Orthogonality — Part 3: Least squares and orthonormal bases

Lecture notes on sections 4.3 and 4.4 of Strang, *Introduction to Linear Algebra*. Read with pen and paper beside you, work each example yourself before reading its solution, and do every exercise before opening the answer.

The chapter in three parts: Part 1 · Orthogonal vectors and subspaces | Part 2 · Projections | Part 3 · Least squares and orthonormal bases

Contents

0. Before you start
1. Least squares
2. Orthonormal bases and Gram–Schmidt
3. Question types and model answers
4. Practice test
5. Formula sheet
6. Common mistakes and self-check

0 Before you start

Why we study this

- **Fitting data.** Measurements never lie exactly on a straight line. **Least squares** chooses the line that fits best. It is the projection of Part 2 applied to data, it is the basis of linear regression, and it is one of the most used calculations in all of science.
- **A convenient basis.** When the basis vectors are perpendicular to each other and have length 1 (an **orthonormal** basis), the coefficient of each basis vector is a single dot product, so no system of equations has to be solved. The **Gram–Schmidt** process produces such a basis from any basis, and it leads to the factorization $A = QR$ that computers use for least squares.

What you need

- From Part 2: the normal equations $A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$, the projection $\mathbf{p} = A\hat{\mathbf{x}}$, the error $\mathbf{e} = \mathbf{b} - \mathbf{p}$ with $A^{\mathsf{T}}\mathbf{e} = \mathbf{0}$, and projection onto a line, $\hat{x} = \mathbf{a}^{\mathsf{T}}\mathbf{b}/\mathbf{a}^{\mathsf{T}}\mathbf{a}$.
- From Part 1: orthogonal vectors (dot product zero), lengths, and the column space.
- From Chapter 2: elimination and back substitution for two equations in two unknowns; the inverse of a matrix.

What you will be able to do

- ☐ Explain why $A\mathbf{x} = \mathbf{b}$ has no solution for noisy data, and what “least squares” minimises.
- ☐ Fit the line $b = C + Dt$ from the five sums, and report the heights, the errors and the minimum E, with checks.

- [ ] Recognise orthonormal columns ($Q^{\mathsf{T}}Q = I$) and orthogonal matrices ($Q^{-1} = Q^{\mathsf{T}}$), and project with them using dot products only.
- [ ] Run Gram–Schmidt on two or three vectors and write $A = QR$.

Notation

Symbol	Meaning
$(t_i, b_i), m$	the data points (time, measurement) and how many there are
C, D	the intercept and slope of the line $b = C + Dt$
$\sum t_i$	$t_1 + t_2 + \cdots + t_m$
p_i, e_i, E	height of the line at t_i; error $b_i - p_i$; total squared error $\sum e_i^2$
$\mathbf{q}_1, \mathbf{q}_2, \ldots; Q$	orthonormal vectors; the matrix with them as columns
$\mathbf{A}, \mathbf{B}, \mathbf{C}; R$	the orthogonal (not yet unit) vectors in Gram–Schmidt; the triangular factor in $A = QR$

1 Least squares (4.3)

Suppose we measure a quantity b at several times t and we believe that b grows along a straight line $b = C + Dt$. Because of measurement noise no line passes through all the points. Which line is *best*? This is the least-squares problem, and we will see that it is a projection in disguise.

> **Historical note.** In 1801 the newly discovered asteroid Ceres was lost behind the Sun after only forty days of observations. The 24-year-old Gauss used least squares on those noisy observations to predict where it would reappear, and it reappeared exactly there. Today the same method is called *linear regression*; it is the first method taught in statistics and machine learning, and a GPS receiver solves a least-squares problem every second.

Why there is no exact solution

With m data points (t_i, b_i) the line $b = C + Dt$ must satisfy m equations in only two unknowns:

$$C + Dt_1 = b_1, \quad C + Dt_2 = b_2, \quad \dots, \quad C + Dt_m = b_m,$$

that is $A\mathbf{x} = \mathbf{b}$ with

$$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} C \\ D \end{bmatrix}.$$

The first column is all ones because it multiplies C; the second column holds the times because they multiply D. The matrix is tall and thin, its two columns span only a plane inside $\mathbf{R}^m$, and a noisy $\mathbf{b}$ lies outside that plane. So $A\mathbf{x} = \mathbf{b}$ has no solution. We cannot make the error $\mathbf{e} = \mathbf{b} - A\mathbf{x}$ zero, but we can make it as short as possible.

> **Definition 1.1.** The **least-squares solution** $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ is the vector that makes the squared length of the error as small as possible:
>
> $$E = \|A\mathbf{x} - \mathbf{b}\|^2 = e_1^2 + e_2^2 + \cdots + e_m^2 \quad \text{is minimised at } \mathbf{x} = \hat{\mathbf{x}}.$$
>
> For a line fit, $e_i = b_i - (C + Dt_i)$ is the vertical distance from the i-th point to the line, and E is the sum of their squares.

> **Theorem 1.2.** The least-squares solution satisfies the normal equations
>
> $$A^{\mathsf{T}} A\,\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}.$$

Proof. As $\mathbf{x}$ varies, $A\mathbf{x}$ runs through every point of the column space $\mathbf{C}(A)$. Minimising $\|A\mathbf{x} - \mathbf{b}\|$ therefore means finding the point of $\mathbf{C}(A)$ closest to $\mathbf{b}$. By Part 2 (projection onto a subspace) that point is the projection $\mathbf{p} = A\hat{\mathbf{x}}$, and the error $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ is perpendicular to every column of A: $A^{\mathsf{T}}(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$, which is $A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$. ∎

Remark. Calculus gives the same equations: setting the partial derivatives $\partial E/\partial C$ and $\partial E/\partial D$ equal to zero produces exactly the two normal equations. The book shows both arguments.

Fitting a straight line

The normal equations for $b = C + Dt$.

$$A^{\mathsf{T}}A = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}, \qquad A^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}, \qquad \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

Only five numbers are needed: m, $\sum t_i$, $\sum t_i^2$, $\sum b_i$, $\sum t_i b_i$.

Method 1.1.

1. Write the m equations $C + Dt_i = b_i$, and so A and $\mathbf{b}$.
2. Compute the five sums and write $A^{\mathsf{T}}A$ and $A^{\mathsf{T}}\mathbf{b}$.
3. Solve the normal equations by elimination: this gives C and D, hence the best line.
4. Compute the heights on the line $p_i = C + Dt_i$, the errors $e_i = b_i - p_i$, and the minimum $E = \sum e_i^2$.
5. **Check:** $\sum e_i = 0$ and $\sum t_i e_i = 0$ (the error is perpendicular to both columns of A).

Example 1.1 (book). Find the closest line to the points $(0,6)$, $(1,0)$, $(2,0)$.

1. The equations $C + 0D = 6$, $C + 1D = 0$, $C + 2D = 0$ give

   $$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}.$$

   No line passes through all three points, so we look for the least-squares solution.

2. $m = 3$, $\sum t = 0 + 1 + 2 = 3$, $\sum t^2 = 0 + 1 + 4 = 5$, $\sum b = 6 + 0 + 0 = 6$, $\sum tb = 0 + 0 + 0 = 0$:

   $$A^{\mathsf{T}}A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}, \qquad A^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}.$$

3. Elimination and back substitution:

   $$\left[\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{cc|c} 3 & 3 & 6 \\ 0 & 2 & -6 \end{array}\right] \qquad \begin{aligned} 2D &= -6 && \Rightarrow\ D = -3 \\ 3C + 3(-3) &= 6 && \Rightarrow\ C = 5 \end{aligned}$$

   The best line is $b = 5 - 3t$.

4. Heights $p_i = 5 - 3t_i$: $\mathbf{p} = (5, 2, -1)$. Errors $e_i = b_i - p_i$: $\mathbf{e} = (6 - 5,\ 0 - 2,\ 0 + 1) = (1, -2, 1)$. Minimum $E = 1 + 4 + 1 = 6$.
5. Checks: $\sum e_i = 1 - 2 + 1 = 0$ and $\sum t_i e_i = 0 - 2 + 2 = 0$. ✓

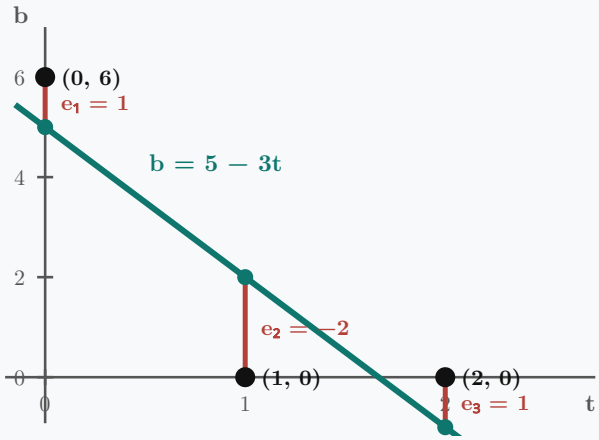


Figure 1.1 — the least-squares line of Example 1.1. Black dots: the data b_i. Teal dots: the heights $p_i = 5 - 3t_i$ on the best line. Red: the vertical errors $e_i = b_i - p_i$. No line passes through all three data points; this line makes $E = e_1^2 + e_2^2 + e_3^2 = 6$ as small as possible.

Example 1.2 (book, Problem Set 4.3). **Fit $b = C + Dt$ to $b = 0, 8, 8, 20$ at $t = 0, 1, 3, 4$.**

1. $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$, $\mathbf{b} = (0, 8, 8, 20)$.

2. $m = 4$, $\sum t = 8$, $\sum t^2 = 0 + 1 + 9 + 16 = 26$, $\sum b = 36$, $\sum tb = 0 + 8 + 24 + 80 = 112$:

$$A^{\mathsf{T}}A = \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix}, \qquad A^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 36 \\ 112 \end{bmatrix}.$$

3. The multiplier is $8 \div 4 = 2$:

$$\left[\begin{array}{cc|c} 4 & 8 & 36 \\ 8 & 26 & 112 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|c} 4 & 8 & 36 \\ 0 & 10 & 40 \end{array}\right] \qquad \begin{aligned} 10D &= 40 \quad &\Rightarrow\ D &= 4 \\ 4C + 8(4) &= 36 \quad &\Rightarrow\ C &= 1 \end{aligned}$$

   The best line is $b = 1 + 4t$.

4. Heights $\mathbf{p} = (1, 5, 13, 17)$. Errors $\mathbf{e} = (0 - 1,\ 8 - 5,\ 8 - 13,\ 20 - 17) = (-1, 3, -5, 3)$. $E = 1 + 9 + 25 + 9 = 44$.

5. Checks: $\sum e_i = -1 + 3 - 5 + 3 = 0$ and $\sum t_i e_i = 0 + 3 - 15 + 12 = 0$. ✓

Example 1.3 (a horizontal line). **Fit $b = C$ (no slope) to measurements $b_1, \dots, b_m$.**

Now A is a single column of ones, so $A^{\mathsf{T}}A = m$ and $A^{\mathsf{T}}\mathbf{b} = \sum b_i$. The normal equation $mC = \sum b_i$ gives

$$C = \frac{b_1 + b_2 + \cdots + b_m}{m},$$

the **average** of the measurements. The mean is a least-squares fit, and therefore also a projection.

Exercise 1. Find the least-squares line through $(1,2)$, $(2,0)$, $(3,4)$. Give C and D, the heights, the errors, E, and the checks.

ANSWER

1. $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, $\mathbf{b} = (2,0,4)$.
2. $m = 3$, $\sum t = 6$, $\sum t^2 = 1+4+9 = 14$, $\sum b = 6$, $\sum tb = 2+0+12 = 14$:
   $A^{\mathsf{T}}A = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$, $A^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$.
3. Multiplier $6 \div 3 = 2$: $\left[\begin{array}{cc|c} 3 & 6 & 6 \\ 6 & 14 & 14 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|c} 3 & 6 & 6 \\ 0 & 2 & 2 \end{array}\right]$, so $D = 1$ and $3C + 6(1) = 6$, $C = 0$. **Best line** $b = t$**.**
4. Heights $p_i = t_i$: $(1,2,3)$. Errors $(2-1,\ 0-2,\ 4-3) = (1,-2,1)$. $E = 1+4+1 = 6$.
5. Checks: $\sum e_i = 0$ and $\sum t_i e_i = 1 - 4 + 3 = 0$. ✓

2 Orthonormal bases and Gram–Schmidt (4.4)

The coordinate axes $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ are perpendicular to each other and have length 1. That is why coordinates are so easy to read off: the component of $\mathbf{b}$ along an axis is a single dot product. This section shows that every subspace has a basis of this kind, how to construct it, and why it makes least squares easy.

Definition 2.1. Vectors $\mathbf{q}_1, \dots, \mathbf{q}_n$ are **orthonormal** when they are perpendicular to each other and each has length 1:

$$\mathbf{q}_i^{\mathsf{T}} \mathbf{q}_j = \begin{cases} 0 & i \neq j \quad \text{(orthogonal)} \\ 1 & i = j \quad \text{(unit length)} \end{cases}$$

If Q is the matrix with the $\mathbf{q}$'s as columns, these n^2 conditions are the single matrix equation

$$Q^{\mathsf{T}} Q = I,$$

because the (i,j) entry of $Q^{\mathsf{T}}Q$ is exactly $\mathbf{q}_i^{\mathsf{T}}\mathbf{q}_j$.

Definition 2.2. A **square** matrix with orthonormal columns is called an **orthogonal matrix**. For such a matrix $Q^{\mathsf{T}}Q = I$ says that Q^{T} is the inverse:

$$Q^{-1} = Q^{\mathsf{T}}.$$

Example 2.1. **Three orthogonal matrices.**

- **Rotation** by θ: $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. The columns have dot product $-\cos\theta\sin\theta + \sin\theta\cos\theta = 0$ and each has squared length $\cos^2\theta + \sin^2\theta = 1$.
- **Permutation**: $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ swaps the two entries of a vector. Its columns are the coordinate axes in a different order.
- **A 3×3 example**: $Q = \dfrac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}$. The integer columns $(1,2,2)$, $(2,1,-2)$, $(2,-2,1)$ have dot products $2+2-4=0$, $2-4+2=0$, $4-2-2=0$, and each has squared length 9. Dividing by 3 makes them unit vectors.

Example 2.2 (book, 4.4 #1). Classify each pair as independent only, orthogonal, or orthonormal: $(1,0)$ and $(-1,1)$; $(.6,.8)$ and $(.4,-.3)$; $(\cos\theta,\sin\theta)$ and $(-\sin\theta,\cos\theta)$.

- $(1,0)\cdot(-1,1)=-1\neq 0$: independent, not orthogonal.
- $(.6,.8)\cdot(.4,-.3)=.24-.24=0$: orthogonal. Lengths: $\sqrt{.36+.64}=1$ but $\sqrt{.16+.09}=.5$. So orthogonal, not orthonormal.
- Dot product $-\cos\theta\sin\theta+\sin\theta\cos\theta=0$, both lengths 1: orthonormal.

Theorem 2.1. Multiplying by a matrix with orthonormal columns does not change lengths: $\|Q\mathbf{x}\|=\|\mathbf{x}\|$ for every $\mathbf{x}$.

Proof. $\|Q\mathbf{x}\|^2=(Q\mathbf{x})^{\mathsf{T}}(Q\mathbf{x})=\mathbf{x}^{\mathsf{T}}Q^{\mathsf{T}}Q\mathbf{x}=\mathbf{x}^{\mathsf{T}}I\mathbf{x}=\|\mathbf{x}\|^2$. ∎

Rotations and reflections are orthogonal matrices; this theorem is why computer graphics is full of them.

Caution. $Q^{\mathsf{T}}Q=I$ holds for every matrix with orthonormal columns, but $QQ^{\mathsf{T}}=I$ only when Q is square. For example $Q=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}$ has $Q^{\mathsf{T}}Q=I_2$ but $QQ^{\mathsf{T}}\neq I_3$.

Least squares with orthonormal columns

If the columns of $A=Q$ are orthonormal, the normal equations $Q^{\mathsf{T}}Q\hat{\mathbf{x}}=Q^{\mathsf{T}}\mathbf{b}$ collapse because $Q^{\mathsf{T}}Q=I$:

Projection with orthonormal columns.

$$\hat{\mathbf{x}}=Q^{\mathsf{T}}\mathbf{b} \qquad (\hat{x}_i=\mathbf{q}_i^{\mathsf{T}}\mathbf{b})$$
$$\mathbf{p}=QQ^{\mathsf{T}}\mathbf{b}=(\mathbf{q}_1^{\mathsf{T}}\mathbf{b})\,\mathbf{q}_1+(\mathbf{q}_2^{\mathsf{T}}\mathbf{b})\,\mathbf{q}_2+\cdots$$

There is nothing to solve: every coefficient is a single dot product.

Example 2.3. Project $\mathbf{b}=(3,0,3)$ onto the plane spanned by $\mathbf{q}_1=\frac{1}{3}(1,2,2)$ and $\mathbf{q}_2=\frac{1}{3}(2,1,-2)$.

$$\hat{x}_1=\mathbf{q}_1^{\mathsf{T}}\mathbf{b}=\tfrac{1}{3}(3+0+6)=3, \qquad \hat{x}_2=\mathbf{q}_2^{\mathsf{T}}\mathbf{b}=\tfrac{1}{3}(6+0-6)=0$$
$$\mathbf{p}=3\mathbf{q}_1+0\mathbf{q}_2=(1,2,2), \qquad \mathbf{e}=\mathbf{b}-\mathbf{p}=(2,-2,1)$$
$$\text{checks: } \mathbf{q}_1^{\mathsf{T}}\mathbf{e}=\tfrac{1}{3}(2-4+2)=0, \qquad \mathbf{q}_2^{\mathsf{T}}\mathbf{e}=\tfrac{1}{3}(4-2-2)=0 \quad ✓$$

Gram–Schmidt: making a basis orthonormal

Start with independent vectors $\mathbf{a},\mathbf{b},\mathbf{c}$. We want orthonormal vectors $\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3$ that span the same space. The idea is one projection at a time: keep $\mathbf{a}$; from $\mathbf{b}$ subtract its projection onto $\mathbf{a}$, because what is

left over is an error vector and error vectors are perpendicular; from $\mathbf{c}$ subtract its projections onto the first two; finally divide each vector by its length.

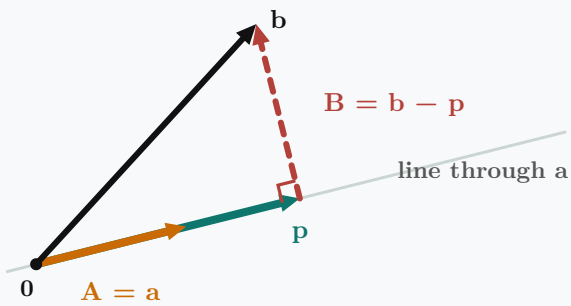


Figure 2.1 — one Gram–Schmidt step is one projection. Project $\mathbf{b}$ onto the line through $\mathbf{A} = \mathbf{a}$ to get $\mathbf{p}$. What is left, $\mathbf{B} = \mathbf{b} - \mathbf{p}$, is the error of that projection, so it is automatically orthogonal to $\mathbf{A}$. Dividing $\mathbf{A}$ and $\mathbf{B}$ by their lengths gives $\mathbf{q}_1$ and $\mathbf{q}_2$.

Method 2.1 (Gram–Schmidt).

1. $\mathbf{A} = \mathbf{a}$.
2. $\mathbf{B} = \mathbf{b} - \dfrac{\mathbf{A}^{\mathsf{T}}\mathbf{b}}{\mathbf{A}^{\mathsf{T}}\mathbf{A}}\,\mathbf{A}$. Check $\mathbf{A}^{\mathsf{T}}\mathbf{B} = 0$.
3. $\mathbf{C} = \mathbf{c} - \dfrac{\mathbf{A}^{\mathsf{T}}\mathbf{c}}{\mathbf{A}^{\mathsf{T}}\mathbf{A}}\,\mathbf{A} - \dfrac{\mathbf{B}^{\mathsf{T}}\mathbf{c}}{\mathbf{B}^{\mathsf{T}}\mathbf{B}}\,\mathbf{B}$. Check $\mathbf{A}^{\mathsf{T}}\mathbf{C} = 0$ and $\mathbf{B}^{\mathsf{T}}\mathbf{C} = 0$.
4. $\mathbf{q}_1 = \dfrac{\mathbf{A}}{\|\mathbf{A}\|}$, $\mathbf{q}_2 = \dfrac{\mathbf{B}}{\|\mathbf{B}\|}$, $\mathbf{q}_3 = \dfrac{\mathbf{C}}{\|\mathbf{C}\|}$.

Example 2.4 (book). Apply Gram–Schmidt to $\mathbf{a} = (1,-1,0)$, $\mathbf{b} = (2,0,-2)$, $\mathbf{c} = (3,-3,3)$.

1. $\mathbf{A} = (1,-1,0)$, with $\mathbf{A}^{\mathsf{T}}\mathbf{A} = 2$.
2. $\mathbf{A}^{\mathsf{T}}\mathbf{b} = 2 + 0 + 0 = 2$, so
$$\mathbf{B} = \mathbf{b} - \frac{2}{2}\mathbf{A} = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$
Check: $\mathbf{A}^{\mathsf{T}}\mathbf{B} = 1 - 1 + 0 = 0$. ✓ Also $\mathbf{B}^{\mathsf{T}}\mathbf{B} = 1 + 1 + 4 = 6$.
3. $\mathbf{A}^{\mathsf{T}}\mathbf{c} = 3 + 3 + 0 = 6$ and $\mathbf{B}^{\mathsf{T}}\mathbf{c} = 3 - 3 - 6 = -6$, so
$$\mathbf{C} = \mathbf{c} - \frac{6}{2}\mathbf{A} - \frac{-6}{6}\mathbf{B} = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$
Checks: $\mathbf{A}^{\mathsf{T}}\mathbf{C} = 1 - 1 + 0 = 0$ and $\mathbf{B}^{\mathsf{T}}\mathbf{C} = 1 + 1 - 2 = 0$. ✓
4. Lengths $\|\mathbf{A}\| = \sqrt{2}$, $\|\mathbf{B}\| = \sqrt{6}$, $\|\mathbf{C}\| = \sqrt{3}$:
$$\mathbf{q}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \qquad \mathbf{q}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \qquad \mathbf{q}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Example 2.5 (book, 4.4 #17). Apply Gram–Schmidt to $\mathbf{a} = (1,1,1)$ and $\mathbf{b} = (1,3,5)$.

1. $\mathbf{A} = (1,1,1)$, $\mathbf{A}^{\mathsf{T}}\mathbf{A} = 3$.
2. $\mathbf{A}^{\mathsf{T}}\mathbf{b} = 1 + 3 + 5 = 9$, so $\mathbf{B} = \mathbf{b} - \frac{9}{3}\mathbf{A} = (1,3,5) - (3,3,3) = (-2,0,2)$. Check: $\mathbf{A}^{\mathsf{T}}\mathbf{B} = -2 + 0 + 2 = 0$. ✓
3. $\|\mathbf{A}\| = \sqrt{3}$ and $\|\mathbf{B}\| = \sqrt{8} = 2\sqrt{2}$:
$$\mathbf{q}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{q}_2 = \frac{1}{2\sqrt{2}}\begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

The factorization $A = QR$

The original vectors are combinations of the $\mathbf{q}$'s, and each one uses only the *earlier* $\mathbf{q}$'s: $\mathbf{a}$ is a multiple of $\mathbf{q}_1$; $\mathbf{b}$ is a combination of $\mathbf{q}_1, \mathbf{q}_2$; $\mathbf{c}$ uses all three. So the matrix connecting them is upper triangular.

$A = QR$. With $A = [\,\mathbf{a}\ \ \mathbf{b}\ \ \mathbf{c}\,]$ and $Q = [\,\mathbf{q}_1\ \ \mathbf{q}_2\ \ \mathbf{q}_3\,]$,

$$A = QR, \qquad R = Q^{\mathsf{T}}A = \begin{bmatrix} \mathbf{q}_1^{\mathsf{T}}\mathbf{a} & \mathbf{q}_1^{\mathsf{T}}\mathbf{b} & \mathbf{q}_1^{\mathsf{T}}\mathbf{c} \\ 0 & \mathbf{q}_2^{\mathsf{T}}\mathbf{b} & \mathbf{q}_2^{\mathsf{T}}\mathbf{c} \\ 0 & 0 & \mathbf{q}_3^{\mathsf{T}}\mathbf{c} \end{bmatrix}.$$

The diagonal of R holds the lengths $\|\mathbf{A}\|, \|\mathbf{B}\|, \|\mathbf{C}\|$.

Example 2.6. Write $A = QR$ for the vectors of Example 2.4.

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{bmatrix}$$

The entries of R are dot products, for example $\mathbf{q}_1^{\mathsf{T}}\mathbf{b} = \dfrac{2+0+0}{\sqrt{2}} = \sqrt{2}$, $\mathbf{q}_1^{\mathsf{T}}\mathbf{c} = \dfrac{3+3+0}{\sqrt{2}} = \dfrac{6}{\sqrt{2}} = \sqrt{18}$, $\mathbf{q}_2^{\mathsf{T}}\mathbf{c} = \dfrac{3-3-6}{\sqrt{6}} = -\sqrt{6}$.

Remark (why computers use QR). With $A = QR$ the normal equations become $R^{\mathsf{T}}R\hat{\mathbf{x}} = R^{\mathsf{T}}Q^{\mathsf{T}}\mathbf{b}$, that is $R\hat{\mathbf{x}} = Q^{\mathsf{T}}\mathbf{b}$: a triangular system solved by back substitution, with much smaller rounding errors than forming $A^{\mathsf{T}}A$. Every serious least-squares program works this way.

Exercise 2.

a. Is $\dfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ an orthogonal matrix? If so, write its inverse.

b. Apply Gram–Schmidt to $\mathbf{a} = (1,1,0)$, $\mathbf{b} = (2,0,2)$, and write R.

ANSWER

a. The columns are $\frac{1}{\sqrt{2}}(1,1)$ and $\frac{1}{\sqrt{2}}(-1,1)$. Dot product $\frac{1}{2}(-1+1) = 0$; squared lengths $\frac{1}{2}(1+1) = 1$ each. Orthonormal columns in a square matrix, so yes, it is an orthogonal matrix and $Q^{-1} = Q^{\mathsf{T}} = \dfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.

b. $\mathbf{A} = (1,1,0)$, $\mathbf{A}^{\mathsf{T}}\mathbf{A} = 2$, $\mathbf{A}^{\mathsf{T}}\mathbf{b} = 2+0+0 = 2$, so $\mathbf{B} = \mathbf{b} - \frac{2}{2}\mathbf{A} = (2,0,2) - (1,1,0) = (1,-1,2)$; check $\mathbf{A}^{\mathsf{T}}\mathbf{B} = 1-1+0 = 0$. ✓ Lengths $\|\mathbf{A}\| = \sqrt{2}$, $\|\mathbf{B}\| = \sqrt{1+1+4} = \sqrt{6}$, so $\mathbf{q}_1 = \frac{1}{\sqrt{2}}(1,1,0)$ and $\mathbf{q}_2 = \frac{1}{\sqrt{6}}(1,-1,2)$. Then $R = \begin{bmatrix} \mathbf{q}_1^{\mathsf{T}}\mathbf{a} & \mathbf{q}_1^{\mathsf{T}}\mathbf{b} \\ 0 & \mathbf{q}_2^{\mathsf{T}}\mathbf{b} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2/\sqrt{2} \\ 0 & 6/\sqrt{6} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{6} \end{bmatrix}$.

3 Question types and model answers

Almost every question on this part is one of the four types below. Each model answer is written out the way I expect to see it on paper: the formula first, then the numbers, the arithmetic, the check, and a one-line conclusion.

Type 1 — Least-squares line

Question. **Find the least-squares line $b = C + Dt$ through $(0,6), (1,0), (2,0)$. Give the heights, the errors and E.**

$$C + 0D = 6, \quad C + 1D = 0, \quad C + 2D = 0 \qquad A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

$$m = 3, \quad \sum t = 3, \quad \sum t^2 = 5, \quad \sum b = 6, \quad \sum tb = 0$$

$$A^{\mathsf{T}}A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}, \qquad A^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{cc|c} 3 & 3 & 6 \\ 0 & 2 & -6 \end{array}\right]$$

$$2D = -6 \;\Rightarrow\; D = -3, \qquad 3C + 3(-3) = 6 \;\Rightarrow\; C = 5$$

$$\text{best line: } b = 5 - 3t$$

$$\mathbf{p} = (5, 2, -1), \qquad \mathbf{e} = \mathbf{b} - \mathbf{p} = (1, -2, 1), \qquad E = 1 + 4 + 1 = 6$$

$$\text{checks: } \sum e_i = 1 - 2 + 1 = 0, \quad \sum t_i e_i = 0 - 2 + 2 = 0 \;\checkmark$$

Type 2 — Orthogonal matrix

Question. **Is $Q = \frac{1}{5}\begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$ an orthogonal matrix? If so, write Q^{-1}.**

$$\text{columns } \mathbf{q}_1 = \tfrac{1}{5}(3,4),\ \mathbf{q}_2 = \tfrac{1}{5}(-4,3)$$

$$\mathbf{q}_1^{\mathsf{T}}\mathbf{q}_2 = \tfrac{1}{25}(-12 + 12) = 0$$

$$\|\mathbf{q}_1\|^2 = \tfrac{9+16}{25} = 1, \qquad \|\mathbf{q}_2\|^2 = \tfrac{16+9}{25} = 1$$

$$\text{orthonormal columns, and } Q \text{ is square: } Q \text{ is an orthogonal matrix}$$

$$Q^{-1} = Q^{\mathsf{T}} = \frac{1}{5}\begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}$$

Type 3 — Projection with orthonormal vectors

Question. **Project $\mathbf{b} = (3, 0, 3)$ onto the plane spanned by the orthonormal vectors $\mathbf{q}_1 = \frac{1}{3}(1, 2, 2)$ and $\mathbf{q}_2 = \frac{1}{3}(2, 1, -2)$.**

$$\hat{x}_1 = \mathbf{q}_1^{\mathsf{T}}\mathbf{b} = \tfrac{1}{3}(3 + 0 + 6) = 3$$
$$\hat{x}_2 = \mathbf{q}_2^{\mathsf{T}}\mathbf{b} = \tfrac{1}{3}(6 + 0 - 6) = 0$$
$$\mathbf{p} = \hat{x}_1\mathbf{q}_1 + \hat{x}_2\mathbf{q}_2 = 3 \cdot \tfrac{1}{3}(1, 2, 2) + 0 = (1, 2, 2)$$
$$\mathbf{e} = \mathbf{b} - \mathbf{p} = (2, -2, 1)$$
$$\text{checks: } \mathbf{q}_1^{\mathsf{T}}\mathbf{e} = \tfrac{1}{3}(2 - 4 + 2) = 0, \qquad \mathbf{q}_2^{\mathsf{T}}\mathbf{e} = \tfrac{1}{3}(4 - 2 - 2) = 0 \ \checkmark$$

Type 4 — Gram–Schmidt and $A = QR$

Question. **Apply Gram–Schmidt to $\mathbf{a} = (1, 1, 1)$, $\mathbf{b} = (1, 3, 5)$ and write $A = QR$.**

$$\mathbf{A} = \mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{A}^{\mathsf{T}}\mathbf{b} = 1 + 3 + 5 = 9, \qquad \mathbf{A}^{\mathsf{T}}\mathbf{A} = 3$$

$$\mathbf{B} = \mathbf{b} - \frac{9}{3}\mathbf{A} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}, \qquad \text{check: } \mathbf{A}^{\mathsf{T}}\mathbf{B} = -2 + 0 + 2 = 0 \ \checkmark$$

$$\|\mathbf{A}\| = \sqrt{3}, \qquad \|\mathbf{B}\| = \sqrt{8} = 2\sqrt{2}, \qquad \mathbf{q}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{q}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$R = \begin{bmatrix} \mathbf{q}_1^{\mathsf{T}}\mathbf{a} & \mathbf{q}_1^{\mathsf{T}}\mathbf{b} \\ 0 & \mathbf{q}_2^{\mathsf{T}}\mathbf{b} \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 9/\sqrt{3} \\ 0 & 4/\sqrt{2} \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 3\sqrt{3} \\ 0 & 2\sqrt{2} \end{bmatrix}, \qquad A = QR$$

4 Practice test

Six questions, about sixty minutes of work. Write full answers on paper first and open each answer only afterwards. Question 6 is theory.

Q1. Find the least-squares line through $(0,2)$, $(1,0)$, $(2,4)$. Give C, D, the heights, the errors and E.

ANSWER

$m=3$, $\sum t = 0+1+2 = 3$, $\sum t^2 = 0+1+4 = 5$, $\sum b = 2+0+4 = 6$, $\sum tb = 0+0+8 = 8$.

$\left[\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 8 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{cc|c} 3 & 3 & 6 \\ 0 & 2 & 2 \end{array}\right]$, so $D = 1$ and $3C + 3(1) = 6$, $C = 1$. **Best line** $b = 1 + t$**.**

Heights $p_i = 1 + t_i$: $(1, 2, 3)$. Errors $e_i = b_i - p_i$: $(1, -2, 1)$. $E = 1 + 4 + 1 = 6$.

Checks: $\sum e_i = 0$ and $\sum t_i e_i = 0 - 2 + 2 = 0$. ✓

Q2. Fit $b = C + Dt$ to $b = 1, 3, 4, 4$ at $t = 0, 1, 2, 3$. Give the line, the heights, the errors and E.

ANSWER

$m = 4$, $\sum t = 6$, $\sum t^2 = 0+1+4+9 = 14$, $\sum b = 12$, $\sum tb = 0+3+8+12 = 23$.

Multiplier $6 \div 4 = 1.5$: $\left[\begin{array}{cc|c} 4 & 6 & 12 \\ 6 & 14 & 23 \end{array}\right] \xrightarrow{R_2 - 1.5R_1} \left[\begin{array}{cc|c} 4 & 6 & 12 \\ 0 & 5 & 5 \end{array}\right]$, so $D = 1$ and $4C + 6(1) = 12$, $C = 1.5$. **Best line** $b = 1.5 + t$**.**

Heights $(1.5, 2.5, 3.5, 4.5)$. Errors $(1 - 1.5,\ 3 - 2.5,\ 4 - 3.5,\ 4 - 4.5) = (-0.5, 0.5, 0.5, -0.5)$. $E = 4 \times 0.25 = 1$.

Checks: $\sum e_i = 0$ and $\sum t_i e_i = 0 + 0.5 + 1 - 1.5 = 0$. ✓

Q3. Is $Q = \dfrac{1}{5}\begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$ an orthogonal matrix? Write Q^{-1} without elimination.

ANSWER

Columns $\mathbf{q}_1 = \frac{1}{5}(3, 4)$, $\mathbf{q}_2 = \frac{1}{5}(-4, 3)$. Dot product $\frac{1}{25}(-12 + 12) = 0$; squared lengths $\frac{9+16}{25} = 1$ and $\frac{16+9}{25} = 1$. Orthonormal columns in a square matrix, so Q is an orthogonal matrix and $Q^{-1} = Q^{\mathsf{T}} = \dfrac{1}{5}\begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}$. (Check: $Q^{\mathsf{T}}Q = \frac{1}{25}\begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix} = I$.)

Q4. Project $\mathbf{b} = (1, 1, 1)$ onto the plane spanned by the orthonormal vectors $\mathbf{q}_1 = \frac{1}{3}(1, 2, 2)$ and $\mathbf{q}_2 = \frac{1}{3}(2, -2, 1)$. Give $\mathbf{p}$, $\mathbf{e}$ and the checks.

ANSWER

$\hat{x}_1 = \mathbf{q}_1^{\mathsf{T}}\mathbf{b} = \frac{1}{3}(1 + 2 + 2) = \frac{5}{3}$, $\hat{x}_2 = \mathbf{q}_2^{\mathsf{T}}\mathbf{b} = \frac{1}{3}(2 - 2 + 1) = \frac{1}{3}$.
$\mathbf{p} = \frac{5}{3}\mathbf{q}_1 + \frac{1}{3}\mathbf{q}_2 = \frac{5}{9}(1, 2, 2) + \frac{1}{9}(2, -2, 1) = \left(\frac{7}{9}, \frac{8}{9}, \frac{11}{9}\right)$.
$\mathbf{e} = \mathbf{b} - \mathbf{p} = \left(\frac{2}{9}, \frac{1}{9}, -\frac{2}{9}\right)$.
Checks: $\mathbf{q}_1^{\mathsf{T}}\mathbf{e} = \frac{1}{27}(2 + 2 - 4) = 0$ and $\mathbf{q}_2^{\mathsf{T}}\mathbf{e} = \frac{1}{27}(4 - 2 - 2) = 0$. ✓

Q5. Apply Gram–Schmidt to $\mathbf{a} = (1, 1, 0)$, $\mathbf{b} = (2, 0, 2)$: find $\mathbf{q}_1$, $\mathbf{q}_2$ and R.

ANSWER

$\mathbf{A} = \mathbf{a} = (1, 1, 0)$, $\mathbf{A}^{\mathsf{T}}\mathbf{A} = 2$. $\mathbf{A}^{\mathsf{T}}\mathbf{b} = 2 + 0 + 0 = 2$, so
$\mathbf{B} = \mathbf{b} - \frac{2}{2}\mathbf{A} = (2, 0, 2) - (1, 1, 0) = (1, -1, 2)$; check $\mathbf{A}^{\mathsf{T}}\mathbf{B} = 1 - 1 + 0 = 0$. ✓
Lengths $\|\mathbf{A}\| = \sqrt{2}$, $\|\mathbf{B}\| = \sqrt{1 + 1 + 4} = \sqrt{6}$, so $\mathbf{q}_1 = \frac{1}{\sqrt{2}}(1, 1, 0)$, $\mathbf{q}_2 = \frac{1}{\sqrt{6}}(1, -1, 2)$.

$$R = \begin{bmatrix} \mathbf{q}_1^{\mathsf{T}}\mathbf{a} & \mathbf{q}_1^{\mathsf{T}}\mathbf{b} \\ 0 & \mathbf{q}_2^{\mathsf{T}}\mathbf{b} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2/\sqrt{2} \\ 0 & 6/\sqrt{6} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{6} \end{bmatrix}.$$

Q6 (theory). (a) Explain why the least-squares solution satisfies $A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$. (b) Prove that $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ when Q has orthonormal columns. (c) Why do the normal equations become $\hat{\mathbf{x}} = Q^{\mathsf{T}}\mathbf{b}$ when $A = Q$?

ANSWER

(a) As $\mathbf{x}$ varies, $A\mathbf{x}$ runs through the column space, so minimising $\|A\mathbf{x} - \mathbf{b}\|$ means finding the point of $\mathbf{C}(A)$ closest to $\mathbf{b}$. That point is the projection $\mathbf{p} = A\hat{\mathbf{x}}$, whose error $\mathbf{b} - A\hat{\mathbf{x}}$ is perpendicular to every column: $A^{\mathsf{T}}(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$, i.e. $A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$.
(b) $\|Q\mathbf{x}\|^2 = (Q\mathbf{x})^{\mathsf{T}}(Q\mathbf{x}) = \mathbf{x}^{\mathsf{T}}Q^{\mathsf{T}}Q\mathbf{x} = \mathbf{x}^{\mathsf{T}}I\mathbf{x} = \|\mathbf{x}\|^2$. ■
(c) With $A = Q$ the normal equations read $Q^{\mathsf{T}}Q\hat{\mathbf{x}} = Q^{\mathsf{T}}\mathbf{b}$, and $Q^{\mathsf{T}}Q = I$ because the columns are orthonormal, so $\hat{\mathbf{x}} = Q^{\mathsf{T}}\mathbf{b}$: each coefficient is one dot product $\mathbf{q}_i^{\mathsf{T}}\mathbf{b}$.

5 Formula sheet

Topic	Formulas
Least squares	$\hat{\mathbf{x}}$ minimises $E = \|A\mathbf{x} - \mathbf{b}\|^2 = e_1^2 + \cdots + e_m^2$ solve the normal equations $A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$
Best line $b = C + Dt$	$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}$, $A^{\mathsf{T}}A = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}$, $A^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}$ $p_i = C + Dt_i$, $e_i = b_i - p_i$, $E = \sum e_i^2$ checks $\sum e_i = 0$, $\sum t_i e_i = 0$
Orthonormal columns	$\mathbf{q}_i^{\mathsf{T}}\mathbf{q}_j = 0$ for $i \neq j$, $\mathbf{q}_i^{\mathsf{T}}\mathbf{q}_i = 1$; $Q^{\mathsf{T}}Q = I$ square Q (orthogonal matrix): $Q^{-1} = Q^{\mathsf{T}}$; $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ least squares: $\hat{\mathbf{x}} = Q^{\mathsf{T}}\mathbf{b}$, $\mathbf{p} = QQ^{\mathsf{T}}\mathbf{b} = (\mathbf{q}_1^{\mathsf{T}}\mathbf{b})\mathbf{q}_1 + (\mathbf{q}_2^{\mathsf{T}}\mathbf{b})\mathbf{q}_2 + \cdots$
Gram–Schmidt	$\mathbf{A} = \mathbf{a}$ $\mathbf{B} = \mathbf{b} - \dfrac{\mathbf{A}^{\mathsf{T}}\mathbf{b}}{\mathbf{A}^{\mathsf{T}}\mathbf{A}}\mathbf{A}$ $\mathbf{C} = \mathbf{c} - \dfrac{\mathbf{A}^{\mathsf{T}}\mathbf{c}}{\mathbf{A}^{\mathsf{T}}\mathbf{A}}\mathbf{A} - \dfrac{\mathbf{B}^{\mathsf{T}}\mathbf{c}}{\mathbf{B}^{\mathsf{T}}\mathbf{B}}\mathbf{B}$ $\mathbf{q}_1 = \mathbf{A}/\|\mathbf{A}\|$, $\mathbf{q}_2 = \mathbf{B}/\|\mathbf{B}\|$, $\mathbf{q}_3 = \mathbf{C}/\|\mathbf{C}\|$
$A = QR$	$R = Q^{\mathsf{T}}A$, upper triangular, diagonal entries $\|\mathbf{A}\|, \|\mathbf{B}\|, \|\mathbf{C}\|$ least squares via $R\hat{\mathbf{x}} = Q^{\mathsf{T}}\mathbf{b}$

6 Common mistakes and self-check

- Trying to solve $A\mathbf{x} = \mathbf{b}$ directly when it has no solution. Multiply both sides by A^{T} first.
- Putting the t's in the first column of A. The first column is all ones (it multiplies C); the second holds the t's (it multiplies D).
- A slip in one of the five sums. Write them out term by term, and use the check $\sum e_i = 0$ at the end.
- Calling vectors “orthonormal” when they are only orthogonal. Check the lengths as well.
- Stopping Gram–Schmidt before dividing by the lengths, or dividing before checking $\mathbf{A}^{\mathsf{T}}\mathbf{B} = 0$.
- Claiming $Q^{-1} = Q^{\mathsf{T}}$ for a rectangular Q. That needs a square Q; for a rectangular one only $Q^{\mathsf{T}}Q = I$ holds.
- Writing R with the lengths in the wrong places. The diagonal of R holds $\|\mathbf{A}\|, \|\mathbf{B}\|, \ldots$; the entries above it are dot products $\mathbf{q}_i^{\mathsf{T}}(\text{later column})$.

Can you now…

- ☐ fit a least-squares line from the five sums and report the heights, the errors and E, with both checks?
- ☐ explain why the least-squares solution satisfies the normal equations?
- ☐ tell orthogonal from orthonormal, invert an orthogonal matrix by transposing, and project with $\hat{x}_i = \mathbf{q}_i^{\mathsf{T}}\mathbf{b}$?
- ☐ run Gram–Schmidt on two or three vectors and write $A = QR$?

Every one of these is a numbered question on the **practice page**, with new numbers each time and a worked solution. Practise until every chip shows a tick, then take the quiz.