$2x - 3y + 2z = 0$$

$$x + 7y - 3z = -8$$

The coefficient determinant.

$$\det A = \begin{vmatrix} 4 & 8 & 1 \\ 2 & -3 & 2 \\ 1 & 7 & -3 \end{vmatrix} = 4 \begin{vmatrix} -3 & 2 \\ 7 & -3 \end{vmatrix} - 8 \begin{vmatrix} 2 & 2 \\ 1 & -3 \end{vmatrix} + 1 \begin{vmatrix} 2 & -3 \\ 1 & 7 \end{vmatrix}$$

$$= 4(9 - 14) - 8(-6 - 2) + (14 + 3) = 4(-5) + 64 + 17 = -20 + 64 + 17 = 61.$$

Nonzero, so Cramer applies.

Replace column 1 by $b = (-6, 0, -8)$:

$$\begin{aligned}\det A_1 &= \begin{vmatrix} -6 & 8 & 1 \\ 0 & -3 & 2 \\ -8 & 7 & -3 \end{vmatrix} = -6 \begin{vmatrix} -3 & 2 \\ 7 & -3 \end{vmatrix} - 8 \begin{vmatrix} 0 & 2 \\ -8 & -3 \end{vmatrix} + 1 \begin{vmatrix} 0 & -3 \\ -8 & 7 \end{vmatrix} \\ &= -6(9 - 14) - 8(0 + 16) + (0 - 24) = 30 - 128 - 24 = -122.\end{aligned}$$

Replace column 2:

$$\begin{aligned}\det A_2 &= \begin{vmatrix} 4 & -6 & 1 \\ 2 & 0 & 2 \\ 1 & -8 & -3 \end{vmatrix} = 4 \begin{vmatrix} 0 & 2 \\ -8 & -3 \end{vmatrix} + 6 \begin{vmatrix} 2 & 2 \\ 1 & -3 \end{vmatrix} + 1 \begin{vmatrix} 2 & 0 \\ 1 & -8 \end{vmatrix} \\ &= 4(0 + 16) + 6(-6 - 2) + (-16 - 0) = 64 - 48 - 16 = 0.\end{aligned}$$

Replace column 3:

$$\begin{aligned}\det A_3 &= \begin{vmatrix} 4 & 8 & -6 \\ 2 & -3 & 0 \\ 1 & 7 & -8 \end{vmatrix} = 4 \begin{vmatrix} -3 & 0 \\ 7 & -8 \end{vmatrix} - 8 \begin{vmatrix} 2 & 0 \\ 1 & -8 \end{vmatrix} - 6 \begin{vmatrix} 2 & -3 \\ 1 & 7 \end{vmatrix} \\ &= 4(24 - 0) - 8(-16 - 0) - 6(14 + 3) = 96 + 128 - 102 = 122.\end{aligned}$$

Solution.

$$x = \frac{-122}{61} = -2, \quad y = \frac{0}{61} = 0, \quad z = \frac{122}{61} = 2.$$

Check in equation 3: $(-2) + 7(0) - 3(2) = -2 - 6 = -8$. ✓

Common Pitfall • Cramer's rule is a formula, not a method

It requires $n + 1$ determinants of size n , costing roughly $(n + 1)!$ operations by cofactors — astronomically worse than elimination's $\frac{1}{3}n^3$. Use it when

- $n \leq 3$ and you want a quick answer,
- you need only *one* component x_j ,
- you want a *symbolic* formula showing how x_j depends on the data.

Otherwise, eliminate. It also fails outright when $\det A = 0$ — it says nothing about the singular case.

7.7 Determinants and singularity with a parameter

Worked Example 7.9 • Tutorial 05, Q7 — determinant as a function of a

Let $A = \begin{bmatrix} 1 & 2 & -3 \\ 3 & -1 & 5 \\ 4 & 1 & a^2 - 14 \end{bmatrix}$. Find $\det A$ in terms of a , then determine for which a the system

$$\begin{aligned} x + 2y - 3z &= 4 \\ 3x - y + 5z &= 2 \\ 4x + y + (a^2 - 14)z &= a + 2 \end{aligned}$$

has (i) exactly one solution, (ii) infinitely many, (iii) none.

Determinant. Expand along row 1:

$$\begin{aligned} \det A &= 1 \begin{vmatrix} -1 & 5 \\ 1 & a^2 - 14 \end{vmatrix} - 2 \begin{vmatrix} 3 & 5 \\ 4 & a^2 - 14 \end{vmatrix} + (-3) \begin{vmatrix} 3 & -1 \\ 4 & 1 \end{vmatrix} \\ &= (-(a^2 - 14) - 5) - 2(3(a^2 - 14) - 20) - 3(3 + 4) \\ &= (-a^2 + 9) - 2(3a^2 - 62) - 21 = -a^2 + 9 - 6a^2 + 124 - 21 = -7a^2 + 112. \end{aligned}$$

So

$$\det A = -7(a^2 - 16) = -7(a - 4)(a + 4).$$

(i) **Exactly one solution** when $\det A \neq 0$, i.e. $a \neq \pm 4$.

The singular cases. When $a^2 = 16$ we have $a^2 - 14 = 2$, so the matrix becomes

$\begin{bmatrix} 1 & 2 & -3 \\ 3 & -1 & 5 \\ 4 & 1 & 2 \end{bmatrix}$ and row 3 equals row 1 + row 2. Consistency therefore requires the same relation on the right-hand side:

$$b_3 = b_1 + b_2 \iff a + 2 = 4 + 2 = 6 \iff a = 4.$$

(ii) **Infinitely many solutions** when $a = 4$: the third equation is redundant, leaving 2 pivots and 1 free variable.

(iii) **No solution** when $a = -4$: then $a + 2 = -2 \neq 6$, so the third equation contradicts the first two.

Worked Example 7.10 • Tutorial 05, Q5(b) — forcing singularity

Find all x making $A = \begin{bmatrix} x & 1 & 1 \\ 1 & x & 1 \\ 1 & 1 & x \end{bmatrix}$ singular.

Singular $\iff \det A = 0$. Expanding,

$$\det A = x(x^2 - 1) - 1(x - 1) + 1(1 - x) = x^3 - x - x + 1 + 1 - x = x^3 - 3x + 2.$$

Factor: $x = 1$ is a root, so

$$x^3 - 3x + 2 = (x - 1)(x^2 + x - 2) = (x - 1)(x - 1)(x + 2) = (x - 1)^2(x + 2).$$

Hence A is singular for $x = 1$ or $x = -2$.

Sanity check. At $x = 1$ all three rows are $(1, 1, 1)$ — rank 1, certainly singular. At $x = -2$ the rows are $(-2, 1, 1), (1, -2, 1), (1, 1, -2)$, which sum to $(0, 0, 0)$ — a dependency, again singular.

7.8 Exam preparation

Core formulas

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \quad C_{ij} = (-1)^{i+j} M_{ij}, \quad \det A = \sum_k a_{ik} C_{ik}$$

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A, \quad \operatorname{adj} A = C^T, \quad x_j = \frac{\det A_j}{\det A} \text{ (Cramer)}$$

$$\det(AB) = \det A \det B, \quad \det(A^T) = \det A, \quad \det(cA) = c^n \det A, \quad \det(A^{-1}) = \frac{1}{\det A}$$

Row operations and det

$R_i \rightarrow R_i + cR_j$	det unchanged
$R_i \leftrightarrow R_j$	det changes sign
$R_i \rightarrow cR_i$	det multiplied by c

Pitfalls

- $\det(A + B) \neq \det A + \det B$.
- $\det(cA) = c^n \det A$ — the exponent is the matrix size.
- Forgetting the transpose in $\operatorname{adj} A = C^T$.
- Sign errors in $(-1)^{i+j}$. Write out the checkerboard first.
- Sarrus' rule beyond 3×3 .
- Applying Cramer's rule when $\det A = 0$.

Self-check

1. Compute $\begin{vmatrix} 2 & 0 & 1 \\ 3 & -1 & 2 \\ 1 & 4 & 0 \end{vmatrix}$ two ways.

2. If A is 4×4 with $\det A = 5$, find $\det(2A)$, $\det(A^T A)$ and $\det(A^{-1})$.

3. Solve $-7x - 6y - 12z = 0$, $5x + 5y + 7z = 0$, $x + 4z = 0$ using Cramer's rule. [Tutorial 05, Q6(b)(ii)]

4. Prove that if A is $n \times n$ and skew-symmetric with n odd, then $\det A = 0$.

Answers

1. Along row 1: $2(0 - 8) - 0 + 1(12 + 1) = -16 + 13 = -3$. By Sarrus: $0 + 0 + 12 - (-1) - 16 - 0 = -3$. ✓

2. $\det(2A) = 2^4 \cdot 5 = 80$; $\det(A^T A) = \det(A^T) \det A = 5 \cdot 5 = 25$; $\det(A^{-1}) = 1/5$.

3. The system is homogeneous. Its determinant is $\begin{vmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{vmatrix} = -7(20 - 0) + 6(20 - 7) - 12(0 - 5) = -140 + 78 + 60 = -2$. Nonzero, so the only solution is the trivial one: $x = y = z = 0$. (Cramer confirms this: each A_j has a zero column, so every $\det A_j = 0$.)

4. $A^T = -A$ gives $\det A = \det(A^T) = \det(-A) = (-1)^n \det A = -\det A$ for odd n . Hence $2 \det A = 0$, so $\det A = 0$.

PART III

Vector Spaces and the Four Subspaces

Vector Spaces and Subspaces

Up to here, “vector” meant a column of numbers. This chapter performs the abstraction that makes linear algebra powerful: we keep only the *rules* of vector arithmetic and forget what the objects are. Anything obeying the ten axioms — polynomials, functions, matrices, signals — becomes a vector space, and every theorem proved from the axioms applies to all of them at once. We then meet spanning sets, independence, bases and dimension, which together answer “how big is this space, and what is the most economical way to describe it?”

Strang §3.1 (“Spaces of Vectors”), §3.2 (“Computing the Nullspace”), §3.4 (“Independence, Basis and Dimension”). Lecture 7; Tutorial 06 Q1(d) uses subspace verification. STRANG | 6th.ed.

8.1 The axioms

Definition 8.1 Vector space

A **vector space** is a non-empty set V together with two operations — *vector addition* $\mathbf{u} + \mathbf{v}$ and *scalar multiplication* $k\mathbf{u}$ with $k \in \mathbb{R}$ — satisfying the following ten axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars $k, c \in \mathbb{R}$.

Addition		Scalar multiplication	
1	$\mathbf{u} + \mathbf{v} \in V$ (closure)	6.	$k\mathbf{u} \in V$ (closure)
2	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (commutative)	7.	$k(c\mathbf{u}) = (kc)\mathbf{u}$ (associative)
3	$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (associative)	8.	$k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$ (distributive)
4	$\exists \mathbf{0} \in V$ with $\mathbf{u} + \mathbf{0} = \mathbf{u}$ (neutral element)	9.	$(k + c)\mathbf{u} = k\mathbf{u} + c\mathbf{u}$ (distributive)
5	$\exists (-\mathbf{u})$ with $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ (additive inverse)	10.	$1\mathbf{u} = \mathbf{u}$ (identity)

Intuition • Lecture 7’s football-field picture

A vector space is a playing field you cannot leave. Add two vectors — still in the field. Scale a vector by any real number — still in the field. If any legal operation kicks you off the field, it is not a vector space. Axioms 1 and 6, the two *closure* axioms, are the ones that do real work; the other eight are usually inherited automatically from the arithmetic of real numbers.

8.1.1 Examples — and non-examples

Set	Vector space?	Why
$\mathbb{R}^n$	yes	the model case
$\mathbb{R}^{m \times n}$ (all $m \times n$ matrices)	yes	add and scale entrywise; $\dim = mn$
$\mathbb{P}_n$ (polynomials of degree $\leq n$)	yes	$\dim = n + 1$, basis $1, t, \dots, t^n$
$C[0, 1]$ (continuous functions)	yes	infinite-dimensional
$\{0\}$	yes	the zero space, $\dim = 0$
Polynomials of degree <i>exactly</i> n	no	$t^n + (-t^n) = 0$ has degree 0: not closed
$\{(x, y) : x \geq 0\}$ (a half-plane)	no	$-1 \cdot (1, 0) = (-1, 0)$ escapes
$\{(x, y) : xy = 0\}$ (the two axes)	no	$(1, 0) + (0, 1) = (1, 1)$ escapes
$\{(x, y) : x + y = 1\}$ (a line off the origin)	no	does not contain 0
Invertible $n \times n$ matrices	no	$A + (-A) = O$ is not invertible

Proposition 8.1 Consequences of the axioms

In any vector space:

$$0u = 0, \quad k0 = 0, \quad (-1)u = -u, \quad ku = 0 \Rightarrow k = 0 \text{ or } u = 0.$$

Proof. For the first: $0u = (0 + 0)u = 0u + 0u$ by axiom 9. Adding $-(0u)$ to both sides gives $0 = 0u$. For the third: $u + (-1)u = 1u + (-1)u = (1 - 1)u = 0u = 0$, so $(-1)u$ is the additive inverse of u . ■

8.2 Subspaces

Definition 8.2 Subspace

A subset $W \subseteq V$ is a **subspace** of the vector space V if W is itself a vector space under the operations inherited from V .

Checking all ten axioms would be tedious; almost all of them are inherited for free. Only closure and the presence of 0 need checking.

Theorem 8.1 Subspace test

A subset $W \subseteq V$ is a subspace **if and only if** all three hold:

(S1) $0 \in W$ (equivalently, W is non-empty)

(S2) $u, v \in W \Rightarrow u + v \in W$ (closed under addition)

(S3) $u \in W, k \in \mathbb{R} \Rightarrow ku \in W$ (closed under scalar multiplication)

Conditions (S2) and (S3) can be merged into the single test

$$u, v \in W \text{ and } k, c \in \mathbb{R} \implies ku + cv \in W.$$

Method Checking a subspace in an exam

1. **Test 0 first.** It is one line, and it eliminates most non-subspaces immediately (anything

defined by an equation with a nonzero constant fails here).

2. Take general $u, v \in W$ — write down what membership *means* for them.
3. Show $u + v$ satisfies the same condition.
4. Show ku does too.

To prove a set is *not* a subspace, one explicit counterexample is enough.

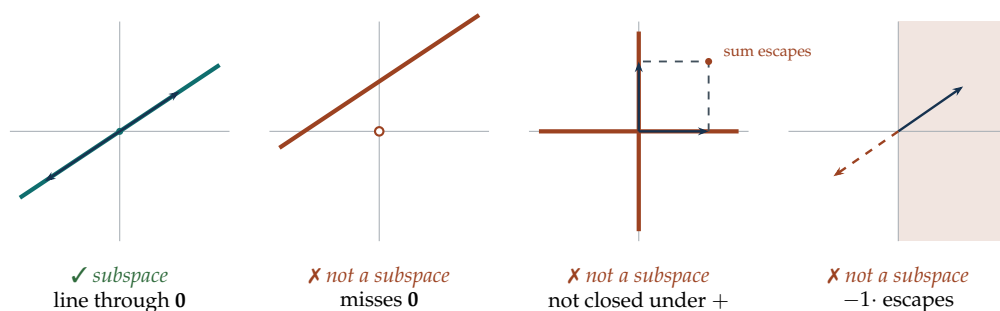


Figure 8.1: Subspaces of $\mathbb{R}^2$ are exactly: $\{0\}$, any line through the origin, and $\mathbb{R}^2$ itself. Nothing else qualifies.

Worked Example 8.1 • Subspace or not?

Decide whether each is a subspace of the given space.

- (a) $W = \{(x, y, z) \in \mathbb{R}^3 : x + 2y - z = 0\}$
- (b) $W = \{(x, y, z) \in \mathbb{R}^3 : x + 2y - z = 3\}$
- (c) $W = \{A \in \mathbb{R}^{2 \times 2} : A^T = A\}$
- (d) $W = \{(x, y) \in \mathbb{R}^2 : x^2 = y^2\}$

(a) Yes. $0 = (0, 0, 0)$ gives $0 + 0 - 0 = 0 \checkmark$. If $u, v \in W$ then $u_1 + 2u_2 - u_3 = 0$ and $v_1 + 2v_2 - v_3 = 0$; adding, $(u_1 + v_1) + 2(u_2 + v_2) - (u_3 + v_3) = 0$, so $u + v \in W$. Scaling by k multiplies the equation by k , still $= 0$. This W is a plane through the origin. *Any* set defined by homogeneous linear equations is a subspace — it is a null space (Chapter 9).

(b) No. 0 gives $0 \neq 3$. Fails at the first hurdle. This is a plane parallel to the one in (a) but shifted off the origin — an *affine* set, not a subspace.

(c) Yes. $O^T = O \checkmark$. If $A^T = A$ and $B^T = B$ then $(A + B)^T = A^T + B^T = A + B \checkmark$, and $(kA)^T = kA^T = kA \checkmark$. The symmetric 2×2 matrices form a 3-dimensional subspace of the 4-dimensional space $\mathbb{R}^{2 \times 2}$ (free choices: the two diagonal entries and the one shared off-diagonal entry).

(d) No. $x^2 = y^2$ means $y = \pm x$ — the union of two lines. $(1, 1)$ and $(1, -1)$ both lie in W , but their sum $(2, 0)$ has $4 \neq 0$. Not closed under addition. (Note the defining equation is not linear, which is the tell.)

Proposition 8.2 Intersections and sums

If U and W are subspaces of V , then

- $U \cap W$ is a subspace;
- $U \cup W$ is **usually not** (see the two-axes picture above);

- $U + W = \{u + w : u \in U, w \in W\}$ is a subspace — the smallest one containing both.

8.3 Span

Definition 8.3 Span

The **span** of vectors $v_1, \dots, v_k \in V$ is the set of all their linear combinations:

$$\text{span}\{v_1, \dots, v_k\} = \{c_1 v_1 + \dots + c_k v_k : c_1, \dots, c_k \in \mathbb{R}\}.$$

If $\text{span}\{v_1, \dots, v_k\} = V$ we call $\{v_1, \dots, v_k\}$ a **spanning set** for V , and say the vectors *span* V .

Proposition 8.3

$\text{span}\{v_1, \dots, v_k\}$ is always a subspace of V — indeed the *smallest* subspace containing all the v_i .

Proof. It contains $\mathbf{0}$ (all $c_i = 0$), and a combination of two combinations is again a combination, so it is closed under both operations. Minimality: any subspace containing every v_i must, by closure, contain all their combinations. ■

8.4 Linear independence

Definition 8.4 Linear independence

Vectors $v_1, \dots, v_k$ are **linearly independent** if

$$c_1 v_1 + c_2 v_2 + \dots + c_k v_k = \mathbf{0} \implies c_1 = c_2 = \dots = c_k = 0.$$

That is, the *only* combination producing $\mathbf{0}$ is the trivial one. Otherwise they are **linearly dependent**, and some non-trivial combination vanishes.

Proposition 8.4 Equivalent characterisations

For $v_1, \dots, v_k$ the following are equivalent:

- (i) they are linearly dependent;
- (ii) at least one v_j is a linear combination of the others;
- (iii) some v_j lies in the span of the preceding ones;
- (iv) the matrix $A = [v_1 \ \dots \ v_k]$ has $\mathbf{N}(A) \neq \{\mathbf{0}\}$;
- (v) $\text{rank } A < k$.

Method Testing independence in $\mathbb{R}^n$

Form the matrix A whose *columns* are the vectors and row reduce.

- k **pivots** (one per column) $\Rightarrow$ independent.

- **Fewer than k pivots** $\Rightarrow$ dependent; the free columns tell you the dependencies.

Shortcuts: any set containing $\mathbf{0}$ is dependent; any set of more than n vectors in $\mathbb{R}^n$ is dependent; two vectors are dependent iff one is a multiple of the other. For $k = n$ (square A), independent $\iff \det A \neq 0$.

Worked Example 8.2 • Assignment 01, Q2(b) — independence of two vectors

Show $\mathbf{a} = (1, 0, -1)$ and $\mathbf{b} = (0, 2, 1)$ are linearly independent.

Suppose $c_1\mathbf{a} + c_2\mathbf{b} = \mathbf{0}$. Componentwise:

$$c_1 = 0, \quad 2c_2 = 0, \quad -c_1 + c_2 = 0.$$

The first gives $c_1 = 0$, the second $c_2 = 0$. Only the trivial combination works, so they are independent.

Faster argument. Two vectors are dependent only if one is a multiple of the other. Here $\mathbf{a}$ has second component 0 while $\mathbf{b}$ has second component 2; any multiple $k\mathbf{a}$ has second component $0 \neq 2$, so $\mathbf{b} \neq k\mathbf{a}$, and $\mathbf{a} \neq \mathbf{0}$ so $\mathbf{a}$ is not a multiple of $\mathbf{b}$ either. Independent.

Common Pitfall • independence is a property of the set

Do not say “ $\mathbf{v}$ is independent”. Independence is about how vectors relate to *each other*. A single nonzero vector forms an independent set; a single zero vector does not.

Also: independent $\neq$ orthogonal. Orthogonal nonzero vectors are always independent (Chapter 10), but independent vectors need not be orthogonal — $(1, 0)$ and $(1, 1)$ are independent at 45° .

8.5 Basis and dimension

Definition 8.5 Basis

A **basis** of a vector space V is a set of vectors that is

- (i) **linearly independent**, and
- (ii) **spans** V .

Equivalently: a basis is a *minimal* spanning set, or a *maximal* independent set.

Theorem 8.2 Unique representation

$\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of V if and only if every $\mathbf{v} \in V$ has **exactly one** expression

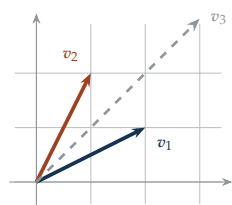
$$\mathbf{v} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n.$$

The scalars c_i are the **coordinates** of $\mathbf{v}$ in this basis.

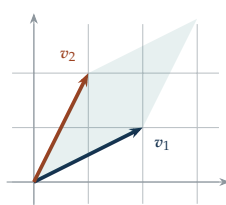
Proof. Spanning gives *at least* one expression. For uniqueness, suppose $\sum c_i\mathbf{v}_i = \sum d_i\mathbf{v}_i$. Subtracting, $\sum (c_i - d_i)\mathbf{v}_i = \mathbf{0}$, and independence forces $c_i - d_i = 0$ for every i . Conversely, if representations are unique then in particular $\mathbf{0}$ has only the trivial one, which is independence. ■

Intuition • Lecture 7's distinction

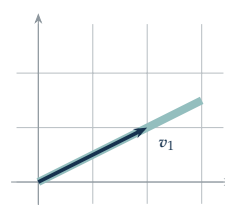
A *spanning set* is any collection of vectors that can generate the whole space. A *basis* is the smallest such collection, with no unnecessary vectors. Spanning sets can be wastefully large; bases are exactly right. Throw away one vector from a basis and you no longer span; add one and you lose independence.



spanning set $\{v_1, v_2, v_3\}$
spans $\mathbb{R}^2$ but is dependent



basis $\{v_1, v_2\}$
spans and is independent



$\{v_1\}$ alone
independent but spans only a line

Figure 8.2: Too many vectors, exactly right, too few. A basis is the balance point between spanning and independence.

Theorem 8.3 Dimension is well defined

Every basis of a vector space V has the *same* number of vectors. That number is the **dimension** $\dim V$.

Sketch — the Steinitz exchange. Suppose $\{u_1, \dots, u_m\}$ spans V and $\{w_1, \dots, w_p\}$ is independent in V . Expressing each w_j in terms of the u_i gives an $m \times p$ matrix; if $p > m$ this matrix has more columns than rows, hence a nonzero null vector, which produces a non-trivial combination of the w_j equal to $\mathbf{0}$ — contradicting independence. So $p \leq m$: any independent set is no larger than any spanning set. Applying this both ways to two bases gives equality. ■

Space	Dimension	A standard basis
$\mathbb{R}^n$	n	$e_1, \dots, e_n$
$\mathbb{R}^{m \times n}$	mn	the matrices E_{ij} with a single 1
$\mathbb{P}_n$	$n + 1$	$1, t, t^2, \dots, t^n$
symmetric $n \times n$	$\frac{n(n+1)}{2}$	E_{ii} and $E_{ij} + E_{ji}$
$\{0\}$	0	the empty set
$C[0, 1]$	∞	no finite basis exists

Theorem 8.4 Counting theorem

Let $\dim V = n$. Then:

- (i) Any independent set has at most n vectors.
- (ii) Any spanning set has at least n vectors.
- (iii) Any independent set of *exactly* n vectors is a basis.
- (iv) Any spanning set of *exactly* n vectors is a basis.
- (v) Any independent set can be *extended* to a basis.
- (vi) Any spanning set can be *reduced* to a basis.

Remark • a genuine labour-saver

Items (iii) and (iv) mean that if you already know the dimension, you only need to check *one* of the two basis conditions. To show three vectors form a basis of $\mathbb{R}^3$, checking independence suffices — and for three vectors in $\mathbb{R}^3$ that is a single determinant.

Worked Example 8.3 • Finding a basis and dimension of a span

Find a basis and the dimension of $W = \text{span}\{(1, 2, 1), (2, 4, 2), (1, 0, 1), (3, 2, 3)\} \subseteq \mathbb{R}^3$.

Put the vectors as *columns* and row reduce:

$$A = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 2 & 4 & 0 & 2 \\ 1 & 2 & 1 & 3 \end{bmatrix} \xrightarrow{R_2 - 2R_1, R_3 - R_1} \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 0 & -2 & -4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_2 / (-2)} \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Pivots are in columns 1 and 3. So:

- **Basis:** the corresponding *original* columns, $\{(1, 2, 1), (1, 0, 1)\}$.
- **Dimension:** $\dim W = 2$ — W is a plane in $\mathbb{R}^3$.

The free columns record the dependencies: column 2 = $2 \times$ column 1, and column 4 = column 1 + $2 \times$ column 3. Check the latter: $(1, 2, 1) + 2(1, 0, 1) = (3, 2, 3)$. ✓

Common Pitfall • take the original columns, not the reduced ones

The pivot columns of the *RREF* are e_1, e_2 — they span a different plane. Row operations preserve which columns are independent, but change what those columns *are*. Always go back to A .

8.6 Exam preparation

Exam Preparation • Chapter 8

Definitions to state verbatim

- **Subspace:** non-empty, closed under $+$ and scalar multiplication (equivalently contains $\mathbf{0}$ and all $ku + cv$).
- **Span:** the set of all linear combinations.
- **Independent:** $\sum c_i v_i = \mathbf{0} \Rightarrow$ all $c_i = 0$.
- **Basis:** independent *and* spanning.
- **Dimension:** the number of vectors in any basis.

Method summary

Question	Do this
Is W a subspace?	check $\mathbf{0} \in W$, then closure
Are these independent?	columns into A , count pivots
Basis for a span?	pivot columns of A (original ones)
Dimension?	number of pivots
n vectors in $\mathbb{R}^n$?	one determinant settles everything

Pitfalls

- Any set defined by a *non-homogeneous* equation misses $\mathbf{0}$ and is not a subspace.
- A union of subspaces is generally not a subspace.
- Using RREF columns instead of original columns as a basis for $\mathbf{C}(A)$.
- Confusing “independent” with “orthogonal”.

Self-check

1. Is $\{(x, y, z) : x = y\}$ a subspace of $\mathbb{R}^3$? What is its dimension?
2. Is the set of 2×2 matrices with $\text{tr } A = 0$ a subspace? Give a basis.
3. Can 4 vectors be independent in $\mathbb{R}^3$? Can 2 vectors span $\mathbb{R}^3$?
4. Show that $\{1 + t, 1 - t, t^2\}$ is a basis of $\mathbb{P}_2$.

Answers

1. Yes — it is $\{(t, t, s)\}$, defined by the homogeneous equation $x - y = 0$. Basis $\{(1, 1, 0), (0, 0, 1)\}$, so $\dim = 2$: a plane.
2. Yes: $\text{tr}(A + B) = \text{tr } A + \text{tr } B$ and $\text{tr}(kA) = k \text{tr } A$, and $\text{tr } O = 0$. Writing $A = \begin{bmatrix} a & b \\ c & -a \end{bmatrix}$, a basis is $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$; so $\dim = 3$.
3. No and no. In $\mathbb{R}^3$ any independent set has ≤ 3 vectors and any spanning set has ≥ 3 .
4. $\dim \mathbb{P}_2 = 3$ and we have 3 polynomials, so by the counting theorem independence suffices. If $a(1 + t) + b(1 - t) + ct^2 = 0$ then comparing coefficients: t^2 gives $c = 0$; t gives $a - b = 0$; constant gives $a + b = 0$. Hence $a = b = c = 0$. Independent, therefore a basis.

Column Space, Null Space and the Four Subspaces

Every matrix A carries four subspaces with it, and together they answer every solvability question you can ask about $Ax = b$. The column space says which b are reachable; the null space says how much freedom there is in x ; the row space and left null space are their mirror images obtained from A^T . Strang calls the relationships between them the *Fundamental Theorem of Linear Algebra*, and it is the conceptual summit of the first half of the course.

Strang §3.1–§3.5, especially §3.5 (“Dimensions of the Four Subspaces”). Lecture 7; Assignment 01 Q1(c)–(e) and Q3 are pure column-space questions. STRANG | 6th ed.

9.1 The column space

Definition 9.1 Column space

The **column space** of $A \in \mathbb{R}^{m \times n}$ is the span of its columns:

$$\mathbf{C}(A) = \{Ax : x \in \mathbb{R}^n\} = \text{span}\{a_1, \dots, a_n\} \subseteq \mathbb{R}^m.$$

Lecture 7's phrasing: *the set of all possible vectors that can be created by combining the columns of the matrix.*

Theorem 9.1 The solvability criterion

$$Ax = b \text{ has a solution} \iff b \in \mathbf{C}(A).$$

Proof. Ax is the combination $x_1 a_1 + \dots + x_n a_n$. A solution x exists precisely when b can be written as such a combination, which is the definition of $b \in \mathbf{C}(A)$. ■

Intuition

$\mathbf{C}(A)$ is the “reachable set” or “output space” of A . Think of A as a machine: you feed in any $x \in \mathbb{R}^n$, and the outputs it can produce form exactly $\mathbf{C}(A)$. If $\mathbf{C}(A)$ is a plane inside $\mathbb{R}^3$, then two-thirds of the target vectors are simply unreachable, no matter what you feed in.

Worked Example 9.1 • Assignment 01, Q3(a)–(b) — a rank-one column space

For $A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix}$, describe $\mathbf{C}(A)$ geometrically.

Every column is a multiple of $\mathbf{a}_1 = (2, 1, -1)$: indeed $\mathbf{a}_2 = 2\mathbf{a}_1$ and $\mathbf{a}_3 = 3\mathbf{a}_1$. So

$$\mathbf{C}(A) = \{t(2, 1, -1) : t \in \mathbb{R}\},$$

a **line** through the origin in $\mathbb{R}^3$ with direction $(2, 1, -1)$. Its dimension is 1, so $\text{rank } A = 1$. Only $\frac{1}{\infty}$ of $\mathbb{R}^3$ is reachable — concretely, $\mathbf{b} = (12, 6, -6)$ lies on this line ($t = 6$) and is reachable, whereas $\mathbf{b} = (1, 0, 0)$ does not and is not.

9.2 The null space

Definition 9.2 Null space

The **null space** (or **kernel**) of $A \in \mathbb{R}^{m \times n}$ is

$$\mathbf{N}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{R}^n.$$

Lecture 7: the set of all vectors that become the zero vector when multiplied by the matrix. Its dimension is called the **nullity**.

Proposition 9.1

$\mathbf{N}(A)$ is a subspace of $\mathbb{R}^n$.

Proof. $A\mathbf{0} = \mathbf{0}$, so $\mathbf{0} \in \mathbf{N}(A)$. If $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$ then $A(c\mathbf{u} + d\mathbf{v}) = cA\mathbf{u} + dA\mathbf{v} = \mathbf{0}$. ■

Common Pitfall • the two spaces live in different worlds

For $A \in \mathbb{R}^{m \times n}$:

$$\mathbf{C}(A) \subseteq \mathbb{R}^m \quad (\text{as many entries as rows}), \quad \mathbf{N}(A) \subseteq \mathbb{R}^n \quad (\text{as many entries as columns}).$$

For a 3×5 matrix, $\mathbf{C}(A)$ sits in $\mathbb{R}^3$ and $\mathbf{N}(A)$ in $\mathbb{R}^5$. They cannot be compared or intersected. Getting this backwards is the most common error in this chapter.

9.2.1 Computing the null space

Method Special solutions

1. Row reduce A to RREF R (no need to augment — the right side stays $\mathbf{0}$).
2. Identify the **pivot columns** and the **free columns**. Number of free columns = $n - r$ where $r = \text{rank } A$.
3. For each free variable in turn: set it to 1, all other free variables to 0, and solve for the

pivot variables. This gives a **special solution**.

4. The $n - r$ special solutions form a **basis** for $\mathbf{N}(A)$, and

$$\mathbf{N}(A) = \text{span}\{s_1, \dots, s_{n-r}\}, \quad \dim \mathbf{N}(A) = n - r.$$

Worked Example 9.2 • Null space of a rank-2 matrix

Find a basis for $\mathbf{N}(A)$ where $A = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix}$.

Reduce.

$$\begin{aligned} A &\xrightarrow{R_2 - 2R_1, R_3 - 3R_1} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{bmatrix} \xrightarrow{R_3 - R_2} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ &\xrightarrow{R_2/2, R_1 - 2R_2} R = \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Pivot columns: 1 and 3. Free columns: 2 and 4. So $r = 2$, $n = 4$, and $\dim \mathbf{N}(A) = 4 - 2 = 2$.

Special solution 1 ($x_2 = 1$, $x_4 = 0$). From R : row 2 gives $x_3 + 2(0) = 0$ so $x_3 = 0$; row 1 gives $x_1 + 2(1) + 0 - 2(0) = 0$ so $x_1 = -2$.

$$s_1 = (-2, 1, 0, 0).$$

Special solution 2 ($x_2 = 0$, $x_4 = 1$). Row 2: $x_3 + 2(1) = 0$ so $x_3 = -2$; row 1: $x_1 + 0 + 0 - 2(1) = 0$ so $x_1 = 2$.

$$s_2 = (2, 0, -2, 1).$$

Check As_2 : first component $1(2) + 2(0) + 2(-2) + 2(1) = 2 - 4 + 2 = 0$ ✓; second $2(2) + 4(0) + 6(-2) + 8(1) = 4 - 12 + 8 = 0$ ✓; third $3(2) + 6(0) + 8(-2) + 10(1) = 6 - 16 + 10 = 0$ ✓.

$$\mathbf{N}(A) = \text{span}\{(-2, 1, 0, 0), (2, 0, -2, 1)\} \subseteq \mathbb{R}^4.$$

9.2.2 The complete solution of $Ax = b$

Theorem 9.2 Particular plus null

If x_p is any one solution of $Ax = b$, then the *complete* solution set is

$$x = x_p + x_n, \quad x_n \in \mathbf{N}(A).$$

Geometrically, the solution set is the null space *translated* by x_p .

Proof. If $Ax = b$ and $Ax_p = b$, subtracting gives $A(x - x_p) = 0$, so $x - x_p \in \mathbf{N}(A)$. Conversely $A(x_p + x_n) = b + 0 = b$. ■

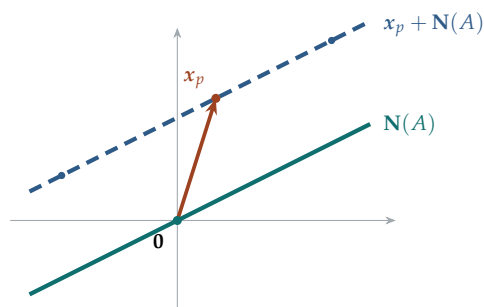


Figure 9.1: The solution set of $Ax = b$ is a translate of $N(A)$. It is a subspace only in the homogeneous case $b = 0$, when $x_p = 0$.

Worked Example 9.3 • Assignment 01, Q3(c) — complete solution

Solve $Ax = (12, 6, -6)$ for $A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix}$.

Reduce the augmented matrix.

$$\left[\begin{array}{ccc|c} 2 & 4 & 6 & 12 \\ 1 & 2 & 3 & 6 \\ -1 & -2 & -3 & -6 \end{array} \right] \xrightarrow{R_1/2} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 1 & 2 & 3 & 6 \\ -1 & -2 & -3 & -6 \end{array} \right] \xrightarrow{R_2 - R_1, R_3 + R_1} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

Consistent (no bad row), $r = 1$, $n = 3$, so $3 - 1 = 2$ free variables (x_2, x_3).

Particular solution: set $x_2 = x_3 = 0$, giving $x_1 = 6$: $x_p = (6, 0, 0)$.

Special solutions: $x_2 = 1, x_3 = 0 \Rightarrow x_1 = -2$: $s_1 = (-2, 1, 0)$. $x_2 = 0, x_3 = 1 \Rightarrow x_1 = -3$: $s_2 = (-3, 0, 1)$.

Complete solution:

$$x = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}.$$

This is a *plane* of solutions in $\mathbb{R}^3$, parallel to $N(A)$ but passing through $(6, 0, 0)$ rather than the origin.

9.3 The rank

Definition 9.3 Rank

The **rank** of A is the number of pivots in any echelon form of A . Equivalently,

$$r = \text{rank } A = \dim C(A) = \dim C(A^T),$$

and this same number is simultaneously

$$(\text{number of independent columns}) = (\text{number of independent rows}).$$

Theorem 9.3 Row rank = column rank

For every matrix, the number of independent rows equals the number of independent columns.

Why. Row reduction changes neither the row space nor the dependencies among columns. In the RREF R , the nonzero rows are visibly independent (each has a leading 1 where the others have 0) and there are r of them; the pivot columns are visibly independent and there are also r of them. Hence both counts equal r . ■

Remark • a genuinely surprising theorem

There is no obvious reason why counting independent rows should give the same answer as counting independent columns — they are computations in different spaces ($\mathbb{R}^n$ and $\mathbb{R}^m$). Yet they always agree. This is the first fact of the Fundamental Theorem.

Theorem 9.4 Rank–Nullity Theorem

For any $A \in \mathbb{R}^{m \times n}$,

$$\boxed{\text{rank } A + \dim \mathbf{N}(A) = n} \quad \text{i.e.} \quad r + (n - r) = n.$$

In words: (*pivot columns*) + (*free columns*) = (*all columns*).

Intuition • conservation of information

A maps $\mathbb{R}^n$ into $\mathbb{R}^m$. Some directions survive (rank r of them), the rest are crushed to zero ($n - r$ of them). Nothing is lost or created: the input dimension n splits exactly into “survives” plus “collapses”.

Worked Example 9.4 • Tutorial 04, Q6 — rank with a parameter

Find the rank of $A = \begin{bmatrix} a & 1 & 2 \\ 1 & 1 & 1 \\ -1 & 1 & 1-a \end{bmatrix}$ for each real a .

Put the simple row on top: $R_1 \leftrightarrow R_2$:

$$\begin{bmatrix} 1 & 1 & 1 \\ a & 1 & 2 \\ -1 & 1 & 1-a \end{bmatrix} \xrightarrow{R_2 - aR_1, R_3 + R_1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1-a & 2-a \\ 0 & 2 & 2-a \end{bmatrix}.$$

Now eliminate in column 2. To avoid dividing by $1-a$, swap rows 2 and 3 first:

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2-a \\ 0 & 1-a & 2-a \end{bmatrix} \xrightarrow{R_3 - \frac{1-a}{2}R_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2-a \\ 0 & 0 & (2-a) - \frac{(1-a)(2-a)}{2} \end{bmatrix}.$$

The (3,3) entry is

$$(2-a) \left[1 - \frac{1-a}{2} \right] = (2-a) \cdot \frac{2-1+a}{2} = \frac{(2-a)(1+a)}{2}.$$

Therefore:

Case	(3,3) entry	Rank
$a \neq 2$ and $a \neq -1$	nonzero	$\text{rank } A = 3$
$a = 2$	0	$\text{rank } A = 2$
$a = -1$	0	$\text{rank } A = 2$

Check via the determinant. Expanding, $\det A = -(a-2)(a+1)$, which vanishes exactly at $a = 2$ and $a = -1$ — matching the pivot analysis. At $a = 2$ the matrix is $\begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix}$, whose first and third columns are equal: rank 2 ✓.

9.4 The four fundamental subspaces

Definition 9.4 The four subspaces

For $A \in \mathbb{R}^{m \times n}$ of rank r :

Subspace	Lives in	Dimension	Basis from elimination
Column space $\mathbf{C}(A)$	$\mathbb{R}^m$	r	pivot columns of A
Null space $\mathbf{N}(A)$	$\mathbb{R}^n$	$n - r$	special solutions
Row space $\mathbf{C}(A^T)$	$\mathbb{R}^n$	r	nonzero rows of R
Left null space $\mathbf{N}(A^T)$	$\mathbb{R}^m$	$m - r$	from $EA = R$, last $m - r$ rows of E

The left null space is so named because $A^T \mathbf{y} = \mathbf{0}$ is equivalent to $\mathbf{y}^T A = \mathbf{0}^T$ — $\mathbf{y}$ multiplies A from the left.

Theorem 9.5 Fundamental Theorem of Linear Algebra

- (i) **Dimensions.** $\dim \mathbf{C}(A) = \dim \mathbf{C}(A^T) = r$, $\dim \mathbf{N}(A) = n - r$, $\dim \mathbf{N}(A^T) = m - r$.
- (ii) **Orthogonality.** In $\mathbb{R}^n$: $\mathbf{N}(A) \perp \mathbf{C}(A^T)$ and together they fill $\mathbb{R}^n$. In $\mathbb{R}^m$: $\mathbf{N}(A^T) \perp \mathbf{C}(A)$ and together they fill $\mathbb{R}^m$. Formally,

$$\mathbf{N}(A) = \mathbf{C}(A^T)^\perp, \quad \mathbf{N}(A^T) = \mathbf{C}(A)^\perp.$$

Proof of the orthogonality. $\mathbf{x} \in \mathbf{N}(A)$ means $A\mathbf{x} = \mathbf{0}$, i.e. every row of A dotted with $\mathbf{x}$ gives 0. So $\mathbf{x}$ is orthogonal to every row, hence to every combination of rows — that is, to all of $\mathbf{C}(A^T)$. The dimension count $(n - r) + r = n$ then shows they fill $\mathbb{R}^n$ between them. Applying the same argument to A^T gives the second pair. ■

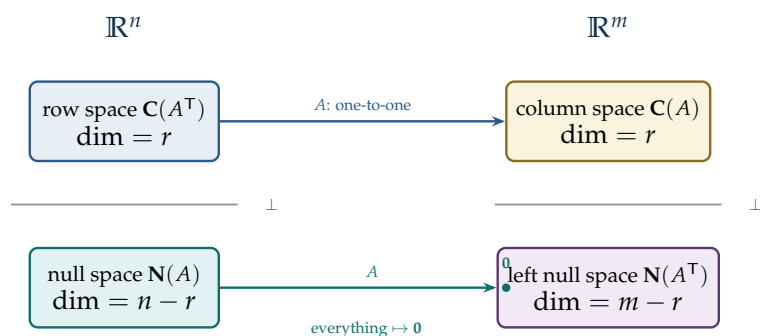


Figure 9.2: The Big Picture. A maps the row space one-to-one onto the column space, and crushes the null space to $\mathbf{0}$. Within each of $\mathbb{R}^n$ and $\mathbb{R}^m$, the two subspaces are orthogonal complements.

Worked Example 9.5 • All four subspaces of one matrix

Find bases for the four subspaces of $A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{bmatrix}$ ($m = 3, n = 4$).

Reduce.

$$A \xrightarrow{R_2 - R_1, R_3 - R_1} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_2 \rightarrow -R_2, R_1 - 2R_2} R = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Pivots in columns 1, 2: so $r = 2$.

1. Column space $\subseteq \mathbb{R}^3$, $\dim = 2$. Basis = pivot columns of the *original* A :

$$\mathbf{C}(A) = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \right\}.$$

2. Row space $\subseteq \mathbb{R}^4$, $\dim = 2$. Basis = nonzero rows of R :

$$\mathbf{C}(A^T) = \text{span}\{(1, 0, 1, 1), (0, 1, 1, 0)\}.$$

3. Null space $\subseteq \mathbb{R}^4$, $\dim = 4 - 2 = 2$. Free variables x_3, x_4 .

• $x_3 = 1, x_4 = 0$: row 2 gives $x_2 = -1$; row 1 gives $x_1 = -1$. $\mathbf{s}_1 = (-1, -1, 1, 0)$.

• $x_3 = 0, x_4 = 1$: row 2 gives $x_2 = 0$; row 1 gives $x_1 = -1$. $\mathbf{s}_2 = (-1, 0, 0, 1)$.

Check $A\mathbf{s}_1$: $(1 - 2 + 3, 1 - 1 + 2 \cdot 1 \dots)$ — first component $1(-1) + 2(-1) + 3(1) + 1(0) = -1 - 2 + 3 = 0$ ✓.

4. Left null space $\subseteq \mathbb{R}^3$, $\dim = 3 - 2 = 1$. Solve $\mathbf{y}^T A = \mathbf{0}$, i.e. find combinations of the *rows* of A giving zero. Rows 1 and 3 of A are identical, so $\text{row}_1 - \text{row}_3 = \mathbf{0}$:

$$\mathbf{N}(A^T) = \text{span}\{(1, 0, -1)\}.$$

Check: $1 \cdot (1, 2, 3, 1) + 0 \cdot (1, 1, 2, 1) - 1 \cdot (1, 2, 3, 1) = (0, 0, 0, 0)$. ✓

Dimension audit. $\mathbf{C}(A)$ and $\mathbf{N}(A^T)$: $2 + 1 = 3 = m$ ✓. $\mathbf{C}(A^T)$ and $\mathbf{N}(A)$: $2 + 2 = 4 = n$ ✓.

9.5 Rank, solvability and the shape of the answer

Theorem 9.6 Complete classification by rank

For $A \in \mathbb{R}^{m \times n}$ with rank r :

Case	Meaning	$Ax = b$
$r = m = n$	square, invertible	exactly one solution, for every b
$r = n < m$	full column rank, tall	0 or 1 solution; $\mathbf{N}(A) = \{\mathbf{0}\}$
$r = m < n$	full row rank, wide	infinitely many solutions, for every b
$r < m, r < n$	rank deficient	0 or infinitely many solutions

Worked Example 9.6 • Assignment 01, Q1(e) — geometry of the column space

For $A(t) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & t & 2 \\ 1 & 2 & t \end{bmatrix}$ we found $\det A(t) = t(t-2)$. Describe $\mathbf{C}(A)$ when $t = 0$ and when $t = 2$.

Case $t = 2$. $A(2) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{bmatrix}$. Rows 2 and 3 coincide, and columns 2,3 coincide.

Reducing:

$$\xrightarrow{R_2 - R_1, R_3 - R_1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{R_3 - R_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Rank 2, so $\mathbf{C}(A)$ is a **plane** through the origin in $\mathbb{R}^3$, spanned by $(1,1,1)$ and $(1,2,2)$. A normal vector is found from $\mathbf{n} \cdot (1,1,1) = 0$ and $\mathbf{n} \cdot (1,2,2) = 0$; subtracting gives $n_2 + n_3 = 0$, and then $n_1 = 0$ is forced by neither — solving properly: $n_1 + n_2 + n_3 = 0$ and $n_1 + 2n_2 + 2n_3 = 0$ give $n_2 + n_3 = 0$ and $n_1 = 0$, so $\mathbf{n} = (0,1,-1)$. The plane is $b_2 = b_3$.

Case $t = 0$. $A(0) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$. Reducing:

$$\xrightarrow{R_2 - R_1, R_3 - R_1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow{R_3 + R_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Rank 2 again — also a **plane**. Here $\text{row}_3 = 2\text{row}_1 - \text{row}_2$ in the original, so the consistency condition is $b_3 = 2b_1 - b_2$.

Summary. For $t \notin \{0,2\}$, $\mathbf{C}(A) = \mathbb{R}^3$ and every b works. For $t = 0$ or $t = 2$, $\mathbf{C}(A)$ collapses to a plane through the origin and only b in that plane are reachable. It is never a line or the zero vector, since the first two columns stay independent for all t .

9.6 Exam preparation

Exam Preparation • Chapter 9

The master identities

$$\text{rank } A + \dim \mathbf{N}(A) = n \quad (\text{rank-nullity})$$

$$\dim \mathbf{C}(A) = \dim \mathbf{C}(A^T) = r, \quad \dim \mathbf{N}(A) = n - r, \quad \dim \mathbf{N}(A^T) = m - r$$

$$\mathbf{N}(A) = \mathbf{C}(A^T)^\perp \subseteq \mathbb{R}^n, \quad \mathbf{N}(A^T) = \mathbf{C}(A)^\perp \subseteq \mathbb{R}^m$$

Solvability

$$Ax = \mathbf{b} \text{ solvable} \iff \mathbf{b} \in \mathbf{C}(A) \iff \text{rank}[A \mid \mathbf{b}] = \text{rank } A.$$

$$\text{complete solution} = \mathbf{x}_p + \mathbf{N}(A).$$

Where each basis comes from

$\mathbf{C}(A)$	pivot columns of the <i>original</i> A
$\mathbf{C}(A^T)$	nonzero rows of the RREF R
$\mathbf{N}(A)$	special solutions (one per free variable)
$\mathbf{N}(A^T)$	row combinations of A that give $\mathbf{0}$

Pitfalls

- $\mathbf{C}(A) \subseteq \mathbb{R}^m$ but $\mathbf{N}(A) \subseteq \mathbb{R}^n$. Different spaces.
- Using RREF columns as a basis for $\mathbf{C}(A)$. Row operations *change* the column space (but not the row space or null space).
- The solution set of $Ax = \mathbf{b}$ is a subspace only when $\mathbf{b} = \mathbf{0}$.
- Forgetting that “number of free variables” means *free columns*, i.e. $n - r$, not $m - r$.

Self-check

1. A is 5×3 with $\text{rank } A = 3$. Give the dimensions of all four subspaces. Is $Ax = \mathbf{b}$ always solvable?
2. A is 3×5 with $\text{rank } A = 3$. Same questions.
3. If $\mathbf{N}(A) = \{\mathbf{0}\}$ for a 4×4 matrix, what is $\mathbf{C}(A)$?
4. Show that $\mathbf{N}(A) \subseteq \mathbf{N}(A^T A)$, and in fact they are equal.

Answers

1. $m = 5, n = 3, r = 3$: $\dim \mathbf{C}(A) = 3$, $\dim \mathbf{N}(A) = 0$, $\dim \mathbf{C}(A^T) = 3$, $\dim \mathbf{N}(A^T) = 2$. Not always solvable — $\mathbf{C}(A)$ is only a 3-dimensional subspace of $\mathbb{R}^5$. But when solvable, the solution is unique.
2. $m = 3, n = 5, r = 3$: $\dim \mathbf{C}(A) = 3 = \mathbb{R}^3$, $\dim \mathbf{N}(A) = 2$, $\dim \mathbf{C}(A^T) = 3$, $\dim \mathbf{N}(A^T) = 0$. Always solvable, never uniquely — ∞ many solutions for every $\mathbf{b}$.
3. $\dim \mathbf{N}(A) = 0$, so $r = 4 - 0 = 4$ and $\mathbf{C}(A) = \mathbb{R}^4$. (A is invertible.)
4. If $Ax = \mathbf{0}$ then $A^T Ax = A^T \mathbf{0} = \mathbf{0}$, giving $\subseteq$. Conversely if $A^T Ax = \mathbf{0}$ then $x^T A^T Ax = 0$, i.e. $\|Ax\|^2 = 0$, so $Ax = \mathbf{0}$. Hence equality — and by rank-nullity, $\text{rank}(A^T A) = \text{rank } A$, a fact used throughout Chapters 10 and 13.

PART IV

Orthogonality

Orthogonality, Projections and Least Squares

Orthogonality is what turns linear algebra from bookkeeping into geometry. Perpendicular vectors are automatically independent, orthonormal bases make coordinates trivial to compute, and orthogonal matrices preserve every length and angle they touch. This chapter also answers the question elimination cannot: what do you do when $Ax = b$ has *no* solution? The answer — project b onto the column space — is least squares, the single most-used idea in applied linear algebra.

Strang §4.1 (“Orthogonality of the Four Subspaces”), §4.2 (“Projections”), §4.3 (“Least Squares”), §4.4 (“Orthonormal Bases and Gram–Schmidt”). Lecture 7 introduces orthogonality and orthonormal matrices; the SVD notes revisit them. STRANG | 6th ed.

10.1 Orthogonal vectors

Definition 10.1 Orthogonal

Vectors $x, y \in \mathbb{R}^n$ are **orthogonal** (perpendicular), written $x \perp y$, if

$$x \cdot y = x^T y = 0.$$

Lecture 7’s summary: *orthogonal = perpendicular; two vectors are orthogonal if the angle between them is 90° , equivalently if their dot product is zero.*

Remark • the zero vector is orthogonal to everything

$0 \cdot y = 0$ for every y , so 0 is perpendicular to all vectors — including itself. This is a convention that keeps the theorems clean; it is the reason “orthogonal” does not by itself imply “independent” unless the vectors are also nonzero.

Theorem 10.1 Pythagoras

For $x, y \in \mathbb{R}^n$,

$$x \perp y \iff \|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Proof. Expand: $\|x + y\|^2 = (x + y)^T(x + y) = \|x\|^2 + 2x^T y + \|y\|^2$. The cross term vanishes exactly when $x^T y = 0$. ■

Theorem 10.2 Orthogonal $\Rightarrow$ independent

If $v_1, \dots, v_k$ are *nonzero* and mutually orthogonal, they are linearly independent.

Proof. Suppose $c_1 v_1 + \dots + c_k v_k = \mathbf{0}$. Take the dot product with v_j . Every term $c_i v_i \cdot v_j$ with $i \neq j$ vanishes by orthogonality, leaving

$$c_j v_j \cdot v_j = c_j \|v_j\|^2 = 0.$$

Since $v_j \neq \mathbf{0}$ we have $\|v_j\|^2 > 0$, so $c_j = 0$. This holds for every j . ■

Intuition

This is why orthogonal bases are so convenient: independence comes free, and — as we see below — so do the coordinates. With a general basis, finding coordinates means solving a linear system; with an orthogonal basis it means computing dot products.

10.2 Orthogonal complements

Definition 10.2 Orthogonal complement

For a subspace $V \subseteq \mathbb{R}^n$, its **orthogonal complement** is

$$V^\perp = \{x \in \mathbb{R}^n : x \cdot v = 0 \text{ for every } v \in V\}.$$

$V^\perp$ is itself a subspace, $\dim V + \dim V^\perp = n$, and $(V^\perp)^\perp = V$.

Remark • the four subspaces again

Chapter 9 already proved the two key instances:

$$N(A) = C(A^T)^\perp \text{ in } \mathbb{R}^n, \quad N(A^T) = C(A)^\perp \text{ in } \mathbb{R}^m.$$

Every vector in $\mathbb{R}^n$ therefore splits uniquely into a row-space part and a null-space part. That splitting is exactly what projection computes.

Worked Example 10.1 • Assignment 01, Q2(a)(e) — a plane as a null space

Let $n = (2, -1, 2)$ and $P = \{x \in \mathbb{R}^3 : n \cdot x = 0\}$. Show $a = (1, 0, -1)$ and $b = (0, 2, 1)$ lie in P , and explain geometrically why P is a plane through the origin.

Membership.

$$n \cdot a = 2(1) + (-1)(0) + 2(-1) = 2 - 2 = 0 \checkmark$$

$$n \cdot b = 2(0) + (-1)(2) + 2(1) = -2 + 2 = 0 \checkmark$$

Why a plane. P is the null space of the 1×3 matrix $n^T = [2 \ -1 \ 2]$, which has rank 1. By rank-nullity, $\dim P = 3 - 1 = 2$ — a two-dimensional subspace of $\mathbb{R}^3$, i.e. a plane through the origin.

Geometrically, $n \cdot x = 0$ says x makes a right angle with the fixed direction n . The set of all directions perpendicular to a single direction in $\mathbb{R}^3$ is precisely a plane, and it must contain

$\mathbf{0}$ because $\mathbf{n} \cdot \mathbf{0} = 0$. In the language above, $P = \text{span}\{\mathbf{n}\}^\perp$, and $\dim \text{span}\{\mathbf{n}\} + \dim P = 1 + 2 = 3$.

Since $\mathbf{a}, \mathbf{b} \in P$ are independent (Chapter 8) and $\dim P = 2$, they form a **basis** of P .

Worked Example 10.2 • Assignment 01, Q2(d) — expressing a vector in the plane

Express $\mathbf{r} = (3, 2, -2)$ as a combination of $\mathbf{a} = (1, 0, -1)$ and $\mathbf{b} = (0, 2, 1)$, and hence explain why $\mathbf{r} \in P$.

Solve $c_1\mathbf{a} + c_2\mathbf{b} = \mathbf{r}$ componentwise:

$$c_1 = 3, \quad 2c_2 = 2, \quad -c_1 + c_2 = -2.$$

The first gives $c_1 = 3$, the second $c_2 = 1$, and the third checks out: $-3 + 1 = -2 \checkmark$. So

$$\mathbf{r} = 3\mathbf{a} + \mathbf{b}.$$

Since P is a subspace containing $\mathbf{a}$ and $\mathbf{b}$, it is closed under linear combinations, so $\mathbf{r} \in P$.

Direct verification: $\mathbf{n} \cdot \mathbf{r} = 2(3) + (-1)(2) + 2(-2) = 6 - 2 - 4 = 0 \checkmark$.

10.3 Orthonormal sets and orthogonal matrices

Definition 10.3 Orthonormal

A set $\{\mathbf{q}_1, \dots, \mathbf{q}_k\}$ is **orthonormal** if

$$\mathbf{q}_i \cdot \mathbf{q}_j = \begin{cases} 1 & i = j \quad (\text{each is a unit vector}) \\ 0 & i \neq j \quad (\text{mutually perpendicular}) \end{cases}$$

Lecture 7 unpacks the word: *ortho* = orthogonal, *normal* = normalized to length 1.

Definition 10.4 Orthogonal matrix

A matrix Q with orthonormal *columns* satisfies

$$Q^T Q = I.$$

If Q is additionally **square**, then $Q^T Q = Q Q^T = I$, so

$$Q^{-1} = Q^T$$

and Q is called an **orthogonal matrix**. (Lecture 7 calls such a matrix *orthonormal*; both names are in use, and Strang prefers “orthogonal matrix” for the square case.)

Common Pitfall • square matters

A tall matrix with orthonormal columns has $Q^T Q = I_n$ but $Q Q^T \neq I_m$ — the latter is a projection matrix, not the identity. Only for square Q do both products give I .

Theorem 10.3 Orthogonal matrices preserve geometry

If $Q^T Q = I$ then for all x, y :

$$(Qx) \cdot (Qy) = x \cdot y, \quad \|Qx\| = \|x\|, \quad \angle(Qx, Qy) = \angle(x, y).$$

Moreover $\det Q = \pm 1$, and every eigenvalue of Q has $|\lambda| = 1$.

Proof. $(Qx)^T(Qy) = x^T Q^T Q y = x^T I y = x^T y$. Setting $y = x$ gives $\|Qx\|^2 = \|x\|^2$; angles follow from the cosine formula since both dot products and lengths are preserved. For the determinant, $\det(Q^T Q) = \det(Q)^2 = \det I = 1$. ■

Intuition • what orthogonal matrices are

$\det Q = +1$: a rotation. $\det Q = -1$: a reflection (or a rotation composed with one). Nothing else. Orthogonal matrices are exactly the rigid motions of $\mathbb{R}^n$ that fix the origin — see Chapter 15, where rotation and reflection matrices appear explicitly, and note that each one satisfies $Q^T Q = I$.

Worked Example 10.3 • Lecture 7, Example 1 — checking orthonormality

$$\text{Is } Q = \frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix} \text{ orthogonal?}$$

Write the columns as $q_1 = \frac{1}{3}(2, 1, 2)$, $q_2 = \frac{1}{3}(-2, 2, 1)$, $q_3 = \frac{1}{3}(1, 2, -2)$.

Unit length.

$$\|q_1\|^2 = \frac{1}{9}(4 + 1 + 4) = \frac{9}{9} = 1, \quad \|q_2\|^2 = \frac{1}{9}(4 + 4 + 1) = 1, \quad \|q_3\|^2 = \frac{1}{9}(1 + 4 + 4) = 1. \checkmark$$

Mutual orthogonality.

$$q_1 \cdot q_2 = \frac{1}{9}(2(-2) + 1(2) + 2(1)) = \frac{1}{9}(-4 + 2 + 2) = 0 \checkmark$$

$$q_1 \cdot q_3 = \frac{1}{9}(2(1) + 1(2) + 2(-2)) = \frac{1}{9}(2 + 2 - 4) = 0 \checkmark$$

$$q_2 \cdot q_3 = \frac{1}{9}(-2(1) + 2(2) + 1(-2)) = \frac{1}{9}(-2 + 4 - 2) = 0 \checkmark$$

So $Q^T Q = I$ and, since Q is square, $Q^{-1} = Q^T$ — the inverse is had for free, with no elimination at all. Here $\det Q = +1$, so Q is a rotation of $\mathbb{R}^3$.

Theorem 10.4 Coordinates in an orthonormal basis

If $q_1, \dots, q_n$ is an orthonormal basis of $\mathbb{R}^n$, then every v expands as

$$v = (v \cdot q_1)q_1 + (v \cdot q_2)q_2 + \dots + (v \cdot q_n)q_n.$$

The coordinates are just dot products — no system to solve.

Proof. Write $v = \sum_i c_i q_i$ (possible since it is a basis) and dot with q_j . Orthonormality kills every term but one, leaving $v \cdot q_j = c_j$. ■

10.4 Projections

Definition 10.5 Projection onto a line

The **projection** of $\mathbf{b}$ onto the line through $\mathbf{a} \neq \mathbf{0}$ is the multiple $\mathbf{p} = \hat{x}\mathbf{a}$ that is closest to $\mathbf{b}$. Requiring the error $\mathbf{e} = \mathbf{b} - \mathbf{p}$ to be perpendicular to $\mathbf{a}$ gives $\mathbf{a}^\top(\mathbf{b} - \hat{x}\mathbf{a}) = 0$, hence

$$\hat{x} = \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}}, \quad \mathbf{p} = \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}} \mathbf{a}, \quad P = \frac{\mathbf{a} \mathbf{a}^\top}{\mathbf{a}^\top \mathbf{a}} \text{ with } \mathbf{p} = P\mathbf{b}.$$

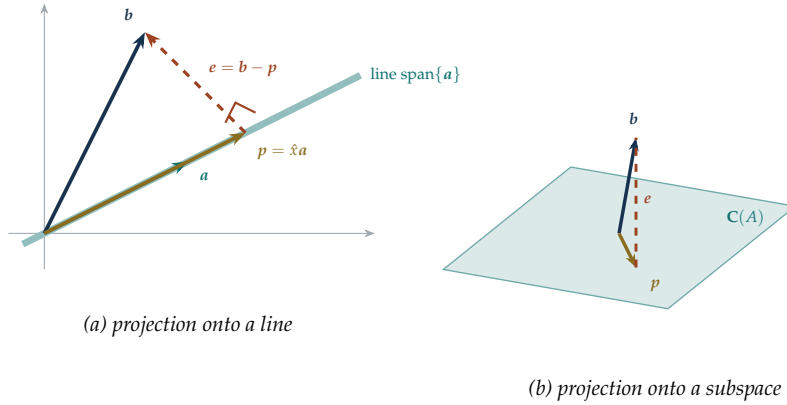


Figure 10.1: Projection drops $\mathbf{b}$ perpendicularly onto the subspace. The error $\mathbf{e} = \mathbf{b} - \mathbf{p}$ is orthogonal to everything in the subspace, and $\mathbf{p}$ is the closest point to $\mathbf{b}$ within it.

Theorem 10.5 Projection onto a subspace

Let $A \in \mathbb{R}^{m \times n}$ have independent columns, so $\mathbf{C}(A)$ is an n -dimensional subspace of $\mathbb{R}^m$. The projection of $\mathbf{b}$ onto $\mathbf{C}(A)$ is $\mathbf{p} = A\hat{x}$ where $\hat{x}$ solves the **normal equations**

$$A^\top A \hat{x} = A^\top \mathbf{b}$$

Explicitly,

$$\hat{x} = (A^\top A)^{-1} A^\top \mathbf{b}, \quad \mathbf{p} = A(A^\top A)^{-1} A^\top \mathbf{b} = P\mathbf{b}, \quad P = A(A^\top A)^{-1} A^\top.$$

Proof. The error $\mathbf{e} = \mathbf{b} - A\hat{x}$ must be orthogonal to every column of A , i.e. $A^\top(\mathbf{b} - A\hat{x}) = \mathbf{0}$, which rearranges to $A^\top A \hat{x} = A^\top \mathbf{b}$. Since A has independent columns, $\mathbf{N}(A) = \{\mathbf{0}\}$ and so (by the self-check at the end of Chapter 9) $\mathbf{N}(A^\top A) = \{\mathbf{0}\}$ too — hence $A^\top A$ is invertible and $\hat{x}$ is unique. ■

Proposition 10.1 Properties of a projection matrix

$P = A(A^\top A)^{-1} A^\top$ satisfies

$$P^\top = P \quad (\text{symmetric}), \quad P^2 = P \quad (\text{idempotent}).$$

Its eigenvalues are all 0 or 1; $\text{rank } P = \dim \mathbf{C}(A)$; and $I - P$ is the projection onto $\mathbf{C}(A)^\perp = \mathbf{N}(A^\top)$.

Proof. $P^\top = (A(A^\top A)^{-1} A^\top)^\top = A((A^\top A)^{-1})^\top A^\top = P$, using symmetry of $A^\top A$. And $P^2 =$

$A(A^T A)^{-1} \underbrace{A^T A(A^T A)^{-1} A^T}_{=I} = P$. Idempotence says what you expect: projecting an already-projected vector changes nothing. ■

Worked Example 10.4 • Projection onto a plane

Project $\mathbf{b} = (1, 1, 1)$ onto the plane spanned by $\mathbf{a}_1 = (1, 0, -1)$ and $\mathbf{a}_2 = (0, 2, 1)$.

With $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ -1 & 1 \end{bmatrix}$:

$$A^T A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix}, \quad A^T \mathbf{b} = \begin{bmatrix} 1+0-1 \\ 0+2+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}.$$

Solve $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$. Since $\det(A^T A) = 10 - 1 = 9$,

$$(A^T A)^{-1} = \frac{1}{9} \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}, \quad \hat{\mathbf{x}} = \frac{1}{9} \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Hence

$$\mathbf{p} = A\hat{\mathbf{x}} = \frac{1}{3} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \\ \frac{1}{3} \end{bmatrix}, \quad \mathbf{e} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Check the error is perpendicular to the plane: $\mathbf{e} \cdot \mathbf{a}_1 = \frac{2}{3} - 0 - \frac{2}{3} = 0$ ✓ and $\mathbf{e} \cdot \mathbf{a}_2 = 0 - \frac{2}{3} + \frac{2}{3} = 0$ ✓.

Note that $\mathbf{e}$ is a multiple of $(2, -1, 2)$ — the normal $\mathbf{n}$ from the Assignment 01 example, exactly as the theory predicts, since $\mathbf{N}(A^T) = \mathbf{C}(A)^\perp = \text{span}\{\mathbf{n}\}$.

10.5 Least squares

Intuition • the problem elimination cannot solve

When A is tall ($m > n$) — more equations than unknowns — the system $A\mathbf{x} = \mathbf{b}$ usually has *no* solution: $\mathbf{b}$ simply does not lie in the n -dimensional column space sitting inside $\mathbb{R}^m$. Rather than give up, we ask for the $\hat{\mathbf{x}}$ making $A\hat{\mathbf{x}}$ as close to $\mathbf{b}$ as possible. Closest means smallest error length, and the closest point in $\mathbf{C}(A)$ to $\mathbf{b}$ is the projection. So least squares *is* projection.

Theorem 10.6 Least squares

The vector $\hat{\mathbf{x}}$ minimising $\|A\mathbf{x} - \mathbf{b}\|^2$ is the solution of the normal equations

$$A^T A \hat{\mathbf{x}} = A^T \mathbf{b}.$$

If A has independent columns this solution is unique.

Method Fitting a straight line to data

To fit $y = c + dt$ through points $(t_1, y_1), \dots, (t_m, y_m)$:

$$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \quad x = \begin{bmatrix} c \\ d \end{bmatrix}, \quad b = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix},$$

then solve $A^T A \hat{x} = A^T b$, a 2×2 system.

Worked Example 10.5 • Least-squares line through four points

Fit the best line $y = c + dt$ to $(0, 1), (1, 3), (2, 4), (3, 4)$.

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}.$$

$$A^T A = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}, \quad A^T b = \begin{bmatrix} 1 + 3 + 4 + 4 \\ 0 + 3 + 8 + 12 \end{bmatrix} = \begin{bmatrix} 12 \\ 23 \end{bmatrix}.$$

The normal equations are

$$4c + 6d = 12, \quad 6c + 14d = 23.$$

From the first, $c = 3 - \frac{3}{2}d$. Substituting: $18 - 9d + 14d = 23$, so $5d = 5$, $d = 1$ and $c = \frac{3}{2}$.

Best line: $y = \frac{3}{2} + t$.

Errors. Predicted values at $t = 0, 1, 2, 3$ are $\frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \frac{9}{2}$, so

$$e = b - A\hat{x} = \left(1 - \frac{3}{2}, 3 - \frac{5}{2}, 4 - \frac{7}{2}, 4 - \frac{9}{2}\right) = \left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right).$$

Check orthogonality to the columns of A : $e \cdot (1, 1, 1, 1) = -\frac{1}{2} + \frac{1}{2} + \frac{1}{2} - \frac{1}{2} = 0 \checkmark$; $e \cdot (0, 1, 2, 3) = 0 + \frac{1}{2} + 1 - \frac{3}{2} = 0 \checkmark$. The total squared error is $\|e\|^2 = 4 \cdot \frac{1}{4} = 1$, and no other line does better.

10.6 Gram–Schmidt and QR

Method Gram–Schmidt orthogonalization

Given independent $a_1, \dots, a_n$, build an orthonormal $q_1, \dots, q_n$ spanning the same space:

$$\begin{aligned} w_1 &= a_1, & q_1 &= w_1 / \|w_1\| \\ w_2 &= a_2 - (a_2 \cdot q_1)q_1, & q_2 &= w_2 / \|w_2\| \\ w_3 &= a_3 - (a_3 \cdot q_1)q_1 - (a_3 \cdot q_2)q_2, & q_3 &= w_3 / \|w_3\| \\ &\vdots & & \end{aligned}$$

At each step: subtract off the projections onto everything already built, then normalise what

remains.

Theorem 10.7 QR factorization

Every $A \in \mathbb{R}^{m \times n}$ with independent columns factors as

$$A = QR$$

with $Q \in \mathbb{R}^{m \times n}$ having orthonormal columns ($Q^T Q = I$) and $R \in \mathbb{R}^{n \times n}$ upper triangular with positive diagonal, where $r_{ij} = q_i \cdot a_j$.

Remark • why QR beats the normal equations

Substituting $A = QR$ into $A^T A \hat{x} = A^T b$ gives $R^T Q^T Q R \hat{x} = R^T Q^T b$, i.e. simply

$$R \hat{x} = Q^T b,$$

a triangular system. This is what numerical libraries actually solve: forming $A^T A$ squares the condition number and can lose half the available precision, whereas QR does not.

Worked Example 10.6 • Gram–Schmidt on two vectors

Orthonormalise $a_1 = (1, 1, 0)$ and $a_2 = (1, 0, 1)$.

Step 1. $w_1 = (1, 1, 0)$, $\|w_1\| = \sqrt{2}$, so

$$q_1 = \frac{1}{\sqrt{2}}(1, 1, 0).$$

Step 2. $a_2 \cdot q_1 = \frac{1}{\sqrt{2}}(1 + 0 + 0) = \frac{1}{\sqrt{2}}$, so

$$w_2 = (1, 0, 1) - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}(1, 1, 0) = (1, 0, 1) - \frac{1}{2}(1, 1, 0) = \left(\frac{1}{2}, -\frac{1}{2}, 1\right).$$

$$\|w_2\| = \sqrt{\frac{1}{4} + \frac{1}{4} + 1} = \sqrt{\frac{3}{2}} = \frac{\sqrt{6}}{2}, \text{ so}$$

$$q_2 = \frac{2}{\sqrt{6}} \left(\frac{1}{2}, -\frac{1}{2}, 1\right) = \frac{1}{\sqrt{6}}(1, -1, 2).$$

Check: $q_1 \cdot q_2 = \frac{1}{\sqrt{12}}(1 - 1 + 0) = 0 \checkmark$, and $\|q_2\|^2 = \frac{1}{6}(1 + 1 + 4) = 1 \checkmark$.
The corresponding R is

$$R = \begin{bmatrix} q_1 \cdot a_1 & q_1 \cdot a_2 \\ 0 & q_2 \cdot a_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} \end{bmatrix}.$$

10.7 Exam preparation

Exam Preparation • Chapter 10

Core formulas

$$x \perp y \iff x^T y = 0, \quad Q^T Q = I \Rightarrow Q^{-1} = Q^T \text{ (square } Q)$$

$$p = \frac{a^T b}{a^T a} a \quad (\text{onto a line}), \quad A^T A \hat{x} = A^T b \quad (\text{normal equations})$$

$$P = A(A^T A)^{-1} A^T, \quad P^T = P, \quad P^2 = P$$

Facts

- Nonzero orthogonal vectors are automatically independent.
- Orthogonal matrices preserve lengths, angles and dot products; $\det Q = \pm 1$.
- In an orthonormal basis, coordinates are dot products.
- Least squares = projection onto $C(A)$; the error $e = b - p$ lies in $N(A^T)$.
- $A^T A$ is invertible $\iff A$ has independent columns.

Pitfalls

- Orthogonal $\neq$ orthonormal — orthonormal also demands unit length.
- For non-square Q with $Q^T Q = I$, $Q Q^T$ is a *projection*, not I .
- Forgetting to normalise at each Gram–Schmidt step (then the formulas need $w^T w$ divisors).
- Writing $\hat{x} = A^{-1} b$ for a rectangular A .

Self-check

1. Project $b = (1, 2, 3)$ onto the line through $a = (1, 1, 1)$.
2. Show that if P is a projection matrix then so is $I - P$, and identify what it projects onto.
3. Are $(1, 2, -1)$, $(2, -1, 0)$, $(1, 2, 5)$ mutually orthogonal? If so, normalise them.
4. Fit the best horizontal line $y = c$ to the data $y = 1, 3, 5$. What is c , and why is that unsurprising?

Answers

1. $a^T b = 6$, $a^T a = 3$, so $\hat{x} = 2$ and $p = 2(1, 1, 1) = (2, 2, 2)$. Error $(-1, 0, 1)$, which indeed satisfies $e \cdot a = 0$.
2. $(I - P)^T = I - P$ and $(I - P)^2 = I - 2P + P^2 = I - 2P + P = I - P$. It projects onto $C(A)^\perp = N(A^T)$, and $b = Pb + (I - P)b$ is the splitting into the two complementary pieces.
3. $(1, 2, -1) \cdot (2, -1, 0) = 2 - 2 + 0 = 0 \checkmark$; $(1, 2, -1) \cdot (1, 2, 5) = 1 + 4 - 5 = 0 \checkmark$; $(2, -1, 0) \cdot (1, 2, 5) = 2 - 2 + 0 = 0 \checkmark$. Yes. Normalising: $\frac{1}{\sqrt{6}}(1, 2, -1)$, $\frac{1}{\sqrt{5}}(2, -1, 0)$, $\frac{1}{\sqrt{30}}(1, 2, 5)$.
4. $A = (1, 1, 1)^T$, so $A^T A = 3$ and $A^T b = 9$, giving $c = 3$ — the *mean* of the data. Least squares with a single constant column always returns the average, which is exactly what “minimise the sum of squared deviations” means in elementary statistics.

PART V

**Eigenvalues, Eigenvectors and
Decompositions**

Eigenvalues and Eigenvectors

Most vectors change direction when multiplied by A . A few special ones do not — they only stretch or shrink. Those are the eigenvectors, and the stretching factors are the eigenvalues. Finding them amounts to finding the coordinate system in which A is at its simplest, and that single idea powers differential equations, Markov chains, vibration analysis, PageRank, PCA and face recognition. This chapter covers the definition, the characteristic polynomial, and the properties needed for Tutorial 06.

Strang §6.1 (“Introduction to Eigenvalues”), §6.2 (“Diagonalizing a Matrix”). Lecture 6; Tutorial 06
STRANG | 6th ed.
 Parts A and B are entirely on this chapter and the next.

11.1 The definition

Definition 11.1 Eigenvalue and eigenvector

Let $A \in \mathbb{R}^{n \times n}$. A scalar λ is an **eigenvalue** of A if there is a **nonzero** vector $x \in \mathbb{R}^n$ with

$$Ax = \lambda x$$

Such an x is an **eigenvector** of A corresponding to λ . The set

$$E_\lambda = \{x \in \mathbb{R}^n : Ax = \lambda x\} = \mathbf{N}(A - \lambda I)$$

is the **eigenspace** of λ ; it consists of all eigenvectors for λ together with 0 .

Common Pitfall • Tutorial 06, Q1(b) — why $x \neq 0$ is essential

The equation $A0 = \lambda 0$ holds for *every* scalar λ and every matrix A . If 0 were allowed as an eigenvector, every real number would be an eigenvalue of every matrix and the definition would carry no information whatsoever. Excluding 0 is what makes “ λ is an eigenvalue” a genuine constraint.

Note the asymmetry: the *eigenvector* must be nonzero, but the *eigenvalue* may perfectly well be 0 — and $\lambda = 0$ occurring means precisely that A is singular.

Intuition • what the equation says

Ax is normally a rotated, stretched, sheared version of x . For an eigenvector, all the rotating and shearing cancels and only pure scaling remains: x and Ax lie on the same line through the origin. The eigenvector picks out a direction that A leaves invariant; the eigenvalue says by how much that direction is stretched ($|\lambda| > 1$), shrunk ($|\lambda| < 1$), flipped ($\lambda < 0$) or

annihilated ($\lambda = 0$).

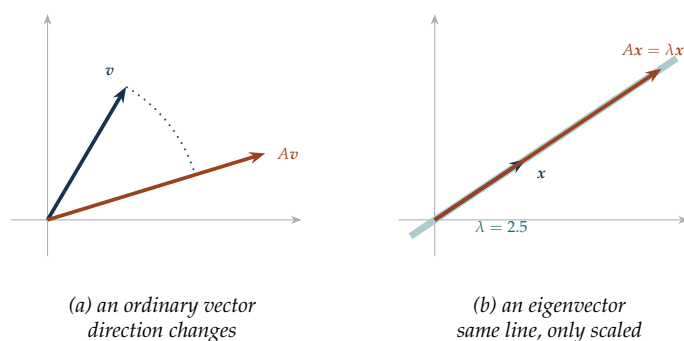


Figure 11.1: Eigenvectors are the directions A does not rotate.

11.2 Finding eigenvalues

Theorem 11.1 Tutorial 06, Q3(a) — the characteristic equation

λ is an eigenvalue of A **if and only if**

$$\det(A - \lambda I) = 0$$

Proof.

$$Ax = \lambda x \iff Ax - \lambda x = \mathbf{0} \iff (A - \lambda I)x = \mathbf{0}.$$

By definition λ is an eigenvalue iff this has a solution $x \neq \mathbf{0}$, i.e. iff $\mathbf{N}(A - \lambda I) \neq \{\mathbf{0}\}$, i.e. iff $A - \lambda I$ is singular — which by the Invertible Matrix Theorem is equivalent to $\det(A - \lambda I) = 0$. ■

Common Pitfall • write λI , not λ

$A - \lambda$ is meaningless: you cannot subtract a scalar from a matrix. The identity matrix must be there. In practice this means subtracting λ from every *diagonal* entry and leaving the rest alone.

Definition 11.2 Characteristic polynomial

$$p_A(\lambda) = \det(A - \lambda I)$$

is a polynomial of degree exactly n in λ . Its roots are the eigenvalues. (Some texts, including Tutorial 06 Q5, use $p_A(t) = \det(tI - A)$ instead, which differs by a factor $(-1)^n$ and is always monic — the roots are identical.)

Corollary 11.1 Tutorial 06, Q3(b)

0 is an eigenvalue of $A \iff \det(A) = 0 \iff A$ is singular.

Proof. Set $\lambda = 0$ in the characteristic equation: $\det(A - 0I) = \det A = 0$. ■

Method Computing eigenvalues and eigenvectors

1. Form $A - \lambda I$ (subtract λ down the diagonal).
2. Compute $\det(A - \lambda I)$ and set it to 0.
3. Solve for the roots $\lambda_1, \dots, \lambda_n$ (with multiplicity).
4. For each λ_i , solve $(A - \lambda_i I)x = \mathbf{0}$ by elimination. The special solutions form a basis of the eigenspace E_{λ_i} .
5. **Check:** $\sum \lambda_i = \text{tr } A$ and $\prod \lambda_i = \det A$.

Worked Example 11.1 • Lecture 6 — a 2×2 example in full

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$.

Characteristic equation.

$$A - \lambda I = \begin{bmatrix} 4 - \lambda & -2 \\ 1 & 1 - \lambda \end{bmatrix},$$

$$\det(A - \lambda I) = (4 - \lambda)(1 - \lambda) - (-2)(1) = 4 - 5\lambda + \lambda^2 + 2 = \lambda^2 - 5\lambda + 6.$$

Factoring, $(\lambda - 2)(\lambda - 3) = 0$, so

$$\lambda_1 = 2, \quad \lambda_2 = 3.$$

Check: $\lambda_1 + \lambda_2 = 5 = \text{tr } A = 4 + 1$ ✓ and $\lambda_1 \lambda_2 = 6 = \det A = 4 - (-2) = 6$ ✓.

Eigenvector for $\lambda_1 = 2$. Solve $(A - 2I)x = \mathbf{0}$:

$$\begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Both rows give the same equation $x_1 = x_2$. Taking $x_2 = 1$:

$$x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$\text{Check: } Ax_1 = \begin{bmatrix} 4 - 2 \\ 1 + 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2x_1 \text{ ✓.}$$

Eigenvector for $\lambda_2 = 3$. Solve $(A - 3I)x = \mathbf{0}$:

$$\begin{bmatrix} 1 & -2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies x_1 = 2x_2.$$

Taking $x_2 = 1$:

$$x_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

$$\text{Check: } Ax_2 = \begin{bmatrix} 8 - 2 \\ 2 + 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3x_2 \text{ ✓.}$$

Common Pitfall • eigenvectors are never unique

Every nonzero multiple of an eigenvector is again an eigenvector for the same λ (Tutorial 06 Q1(c)). So $(1, 1)$, $(2, 2)$ and $(-5, -5)$ are all valid answers above. Quote any convenient one — usually the one with smallest integer entries — and never write “the” eigenvector as though it were unique.

Worked Example 11.2 • A 3×3 example

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$.

Characteristic polynomial. Expand along the first row, which has two zeros:

$$\det(A - \lambda I) = \begin{vmatrix} 2 - \lambda & 0 & 0 \\ 1 & 3 - \lambda & 1 \\ 1 & 1 & 3 - \lambda \end{vmatrix} = (2 - \lambda)[(3 - \lambda)^2 - 1].$$

Now $(3 - \lambda)^2 - 1 = (3 - \lambda - 1)(3 - \lambda + 1) = (2 - \lambda)(4 - \lambda)$, so

$$p_A(\lambda) = (2 - \lambda)^2(4 - \lambda).$$

Eigenvalues: $\lambda = 2$ (algebraic multiplicity 2) and $\lambda = 4$.

Check: $2 + 2 + 4 = 8 = \operatorname{tr} A = 2 + 3 + 3 \checkmark$; $2 \cdot 2 \cdot 4 = 16$, and $\det A = 2(9 - 1) = 16 \checkmark$.

Eigenspace for $\lambda = 4$.

$$A - 4I = \begin{bmatrix} -2 & 0 & 0 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

So $x_1 = 0$ and $x_2 = x_3$. With $x_3 = 1$: $\mathbf{x} = (0, 1, 1)$.

Eigenspace for $\lambda = 2$.

$$A - 2I = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

One pivot, two free variables: $\dim E_2 = 2$. The single equation is $x_1 + x_2 + x_3 = 0$, giving special solutions

$$\mathbf{x} = (-1, 1, 0) \quad \text{and} \quad \mathbf{x} = (-1, 0, 1).$$

Multiplicities. For $\lambda = 2$: algebraic multiplicity 2 (double root) and geometric multiplicity 2 ($\dim E_2 = 2$). They match, so A is diagonalizable — see Chapter 12.

11.3 Properties of eigenvalues

Theorem 11.2 Tutorial 06, Q5 — trace and determinant

If $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A listed with algebraic multiplicity, then

$$\operatorname{tr} A = \lambda_1 + \lambda_2 + \dots + \lambda_n, \quad \det A = \lambda_1 \lambda_2 \dots \lambda_n.$$

These hold whether or not A is diagonalizable.

Proof. Write the characteristic polynomial in the monic form $p_A(t) = \det(tI - A) = t^n - c_1 t^{n-1} + \dots + (-1)^n c_n$.

Expanding $\det(tI - A)$, the only way to reach degree n or $n - 1$ is to take the product of all n diagonal entries $(t - a_{ii})$; any other term in the Leibniz expansion misses at least two diagonal entries and so has degree $\leq n - 2$. Multiplying out,

$$\prod_{i=1}^n (t - a_{ii}) = t^n - \left(\sum_i a_{ii} \right) t^{n-1} + \dots,$$

so $c_1 = \operatorname{tr} A$. Setting $t = 0$ gives $p_A(0) = \det(-A) = (-1)^n \det A$, and from the polynomial form $p_A(0) = (-1)^n c_n$; hence $c_n = \det A$.

On the other hand, factoring over $\mathbb{C}$, $p_A(t) = \prod_i (t - \lambda_i)$, whose t^{n-1} coefficient is $-\sum \lambda_i$ and whose constant term is $(-1)^n \prod \lambda_i$. Comparing coefficients gives both identities.

Part (c). Nothing in this argument used diagonalizability — only the factorization of p_A over $\mathbb{C}$, which the Fundamental Theorem of Algebra always supplies. So the identities survive repeated eigenvalues and defective matrices alike. ■

Remark • the two-second sanity check

Before writing down any eigenvalues, compute $\operatorname{tr} A$ and $\det A$. If your eigenvalues do not sum to the trace and multiply to the determinant, you have made an arithmetic error — go back. For 2×2 matrices this is often faster than the characteristic polynomial: solve $\lambda_1 + \lambda_2 = \operatorname{tr} A$, $\lambda_1 \lambda_2 = \det A$ directly.

Theorem 11.3 Tutorial 06, Q2 — powers and polynomials

Suppose $Ax = \lambda x$ with $x \neq 0$. Then:

- (a) $A^k x = \lambda^k x$ for every positive integer k ;
- (b) $p(A)x = p(\lambda)x$ for every polynomial p ;
- (c) $A^{-1}x = \lambda^{-1}x$ when A is invertible;
- (d) $(A + cI)x = (\lambda + c)x$.

In every case the *eigenvector is unchanged* — only the eigenvalue transforms.

Proof. (a) By induction on k . Base case $k = 1$ is the hypothesis. Assume $A^k x = \lambda^k x$. Then

$$A^{k+1}x = A(A^k x) = A(\lambda^k x) = \lambda^k (Ax) = \lambda^k (\lambda x) = \lambda^{k+1}x,$$

completing the induction.

(b) With $p(t) = a_0 + a_1t + \cdots + a_mt^m$,

$$p(A)x = \sum_{k=0}^m a_k A^k x = \sum_{k=0}^m a_k \lambda^k x = \left(\sum_{k=0}^m a_k \lambda^k \right) x = p(\lambda)x.$$

(Here $A^0 = I$, so the constant term behaves correctly.)

(c) If A is invertible then $\lambda \neq 0$ (else A would be singular). Multiply $Ax = \lambda x$ by A^{-1} : $x = \lambda A^{-1}x$, so $A^{-1}x = \frac{1}{\lambda}x$.

(d) $(A + cI)x = Ax + cx = (\lambda + c)x$. ■

Worked Example 11.3 • Tutorial 06, Q2(c) — the converse fails

Every eigenvector of A is an eigenvector of A^m . Give a counterexample showing an eigenvector of A^m need not be an eigenvector of A .

Take

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{so} \quad A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Every nonzero vector is an eigenvector of $A^2 = I$ (with eigenvalue 1). But the eigenvectors of A itself are only the multiples of $(1, 1)$ (with $\lambda = 1$) and of $(1, -1)$ (with $\lambda = -1$). So $x = (1, 0)$ is an eigenvector of A^2 but not of A : indeed $A(1, 0) = (0, 1)$, which is not a multiple of $(1, 0)$.

Why the asymmetry? Squaring merges information: λ and $-\lambda$ both become λ^2 , so the eigenspaces of A^2 can be strictly larger than those of A . Here $E_1(A)$ and $E_{-1}(A)$ are lines that fuse into all of $\mathbb{R}^2$ for A^2 .

Theorem 11.4 Tutorial 06, Q4 — distinct eigenvalues give independence

Eigenvectors corresponding to *distinct* eigenvalues are linearly independent.

Proof for two, then in general. Two vectors. Let $Ax_1 = \lambda_1 x_1$, $Ax_2 = \lambda_2 x_2$ with $\lambda_1 \neq \lambda_2$, and suppose

$$c_1 x_1 + c_2 x_2 = \mathbf{0}. \quad (*)$$

Apply A : $c_1 \lambda_1 x_1 + c_2 \lambda_2 x_2 = \mathbf{0}$. Now subtract $\lambda_2 \times (*)$:

$$c_1(\lambda_1 - \lambda_2)x_1 = \mathbf{0}.$$

Since $\lambda_1 \neq \lambda_2$ and $x_1 \neq \mathbf{0}$, we get $c_1 = 0$; then $(*)$ forces $c_2 x_2 = \mathbf{0}$ and hence $c_2 = 0$.

General case, by induction on k . Suppose the result holds for $k - 1$ eigenvectors and let $c_1 x_1 + \cdots + c_k x_k = \mathbf{0}$. Apply A and subtract λ_k times the original:

$$\sum_{i=1}^{k-1} c_i(\lambda_i - \lambda_k)x_i = \mathbf{0}.$$

By the inductive hypothesis the $k - 1$ vectors $x_1, \dots, x_{k-1}$ are independent, so $c_i(\lambda_i - \lambda_k) = 0$ for each $i < k$. The eigenvalues are distinct, so $c_i = 0$ for $i < k$, and then $c_k = 0$ too. ■

Corollary 11.2 Tutorial 06, Q4(c)

An $n \times n$ matrix with n *distinct* eigenvalues has n independent eigenvectors, hence is diagonalizable.

Common Pitfall • the converse is false

Distinct eigenvalues are *sufficient* but not *necessary* for diagonalizability. I has the single eigenvalue 1 repeated n times and is already diagonal. See Chapter 12, Tutorial 06 Q6(d).

11.4 Eigenspaces are subspaces

Theorem 11.5 Tutorial 06, Q1(d)

For each eigenvalue λ of A , the eigenspace $E_\lambda = \ker(A - \lambda I)$ is a subspace of $\mathbb{R}^n$.

Proof. E_λ is by definition the null space of the matrix $A - \lambda I$, and null spaces are subspaces (Chapter 9). Directly: $A\mathbf{0} = \mathbf{0} = \lambda\mathbf{0}$, so $\mathbf{0} \in E_\lambda$. If $\mathbf{u}, \mathbf{v} \in E_\lambda$ and $c, d \in \mathbb{R}$ then

$$A(c\mathbf{u} + d\mathbf{v}) = cA\mathbf{u} + dA\mathbf{v} = c\lambda\mathbf{u} + d\lambda\mathbf{v} = \lambda(c\mathbf{u} + d\mathbf{v}),$$

so $c\mathbf{u} + d\mathbf{v} \in E_\lambda$. ■

Definition 11.3 Algebraic and geometric multiplicity

For an eigenvalue λ of A :

- the **algebraic multiplicity** $\text{AM}(\lambda)$ is its multiplicity as a root of p_A ;
- the **geometric multiplicity** $\text{GM}(\lambda)$ is $\dim E_\lambda = n - \text{rank}(A - \lambda I)$.

Always

$$1 \leq \text{GM}(\lambda) \leq \text{AM}(\lambda).$$

When $\text{GM} < \text{AM}$ the eigenvalue is called **defective**.

Worked Example 11.4 • A defective matrix

Analyse $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$.

$$\det(A - \lambda I) = \begin{vmatrix} 3 - \lambda & 1 \\ 0 & 3 - \lambda \end{vmatrix} = (3 - \lambda)^2.$$

So $\lambda = 3$ with $\text{AM} = 2$. For the eigenspace,

$$A - 3I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

which has rank 1, so $\text{GM} = 2 - 1 = 1$. Solving $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0}$ gives $x_2 = 0$, so $E_3 = \text{span}\{(1, 0)\}$ — only a line.

Since $GM = 1 < 2 = AM$, this matrix is **defective** and cannot be diagonalized: there is no basis of $\mathbb{R}^2$ made of its eigenvectors. It is a *shear* (Chapter 15), and a shear genuinely has only one invariant direction.

11.5 Special matrices and their eigenvalues

Matrix type	Eigenvalues	Eigenvectors
Diagonal / triangular	the diagonal entries	read off directly for diagonal
Symmetric $S^T = S$	all real	can be chosen orthonormal
Orthogonal $Q^T Q = I$	$ \lambda = 1$ (possibly complex)	orthogonal
Projection $P^2 = P$	only 0 and 1	$\mathbf{C}(P)$ and $\mathbf{N}(P)$
Singular	includes $\lambda = 0$	$\mathbf{N}(A) = E_0$
Rotation in $\mathbb{R}^2$ ($\theta \neq 0, \pi$)	complex, $e^{\pm i\theta}$	no real eigenvectors
Nilpotent $A^k = O$	all $\lambda = 0$	defective

Proposition 11.1 Triangular matrices

If A is triangular, its eigenvalues are exactly its diagonal entries.

Proof. $A - \lambda I$ is triangular with diagonal entries $a_{ii} - \lambda$, and the determinant of a triangular matrix is the product of its diagonal entries (rule D7). So $p_A(\lambda) = \prod_i (a_{ii} - \lambda)$, whose roots are the a_{ii} . ■

Worked Example 11.5 • A rotation has no real eigenvectors

Find the eigenvalues of the 90° rotation $R = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

$$\det(R - \lambda I) = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0 \implies \lambda = \pm i.$$

There are *no real* eigenvalues — and geometrically that is obvious: a rotation by 90° moves *every* nonzero vector off its own line, so no real direction is preserved.

The eigenvalues satisfy $|\lambda| = 1$, as they must for an orthogonal matrix, and $\lambda_1 \lambda_2 = i(-i) = 1 = \det R$ ✓, $\lambda_1 + \lambda_2 = 0 = \text{tr } R$ ✓.

Contrast the general rotation $R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$, whose eigenvalues are $\cos \theta \pm i \sin \theta = e^{\pm i\theta}$. Real eigenvectors exist only when $\sin \theta = 0$, i.e. $\theta = 0$ (identity) or $\theta = \pi$ (point reflection, $\lambda = -1$ twice).

11.6 Similar matrices

Definition 11.4 Similarity

A and B are **similar**, written $A \sim B$, if there is an invertible P with

$$B = P^{-1}AP.$$

Theorem 11.6 Tutorial 06, Q7 — similarity preserves the spectrum

If $B = P^{-1}AP$ then:

- (a) A and B have the same characteristic polynomial, hence the same eigenvalues with the same algebraic multiplicities;
- (b) if $Ax = \lambda x$ then $P^{-1}x$ is an eigenvector of B for the same λ ;
- (c) A is diagonalizable $\iff B$ is diagonalizable;
- (d) $\text{tr } A = \text{tr } B$ and $\det A = \det B$.

Proof. (a)

$$\begin{aligned} \det(B - \lambda I) &= \det(P^{-1}AP - \lambda P^{-1}P) = \det(P^{-1}(A - \lambda I)P) \\ &= \det(P^{-1}) \det(A - \lambda I) \det(P) = \frac{1}{\det P} \det(A - \lambda I) \det P = \det(A - \lambda I). \end{aligned}$$

(b) Put $y = P^{-1}x$, which is nonzero since P^{-1} is invertible. Then

$$By = P^{-1}APP^{-1}x = P^{-1}Ax = P^{-1}\lambda x = \lambda y.$$

(c) If $A = X\Lambda X^{-1}$ then $B = P^{-1}X\Lambda X^{-1}P = (P^{-1}X)\Lambda(P^{-1}X)^{-1}$, so B is diagonalized by $P^{-1}X$. The converse is symmetric since $A = PBP^{-1}$.

(d) Immediate from (a), or directly: $\text{tr}(P^{-1}AP) = \text{tr}(APP^{-1}) = \text{tr } A$ by the cyclic property, and $\det(P^{-1}AP) = \det A$ as computed above. ■

Intuition • Tutorial 06, Q7(d) — what similarity means

A and $B = P^{-1}AP$ are the **same linear transformation written in two different bases**, with P the change-of-basis matrix. Reading By from right to left: P converts y from the new coordinates into the old ones, A acts, and P^{-1} converts back.

Since the underlying geometric operation is unchanged, any quantity that is intrinsic to the transformation — eigenvalues, trace, determinant, rank — must be identical for A and B . Only basis-dependent things like individual matrix entries differ. This is exactly why diagonalization is worth doing: it hunts for the basis in which the entries are as simple as possible.

11.7 Exam preparation

Exam Preparation • Chapter 11

The procedure

$$\det(A - \lambda I) = 0 \Rightarrow \lambda_i \quad \text{then} \quad (A - \lambda_i I)x = \mathbf{0} \Rightarrow x_i.$$

Identities to check every answer against

$$\sum_i \lambda_i = \operatorname{tr} A, \quad \prod_i \lambda_i = \det A.$$

Transformation rules (eigenvector unchanged)

A^k	λ^k	A^{-1}	$1/\lambda$
$A + cI$	$\lambda + c$	cA	$c\lambda$
$p(A)$	$p(\lambda)$	A^T	λ (same eigenvalues)

Facts

- Eigenvectors must be nonzero; eigenvalues may be zero.
- $\lambda = 0$ occurs $\iff A$ is singular.
- Distinct eigenvalues $\Rightarrow$ independent eigenvectors.
- $1 \leq \operatorname{GM}(\lambda) \leq \operatorname{AM}(\lambda)$.
- Triangular $\Rightarrow$ eigenvalues are the diagonal entries.
- Similar matrices share eigenvalues, trace, determinant and rank — but not eigenvectors.

Pitfalls

- Writing $\det(A - \lambda)$ instead of $\det(A - \lambda I)$.
- Reporting $x = \mathbf{0}$ as an eigenvector.
- Assuming real matrices have real eigenvalues — rotations do not.
- Assuming a repeated eigenvalue means non-diagonalizable.
- Forgetting that A and A^T share eigenvalues but generally *not* eigenvectors.

Self-check

1. Find the eigenvalues and eigenvectors of $\begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}$.
2. A is 3×3 with $\operatorname{tr} A = 6$, $\det A = 6$, and $\lambda = 1$ is an eigenvalue. Find the other two.
3. Prove that A and A^T have the same eigenvalues.
4. If $A^2 = A$ (a projection), what are the possible eigenvalues?

Answers

1. $\operatorname{tr} = 7$, $\det = 10 - 4 = 6$, so $\lambda^2 - 7\lambda + 6 = 0$ gives $\lambda = 1, 6$. For $\lambda = 1$: $\begin{bmatrix} 4 & 4 \\ 1 & 1 \end{bmatrix} x = \mathbf{0}$ gives $x = (1, -1)$. For $\lambda = 6$: $\begin{bmatrix} -1 & 4 \\ 1 & -4 \end{bmatrix} x = \mathbf{0}$ gives $x = (4, 1)$.
2. $1 + \lambda_2 + \lambda_3 = 6$ and $1 \cdot \lambda_2 \lambda_3 = 6$, so $\lambda_2 + \lambda_3 = 5$ and $\lambda_2 \lambda_3 = 6$: the roots of $t^2 - 5t + 6 = 0$, namely 2 and 3.
3. $\det(A^T - \lambda I) = \det((A - \lambda I)^T) = \det(A - \lambda I)$ by rule D9. Identical characteristic polynomials, hence identical eigenvalues.
4. If $Ax = \lambda x$ then $\lambda x = Ax = A^2x = \lambda^2 x$, so $(\lambda^2 - \lambda)x = \mathbf{0}$ with $x \neq \mathbf{0}$. Hence $\lambda(\lambda - 1) = 0$: only $\lambda = 0$ or $\lambda = 1$.

Diagonalization and the Spectral Theorem

Lecture 6 states the purpose plainly: *diagonalizing a matrix simplifies complex linear transformations and heavy matrix computations by converting them into a diagonal matrix*. If $A = X\Lambda X^{-1}$ then $A^{100} = X\Lambda^{100}X^{-1}$, and raising a diagonal matrix to the hundredth power is n scalar exponentiations. This chapter proves when diagonalization is possible, what to do when it is not, and why symmetric matrices are always the best case.

Strang §6.2 (“Diagonalizing a Matrix”), §6.3 (“Symmetric Positive Definite Matrices”). STRANG | 6th ed. Lecture 6, slides 20–29; Tutorial 06 Q6 and Part B.

12.1 The eigen-decomposition

Theorem 12.1 Diagonalization

Let $A \in \mathbb{R}^{n \times n}$ have n linearly independent eigenvectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ with eigenvalues $\lambda_1, \dots, \lambda_n$. Put them into matrices

$$X = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n], \quad \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

Then X is invertible and

$$AX = X\Lambda, \quad X^{-1}AX = \Lambda, \quad A = X\Lambda X^{-1}.$$

Such an A is called **diagonalizable**.

Tutorial 06, Q6(a). Multiply A by X column by column. Using the column view of matrix multiplication,

$$AX = [A\mathbf{x}_1 \ A\mathbf{x}_2 \ \cdots \ A\mathbf{x}_n] = [\lambda_1\mathbf{x}_1 \ \lambda_2\mathbf{x}_2 \ \cdots \ \lambda_n\mathbf{x}_n].$$

On the other hand, multiplying X on the right by the diagonal matrix Λ scales column j of X by λ_j :

$$X\Lambda = [\lambda_1\mathbf{x}_1 \ \lambda_2\mathbf{x}_2 \ \cdots \ \lambda_n\mathbf{x}_n].$$

The two agree, so $AX = X\Lambda$. Because the eigenvectors are independent, X has independent columns and is therefore invertible; multiplying on the left by X^{-1} gives $X^{-1}AX = \Lambda$, and on the right by X^{-1} gives $A = X\Lambda X^{-1}$. ■

Common Pitfall • the order must match

Column j of X must be an eigenvector for the eigenvalue in position j of Λ . Swap two columns of X and you must swap the corresponding diagonal entries of Λ . The decomposi-

tion is not unique — reordering, and rescaling any eigenvector, both give valid alternatives. Also note the side: $A = X\Lambda X^{-1}$, with X on the *left*. The common slip $A = X^{-1}\Lambda X$ diagonalizes a different matrix.

Theorem 12.2 Tutorial 06, Q6(b) — powers

If $A = X\Lambda X^{-1}$ then for every positive integer k ,

$$A^k = X\Lambda^k X^{-1}, \quad \Lambda^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k).$$

If in addition A is invertible, the same formula holds for negative k .

Proof. By induction. For $k = 1$ it is the hypothesis. Assuming $A^k = X\Lambda^k X^{-1}$,

$$A^{k+1} = A^k A = (X\Lambda^k X^{-1})(X\Lambda X^{-1}) = X\Lambda^k \underbrace{X^{-1}X}_I \Lambda X^{-1} = X\Lambda^{k+1} X^{-1}.$$

The inner $X^{-1}X$ collapsing to I is the whole point: in the middle of a long product, all the X 's annihilate one another and only the Λ 's survive. ■

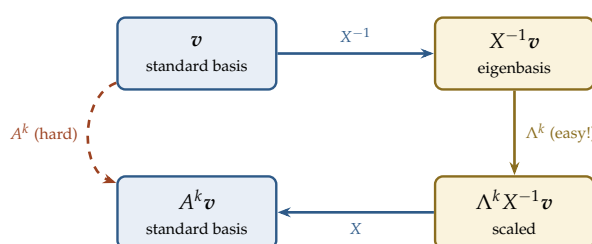


Figure 12.1: Why diagonalization pays. The direct route A^k needs $k - 1$ matrix products; the detour through the eigenbasis needs only n scalar powers.

Worked Example 12.1 • Lecture 6 — diagonalizing and taking a power

Diagonalize $A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$ and compute A^5 .

From Chapter 11: $\lambda_1 = 2$ with $\mathbf{x}_1 = (1, 1)$, and $\lambda_2 = 3$ with $\mathbf{x}_2 = (2, 1)$.

Assemble.

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

Since $\det X = 1 - 2 = -1$,

$$X^{-1} = \frac{1}{-1} \begin{bmatrix} 1 & -2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}.$$

Verify $A = X\Lambda X^{-1}$. First

$$X\Lambda = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 2 & 3 \end{bmatrix},$$

then

$$(X\Lambda)X^{-1} = \begin{bmatrix} 2 & 6 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -2+6 & 4-6 \\ -2+3 & 4-3 \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} = A. \checkmark$$

Fifth power. $\Lambda^5 = \text{diag}(2^5, 3^5) = \text{diag}(32, 243)$, so

$$\begin{aligned} A^5 &= X\Lambda^5X^{-1} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 32 & 0 \\ 0 & 243 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 32 & 486 \\ 32 & 243 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} -32+486 & 64-486 \\ -32+243 & 64-243 \end{bmatrix} = \begin{bmatrix} 454 & -422 \\ 211 & -179 \end{bmatrix}. \end{aligned}$$

Check: $\text{tr } A^5 = 454 - 179 = 275 = 32 + 243 = \lambda_1^5 + \lambda_2^5 \checkmark$, and $\det A^5 = (\det A)^5 = 6^5 = 7776$; directly $454(-179) - (-422)(211) = -81266 + 89042 = 7776 \checkmark$.

12.2 When is a matrix diagonalizable?

Theorem 12.3 Diagonalizability criterion

$A \in \mathbb{R}^{n \times n}$ is diagonalizable **if and only if** it has n linearly independent eigenvectors. Equivalently:

$$A \text{ is diagonalizable} \iff \text{GM}(\lambda) = \text{AM}(\lambda) \text{ for every eigenvalue } \lambda.$$

Situation	Diagonalizable?	Example
n distinct eigenvalues	always yes	$\begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$
Repeated λ , $\text{GM} = \text{AM}$	yes	I , or $\begin{bmatrix} 2 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$
Repeated λ , $\text{GM} < \text{AM}$	no (defective)	$\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$
Symmetric $S^T = S$	always yes , orthogonally	any S

Worked Example 12.2 • Tutorial 06, Q6(c)–(d) — repeated eigenvalues both ways

Explain, with examples, why a repeated eigenvalue does *not* automatically prevent diagonalization.

Repeated but diagonalizable. Take $A = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} = 5I$. Then $p_A(\lambda) = (5 - \lambda)^2$, so $\lambda = 5$ has $\text{AM} = 2$. But $A - 5I = O$, which has rank 0, so $\text{GM} = 2 - 0 = 2$. Every vector in $\mathbb{R}^2$ is an eigenvector; the matrix is already diagonal.
More substantially, the 3×3 example of Chapter 11 with $p_A(\lambda) = (2 - \lambda)^2(4 - \lambda)$ had $\text{GM}(2) = 2 = \text{AM}(2)$ and is diagonalizable, with

$$X = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad \Lambda = \text{diag}(4, 2, 2).$$

Repeated and not diagonalizable. Take $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$. Again $\text{AM} = 2$, but $A - 3I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ has rank 1, so $\text{GM} = 1$. Only one independent eigenvector exists, X cannot be formed, and A is defective.

The moral. What matters is not whether eigenvalues repeat, but whether each eigenvalue supplies *as many independent eigenvectors as its algebraic multiplicity demands*. Repetition is an opportunity for failure, not a guarantee of it.

Remark • when diagonalization fails: the Jordan form

Every square matrix is similar to a **Jordan form** — block diagonal with blocks $\begin{bmatrix} \lambda & 1 \\ & \lambda & \ddots \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$.
 Diagonalizable matrices are exactly those whose Jordan blocks are all 1×1 . This is beyond SCS 1306 but worth knowing exists: it explains what the “missing” eigenvectors are replaced by.

12.3 Symmetric matrices: the spectral theorem

Theorem 12.4 Spectral Theorem

Let $S \in \mathbb{R}^{n \times n}$ be symmetric ($S^T = S$). Then:

- (i) all eigenvalues of S are **real**;
- (ii) eigenvectors for distinct eigenvalues are **orthogonal**;
- (iii) S has an **orthonormal basis** of eigenvectors — even when eigenvalues repeat;
- (iv) consequently

$$S = Q\Lambda Q^T$$

with Q orthogonal ($Q^{-1} = Q^T$) and Λ real diagonal.

Every symmetric matrix is diagonalizable, and diagonalizable *orthogonally*.

Proof of (ii). Let $Sx = \lambda x$ and $Sy = \mu y$ with $\lambda \neq \mu$. Then

$$\lambda(x \cdot y) = (Sx)^T y = x^T S^T y = x^T Sy = \mu(x \cdot y),$$

where symmetry was used in the middle. Hence $(\lambda - \mu)(x \cdot y) = 0$, and since $\lambda \neq \mu$ we get $x \cdot y = 0$. ■

Intuition • the best possible situation

$S = Q\Lambda Q^T$ says: rotate into the eigenvector frame (Q^T), scale each axis independently (Λ), rotate back (Q). No shearing, no distortion of angles — a symmetric matrix acts as pure stretching along n mutually perpendicular axes. This is why $A^T A$ being symmetric (Chapter 6) is the key that unlocks the SVD.

Worked Example 12.3 • Orthogonal diagonalization of a symmetric matrix

Diagonalize $S = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ orthogonally.

Eigenvalues. $\text{tr } S = 6$, $\det S = 9 - 1 = 8$, so $\lambda^2 - 6\lambda + 8 = 0$ gives $\lambda = 2$ and $\lambda = 4$.

Eigenvectors. For $\lambda = 4$: $(S - 4I) = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$, giving $x_1 = x_2$, so $x = (1, 1)$. For $\lambda = 2$:

$(S - 2I) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, giving $x_1 = -x_2$, so $x = (1, -1)$.

Orthogonality check. $(1, 1) \cdot (1, -1) = 1 - 1 = 0$ ✓ — as the spectral theorem guarantees, without our having to arrange it.

Normalise.

$$q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad q_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}.$$

Q is orthogonal: $Q^T Q = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = I$ ✓.

Verify.

$$Q \Lambda Q^T = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 4 & 2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 6 & 2 \\ 2 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = S. \quad \checkmark$$

Note Q here is a reflection ($\det Q = -1$); using $q_2 = (-1, 1)/\sqrt{2}$ instead would give a rotation. Both are valid.

12.4 Positive definite matrices

Definition 12.1 Positive definite / semidefinite

A symmetric S is

- **positive definite** if $x^T S x > 0$ for all $x \neq 0$;
- **positive semidefinite** if $x^T S x \geq 0$ for all x .

Theorem 12.5 Equivalent tests

For a symmetric S , the following are equivalent:

- (1) S is positive definite;
- (2) all eigenvalues $\lambda_i > 0$;
- (3) all pivots in elimination are positive;
- (4) all leading principal minors are positive;
- (5) $S = A^T A$ for some A with independent columns.

Replacing every “ $>$ ” by “ $\geq$ ” gives the semidefinite versions (with (5) becoming $S = A^T A$ for any A).

(1) $\iff$ (2). Write $S = Q\Lambda Q^T$ and substitute $\mathbf{y} = Q^T \mathbf{x}$:

$$\mathbf{x}^T S \mathbf{x} = \mathbf{x}^T Q \Lambda Q^T \mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y} = \sum_{i=1}^n \lambda_i y_i^2.$$

Since Q is invertible, $\mathbf{x} \neq \mathbf{0} \iff \mathbf{y} \neq \mathbf{0}$. The sum is positive for all such $\mathbf{y}$ exactly when every $\lambda_i > 0$ (take $\mathbf{y} = \mathbf{e}_i$ to see the necessity). ■

Worked Example 12.4 • Testing positive definiteness

Is $S = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ positive definite?

By leading minors.

$$\Delta_1 = 2 > 0, \quad \Delta_2 = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 4 - 1 = 3 > 0,$$

$$\Delta_3 = \det S = 2(4 - 1) - (-1)(-2 - 0) = 6 - 2 = 4 > 0.$$

All positive, so S is positive definite.

By pivots. Elimination gives pivots $2, \frac{3}{2}, \frac{4}{3}$ — all positive, and their product $2 \cdot \frac{3}{2} \cdot \frac{4}{3} = 4 = \det S$ ✓.

By eigenvalues. They are $2 - \sqrt{2} \approx 0.586, 2, 2 + \sqrt{2} \approx 3.414$ — all positive. Their sum is $6 = \text{tr } S$ ✓ and their product is $(4 - 2) \cdot 2 = 4 = \det S$ ✓.

This S is Strang's second-difference matrix, which appeared in Chapter 2. It is positive definite for every size n .

12.5 An application: Markov chains

Worked Example 12.5 • Long-run behaviour of a Markov matrix

A city's population moves between two districts each year: 10% of district A moves to B, and 20% of B moves to A. With $\mathbf{p}_k = (a_k, b_k)$ the populations after k years,

$$\mathbf{p}_{k+1} = A \mathbf{p}_k, \quad A = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}.$$

What happens as $k \rightarrow \infty$?

Eigenvalues. $\text{tr } A = 1.7, \det A = 0.72 - 0.02 = 0.7$, so $\lambda^2 - 1.7\lambda + 0.7 = 0$, giving

$$\lambda_1 = 1, \quad \lambda_2 = 0.7.$$

($\lambda = 1$ is automatic for any Markov matrix, since the columns sum to 1 and hence A^T has eigenvector $(1, 1)$ for $\lambda = 1$.)

Eigenvectors. For $\lambda_1 = 1$: $(A - I) = \begin{bmatrix} -0.1 & 0.2 \\ 0.1 & -0.2 \end{bmatrix}$ gives $a = 2b$, so $\mathbf{x}_1 = (2, 1)$. For

$\lambda_2 = 0.7$: $(A - 0.7I) = \begin{bmatrix} 0.2 & 0.2 \\ 0.1 & 0.1 \end{bmatrix}$ gives $a = -b$, so $x_2 = (1, -1)$.

Long run. Writing $p_0 = c_1x_1 + c_2x_2$,

$$p_k = A^k p_0 = c_1(1)^k x_1 + c_2(0.7)^k x_2 \xrightarrow{k \rightarrow \infty} c_1 x_1,$$

because $0.7^k \rightarrow 0$. The population settles into the *steady state* proportional to $(2, 1)$ — two thirds in district A, one third in B — regardless of where it started.

This is the entire theory of Markov chains in one calculation: the eigenvalue 1 gives the equilibrium, and every other eigenvalue has $|\lambda| < 1$ and dies out. The *second largest* $|\lambda|$ controls how *fast* it converges.

12.6 Exam preparation

Exam Preparation • Chapter 12

The decompositions

$$A = X\Lambda X^{-1} \quad (\text{general, needs } n \text{ independent eigenvectors})$$

$$S = Q\Lambda Q^T \quad (\text{symmetric, always exists, } Q^{-1} = Q^T)$$

$$A^k = X\Lambda^k X^{-1}, \quad \Lambda^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$$

Diagonalizability checklist

n distinct eigenvalues	diagonalizable (sufficient, not necessary)
$\text{GM}(\lambda) = \text{AM}(\lambda)$ for all λ	diagonalizable (the real criterion)
$S^T = S$	diagonalizable, and orthogonally so
some $\text{GM} < \text{AM}$	defective, not diagonalizable

Pitfalls

- Mismatching the order of columns in X with the diagonal of Λ .
- Writing $A = X^{-1}\Lambda X$.
- Concluding “repeated eigenvalue $\Rightarrow$ not diagonalizable”. Check GM.
- For symmetric S , forgetting to *normalise* the eigenvectors before claiming Q is orthogonal.
- Assuming A diagonalizable $\Rightarrow A$ invertible. A diagonalizable matrix can have $\lambda = 0$ (e.g. any projection).

Self-check

1. Diagonalize $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$ and use it to find A^4 .
2. Show that if A is diagonalizable with all $|\lambda_i| < 1$, then $A^k \rightarrow O$.
3. Prove that a projection matrix P ($P^T = P$, $P^2 = P$) is diagonalizable with Λ containing only 0s and 1s.
4. Is $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ diagonalizable?

Answers

1. Triangular, so $\lambda = 1, 3$. For $\lambda = 1$: $\begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} x = 0$ gives $x = (1, 0)$. For $\lambda = 3$: $\begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix} x = 0$ gives $x = (1, 1)$. So $X = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $X^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, and $A^4 = X \text{diag}(1, 81) X^{-1} = \begin{bmatrix} 1 & 80 \\ 0 & 81 \end{bmatrix}$.

2. $A^k = X\Lambda^k X^{-1}$ and every $\lambda_i^k \rightarrow 0$, so $\Lambda^k \rightarrow O$ and hence $A^k \rightarrow XOX^{-1} = O$.
3. P is symmetric, so the spectral theorem applies and $P = Q\Lambda Q^T$. From $P^2 = P$ we get $\lambda^2 = \lambda$ for every eigenvalue (Chapter 11 self-check), so each $\lambda \in \{0, 1\}$. The count of 1s equals $\text{rank } P$.
4. No. It is triangular with $\lambda = 2$ of AM = 3, but $A - 2I = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ has rank 2, so GM = $3 - 2 = 1 < 3$. This is a single Jordan block — maximally defective.

The Singular Value Decomposition

Diagonalization only works for square matrices with enough eigenvectors. The SVD works for *every* matrix — rectangular, singular, defective, anything — and it does so with orthonormal bases on both sides. The lecture notes open with the key manoeuvre: *we need a square matrix; $A^T A$ and AA^T will be square, and they will also be symmetric.* Apply the spectral theorem to those, transfer the result back, and the SVD falls out. It is the climax of the course and the foundation of PCA, image compression and low-rank approximation.

Strang §7.1 (“Singular Values and Singular Vectors”), §7.2 (“Bases and Matrices in the SVD”), §7.3 (“Principal Component Analysis”), §7.4 (“The Geometry of the SVD”). The supplementary SVD lecture notes follow this route exactly.

13.1 The statement

Theorem 13.1 Singular Value Decomposition

Every matrix $A \in \mathbb{R}^{m \times n}$ of rank r factors as

$$A = U \Sigma V^T$$

where

- $U \in \mathbb{R}^{m \times m}$ is orthogonal; its columns $u_1, \dots, u_m$ are the **left singular vectors**;
- $V \in \mathbb{R}^{n \times n}$ is orthogonal; its columns $v_1, \dots, v_n$ are the **right singular vectors**;
- $\Sigma \in \mathbb{R}^{m \times n}$ is “diagonal” with non-negative entries

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots,$$

the **singular values**.

Exactly $r = \text{rank } A$ singular values are nonzero.

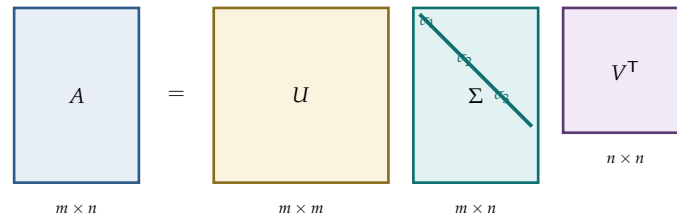


Figure 13.1: Shapes in the SVD. U and V are square and orthogonal; Σ has the same shape as A and carries the singular values on its main diagonal in decreasing order.

Intuition • what the three factors do

Read $Ax = U\Sigma V^T x$ from right to left:

1. V^T **rotates** (or reflects) the input into the frame of the right singular vectors — no stretching, since V is orthogonal;
2. Σ **stretches** each of those axes by σ_i , and may pad with zeros or drop coordinates to change the dimension;
3. U **rotates** the result into the output frame.

So **every** linear map, however complicated, is a rotation followed by an axis-aligned stretch followed by another rotation. There is nothing else a matrix can do.

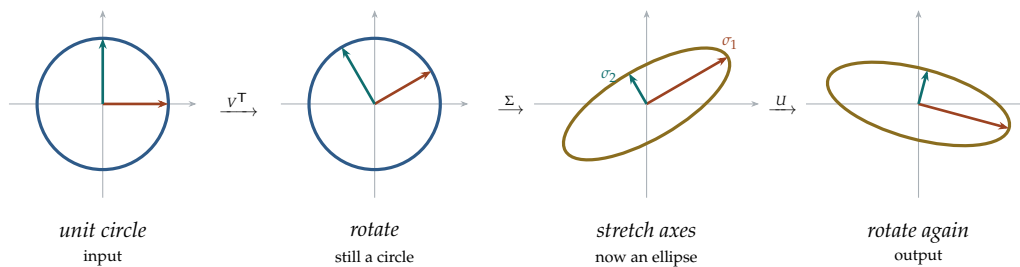


Figure 13.2: The geometry of the SVD. Any matrix maps the unit sphere to an ellipsoid; the singular values are the lengths of its semi-axes, and the left singular vectors are their directions.

13.2 Why it works

Theorem 13.2 Construction from $A^T A$ and AA^T

Let $A \in \mathbb{R}^{m \times n}$.

- (i) $A^T A$ is symmetric positive semidefinite; by the spectral theorem $A^T A = V\Lambda V^T$ with V orthogonal and $\lambda_1 \geq \dots \geq \lambda_n \geq 0$.
- (ii) The singular values are $\sigma_i = \sqrt{\lambda_i}$.
- (iii) The right singular vectors v_i are the eigenvectors of $A^T A$.
- (iv) For $i \leq r$, the left singular vectors are

$$u_i = \frac{Av_i}{\sigma_i},$$

and these are automatically orthonormal eigenvectors of AA^T .

Why the u_i come out orthonormal — the crucial step.

$$u_i \cdot u_j = \frac{(Av_i)^T(Av_j)}{\sigma_i\sigma_j} = \frac{v_i^T(A^T A)v_j}{\sigma_i\sigma_j} = \frac{v_i^T(\lambda_j v_j)}{\sigma_i\sigma_j} = \frac{\lambda_j(v_i \cdot v_j)}{\sigma_i\sigma_j}.$$

If $i \neq j$ then $v_i \cdot v_j = 0$ and the whole thing is 0. If $i = j$ then $v_i \cdot v_i = 1$ and the value is $\lambda_i/\sigma_i^2 = 1$. So $\{u_i\}$ is orthonormal.

Finally, $Av_i = \sigma_i u_i$ for every i says exactly $AV = U\Sigma$, and multiplying on the right by $V^T = V^{-1}$ gives $A = U\Sigma V^T$. ■

Proposition 13.1 Relation to eigenvalues

$$A^T A = V\Sigma^T \Sigma V^T, \quad AA^T = U\Sigma \Sigma^T U^T.$$

So $A^T A$ and AA^T have the same nonzero eigenvalues $\sigma_1^2, \dots, \sigma_r^2$; they differ only in how many zeros they carry.

Common Pitfall • singular values are not eigenvalues

For a general A the two are unrelated: $\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ has both eigenvalues 0 but singular values 2 and 0. They coincide only when A is symmetric positive semidefinite; for symmetric A in general, $\sigma_i = |\lambda_i|$.

13.3 Computing an SVD

Method SVD by hand

1. Form $A^T A$ (choose the smaller of $A^T A$, AA^T).
2. Find its eigenvalues λ_i and orthonormal eigenvectors v_i ; order them so $\lambda_1 \geq \lambda_2 \geq \dots$
3. Set $\sigma_i = \sqrt{\lambda_i}$ and build Σ .
4. For each $\sigma_i > 0$, set $u_i = Av_i/\sigma_i$.
5. If $m > r$, extend $\{u_i\}$ to an orthonormal basis of $\mathbb{R}^m$ (any basis of $N(A^T)$ will do).
6. Assemble $A = U\Sigma V^T$ and verify.

Worked Example 13.1 • A full 2×2 SVD

Find the SVD of $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$.

Step 1.

$$A^T A = \begin{bmatrix} 3 & 4 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}.$$

Step 2 — eigenvalues. $\text{tr} = 50$, $\det = 625 - 400 = 225$, so $\lambda^2 - 50\lambda + 225 = 0$ gives

$$\lambda = \frac{50 \pm \sqrt{2500 - 900}}{2} = \frac{50 \pm 40}{2} \implies \lambda_1 = 45, \lambda_2 = 5.$$

Eigenvectors of the symmetric matrix $\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$: for $\lambda_1 = 45$, $(A^T A - 45I) = \begin{bmatrix} -20 & 20 \\ 20 & -20 \end{bmatrix}$ gives $v_1 \propto (1, 1)$; for $\lambda_2 = 5$, $\begin{bmatrix} 20 & 20 \\ 20 & 20 \end{bmatrix}$ gives $v_2 \propto (1, -1)$. Normalising,

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Step 3 — singular values.

$$\sigma_1 = \sqrt{45} = 3\sqrt{5}, \quad \sigma_2 = \sqrt{5}, \quad \Sigma = \begin{bmatrix} 3\sqrt{5} & 0 \\ 0 & \sqrt{5} \end{bmatrix}.$$

Step 4 — left singular vectors.

$$Av_1 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 3 \\ 9 \end{bmatrix}, \quad u_1 = \frac{Av_1}{3\sqrt{5}} = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ 9 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

$$Av_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \quad u_2 = \frac{Av_2}{\sqrt{5}} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

Check orthonormality: $u_1 \cdot u_2 = \frac{1}{10}(3 - 3) = 0 \checkmark$, $\|u_1\|^2 = \frac{1}{10}(1 + 9) = 1 \checkmark$.

Result.

$$A = \underbrace{\frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 3\sqrt{5} & 0 \\ 0 & \sqrt{5} \end{bmatrix}}_\Sigma \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{V^T}.$$

(V is symmetric here, so $V^T = V$.)

Sanity checks. $\sigma_1 \sigma_2 = 3\sqrt{5} \cdot \sqrt{5} = 15 = |\det A| = |15 - 0| \checkmark$. $\sigma_1^2 + \sigma_2^2 = 45 + 5 = 50 = 3^2 + 0^2 + 4^2 + 5^2 \checkmark$ (the sum of squared singular values equals the sum of squared entries).

Worked Example 13.2 • A rectangular, rank-deficient example

Find the SVD of $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$.

Step 1. $A^T A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$.

Step 2. $\text{tr} = 4$, $\det = 0$, so $\lambda_1 = 4$, $\lambda_2 = 0$. Eigenvectors: $v_1 = \frac{1}{\sqrt{2}}(1, 1)$ for $\lambda = 4$, and $v_2 = \frac{1}{\sqrt{2}}(1, -1)$ for $\lambda = 0$.

Step 3. $\sigma_1 = 2$, $\sigma_2 = 0$, so $r = \text{rank } A = 1$ and

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{3 \times 2}.$$

Step 4.

$$Av_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \quad u_1 = \frac{1}{2} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Step 5 — extend U . We need two more orthonormal vectors in $\mathbb{R}^3$, spanning $\mathbf{N}(A^T)$.

Since $A^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$, its null space is $\{(a, b, c) : a + b = 0\} = \text{span}\{(1, -1, 0), (0, 0, 1)\}$.

Normalising,

$$u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Result.

$$A = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_\Sigma \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{V^T}.$$

Note how gracefully the SVD handles rank deficiency: the zero singular value simply reports $\text{rank } A = 1$, where diagonalization would have been undefined (the matrix is not even square).

13.4 The SVD and the four subspaces

Theorem 13.3 Orthonormal bases for everything

With $A = U\Sigma V^T$ of rank r :

Subspace	Orthonormal basis	Dimension
Column space $\mathbf{C}(A)$	$u_1, \dots, u_r$	r
Left null space $\mathbf{N}(A^T)$	$u_{r+1}, \dots, u_m$	$m - r$
Row space $\mathbf{C}(A^T)$	$v_1, \dots, v_r$	r
Null space $\mathbf{N}(A)$	$v_{r+1}, \dots, v_n$	$n - r$

Remark • the Big Picture, perfected

Chapter 9 gave bases for the four subspaces via elimination — but those bases were neither orthogonal nor normalised. The SVD supplies orthonormal bases for all four *simultaneously*, and in these bases A acts as simply as it possibly can:

$$Av_i = \sigma_i u_i \quad (i \leq r), \quad Av_i = \mathbf{0} \quad (i > r).$$

It is the diagonalization that always exists.

13.5 Low-rank approximation

Theorem 13.4 SVD as a sum of rank-one pieces

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T.$$

Each term is a rank-one matrix, and the terms are ordered by importance.

Theorem 13.5 Eckart–Young

Truncating after k terms,

$$A_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T,$$

gives the **best possible** rank- k approximation to A : for every matrix B of rank $\leq k$,

$$\|A - A_k\| \leq \|A - B\|$$

in both the spectral and Frobenius norms. The error is $\|A - A_k\|_2 = \sigma_{k+1}$.

Intuition • why this is compression

Storing A costs mn numbers. Storing A_k costs $k(m + n + 1)$ — the k vectors $\mathbf{u}_i$, the k vectors $\mathbf{v}_i$, and k scalars. For a 1000×1000 image, keeping $k = 50$ terms costs about 100,050 numbers instead of 1,000,000: a $10\times$ saving, and if the singular values decay quickly the picture is nearly indistinguishable. Lecture 6 makes the point: *keeping only the largest eigenvalues allows us to reconstruct images using much less storage.*

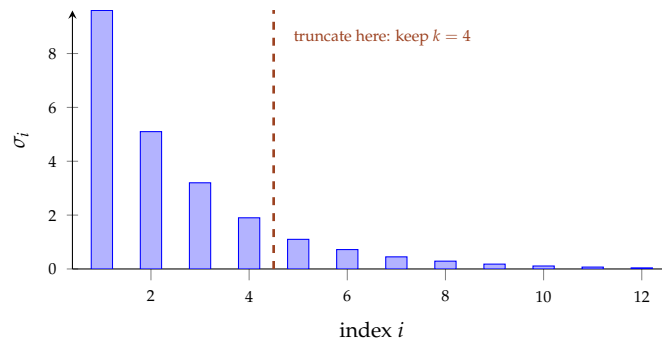


Figure 13.3: Typical singular value spectrum. Real data almost always has rapidly decaying σ_i , which is precisely why low-rank approximation works so well in practice.

Worked Example 13.3 • Rank-one approximation

For $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ above, find the best rank-one approximation.

Keep only the first term:

$$A_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = 3\sqrt{5} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} = \frac{3\sqrt{5}}{\sqrt{20}} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 1.5 & 1.5 \\ 4.5 & 4.5 \end{bmatrix}.$$

The error is

$$A - A_1 = \begin{bmatrix} 1.5 & -1.5 \\ -0.5 & 0.5 \end{bmatrix},$$

whose largest singular value is $\sigma_2 = \sqrt{5} \approx 2.236$ — exactly as Eckart–Young predicts.

Check: $\|A - A_1\|_F^2 = 2.25 + 2.25 + 0.25 + 0.25 = 5 = \sigma_2^2$ ✓.

No other rank-one matrix comes closer to A .

13.6 The pseudoinverse

Definition 13.1 Moore–Penrose pseudoinverse

For $A = U\Sigma V^T$ of rank r , define

$$A^+ = V\Sigma^+U^T, \quad \Sigma^+ = \text{diag}\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_r}, 0, \dots, 0\right)^T.$$

(Invert the nonzero singular values; leave the zeros alone; transpose the shape.)

Proposition 13.2

A^+ generalises the inverse:

- if A is invertible, $A^+ = A^{-1}$;
- if A has independent columns, $A^+ = (A^T A)^{-1} A^T$ — the least-squares formula of Chapter 10;
- in general $x^+ = A^+ b$ is the *shortest* least-squares solution.

13.7 PCA versus SVD

The supplementary lecture notes include a slide comparing the two. The relationship is worth stating precisely, since it is a common exam question.

	SVD	PCA
Applies to	any matrix A	a data matrix, after centring
Computes	$A = U\Sigma V^T$	eigen-decomposition of the covariance matrix
Requires	nothing	columns must be mean-centred first
Output	singular values and vectors	principal components and explained variance
Relation	—	PCA is the SVD of the centred data
Numerics	stable	unstable if done via $A^T A$ explicitly

Remark

In practice, *every* serious PCA implementation computes the SVD of the centred data matrix rather than forming the covariance matrix and diagonalizing it. Forming $A^T A$ squares

the condition number and can destroy small singular values entirely. Chapter 14 works through the details.

13.8 Exam preparation

Exam Preparation • Chapter 13

The decomposition

$$A = U\Sigma V^T, \quad U^T U = I_m, \quad V^T V = I_n, \quad \sigma_1 \geq \dots \geq \sigma_r > 0$$

$$A^T A = V\Sigma^T \Sigma V^T, \quad \sigma_i = \sqrt{\lambda_i(A^T A)}, \quad \mathbf{u}_i = \frac{A\mathbf{v}_i}{\sigma_i}$$

Reading facts off the SVD

rank A	number of nonzero σ_i
$\mathbf{C}(A)$	$\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$
$\mathbf{N}(A)$	$\text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$
$ \det A $ (square)	$\sigma_1 \sigma_2 \dots \sigma_n$
$\ A\ _2$	σ_1
condition number	σ_1 / σ_r
$\ A\ _F^2$	$\sum_i \sigma_i^2 = \sum_{i,j} a_{ij}^2$

Pitfalls

- Confusing σ_i with λ_i . Singular values are square roots of eigenvalues of $A^T A$, and are always ≥ 0 .
- Forgetting to order $\sigma_1 \geq \sigma_2 \geq \dots$.
- Forgetting to normalise the $\mathbf{v}_i$ (then the $\mathbf{u}_i$ come out wrong too).
- Writing $A = U\Sigma V$ instead of $U\Sigma V^T$.
- Trying to compute $\mathbf{u}_i = A\mathbf{v}_i / \sigma_i$ when $\sigma_i = 0$ — those $\mathbf{u}_i$ must come from $\mathbf{N}(A^T)$ instead.

Self-check

1. Find the singular values of $\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$.
2. If A is 5×3 with $\sigma = (4, 2, 0)$, give the dimensions of all four subspaces.
3. Show that $\|A\mathbf{x}\| \leq \sigma_1 \|\mathbf{x}\|$ for all $\mathbf{x}$, with equality attained.
4. For symmetric positive definite S , how do the SVD and the spectral decomposition compare?

Answers

1. $A^T A = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$, so $\lambda = 9, 4$ and $\sigma_1 = 3, \sigma_2 = 2$. Note $\sigma_i = |\lambda_i(A)|$, since A is symmetric here.
2. $r = 2, m = 5, n = 3$: $\dim \mathbf{C}(A) = 2, \dim \mathbf{N}(A^T) = 3, \dim \mathbf{C}(A^T) = 2, \dim \mathbf{N}(A) = 1$.
3. Write $\mathbf{x} = \sum c_i \mathbf{v}_i$, so $\|\mathbf{x}\|^2 = \sum c_i^2$. Then $A\mathbf{x} = \sum_{i \leq r} c_i \sigma_i \mathbf{u}_i$ and $\|A\mathbf{x}\|^2 = \sum_{i \leq r} c_i^2 \sigma_i^2 \leq \sigma_1^2 \sum c_i^2 = \sigma_1^2 \|\mathbf{x}\|^2$. Equality holds at $\mathbf{x} = \mathbf{v}_1$.
4. They are identical: $S = Q\Lambda Q^T$ with all $\lambda_i > 0$ is already of the form $U\Sigma V^T$ with $U = V = Q$ and $\Sigma = \Lambda$ (after ordering λ decreasingly). For symmetric S with negative eigenvalues, $\sigma_i = |\lambda_i|$ and the sign is absorbed into the corresponding column of U .

Principal Component Analysis

PCA is the SVD applied to data. Given measurements of many correlated variables, it finds the few directions along which the data actually varies, and lets you throw the rest away. Lecture 6 motivates it exactly this way: *instead of studying height and weight separately, eigenvectors identify the principal direction where the data varies the most*. This chapter derives PCA from the covariance matrix, connects it to the SVD, and works a complete numerical example.

Strang §7.3 (“Principal Component Analysis”) and §7.5. The supplementary *Principle Component Analysis* lecture notes; Lecture 6 slides 7–10. STRANG | 6th ed.

14.1 The problem PCA solves

Intuition • correlated measurements waste dimensions

Record the height and weight of 200 students. Plotted, the points form an elongated cloud tilted at roughly 45° — taller students tend to be heavier, so the two variables carry overlapping information. Describing each student by two numbers is wasteful: one number along the cloud’s long axis captures most of what distinguishes them.

PCA finds that long axis automatically. In two dimensions the saving is modest; with 1000 correlated sensor readings it can be enormous. Lecture 6’s example: *original data 1000 features; after PCA, 20 principal components — much faster machine learning with little loss of information*.

Definition 14.1 Goal of PCA

Given data $x_1, \dots, x_N \in \mathbb{R}^p$, find an orthonormal basis $w_1, w_2, \dots$ of $\mathbb{R}^p$ such that

- w_1 is the direction of *maximum variance* in the data;
- w_2 is the direction of maximum variance among those orthogonal to w_1 ;
- and so on.

These are the **principal components**. Projecting onto the first k of them gives the best k -dimensional summary of the data.

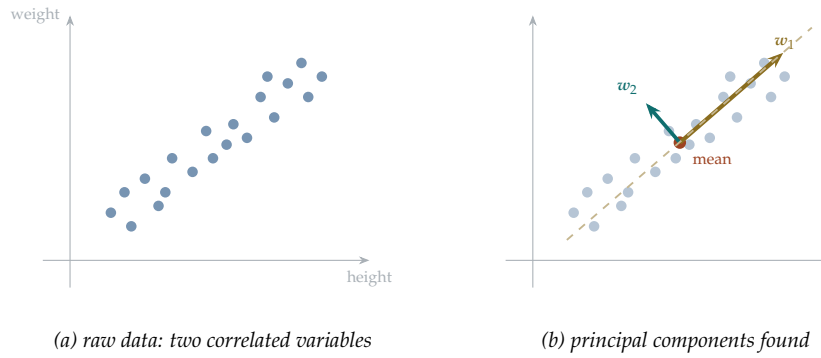


Figure 14.1: PCA finds the axes of the data cloud. w_1 points along the greatest spread; w_2 is perpendicular to it and captures the residual.

14.2 Centring and the covariance matrix

Definition 14.2 Centred data matrix

Stack the observations as *rows* of $X \in \mathbb{R}^{N \times p}$ (N samples, p variables). Compute the column means $\bar{x}_j = \frac{1}{N} \sum_i x_{ij}$ and subtract them:

$$B = X - \mathbf{1}\bar{x}^\top, \quad \text{so every column of } B \text{ has mean 0.}$$

B is the **centred** data matrix.

Common Pitfall • centring is not optional

Skip the centring and the first “principal component” will simply point from the origin towards the data cloud’s centre, telling you nothing about its shape. Every PCA begins by subtracting the mean. (Whether to also divide by the standard deviation — *standardising* — depends on whether the variables share units; if they do not, standardise.)

Definition 14.3 Sample covariance matrix

$$S = \frac{1}{N-1} B^\top B \in \mathbb{R}^{p \times p}.$$

Entry (j, k) is the sample covariance of variables j and k ; the diagonal entries are the variances. S is symmetric and positive semidefinite (Chapter 6).

Theorem 14.1 PCA via eigen-decomposition

Let $S = Q\Lambda Q^\top$ be the spectral decomposition, with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Then:

- (i) the columns q_i of Q are the principal components (loadings);
- (ii) λ_i is the variance of the data along q_i ;
- (iii) the total variance is $\sum_i \lambda_i = \text{tr } S$;
- (iv) the **proportion of variance explained** by the first k components is

$$\frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_p}.$$

Why the top eigenvector maximises variance. The variance of the data projected onto a unit vector w is

$$\frac{1}{N-1} \|Bw\|^2 = \frac{1}{N-1} w^T B^T B w = w^T S w.$$

Writing $w = \sum_i c_i q_i$ with $\sum c_i^2 = 1$,

$$w^T S w = \sum_i \lambda_i c_i^2 \leq \lambda_1 \sum_i c_i^2 = \lambda_1,$$

with equality exactly when $c_1 = \pm 1$, i.e. $w = \pm q_1$. Restricting to $w \perp q_1$ forces $c_1 = 0$ and repeats the argument with λ_2 , and so on. ■

Theorem 14.2 PCA is the SVD

Let $B = U\Sigma V^T$ be the SVD of the centred data matrix. Then

$$S = \frac{1}{N-1} B^T B = \frac{1}{N-1} V \Sigma^T \Sigma V^T,$$

so

- the principal components are the **right singular vectors** v_i ;
- the variances are $\lambda_i = \frac{\sigma_i^2}{N-1}$;
- the **scores** (coordinates of the data in the new basis) are the columns of $BV = U\Sigma$.

Remark • which route to take

Mathematically the two routes are identical. Numerically the SVD of B is strictly better, because forming $B^T B$ squares the condition number — a singular value of 10^{-8} becomes an eigenvalue of 10^{-16} , below double precision. Every library computes PCA by SVD.

14.3 The PCA procedure

Method PCA, step by step

1. **Arrange** the data as $X \in \mathbb{R}^{N \times p}$: one row per observation, one column per variable.
2. **Centre**: subtract each column's mean to get B . (Standardise as well if the units differ.)
3. **Covariance**: $S = \frac{1}{N-1} B^T B$ — or go straight to the SVD of B .
4. **Decompose**: find eigenvalues λ_i and orthonormal eigenvectors q_i of S , sorted with $\lambda_1 \geq \lambda_2 \geq \dots$
5. **Choose** k : keep enough components to explain the desired fraction of variance (often 90–95%), or use the “elbow” of the scree plot.
6. **Project**: the reduced data is $Z = BQ_k \in \mathbb{R}^{N \times k}$, where Q_k holds the first k eigenvectors.
7. **Reconstruct** (if needed): $\hat{B} = ZQ_k^T$, and add the means back.

Worked Example 14.1 • A complete PCA by hand

Five students are measured on two (standardised) variables:

$$X = \begin{bmatrix} 2 & 1 \\ 3 & 3 \\ 4 & 4 \\ 5 & 6 \\ 6 & 6 \end{bmatrix}.$$

Carry out PCA.

Step 1 — means. $\bar{x}_1 = \frac{2+3+4+5+6}{5} = 4$, $\bar{x}_2 = \frac{1+3+4+6+6}{5} = 4$.

Step 2 — centre.

$$B = \begin{bmatrix} -2 & -3 \\ -1 & -1 \\ 0 & 0 \\ 1 & 2 \\ 2 & 2 \end{bmatrix}.$$

Step 3 — covariance.

$$B^T B = \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -3 & -1 & 0 & 2 & 2 \end{bmatrix} \text{ times } B = \begin{bmatrix} 4+1+0+1+4 & 6+1+0+2+4 \\ 6+1+0+2+4 & 9+1+0+4+4 \end{bmatrix} = \begin{bmatrix} 10 & 13 \\ 13 & 18 \end{bmatrix}.$$

With $N - 1 = 4$,

$$S = \frac{1}{4} \begin{bmatrix} 10 & 13 \\ 13 & 18 \end{bmatrix} = \begin{bmatrix} 2.5 & 3.25 \\ 3.25 & 4.5 \end{bmatrix}.$$

Step 4 — eigenvalues. $\text{tr } S = 7$ and $\det S = 2.5(4.5) - 3.25^2 = 11.25 - 10.5625 = 0.6875$. So

$$\lambda^2 - 7\lambda + 0.6875 = 0, \quad \lambda = \frac{7 \pm \sqrt{49 - 2.75}}{2} = \frac{7 \pm \sqrt{46.25}}{2}.$$

$\sqrt{46.25} \approx 6.8007$, giving

$$\lambda_1 \approx 6.9004, \quad \lambda_2 \approx 0.0996.$$

Check: $\lambda_1 + \lambda_2 \approx 7.000 = \text{tr } S$ ✓ and $\lambda_1 \lambda_2 \approx 0.6873 \approx \det S$ ✓.

Eigenvectors. For λ_1 : $(S - \lambda_1 I)w = \mathbf{0}$ gives $(2.5 - 6.9004)w_1 + 3.25w_2 = 0$, so $w_2 = \frac{4.4004}{3.25}w_1 \approx 1.3540w_1$. Taking $w_1 = 1$ and normalising by $\sqrt{1 + 1.3540^2} = \sqrt{2.8333} \approx 1.6833$:

$$q_1 \approx \begin{bmatrix} 0.5941 \\ 0.8044 \end{bmatrix}, \quad q_2 \approx \begin{bmatrix} -0.8044 \\ 0.5941 \end{bmatrix}$$

(q_2 perpendicular to q_1 , as the spectral theorem guarantees).

Step 5 — variance explained.

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{6.9004}{7.0000} \approx 0.9858 = 98.6\%.$$

The first component alone captures almost 99% of the variance. Reducing from two variables to one loses only 1.4% of the information.

Step 6 — scores. Projecting each centred row onto q_1 :

$$\begin{aligned}
 (-2, -3) &\mapsto -1.188 - 2.413 = -3.601 \\
 (-1, -1) &\mapsto -0.594 - 0.804 = -1.398 \\
 z_i = b_i \cdot q_1 : \quad (0, 0) &\mapsto 0 \\
 (1, 2) &\mapsto 0.594 + 1.609 = 2.203 \\
 (2, 2) &\mapsto 1.188 + 1.609 = 2.797
 \end{aligned}$$

Each student is now described by a single number instead of two, and the ordering along the “size” axis is preserved.

14.4 Choosing how many components

Criterion	Rule
Cumulative variance	keep the smallest k with $\sum_{i \leq k} \lambda_i / \sum_i \lambda_i \geq 0.90$ (or 0.95)
Scree plot elbow	plot λ_i against i and cut where the curve flattens
Kaiser criterion	on standardised data, keep $\lambda_i > 1$ (a component must explain more than one original variable’s worth)
Downstream task	choose k by cross-validation on the model that consumes the components

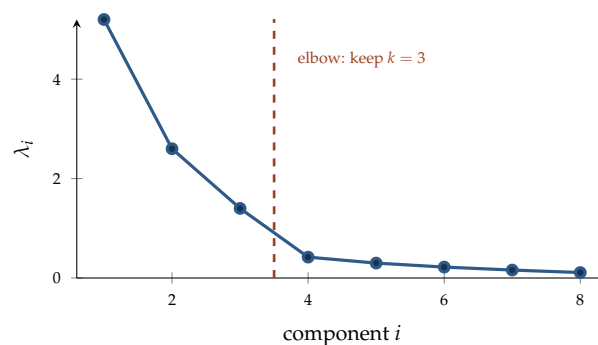


Figure 14.2: A scree plot. The eigenvalues drop sharply and then level off; components past the elbow mostly describe noise.

14.5 What PCA is used for

Lecture 6 lists the applications; here is what each one is doing mathematically.

Application	What PCA contributes
Dimensionality reduction	replace p correlated variables by $k \ll p$ uncorrelated ones, speeding up every downstream algorithm
Data visualisation	project onto the first 2 or 3 components to plot high-dimensional data and spot clusters
Noise reduction	discard the small- λ directions, which are typically dominated by measurement noise
Feature extraction	the components are new features, each a linear combination of the originals, ranked by informativeness
Image compression	eigenfaces: represent each face image by its coefficients on a small set of principal face-images
Decorrelation	the transformed variables have diagonal covariance — zero correlation by construction

Worked Example 14.2 • Eigenfaces — the idea

Take N face photographs, each 100×100 pixels. Flatten each into a vector of $p = 10,000$ numbers and stack them into $X \in \mathbb{R}^{N \times 10000}$.

1. Centre: subtract the “average face” from each image.
2. Take the SVD of the centred B .
3. The right singular vectors $v_1, v_2, \dots$ are themselves 10,000-vectors — reshape each back into a 100×100 image. These are the **eigenfaces**: ghostly composite patterns capturing the main ways faces differ (lighting, face width, presence of glasses, and so on).
4. Every face in the dataset is then

$$\text{face} \approx \text{average face} + c_1 v_1 + \dots + c_k v_k,$$

so it is stored as $k \approx 50$ coefficients instead of 10,000 pixels — a $200\times$ compression.

5. Recognition compares the coefficient vectors, which is fast and surprisingly robust.

This was the first practical face-recognition method (Turk and Pentland, 1991) and it is nothing more than the SVD of a data matrix.

Common Pitfall • what PCA cannot do

- PCA is **linear**. Data lying on a curved surface (a spiral, a sphere) is not well summarised by straight axes.
- PCA is **unsupervised**. The direction of greatest variance is not necessarily the direction that best separates your classes — for that, use LDA.
- PCA is **scale sensitive**. Measuring height in millimetres instead of metres changes the answer, unless you standardise.
- Components are **hard to interpret**. Each is a mixture of all original variables, so “component 1” rarely has a clean meaning.

14.6 Exam preparation

Exam Preparation • Chapter 14

The pipeline

$$X \xrightarrow{\text{centre}} B \xrightarrow{S = \frac{1}{N-1} B^T B} S = Q \Lambda Q^T \xrightarrow{\text{keep } k} Z = B Q_k$$

Equivalently and better: $B = U \Sigma V^T$, components = v_i , $\lambda_i = \sigma_i^2 / (N - 1)$, scores = $U \Sigma$.

Facts

- Principal components are orthonormal eigenvectors of the covariance matrix, ordered by decreasing eigenvalue.
- λ_i is the variance along component i ; $\text{tr } S = \sum \lambda_i$ is the total variance.
- Variance explained by k components = $\sum_{i \leq k} \lambda_i / \text{tr } S$.
- PCA of centred data = SVD of centred data.
- The transformed variables are uncorrelated.

Pitfalls

- Forgetting to centre.
- Forgetting to sort eigenvalues in decreasing order.
- Confusing λ_i (variance) with σ_i (singular value): $\lambda_i = \sigma_i^2 / (N - 1)$.
- Using PCA on non-linear structure and expecting it to work.
- Mixing up the two data layouts. Here rows are observations; some texts use columns, which transposes everything.

Self-check

1. Data has covariance eigenvalues 8, 3, 2, 1. How many components are needed to explain $\geq 90\%$ of the variance?
2. Why must the covariance matrix always be symmetric positive semidefinite?
3. If all p variables are uncorrelated with equal variance, what does PCA return?
4. A dataset has $N = 50$ samples and $p = 1000$ variables. What is the maximum possible number of nonzero eigenvalues of S ?

Answers

1. Total = 14. One component: $8/14 = 57\%$. Two: $11/14 = 79\%$. Three: $13/14 = 93\% \geq 90\%$. So $k = 3$.
2. $S = \frac{1}{N-1} B^T B$, and $A^T A$ is always symmetric (Chapter 6) with $x^T A^T A x = \|Ax\|^2 \geq 0$.
3. $S = cI$, so every eigenvalue equals c and *every* orthonormal basis is a valid set of components. PCA gives no preferred direction — correctly, since the data cloud is spherical and there is nothing to find.
4. $\text{rank } S = \text{rank } B \leq \min(N - 1, p) = 49$ (the -1 because centring removes one degree of freedom). So at most 49 nonzero eigenvalues, and at least 951 zeros. This “ $p \gg N$ ” regime is exactly where PCA is most valuable — and why one computes the SVD of B rather than the 1000×1000 matrix S .

PART VI

Linear Transformations and Applications

Linear Transformations, Translations and Affine Maps

A matrix is a linear transformation in disguise. This chapter makes the connection explicit, catalogues the geometric transformations of the lecture notes — rotation, scaling, reflection, shear — in both 2D and 3D, and then confronts the awkward one: *translation*. Translation is not linear, because it moves the origin. The fix, homogeneous coordinates, embeds $\mathbb{R}^n$ in $\mathbb{R}^{n+1}$ so that translations become matrix multiplications too, and every affine transformation — linear map plus shift — becomes a single matrix. This is exactly how computer graphics pipelines work.

Strang §8.1 (“The Idea of a Linear Transformation”), §8.2 (“The Matrix of a Linear Transformation”), §8.3 (“The Search for a Good Basis”). The supplementary *Linear Transformations* lecture notes cover geometric transformations in 2D and 3D. STRANG | 6th ed.

15.1 Linear transformations

Definition 15.1 Linear transformation

A map $T : V \rightarrow W$ between vector spaces is a **linear transformation** if for all $u, v \in V$ and all scalars c :

$$(L1) \quad T(u + v) = T(u) + T(v), \quad (L2) \quad T(cu) = cT(u).$$

The two conditions combine into the single requirement

$$T(cu + dv) = cT(u) + dT(v).$$

Proposition 15.1 Immediate consequences

Every linear transformation satisfies

$$T(\mathbf{0}) = \mathbf{0}, \quad T(-v) = -T(v), \quad T\left(\sum_i c_i v_i\right) = \sum_i c_i T(v_i).$$

Proof. Put $c = 0$ in (L2): $T(\mathbf{0}) = T(0v) = 0T(v) = \mathbf{0}$. The others follow by induction on the number of terms. ■

Common Pitfall • $T(\mathbf{0}) = \mathbf{0}$ is the quickest test

If a candidate map does not send $\mathbf{0}$ to $\mathbf{0}$, it is *not* linear — no further work needed. This

immediately disqualifies translation, which is the subject of Section 15.4.

Theorem 15.1 Every linear map on $\mathbb{R}^n$ is a matrix

$T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear **if and only if** there is a unique matrix $A \in \mathbb{R}^{m \times n}$ with

$$T(x) = Ax \quad \text{for all } x \in \mathbb{R}^n.$$

The matrix is built from the images of the standard basis vectors:

$$A = [T(e_1) \ T(e_2) \ \cdots \ T(e_n)]$$

Proof. Multiplication by a matrix is linear by the distributive rules of Chapter 3. Conversely, given linear T , write $x = x_1 e_1 + \cdots + x_n e_n$; then

$$T(x) = x_1 T(e_1) + \cdots + x_n T(e_n) = [T(e_1) \ \cdots \ T(e_n)] x.$$

Uniqueness: applying both candidate matrices to e_j shows their j th columns agree. ■

Intuition • the whole map from n vectors

A linear transformation of $\mathbb{R}^n$ is completely determined by what it does to the n basis vectors. Everything else follows by linearity. That is why matrices — which store exactly those n images as columns — are adequate to describe every linear map.

Worked Example 15.1 • Linear or not?

Decide which of these maps $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ are linear.

- (a) $T(x, y) = (2x - y, x + 3y)$
- (b) $T(x, y) = (x + 1, y)$
- (c) $T(x, y) = (x^2, y)$
- (d) $T(x, y) = (y, x)$
- (e) $T(x, y) = (\|(x, y)\|, 0)$

(a) Linear. $T(0) = (0, 0) \checkmark$, and it is visibly of the form Ax with $A = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}$.

(b) Not linear. $T(0, 0) = (1, 0) \neq 0$. This is a *translation* by $(1, 0)$ — affine, not linear.

(c) Not linear. $T(2, 0) = (4, 0)$ but $2T(1, 0) = 2(1, 0) = (2, 0)$. Fails (L2). A squared term can never be linear.

(d) Linear. It swaps coordinates: $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, the reflection in the line $y = x$.

(e) Not linear. $T(-1, 0) = (1, 0)$ but $-T(1, 0) = (-1, 0)$. Norms are not linear (they satisfy the triangle *inequality*, not equality).

Definition 15.2 Kernel and range

For linear $T : V \rightarrow W$,

$$\ker T = \{v \in V : T(v) = 0\}, \quad \text{range } T = \{T(v) : v \in V\}.$$

When $T(x) = Ax$ these are exactly $\mathbf{N}(A)$ and $\mathbf{C}(A)$, and the rank–nullity theorem reads

$$\dim \ker T + \dim \text{range } T = \dim V.$$

15.2 Geometric transformations in 2D

The lecture notes classify the key types. All of the following (except translation) are linear, because they fix the origin.

Type	Matrix	det	Effect
Rotation by θ	$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$	+1	turns, preserves shape
Scaling	$\begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}$	$s_x s_y$	stretches along axes
Reflection in x -axis	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	-1	flips vertically
Reflection in y -axis	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	-1	flips horizontally
Shear along x	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	+1	slants, preserves area
Shear along y	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$	+1	slants, preserves area
Projection onto x	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	0	flattens (not invertible)

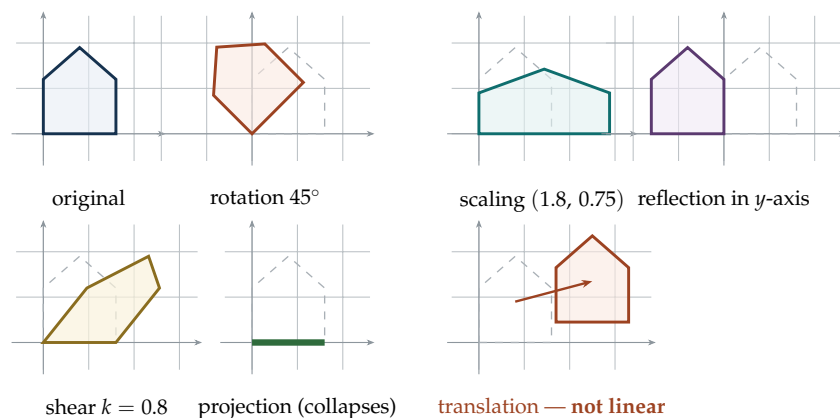


Figure 15.1: The standard geometric transformations. All but the last fix the origin and are therefore linear; translation moves the origin and is not.

15.2.1 Rotation

Definition 15.3 Rotation matrix

Counter-clockwise rotation by θ about the origin:

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

The lecture notes instead write

$$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} = R_{-\theta} = R_{\theta}^T,$$

which is the same map with θ measured *clockwise*. Either convention is fine provided you state which one you are using.

Proposition 15.2

R_{θ} is orthogonal: $R_{\theta}^T R_{\theta} = I$, $\det R_{\theta} = 1$, and

$$R_{\theta}^{-1} = R_{-\theta} = R_{\theta}^T, \quad R_{\alpha} R_{\beta} = R_{\alpha+\beta}.$$

Proof. $R_{\theta}^T R_{\theta} = \begin{bmatrix} \cos^2 \theta + \sin^2 \theta & 0 \\ 0 & \sin^2 \theta + \cos^2 \theta \end{bmatrix} = I$. For composition, multiply out and use the angle-addition formulas: the $(1,1)$ entry of $R_{\alpha} R_{\beta}$ is $\cos \alpha \cos \beta - \sin \alpha \sin \beta = \cos(\alpha + \beta)$, and similarly for the rest. ■

Worked Example 15.2 • Rotating a point

Rotate $(3, 1)$ by 90° counter-clockwise, then by a further 30° .

$$R_{90^\circ} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \text{ so}$$

$$R_{90^\circ} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$$

The composition $R_{30^\circ} R_{90^\circ} = R_{120^\circ}$, and

$$R_{120^\circ} = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}, \quad R_{120^\circ} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} - \frac{\sqrt{3}}{2} \\ \frac{3\sqrt{3}}{2} - \frac{1}{2} \end{bmatrix} \approx \begin{bmatrix} -2.366 \\ 2.098 \end{bmatrix}.$$

Length check: $\|(3, 1)\| = \sqrt{10} \approx 3.162$ and $\sqrt{2.366^2 + 2.098^2} = \sqrt{5.598 + 4.402} = \sqrt{10} \checkmark$ — rotations preserve length, as every orthogonal matrix must.

15.2.2 Reflection in an arbitrary line

Proposition 15.3

Reflection in the line through the origin at angle θ to the x -axis is

$$F_{\theta} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}, \quad \det F_{\theta} = -1, \quad F_{\theta}^2 = I.$$

Special cases: $\theta = 0$ gives reflection in the x -axis $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$; $\theta = \frac{\pi}{2}$ gives $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$; $\theta = \frac{\pi}{4}$

gives $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, the swap.

Remark • the sign of the determinant is the orientation

$\det = +1$ means orientation is preserved (a rotation); $\det = -1$ means it is reversed (a reflection — a left hand becomes a right hand). Since $\det(AB) = \det A \det B$, composing two reflections gives $\det = +1$: two reflections always make a rotation. This is why mirrors come in pairs in a kaleidoscope.

15.2.3 Shear

Definition 15.4 Shear

A **shear** slants a shape by shifting coordinates along one axis in proportion to the other:

$$S_x(k) = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \quad (\text{shift } x \text{ by } k \cdot y), \quad S_y(k) = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \quad (\text{shift } y \text{ by } k \cdot x).$$

Worked Example 15.3 • A shear is defective

$S_x(k)$ has $\det = 1$, so it preserves area, and its only eigenvalue is $\lambda = 1$ (repeated).

$S_x(k) - I = \begin{bmatrix} 0 & k \\ 0 & 0 \end{bmatrix}$ has rank 1 (for $k \neq 0$), so $\text{GM}(1) = 1 < 2 = \text{AM}(1)$: the shear is **defective** and cannot be diagonalized (Chapter 12).

Geometrically this is exactly right — a shear leaves only the x -axis pointwise fixed and tilts every other direction, so there is genuinely only one invariant line. The algebra and the picture agree.

15.3 Geometric transformations in 3D

Transformation	Matrix
Rotation about x -axis	$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}$
Rotation about y -axis	$R_y(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}$
Rotation about z -axis	$R_z(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Scaling	$\text{diag}(s_x, s_y, s_z)$
Reflection in the xy -plane	$\text{diag}(1, 1, -1)$
Reflection in the yz -plane	$\text{diag}(-1, 1, 1)$
Reflection in the xz -plane	$\text{diag}(1, -1, 1)$
Reflection through the origin	$\text{diag}(-1, -1, -1) = -I$

Common Pitfall • 3D rotations do not commute

$R_x(\theta)R_y(\phi) \neq R_y(\phi)R_x(\theta)$. Rotate a book 90° about the vertical axis then 90° about the horizontal, and compare with the opposite order — the results differ. This is the non-commutativity of Chapter 3 made physical, and it is why 3D orientation is usually stored as a quaternion rather than as three separate angles.

Worked Example 15.4 • Composing 3D rotations

Rotate $(1, 0, 0)$ by 90° about the z -axis, then 90° about the x -axis. Compare with the reverse order.

$$R_z(90^\circ) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad R_x(90^\circ) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Order 1: z then x .

$$R_x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad R_z \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Order 2: x then z .

$$R_z \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad R_x \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The results $(0, 0, 1)$ and $(0, 1, 0)$ are different. Order matters.

15.4 Translation and why it breaks linearity

Definition 15.5 Translation

Translation by a fixed vector $t \in \mathbb{R}^n$ is the map

$$T(x) = x + t.$$

It moves every point by the same displacement, changing position but not orientation, size or shape.

Theorem 15.2 Translation is not linear

If $t \neq 0$, translation by t is **not** a linear transformation.

Proof. $T(0) = 0 + t = t \neq 0$, violating the requirement $T(0) = 0$. Concretely, additivity also fails:

$$T(u + v) = u + v + t \quad \text{but} \quad T(u) + T(v) = u + v + 2t,$$

which differ by t . ■

Remark • the lecture notes' phrasing

Translation is not a linear transformation by itself because it does not preserve the origin, but it can be represented using matrices in homogeneous coordinates. That is the plan for the rest of this chapter.

Intuition • why this matters in practice

Graphics, robotics and vision pipelines need to chain together dozens of operations: move the object here, rotate it, scale it, move the camera, project. If translations could not be written as matrices, every chain would be an awkward alternation of matrix multiplications and vector additions. Homogeneous coordinates make the whole chain a single matrix product — which hardware can execute in one instruction.

15.5 Homogeneous coordinates

Definition 15.6 Homogeneous coordinates

Embed $\mathbb{R}^2$ into $\mathbb{R}^3$ by appending a 1:

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

More generally $\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+1}$ via $x \mapsto \begin{bmatrix} x \\ 1 \end{bmatrix}$. A point (x, y, w) with $w \neq 0$ represents the ordinary point $(x/w, y/w)$; points with $w = 0$ represent *directions* (points at infinity).

Theorem 15.3 Translation as a matrix

In homogeneous coordinates, translation by (t_x, t_y) is

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}, \quad T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x + t_x \\ y + t_y \\ 1 \end{bmatrix}.$$

The map is now a genuine matrix multiplication — and it *is* linear as a map on $\mathbb{R}^3$; it is only the restriction to the plane $w = 1$ that looks like a shift.

Intuition • what the trick really does

T is a *shear* of $\mathbb{R}^3$ along the w direction. Slicing that shear at the height $w = 1$ produces a translation of the plane. So translation is not linear in $\mathbb{R}^2$, but it is the shadow of something linear in $\mathbb{R}^3$ — which is exactly what the extra coordinate buys.

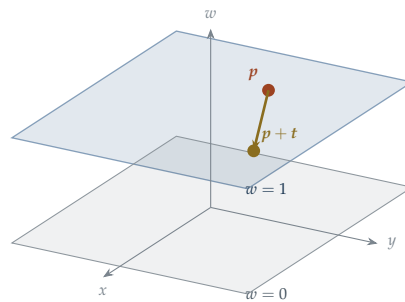


Figure 15.2: Homogeneous coordinates. The plane $\mathbb{R}^2$ is lifted to the slice $w = 1$ of $\mathbb{R}^3$; a linear shear of $\mathbb{R}^3$ restricts to a translation of that slice.

Definition 15.7 Affine transformation

An **affine transformation** of $\mathbb{R}^n$ is a linear map followed by a translation:

$$T(x) = Ax + t.$$

In homogeneous coordinates it is the single $(n+1) \times (n+1)$ matrix

$$M = \left[\begin{array}{c|c} A & t \\ \hline \mathbf{0}^T & 1 \end{array} \right], \quad M \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} Ax + t \\ 1 \end{bmatrix}.$$

Affine maps preserve straight lines, parallelism, and ratios of lengths along a line — but not necessarily angles, lengths or the origin.

2D transformation	Homogeneous matrix
Translation (t_x, t_y)	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$
Rotation by θ	$\begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Scaling (s_x, s_y)	$\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Shear along x	$\begin{bmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Reflection in x -axis	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

15.6 Composing transformations

Theorem 15.4 Composition is matrix multiplication

If S and T are linear (or affine, in homogeneous form) with matrices A and B , then the composition $S \circ T$ — “do T first, then S ” — has matrix

$$AB.$$

Common Pitfall • read composed matrices right to left

$M = M_3 M_2 M_1$ applies M_1 *first*. The rightmost factor acts first, because it is the one adjacent to the vector in $M_3 M_2 M_1 x$. Getting this backwards produces objects in wildly wrong places — the classic graphics bug.

Worked Example 15.5 • Rotation about an arbitrary point

Rotate the plane by θ about the point $c = (c_x, c_y)$, not the origin.

Rotation matrices only rotate about the origin, so we move the centre there, rotate, and move back. Three steps:

1. translate by $-c$: $T_{-c} = \begin{bmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & 1 \end{bmatrix}$

2. rotate by θ : $R = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

3. translate back by $+c$: $T_c = \begin{bmatrix} 1 & 0 & c_x \\ 0 & 1 & c_y \\ 0 & 0 & 1 \end{bmatrix}$

The composite (remembering right-to-left) is

$$M = T_c R T_{-c} = \begin{bmatrix} \cos \theta & -\sin \theta & c_x - c_x \cos \theta + c_y \sin \theta \\ \sin \theta & \cos \theta & c_y - c_x \sin \theta - c_y \cos \theta \\ 0 & 0 & 1 \end{bmatrix}.$$

Numerical check. Take $\theta = 90^\circ$ and $c = (1, 1)$. Then $\cos \theta = 0, \sin \theta = 1$, so

$$M = \begin{bmatrix} 0 & -1 & 1 - 0 + 1 \\ 1 & 0 & 1 - 1 - 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 2 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The centre must be fixed: $M(1, 1, 1)^T = (0 - 1 + 2, 1 + 0 + 0, 1) = (1, 1, 1) \checkmark$. And the origin should go to $(2, 0)$: $M(0, 0, 1)^T = (2, 0, 1) \checkmark$ — rotating the origin 90° about $(1, 1)$ does indeed land at $(2, 0)$.

Worked Example 15.6 • Scaling about a point, and why order matters

Scale by 2 about the point $(3, 4)$, and compare with “scale then translate”.

Correct (scale about $(3, 4)$):

$$M = T_{(3,4)} S_2 T_{(-3,-4)} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -3 \\ 0 & 2 & -4 \\ 0 & 0 & 1 \end{bmatrix}.$$

Check the fixed point: $M(3, 4, 1)^T = (6 - 3, 8 - 4, 1) = (3, 4, 1) \checkmark$.

Wrong order (scale first, then translate by $(3, 4)$):

$$M' = T_{(3,4)} S_2 = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 1 \end{bmatrix},$$

which sends $(3, 4)$ to $(9, 12)$ — the point is not fixed at all. Same ingredients, completely different result.

15.7 Change of basis

Definition 15.8 Matrix of T in a basis

Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be linear and let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis with $P = [\mathbf{b}_1 \cdots \mathbf{b}_n]$. If A is the matrix of T in the standard basis, its matrix in $\mathcal{B}$ is

$$[T]_{\mathcal{B}} = P^{-1}AP.$$

This is exactly the similarity relation of Chapter 11, now with its intended meaning: *similar matrices are the same transformation seen from different coordinate systems*. Diagonalization $A = X\Lambda X^{-1}$ is the search for a basis in which T is pure scaling; the SVD is the search for two bases (input and output) in which T is pure scaling even when no single one exists.

Worked Example 15.7 • A transformation looks simplest in its own basis

$A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$ acts confusingly in standard coordinates. In the eigenbasis $x_1 = (1, 1)$, $x_2 = (2, 1)$ (Chapter 11),

$$[T]_{\mathcal{B}} = X^{-1}AX = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

So T is really just “stretch by 2 along $(1, 1)$ and by 3 along $(2, 1)$ ”. The complicated-looking entries 4, -2 , 1, 1 are an artifact of insisting on the standard basis, which happens to be badly aligned with what this transformation actually does.

15.8 Exam preparation

Exam Preparation • Chapter 15

Definitions

$$T(cu + dv) = cT(u) + dT(v), \quad A = [T(e_1) \cdots T(e_n)]$$

Linear $\Rightarrow T(0) = 0$. Affine: $T(x) = Ax + t$.

2D matrices to memorise

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

Homogeneous form

$$\left[\begin{array}{c|c} A & t \\ \hline \mathbf{0}^T & 1 \end{array} \right] \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} Ax + t \\ 1 \end{bmatrix}$$

About-a-point: $M = T_c(\text{transform})T_{-c}$.

Determinant tells you what kind

$\det = +1$	rotation / shear — area and orientation preserved
$\det = -1$	reflection — orientation reversed
$ \det = k$	areas scaled by k
$\det = 0$	collapses dimension; not invertible

Pitfalls

- Calling translation linear. It is affine.
- Composing in the wrong order — the rightmost matrix acts first.
- Forgetting the translate–transform–translate-back sandwich for operations about a point other than the origin.
- Assuming 3D rotations commute.
- Forgetting the trailing 1 when converting a point to homogeneous coordinates (and that *directions* get a trailing 0 instead, so translations correctly leave them alone).

Self-check

1. Write the matrix that reflects in the line $y = x$ then rotates by 90° . Is the result a rotation or a reflection?
2. Give the homogeneous matrix that translates by $(2, -1)$, then scales by 3 about the origin.
3. Is $T(x, y) = (x + y, xy)$ linear? Justify.
4. Find the matrix of the shear that maps $(1, 0) \mapsto (1, 0)$ and $(0, 1) \mapsto (2, 1)$. What are its eigenvalues?

Answers

1. Reflection in $y = x$ is $F = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$; $R_{90} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Reflect first, so $M = R_{90}F = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. $\det M = -1$, so it is a *reflection* — in the y -axis.
2. Scale acts second, so $M = S T = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 6 \\ 0 & 3 & -3 \\ 0 & 0 & 1 \end{bmatrix}$.
3. No. $T(0) = (0, 0)$ passes, but $T(2, 2) = (4, 4)$ while $2T(1, 1) = 2(2, 1) = (4, 2)$. The product xy breaks (L2).
4. The columns are the images of e_1, e_2 : $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. Triangular, so $\lambda = 1$ twice; $\text{GM} = 1$, so it is defective — as every nontrivial shear is.

Applications of Linear Algebra

Lecture 1 justifies the course by pointing at the places linear algebra turns up: graphics, computer vision, machine learning, data representation, cryptography. This chapter follows through on those promises. Each section is short and self-contained, and each one is nothing more than a re-reading of a technique already covered — which is the point. The subject is worth learning once because it is then usable everywhere.

Strang Chapter 9 (“Linear Algebra in Optimization”) and Chapter 10 (“Learning from Data”), plus §9.2 on graphs and networks. Lecture 1 slides 9–18 motivate all of these.

16.1 Computer graphics

Every frame a game renders is a chain of matrix multiplications. A vertex p in an object’s own coordinates travels through:

$$p_{\text{screen}} = \underbrace{M_{\text{proj}}}_{\text{camera lens}} \underbrace{M_{\text{view}}}_{\text{camera pose}} \underbrace{M_{\text{model}}}_{\text{object pose}} p_{\text{object}}$$

all in homogeneous coordinates (Chapter 15), so that translations join rotations and scalings inside a single 4×4 matrix.

Worked Example 16.1 • Building a model matrix

Place a cube: scale it to half size, rotate it 45° about the z-axis, then move it to $(5, 0, 2)$.

In 4×4 homogeneous form, with $c = s = \frac{\sqrt{2}}{2}$:

$$M_{\text{model}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{translate}} \underbrace{\begin{bmatrix} c & -s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{rotate}} \underbrace{\begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{scale}}.$$

Remember the order: the rightmost factor acts first, so the cube is scaled before it is rotated and moved. Multiplying out,

$$M_{\text{model}} = \begin{bmatrix} 0.5c & -0.5s & 0 & 5 \\ 0.5s & 0.5c & 0 & 0 \\ 0 & 0 & 0.5 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \approx \begin{bmatrix} 0.354 & -0.354 & 0 & 5 \\ 0.354 & 0.354 & 0 & 0 \\ 0 & 0 & 0.5 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

A corner at $(1, 1, 1)$ maps to

$$(0.354 - 0.354 + 0 + 5, 0.354 + 0.354 + 0, 0.5 + 2, 1) = (5, 0.707, 2.5, 1).$$

The GPU does this for millions of vertices per frame — which is exactly why graphics hardware is built around 4×4 matrix multiplication.

Remark • why the projection matrix needs w

Perspective projection — distant objects appearing smaller — is achieved by writing a -1 into the bottom row so that the output has $w = -z$. Dividing through by w (the “perspective divide”) then scales x and y by $1/\text{depth}$. This is genuinely non-affine, and it is the one place where the w coordinate does real work rather than just enabling translation.

16.2 Image compression

Lecture 6 puts it directly: *digital images contain many pixels; eigenvectors identify the most important patterns; keeping only the largest eigenvalues allows us to reconstruct images using much less storage.*

Worked Example 16.2 • Compressing a greyscale image with the SVD

A 512×512 greyscale photograph is a matrix $A \in \mathbb{R}^{512 \times 512}$ of pixel intensities. Compress it.

Take the SVD $A = U\Sigma V^T$ and keep the largest k terms:

$$A_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

By Eckart–Young (Chapter 13) this is the best rank- k approximation available.

Storage.

	Numbers stored	vs. original	Quality
Full image	$512 \times 512 = 262,144$	100%	exact
$k = 5$	$5(512 + 512 + 1) = 5,125$	2.0%	blocky outline
$k = 20$	20,500	7.8%	recognisable
$k = 50$	51,250	19.6%	near-indistinguishable
$k = 100$	102,500	39.1%	visually identical

The reason this works is that photographs are far from random: neighbouring pixels are strongly correlated, so the singular values decay quickly and almost all the “energy” $\|A\|_F^2 = \sum_i \sigma_i^2$ sits in the first few terms.

Remark • SVD is not actually how JPEG works

JPEG uses the discrete cosine transform on 8×8 blocks — a *fixed* orthogonal basis, so it need not be transmitted with the image. The SVD adapts its basis to each image, which compresses better in principle but costs more to compute and requires storing U and V . The underlying idea is identical: change to a basis where most coefficients are negligible, then discard them.

16.3 Machine learning and data science

Technique	The linear algebra inside
Linear regression	least squares: $A^\top A \hat{x} = A^\top b$ (Chapter 10)
Ridge regression	$(A^\top A + \alpha I) \hat{x} = A^\top b$ — adding αI shifts every eigenvalue up by α , guaranteeing invertibility
PCA	eigen-decomposition of the covariance matrix (Chapter 14)
Neural network layer	$a^{(\ell+1)} = \sigma(Wa^{(\ell)} + b)$ — an affine map (Chapter 15) followed by a nonlinearity
Word embeddings	low-rank factorization of a word co-occurrence matrix
Recommender systems	matrix completion: fill in a sparse ratings matrix by assuming it is approximately low rank
Gradient descent	convergence rate governed by the eigenvalues of the Hessian; the condition number σ_1/σ_n decides how slow it is

Worked Example 16.3 • Why a neural network needs its nonlinearity

Show that stacking two linear layers gains nothing.

Two layers without activation functions give

$$y = W_2(W_1x + b_1) + b_2 = (W_2W_1)x + (W_2b_1 + b_2) = Wx + b,$$

which is a *single* affine map with $W = W_2W_1$ and $b = W_2b_1 + b_2$. A hundred stacked linear layers collapse the same way — the composition of linear maps is linear (Chapter 15). So depth buys nothing at all without the nonlinearity σ . It is precisely the failure of σ to be linear that makes deep networks more expressive than shallow ones. Notice also $\text{rank}(W_2W_1) \leq \min(\text{rank } W_2, \text{rank } W_1)$: a narrow layer imposes a hard rank ceiling on everything downstream, which is the mathematics behind “bottleneck” architectures and autoencoders.

Worked Example 16.4 • Ridge regression through eigenvalues

Why does adding αI fix an ill-conditioned least-squares problem?

Let $A^\top A = Q\Lambda Q^\top$ with eigenvalues $\lambda_i = \sigma_i^2 \geq 0$. Then

$$A^\top A + \alpha I = Q(\Lambda + \alpha I)Q^\top,$$

whose eigenvalues are $\lambda_i + \alpha$. Every one is now strictly positive, so the matrix is invertible even when A has dependent columns and $\lambda_i = 0$.

The effect on the solution is a shrinkage. Writing $\hat{x} = \sum_i \frac{\sigma_i}{\sigma_i^2 + \alpha} (u_i \cdot b) v_i$ in SVD terms:

- directions with $\sigma_i \gg \sqrt{\alpha}$ are barely touched — the well-determined part of the fit survives;
 - directions with $\sigma_i \ll \sqrt{\alpha}$ are strongly damped — the noise-dominated part is suppressed.
- Regularisation is thus a filter on singular values, and α sets the cutoff.

16.4 Networks and graphs

Definition 16.1 Adjacency and incidence matrices

For a graph with n nodes and m edges:

- the **adjacency matrix** $A \in \mathbb{R}^{n \times n}$ has $a_{ij} = 1$ when nodes i and j are joined, else 0;
- the **incidence matrix** $E \in \mathbb{R}^{m \times n}$ has, in the row for the edge from i to j , a -1 in column i and a $+1$ in column j ;
- the **Laplacian** is $L = E^T E = D - A$, where D is the diagonal matrix of node degrees.

Proposition 16.1 Counting walks

$(A^k)_{ij}$ is the number of walks of length exactly k from node i to node j .

Proof. Induction. For $k = 1$ it is the definition. A walk of length $k + 1$ from i to j is a walk of length k from i to some ℓ followed by an edge $\ell \rightarrow j$; summing over ℓ gives $(A^{k+1})_{ij} = \sum_{\ell} (A^k)_{i\ell} a_{\ell j}$, which is matrix multiplication. ■

Worked Example 16.5 • A small network

Four routers are connected: 1–2, 2–3, 3–4, 1–3. How many length-2 paths run from router 1 to router 4?

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

$$A^2 = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}.$$

Entry $(1, 4)$ is 1: exactly one two-hop path, namely $1 \rightarrow 3 \rightarrow 4$. ✓

Bonus reading. The diagonal of A^2 gives the degrees: router 3 has degree 3, the others 2, 2, 1 — consistent with the edge list. And $\text{tr } A^3 / 6$ counts triangles; here A contains the single triangle 1–2–3.

Proposition 16.2 Laplacian and connectivity

The graph Laplacian $L = D - A$ is symmetric positive semidefinite. The multiplicity of the eigenvalue 0 equals the number of connected components. The second-smallest eigenvalue — the **algebraic connectivity** or Fiedler value — measures how well connected the graph is, and the sign pattern of its eigenvector gives a good two-way partition (spectral clustering).

Remark • PageRank

Google's original ranking algorithm builds a column-stochastic matrix M from the web's

link structure and finds the eigenvector for $\lambda = 1$:

$$Mr = r.$$

The entries of r are the page ranks. Existence of that eigenvector follows from the Perron–Frobenius theorem, and computing it is done by repeated multiplication $r_{k+1} = Mr_k$ — the power method, which converges because every other eigenvalue has $|\lambda| < 1$ (exactly the Markov chain argument of Chapter 12).

16.5 Cryptography

Worked Example 16.6 • The Hill cipher

Encrypt the message `HELP` using the key $K = \begin{bmatrix} 3 & 3 \\ 2 & 5 \end{bmatrix}$ over $\mathbb{Z}_{26}$.

Encode letters. $A = 0, B = 1, \dots, Z = 25$, so `HE` = (7, 4) and `LP` = (11, 15).

Encrypt by $c = Kp \bmod 26$:

$$K \begin{bmatrix} 7 \\ 4 \end{bmatrix} = \begin{bmatrix} 21 + 12 \\ 14 + 20 \end{bmatrix} = \begin{bmatrix} 33 \\ 34 \end{bmatrix} \equiv \begin{bmatrix} 7 \\ 8 \end{bmatrix} \pmod{26} \Rightarrow \text{HI}$$

$$K \begin{bmatrix} 11 \\ 15 \end{bmatrix} = \begin{bmatrix} 33 + 45 \\ 22 + 75 \end{bmatrix} = \begin{bmatrix} 78 \\ 97 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 19 \end{bmatrix} \pmod{26} \Rightarrow \text{AT}$$

Ciphertext: `HIAT`.

Decrypt with K^{-1} modulo 26. First $\det K = 15 - 6 = 9$, and 9 is invertible mod 26 because $\gcd(9, 26) = 1$; indeed $9 \cdot 3 = 27 \equiv 1$, so $9^{-1} \equiv 3$. Then

$$K^{-1} \equiv 3 \begin{bmatrix} 5 & -3 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 15 & -9 \\ -6 & 9 \end{bmatrix} \equiv \begin{bmatrix} 15 & 17 \\ 20 & 9 \end{bmatrix} \pmod{26}.$$

Check: $K^{-1}(7, 8)^T = (105 + 136, 140 + 72) = (241, 212)$, and $241 = 9(26) + 7$, $212 = 8(26) + 4$, giving `(7, 4)` = `HE` ✓.

The key condition. K must be invertible *modulo* 26, which requires $\gcd(\det K, 26) = 1$ — a stricter demand than $\det K \neq 0$ over $\mathbb{R}$. (The Hill cipher is not secure by modern standards — it falls to a known-plaintext attack, since enough plaintext-ciphertext pairs let you solve for K by ordinary elimination. It is a nice illustration, not a recommendation.)

16.6 Differential equations

Theorem 16.1 Solving $\frac{du}{dt} = Au$

If $A = X\Lambda X^{-1}$ is diagonalizable with eigenvalues λ_i and eigenvectors x_i , the solution with $u(0) = u_0$ is

$$u(t) = c_1 e^{\lambda_1 t} x_1 + \cdots + c_n e^{\lambda_n t} x_n,$$

where the c_i come from expanding $u_0 = \sum c_i x_i$. Compactly, $u(t) = e^{At} u_0$ with $e^{At} = X e^{\Lambda t} X^{-1}$.

Intuition • eigenvalues decide stability

Along the eigenvector x_i the system evolves as the scalar equation $\dot{u} = \lambda_i u$, whose solution is $e^{\lambda_i t}$. So the eigenvalues are exactly the growth or decay rates:

- all $\text{Re}(\lambda_i) < 0$: every mode decays, the system is **stable**;
- some $\text{Re}(\lambda_i) > 0$: that mode blows up, **unstable**;
- purely imaginary λ : oscillation without growth.

Diagonalization decouples n interacting equations into n independent scalar ones — that is the whole reason for doing it.

Worked Example 16.7 • A two-species model

Solve $\dot{u} = Au$ with $A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$ and $u(0) = (4, 3)$.

From Chapter 11: $\lambda_1 = 2$ with $x_1 = (1, 1)$, and $\lambda_2 = 3$ with $x_2 = (2, 1)$.

Expand the initial condition. Solve $c_1(1, 1) + c_2(2, 1) = (4, 3)$:

$$c_1 + 2c_2 = 4, \quad c_1 + c_2 = 3 \quad \implies \quad c_2 = 1, \quad c_1 = 2.$$

Solution.

$$u(t) = 2e^{2t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + e^{3t} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2e^{2t} + 2e^{3t} \\ 2e^{2t} + e^{3t} \end{bmatrix}.$$

Check at $t = 0$: $(2 + 2, 2 + 1) = (4, 3) \checkmark$. And $\dot{u}(0) = (4 + 6, 4 + 3) = (10, 7)$, while $Au(0) = (16 - 6, 4 + 3) = (10, 7) \checkmark$.

Both eigenvalues are positive, so the system grows without bound, and for large t the e^{3t} term dominates: the state aligns with the *dominant eigenvector* $(2, 1)$ regardless of where it started. That asymptotic alignment is the same phenomenon as the Markov steady state, with growth instead of decay.

16.7 Numerical computing with Octave

The Wednesday practicals (Tutorial 03) use GNU Octave. For reference, here is the whole course in commands.

Command	Does
<code>A = [1 2; 3 4]</code>	build a matrix (; separates rows)
<code>A'</code>	transpose A^T
<code>A*B</code> vs <code>A.*B</code>	matrix product vs. entrywise product
<code>x = A\b</code>	solve $Ax = b$ — use this
<code>inv(A)</code>	inverse (avoid for solving)
<code>det(A), rank(A), trace(A)</code>	the obvious things
<code>rref([A b])</code>	reduced row echelon form
<code>[L,U,P] = lu(A)</code>	$PA = LU$
<code>[Q,R] = qr(A)</code>	QR factorization
<code>[V,D] = eig(A)</code>	eigenvectors in V , eigenvalues on $\text{diag}(D)$
<code>[U,S,V] = svd(A)</code>	$A = U\Sigma V^T$
<code>null(A), orth(A)</code>	orthonormal bases for $N(A)$ and $C(A)$
<code>norm(v), dot(u,v)</code>	length and dot product
<code>eye(n), zeros(m,n), ones(m,n)</code>	standard matrices

Common Pitfall • `A\b` versus `inv(A)*b`

Chapter 5 explained why: the backslash operator runs elimination with partial pivoting, costing about $\frac{1}{3}n^3$ and choosing a suitable algorithm automatically (it even returns a least-squares solution if A is rectangular). Forming `inv(A)` costs three times as much and loses accuracy. In Octave and MATLAB alike, `inv` appearing in code is nearly always a mistake.

16.8 Exam preparation

Exam Preparation • Chapter 16

Which tool for which problem

Problem	Tool
Solve $Ax = b$ exactly	elimination / LU
Solve when there is no exact solution	least squares, $A^T A \hat{x} = A^T b$
Repeated application A^k	diagonalization $X\Lambda^k X^{-1}$
Long-run behaviour	dominant eigenvalue and eigenvector
Reduce dimensions of data	PCA / SVD
Best low-rank approximation	truncated SVD (Eckart–Young)
Move, rotate or scale objects	homogeneous transformation matrices
Count paths in a network	powers of the adjacency matrix
Decouple linear ODEs	eigen-decomposition

Recurring themes worth stating in an essay answer

- **Change of basis.** Nearly every technique amounts to finding coordinates in which the problem is diagonal.
- **Low rank.** Real data is rarely full rank; exploiting that is what compression, recommendation and PCA all do.
- **Orthogonality.** Perpendicular bases make coordinates cheap and computations stable.
- **Eigenvalues govern dynamics.** Growth, decay, stability and convergence rates are all spectral questions.

Self-check

1. A 1000×800 image is approximated with $k = 40$ singular values. What fraction of the original storage is needed?
2. Why can a stack of 10 linear layers in a neural network be replaced by one?
3. In a Markov chain with eigenvalues $1, 0.9, 0.3$, which controls how fast the chain reaches equilibrium?
4. The Hill cipher key has $\det K = 13$. Can it be used over $\mathbb{Z}_{26}$?

Answers

1. Original = $1000 \times 800 = 800,000$. Compressed = $40(1000 + 800 + 1) = 72,040$, which is 9.0%.
2. The composition of affine maps is affine: $W_{10} \cdots W_1 x + \mathbf{b} = Wx + \mathbf{b}$. Depth without nonlinearity adds no expressive power.
3. The *second largest* magnitude, 0.9. The error after k steps decays like 0.9^k ; the eigenvalue 1 merely fixes what the limit is.
4. No. $\gcd(13, 26) = 13 \neq 1$, so 13 has no inverse modulo 26 and K^{-1} does not exist over $\mathbb{Z}_{26}$ — decryption would be impossible. The key determinant must be coprime to 26.

PART VII

Appendices

Worked Tutorial Solutions

This appendix works through a representative selection from Tutorials 01–06, with full arithmetic. Problems already solved inside the main chapters are cross-referenced rather than repeated. Where a tutorial question is drawn verbatim from Strang, the corresponding textbook section is noted.

A.1 Tutorial 01 — Matrix arithmetic

Worked Example A.1 • T01 Q2 — transposes and traces, complete

$$A = \begin{bmatrix} -5 & 2 & 0 \\ 7 & -3 & 4 \\ -1 & 3 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & -1 & 8 \\ 6 & -14 & 2 \\ 9 & 5 & 1 \end{bmatrix}.$$

(i) Transposes.

$$A^T = \begin{bmatrix} -5 & 7 & -1 \\ 2 & -3 & 3 \\ 0 & 4 & 2 \end{bmatrix}, \quad B^T = \begin{bmatrix} 5 & 6 & 9 \\ -1 & -14 & 5 \\ 8 & 2 & 1 \end{bmatrix}.$$

(ii)–(iii) $(A + B)^T = A^T + B^T$.

$$A + B = \begin{bmatrix} 0 & 1 & 8 \\ 13 & -17 & 6 \\ 8 & 8 & 3 \end{bmatrix}, \quad (A + B)^T = \begin{bmatrix} 0 & 13 & 8 \\ 1 & -17 & 8 \\ 8 & 6 & 3 \end{bmatrix} = A^T + B^T. \checkmark$$

(iv) $(AB)^T = B^T A^T$. First

$$AB = \begin{bmatrix} -13 & -23 & -36 \\ 53 & 55 & 54 \\ 31 & -31 & 0 \end{bmatrix}.$$

(Sample entries: $(2,1)$ is $7(5) + (-3)(6) + 4(9) = 35 - 18 + 36 = 53$; $(2,2)$ is $7(-1) + (-3)(-14) + 4(5) = -7 + 42 + 20 = 55$.) Then

$$(AB)^T = \begin{bmatrix} -13 & 53 & 31 \\ -23 & 55 & -31 \\ -36 & 54 & 0 \end{bmatrix}.$$

Computing $B^T A^T$ gives the same matrix. For instance its $(1,2)$ entry is $\text{row}_1(B^T) = (5, 6, 9)$ dotted with $\text{col}_2(A^T) = (7, -3, 4)$, giving $35 - 18 + 36 = 53 \checkmark$, matching $(AB)^T_{12}$.

$$\text{(v)} (-A)^T = -A^T. \quad (-A)^T = \begin{bmatrix} 5 & -7 & 1 \\ -2 & 3 & -3 \\ 0 & -4 & -2 \end{bmatrix}.$$

(vi) $(2A - 5BA)^T$. First

$$BA = \begin{bmatrix} -40 & 37 & 12 \\ -130 & 60 & -52 \\ -11 & 6 & 22 \end{bmatrix},$$

(sample entry: $(1,2)$ is $5(2) + (-1)(-3) + 8(3) = 10 + 3 + 24 = 37$). Then

$$2A - 5BA = \begin{bmatrix} -10 & 4 & 0 \\ 14 & -6 & 8 \\ -2 & 6 & 4 \end{bmatrix} - \begin{bmatrix} -200 & 185 & 60 \\ -650 & 300 & -260 \\ -55 & 30 & 110 \end{bmatrix} = \begin{bmatrix} 190 & -181 & -60 \\ 664 & -306 & 268 \\ 53 & -24 & -106 \end{bmatrix},$$

so

$$(2A - 5BA)^T = \begin{bmatrix} 190 & 664 & 53 \\ -181 & -306 & -24 \\ -60 & 268 & -106 \end{bmatrix}.$$

(vii) $\text{tr}(A + B)$. $0 - 17 + 3 = -14$. Alternatively $\text{tr } A + \text{tr } B = -6 + (-8) = -14 \checkmark$.

(viii) $\text{tr}((2A - 5BA)^T)$. Trace is transpose-invariant, so it equals

$$\text{tr}(2A - 5BA) = 190 - 306 - 106 = -222.$$

Check by linearity: $2 \text{tr } A - 5 \text{tr}(BA) = 2(-6) - 5(42) = -12 - 210 = -222 \checkmark$.

(ix)–(x) $\text{tr}(AB)$ and $\text{tr}((AB)^T)$.

$$\text{tr}(AB) = -13 + 55 + 0 = 42, \quad \text{tr}((AB)^T) = \text{tr}(AB) = 42.$$

Remark • the built-in check

Independently, $\text{tr}(BA) = -40 + 60 + 22 = 42$. The agreement $\text{tr}(AB) = \text{tr}(BA) = 42$ is guaranteed by the cyclic property (Chapter 3) even though $AB \neq BA$ — compare the two products above, which share not a single off-diagonal entry. Use this identity as a free check on every product you compute by hand.

Worked Example A.2 • T01 Q4(c) — a matrix square root

$$\text{Find } X \in \mathbb{R}^{2 \times 2} \text{ with } X^2 - 2X = \begin{bmatrix} -1 & 0 \\ 6 & 3 \end{bmatrix}.$$

Strategy. Completing the square, $X^2 - 2X + I = (X - I)^2$, so with $Y = X - I$,

$$Y^2 = \begin{bmatrix} -1 & 0 \\ 6 & 3 \end{bmatrix} + I = \begin{bmatrix} 0 & 0 \\ 6 & 4 \end{bmatrix}.$$

Look for lower triangular $Y = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix}$. Then $Y^2 = \begin{bmatrix} a^2 & 0 \\ c(a+d) & d^2 \end{bmatrix}$, so

$$a^2 = 0 \Rightarrow a = 0, \quad d^2 = 4 \Rightarrow d = \pm 2, \quad c(a+d) = 6.$$

Taking $d = 2$: $2c = 6$, so $c = 3$. Hence $Y = \begin{bmatrix} 0 & 0 \\ 3 & 2 \end{bmatrix}$ and

$$X = Y + I = \begin{bmatrix} 1 & 0 \\ 3 & 3 \end{bmatrix}.$$

Verify. $X^2 = \begin{bmatrix} 1 & 0 \\ 12 & 9 \end{bmatrix}$, so $X^2 - 2X = \begin{bmatrix} 1-2 & 0 \\ 12-6 & 9-6 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 6 & 3 \end{bmatrix}$. ✓

Taking $d = -2$ gives $c = -3$ and the second solution $X = \begin{bmatrix} 1 & 0 \\ -3 & -1 \end{bmatrix}$; matrix equations of this type generally have several solutions.

A.2 Tutorial 02 — Vectors and combinations (Strang §1.1–1.2)

Questions 1–10 of Tutorial 02 are worked in Chapter 1 (Q1, Q2, Q3, Q5, Q8, Q10, Q14, Q17, Q24, Q27). The remainder:

Worked Example A.3 • T02 Q11 — fourth corners of a parallelogram

Three corners of a parallelogram are $(1,1)$, $(4,2)$, $(1,3)$. Find all three possible fourth corners.

Label $a = (1,1)$, $b = (4,2)$, $c = (1,3)$. Each choice of which corner is “opposite” the missing one gives a different answer, and the rule is that the diagonals of a parallelogram bisect each other:

$$d = b + c - a = (4 + 1 - 1, 2 + 3 - 1) = (4, 4),$$

$$d = a + c - b = (1 + 1 - 4, 1 + 3 - 2) = (-2, 2),$$

$$d = a + b - c = (1 + 4 - 1, 1 + 2 - 3) = (4, 0).$$

So the three possibilities are $(4, 4)$, $(-2, 2)$ and $(4, 0)$.

Check the first. Sides $b - a = (3, 1)$ and $d - c = (3, 1)$ are equal, so those two sides are parallel and of equal length — a parallelogram. ✓

Worked Example A.4 • T02 Q25 — two different combinations

Find two different combinations of $u = (1,3)$, $v = (2,7)$, $w = (1,5)$ producing $b = (0,1)$.

We need $cu + dv + ew = (0,1)$, i.e.

$$c + 2d + e = 0, \quad 3c + 7d + 5e = 1.$$

Two equations, three unknowns — one free variable, hence infinitely many solutions.

Solution A (set $e = 0$): $c + 2d = 0$ and $3c + 7d = 1$. From the first $c = -2d$; substituting, $-6d + 7d = 1$, so $d = 1$ and $c = -2$:

$$-2u + 1v + 0w = (-2 + 2, -6 + 7) = (0, 1). \quad \checkmark$$

Solution B (set $d = 0$): $c + e = 0$ and $3c + 5e = 1$. From the first $c = -e$; substituting, $-3e + 5e = 1$, so $e = \frac{1}{2}$ and $c = -\frac{1}{2}$:

$$-\frac{1}{2}\mathbf{u} + \frac{1}{2}\mathbf{w} = \left(-\frac{1}{2} + \frac{1}{2}, -\frac{3}{2} + \frac{5}{2}\right) = (0, 1). \checkmark$$

The delicate part. Will *any* three vectors in the plane admit two different combinations for a given $\mathbf{b}$? Not always. Three vectors in $\mathbb{R}^2$ are certainly dependent, so if $\mathbf{b}$ is reachable at all there are infinitely many recipes. But $\mathbf{b}$ need not be reachable: if all three vectors lie on one line (say all multiples of $(1, 1)$) and $\mathbf{b}$ is off that line, there are *zero* combinations. So the correct statement is: either none, or infinitely many.

A.3 Tutorial 04 — Inverses, rank and elimination

Q1, Q4, Q5(a), Q6, Q10(a), Q11, Q12(a), Q13 are worked in Chapters 4–9. Additional solutions:

Worked Example A.5 • T04 Q2 — reducing to triangular and to I

(a) Reduce $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 1 & 2 \end{bmatrix}$ to upper triangular form.

$$A \xrightarrow{R_2 - 2R_1, R_3 - 3R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & -5 & -7 \end{bmatrix} \xrightarrow{R_3 + 5R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -2 \end{bmatrix} = U.$$

Pivots 1, 1, -2 ; $\det A = 1 \cdot 1 \cdot (-2) = -2$.

(b) Reduce $B = \begin{bmatrix} 1 & 3 & 3 \\ 2 & 4 & 10 \\ 3 & 8 & 4 \end{bmatrix}$ to the identity.

$$B \xrightarrow{R_2 - 2R_1, R_3 - 3R_1} \begin{bmatrix} 1 & 3 & 3 \\ 0 & -2 & 4 \\ 0 & -1 & -5 \end{bmatrix} \xrightarrow{R_2 / (-2)} \begin{bmatrix} 1 & 3 & 3 \\ 0 & 1 & -2 \\ 0 & -1 & -5 \end{bmatrix} \xrightarrow{R_3 + R_2} \begin{bmatrix} 1 & 3 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & -7 \end{bmatrix}$$

$$\xrightarrow{R_3 / (-7)} \begin{bmatrix} 1 & 3 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_2 + 2R_3, R_1 - 3R_3} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_1 - 3R_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I.$$

Since B reduces to I , it is invertible; $\det B = 1(-2)(-7) = 14$.

Worked Example A.6 • T04 Q5(b) — a 4×4 inverse

Find B^{-1} for $B = \begin{bmatrix} 2 & -6 & -2 & -3 \\ 5 & -13 & -4 & -7 \\ -1 & 4 & 1 & 2 \\ 0 & 1 & 0 & 1 \end{bmatrix}$.

Row reduce $[B \mid I]$. Begin by using the third row (which has a convenient -1) as the first pivot row: swap $R_1 \leftrightarrow R_3$ and negate.

$$\left[\begin{array}{cccc|cccc} 1 & -4 & -1 & -2 & 0 & 0 & -1 & 0 \\ 5 & -13 & -4 & -7 & 0 & 1 & 0 & 0 \\ 2 & -6 & -2 & -3 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array} \right]$$

$R_2 \rightarrow R_2 - 5R_1, R_3 \rightarrow R_3 - 2R_1$:

$$\left[\begin{array}{cccc|cccc} 1 & -4 & -1 & -2 & 0 & 0 & -1 & 0 \\ 0 & 7 & 1 & 3 & 0 & 1 & 5 & 0 \\ 0 & 2 & 0 & 1 & 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array} \right]$$

Swap $R_2 \leftrightarrow R_4$ to get a leading 1:

$$\left[\begin{array}{cccc|cccc} 1 & -4 & -1 & -2 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 & 1 & 0 & 2 & 0 \\ 0 & 7 & 1 & 3 & 0 & 1 & 5 & 0 \end{array} \right] \xrightarrow{R_3 - 2R_2, R_4 - 7R_2} \left[\begin{array}{cccc|cccc} 1 & -4 & -1 & -2 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 & 0 & 2 & -2 \\ 0 & 0 & 1 & -4 & 0 & 1 & 5 & -7 \end{array} \right]$$

Swap $R_3 \leftrightarrow R_4$, then negate the new R_4 :

$$\left[\begin{array}{cccc|cccc} 1 & -4 & -1 & -2 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -4 & 0 & 1 & 5 & -7 \\ 0 & 0 & 0 & 1 & -1 & 0 & -2 & 2 \end{array} \right]$$

Now clear upwards: $R_3 \rightarrow R_3 + 4R_4$, then $R_2 \rightarrow R_2 - R_4$, then $R_1 \rightarrow R_1 + 2R_4$, then $R_1 \rightarrow R_1 + R_3$, and finally $R_1 \rightarrow R_1 + 4R_2$. The left block becomes I and the right block is

$$B^{-1} = \begin{bmatrix} -2 & 1 & 0 & 1 \\ 1 & 0 & 2 & -1 \\ -4 & 1 & -3 & 1 \\ -1 & 0 & -2 & 2 \end{bmatrix}.$$

Verification — do this every time. Row 1 of B is $(2, -6, -2, -3)$.

- Against column 1 of B^{-1} , $(-2, 1, -4, -1)$: $-4 - 6 + 8 + 3 = 1$ ✓
- Against column 2, $(1, 0, 1, 0)$: $2 + 0 - 2 + 0 = 0$ ✓
- Against column 3, $(0, 2, -3, -2)$: $0 - 12 + 6 + 6 = 0$ ✓
- Against column 4, $(1, -1, 1, 2)$: $2 + 6 - 2 - 6 = 0$ ✓

Row 4 of B is $(0, 1, 0, 1)$; against column 4: $0 + (-1) + 0 + 2 = 1$ ✓.

Consistency. $\det B = -1$, and $\det(B^{-1}) = 1/\det B = -1$ as well — both matrices have integer entries precisely because $\det B = \pm 1$.

Remark • strategy for a 4×4 by hand

This computation is long and error-prone. Two habits make it survivable: start by manoeuvring a row containing ± 1 into the first pivot position (here row 3), which keeps every subsequent multiplier an integer; and *always* spot-check several entries of BB^{-1} at the end. Under exam conditions, prefer Gauss–Jordan over the adjugate for 4×4 ; in practice use Octave’s `B\eye(4)`.

Worked Example A.7 • T04 Q9(a) — echelon form, basic columns, rank

For $A = \begin{bmatrix} 2 & 2 & -1 \\ 4 & 0 & 2 \\ 0 & 6 & -3 \end{bmatrix}$, find the row echelon form, the basic columns, and the rank.

$$A \xrightarrow{R_2 - 2R_1} \begin{bmatrix} 2 & 2 & -1 \\ 0 & -4 & 4 \\ 0 & 6 & -3 \end{bmatrix} \xrightarrow{R_3 + \frac{3}{2}R_2} \begin{bmatrix} 2 & 2 & -1 \\ 0 & -4 & 4 \\ 0 & 0 & 3 \end{bmatrix}.$$

Row echelon form: $\begin{bmatrix} 2 & 2 & -1 \\ 0 & -4 & 4 \\ 0 & 0 & 3 \end{bmatrix}.$

Pivots sit in columns 1, 2, 3, so all three columns are **basic** and

$$\text{rank } A = 3.$$

A is therefore invertible; indeed $\det A = 2(-4)(3) = -24 \neq 0$.

Worked Example A.8 • T04 Q9(c) — a rank-deficient case

For $C = \begin{bmatrix} 1 & 2 & 1 & -1 \\ 9 & 5 & 2 & 2 \\ 7 & 1 & 0 & 4 \end{bmatrix}$, find the echelon form, basic columns and rank.

$$C \xrightarrow{R_2 - 9R_1, R_3 - 7R_1} \begin{bmatrix} 1 & 2 & 1 & -1 \\ 0 & -13 & -7 & 11 \\ 0 & -13 & -7 & 11 \end{bmatrix} \xrightarrow{R_3 - R_2} \begin{bmatrix} 1 & 2 & 1 & -1 \\ 0 & -13 & -7 & 11 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Pivots: columns 1 and 2. **Basic columns:** (1, 9, 7) and (2, 5, 1) — the *original* columns. **Rank:** $\text{rank } C = 2$.

Consequently $\dim \mathbf{N}(C) = 4 - 2 = 2$ and $\dim \mathbf{N}(C^T) = 3 - 2 = 1$. The zero row appeared because row 3 of C equals row 2 minus $2 \times$ row 1: check $(9, 5, 2, 2) - 2(1, 2, 1, -1) = (7, 1, 0, 4) \checkmark$.

Worked Example A.9 • T04 Q7 — normal form PAQ

For $A = \begin{bmatrix} 2 & 1 & -3 & -6 \\ 3 & -3 & 1 & 2 \\ 1 & 1 & 1 & 2 \end{bmatrix}$, find non-singular P and Q with PAQ in normal form $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$.

Step 1 — row reduce (this builds P). Put the simple row first, $R_1 \leftrightarrow R_3$:

$$\begin{aligned} & \begin{bmatrix} 1 & 1 & 1 & 2 \\ 3 & -3 & 1 & 2 \\ 2 & 1 & -3 & -6 \end{bmatrix} \xrightarrow{R_2 - 3R_1, R_3 - 2R_1} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & -6 & -2 & -4 \\ 0 & -1 & -5 & -10 \end{bmatrix} \\ & \xrightarrow{R_2 \leftrightarrow R_3, R_2 \rightarrow -R_2} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 5 & 10 \\ 0 & -6 & -2 & -4 \end{bmatrix} \xrightarrow{R_3 + 6R_2} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 5 & 10 \\ 0 & 0 & 28 & 56 \end{bmatrix} \\ & \xrightarrow{R_3/28} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 5 & 10 \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{\text{clear above}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}. \end{aligned}$$

So rank $A = 3$.

Step 2 — column operations (these build Q). One column operation finishes the job: $C_4 \rightarrow C_4 - 2C_3$ removes the last entry, giving

$$PAQ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = [I_3 \mid \mathbf{0}],$$

the normal form with $r = 3$. Here Q is the identity with -2 in position $(3, 4)$:

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and P is the product of the elementary row matrices used, obtainable by applying the same row operations to I_3 .

Remark

The normal form theorem says every $m \times n$ matrix of rank r is equivalent to $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$. Rank is the *only* invariant under row *and* column operations together — which is why it is such a fundamental quantity.

A.4 Tutorial 05 — Determinants

Q1, Q2(a), Q3, Q4, Q5, Q6(b)(i), Q7 are worked in Chapters 6 and 7. Additionally:

Worked Example A.10 • T05 Q2(b) — verifying a 3×3 inverse pair

Verify that $A = \begin{bmatrix} 1 & 3 & 3 \\ 1 & 3 & 4 \\ 1 & 4 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & -3 & -3 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix}$ are inverses.

Compute AB row by row.

Row 1 of A is $(1, 3, 3)$:

$$(1, 3, 3) \cdot (7, -1, -1) = 7 - 3 - 3 = 1,$$

$$(1, 3, 3) \cdot (-3, 0, 1) = -3 + 0 + 3 = 0,$$

$$(1, 3, 3) \cdot (-3, 1, 0) = -3 + 3 + 0 = 0.$$

Row 2 of A is $(1, 3, 4)$:

$$(1, 3, 4) \cdot (7, -1, -1) = 7 - 3 - 4 = 0, \quad (1, 3, 4) \cdot (-3, 0, 1) = -3 + 0 + 4 = 1,$$

$$(1, 3, 4) \cdot (-3, 1, 0) = -3 + 3 + 0 = 0.$$

Row 3 of A is $(1, 4, 3)$:

$$(1, 4, 3) \cdot (7, -1, -1) = 7 - 4 - 3 = 0, \quad (1, 4, 3) \cdot (-3, 0, 1) = -3 + 0 + 3 = 0,$$

$$(1, 4, 3) \cdot (-3, 1, 0) = -3 + 4 + 0 = 1.$$

So $AB = I$ ✓, and since both are square this suffices (Chapter 5) — $B = A^{-1}$.

Cross-check via the determinant. Expanding A along row 1: $1(9 - 16) - 3(3 - 4) + 3(4 - 3) = -7 + 3 + 3 = -1$. And $\det B = 7(0 - 1) + 3(0 + 1) - 3(-1 - 0) = -7 + 3 + 3 = -1$. Consistently $\det A \det B = 1 = \det I$ ✓.

Worked Example A.11 • T05 Q6(a)–(b)(ii) — Cramer's rule, statement and use

(a) State Cramer's rule. If $Ax = b$ with A an $n \times n$ matrix and $\det A \neq 0$, the unique solution has components

$$x_j = \frac{\det A_j}{\det A}, \quad j = 1, \dots, n,$$

where A_j is A with its j th column replaced by b .

(b)(ii) Solve $-7x - 6y - 12z = 0$, $5x + 5y + 7z = 0$, $x + 4z = 0$.

The system is homogeneous. Its determinant is

$$\det A = \begin{vmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{vmatrix}.$$

Expanding along row 3 (which contains a zero):

$$= 1 \begin{vmatrix} -6 & -12 \\ 5 & 7 \end{vmatrix} - 0 + 4 \begin{vmatrix} -7 & -6 \\ 5 & 5 \end{vmatrix} = (-42 + 60) + 4(-35 + 30) = 18 - 20 = -2.$$

Since $\det A = -2 \neq 0$, A is nonsingular and the only solution is the trivial one:

$$x = y = z = 0.$$

Cramer's rule confirms it directly — each A_j has a column of zeros (the replaced column is $\mathbf{b} = \mathbf{0}$), so $\det A_j = 0$ and $x_j = 0/(-2) = 0$ for every j .

Remark

A homogeneous system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution *if and only if* $\det A = 0$. Computing the determinant answers the question before any solving begins.

A.5 Tutorial 06 — Eigenvalues (Part A)

Q1–Q7 of Part A are proved in full in Chapters 11 and 12:

Question	Content	Where
Q1(a)	definitions of eigenvalue, eigenvector, eigenspace	Ch. 11, Def. 11.1
Q1(b)	why $\mathbf{x} \neq \mathbf{0}$ is essential	Ch. 11, pitfall box
Q1(c)	$c\mathbf{x}$ is also an eigenvector	Ch. 11, pitfall box
Q1(d)	$E_\lambda = \ker(A - \lambda I)$ is a subspace	Ch. 11, Thm. 11.6
Q2(a)	$A^k \mathbf{x} = \lambda^k \mathbf{x}$ by induction	Ch. 11, Thm. 11.4
Q2(b)	$p(A)\mathbf{x} = p(\lambda)\mathbf{x}$	Ch. 11, Thm. 11.4
Q2(c)	counterexample for the converse	Ch. 11, Ex. 11.4
Q3	$\det(A - \lambda I) = 0$; $\lambda = 0 \iff$ singular	Ch. 11, Thm. 11.1
Q4	distinct eigenvalues $\Rightarrow$ independence	Ch. 11, Thm. 11.5
Q5	$\text{tr } A = \sum \lambda_i$, $\det A = \prod \lambda_i$	Ch. 11, Thm. 11.3
Q6	$AX = X\Lambda$, $A^k = X\Lambda^k X^{-1}$, multiplicities	Ch. 12
Q7	similar matrices	Ch. 11, Thm. 11.7

Worked Example A.12 • T06 Q1(a) — the three definitions, stated for the exam

Let $A \in \mathbb{R}^{n \times n}$.

Eigenvalue. A scalar $\lambda \in \mathbb{R}$ is an eigenvalue of A if there exists a vector $\mathbf{x} \in \mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{0}$ such that $A\mathbf{x} = \lambda\mathbf{x}$.

Eigenvector. Such a vector $\mathbf{x}$ is called an eigenvector of A corresponding to the eigenvalue λ . (It is required to be nonzero; it is not unique, since every nonzero multiple works too.)

Eigenspace. For an eigenvalue λ of A , the eigenspace is

$$E_\lambda = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \lambda\mathbf{x}\} = \ker(A - \lambda I) = \mathbf{N}(A - \lambda I).$$

It is a subspace of $\mathbb{R}^n$, consisting of all eigenvectors for λ together with the zero vector. Its dimension is the geometric multiplicity of λ , and $\dim E_\lambda = n - \text{rank}(A - \lambda I)$.

Worked Example A.13 • T06 Q1(c) — scalar multiples of eigenvectors

Prove that if $\mathbf{x}$ is an eigenvector of A for λ , then so is $c\mathbf{x}$ for every scalar $c \neq 0$.

Since $\mathbf{x} \neq \mathbf{0}$ and $c \neq 0$, the vector $c\mathbf{x}$ is nonzero, so it is eligible to be an eigenvector. And

$$A(c\mathbf{x}) = c(A\mathbf{x}) = c(\lambda\mathbf{x}) = \lambda(c\mathbf{x}),$$

using linearity of matrix multiplication in the first step and commutativity of scalar multiplication in the last. Hence cx satisfies the eigenvector equation for the same λ .

Why $c \neq 0$ matters. For $c = 0$ the calculation still holds numerically, but $0x = \mathbf{0}$ is excluded by definition from being an eigenvector — which is precisely the point of Q1(b).

Assignment 01 — Full Solutions

Assignment 01 (duration 1 hour 45 minutes) has three questions worth 40, 30 and 30 marks. Between them they test cofactor expansion, row reduction with a parameter, consistency conditions, column-space geometry, dot products, independence and rank — essentially the whole first half of the course. Complete solutions follow, with the marking logic indicated.

B.1 Question 1 (40 marks)

Let

$$A(t) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & t & 2 \\ 1 & 2 & t \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

- Find $\det(A(t))$ using cofactor expansion.
- For which values of t does $A(t)\mathbf{x} = \mathbf{b}$ have a unique solution for every $\mathbf{b} \in \mathbb{R}^3$?
- When $t = 0$, use row reduction to find the condition on b_1, b_2, b_3 for consistency.
- When $t = 2$, do the same.
- Explain geometrically what happens to the column space when $t = 0$ or $t = 2$. Is it all of $\mathbb{R}^3$, a plane, a line, or only the zero vector?

(a) The determinant

Expand along the first row. The three minors are

$$M_{11} = \begin{vmatrix} t & 2 \\ 2 & t \end{vmatrix} = t^2 - 4, \quad M_{12} = \begin{vmatrix} 1 & 2 \\ 1 & t \end{vmatrix} = t - 2, \quad M_{13} = \begin{vmatrix} 1 & t \\ 1 & 2 \end{vmatrix} = 2 - t.$$

With the checkerboard signs $+, -, +$:

$$\det A(t) = 1(t^2 - 4) - 1(t - 2) + 1(2 - t).$$

Expanding and collecting,

$$\det A(t) = t^2 - 4 - t + 2 + 2 - t = t^2 - 2t,$$

so

$\det A(t) = t(t - 2)$

Remark • a useful cross-check

Setting $t = 1$ makes rows 2 and 3 read $(1, 1, 2)$ and $(1, 2, 1)$, and the formula gives $\det = 1(1 - 2) = -1$. Direct computation of $\det \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$ by Sarrus: $1 + 2 + 2 - 1 - 4 - 1 = -1 \checkmark$.

(b) When the solution is unique for every $\mathbf{b}$

$A(t)\mathbf{x} = \mathbf{b}$ has a unique solution for *every* $\mathbf{b} \in \mathbb{R}^3$ precisely when $A(t)$ is invertible (Invertible Matrix Theorem, Chapter 5), i.e. when $\det A(t) \neq 0$:

$$t(t - 2) \neq 0 \quad \Longleftrightarrow \quad \boxed{t \neq 0 \text{ and } t \neq 2}$$

For $t = 0$ or $t = 2$ the matrix is singular: some $\mathbf{b}$ have no solution and the rest have infinitely many.

(c) The case $t = 0$

$$A(0) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}.$$

Row reduce the augmented matrix:

$$\begin{aligned} & \left[\begin{array}{ccc|c} 1 & 1 & 1 & b_1 \\ 1 & 0 & 2 & b_2 \\ 1 & 2 & 0 & b_3 \end{array} \right] \xrightarrow{R_2 - R_1, R_3 - R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & b_1 \\ 0 & -1 & 1 & b_2 - b_1 \\ 0 & 1 & -1 & b_3 - b_1 \end{array} \right] \\ & \xrightarrow{R_3 + R_2} \left[\begin{array}{ccc|c} 1 & 1 & 1 & b_1 \\ 0 & -1 & 1 & b_2 - b_1 \\ 0 & 0 & 0 & b_3 - b_1 + b_2 - b_1 \end{array} \right] = \left[\begin{array}{ccc|c} 1 & 1 & 1 & b_1 \\ 0 & -1 & 1 & b_2 - b_1 \\ 0 & 0 & 0 & b_2 + b_3 - 2b_1 \end{array} \right] \end{aligned}$$

The last row reads $0 = b_2 + b_3 - 2b_1$. Hence the system is consistent if and only if

$$\boxed{b_2 + b_3 - 2b_1 = 0, \quad \text{i.e.} \quad b_3 = 2b_1 - b_2}$$

When this holds there are only 2 pivots and 3 unknowns, so there is one free variable and infinitely many solutions.

Verification. Take $b_1 = 1, b_2 = 1$, giving $b_3 = 1$. Then $\mathbf{b} = (1, 1, 1)$ and $\mathbf{x} = (1, 0, 0)$ solves the system: $A(0)(1, 0, 0)^T = (1, 1, 1)^T \checkmark$. Whereas $\mathbf{b} = (1, 0, 0)$ has $b_2 + b_3 - 2b_1 = -2 \neq 0$ and is unreachable.

(d) The case $t = 2$

$$A(2) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

Rows 2 and 3 are identical, which makes the answer nearly immediate, but reducing properly:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & b_1 \\ 1 & 2 & 2 & b_2 \\ 1 & 2 & 2 & b_3 \end{array} \right] \xrightarrow{R_2 - R_1, R_3 - R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & b_1 \\ 0 & 1 & 1 & b_2 - b_1 \\ 0 & 1 & 1 & b_3 - b_1 \end{array} \right]$$

$$\xrightarrow{R_3 - R_2} \left[\begin{array}{ccc|c} 1 & 1 & 1 & b_1 \\ 0 & 1 & 1 & b_2 - b_1 \\ 0 & 0 & 0 & b_3 - b_2 \end{array} \right]$$

The last row gives $0 = b_3 - b_2$, so the consistency condition is

$$\boxed{b_2 = b_3}$$

Again 2 pivots, so when it is satisfied there are infinitely many solutions with one free variable.

Verification. $\mathbf{b} = (1, 2, 2)$ satisfies $b_2 = b_3$, and $\mathbf{x} = (0, 1, 0)$ gives $A(2)(0, 1, 0)^T = (1, 2, 2)^T \checkmark$.

(e) Geometry of the column space

In both singular cases the answer is the same in kind: $\mathbf{C}(A)$ is a **plane through the origin** in $\mathbb{R}^3$.

Why a plane and not a line. In each case the reduction produced exactly 2 pivots, so $\text{rank } A = 2$ and $\dim \mathbf{C}(A) = 2$ — a two-dimensional subspace of $\mathbb{R}^3$, which is a plane through the origin. It cannot be a line (that would need rank 1) or $\{\mathbf{0}\}$ (rank 0), because the first two columns are independent for every t : no scalar multiple of $(1, 1, 1)$ equals $(1, t, 2)$ unless $t = 1$ and $2 = 1$, which is impossible.

Which plane. The consistency conditions found in (c) and (d) are the equations of these planes:

	rank A	$\mathbf{C}(A)$	Normal vector
$t \notin \{0, 2\}$	3	all of $\mathbb{R}^3$	—
$t = 0$	2	plane $2b_1 - b_2 - b_3 = 0$	$(2, -1, -1)$
$t = 2$	2	plane $b_2 - b_3 = 0$	$(0, 1, -1)$

Interpretation. For generic t the three columns are independent and span all of $\mathbb{R}^3$: every target $\mathbf{b}$ is reachable, by exactly one recipe. At $t = 0$ and $t = 2$ one column falls into the plane of the other two — the third direction is lost. The set of reachable $\mathbf{b}$ collapses from a solid three-dimensional space to a flat two-dimensional sheet, and only $\mathbf{b}$ lying exactly on that sheet can be produced. For those that do, the lost direction reappears as a line of freedom in $\mathbf{x}$ (the null space), which is why the solution becomes non-unique at the same moment it becomes conditional.

Remark • the dimension audit

At $t = 0$ or $t = 2$: $\dim \mathbf{C}(A) = 2$ and $\dim \mathbf{N}(A^T) = 3 - 2 = 1$ (the normal direction), summing to $3 = m \checkmark$; $\dim \mathbf{C}(A^T) = 2$ and $\dim \mathbf{N}(A) = 3 - 2 = 1$, summing to $3 = n \checkmark$.

B.2 Question 2 (30 marks)

Let $\mathbf{n} = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$, and consider the plane $P = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{n} \cdot \mathbf{x} = 0\}$.

- Show that both $\mathbf{a}$ and $\mathbf{b}$ lie in P .
- Show that $\mathbf{a}$ and $\mathbf{b}$ are linearly independent.
- Find the angle between $\mathbf{a}$ and $\mathbf{b}$.
- Express $\mathbf{r} = (3, 2, -2)$ as a combination of $\mathbf{a}$ and $\mathbf{b}$; hence explain why $\mathbf{r} \in P$.
- Explain geometrically why all $\mathbf{x}$ with $\mathbf{n} \cdot \mathbf{x} = 0$ form a plane through the origin.

(a) Membership

$$\mathbf{n} \cdot \mathbf{a} = 2(1) + (-1)(0) + 2(-1) = 2 - 0 - 2 = 0 \Rightarrow \mathbf{a} \in P \checkmark$$

$$\mathbf{n} \cdot \mathbf{b} = 2(0) + (-1)(2) + 2(1) = 0 - 2 + 2 = 0 \Rightarrow \mathbf{b} \in P \checkmark$$

(b) Independence

Suppose $c_1\mathbf{a} + c_2\mathbf{b} = \mathbf{0}$. Componentwise:

$$(1\text{st}) \quad c_1 = 0, \quad (2\text{nd}) \quad 2c_2 = 0, \quad (3\text{rd}) \quad -c_1 + c_2 = 0.$$

The first equation forces $c_1 = 0$ and the second forces $c_2 = 0$ (the third is then automatically satisfied). Only the trivial combination gives $\mathbf{0}$, so $\mathbf{a}$ and $\mathbf{b}$ are **linearly independent**.

Alternative one-line argument. Two vectors are dependent only if one is a scalar multiple of the other. Every multiple $k\mathbf{a}$ has second component 0, whereas $\mathbf{b}$ has second component $2 \neq 0$; and $\mathbf{a} \neq \mathbf{0}$, so $\mathbf{a}$ is not a multiple of $\mathbf{b}$ either.

Consequence. Since $\dim P = 2$ (see part (e)) and $\mathbf{a}, \mathbf{b}$ are two independent vectors lying in P , they form a *basis* of P .

(c) The angle

$$\mathbf{a} \cdot \mathbf{b} = 1(0) + 0(2) + (-1)(1) = -1,$$

$$\|\mathbf{a}\| = \sqrt{1^2 + 0^2 + (-1)^2} = \sqrt{2}, \quad \|\mathbf{b}\| = \sqrt{0^2 + 2^2 + 1^2} = \sqrt{5}.$$

Therefore

$$\cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} = \frac{-1}{\sqrt{2} \cdot \sqrt{5}} = \frac{-1}{\sqrt{10}} \approx -0.31623,$$

$$\theta = \arccos\left(-\frac{1}{\sqrt{10}}\right) \approx 108.43^\circ \quad (\approx 1.8925 \text{ rad})$$

The negative dot product signals an obtuse angle before any arithmetic is done. Note also that $\theta \neq 90^\circ$: the two vectors span the plane but are not perpendicular, so they form a basis of P that is not an *orthogonal* basis.

(d) Expressing r

Solve $c_1 \mathbf{a} + c_2 \mathbf{b} = \mathbf{r}$, i.e.

$$c_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -2 \end{bmatrix}.$$

Componentwise:

$$(1\text{st}) \quad c_1 = 3, \quad (2\text{nd}) \quad 2c_2 = 2 \Rightarrow c_2 = 1, \quad (3\text{rd}) \quad -c_1 + c_2 = -3 + 1 = -2 \checkmark$$

The third equation is consistent, so

$$\boxed{\mathbf{r} = 3\mathbf{a} + \mathbf{b}}$$

Why $\mathbf{r} \in P$. P is a subspace (it is the null space of the 1×3 matrix $\mathbf{n}^\top$), so it is closed under linear combinations. Since $\mathbf{a}, \mathbf{b} \in P$ by part (a), any combination of them — in particular $3\mathbf{a} + \mathbf{b} = \mathbf{r}$ — also lies in P .

Direct confirmation. $\mathbf{n} \cdot \mathbf{r} = 2(3) + (-1)(2) + 2(-2) = 6 - 2 - 4 = 0 \checkmark$.

(e) Why $\mathbf{n} \cdot \mathbf{x} = 0$ is a plane through the origin

Geometric argument. The condition $\mathbf{n} \cdot \mathbf{x} = 0$ says $\mathbf{x}$ makes an angle of 90° with the fixed nonzero direction $\mathbf{n}$. In $\mathbb{R}^3$, the set of *all* directions perpendicular to one given direction sweeps out a flat two-dimensional sheet — a plane. It passes through the origin because $\mathbf{n} \cdot \mathbf{0} = 0$, so $\mathbf{0}$ always satisfies the condition.

Algebraic confirmation. $P = \mathbf{N}(\mathbf{n}^\top)$ where $\mathbf{n}^\top = [2 \ -1 \ 2]$ is a 1×3 matrix of rank 1 (it is nonzero). By the rank–nullity theorem (Chapter 9),

$$\dim P = 3 - \text{rank}(\mathbf{n}^\top) = 3 - 1 = 2.$$

A two-dimensional subspace of $\mathbb{R}^3$ is exactly a plane through the origin.

Subspace check, explicitly. $\mathbf{0} \in P$; and if $\mathbf{x}, \mathbf{y} \in P$ with scalars c, d , then

$$\mathbf{n} \cdot (c\mathbf{x} + d\mathbf{y}) = c(\mathbf{n} \cdot \mathbf{x}) + d(\mathbf{n} \cdot \mathbf{y}) = c(0) + d(0) = 0,$$

so $c\mathbf{x} + d\mathbf{y} \in P$. In the language of Chapter 10, $P = \text{span}\{\mathbf{n}\}^\perp$, and $\dim \text{span}\{\mathbf{n}\} + \dim P = 1 + 2 = 3 \checkmark$.

B.3 Question 3 (30 marks)

Let

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix}.$$

- (a) Show that all columns of A lie on the same line. State rank A .
- (b) Describe $\mathbf{C}(A)$ geometrically.
- (c) Determine whether $A\mathbf{x} = (12, 6, -6)$ has no solution, one solution, or infinitely many. Find all solutions if they exist.
- (d) Determine whether $A\mathbf{x} = (1, 0, 0)$ is solvable, justifying via the column space.
- (e) Use cofactor expansion to compute $\det A$. Explain why A is not invertible.

(a) All columns on one line

The columns are

$$\mathbf{a}_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 6 \\ 3 \\ -3 \end{bmatrix}.$$

Directly,

$$\mathbf{a}_2 = 2\mathbf{a}_1, \quad \mathbf{a}_3 = 3\mathbf{a}_1.$$

Every column is a scalar multiple of $\mathbf{a}_1$, so all three lie on the line through the origin in the direction $(2, 1, -1)$.

Only one column is independent, so

$$\boxed{\text{rank } A = 1}$$

(b) The column space

$$\mathbf{C}(A) = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} = \text{span}\{\mathbf{a}_1\} = \left\{ t \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} : t \in \mathbb{R} \right\}.$$

This is a **line through the origin** in $\mathbb{R}^3$ with direction vector $(2, 1, -1)$ — a one-dimensional subspace, consistent with $\dim \mathbf{C}(A) = \text{rank } A = 1$.

Equivalently, $\mathbf{b} = (b_1, b_2, b_3)$ lies in $\mathbf{C}(A)$ if and only if

$$\frac{b_1}{2} = \frac{b_2}{1} = \frac{b_3}{-1}, \quad \text{i.e.} \quad b_1 = 2b_2 \text{ and } b_3 = -b_2.$$

(c) Solving $A\mathbf{x} = (12, 6, -6)$

Is it consistent? $(12, 6, -6) = 6(2, 1, -1) = 6\mathbf{a}_1$, which lies on the line $\mathbf{C}(A)$. So yes — the system is consistent.

Row reduce.

$$\left[\begin{array}{ccc|c} 2 & 4 & 6 & 12 \\ 1 & 2 & 3 & 6 \\ -1 & -2 & -3 & -6 \end{array} \right] \xrightarrow{R_1 \rightarrow \frac{1}{2}R_1} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 1 & 2 & 3 & 6 \\ -1 & -2 & -3 & -6 \end{array} \right] \xrightarrow{R_2 - R_1, R_3 + R_1} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

One pivot, no contradictory row. With $n = 3$ and $r = 1$ there are $3 - 1 = 2$ free variables, hence **infinitely many solutions**.

Complete solution. The single equation is $x_1 + 2x_2 + 3x_3 = 6$. Take $x_2 = s$ and $x_3 = t$ free; then $x_1 = 6 - 2s - 3t$:

$$\boxed{\mathbf{x} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}}$$

This is a plane of solutions in $\mathbb{R}^3$: a particular solution $(6, 0, 0)$ plus the two-dimensional null space (Chapter 9).

Check. With $s = 1, t = 0$: $\mathbf{x} = (4, 1, 0)$ and $A\mathbf{x} = (8 + 4 + 0, 4 + 2 + 0, -4 - 2 - 0) = (12, 6, -6) \checkmark$.

(d) Solving $Ax = (1, 0, 0)$

$Ax = b$ is solvable if and only if $b \in C(A)$ (Chapter 9). Here $C(A)$ is the line $\{t(2, 1, -1)\}$. For $(1, 0, 0)$ to lie on it we would need

$$2t = 1, \quad t = 0, \quad -t = 0,$$

and the first requires $t = \frac{1}{2}$ while the second requires $t = 0$ — a contradiction. So

$$b = (1, 0, 0) \notin C(A) : \text{the system has no solution}$$

Confirmation by elimination.

$$\left[\begin{array}{ccc|c} 2 & 4 & 6 & 1 \\ 1 & 2 & 3 & 0 \\ -1 & -2 & -3 & 0 \end{array} \right] \xrightarrow{R_1 - 2R_2, R_3 + R_2} \left[\begin{array}{ccc|c} 0 & 0 & 0 & 1 \\ 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

The row $(0 \ 0 \ 0 \mid 1)$ reads $0 = 1$ — inconsistent. In rank terms, $\text{rank } A = 1$ but $\text{rank}[A \mid b] = 2$, and $\text{rank}[A \mid b] > \text{rank } A$ is exactly the signature of inconsistency.

(e) The determinant

Expand along the first row:

$$\det A = 2 \begin{vmatrix} 2 & 3 \\ -2 & -3 \end{vmatrix} - 4 \begin{vmatrix} 1 & 3 \\ -1 & -3 \end{vmatrix} + 6 \begin{vmatrix} 1 & 2 \\ -1 & -2 \end{vmatrix}.$$

Each 2×2 determinant vanishes:

$$\begin{vmatrix} 2 & 3 \\ -2 & -3 \end{vmatrix} = 2(-3) - 3(-2) = -6 + 6 = 0,$$

$$\begin{vmatrix} 1 & 3 \\ -1 & -3 \end{vmatrix} = 1(-3) - 3(-1) = -3 + 3 = 0,$$

$$\begin{vmatrix} 1 & 2 \\ -1 & -2 \end{vmatrix} = 1(-2) - 2(-1) = -2 + 2 = 0.$$

Hence

$$\det A = 2(0) - 4(0) + 6(0) = 0$$

Why A is not invertible. Several equivalent reasons, any one of which earns the marks:

- $\det A = 0$, and a matrix is invertible if and only if its determinant is nonzero (Chapter 7).
- The columns are linearly dependent — $a_2 = 2a_1$ and $a_3 = 3a_1$ — so they cannot form a basis of $\mathbb{R}^3$.
- $\text{rank } A = 1 < 3$, so A does not have full rank.
- $C(A)$ is a line, not all of $\mathbb{R}^3$, so $Ax = b$ fails to be solvable for some b — as part (d) exhibited concretely.
- $N(A) \neq \{0\}$: for instance $A(-2, 1, 0)^T = 0$, so A cannot be one-to-one.

Geometrically, A crushes all of $\mathbb{R}^3$ onto a single line. Such a collapse destroys information irreversibly, and no matrix can undo it.

B.4 Marks summary and common errors

Part	Tests	Where marks are typically lost
Q1(a)	cofactor expansion	sign errors in $(-1)^{i+j}$; not simplifying to factored form
Q1(b)	Invertible Matrix Theorem	answering " $t = 0, 2$ " instead of " $t \neq 0, 2$ "
Q1(c,d)	row reduction with symbols	dropping the b_i through the operations
Q1(e)	column space geometry	saying "line" without checking the rank
Q2(a,b)	dot product, independence	asserting independence without proof
Q2(c)	angle formula	forgetting that a negative $\cos \theta$ gives an obtuse angle
Q2(d)	linear combination	not justifying $r \in P$ via closure
Q2(e)	subspace / rank-nullity	describing rather than proving
Q3(a,b)	rank and column space	claiming rank 2 by miscounting
Q3(c)	complete solution	giving one solution instead of all
Q3(d)	solvability	answering without reference to $\mathbf{C}(A)$
Q3(e)	determinant	computing $\det$ correctly but not explaining non-invertibility

Remark • the thread running through the paper

Every question is the same question in different clothing: *what does A reach, and how much freedom is left over?* Q1 varies a parameter until the reachable set collapses from $\mathbb{R}^3$ to a plane. Q2 constructs a plane directly and works inside it. Q3 gives a matrix whose reachable set is only a line. Rank is the number that governs all three.

Formula Reference Sheet

Everything worth memorising, on a few pages. Use this for last-minute revision; use the chapters for understanding.

C.1 Vectors

Dot product	$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = \sum_i v_i w_i$ (a scalar)
Length	$\ \mathbf{v}\ = \sqrt{\mathbf{v} \cdot \mathbf{v}}, \quad \ \mathbf{v}\ ^2 = \mathbf{v} \cdot \mathbf{v}$
Unit vector	$\mathbf{u} = \mathbf{v} / \ \mathbf{v}\ $
Angle	$\cos \theta = \frac{\mathbf{v} \cdot \mathbf{w}}{\ \mathbf{v}\ \ \mathbf{w}\ }$
Orthogonal	$\mathbf{v} \perp \mathbf{w} \iff \mathbf{v} \cdot \mathbf{w} = 0$
Expansion	$\ \mathbf{v} \pm \mathbf{w}\ ^2 = \ \mathbf{v}\ ^2 \pm 2\mathbf{v} \cdot \mathbf{w} + \ \mathbf{w}\ ^2$
Cauchy–Schwarz	$ \mathbf{v} \cdot \mathbf{w} \leq \ \mathbf{v}\ \ \mathbf{w}\ $
Triangle	$\ \mathbf{v} + \mathbf{w}\ \leq \ \mathbf{v}\ + \ \mathbf{w}\ $
Transpose rule	$(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A^T \mathbf{y})$

C.2 Matrix algebra

Product	$(AB)_{ij} = \sum_k a_{ik} b_{kj}, \quad (m \times n)(n \times p) = (m \times p)$
Column view	$AB = [A\mathbf{b}_1 \cdots A\mathbf{b}_p]; \quad A\mathbf{x} = \sum_j x_j \mathbf{a}_j$
Associative / distributive	$(AB)C = A(BC), \quad A(B + C) = AB + AC$
Not commutative	$AB \neq BA$ in general
Transpose	$(AB)^T = B^T A^T, \quad (A^T)^T = A$
Inverse	$(AB)^{-1} = B^{-1} A^{-1}, \quad (A^T)^{-1} = (A^{-1})^T$
2×2 inverse	$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
Trace	$\text{tr } A = \sum_i a_{ii}, \quad \text{tr}(AB) = \text{tr}(BA), \quad \text{tr}(A^T) = \text{tr } A$
Symmetric	$S^T = S; \quad A^T A$ and AA^T always symmetric psd
Orthogonal	$Q^T Q = I \Rightarrow Q^{-1} = Q^T$ (square Q)

C.3 Elimination and factorization

Row operations	$R_i \rightarrow R_i - \ell R_j, R_i \leftrightarrow R_j, R_i \rightarrow cR_i (c \neq 0)$
Multiplier	$\ell_{ij} = \frac{\text{entry to clear}}{\text{pivot}}$
LU	$A = LU$; L unit lower triangular holding the ℓ_{ij}
$PA = LU$	with row exchanges; $P^{-1} = P^T$
QR	$A = QR, Q^T Q = I, R$ upper triangular
Back substitution	$x_i = \frac{1}{u_{ii}} \left(c_i - \sum_{j>i} u_{ij} x_j \right)$
Gauss–Jordan inverse	$[A \mid I] \rightarrow [I \mid A^{-1}]$
Cost	elimination $\approx \frac{1}{3}n^3$; inverse $\approx n^3$

C.4 Determinants

2×2	$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$
3×3 (Sarrus)	$aei + bfg + cdh - ceg - afh - bdi$
Cofactor	$C_{ij} = (-1)^{i+j} M_{ij}$
Expansion	$\det A = \sum_k a_{ik} C_{ik}$ (any row or column)
Adjugate	$\text{adj } A = C^T, \quad A^{-1} = \frac{1}{\det A} \text{adj } A$
Cramer	$x_j = \frac{\det A_j}{\det A}$
Pivots	$\det A = (-1)^s \prod (\text{pivots}), s = \# \text{ row swaps}$
Product	$\det(AB) = \det A \det B$
Transpose	$\det(A^T) = \det A$
Scalar	$\det(cA) = c^n \det A$
Inverse	$\det(A^{-1}) = 1 / \det A$
Triangular	$\det A = \prod_i a_{ii}$
Warning	$\det(A + B) \neq \det A + \det B$

C.5 Vector spaces and the four subspaces

Subspace test	$\mathbf{0} \in W$ and $ku + cv \in W$
Independence	$\sum c_i v_i = \mathbf{0} \Rightarrow \text{all } c_i = 0$
Basis	independent <i>and</i> spanning
Dimension	size of any basis
Rank	$\# \text{ pivots} = \dim \mathbf{C}(A) = \dim \mathbf{C}(A^T)$
Rank–nullity	$\text{rank } A + \dim \mathbf{N}(A) = n$
Solvability	$Ax = b$ solvable $\iff b \in \mathbf{C}(A) \iff \text{rank}[A b] = \text{rank } A$
Complete solution	$x = x_p + \mathbf{N}(A)$

Subspace	Lives in	Dim	Basis from
$\mathbf{C}(A)$	$\mathbb{R}^m$	r	pivot columns of A
$\mathbf{N}(A)$	$\mathbb{R}^n$	$n - r$	special solutions
$\mathbf{C}(A^T)$	$\mathbb{R}^n$	r	nonzero rows of RREF
$\mathbf{N}(A^T)$	$\mathbb{R}^m$	$m - r$	row combinations giving $\mathbf{0}$

$$\mathbf{N}(A) = \mathbf{C}(A^T)^\perp, \quad \mathbf{N}(A^T) = \mathbf{C}(A)^\perp$$

C.6 Orthogonality and least squares

Projection onto a line	$p = \frac{a^T b}{a^T a} a, \quad P = \frac{a a^T}{a^T a}$
Normal equations	$A^T A \hat{x} = A^T b$
Projection matrix	$P = A(A^T A)^{-1} A^T; P^T = P, P^2 = P$
Complement	$I - P$ projects onto $\mathbf{N}(A^T)$
Gram-Schmidt	$w_k = a_k - \sum_{i < k} (a_k \cdot q_i) q_i, q_k = w_k / \ w_k\ $
QR least squares	$R \hat{x} = Q^T b$
Orthonormal coords	$v = \sum_i (v \cdot q_i) q_i$

C.7 Eigenvalues and diagonalization

Definition	$Ax = \lambda x, x \neq 0$
Characteristic equation	$\det(A - \lambda I) = 0$
Eigenspace	$E_\lambda = \mathbf{N}(A - \lambda I), \dim E_\lambda = n - \text{rank}(A - \lambda I)$
Checks	$\sum \lambda_i = \text{tr } A, \prod \lambda_i = \det A$
Multiplicities	$1 \leq \text{GM}(\lambda) \leq \text{AM}(\lambda)$
Diagonalization	$A = X \Lambda X^{-1}$ (needs n independent eigenvectors)
Powers	$A^k = X \Lambda^k X^{-1}$
Spectral theorem	$S^T = S \Rightarrow S = Q \Lambda Q^T$, real λ , orthonormal Q
Similarity	$B = P^{-1} A P$: same λ , tr, det, rank
Positive definite	$S^T = S$ and all $\lambda_i > 0 \iff$ all pivots $> 0 \iff$ all leading minors > 0

Matrix	A^k	A^{-1}	$A + cI$	cA	A^T
Eigenvalue	λ^k	$1/\lambda$	$\lambda + c$	$c\lambda$	λ
Eigenvector	same	same	same	same	<i>different</i>

C.8 SVD and PCA

SVD	$A = U\Sigma V^T, U^T U = I, V^T V = I$
Singular values	$\sigma_i = \sqrt{\lambda_i(A^T A)} \geq 0$, sorted decreasing
Right singular vectors	eigenvectors of $A^T A$
Left singular vectors	$u_i = Av_i / \sigma_i$ ($\sigma_i > 0$)
Rank	# nonzero σ_i
Outer form	$A = \sum_{i=1}^r \sigma_i u_i v_i^T$
Best rank k	$A_k = \sum_{i \leq k} \sigma_i u_i v_i^T; \ A - A_k\ _2 = \sigma_{k+1}$
Norms	$\ A\ _2 = \sigma_1, \ A\ _F^2 = \sum_i \sigma_i^2$
Condition number	$\kappa = \sigma_1 / \sigma_r$
Pseudoinverse	$A^+ = V\Sigma^+ U^T$
PCA: centre	$B = X - \mathbf{1}\bar{x}^T$
PCA: covariance	$S = \frac{1}{N-1} B^T B$
PCA: components	eigenvectors of S = right singular vectors of B
PCA: variance	$\lambda_i = \sigma_i^2 / (N-1)$; explained = $\sum_{i \leq k} \lambda_i / \text{tr } S$
PCA: scores	$Z = BQ_k = U\Sigma$ (first k columns)

C.9 Transformations

Linear	$T(cu + dv) = cT(u) + dT(v); T(\mathbf{0}) = \mathbf{0}$
Matrix of T	$A = [T(e_1) \cdots T(e_n)]$
Rotation (2D)	$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \det = 1$
Scaling	$\text{diag}(s_x, s_y), \det = s_x s_y$
Reflection (x -axis)	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \det = -1$
Reflection (line at θ)	$\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$
Shear (x)	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \det = 1$
Translation	not linear ; affine
Homogeneous affine	$\left[\begin{array}{c c} A & t \\ \hline \mathbf{0}^T & 1 \end{array} \right] \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} Ax + t \\ 1 \end{bmatrix}$
About a point c	$M = T_c(\text{transform}) T_{-c}$
Composition	rightmost matrix acts <i>first</i>
Change of basis	$[T]_B = P^{-1}AP$

C.10 The Invertible Matrix Theorem

For square $A \in \mathbb{R}^{n \times n}$, all equivalent:

- (1) A is invertible
- (2) $Ax = b$ has a unique solution $\forall b$
- (3) $Ax = \mathbf{0}$ only for $x = \mathbf{0}$
- (4) columns are independent
- (5) columns span $\mathbb{R}^n$
- (6) columns form a basis
- (7) rows are independent
- (8) rank $A = n$

- (9) n nonzero pivots
 (10) RREF of A is I
 (11) $\det A \neq 0$
 (12) $\mathbf{N}(A) = \{\mathbf{0}\}$
 (13) $\mathbf{C}(A) = \mathbb{R}^n$
 (14) 0 is not an eigenvalue
 (15) all $\sigma_i > 0$
 (16) product of elementary matrices

C.11 Decision table: which method?

Task	Method
Solve $Ax = b$, one right side	Gaussian elimination + back substitution
Solve for many right sides	factor $A = LU$ once, substitute repeatedly
Solve, no exact solution exists	least squares $A^T A \hat{x} = A^T b$
Find A^{-1} , $n = 2$	the 2×2 formula
Find A^{-1} , $n = 3$	adjugate, or Gauss–Jordan
Find A^{-1} , $n \geq 4$	Gauss–Jordan (not the adjugate)
Just one component x_j	Cramer’s rule
$\det A$, $n \leq 3$	Sarrus or cofactors
$\det A$, $n \geq 4$	row reduce, multiply the pivots
Test invertibility	$\det A \neq 0$, or count pivots
Test independence of n vectors in $\mathbb{R}^n$	one determinant
Test independence otherwise	columns into A , count pivots
Basis for $\mathbf{C}(A)$	pivot columns of the original A
Basis for $\mathbf{N}(A)$	special solutions from RREF
Compute A^{100}	diagonalize, $X\Lambda^{100}X^{-1}$
Long-run / steady state	dominant eigenvalue and eigenvector
Reduce data dimension	PCA (via SVD of centred data)
Best low-rank approximation	truncated SVD
Rectangular or singular matrix	SVD (always exists)

C.12 Twelve mistakes that cost marks

- $(AB)^T = A^T B^T$ ✗ — it is $B^T A^T$; likewise $(AB)^{-1} = B^{-1} A^{-1}$.
- $\det(cA) = c \det A$ ✗ — it is $c^n \det A$.
- $\det(A + B) = \det A + \det B$ ✗ — no such rule exists.
- Forgetting the transpose in $\text{adj } A = C^T$.
- Writing $\det(A - \lambda)$ instead of $\det(A - \lambda I)$.
- Reporting $x = \mathbf{0}$ as an eigenvector.
- Assuming a repeated eigenvalue means not diagonalizable — check GM against AM.
- Using RREF columns as a basis for $\mathbf{C}(A)$ instead of the original columns.
- Confusing $\mathbf{C}(A) \subseteq \mathbb{R}^m$ with $\mathbf{N}(A) \subseteq \mathbb{R}^n$.
- Cancelling A in $AB = AC$ without checking A is invertible.
- Calling translation a linear transformation.
- Composing transformations left to right — the *rightmost* matrix acts first.

C.13 Greek letters used in this course

λ	lambda	eigenvalue	Λ diagonal matrix of eigenvalues
σ	sigma	singular value	Σ diagonal matrix of singular values
θ	theta	angle	
ℓ	ell	elimination multiplier	
$\perp$	perp	orthogonal to	
$\oplus$	—	direct sum	

Index

- adjacency matrix, 147
- adjugate, 62
- affine transformation, 139
- algebraic connectivity, 147
- algebraic multiplicity, 104
- augmented matrix, 16

- basis, 74
- blocks, 27

- centred, 125
- coefficient matrix, 16
- coefficients, 4, 11
- cofactor, 58
- cofactor matrix, 62
- column picture, 14
- column space, 78
- components, 2
- consistent, 12
- constant term, 11
- coordinates, 74

- defective, 104
- determinant, 56
- diagonalizable, 108
- dimension, 75

- eigenfaces, 129
- eigenspace, 98
- eigenvalue, 98
- eigenvector, 98

- free columns, 34

- Gaussian elimination, 30
- geometric multiplicity, 104

- homogeneous, 12
- hyperplane, 12

- incidence matrix, 147
- inconsistent, 12
- inverse, 40
- invertible, 18, 40

- Jordan form, 111

- kernel, 79

- Laplacian, 147
- left singular vectors, 116

- length, 7
- linear, 11
- linear combination, 4
- linear transformation, 132
- linearly dependent, 6, 73
- linearly independent, 6, 73

- matrix, 20
- minor, 58
- multiplier, 30

- nonsingular, 18, 40
- normal equations, 92
- normalizing, 7
- null space, 79
- nullity, 79

- orthogonal, 7, 88
- orthogonal complement, 89
- orthogonal matrix, 90
- orthonormal, 90

- partial pivoting, 33
- Permanent breakdown, 33
- permutation matrix, 53
- pivot, 30, 34
- pivot columns, 34
- positive definite, 112
- positive semidefinite, 52, 112
- principal components, 124
- projection, 92
- proportion of variance explained, 125

- rank, 81
- reduced row echelon form, 34
- right singular vectors, 116
- row echelon form, 34
- row equivalent, 29
- row picture, 12

- scores, 126
- shear, 136
- similar, 106
- singular, 18, 40
- singular values, 116
- skew-symmetric, 51
- solution, 12
- solution set, 12
- span, 73
- spanning set, 73

special solution, 80

square, 18

standard basis, 9

subspace, 71

symmetric, 51

system, 12

Temporary breakdown, 33

trace, 25

transpose, 49

unit vector, 7

upper triangular, 30

vector, 2

vector space, 70

weights, 4