LECTURE 3 · LINEAR ALGEBRA III · SEPTEMBER 16

The Closest Reachable Vector

From Exact Solutions to Best Approximations

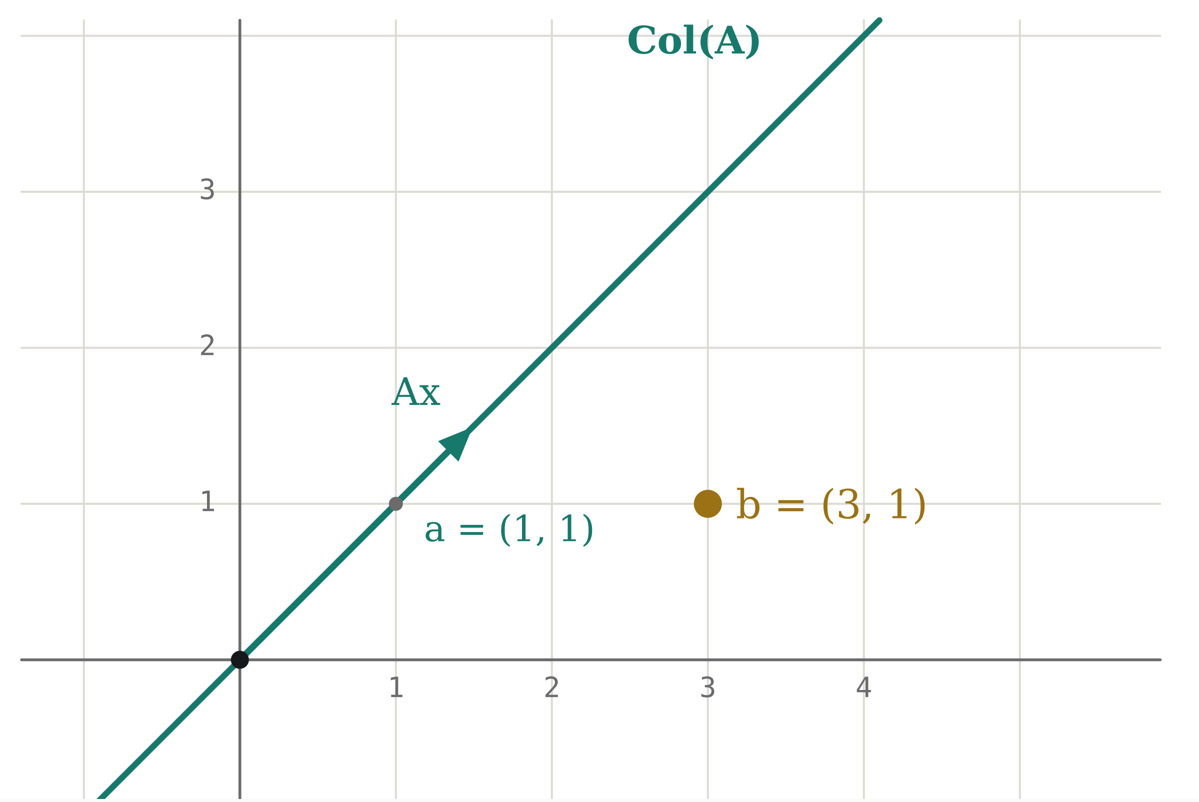
$$Ax = x_1 a_1 + \cdots + x_n a_n$$

$A \in \mathbb{R}^{m\times n}$ choose $x \in \mathbb{R}^n$ — parameter space

$a_j, Ax, b \in \mathbb{R}^m$ compare here — output space

Exact solution iff $b \in \mathrm{Col}(A)$.

> What if the target is outside the reachable space?



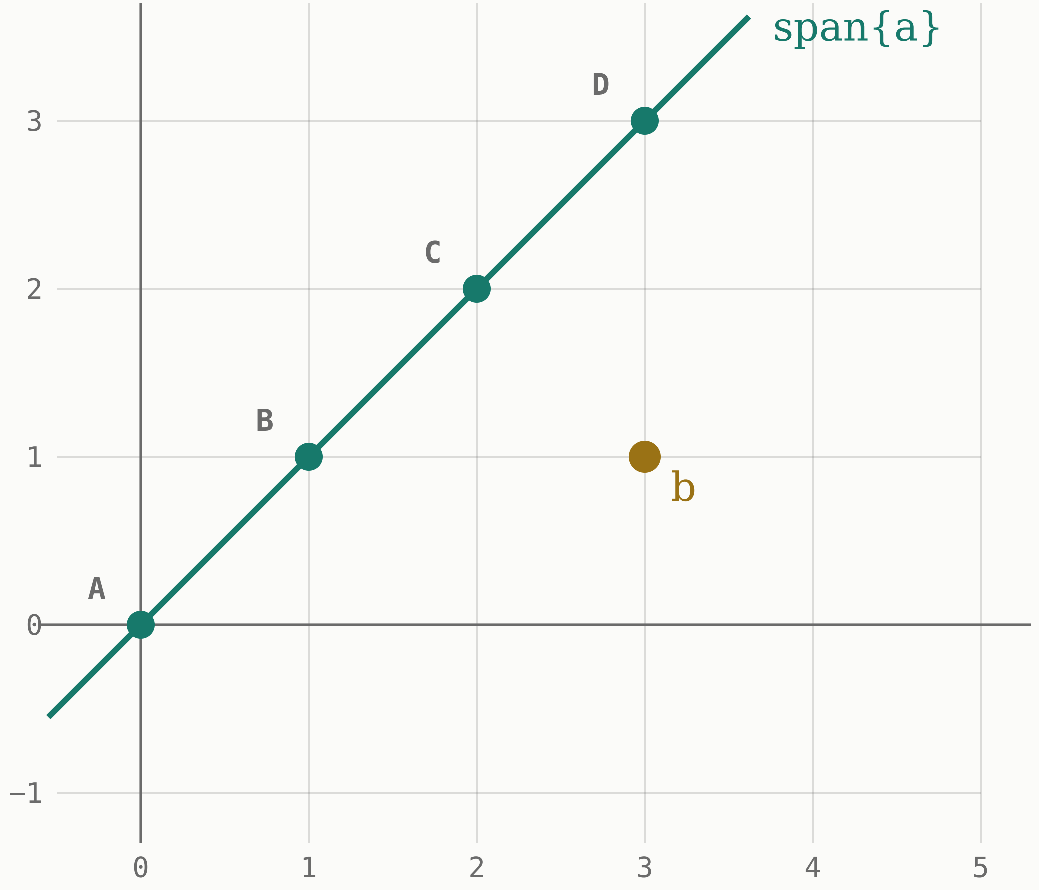
VOTE

The Closest Reachable Vector

Which Reachable Vector Would You Choose?

Which point on the line is closest to b?

- (A) $(0, 0)$
- (B) $(1, 1)$
- (C) $(2, 2)$
- (D) $(3, 3)$



DEFINITION

The Closest Reachable Vector

Difference, Norm, Squared Norm

So far: *representable? unique?* — *closest* needs a distance.

$$r = u - v$$

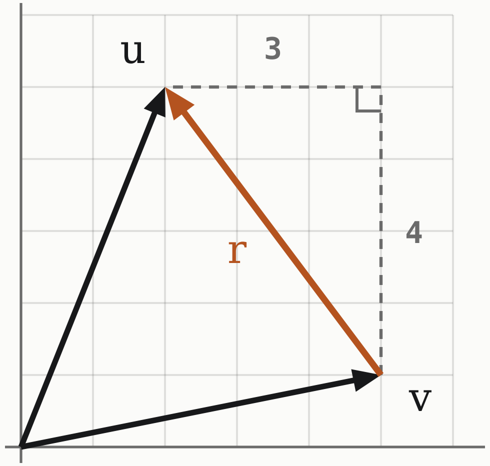
$$\|r\|_2^2 = r^\mathsf{T} r = \sum_{i=1}^{m} r_i^2$$

$$d(u,v) = \|u - v\|_2 = \sqrt{r^\mathsf{T} r}$$

$u, v, r \in \mathbb{R}^m$ r = difference vector

$\|r\|_2$ a number ≥ 0; $= 0$ iff $r = 0$

$\|r\|_2^2$ a number too, no square root



r = (−3, 4)

$\|r\|^2 = (-3)^2 + 4^2 = 25$

$\|r\| = \sqrt{25} = 5$

PREDICT

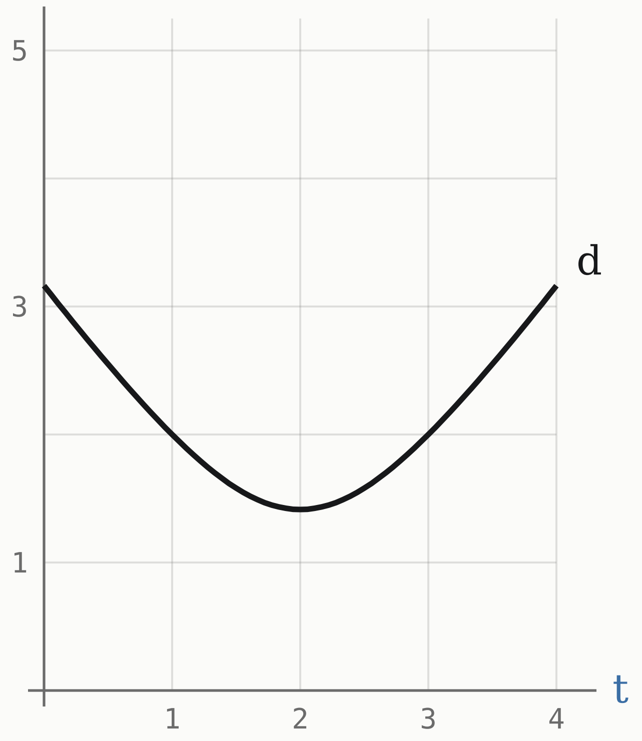
The Closest Reachable Vector

Does Squaring Move the Best Choice?

Candidates $y = (t, t)$ on the same line, $d(t) = \|b - y\|_2$.

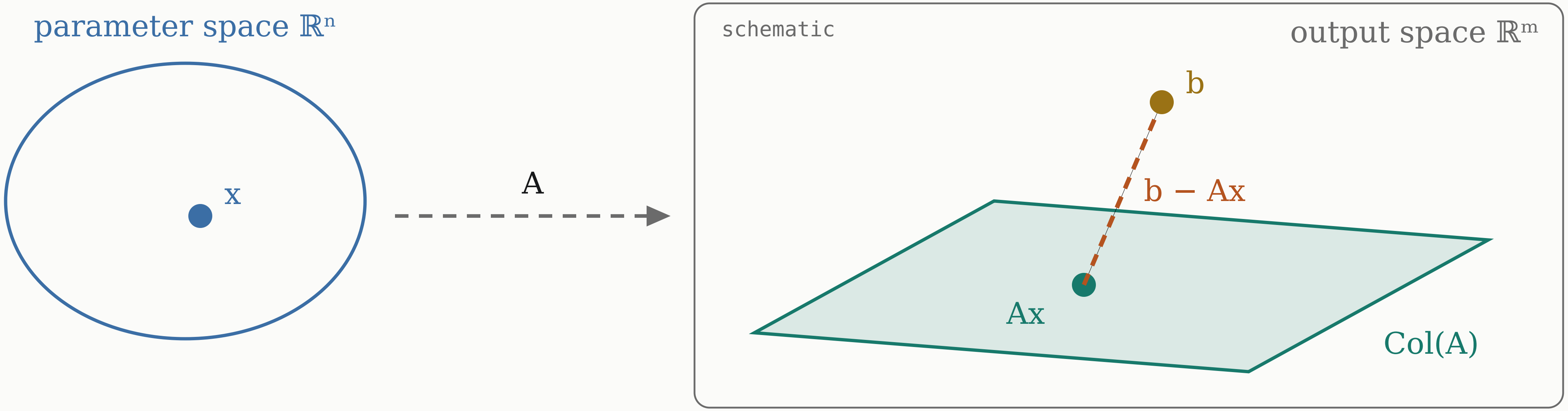
We minimise d^2 instead of d. Does the best t move?

- (A) yes
- (B) no



The Closest Reachable Vector

Today's Roadmap



> Find the best output first, then find parameters that produce it; the error is always measured in output space.

1. Judge orthogonality
2. Compute a projection
3. Set up least squares
4. Gram–Schmidt and QR
5. A robot step

DEFINITION

Orthogonality and Orthonormal Bases

The Euclidean Inner Product

$$u^{\mathsf{T}} v = \sum_{i=1}^{m} u_i v_i$$

$u, v \in \mathbb{R}^m$ real; $u^{\mathsf{T}} v$ is a number, not a vector

Lecture 2 wrote $\langle u, v \rangle$

$$u^{\mathsf{T}} v = v^{\mathsf{T}} u$$

$$(c_1 u + c_2 u')^{\mathsf{T}} v = c_1 u^{\mathsf{T}} v + c_2 u'^{\mathsf{T}} v$$

symmetric each $u_i v_i = v_i u_i$

linear $u' \in \mathbb{R}^m$ any other vector, c_1, c_2 any real scalars; slot 2 by symmetry

$$\begin{array}{ccccc} u & & v & & u_i v_i \\ \begin{bmatrix} 1 \\ 2 \end{bmatrix} & \begin{matrix} \times \\ \times \end{matrix} & \begin{bmatrix} 3 \\ -1 \end{bmatrix} & \begin{matrix} = \\ = \end{matrix} & \begin{matrix} 3 \\ -2 \end{matrix} \\ \hline & & & & u^{\mathsf{T}} v = 1 \end{array}$$

Orthogonality and Orthonormal Bases

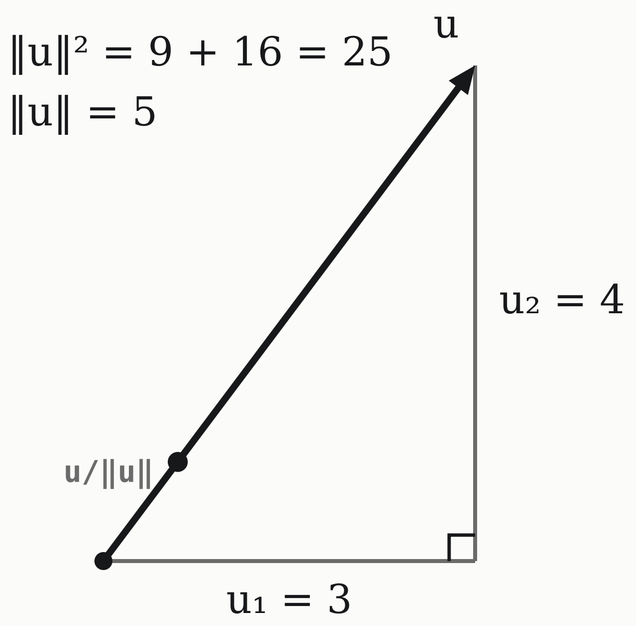
From Inner Product to Length

$$\|u\|^2 = u^\mathsf{T} u = \sum_{i=1}^{m} u_i^2 \;\geq 0$$

$= 0$ iff $u = 0$	each term is a square, so 0 forces every $u_i = 0$
same norm	the distance $\|r\|_2$, with r for u

$$\|u\| = \sqrt{u^\mathsf{T} u}, \qquad \left\| \frac{u}{\|u\|} \right\| = 1$$

$\|u\|$	the Euclidean length, a number
unit vector	$\|u\| = 1$; dividing by $\|u\|$ needs $u \neq 0$



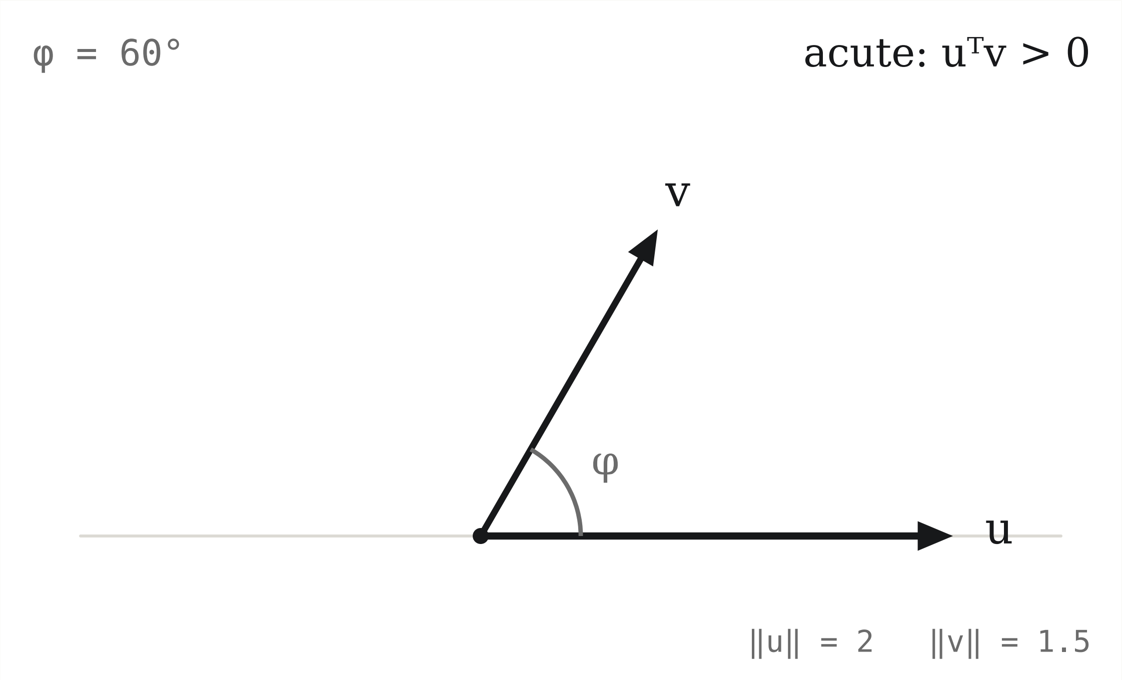
Orthogonality and Orthonormal Bases

The Dot Product Measures Angle

$$u^\mathsf{T} v = \|u\|\,\|v\|\cos\varphi$$

$\varphi \in [0, \pi]$	the angle between nonzero u, v; no angle if one is 0, and then $u^\mathsf{T} v = 0$
why	a rotation R keeps $u^\mathsf{T} v$; turn both by R with $Ru = \|u\|e_1$: $u^\mathsf{T} v = \|u\|(Rv)_1$
sign	$\|u\|\|v\| > 0$, so only $\cos\varphi$ can turn it

$\varphi < 90^\circ$	$\varphi = 90^\circ$	$\varphi > 90^\circ$
$u^\mathsf{T} v > 0$	$u^\mathsf{T} v = 0$	$u^\mathsf{T} v < 0$

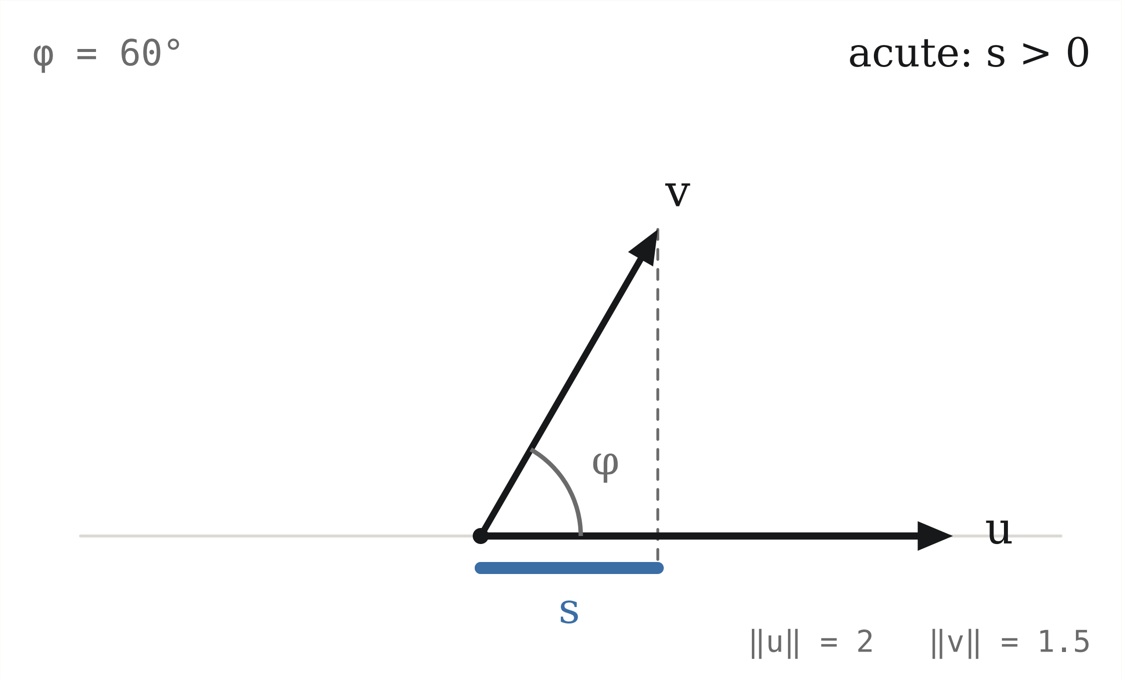


Orthogonality and Orthonormal Bases

The Signed Shadow

$$s = \frac{u^{\mathsf{T}} v}{\|u\|} = \|v\| \cos \varphi$$

s	a number ($u \neq 0$): the signed component of v along the unit direction of u
$\|v\| \cos \varphi$	also needs $v \neq 0$: no angle when $v = 0$, though $s = 0$
divisor	$\|u\|$, never $\|v\|$ — measured along u
$\|s\|$	the shadow's length; $s > 0$ same side as u, $s < 0$ opposite, $s = 0$ perpendicular



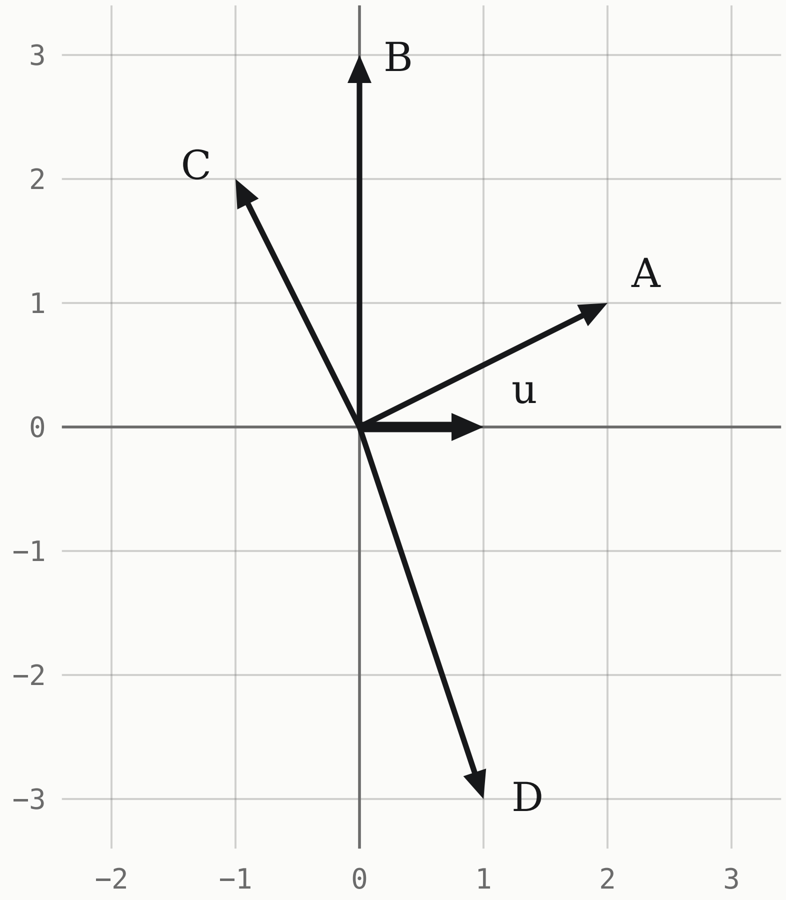
VOTE · ALL THAT APPLY

Orthogonality and Orthonormal Bases

Predict the Sign

Which have a positive dot product with $u = (1, 0)$?

- **A** $v_A = (2, 1)$
- **B** $v_B = (0, 3)$
- **C** $v_C = (-1, 2)$
- **D** $v_D = (1, -3)$



Orthogonality and Orthonormal Bases

Orthogonality and the Pythagorean Theorem

$$u \perp v \iff u^{\mathsf{T}} v = 0$$

$\perp$ orthogonal: dot product zero

$0^{\mathsf{T}} v = 0$ so $0 \perp$ every v (no angle defined)

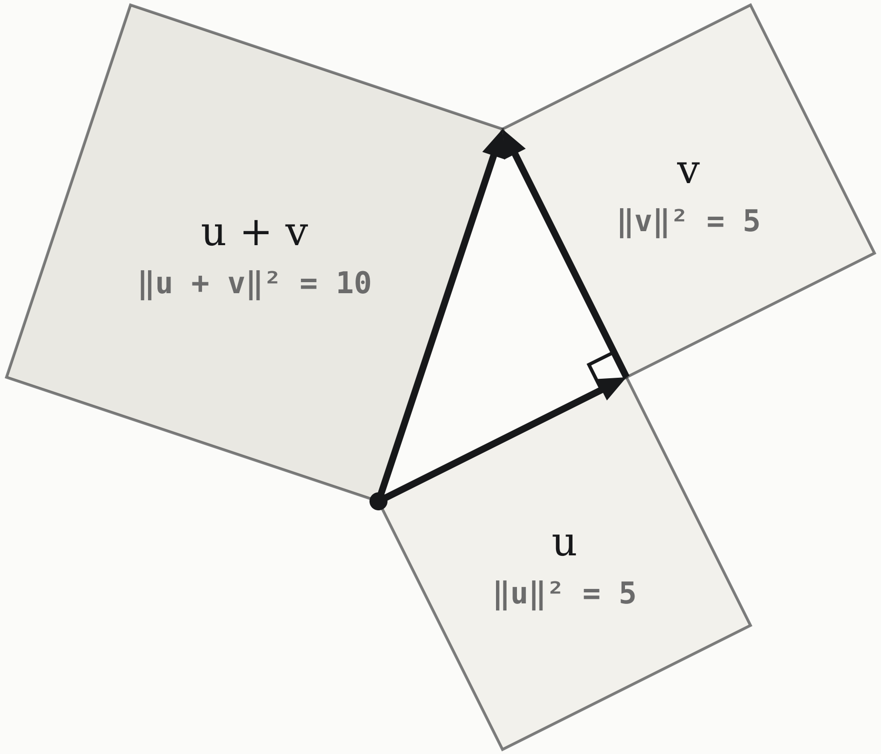
$$\|u + v\|^2 = \|u\|^2 + 2\, u^{\mathsf{T}} v + \|v\|^2$$

why expand $(u+v)^{\mathsf{T}}(u+v)$ by linearity

$2\, u^{\mathsf{T}} v$ cross term: 0 iff $u \perp v$

WORKED EXAMPLE

$u = (2, 1)$, $v = (-1, 2)$: $u^{\mathsf{T}} v = -2 + 2 = 0$, so $\|(1, 3)\|^2 = 10 = 5 + 0 + 5$.



DEFINITION

Orthogonality and Orthonormal Bases

Orthogonal to a Whole Subspace

$$r \perp S \iff r^\top y = 0 \text{ for every } y \in S$$

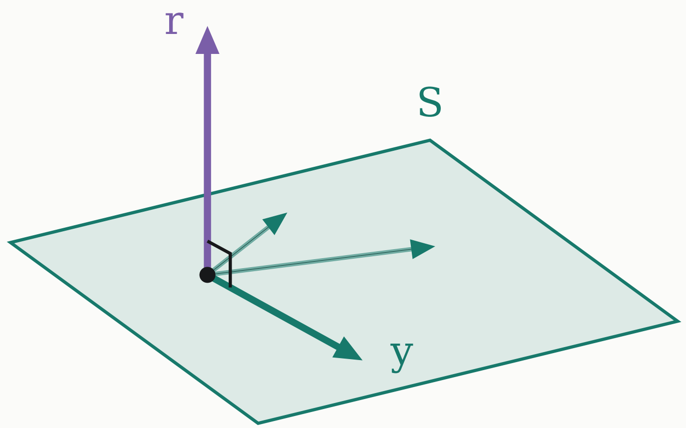
$S \subseteq \mathbb{R}^m$ a subspace; y is every vector of it

$r^\top y$ one number for each y in S

$$S^\perp = \{r \in \mathbb{R}^m : r^\top y = 0 \ \forall y \in S\}$$

a subspace $0 \in S^\perp$; $(c_1 r_1 + c_2 r_2)^\top y = 0$ too, by linearity

edges $\{0\}^\perp = \mathbb{R}^m$, $(\mathbb{R}^m)^\perp = \{0\}$: $r^\top r = 0 \Rightarrow r = 0$



schematic

VOTE

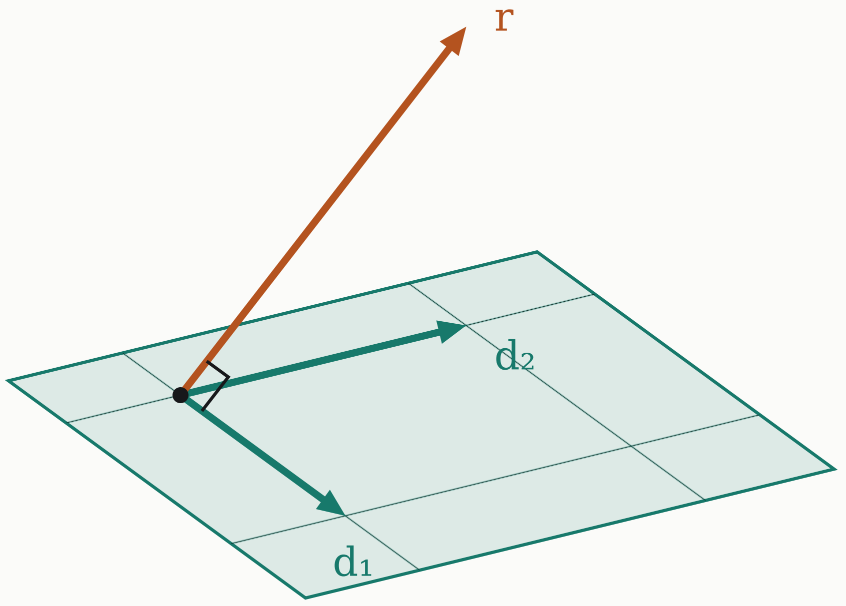
Orthogonality and Orthonormal Bases

Perpendicular to a Line, or to the Plane?

$d_1 = (1, 0, 0)$, $d_2 = (0, 1, 0)$, $S = \text{span}\{d_1, d_2\}$, $r = (0, 1, 1)$

$r \perp d_1$. Which is true?

- **A** $r \perp S$
- **B** $r/\|r\| \perp S$
- **C** $r \not\perp S$: $r^\top d_2 = 1$
- **D** $r^\top d_1 = 0$ gives $r \perp S$



S = span{d₁, d₂} 3D view: trust the number

Orthogonality and Orthonormal Bases

Testing a Spanning Set Is Enough

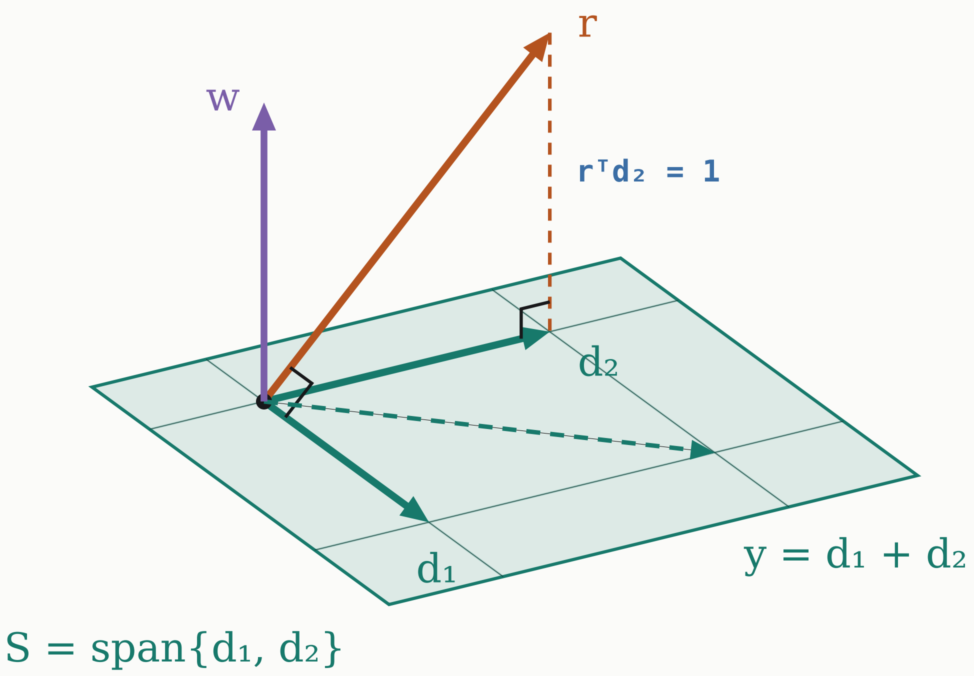
$$r^{\top}y = \sum_{i=1}^{k} c_i\,(r^{\top}d_i)$$

$y \in S$	$S = \mathrm{span}\{d_1, \ldots, d_k\}$, so $y = \sum c_i d_i$
k tests	settle every y: $r^{\top}y$ is linear in y
span only	independence not needed

WORKED EXAMPLE

From the vote: $y = d_1 + d_2$ gives $r^{\top}y = 1$, $w^{\top}y = 0$; d_1 alone misses it.

$r \perp S \Leftrightarrow r^{\top}d_i = 0$ for every i.



Orthogonality and Orthonormal Bases

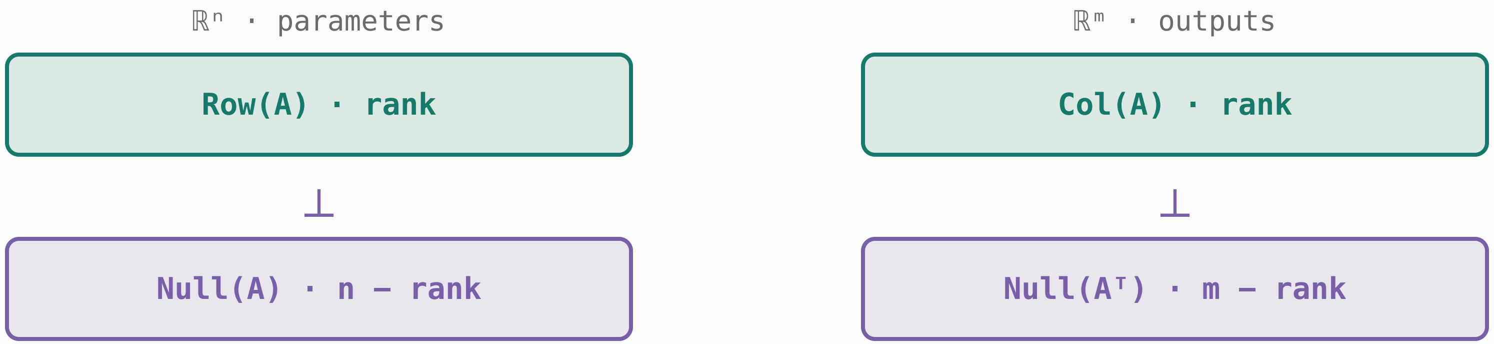
The Four Fundamental Subspaces Revisited

$A^\mathsf{T} r = (a_1^\mathsf{T} r, \ldots, a_n^\mathsf{T} r) \in \mathbb{R}^n$, so $r \perp \mathrm{Col}(A) \Leftrightarrow A^\mathsf{T} r = 0$.

$$\mathrm{Col}(A)^\perp = \mathrm{Null}(A^\mathsf{T}) \subseteq \mathbb{R}^m, \qquad \mathrm{Null}(A) = \mathrm{Row}(A)^\perp \subseteq \mathbb{R}^n$$

a_j columns of $A \in \mathbb{R}^{m \times n}$: they span $\mathrm{Col}(A)$

`input` same on A^T; Lecture 1 wrote $\mathcal{C}, \mathcal{N}$



Strang, G. ***MIT 18.06 Linear Algebra***, Lecture 14, orthogonal vectors and subspaces. MIT OpenCourseWare, 2010.

SORT

Orthogonality and Orthonormal Bases

Orthogonal Is Not the Same as Orthonormal

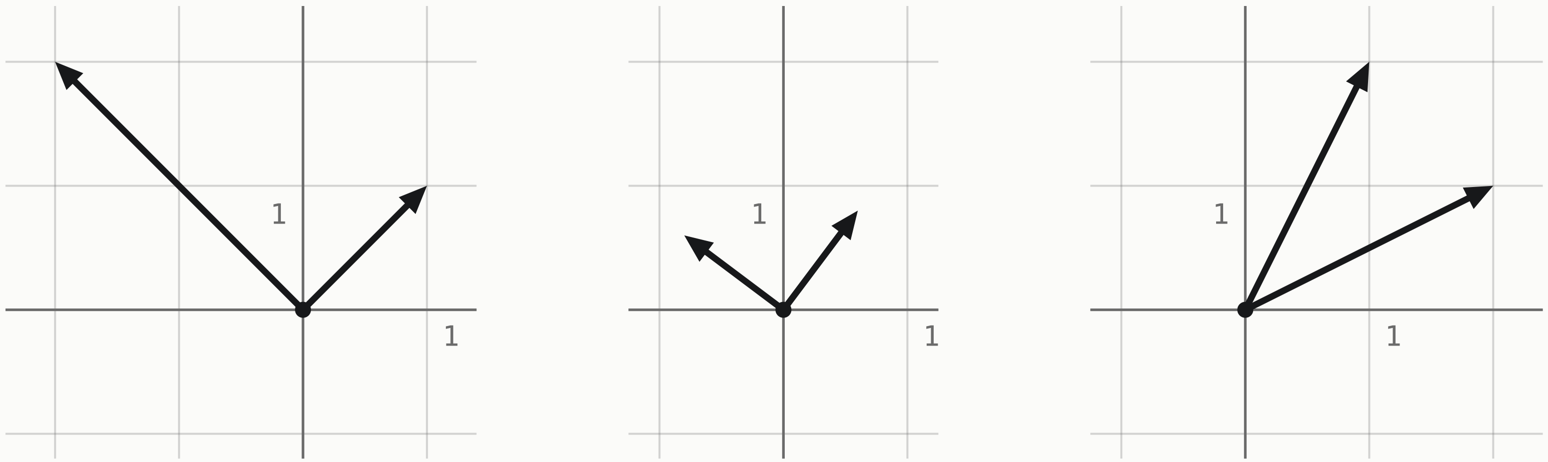
$$q_i^\top q_j = \delta_{ij}$$

δ_{ij} 1 if $i = j$ (unit), else 0 (orthogonal)

Each pair is a basis. Sort:

- A independent only
- B orthogonal, not unit
- C orthonormal

1 2 3



Orthogonality and Orthonormal Bases

Right Angles, Unit Lengths, Independence

RIGHT ANGLES

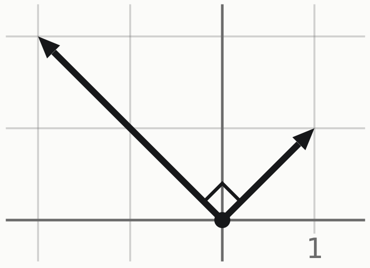
$$\sum_i c_i v_i = 0 \;\Rightarrow\; v_j^\top \sum_i c_i v_i = c_j \|v_j\|^2 = 0$$

v_i	pairwise $\perp$, nonzero: $\|v_j\|^2 > 0$
`all` j	$c_j = 0$: independent
`fails`	drop nonzero: $0 \perp$ all

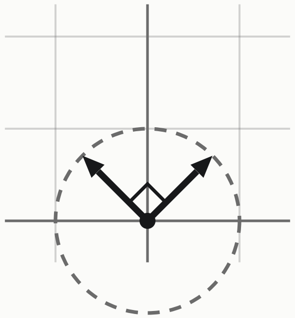
UNIT LENGTHS

Not for independence; with $\perp$: $v = \sum c_i v_i \Rightarrow c_j = v_j^\top v$.

> Orthonormal = both. Nonzero right angles $\Rightarrow$ independent, not conversely.



orthogonal



orthonormal

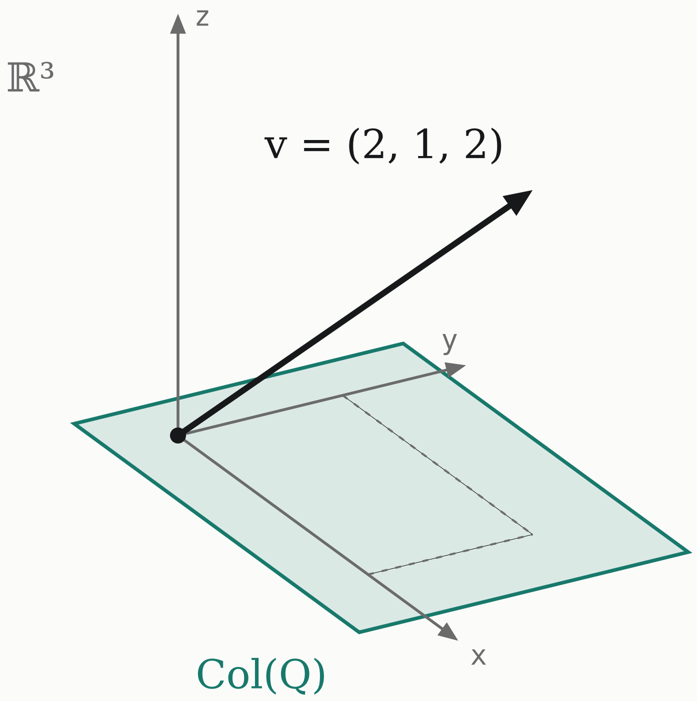
VOTE · ALL THAT APPLY

Orthogonality and Orthonormal Bases

What Does $Q^\top Q = I$ Actually Mean?

For $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, which are true?

- (A) $Q^\top Q = I_2$
- (B) $QQ^\top = I_3$
- (C) $\|Qz\| = \|z\|$ for every $z \in \mathbb{R}^2$
- (D) $Q^\top v$ keeps all of any $v \in \mathbb{R}^3$



Orthogonality and Orthonormal Bases

Q Embeds, $Q^\top$ Extracts

Q **EMBEDS**

$$\|Qz\|^2 = z^\top Q^\top Q z = \|z\|^2$$

Q $m \times k$, columns q_i, $Q^\top Q = I_k$; $S = \mathrm{Col}(Q)$

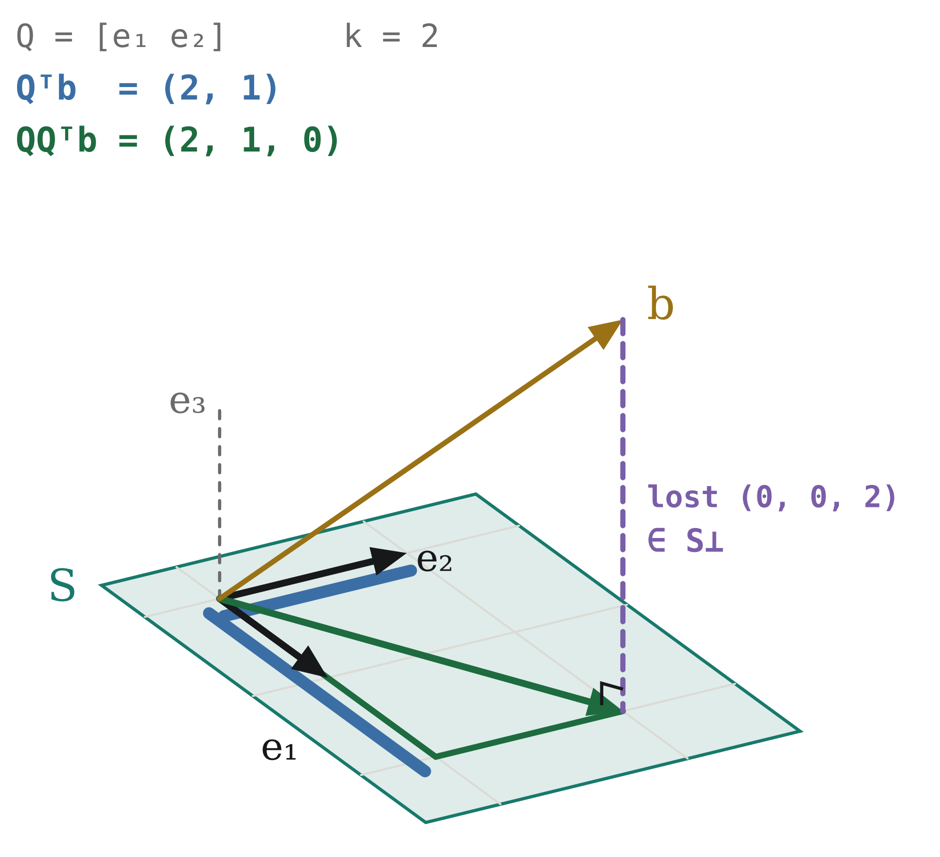
$z \in \mathbb{R}^k$ $Qz \in \mathbb{R}^m$: lengths, angles kept

$Q^\top$ **EXTRACTS**

$(Q^\top b)_i = q_i^\top b$: coordinates of the projection.

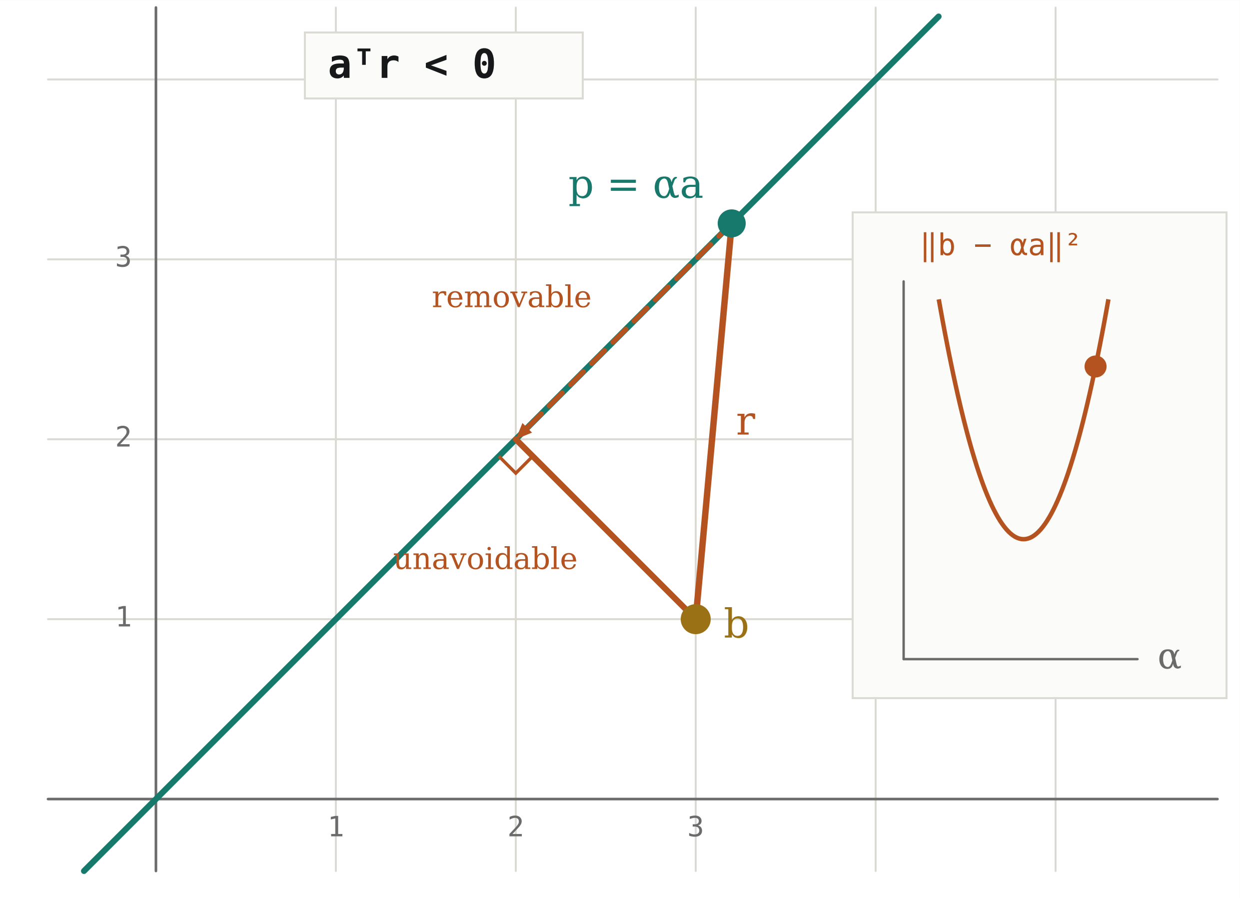
	Full orthonormal basis	**Orthonormal basis of a proper subspace**
$Q^\top Q$	I_m	I_k
$QQ^\top b$	b	projection
recover b	yes	iff $b \in S$

> $Q^\top$ loses b's $S^\perp$ component.



Orthogonal Projection

The Residual Must Be Perpendicular



$$r = b - p, \qquad p = \alpha a$$

b the target in $\mathbb{R}^m$, off the line

$p = \alpha a$ a candidate; α is our only choice

r what p fails to explain

$$\text{lowest } \|b - \alpha a\|^2 \;\Rightarrow\; a^\mathsf{T} r = 0$$

$a^\mathsf{T} r \neq 0$ r keeps a part along the line: removable, so p is not lowest

$a^\mathsf{T} r = 0$ nothing left along the line — why is that enough?

DEFINITION

Orthogonal Projection

Solving the Perpendicularity Equation

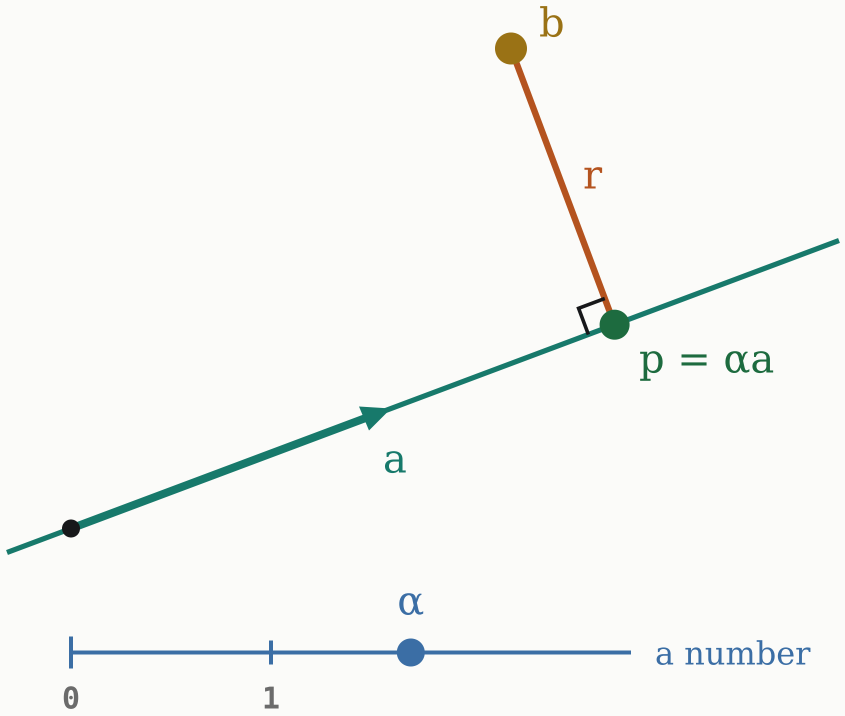
$a \neq 0$, $b \in \mathbb{R}^m$. Candidate on span$\{a\}$: $p = \alpha a$

Ask for a perpendicular residual: $a^\top(b - \alpha a) = 0$

Expand: $a^\top b - \alpha\, a^\top a = 0$ — one equation, one unknown

$$\alpha = \frac{a^\top b}{a^\top a}, \qquad p = \alpha a$$

- α a number: how many copies of a
- p a vector in $\mathbb{R}^m$
- $a^\top a$ $= \|a\|^2 > 0$ as $a \neq 0$; $a = 0$: no line



Strang. ***MIT 18.06***, L15 projections onto subspaces. MIT OpenCourseWare, 2010.

Orthogonal Projection

The Same Projection as a Matrix

Substitute α back: the same p, now one map.

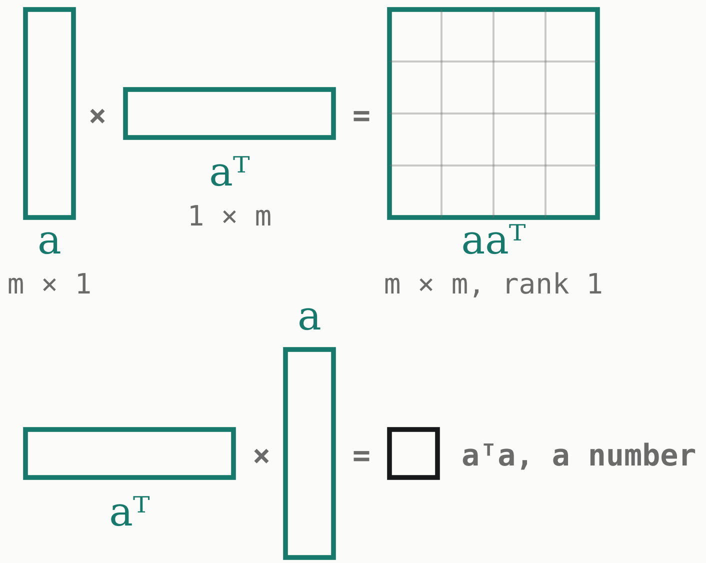
$$p = a\,\frac{a^\top b}{a^\top a} = \frac{a a^\top}{a^\top a}\,b$$

$a^\top b$ $1 \times m$ by $m \times 1$: a number

$aa^\top$ $m \times 1$ by $1 \times m$: $m \times m$, rank 1 if $a \neq 0$, symmetric

$a = 0$ divisor 0: quotient undefined

Types: α number · p vector · $\dfrac{aa^\top}{a^\top a}$ matrix



$a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $b = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

Orthogonal Projection

Compute and Check the Projection

Alone, no arithmetic: which way does r point from p to b, and is α above or below 1?

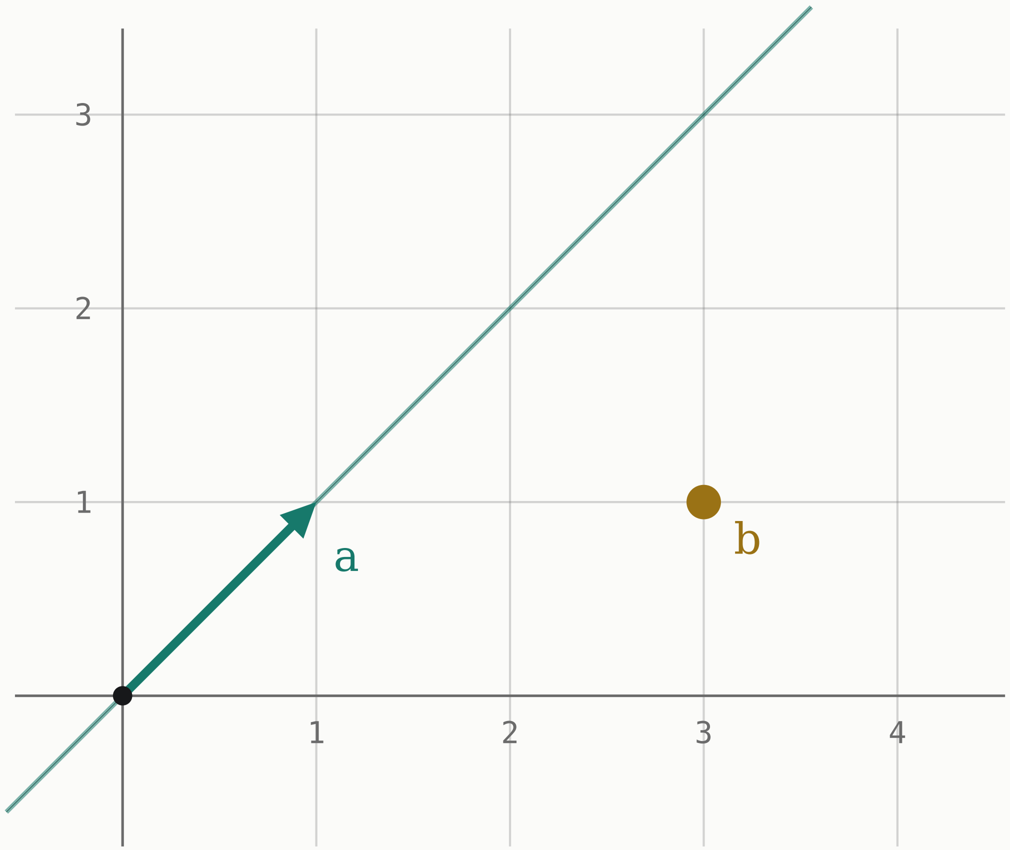
Then in pairs: α, p, r, check r.

1 α

2 p

3 r

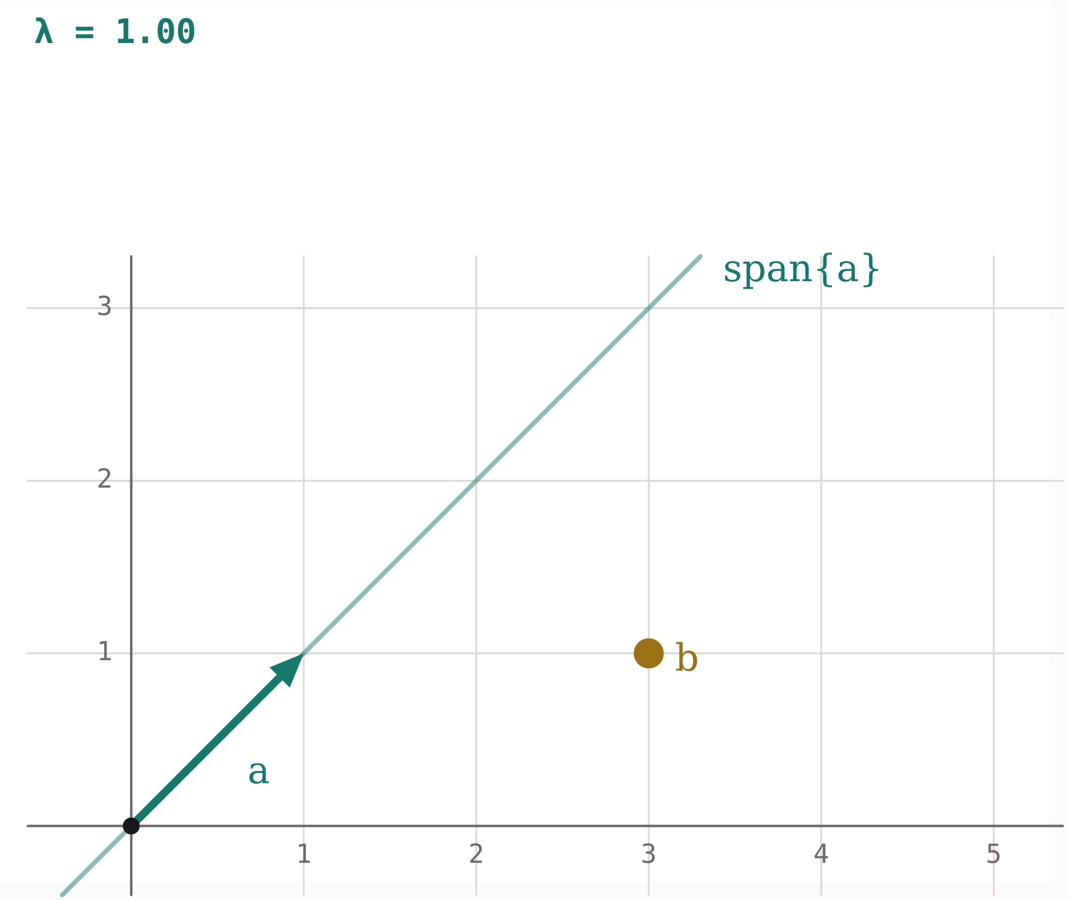
CHECK



VOTE

Orthogonal Projection

Does Rescaling the Basis Change the Projection?



Replace a by λa, $\lambda > 0$. What happens to $a^\top b$, α, p?

- (A) $a^\top b \times\lambda$, $\alpha \div\lambda$, p same
- (B) all the same
- (C) $a^\top b \times\lambda$, $\alpha \times\lambda$, $p \times\lambda^2$
- (D) $a^\top b$ same, $\alpha \div\lambda$, p same

Orthogonal Projection

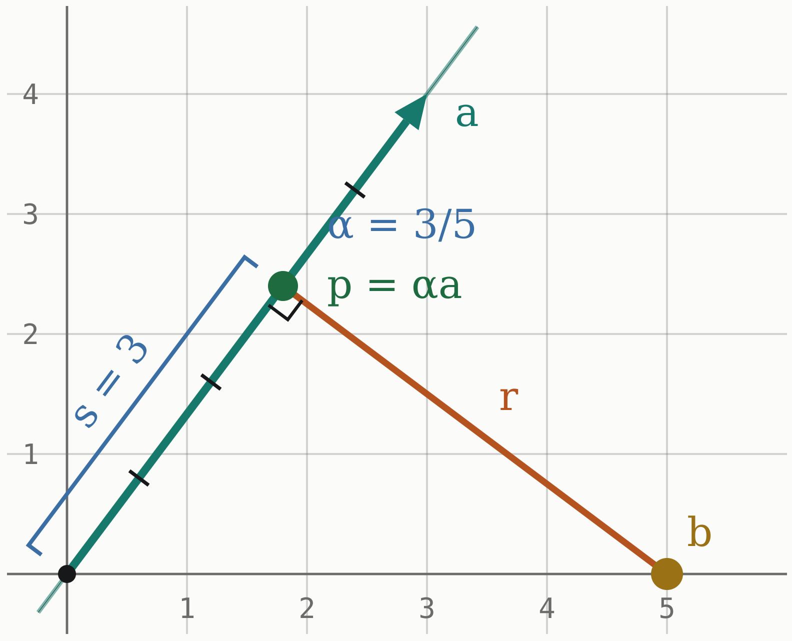
Component, Coefficient, Projection

$a \neq 0$		$\lambda a,\ \lambda > 0$
$s = a^\top b / \|a\|$	signed shadow	same
$\alpha = a^\top b / \|a\|^2$	copies of a	$\div \lambda$
$p = \alpha a$	vector in $\mathbb{R}^m$	same

Unit a: $s = \alpha$, the coordinate of Lecture 2; else equal only if $a^\top b = 0$.

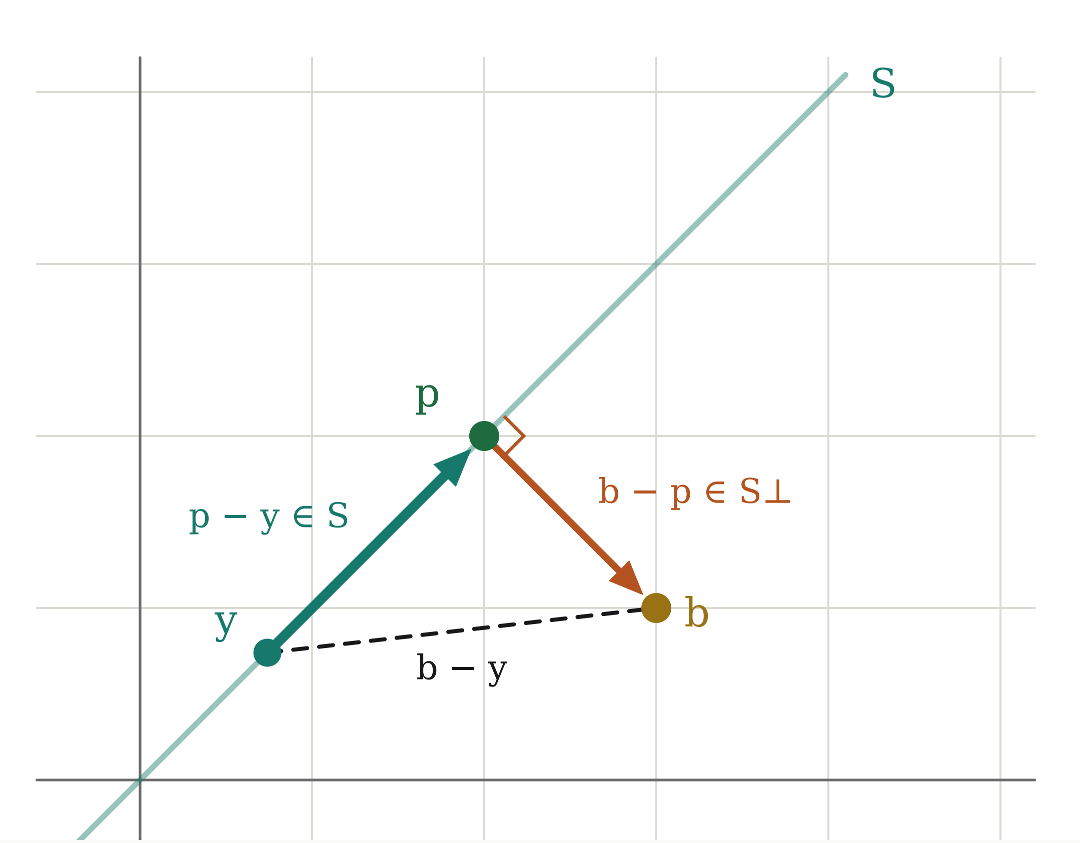
WORKED EXAMPLE

$a = (3,4)$, $b = (5,0)$: $a^\top b = 15$, $\|a\| = 5$, $s = 3$, $\alpha = 3/5$, $p = (9/5, 12/5)$, $r = (16/5, -12/5)$, $a^\top r = 0$



Orthogonal Projection

Split the Error at Any Candidate



Why is a perpendicular residual enough to make p the closest point?

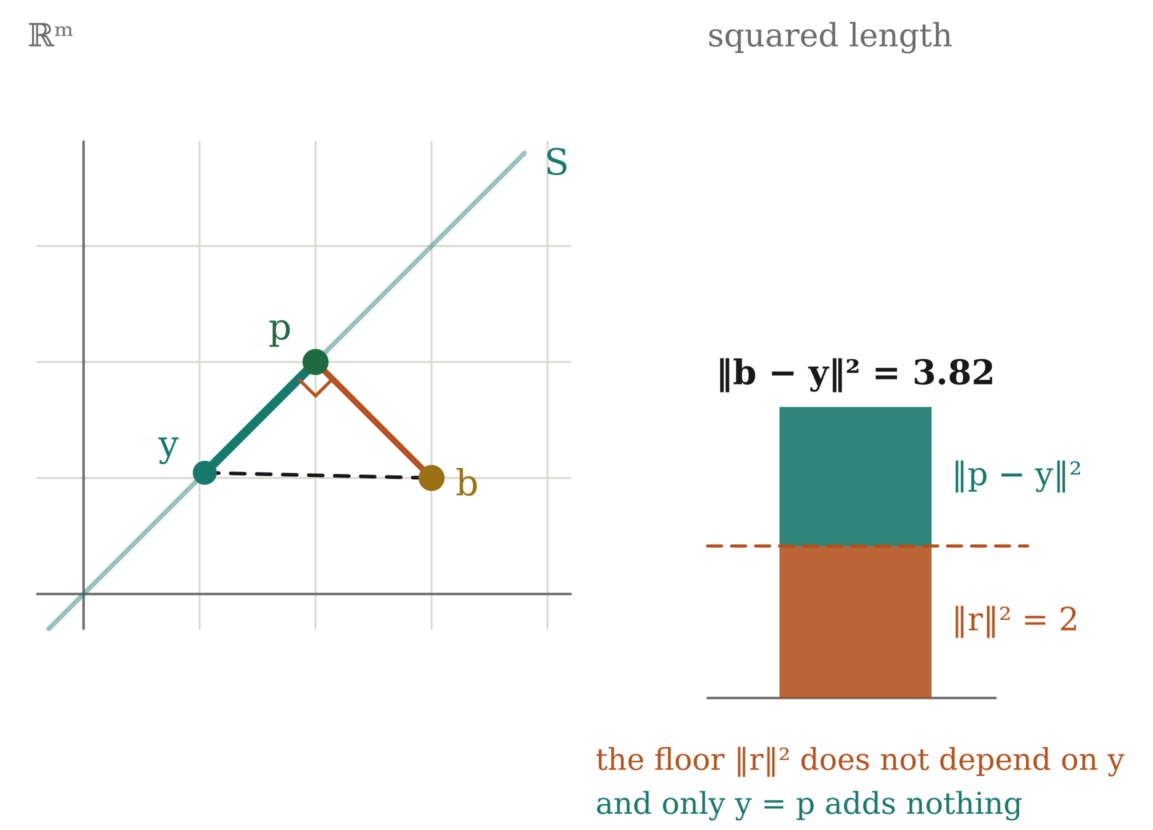
Premise: S any subspace, $p \in S$, $r = b - p \in S^{\perp}$. Take any other $y \in S$.

$$b - y = (b - p) + (p - y)$$

$p - y$	lies in S: a subspace is closed under subtraction
$b - p$	lies in $S^{\perp}$, so $(b-p)^{\top}(p-y) = 0$
`right angle`	the two pieces are the legs, $b - y$ the hypotenuse

Orthogonal Projection

Yes: the Perpendicular Test Is Enough



Yes. Take squared lengths on that right triangle.

$$\|b-y\|^2 = \|r\|^2 + \|p-y\|^2 \;\geq\; \|r\|^2$$

$\|r\|^2$ unavoidable: it does not depend on y

$\|p-y\|^2$ the extra, ≥ 0, zero only at $y=p$

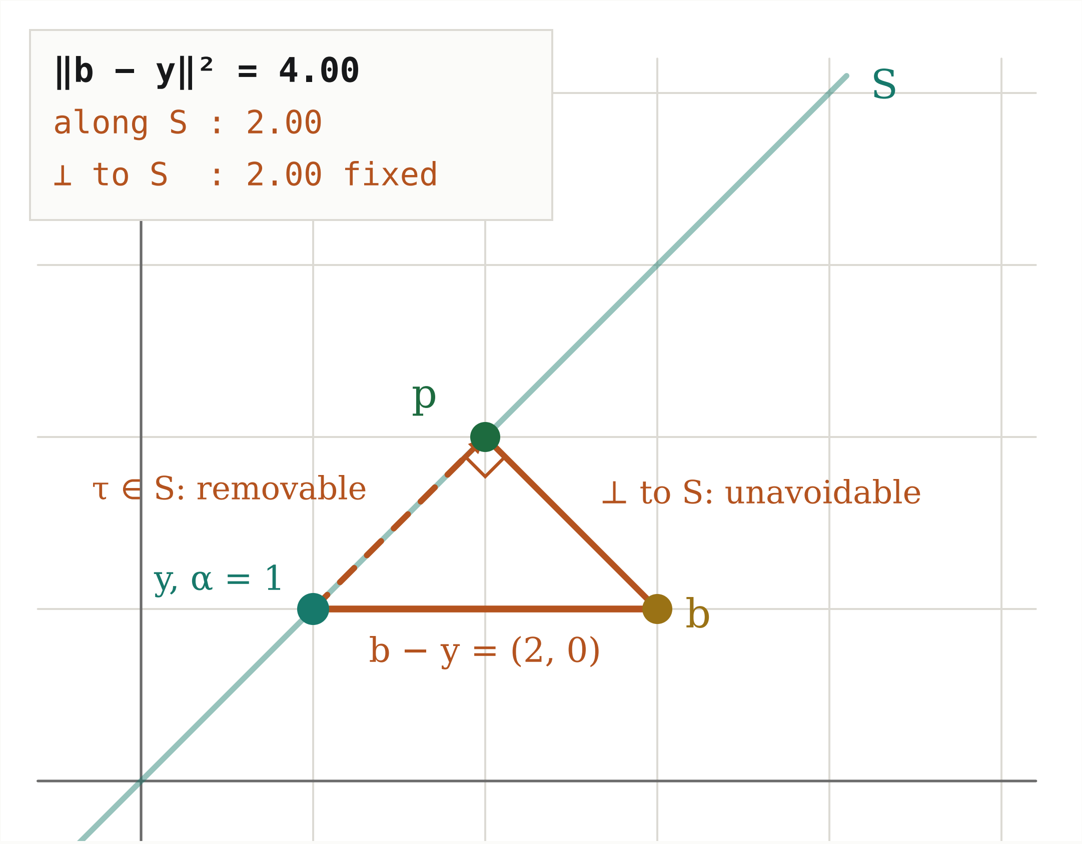
Optimal · every $y \neq p$ adds a positive square

Unique · nothing else ties p

Built for a line; a general S still needs a construction.

Orthogonal Projection

A Candidate That Is Not Yet Best



An illustration of the converse: a residual still leaning along S.

$$\|(b-y)-\tau\|^2 = \|b-y\|^2 - \|\tau\|^2$$

- τ — the orthogonal projection of $b-y$ onto the line; step to $y+\tau \in S$
- $b-y-\tau$ — the perpendicular part, untouched
- why — Pythagoras: $\tau \perp (b-y-\tau)$

> At $\alpha = 1$: $b-y=(2,0)$, $\tau=(1,1)$; at $y+\tau=p$: $4-2=2$, nothing left to improve.

DEFINITION

Orthogonal Projection

The Projection Theorem

$S \subseteq \mathbb{R}^m$ a subspace: each $b \in \mathbb{R}^m$ splits one way only, and that p is closest.

$$b = p + r, \qquad p \in S, \quad r \in S^\perp$$

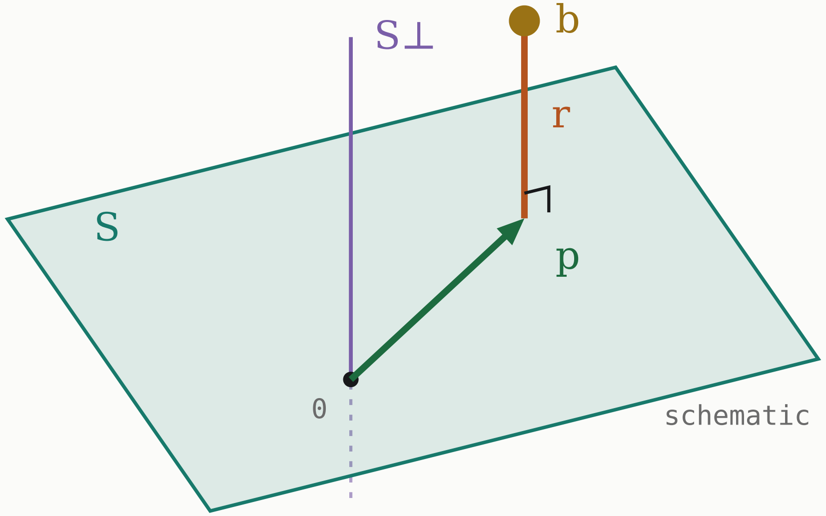
unique $p + r = p' + r' \Rightarrow p - p' = r' - r \in S \cap S^\perp = \{0\}$ ($\perp$ itself)

closest proved, for any $y \in S$

exists construction follows, from an orthonormal basis Q

edges $b \in S \Rightarrow r = 0$; $b \in S^\perp \Rightarrow p = 0$

Lecture 1 $S =$Col(A): $b = \hat{b} + r$, $r \in$Null$(A^\top)$, dims add



Orthogonal Projection

Projection onto a Subspace: $p = QQ^\mathsf{T} b$

$Q = \begin{bmatrix} 1 & 0 \\ 0 & 4/5 \\ 0 & 3/5 \end{bmatrix}$ $b = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$

$Q \in \mathbb{R}^{m\times k}$ with $Q^\mathsf{T} Q = I_k$: columns q_i orthonormal, spanning S.

$$c = Q^\mathsf{T} b \in \mathbb{R}^k, \qquad p = Qc = QQ^\mathsf{T} b$$

extract $c_i = q_i^\mathsf{T} b$: the part of b along q_i

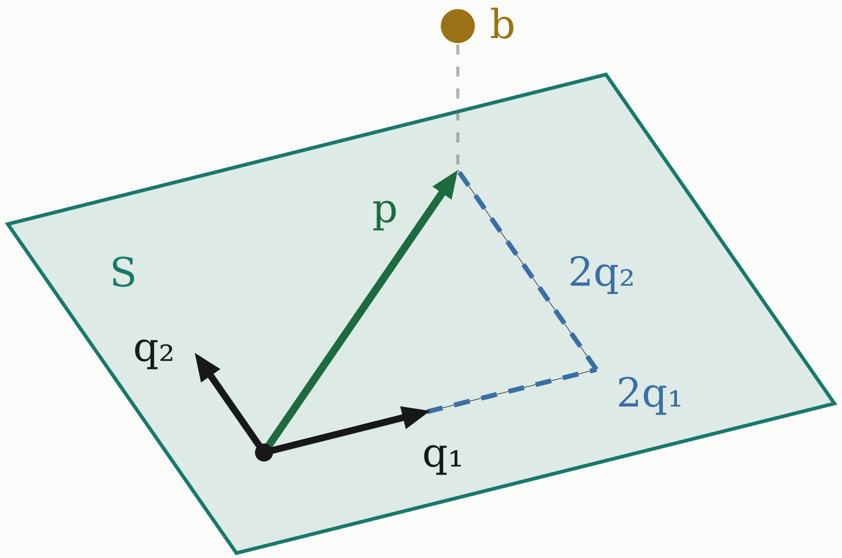
rebuild $p = c_1 q_1 + \cdots + c_k q_k \in S$

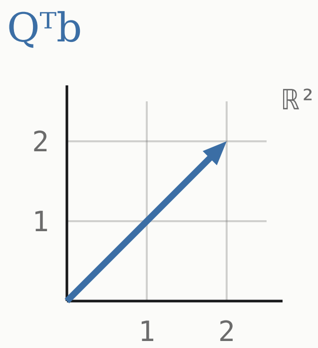
c the coordinates of p; of b too iff $b \in S$

sizes $m \times k$ times $\mathbb{R}^k$: $p \in \mathbb{R}^m$

WORKED EXAMPLE

$Q^\mathsf{T} b = (2, 2)$, $p = (2, 8/5, 6/5)$





Orthogonal Projection

Checking the Two Requirements

$Q = \begin{bmatrix} 1 & 0 \\ 0 & 4/5 \\ 0 & 3/5 \end{bmatrix}$ $b = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$

Both requirements, one at a time.

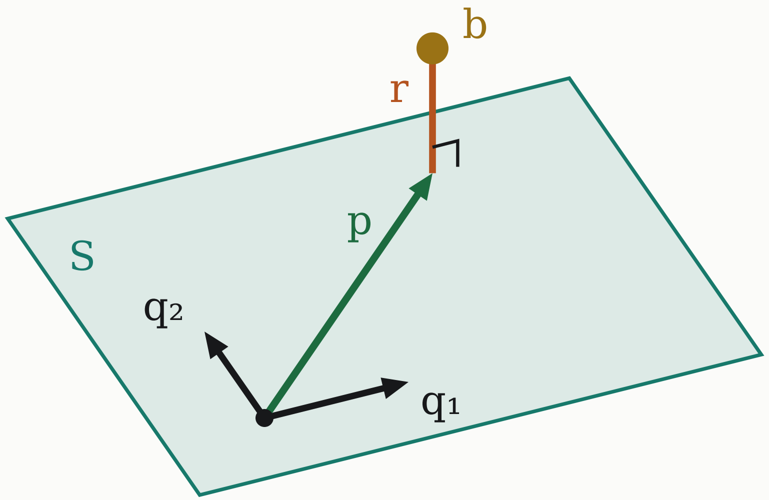
1. $p \in S$: $p = Q(Q^\top b)$, Q times a vector of $\mathbb{R}^k$.

$$Q^\top(b - p) = Q^\top b - (Q^\top Q)Q^\top b = 0$$

2. $b - p \perp$ each $q_i \Rightarrow \perp S$ (spanning test) $\Rightarrow p$ is **the** closest point

needs $Q^\top Q = I_k$; a general basis A needs a correction factor, not plain $AA^\top$

here $r = (0, -3/5, 4/5)$: $q_1^\top r = 0$, $q_2^\top r = -12/25 + 12/25 = 0$

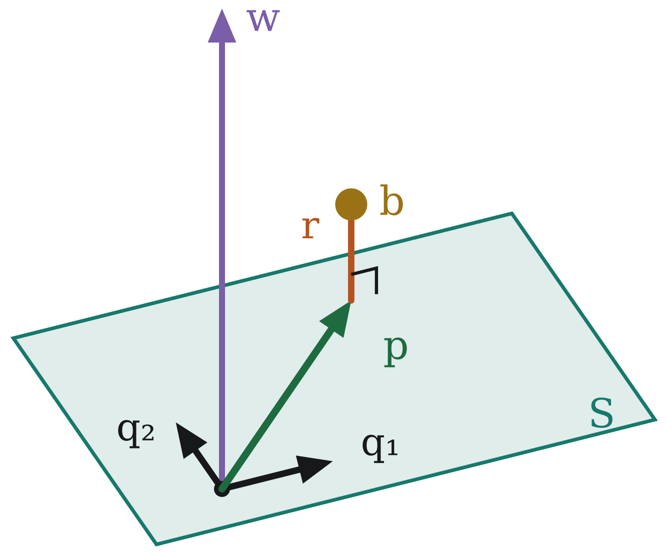


VOTE

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 4/5 \\ 0 & 3/5 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

Orthogonal Projection

Does Moving b Out of the Plane Change the Prediction?



Same S and b. Add $w = (0, -3, 4)$, $\perp$ q_1 and q_2. For $b + w = (2, -2, 6)$:

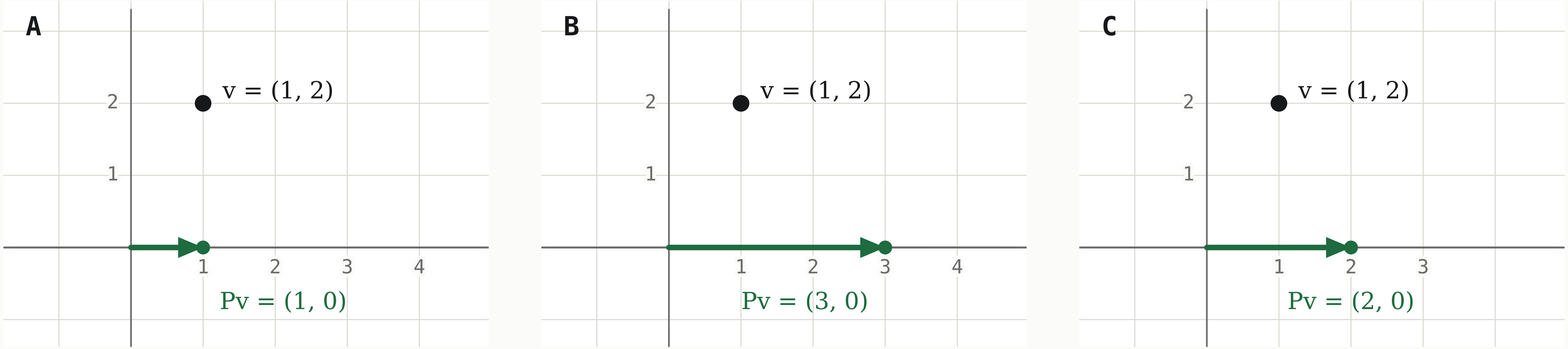
- (A) projection changes, residual same
- (B) projection same, residual changes
- (C) both stay the same
- (D) both change

VOTE

Orthogonal Projection

Which Matrix Is an Orthogonal Projector?

Each of these flattens the plane onto the axis. Which sends every v to its orthogonal projection? On $v = (1, 2)$: does a second pass move Pv? Is the discarded direction $\perp$ the axis?



(A) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

(B) $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

(C) $\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$

DEFINITION

Orthogonal Projection

What the Projection Matrix Does

Orthogonal projector onto S: a matrix P with $Pb \in S$ and $b - Pb \in S^\perp$ for **every** $b \in \mathbb{R}^m$ (the two requirements checked).

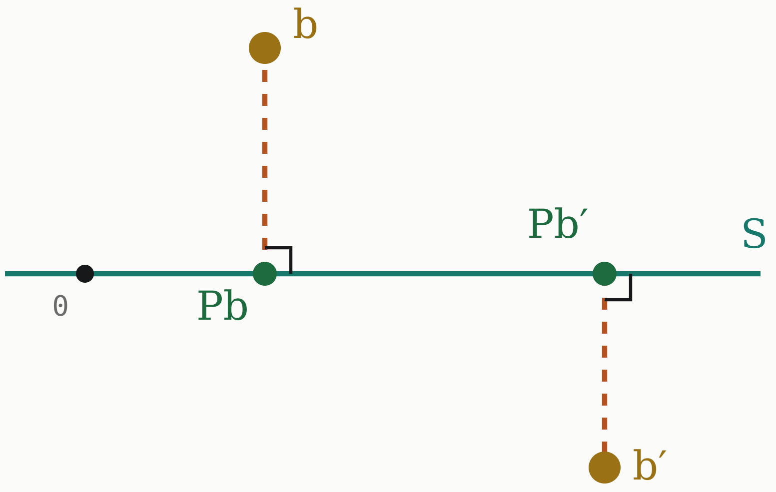
$$P = QQ^\mathsf{T} \in \mathbb{R}^{m\times m}, \qquad Pb = p$$

why QQ^T passes both; the closest point is unique

size Q is $m \times k$: P is $m \times m$, rank k, symmetric

unique any orthonormal basis of S gives the same P

edges $S = \{0\}$: $P = 0$; $S = \mathbb{R}^m$: $P = I$



Orthogonal Projection

Applying It Twice Changes Nothing

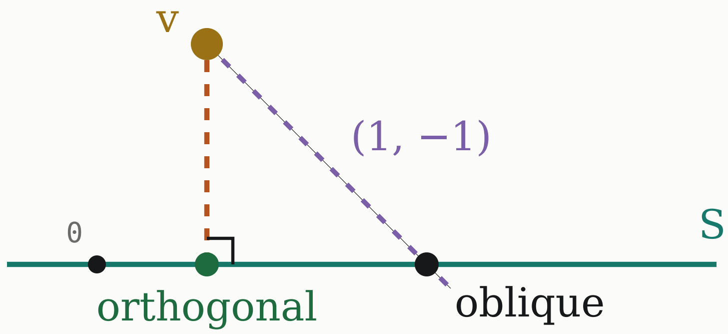
$$P^2 = Q(Q^\top Q)Q^\top = QQ^\top = P$$

`reason` $Pb \in S$, and the closest point of S to a point of S is itself

`reads` $P(Pb) = Pb$ for every b: the second pass moves nothing

`uses` $Q^\top Q = I_k$, nothing else

> $P^2 = P$ alone means **a** projection, not necessarily orthogonal: the vote's P_B is idempotent, yet $(1, -1)$, the direction it discards, is not $\perp$ the axis.



Orthogonal Projection

Symmetry Is the Missing Half

$$P^\top = (QQ^\top)^\top = QQ^\top = P$$

`why` $(UV)^\top = V^\top U^\top$, $(Q^\top)^\top = Q$

`so far` $QQ^\top$: symmetric, idempotent

> $P^\top = P = P^2$ ⇔ orthogonal projector onto $\mathrm{Col}(P)$

(⇒) $Pb \in \mathrm{Col}(P)$; $P^\top(b - Pb) = P(b - Pb) = Pb - P^2 b = 0$: $b - Pb \in \mathrm{Col}(P)^\perp$

(⇐) P fixes $\mathrm{Col}(P)$: $P^2 = P$. Any a, b: $a^\top Pb = (Pa)^\top Pb$ $[a - Pa \perp Pb]$ $= (Pa)^\top b$ $[b - Pb \perp Pa]$ $= a^\top P^\top b$: $P^\top = P$

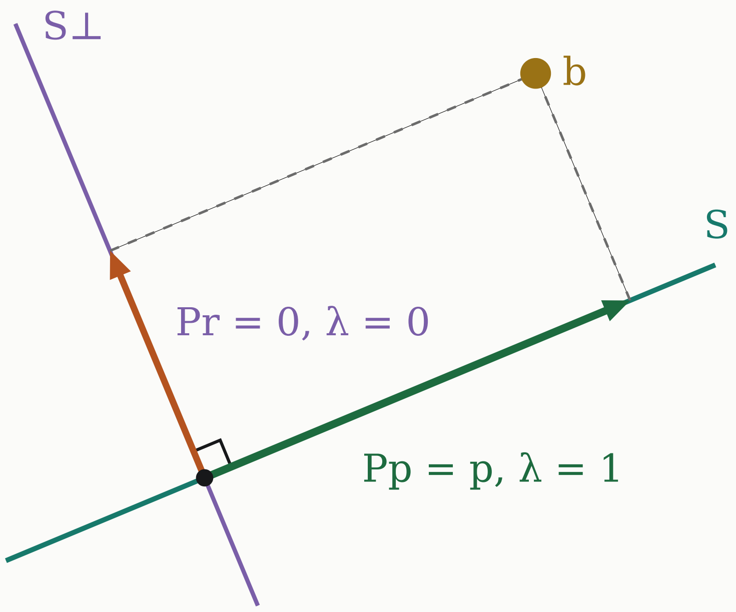
vote: $P_B \neq P_B^\top$, $P_C^2 \neq P_C$

Orthogonal Projection

What P Keeps and What It Discards

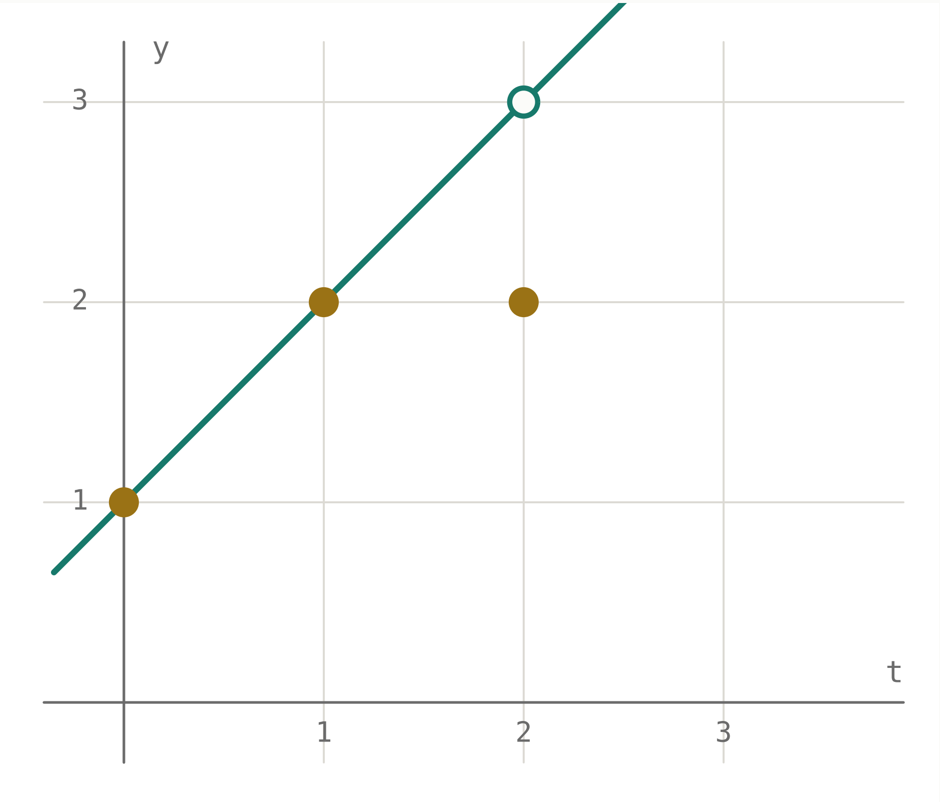
$$(I - P)^{\mathsf{T}} = I - P, \qquad (I - P)^2 = I - 2P + P^2 = I - P$$

$I - P$	orthogonal projector onto $S^{\perp}$: $Q^{\mathsf{T}}(I - P)b = 0$, $P(I - P) = 0$
$b = p + r$	$Pp = p$, $Pr = 0$
λ	1 on kept $p \neq 0$, 0 on discarded $r \neq 0$ (Lecture 2)
$\lambda = 0$	P singular: $P(b + w) = Pb$ for $w \in S^{\perp}$, so Pb does not determine the $S^{\perp}$ part of b
`edges`	$P = 0$: no $\lambda = 1$; $P = I$: no $\lambda = 0$



Least Squares

Three Measurements, Two Unknowns



Model $y \approx \beta_0 + \beta_1 t$ at inputs t_i.

all ones · β_0, the shared height

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

the inputs t_i · β_1, the slope

$A \in \mathbb{R}^{3\times 2}$	$m = 3$ readings, $n = 2$ parameters
$x = (\beta_0, \beta_1)$	intercept and slope, the unknowns
b	the readings y_i, one per input t_i
$b \notin \texttt{Col}(A)$	no exact x

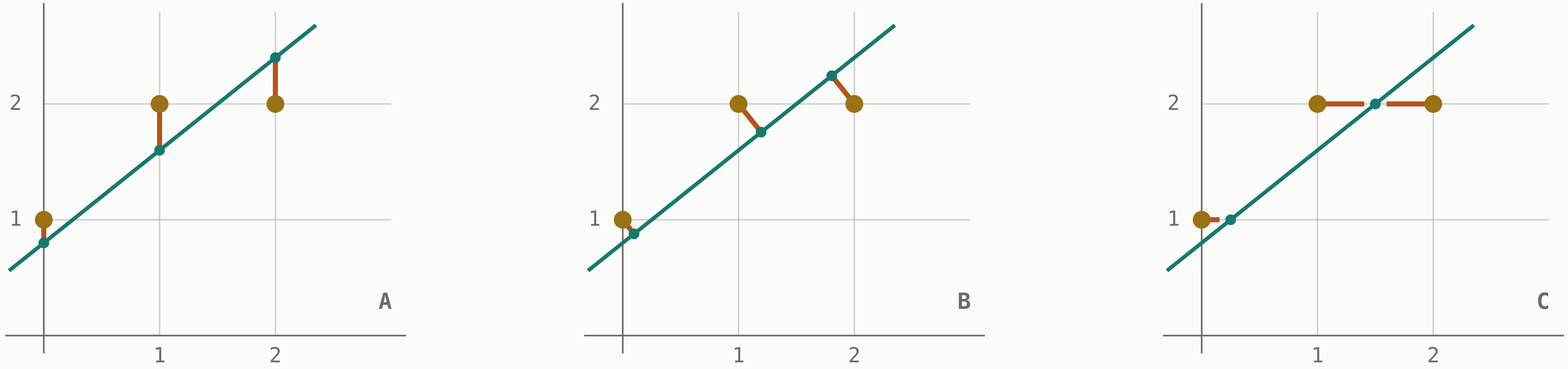
VOTE

$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ $b = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

Least Squares

Which Errors Are We Minimizing?

For $y_i \approx \beta_0 + \beta_1 t_i$, which errors does ordinary least squares minimize?



(A) vertical distances, squared

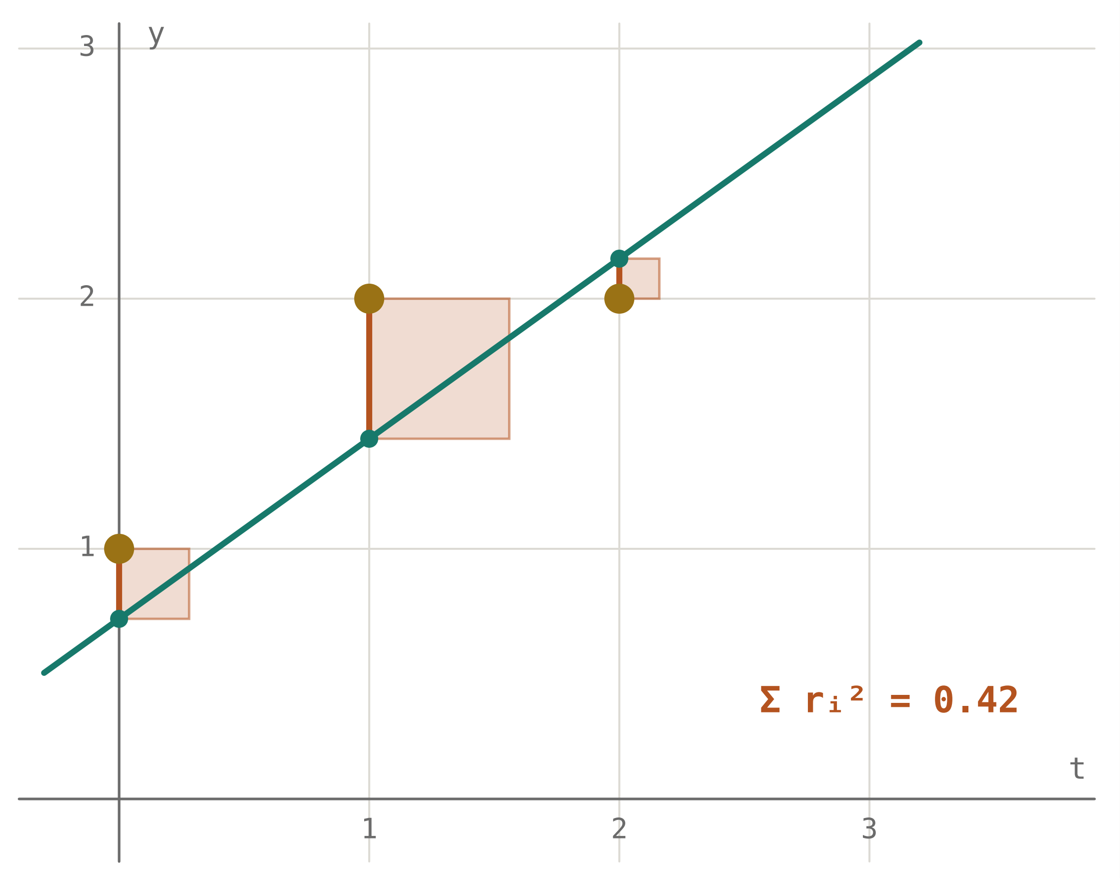
(B) shortest distances, squared

(C) horizontal distances, squared

Least Squares

The Objective and Its Objects

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

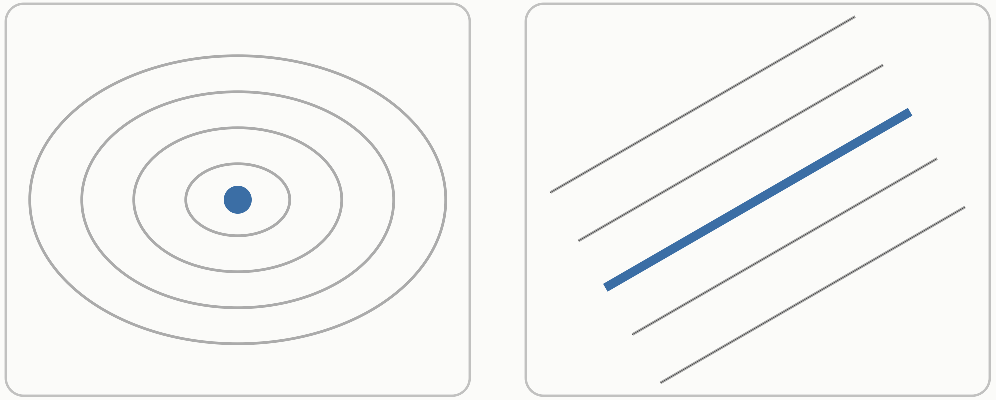


$$r = b - Ax, \qquad f(x) = \|Ax - b\|_2^2 = \|r\|_2^2$$

$x \in \mathbb{R}^n$	the parameters (β_0, β_1) we choose
$Ax, b, r \in \mathbb{R}^m$	prediction, readings, residual: measurement space
$f(x)$	one number, ≥ 0; any A, no rank assumption
$\lvert r_i \rvert$	the side of square i — a side cannot be negative

Least Squares

Minimum and Minimiser



one bottom a flat valley

both in parameter space $\mathbb{R}^n$ schematic

$$\min_x f(x) \in \mathbb{R}, \qquad \hat{x} \in \arg\min_x f(x) \subseteq \mathbb{R}^n$$

- **min** the smallest value f reaches — one number (ours: p53)
- **arg min** the set of x reaching it: one point or many
- $\hat{x}$ any member of that set; any A, any rank

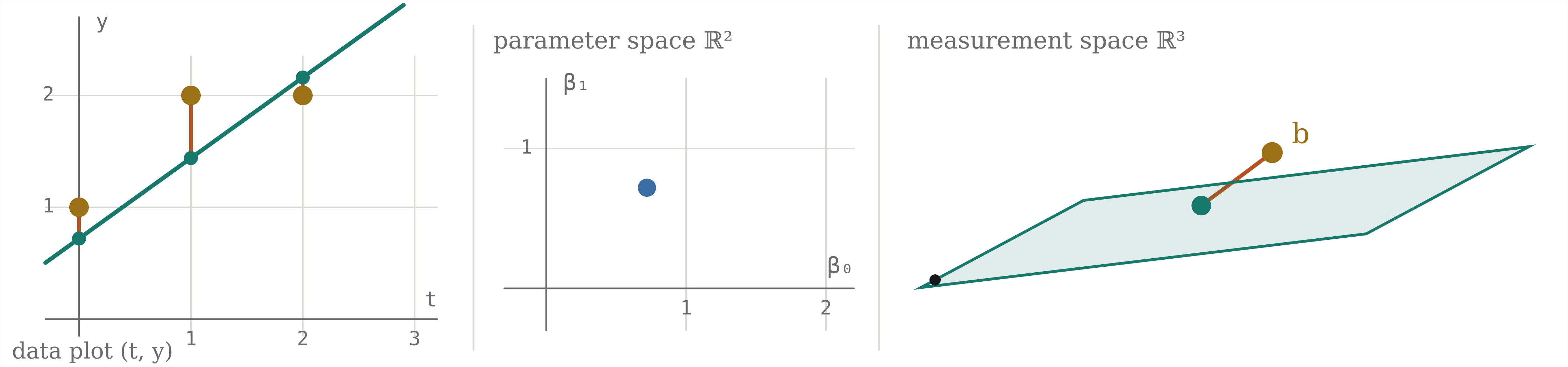
> The best output $A\hat{x}$ is always unique; the best parameter vector $\hat{x}$ need not be.

MATCH · PAIRS

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Least Squares

Fitting Is Projection in Measurement Space



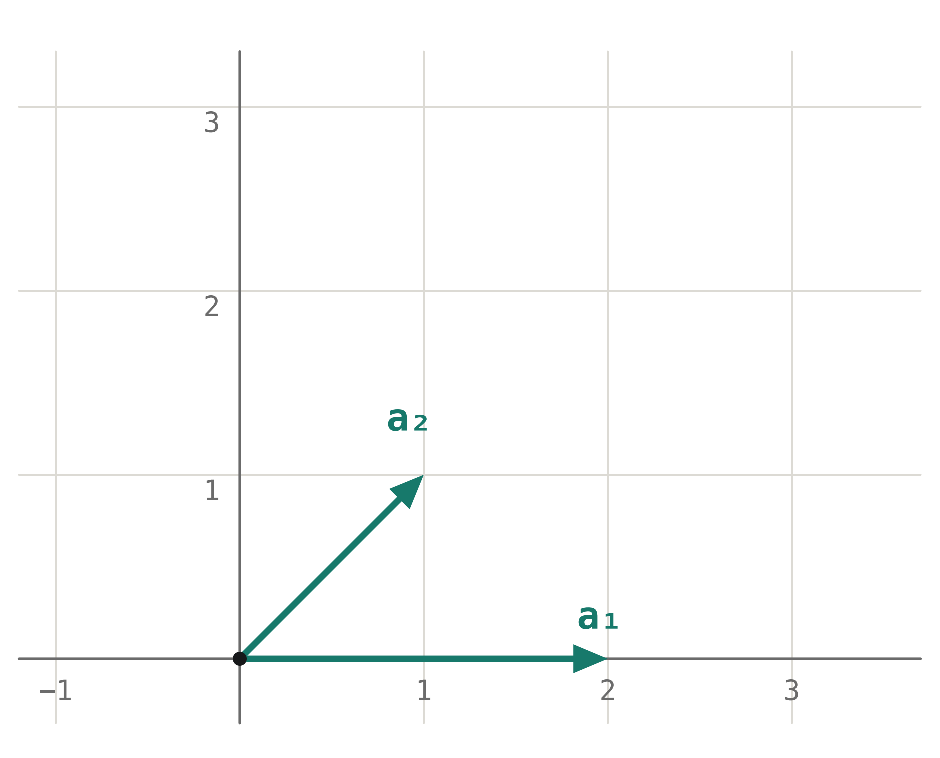
What kind of object is each, a line, a point or a plane, and in which space?

- (A) the numbers (β_0, β_1)
- (B) Ax
- (C) the candidate line
- (D) every possible Ax

VOTE

Least Squares

Perpendicular Columns: Which Gram Entries Change?



From Lecture 1: for $G = A^{\top}A$ with columns a_1, a_2, the entry is $G_{ij} = a_i^{\top}a_j$.

a_2 is turned until perpendicular to a_1 and then doubled in length. What happens to G_{11}, G_{12}, G_{22}?

- A unchanged, 0, $\times 4$
- B 1, 0, 1
- C unchanged, 0, $\times 2$
- D 0, unchanged, $\times 4$

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

DEFINITION

Least Squares

The Gram Matrix

$$G = A^\mathsf{T} A \in \mathbb{R}^{n\times n}, \qquad G_{ij} = a_i^\mathsf{T} a_j$$

a_j column j of $A \in \mathbb{R}^{m\times n}$ — any A, any rank

G_{ij} row i, column j: column i dotted with column j

$G^\mathsf{T} = G$ symmetric, because $a_i^\mathsf{T} a_j = a_j^\mathsf{T} a_i$

n the number of parameters, not of measurements

WORKED EXAMPLE

$a_1 = (1,1,1)$, $a_2 = (0,1,2)$

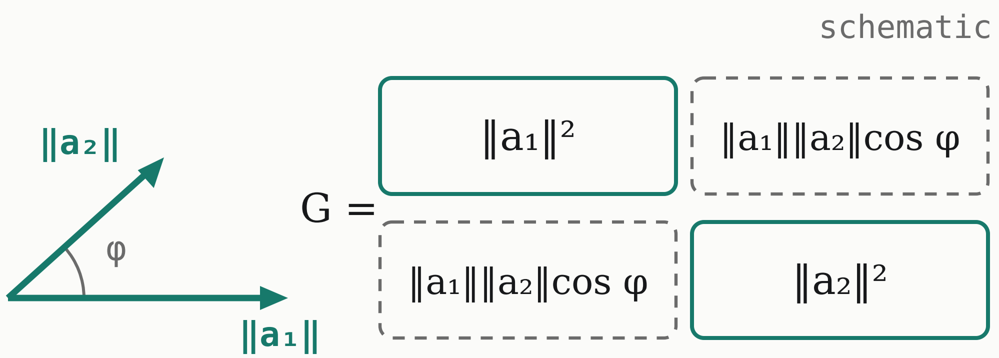
$$G = \begin{bmatrix} \boxed{a_1^\mathsf{T} a_1 \\ 1+1+1 = 3} & a_1^\mathsf{T} a_2 \\ 0+1+2 = 3 \\ a_2^\mathsf{T} a_1 \\ 3 & \boxed{a_2^\mathsf{T} a_2 \\ 0+1+4 = 5} \end{bmatrix}$$

$a_1^\mathsf{T} a_1$: $1+1+1 = 3$	$a_1^\mathsf{T} a_2$: $0+1+2 = 3$
$a_2^\mathsf{T} a_1$: 3	$a_2^\mathsf{T} a_2$: $0+1+4 = 5$

Boyd, S. and Vandenberghe, L. ***Introduction to Applied Linear Algebra***, least squares. Cambridge, 2018.

Least Squares

Reading the Entries



$$G_{ii} = \|a_i\|^2, \qquad G_{ij} = \|a_i\|\,\|a_j\|\cos\varphi$$

G_{ii} squared length of column i

G_{ij} lengths **and** angle at once (p8)

φ between a_i, a_j, both $\neq 0$

> A large $|G_{ij}|$ needs long columns **and** $|\cos\varphi|$ not small. $G = I$ exactly for orthonormal columns, singular exactly for dependent ones (p55).

Break

10 minutes

10 min

The Gram matrix records every dot product between two columns of A.

> Which equations does it give for the best-fit parameters?

For our three readings: $G = A^\top A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$

Least Squares

The Normal Equations Are Necessary

NECESSARY

If $\hat{x}$ minimises, $r = b - A\hat{x} \perp \mathrm{Col}(A)$ (p29)

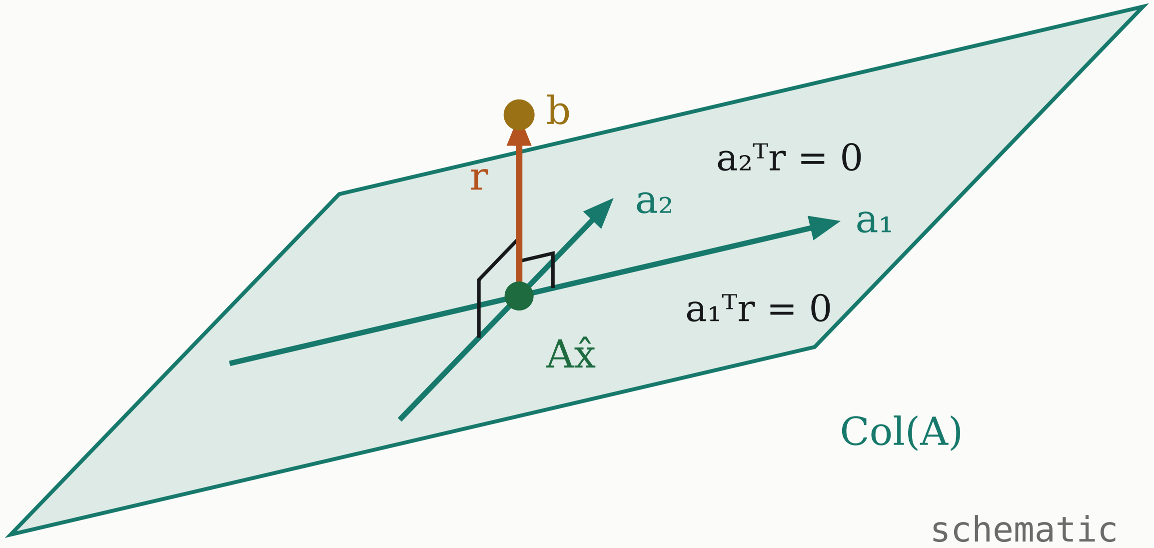
$\perp \mathrm{Col}(A) \Leftrightarrow a_i^\mathsf{T} r = 0$ for every column (p14)

so $A^\mathsf{T} r = 0$: $r \in \mathrm{Null}(A^\mathsf{T})$ (p15)

$$A^\mathsf{T}(b - A\hat{x}) = 0 \Rightarrow \boxed{A^\mathsf{T} A\hat{x} = A^\mathsf{T} b}$$

$A^\mathsf{T} A$ — $n \times n$, any A; $A^\mathsf{T} b \in \mathbb{R}^n$

necessary — every minimiser solves it; the converse: p50–51



Least Squares

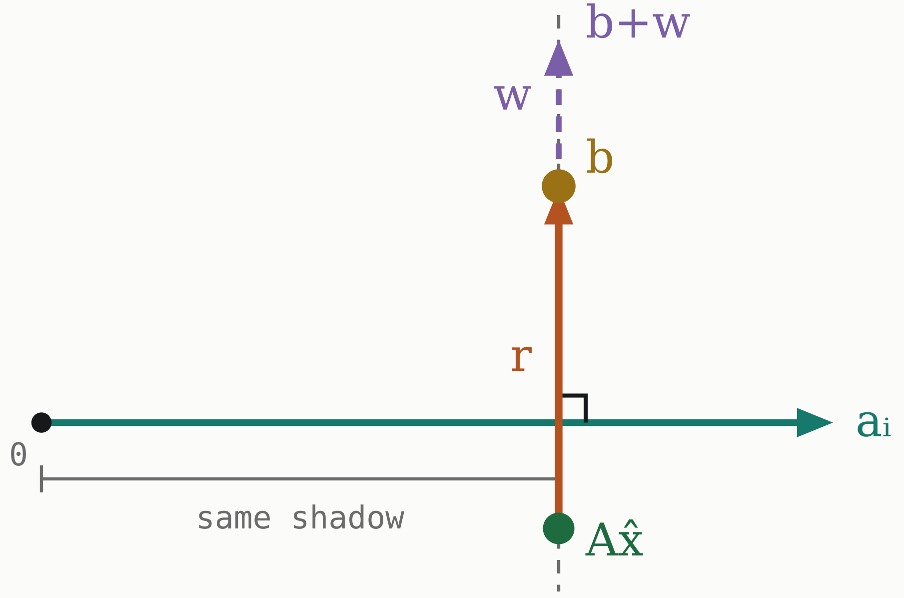
Reading the n Equations

$$a_i^\top(A\hat{x}) = \sum_j (a_i^\top a_j)\hat{x}_j = a_i^\top b$$

row i — what column i measures is matched; the leftover is invisible to it

$n \times n$ — n equations, n unknowns, any A; solvable: take $Ax = \hat{b}$; $A^\top(b - \hat{b}) = 0$ (p29); only one? p54

> Move b by any $w \in \mathrm{Null}(A^\top) = \mathrm{Col}(A)^\perp$ (p15): $A^\top b$, and the equations, do not change (p32).



schematic

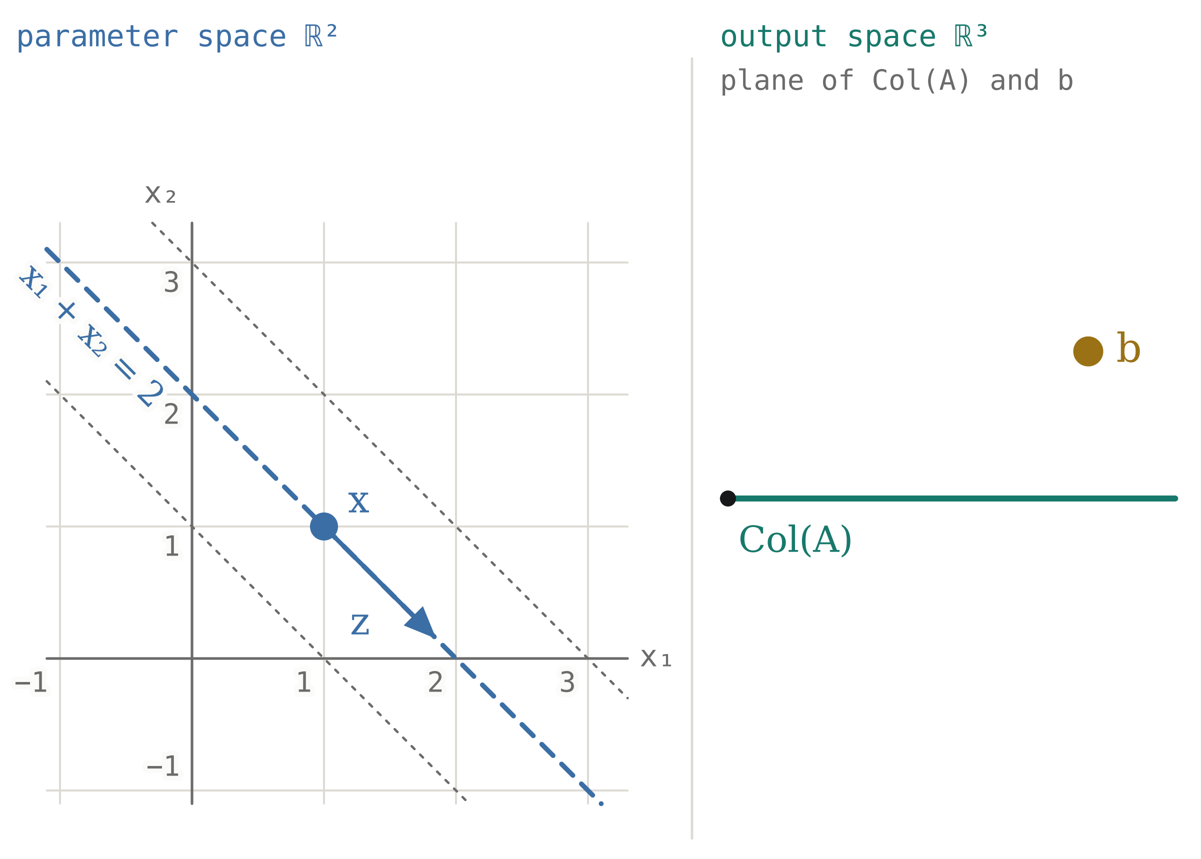
VOTE · ALL THAT APPLY

Least Squares

What Changes Along the Valley?

A is 3×2, all ones; $b = (1, 2, 3)$. Move x along $z = (1, -1)$. Which change?

- (A) parameters x
- (B) prediction Ax
- (C) residual $r = b - Ax$
- (D) objective $f = \|Ax - b\|^2$



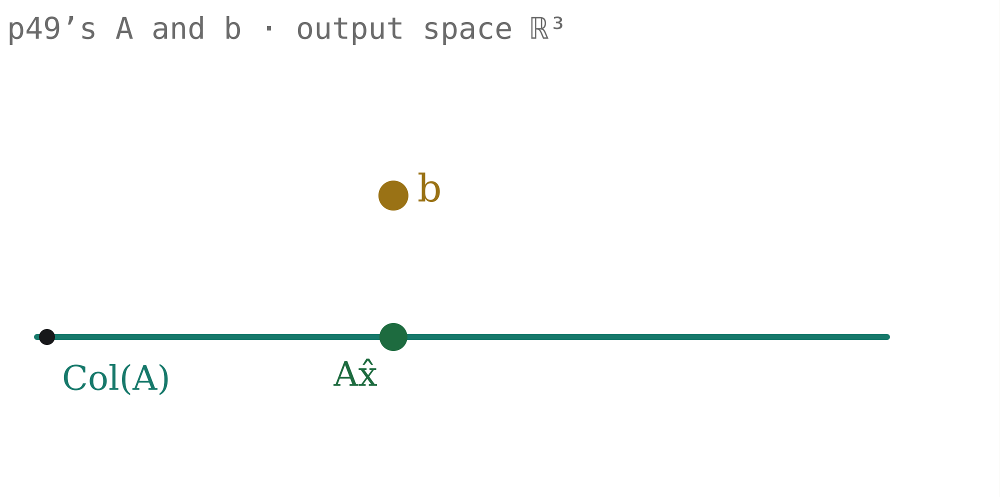
PREDICT

Least Squares

Are the Normal Equations Enough?

$\hat{x}$ solves $A^\top A\hat{x} = A^\top b$. For every A, which is guaranteed? $\hat{x}$ is…

- (A) a global minimiser
- (B) flat here, but lower somewhere else
- (C) the only minimiser



Least Squares

Why They Are Enough

SUFFICIENT

Take any other $\hat{x} + z$:

$$A(\hat{x} + z) - b = Az - r$$

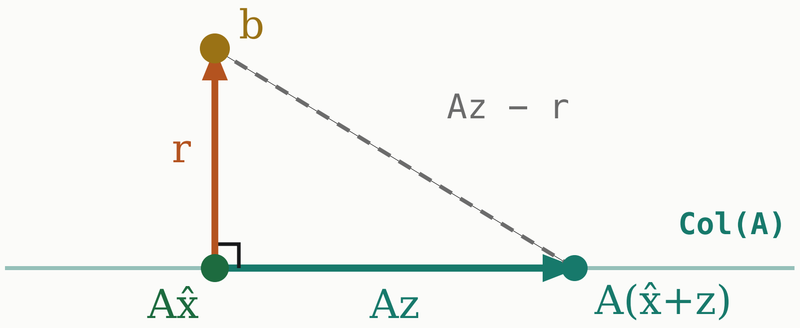
$z \in \mathbb{R}^n$ any step in parameter space; $r = b - A\hat{x}$

$$\|Az - r\|^2 = \|Az\|^2 - 2z^\mathsf{T}(A^\mathsf{T} r) + \|r\|^2$$

cross term $A^\mathsf{T} r = 0$ is the equation $\hat{x}$ solves, so it is 0

$$\|A(\hat{x} + z) - b\|^2 = \|r\|^2 + \|Az\|^2 \geq \|r\|^2$$

both squares global minimum; equality exactly when $Az = 0$



schematic

Least Squares

The Whole Set of Minimisers

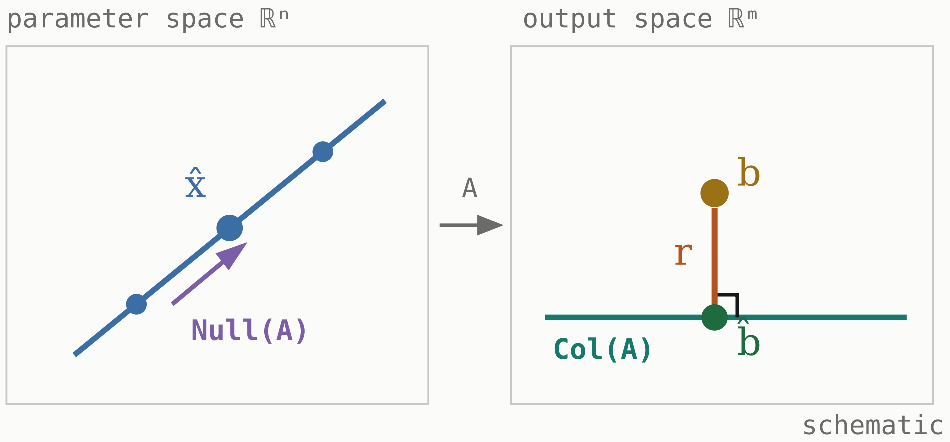
$$\boxed{\hat{x} \text{ minimises } \iff A^\mathsf{T} A\hat{x} = A^\mathsf{T} b}$$

`every` A $\Rightarrow$ p47 (necessary), $\Leftarrow$ p51 (enough); no rank assumption

$$\operatorname{argmin} f = \hat{x} + \mathrm{Null}(A)$$

`Lecture 1` $x_p + \mathcal{N}(A)$; all share one $\hat{b}$, one r

> p49: $\hat{x} = (1,1)$, $z = (1,0) \notin \mathrm{Null}(A)$: $Az = (1,1,1)$, $f = 2 + 3 = 5$.



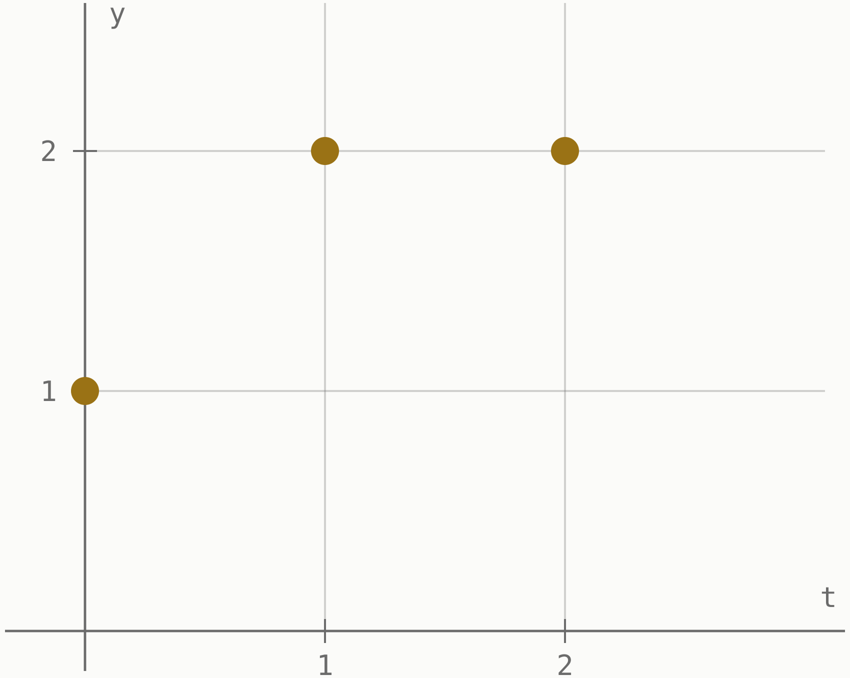
GROUPS · PREDICT THEN COMPUTE

$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$ $b = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$

Least Squares

Solve the Three-Point Fit

$G = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$, $A^\mathsf{T} b = (5, 6)$. Predict the signs of r, then find $\hat{x}, \hat{b}, r$.



Least Squares

When Is the Minimiser Unique?

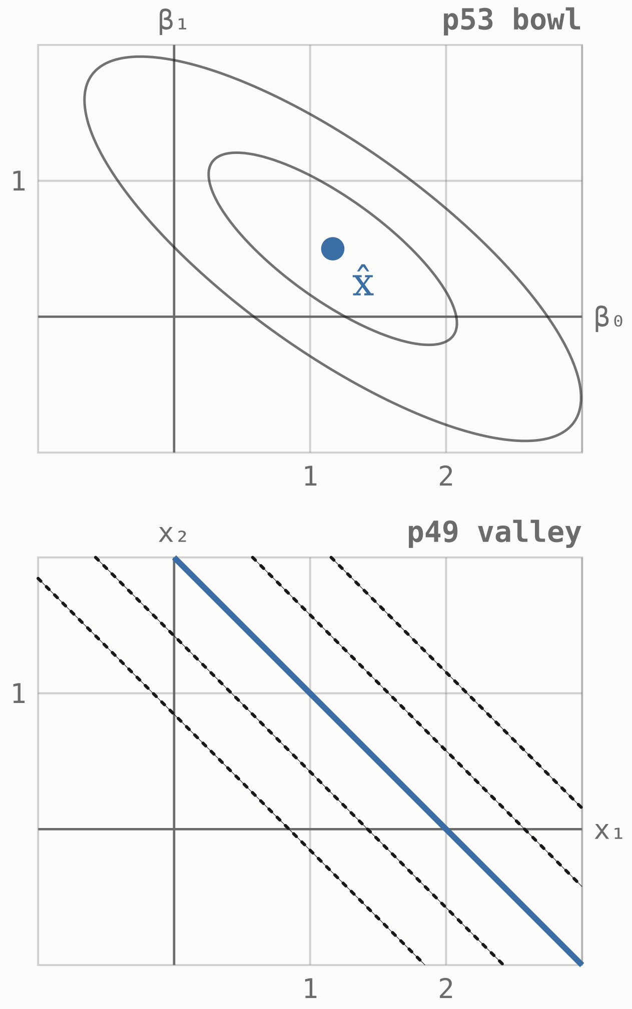
$$\hat{x} + \mathrm{Null}(A) = \{\hat{x}\} \iff \mathrm{Null}(A) = \{0\} \iff \mathrm{rank}\, A = n$$

argmin $= \hat{x}+\mathrm{Null}(A)$ (p52); one point ⇔ independent columns

rank-nullity dim $\mathrm{Null}(A) = n - \mathrm{rank}\, A$

$m \geq n$ from rank $A = n \leq m$; not assumed

> p53's fit: independent columns, one bottom. p49: a valley of bottoms, one $\hat{b}$.



Least Squares

Gram Positivity and the Equivalent Conditions

$$z^\mathsf{T} G z = z^\mathsf{T} A^\mathsf{T} A z = \|Az\|^2 \geq 0$$

$G = A^\mathsf{T} A$ symmetric $n \times n$, any A

> **Positive definite**: symmetric G, $z^\mathsf{T} G z > 0$ for $z \neq 0$.

$\mathrm{Null}(A) = \{0\} \iff G$ pos. definite — $\|Az\|^2 > 0$

$\iff G$ invertible — $Az = 0 \Rightarrow Gz = 0$, $Gz = 0 \Rightarrow Az = 0$: Null($G$) = Null($A$)

$\iff$ one $\hat{x}$ — argmin $= \hat{x}+\mathrm{Null}(A)$ (p52)

Fails: G singular, flat valley.

Boyd and Vandenberghe. ***Introduction to Applied Linear Algebra***, least squares. Cambridge, 2018.

Least Squares

Closed Forms Under Full Column Rank

RANK $A = n$ **(P54)**

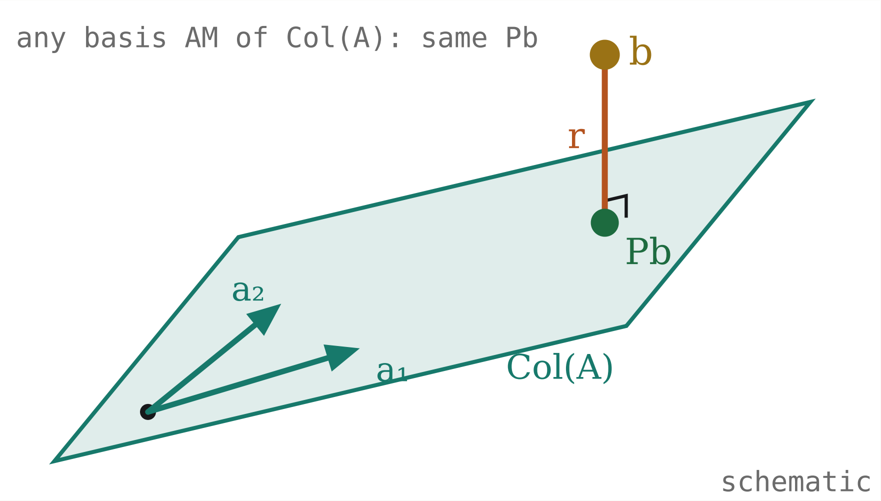
$$\hat{x} = (A^\mathsf{T} A)^{-1} A^\mathsf{T} b, \quad P = A(A^\mathsf{T} A)^{-1} A^\mathsf{T}$$

$\hat{x} \in \mathbb{R}^n$ — $A^\mathsf{T} A$ invertible (p55); in code, solve the system

P — $m \times m$, symmetric, rank n; rank $< n$: next lecture

$Pb \in \mathrm{Col}(A)$, $A^\mathsf{T}(b - Pb) = 0$: **the** projection (p29) — P depends only on $\mathrm{Col}(A)$.

$A = Q$ orthonormal: $Q^\mathsf{T} Q = I$, $P = QQ^\mathsf{T}$ (p30). Otherwise the Gram factor corrects the basis.



Gram–Schmidt Orthogonalization

Why We Want an Orthonormal Basis

ANY BASIS

$$P = A(A^\top A)^{-1} A^\top$$

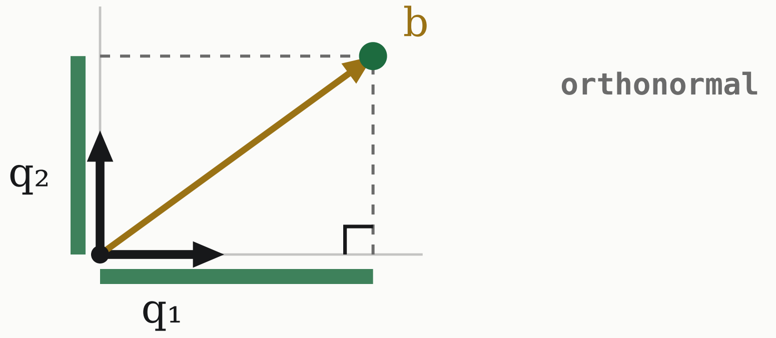
`cost` build $A^\top A$, then solve

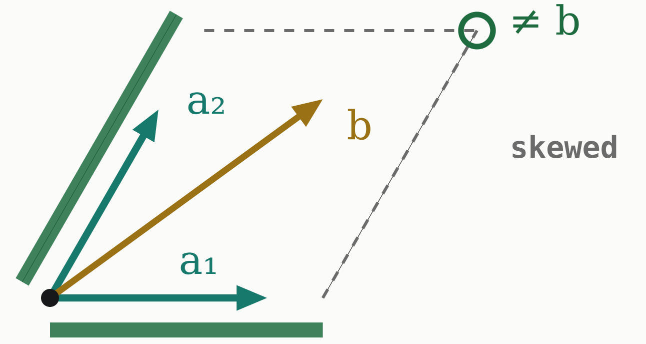
ORTHONORMAL BASIS

$$P = QQ^\top$$

$Q^\top b$ coordinates of $Pb = \sum_i q_i (q_i^\top b)$: one each

Can we build such a Q for the same $\mathrm{Col}(A)$?





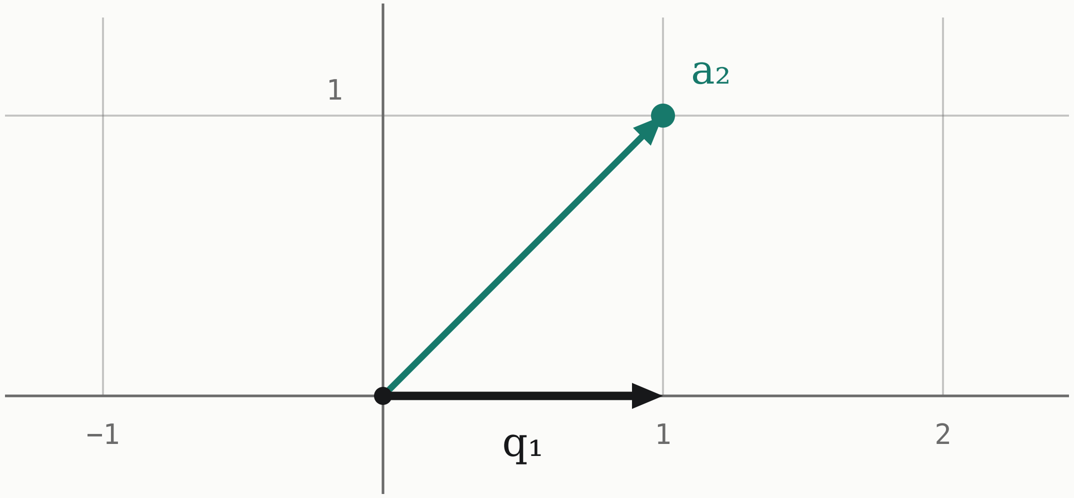
VOTE

Gram-Schmidt Orthogonalization

Which Subtraction Makes It Orthogonal?

$q_1 = (1, 0)$, $a_2 = (1, 1)$. Sketch each result: which removes exactly the part along q_1?

- (A) $a_2 - q_1(q_1^\top a_2)$
- (B) $a_2 - q_1\|a_2\|$
- (C) $a_2/\|a_2\|$
- (D) $a_2 + q_1(q_1^\top a_2)$



WORKED EXAMPLE

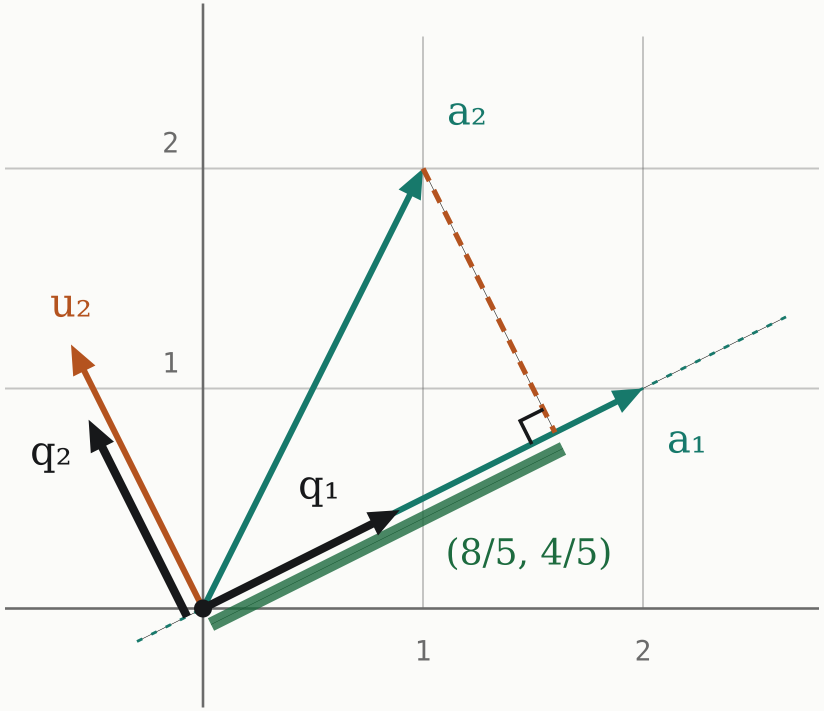
Gram–Schmidt Orthogonalization

Gram–Schmidt on Two Vectors

THE COLUMNS FROM LECTURE 2

$a_1 = (2, 1)$, $a_2 = (1, 2)$

NORMALISE	$\|a_1\| = \sqrt{5} \neq 0$, $q_1 = (2, 1)/\sqrt{5}$
COMPONENT	$q_1^{\mathsf{T}} a_2 = 4/\sqrt{5}$ (becomes r_{12})
REMAINDER	$u_2 = (1, 2) - (8/5, 4/5) = (-3/5,\ 6/5)$
NORMALISE	$\|u_2\| = 3/\sqrt{5} \neq 0$, $q_2 = (-1, 2)/\sqrt{5}$



DEFINITION

Gram–Schmidt Orthogonalization

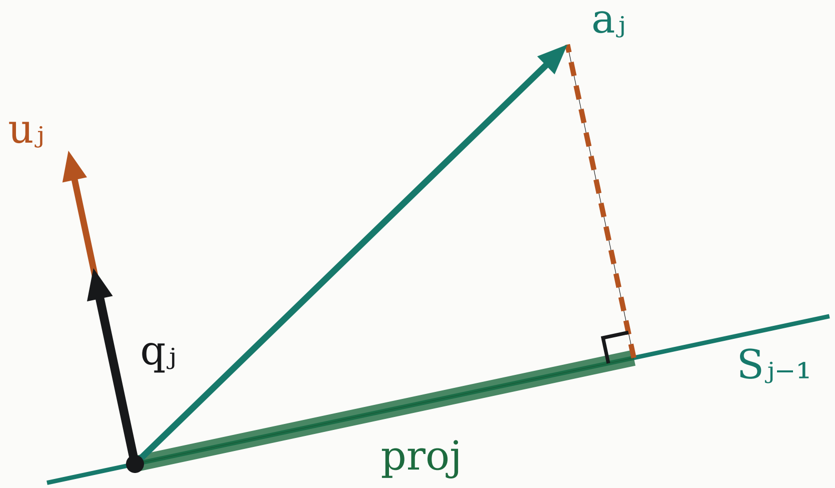
Keep the New Direction, Remove the Overlap

$$u_j = a_j - \sum_{i<j} q_i(q_i^\top a_j), \quad q_j = \frac{u_j}{\|u_j\|}$$

S_{j-1} span of the orthonormal $q_1, \ldots, q_{j-1}$; $j = 1$: empty sum, $S_0 = \{0\}$

$\sum$ $=$ projection onto S_{j-1} only because the q_i are orthonormal

$u_j \in \mathbb{R}^m$ remainder; dividing needs $u_j \neq 0$ ($j = 1$: $u_1 = a_1 \neq 0$)



Axler. ***Linear Algebra Done Right***, 4th ed., 6B. Springer, 2024.

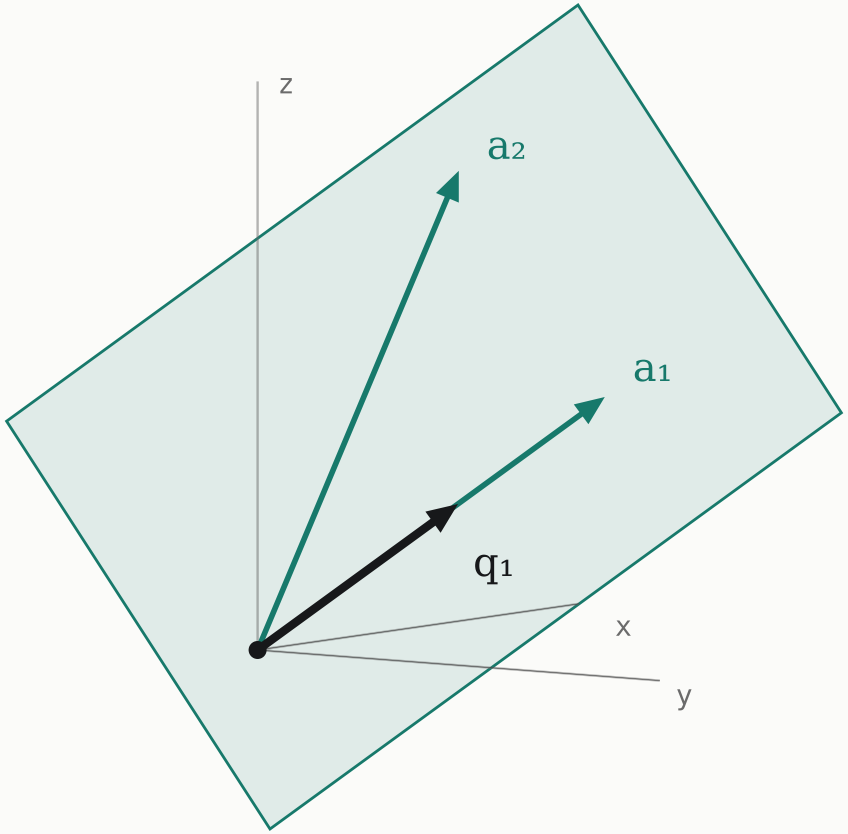
PAIR · PREDICT THEN COMPUTE

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

Gram-Schmidt Orthogonalization

Orthogonalize Our Fitting Matrix

$q_1 = (1,1,1)/\sqrt{3}$, $u_2 \in \operatorname{span}\{a_1, a_2\}$.
Predict its entries and the checks.



Gram–Schmidt Orthogonalization

The General Procedure

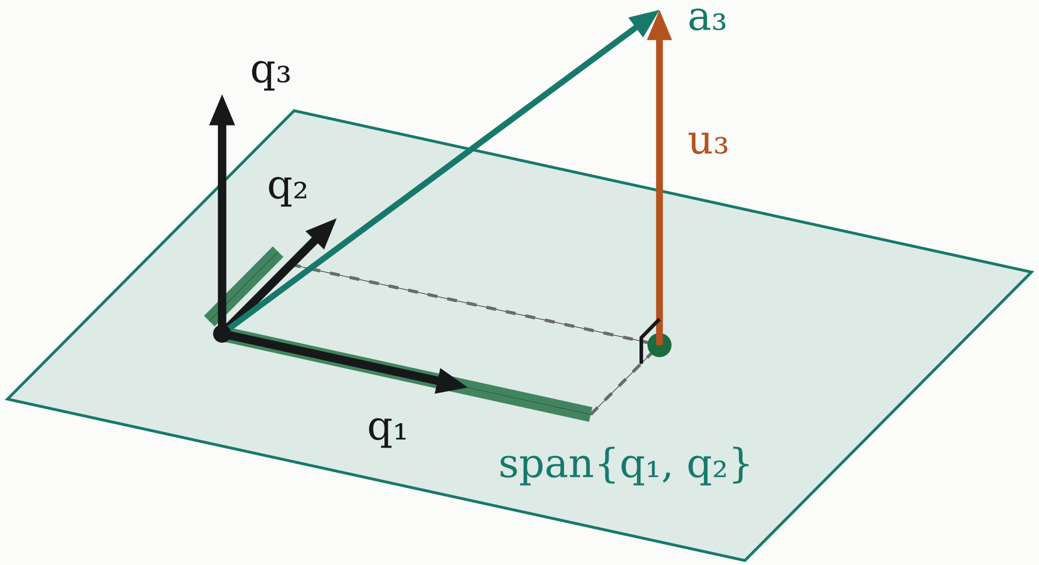
Scope: $a_1, \ldots, a_n$ independent — A of full column rank.

$$\text{for } j = 1, \ldots, n:$$
$$u_j = a_j - \sum_{i<j} q_i(q_i^\top a_j)$$
$$q_j = u_j / \|u_j\|$$

a_j column j of $A \in \mathbb{R}^{m \times n}$

$q_i^\top a_j$ coefficient on earlier q_i

Under that scope $\|u_j\| > 0$ at every step, so each division is legal.



Gram–Schmidt Orthogonalization

Why the Result Is Orthogonal

$$q_i^\top u_j = q_i^\top a_j - \sum_{l<j} (q_i^\top q_l)(q_l^\top a_j)$$

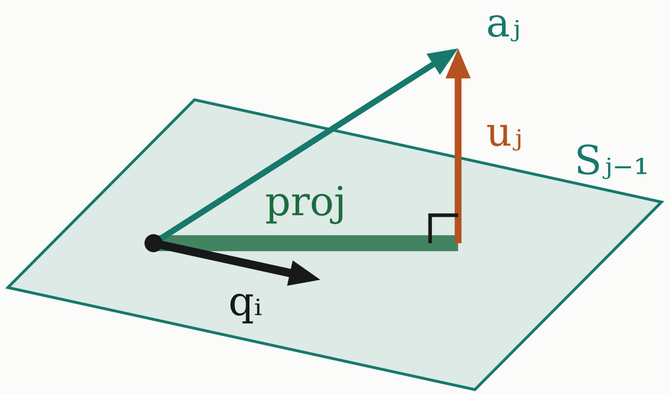
`assume` $q_1, \ldots, q_{j-1}$ already orthonormal — the induction

$i < j$ $q_i^\top q_l = 1$ at $l = i$, else 0: one term survives

u_j the remainder before normalising, in $\mathbb{R}^m$

$= q_i^\top a_j - q_i^\top a_j = 0$ for every $i < j$.

$q_j = u_j / \|u_j\|$, so $q_i^\top q_j = 0$ too: scaling cannot change a zero.



Gram–Schmidt Orthogonalization

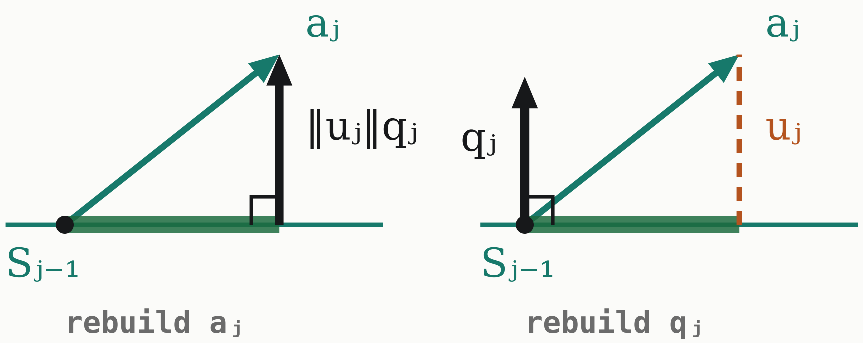
Why the Span Is Unchanged

After subtracting the projection, can we still rebuild a_j?

$$a_j = \sum_{i<j} (q_i^\top a_j) q_i + \|u_j\| q_j$$

$\|u_j\|$ > 0 by the scope

> span$\{q_1..q_j\}$ =span$\{a_1..a_j\}$
> $\supseteq$: this identity at each step $i \le j$
> $\subseteq$: solve for q_j: from a_j and earlier q's (induction)



PREDICT

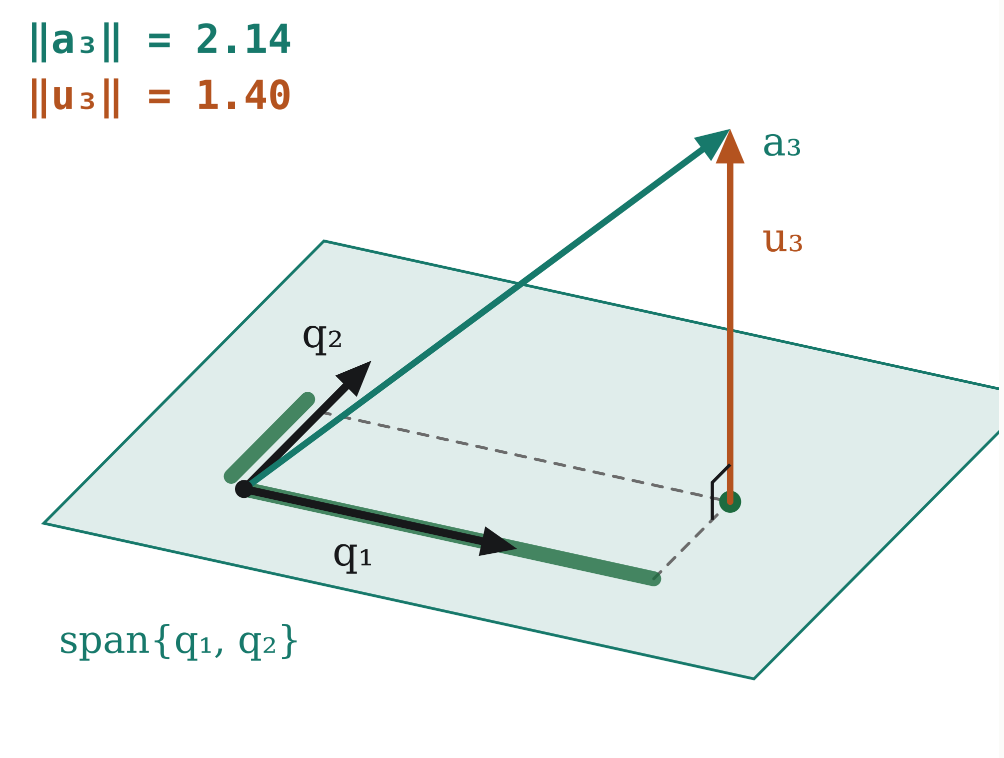
Gram-Schmidt Orthogonalization

What If Nothing Remains?

Lower a_3 into the plane: which hits 0, a_3 or u_3?

$u_3 = 0$ means:

- (A) $a_3 \perp q_1, q_2$
- (B) $a_3 \in \text{span}\{q_1, q_2\}$
- (C) $a_3 = 0$
- (D) q_1, q_2 invalid



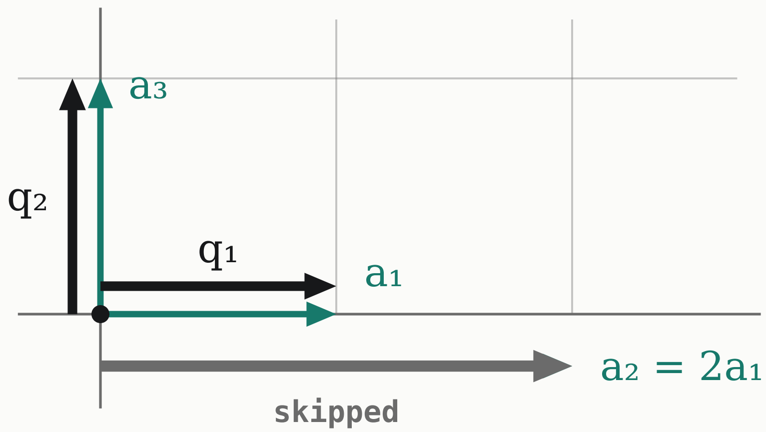
Gram–Schmidt Orthogonalization

What the Procedure Does Not Do

Why we could divide: $u_j = 0$ would write a_j from the earlier q's, which independence forbids.

Without it, halting at $u_j = 0$ is **not** basis extraction: $[e_1, 2e_1, e_2]$ stops at $j = 2$, missing e_2.

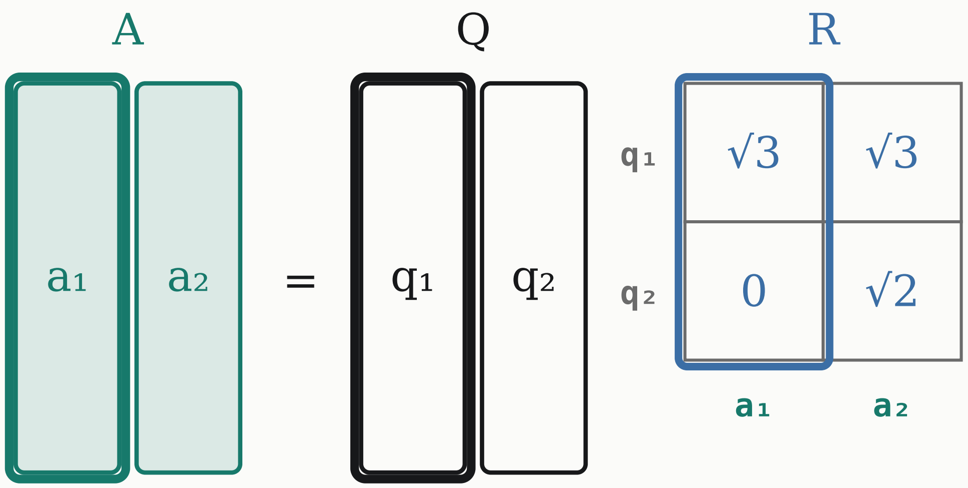
> Skip it, carry on in the original order, index the kept q's.
> Pivoting instead reorders columns: $A\Pi = QR$ (Lecture 4).



$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad Q = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix}$$

QR Factorization and Least Squares

Record the Coefficients: $A = QR$



$$a_j = \sum_{i \le j} r_{ij} q_i$$

$r_{ij} \quad = q_i^\top a_j,\ i < j$

$r_{jj} \quad = q_j^\top a_j = \|u_j\| > 0$: columns independent

$$Q^\top A = (Q^\top Q)R = R$$

$Q^\top Q \quad = I_n$ (p19)

$R \quad$ column j: a_j in the q-basis

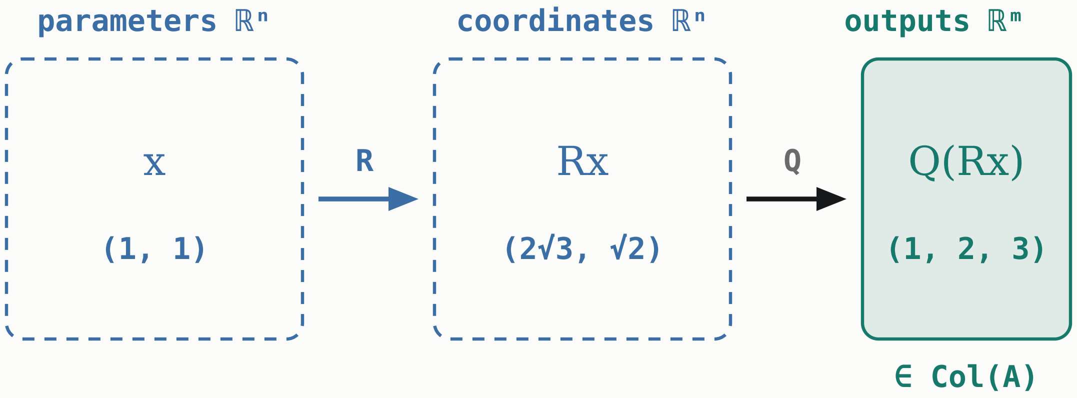
WORKED EXAMPLE

$q_1^\top a_2 = \sqrt{3}$, $\|u_2\| = \sqrt{2}$; check $\sqrt{3}q_1 + \sqrt{2}q_2 = a_2$

$R = \begin{bmatrix} \sqrt{3} & \sqrt{3} \\ 0 & \sqrt{2} \end{bmatrix} \quad G = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$

QR Factorization and Least Squares

What $Ax = Q(Rx)$ Means



$$Ax = Q(Rx)$$

x parameters, $\in \mathbb{R}^n$

Rx Ax in the q-basis, also $\in \mathbb{R}^n$

Q keeps lengths (p19); R need not

$$R^\top R = A^\top A = G$$

always $R^\top R = (QR)^\top QR$

here $G \neq I$: R is not a rotation

FILL IN

QR Factorization and Least Squares

Why Is R Upper Triangular?

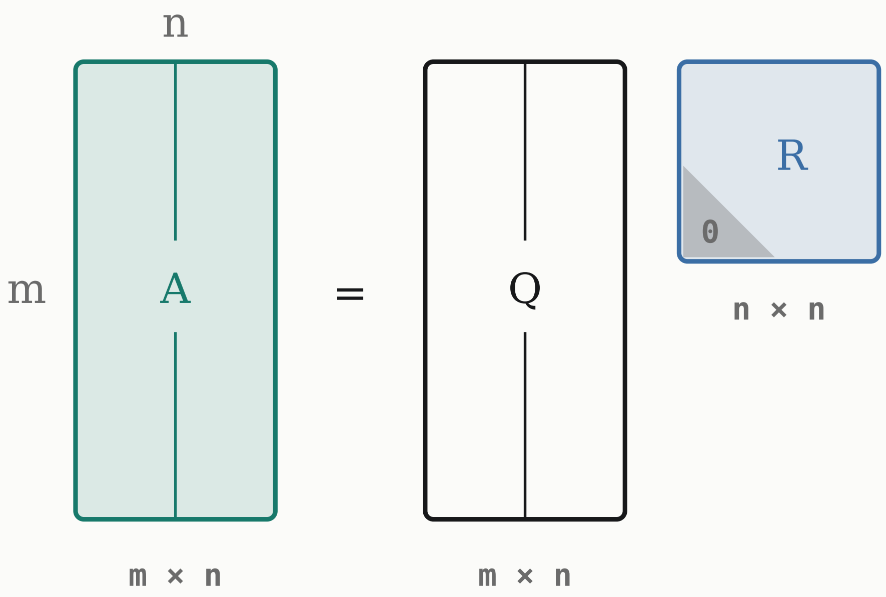
	a_1	a_2	a_3
q_1			
q_2			
q_3			

a_j uses only $q_1, \ldots, q_j$ (p67). Why is $r_{ij} = q_i^\top a_j = 0$ for $i > j$?

- (A) The columns of A are orthogonal
- (B) $q_i \perp q_1, \ldots, q_j$, so $\perp$ their span
- (C) Orthonormal Q forces triangular R
- (D) Only when A is square

QR Factorization and Least Squares

Read the Shapes Before the Entries



Here $m \geq n$ and A has full column rank.

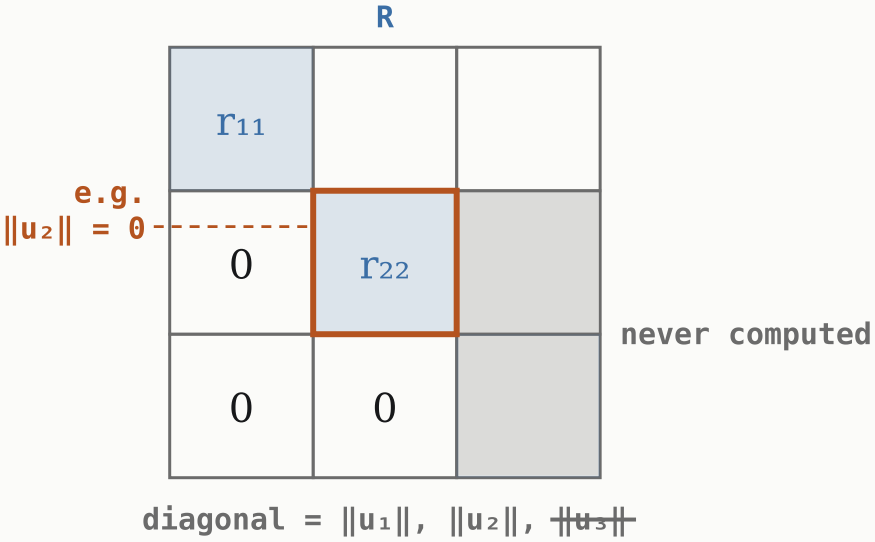
$$Q \in \mathbb{R}^{m\times n},\ R \in \mathbb{R}^{n\times n},\ Q^{\mathsf{T}}Q = I_n$$

thin	Q is as tall as A (p19)
shapes	$(m \times n)(n \times n) = (m \times n)$
QQ^{T}	$\neq I_m$ once $n < m$ (p18)

> We assume full column rank so this construction cannot break down and R is invertible.

QR Factorization and Least Squares

Why R Is Invertible



$$\det R = r_{11} \cdots r_{nn} \neq 0$$

$r_{jj} > 0$ $r_{jj} = \|u_j\|$, and $u_j = 0$ would put a_j in the span of the earlier columns (p66)

$\neq 0$ triangular; ours $\sqrt{3}\sqrt{2} = \sqrt{6}$

> Rank $< n$: Gram–Schmidt stalls; any R has an $r_{jj} = 0$, so $\det R = 0$ and the solve divides by zero.

QR Factorization and Least Squares

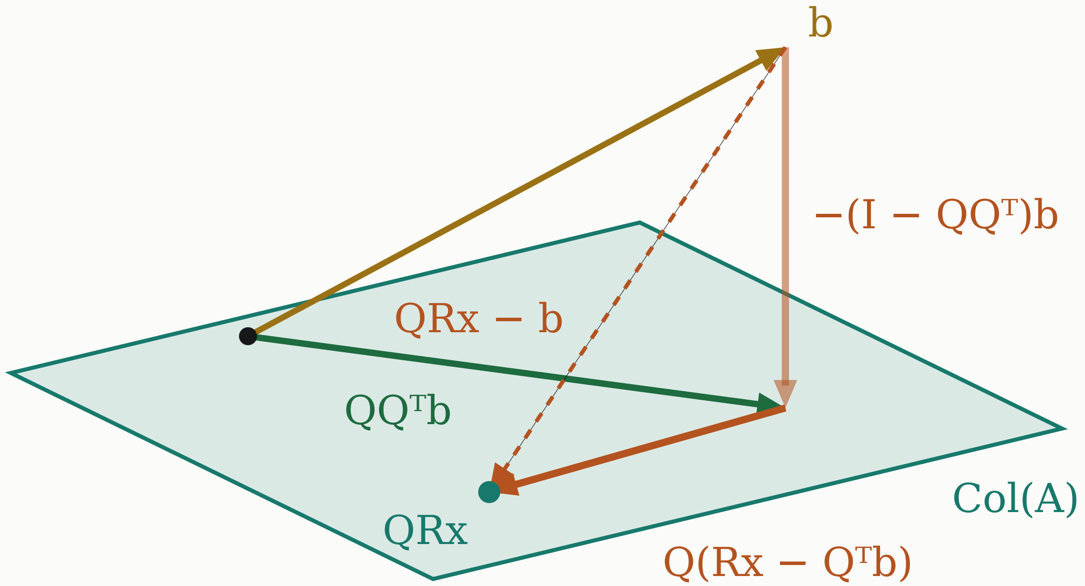
QR Is Not Diagonalization

	$A = V\Lambda V^{-1}$	$A = QR$
DIRECTIONS	eigenvectors	orthonormal columns
MIDDLE	diagonal	upper triangular
USE	modes, powers	outputs, least squares
NEEDS SQUARE A	yes	no

Is q_1 an eigenvector of A?

QR Factorization and Least Squares

Split the Error in the New Coordinates



Throughout $r = b - Ax$, so $QRx - b = -r$ and $\|QRx - b\| = \|r\|$: the same error, drawn from b to the prediction.

> Do not start from $QRx = b$: generally false.

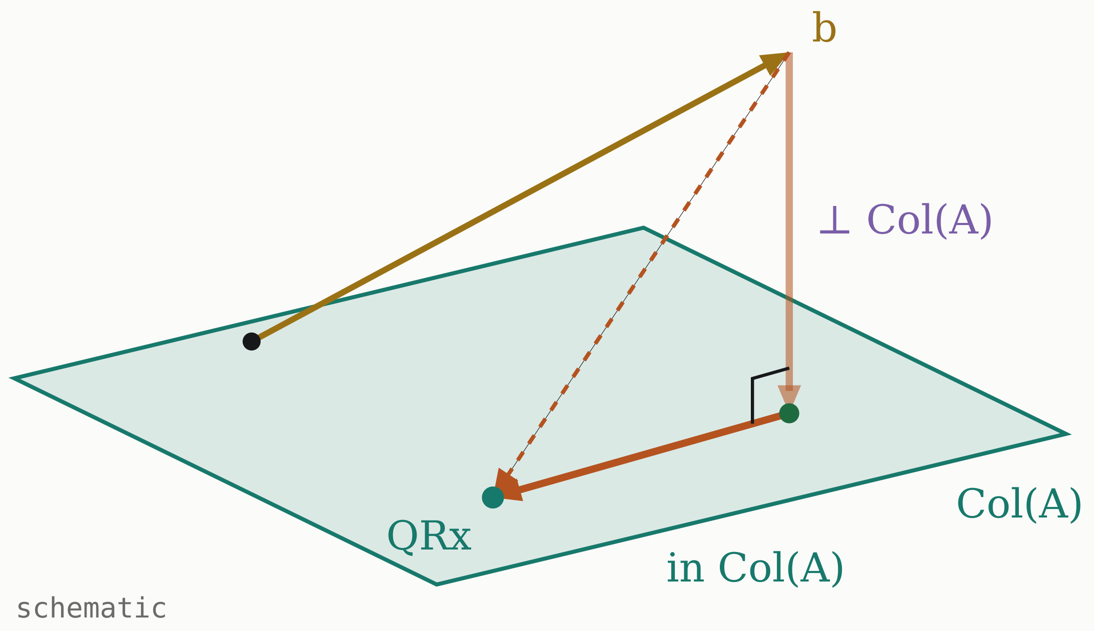
$$QRx - b = Q(Rx - Q^\top b) - (I - QQ^\top)b$$

check $= QRx - QQ^\top b - b + QQ^\top b$

goal $\min_x \|r\|^2$

QR Factorization and Least Squares

Why the Two Parts Are Perpendicular



The two pieces of $QRx - b$:

$$Q(Rx - Q^\top b) \in \mathrm{Col}(Q) = \mathrm{Col}(A)$$

$=$ the span was kept

$$Q^\top(I - QQ^\top)b = Q^\top b - Q^\top b = 0$$

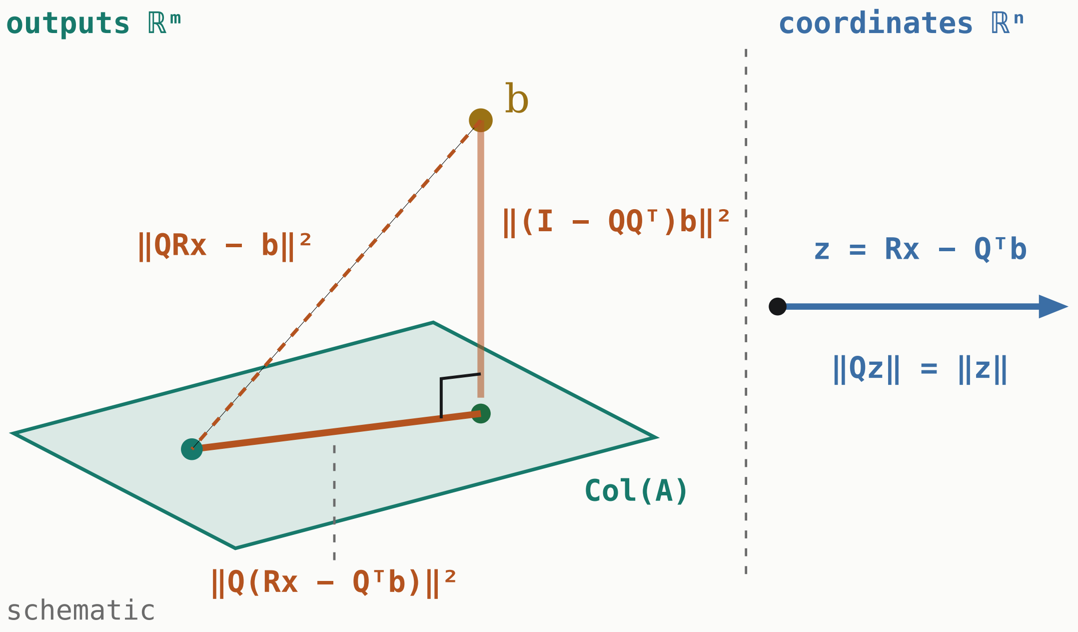
$\Rightarrow$ $\perp$ every q_i, so $\perp \mathrm{Col}(A)$

$(I - QQ^\top)b$ $= (I - P)b$, the projection's residual; drawn leg $-(I - P)b$

> The same perpendicular-error decomposition, in the same colours.

QR Factorization and Least Squares

Pythagoras in the New Coordinates



The two parts are perpendicular (p74):

$$\|QRx - b\|^2 = \|Q(Rx - Q^\top b)\|^2 + \|(I - QQ^\top)b\|^2$$

$+$ cross term $= 0$: squares add (p11)

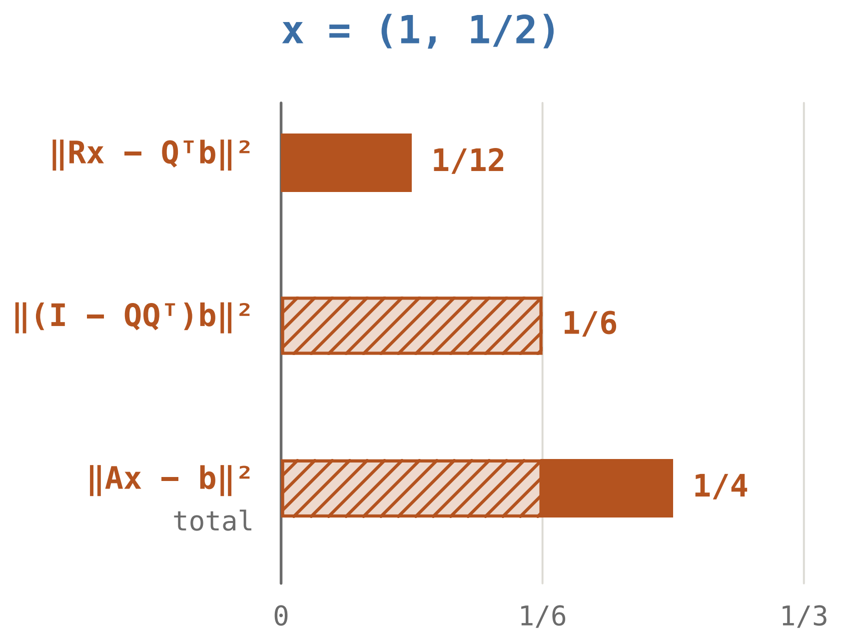
$$\|Ax - b\|^2 = \|Rx - Q^\top b\|^2 + \|(I - QQ^\top)b\|^2$$

$z = Rx - Q^\top b$ $\|Qz\|^2 = z^\top Q^\top Qz = \|z\|^2$ (p19)

QR Factorization and Least Squares

Adjustable and Unavoidable Error

$b = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$



At the solution, which part must be zero?

$$\|Ax - b\|^2 = \|Rx - Q^\top b\|^2 + \|(I - QQ^\top)b\|^2$$

PAIR

$b = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$

QR Factorization and Least Squares

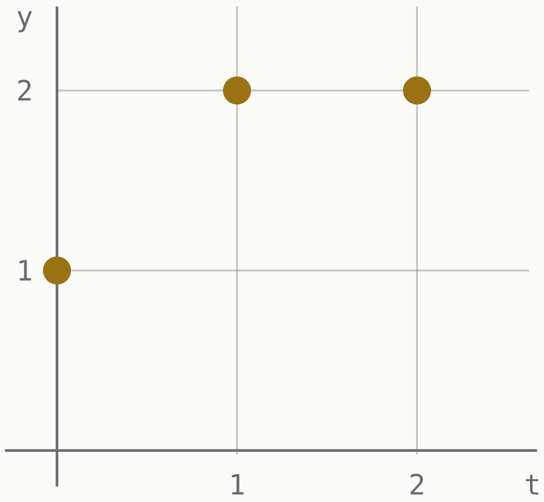
Solve the Same Fit with QR

Which unknown comes first, and why?

$$\begin{bmatrix} \sqrt{3} & \sqrt{3} \\ 0 & \sqrt{2} \end{bmatrix} \hat{x} = \begin{bmatrix} 5/\sqrt{3} \\ 1/\sqrt{2} \end{bmatrix}$$

$\hat{x}$ (β_0, β_1)

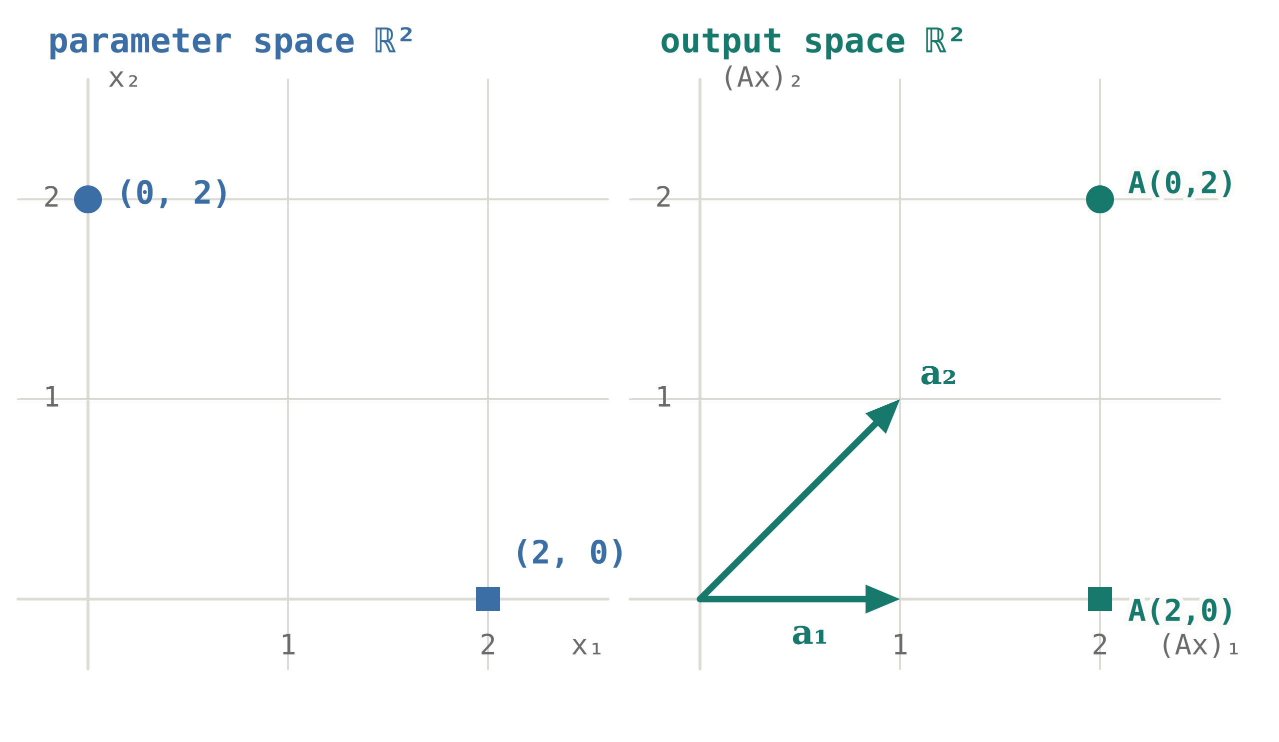
R $r_{jj} > 0$



PREDICT

QR Factorization and Least Squares

Very Different Parameters, Almost the Same Prediction



What becomes hard when we infer parameters from measurements?

QR Factorization and Least Squares

Largest Stretch over Smallest Stretch

$$\kappa_2(A) = \frac{\max_{\|z\|=1} \|Az\|_2}{\min_{\|z\|=1} \|Az\|_2}$$

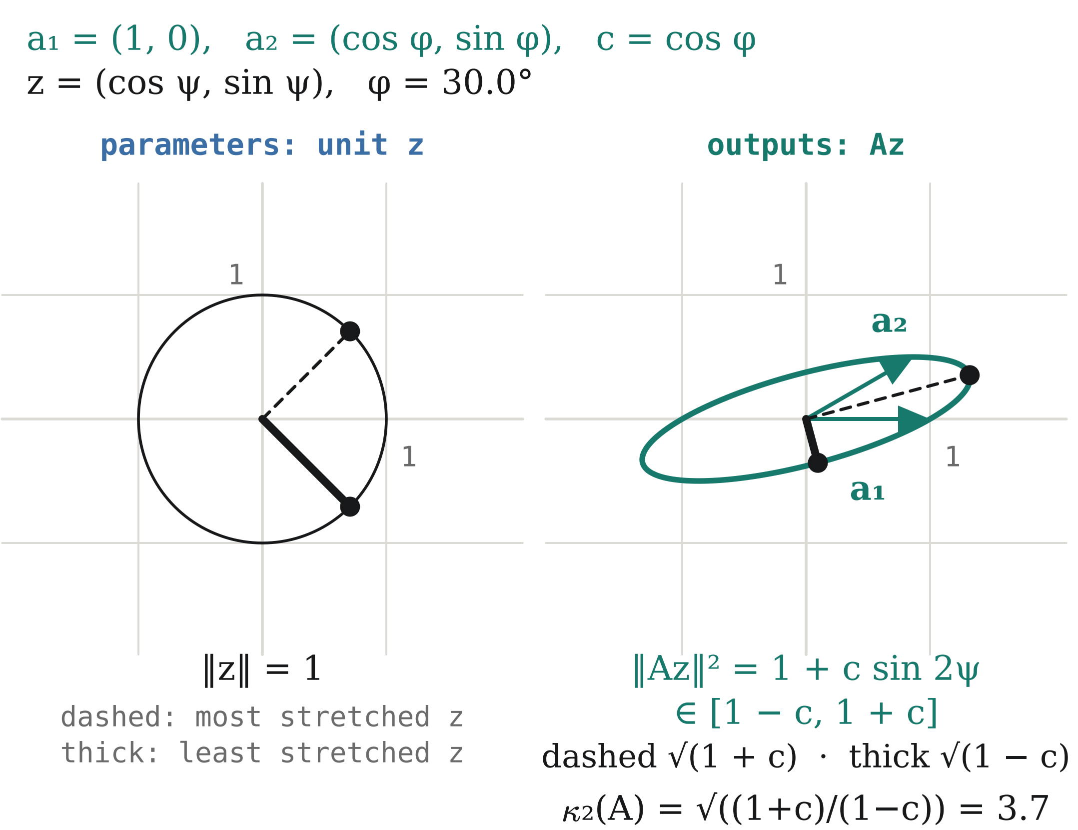
A $m \times n$, $m \geq n$

z parameter direction, $\|z\|_2 = 1$

The second lecture measured how a matrix stretches directions; the new object is the ratio.

min $> 0 \Leftrightarrow \mathrm{Null}(A) = \{0\}$: no unit z is flattened.

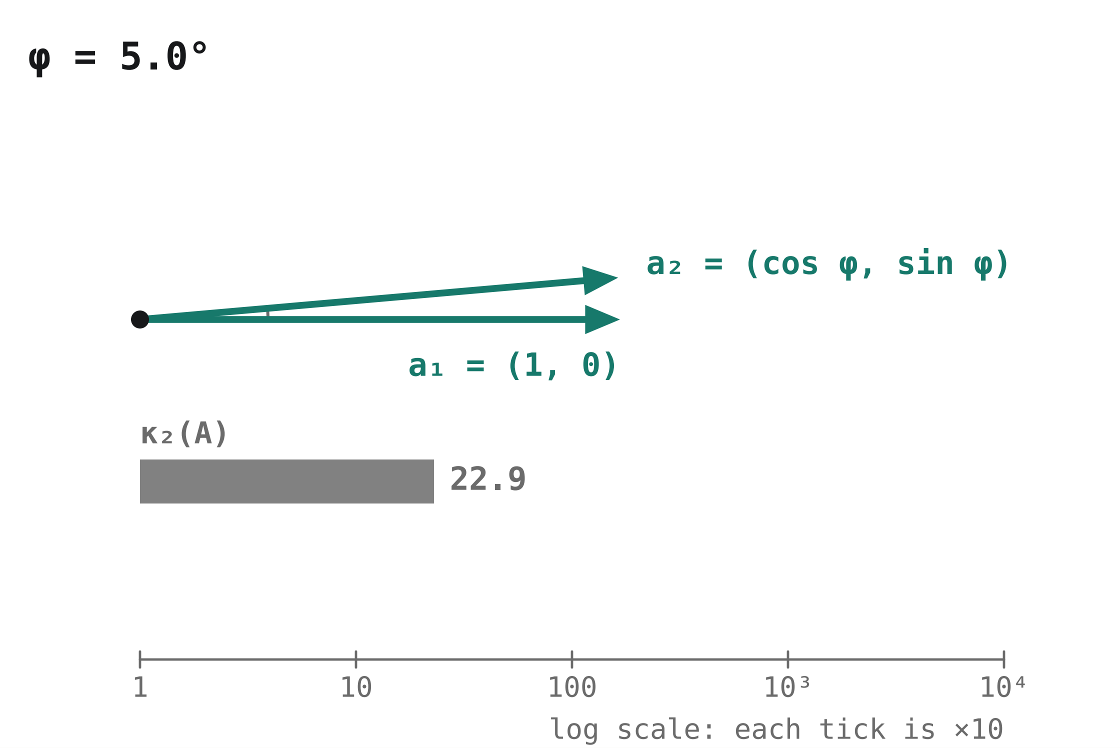
Rank lost: some unit z has $Az = 0$, so min $= 0$ and $\kappa_2(A) = \infty$ by convention.



DISCUSS

QR Factorization and Least Squares

Which Computational Route Would You Choose?



Which is correct?

- (A) QR cures the sensitivity
- (B) Forming $A^\top A$ can hurt the numerics
- (C) Normal equations fail even exactly
- (D) Every QR code is equally stable

QR Factorization and Least Squares

Squaring the Ratio

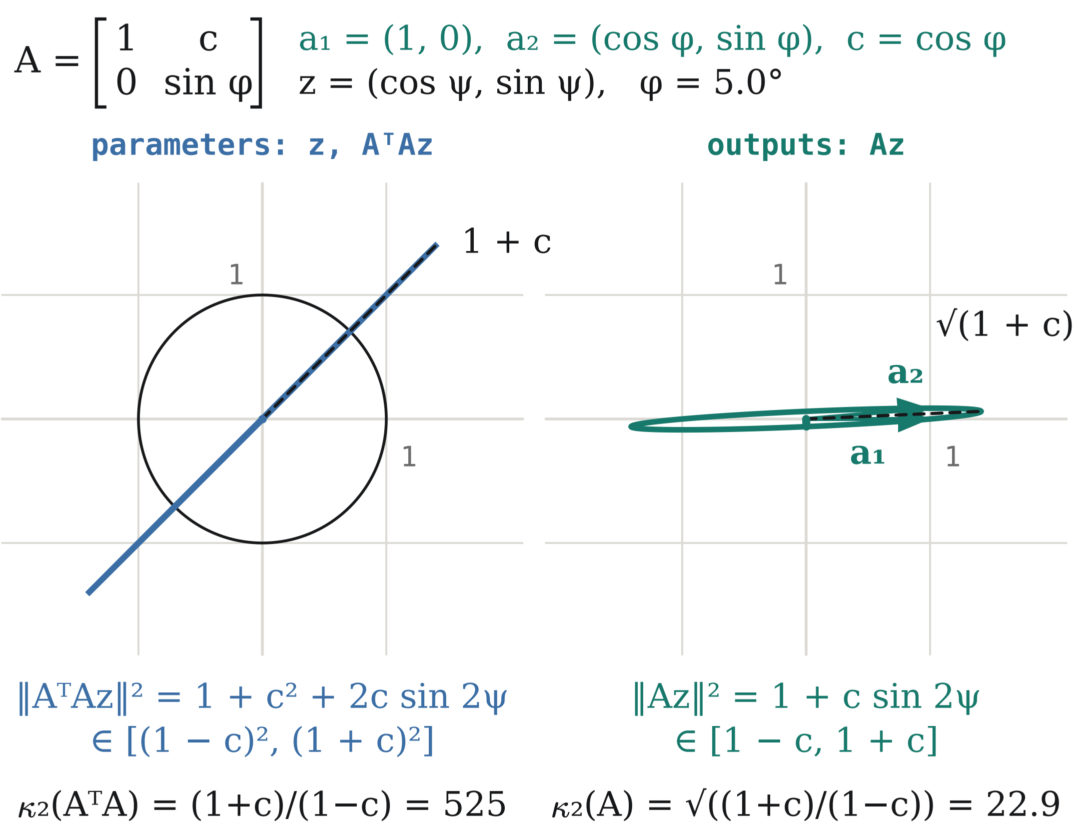
$$\kappa_2(R) = \kappa_2(A), \qquad \kappa_2(A^\top A) = \kappa_2(A)^2$$

R $\|Rz\| = \|QRz\| = \|Az\|$, Q an isometry

$A^\top A$ full column rank; proved next lecture

The triangular coefficient matrix has A's stretch ratio; the normal equations square it.

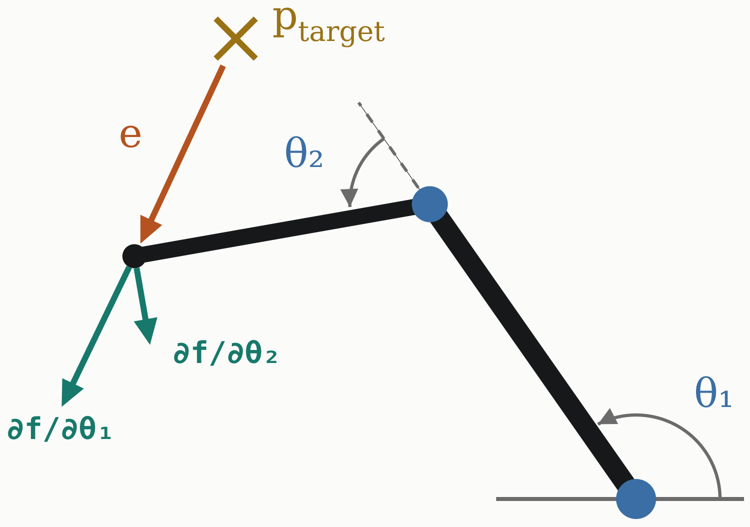
> Rules of thumb, not bounds: about $\log_{10} \kappa_2$ digits lost; sensitivity also depends on the residual, so $\kappa_2(A)$ alone does not settle it.



Trefethen, L. N. and Bau, D. ***Numerical Linear Algebra***, Lecture 18 conditioning of least squares problems. SIAM, 1997.

Robotics Application

The Local Model



$$e(\theta) = f(\theta) - p_{\text{target}} \in \mathbb{R}^2$$

$$e(\theta + \Delta\theta) \approx e(\theta) + J(\theta)\Delta\theta$$

f	tip position, nonlinear in θ
$\theta = (\theta_1, \theta_2)$	joint angles in radians; $\Delta\theta$ small
$J(\theta)$	$\partial f / \partial \theta$; column j: tip velocity per unit joint rate
$J\Delta\theta$	a displacement, not a velocity

> Near a singularity J can be ill-conditioned; at a singularity it loses rank.

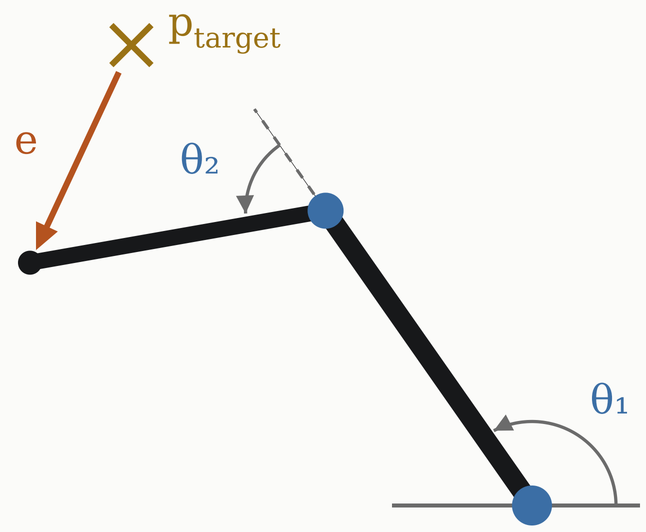
$\det J = l_1 l_2 \sin\theta_2$, link lengths $l_1, l_2 > 0$

Lynch, K. M. and Park, F. C. ***Modern Robotics***, 5.3 singularities and 6.2 numerical inverse kinematics. Cambridge, 2017.

TRANSFER

Robotics Application

One Least-Squares Step



$A =$ | $x =$ | $b =$

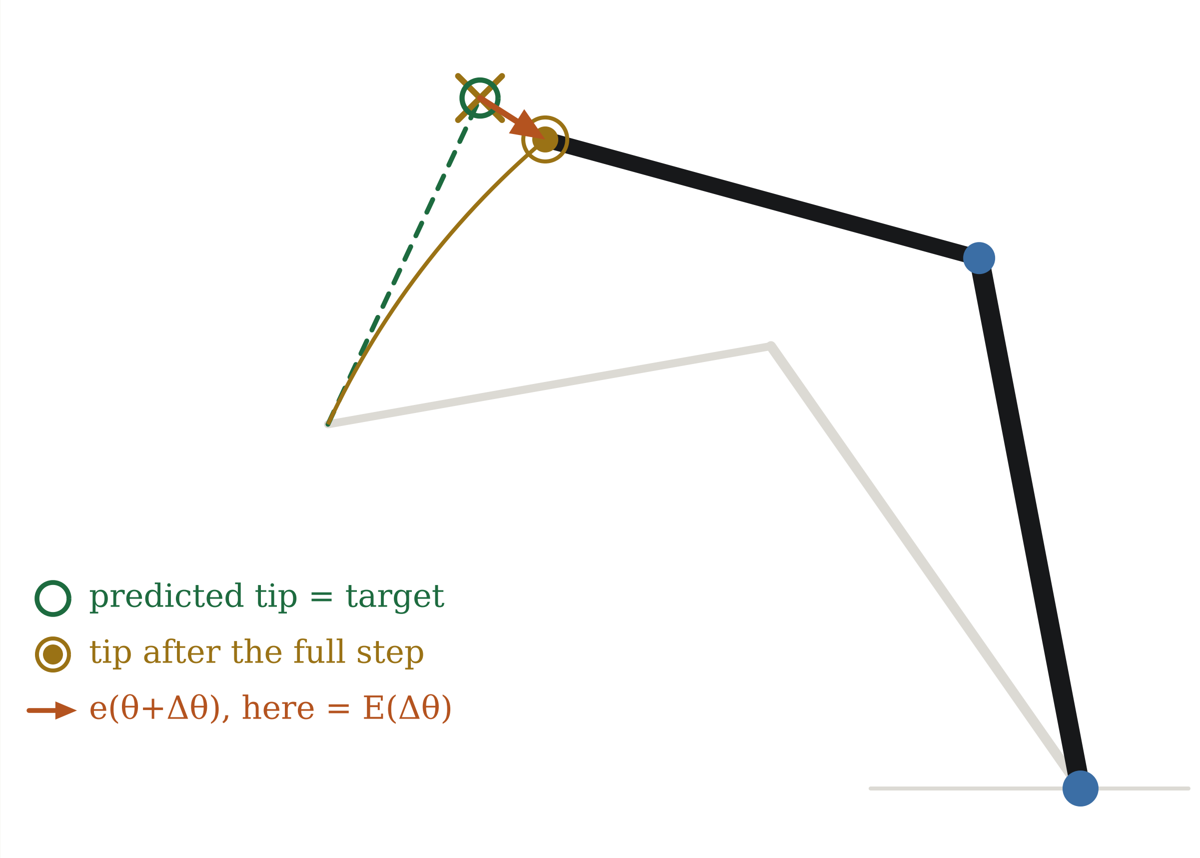
$$\min_{\Delta\theta} \|J\Delta\theta + e\|^2$$

$J\Delta\theta + e$ the model's error after the step

J full column rank here

Robotics Application

Predicted Motion and Actual Motion



$$e(\theta) + J\Delta\theta = 0 \ \text{ but } \ e(\theta + \Delta\theta) \neq 0$$

in this example

$e(\theta) + J\Delta\theta$	the model's error after the step; 0 here
$e(\theta + \Delta\theta)$	the true error after moving
$E(\Delta\theta)$	their difference: model discrepancy, $O(\|\Delta\theta\|^2)$

> Zero linearised error need not mean zero true error.

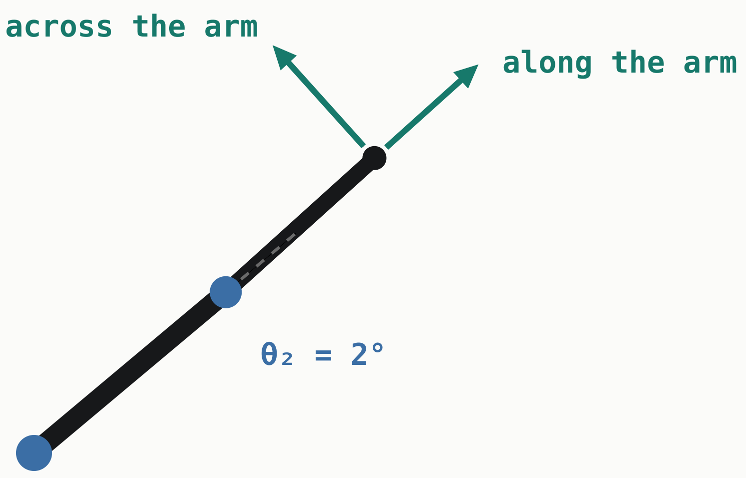
Linearise, step, measure, relinearise, repeat — each step small enough that the linear model stays accurate.

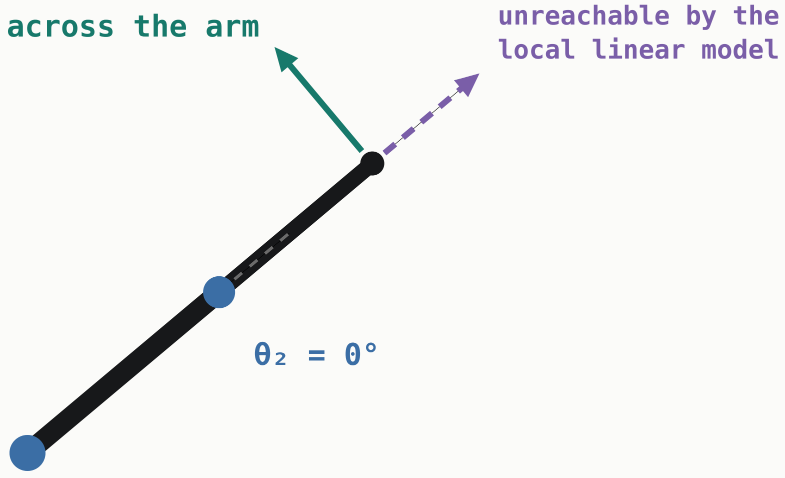
Robotics Application

Ill-Conditioned or Singular?

$$\det J = l_1 l_2 \sin\theta_2$$

l_1, l_2 link lengths; θ_2 elbow angle





$\det J \neq 0$: full rank, R invertible

QR runs; step size depends on e's direction

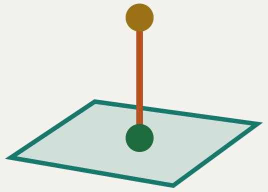
$\det J = 0$: rank 1, $r_{22} = 0$

Ordinary solve fails, yet minimisers exist; no joint rate gives along-arm tip velocity

Synthesis and Exit Ticket

One Problem, Four Views

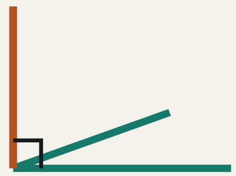
$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ $b = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$



GEOMETRY `any A`

$\hat{b} = \mathrm{proj}_{\mathrm{Col}(A)}\ b$

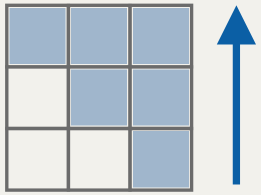
the closest reachable output



ALGEBRA `any A`

$A^{\top} A\hat{x} = A^{\top} b$

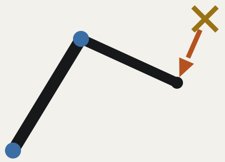
the equations every minimiser satisfies



COMPUTATION `full column rank`

$A = QR$

the minimiser satisfies $R\hat{x} = Q^{\top} b$



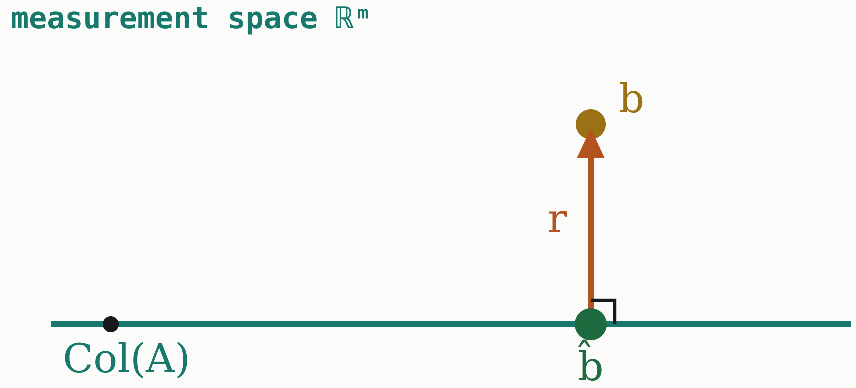
ROBOTICS `full column rank · local model`

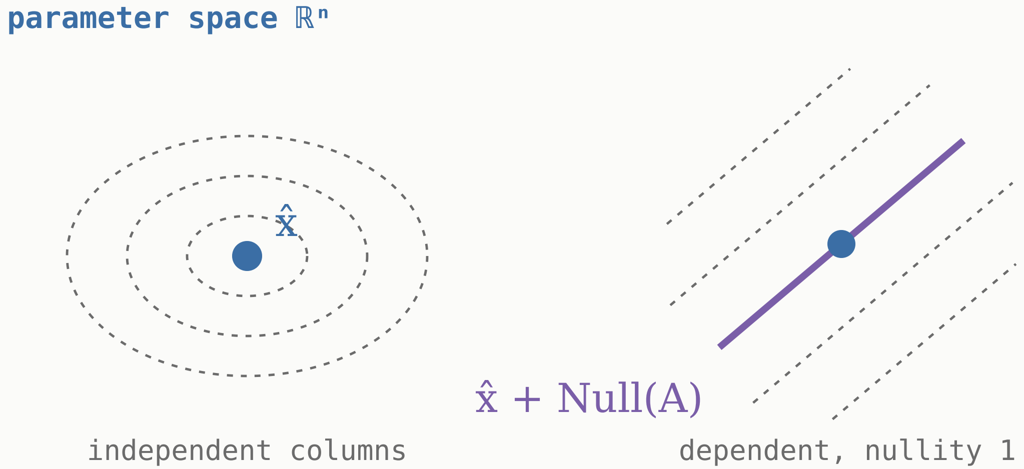
$R\Delta\theta = -Q^{\top} e$

one step, then relinearise

Synthesis and Exit Ticket

One Best Output, Sometimes Many Parameter Vectors





Every A and b: one closest $\hat{b}$, one $r \perp \mathrm{Col}(A)$.

$\{x : Ax = \hat{b}\} = \hat{x} + \mathcal{N}(A)$: one point exactly when the columns are independent.

Projection: no assumption. $R\hat{x} = Q^\top b$ and $(A^\top A)^{-1}$: full column rank.

PAIR

Synthesis and Exit Ticket

Find the Incorrect Claim

Which claim is false? Sketch Col(A) and b so it fails.

- **A** $r = b - A\hat{x}$ is $\perp$ each column of A.
- **B** If $b \in \text{Col}(A)$, the minimum residual is 0.
- **C** A nonzero minimum residual forces a non-unique $\hat{x}$.
- **D** Full column rank makes $\hat{x}$ unique.

full column rank?

b ∈ Col(A)?	yes	no
yes		
no		

EXIT TICKET · ALONE

Synthesis and Exit Ticket

Exit Ticket

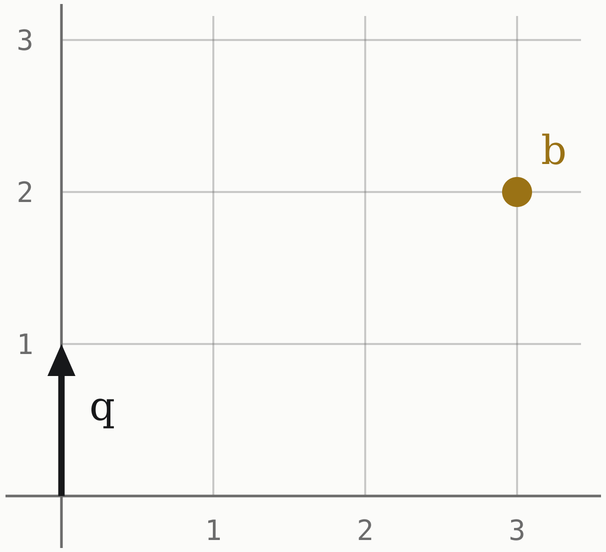
1. $q = (0, 1)$, $b = (3, 2)$. Projection p? Residual r?

2. Why does $A^\top r = 0$ not force $r = 0$? **One sentence of reasoning.**

3. Thin QR, full rank: which part of $\|QRx - b\|^2$ can no x change?

Hand in before the answers.

> Nearly dependent or rank-deficient columns: next lecture.



Courses and books

1. Axler, S. *Linear Algebra Done Right*, 4e, 6A–6C. Springer, 2024.
2. Strang, G. *18.06 Linear Algebra*, L14–L17. MIT OCW, 2010.
3. Trefethen & Bau. *Numerical Linear Algebra*, Lec. 6–8, 10–11, 18–19. SIAM, 1997.
4. Boyd & Vandenberghe. *Introduction to Applied Linear Algebra*, 5.4, 10.4, 12–13, 18. Cambridge, 2018.
5. Lynch & Park. *Modern Robotics*, 5.3, 6.2. Cambridge, 2017.