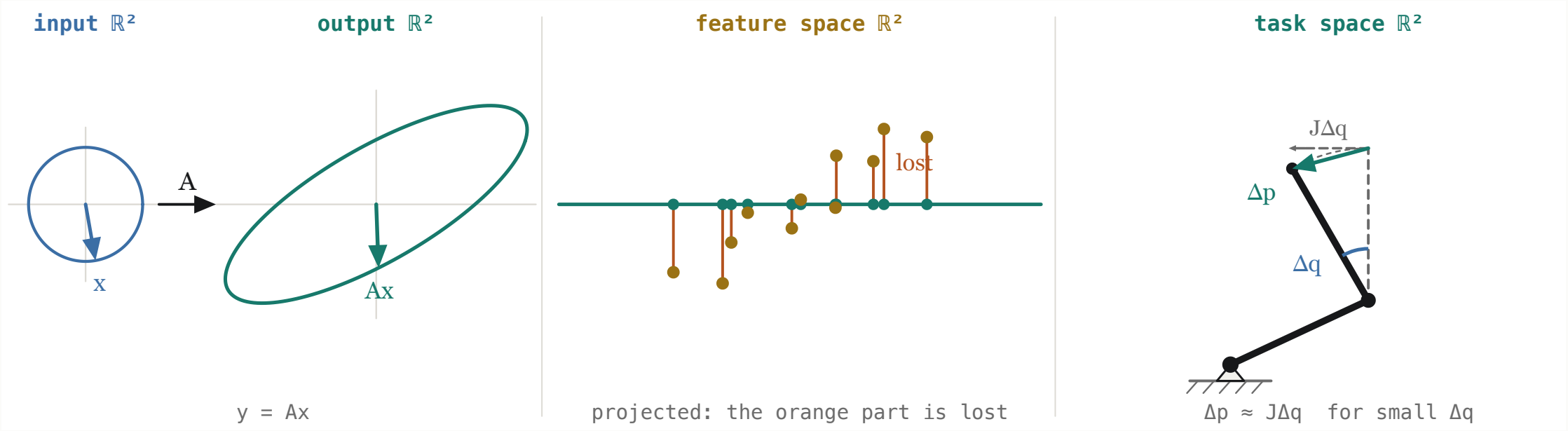
From Projections to Matrix Structure

PSD, SVD, PCA, Pseudoinverse, Jacobian and Hessian



$$y = Ax, \qquad \Delta p \approx J\,\Delta q$$

$A \in \mathbb{R}^{m\times n}$ input $x \in \mathbb{R}^n$ to output $y \in \mathbb{R}^m$

J joint change to tip change; $\approx$ needs small Δq

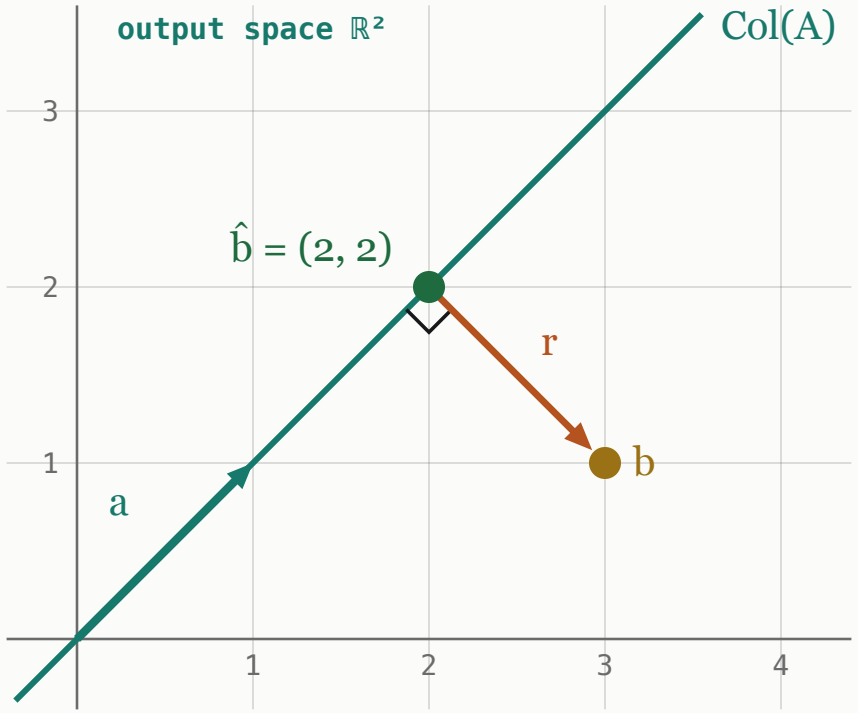
> Which directions matter? What can we recover?

RECALL

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

From Projections to Matrix Structure

Least Squares: Nearest Reachable Output



$$\hat{b} = A\hat{x}, \qquad r = b - \hat{b}, \qquad A^{\mathsf{T}} r = 0$$

- $\hat{b}$ nearest reachable output
- r residual, $\hat{b} \to b$
- $A^{\mathsf{T}} r = 0$ $r \perp$ every column of A

WORKED EXAMPLE

$$A = [\,a\,] \in \mathbb{R}^{2\times 1},\ \hat{x} \in \mathbb{R}\text{: } \hat{x} = \frac{a^{\mathsf{T}} b}{a^{\mathsf{T}} a} = \frac{4}{2} = 2,\ r = b - 2a = (1, -1).$$

Tuples denote columns.

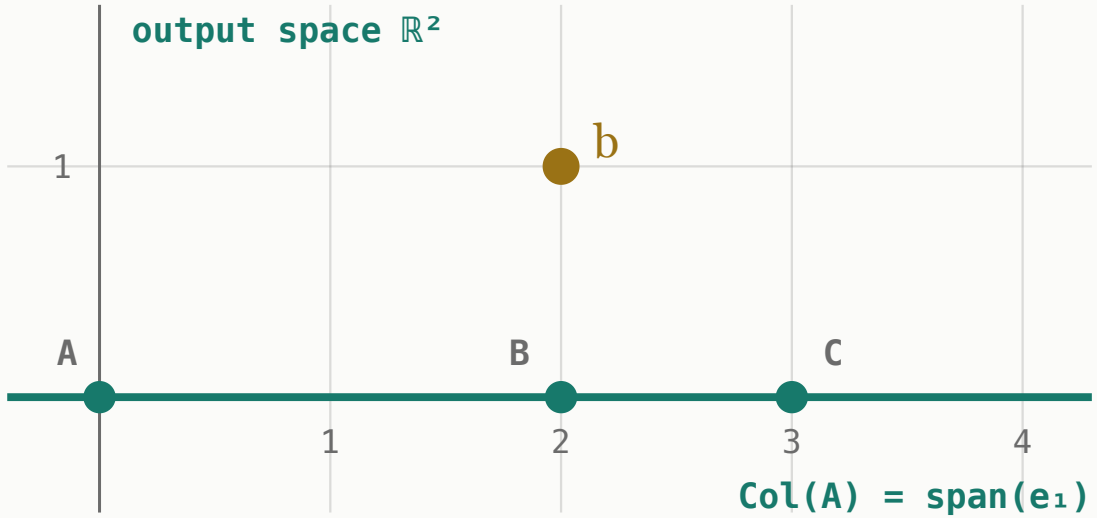
VOTE

From Projections to Matrix Structure

Which Candidate Output Is Closest?

$\mathrm{Col}(A) = \mathrm{span}(e_1)$ and $b = (2, 1)$. Which candidate output is closest to b?

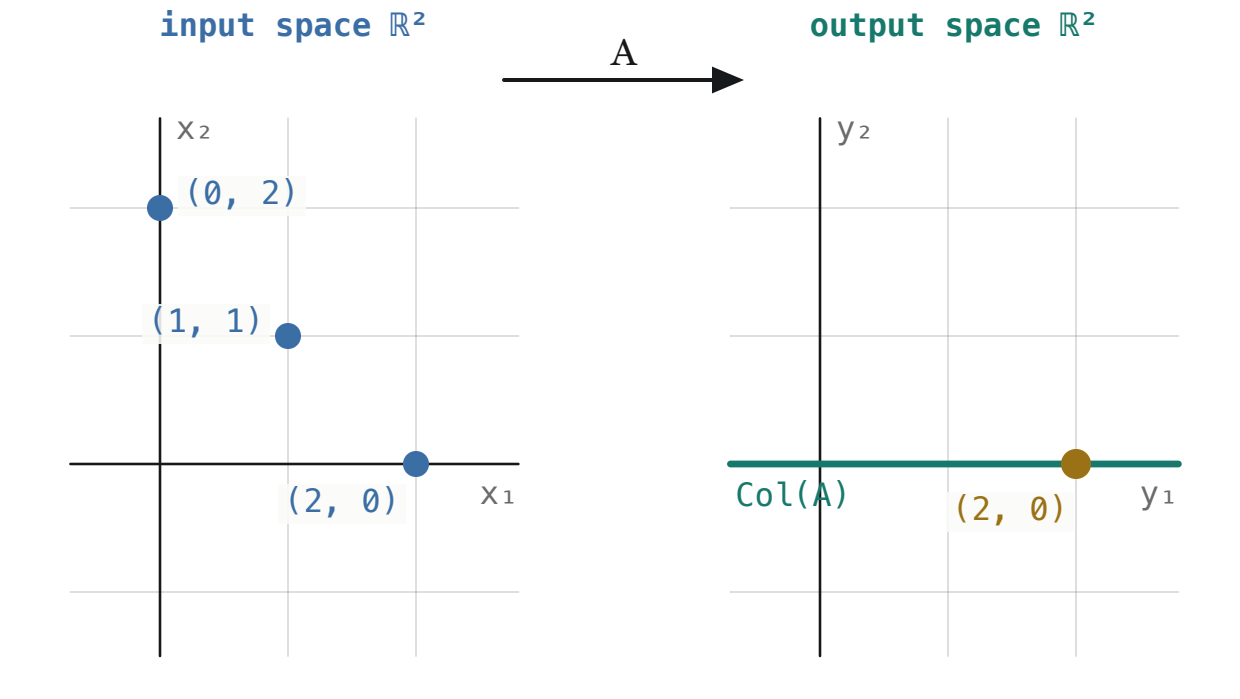
- **A** $(0, 0)$
- **B** $(2, 0)$
- **C** $(3, 0)$



VOTE

From Projections to Matrix Structure

Which Inputs Produce This Output?



$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. Which inputs give the output $(2, 0)$?

A only $(1, 1)$

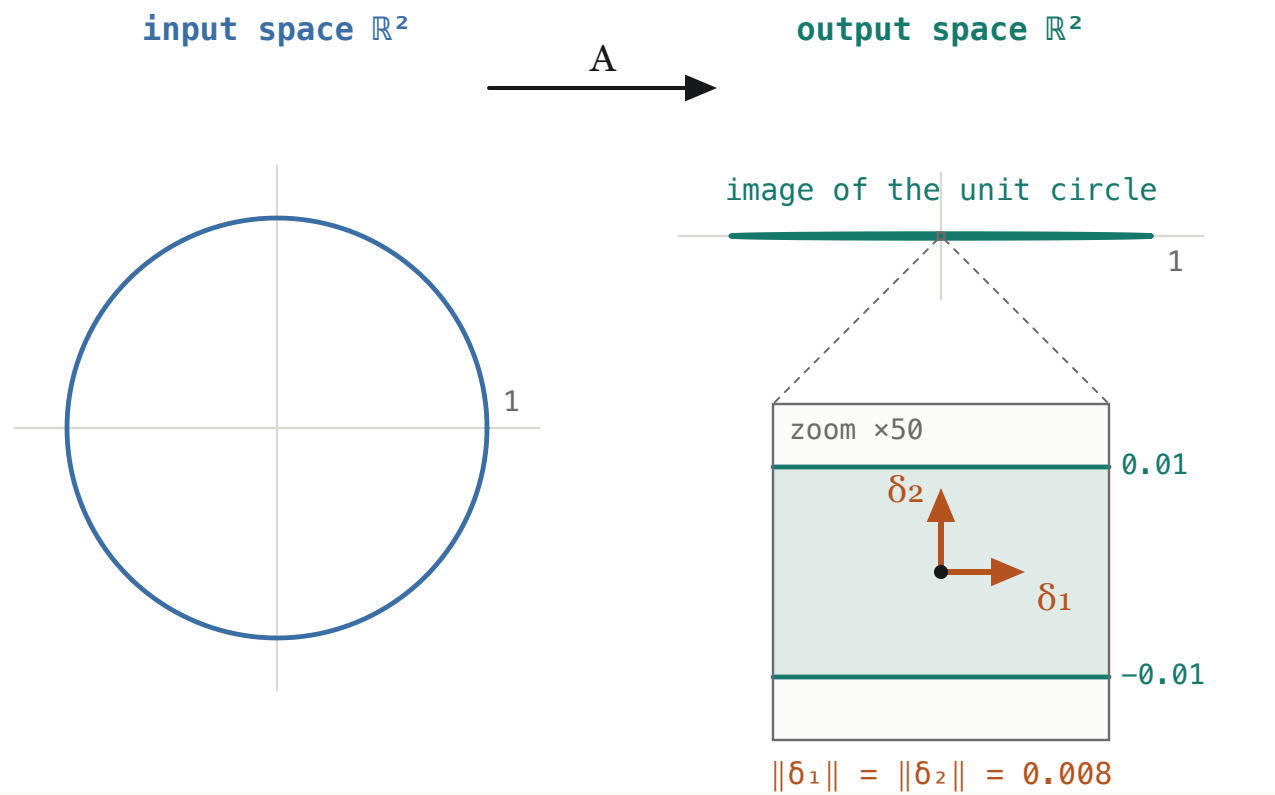
B only $(2, 0)$

C all three

VOTE

From Projections to Matrix Structure

Which Measurement Error Grows When We Invert?



$A = \mathrm{diag}(1,\ 0.01)$. Two output errors have equal length. Which grows more when we recover $x = A^{-1}y$?

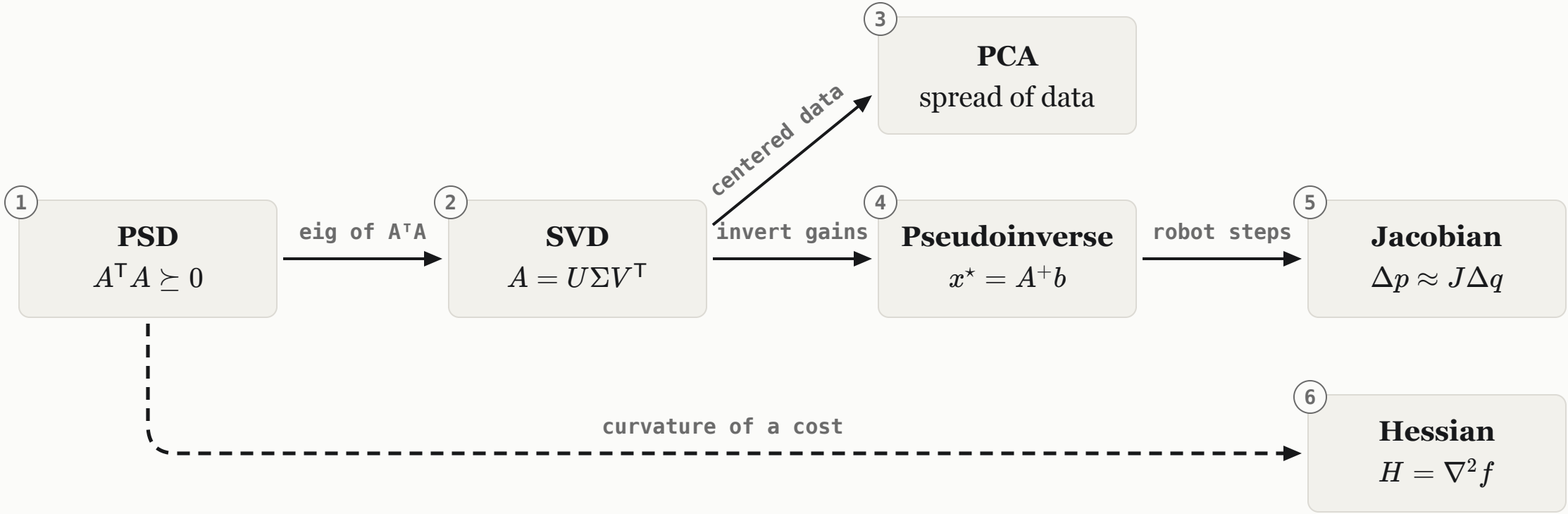
×50 zoom: read the numbers, not the lengths.

- **A** horizontal δ_1
- **B** vertical δ_2
- **C** equally

CONCEPT MAP

From Projections to Matrix Structure

One Structure, Six Tools



Why loop back: near a stationary point, a smooth cost changes like the quadratic form of its Hessian.

DEFINITION

Positive Semidefinite Matrices

A Quadratic Form Gives Each Displacement a Height

$$q(x) = x^\mathsf{T} H x = ax_1^2 + 2bx_1x_2 + cx_2^2$$

$x \in \mathbb{R}^2$ the input displacement

$H = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ real and symmetric, $H = H^\mathsf{T}$

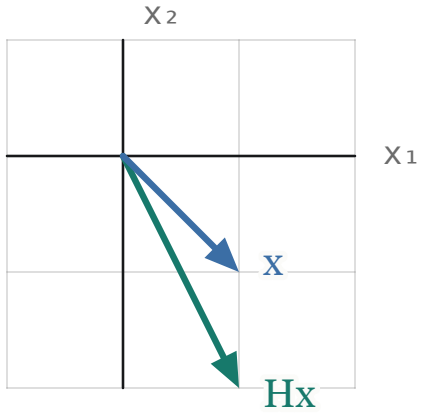
$Hx \in \mathbb{R}^2$ a vector

$q(x) \in \mathbb{R}$ one number: the height, or cost, of x

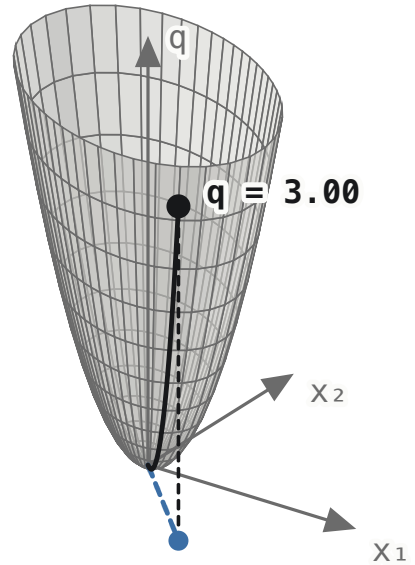
WORKED EXAMPLE

$H = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$, $x = (1, -1)$: $Hx = (1, -2)$, so $x^\mathsf{T}(Hx) = 1 + 2 = 3$. Check: $2 - 2 + 3 = 3$.

input space $\mathbb{R}^2$



height q(x) over $\mathbb{R}^2$



x = (1.00, −1.00)

Hx = (1.00, −2.00) a vector

xᵀ(Hx) = 3.00 one number

VOTE

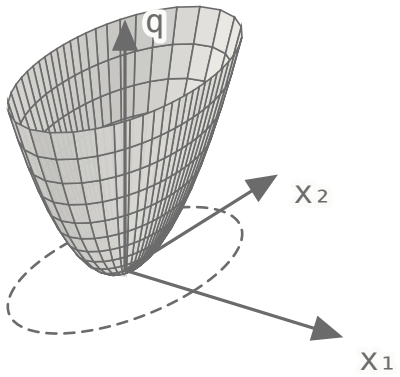
Positive Semidefinite Matrices

Which Surface Matches the Formula?

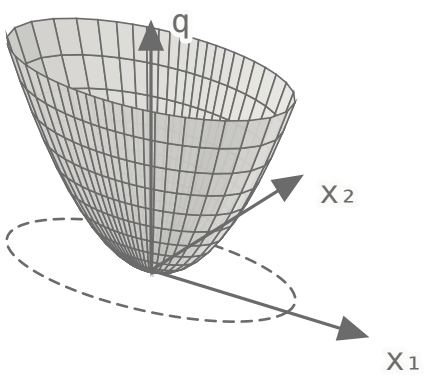
Which surface is the graph of $q(x) = 4x_1^2 + x_2^2$? Same camera, range and scales in all three.

`floor: input space ℝ², axes x₁, x₂ · height: q(x)`

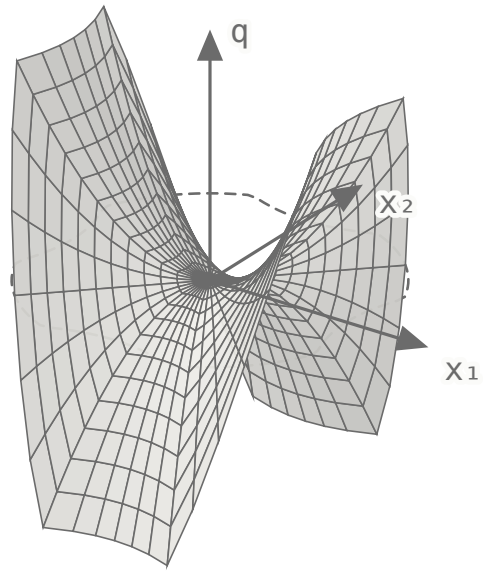
A



B



C



A surface A

B surface B

C surface C

DEFINITION

Positive Semidefinite Matrices

Positive Definite and Positive Semidefinite

$$H \succ 0 \iff x^\mathsf{T} H x > 0 \text{ for all } x \neq 0$$

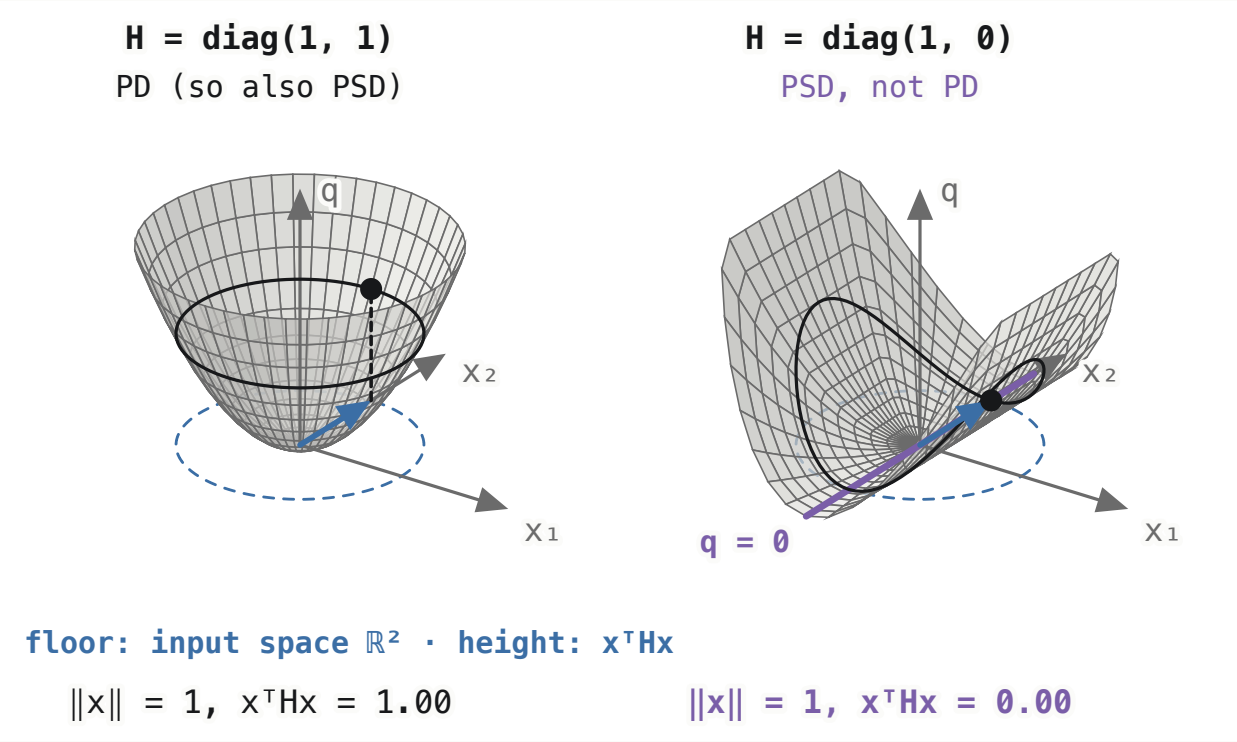
$$H \succeq 0 \iff x^\mathsf{T} H x \geq 0 \text{ for all } x$$

$H = H^\mathsf{T}$ real $n \times n$; both tests assume it

$H \succ 0$ positive definite (PD)

$H \succeq 0$ positive semidefinite (PSD)

Every PD matrix is PSD. The trough is *one* PSD matrix that is not PD; the zero matrix is another.



VOTE

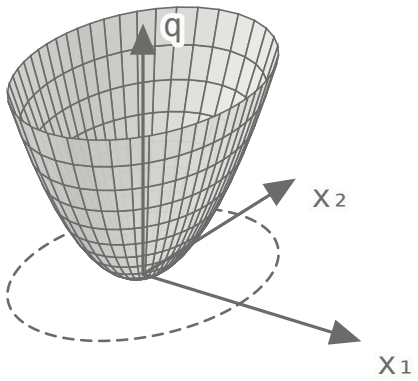
Positive Semidefinite Matrices

Which Matrix Is PSD but Not PD?

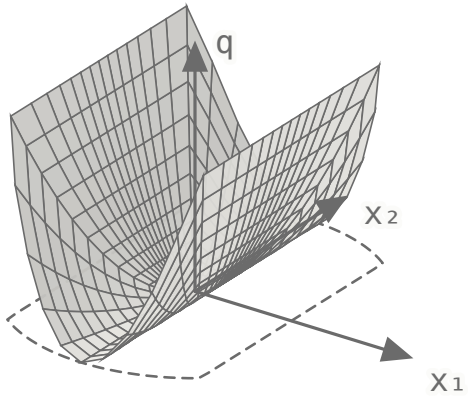
Each surface is the graph of $x^\top H x$, one camera, one scale. Which H is PSD but not PD?

`floor: input space ℝ², axes x₁, x₂ · height: q(x) = xᵀHx`

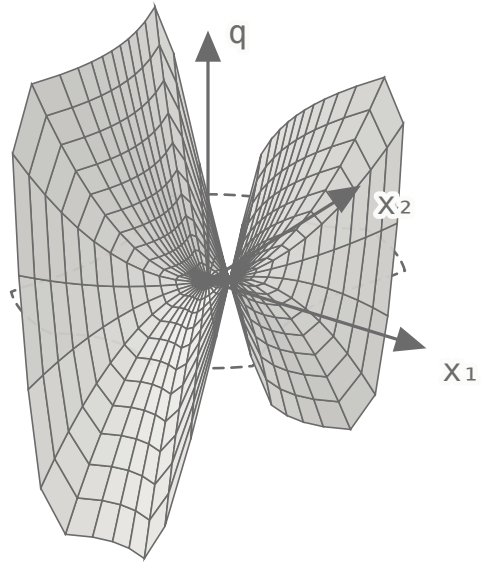
A



B



C



(A) $H = \mathrm{diag}(2, 1)$

(B) $H = \mathrm{diag}(2, 0)$

(C) $H = \mathrm{diag}(2, -1)$

THEOREM

Positive Semidefinite Matrices

Eigenvectors Rotate the Form Into Principal Axes

Spectral theorem: real symmetric H has real λ_i and orthonormal eigenvectors v_i.

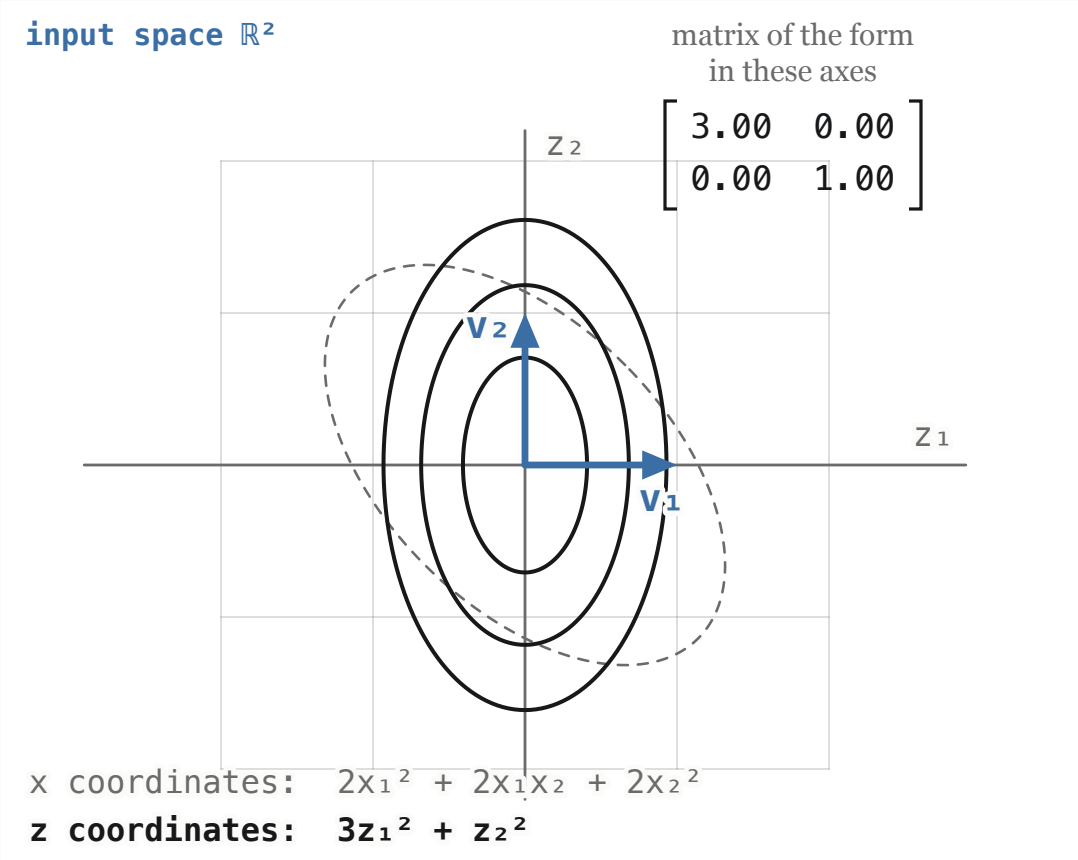
$$H = Q\Lambda Q^{\mathsf{T}}, \qquad x^{\mathsf{T}}Hx = z^{\mathsf{T}}\Lambda z = \sum_i \lambda_i z_i^2$$

$Q = [v_1 \cdots v_n]$ $Q^{\mathsf{T}}Q = I$: rotation or reflection

Λ diagonal λ_i

$z = Q^{\mathsf{T}}x$ $z_i = v_i^{\mathsf{T}}x$

> $H \succeq 0 \iff$ all $\lambda_i \geq 0$
>
> $H \succ 0 \iff$ all $\lambda_i > 0$



VOTE

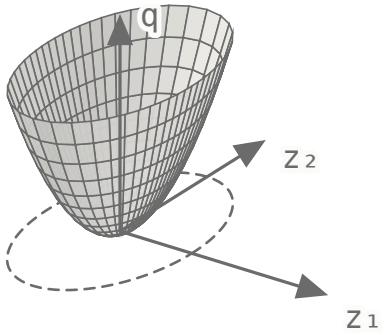
Positive Semidefinite Matrices

What Shape Do These Eigenvalues Make?

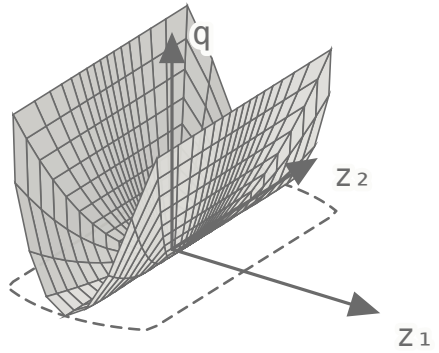
A symmetric 2×2 matrix H has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = -1$. Which shape is the graph of $x^\top H x$?

`floor: input space ℝ², principal coordinates z₁, z₂ · height: xᵀHx`

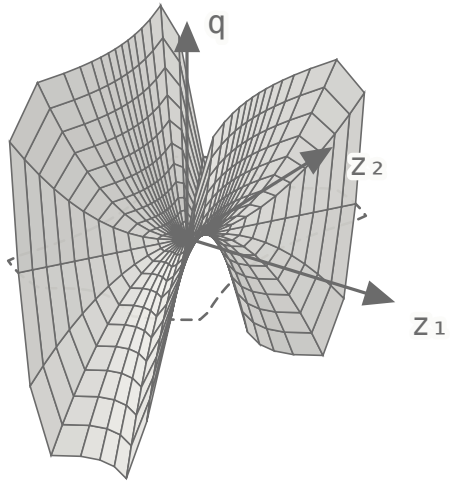
A



B



C



(A) bowl

(B) flat trough

(C) saddle

DERIVATION

Positive Semidefinite Matrices

Every Gram Matrix Is PSD

$$x^\mathsf{T} A^\mathsf{T} A x = (Ax)^\mathsf{T}(Ax) = \|Ax\|^2 \geq 0$$

$A \in \mathbb{R}^{m \times n}$ any shape: tall, wide or square

$A^\mathsf{T} A$ Gram matrix, $n \times n$: $(A^\mathsf{T} A)^\mathsf{T} = A^\mathsf{T} A$

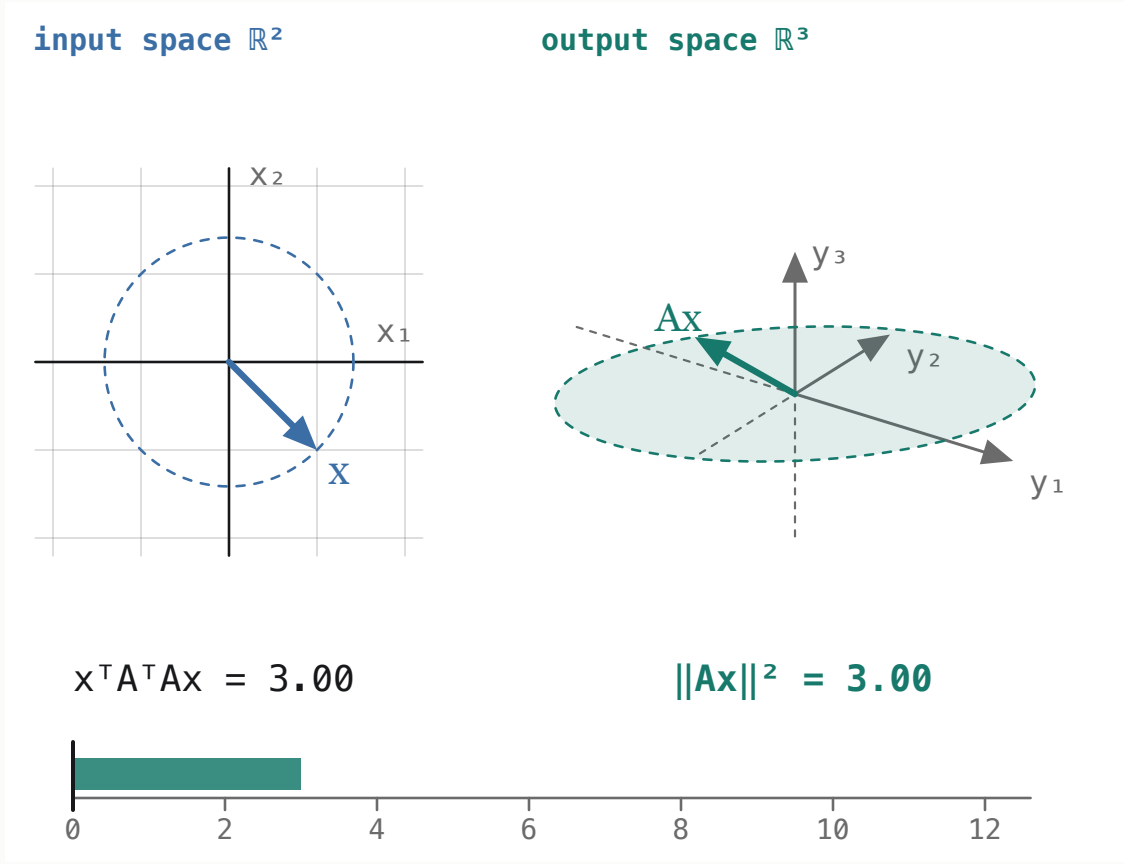
≥ 0 for every $x \in \mathbb{R}^n$, so $A^\mathsf{T} A \succeq 0$

WORKED EXAMPLE

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, A^\mathsf{T} A = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$$

$x = (1, -1)$: $\|Ax\|^2 = \|(-1, -1, 1)\|^2 = 3$

$x^\mathsf{T} A^\mathsf{T} A x = 2 - 4 + 5 = 3$

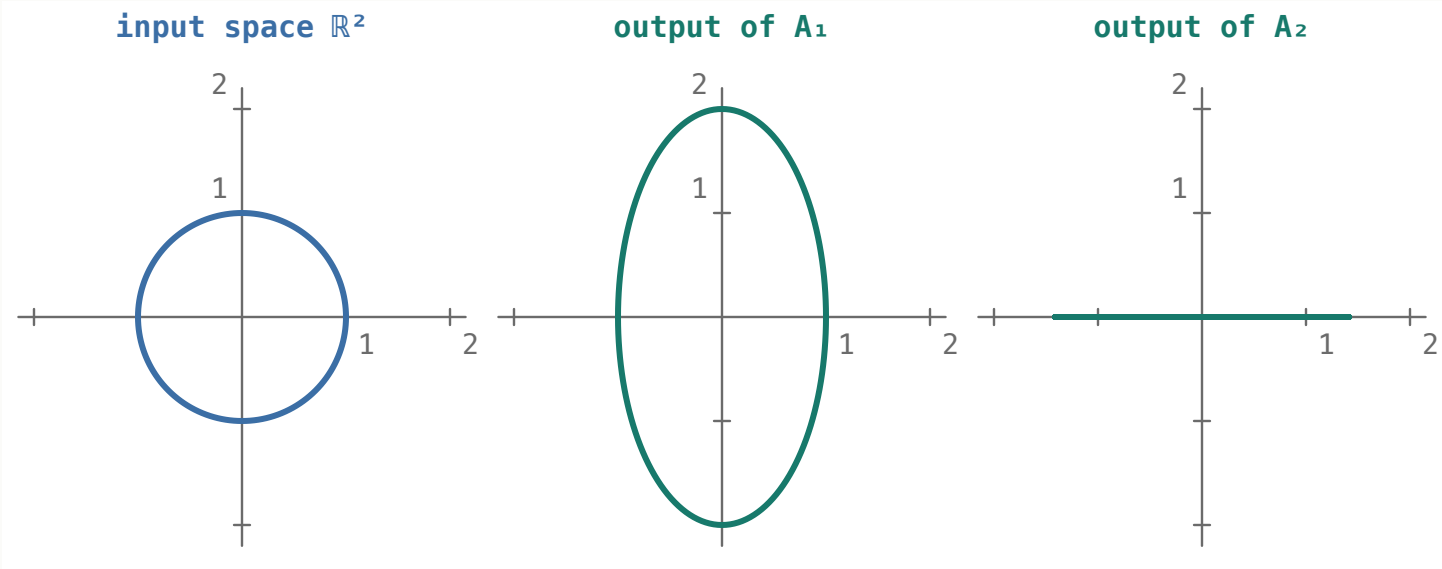


VOTE

Positive Semidefinite Matrices

Which Gram Matrix Is Positive Definite?

$A_1 = \mathrm{diag}(1, 2)$ and $A_2 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ map the unit circle as shown. Which $A^\top A$ is PD?



A only $A_1^\top A_1$

B only $A_2^\top A_2$

C both

PICTURE

$A_2 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

Positive Semidefinite Matrices

Flat Directions Are Null Directions

$$x^\mathsf{T} A^\mathsf{T} A x = \|Ax\|^2 = 0 \iff Ax = 0$$

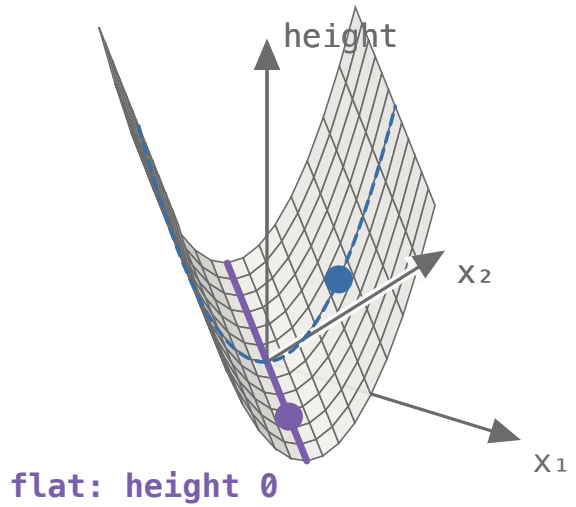
$$\mathrm{Null}(A^\mathsf{T} A) = \mathrm{Null}(A)$$

$\iff$ a squared length is 0 only for the zero vector

$\mathrm{Null}(A)$ all $x \in \mathbb{R}^n$ with $Ax = 0$

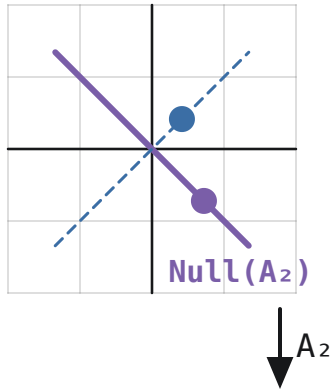
why $A^\mathsf{T} Ax = 0 \Rightarrow x^\mathsf{T} A^\mathsf{T} Ax = 0 \Rightarrow Ax = 0$;
conversely $Ax = 0 \Rightarrow A^\mathsf{T} Ax = 0$

height $x^\mathsf{T}A_2^\mathsf{T}A_2x = (x_1 + x_2)^2$
over the input space $\mathbb{R}^2$



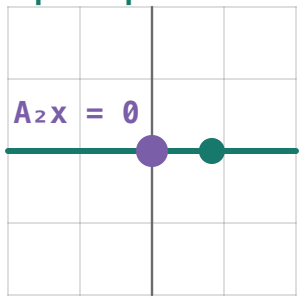
flat: height 0

input space $\mathbb{R}^2$



A_2

output space $\mathbb{R}^2$



VOTE

Positive Semidefinite Matrices

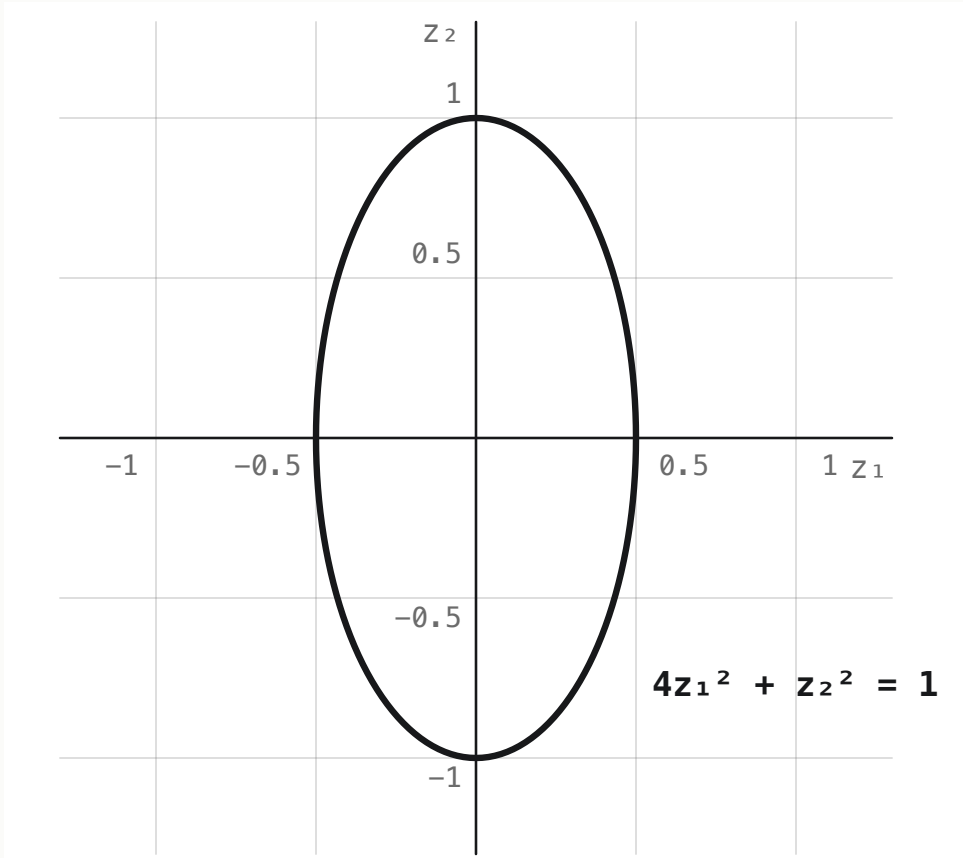
Which Eigenvalue Sets the Long Axis?

$4z_1^2 + z_2^2 = 1$ is the level set $x^\mathsf{T} H x = 1$ in principal coordinates. Which eigenvalue sets its long axis?

A the larger

B the smaller

C neither



DEFINITION

Positive Semidefinite Matrices

The PSD Square Root

$$H^{1/2} = Q\Lambda^{1/2}Q^\mathsf{T}, \qquad x^\mathsf{T}Hx = \|H^{1/2}x\|^2$$

$H = Q\Lambda Q^\mathsf{T}$ PSD: $\lambda_i \geq 0$, $Q = [v_1\ v_2]$

$\Lambda^{1/2}$ $\mathrm{diag}(\sqrt{\lambda_i})$: real as $\lambda_i \geq 0$

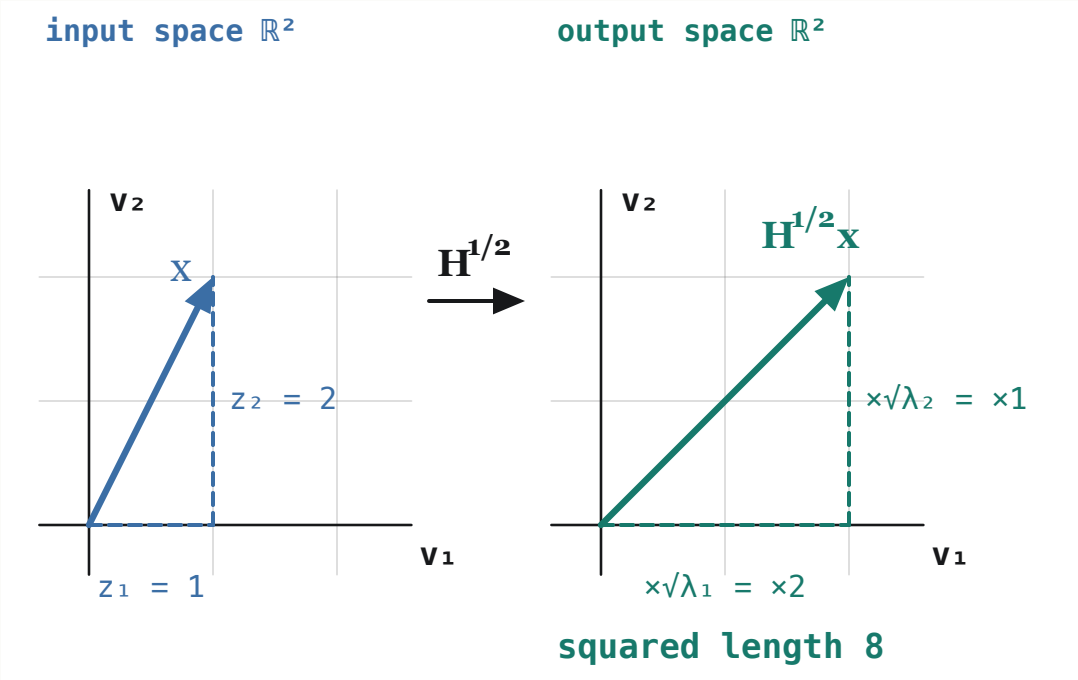
$H^{1/2}$ unique symmetric PSD: $H^{1/2}H^{1/2} = H$

WORKED EXAMPLE

$H = \mathrm{diag}(4, 1)$: $Q = I$, $H^{1/2} = \mathrm{diag}(2, 1)$

$x = (1, 2)$: $H^{1/2}x = (2, 2)$

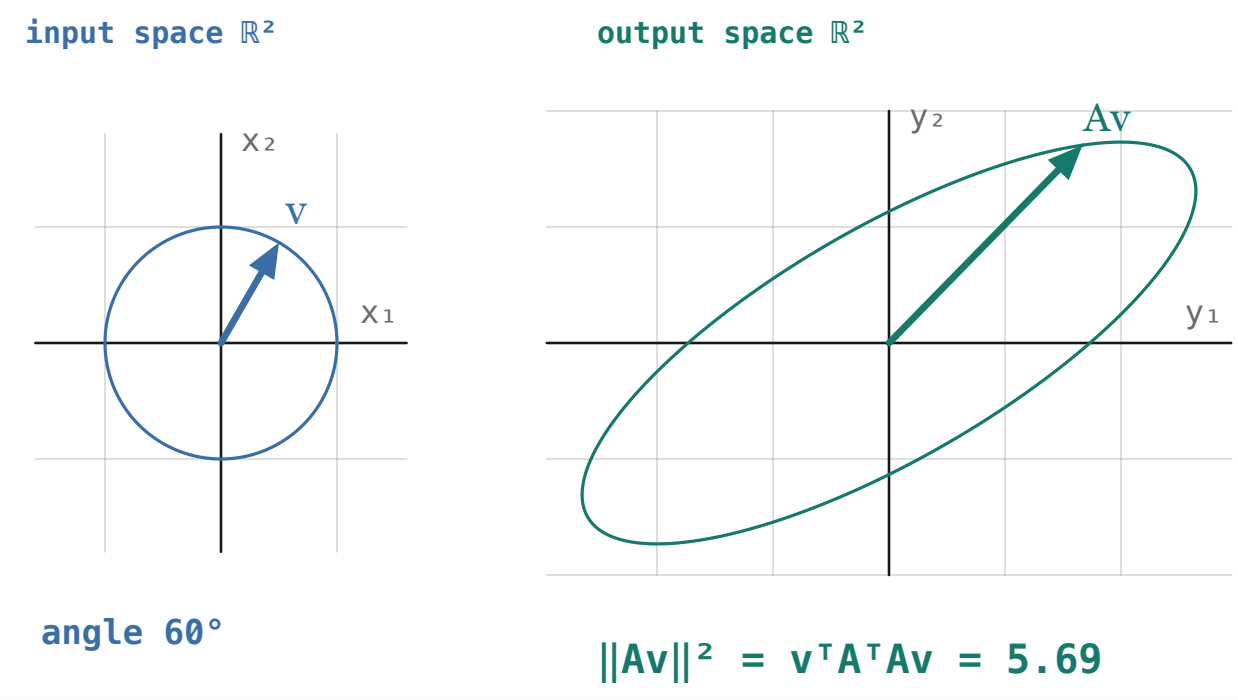
$\|H^{1/2}x\|^2 = 8 = x^\mathsf{T}Hx$



LOOKING AHEAD

Positive Semidefinite Matrices

Which Input Direction Stretches Most?



$$\|Av\|^2 = v^\mathsf{T} A^\mathsf{T} A v, \qquad \|v\| = 1$$

$A \approx \begin{bmatrix} 2.61 & 0.42 \\ 1.11 & 1.33 \end{bmatrix}$, not symmetric

v a unit input direction, turning

$A^\mathsf{T}A$ its Gram matrix: symmetric PSD

> Which v makes $\|Av\|^2$ largest? Which makes it smallest?

DEFINITION

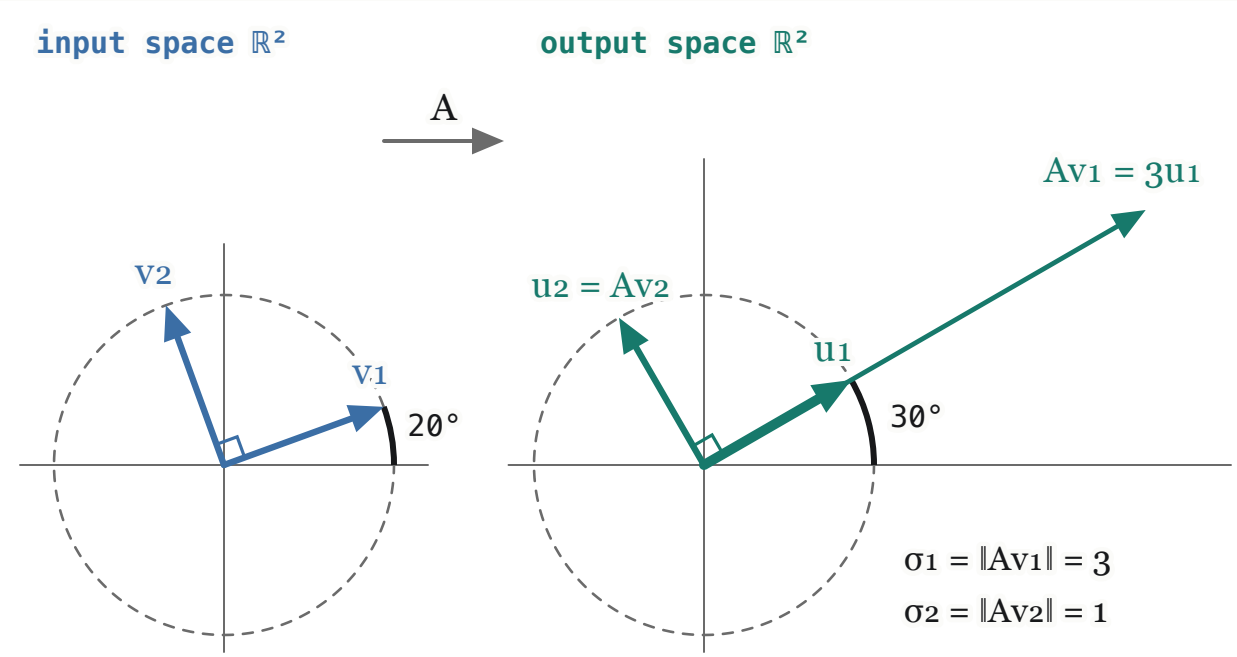
Singular Value Decomposition

Input Directions Paired With Output Directions

$$A = 2.61\ 0.42 \\ 1.11\ 1.33$$

$$Av_i = \sigma_i u_i, \qquad \|v_i\| = \|u_i\| = 1$$

- A — $m \times n$: input $\mathbb{R}^n$ to output $\mathbb{R}^m$
- $v_i \in \mathbb{R}^n$ — input direction; the v_i are orthonormal
- $u_i \in \mathbb{R}^m$ — output direction; the u_i are orthonormal
- $\sigma_i \geq 0$ — length gain: $\sigma_i = \|Av_i\|$, $i \leq \min(m, n)$



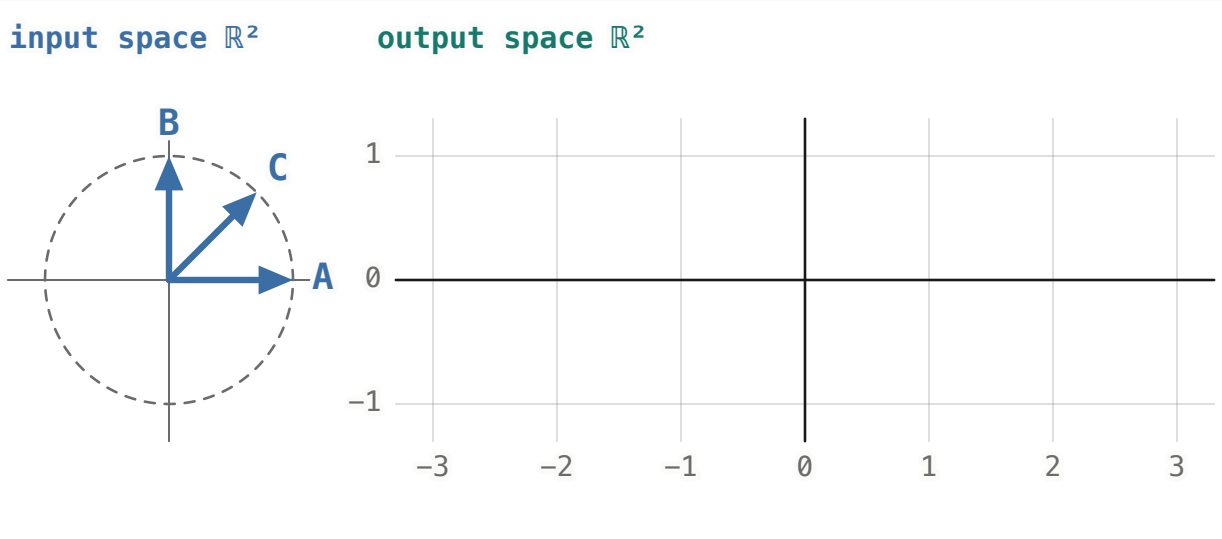
VOTE

Singular Value Decomposition

Which Unit Input Comes Out Longest?

$A = \mathrm{diag}(3, 1)$. Which unit input has the longest output Ax?

- (A) horizontal, e_1
- (B) vertical, e_2
- (C) diagonal, $(e_1 + e_2)/\sqrt{2}$



DEFINITION

Singular Value Decomposition

The Full SVD and Its Shapes

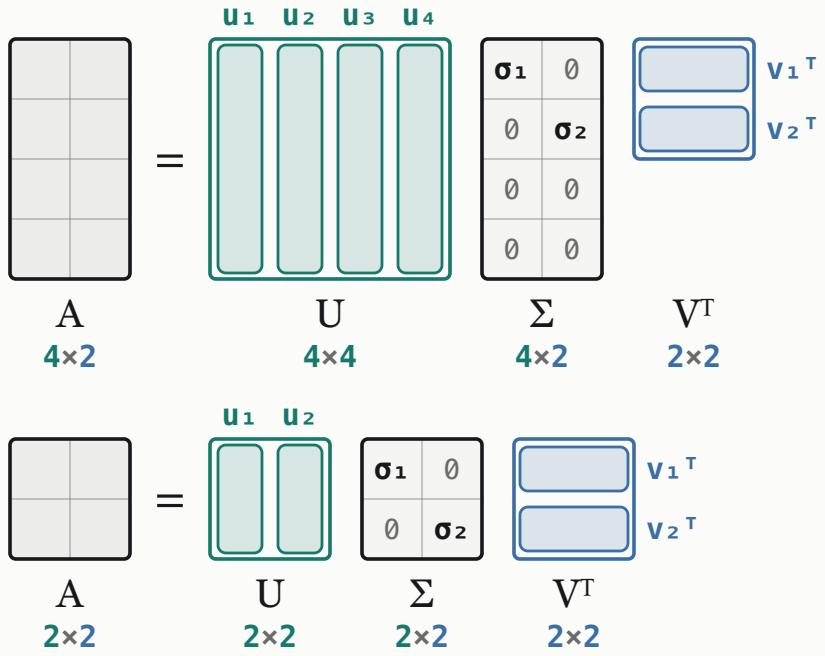
$$A_{m\times n} = U_{m\times m}\,\Sigma_{m\times n}\,V^{\mathsf{T}}_{n\times n}$$

$U^{\mathsf{T}}U = I_m$ orthonormal columns u_i: output $\mathbb{R}^m$

$V^{\mathsf{T}}V = I_n$ orthonormal columns v_i: input $\mathbb{R}^n$

Σ $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ on the diagonal, else 0

> Every real matrix has a full SVD.

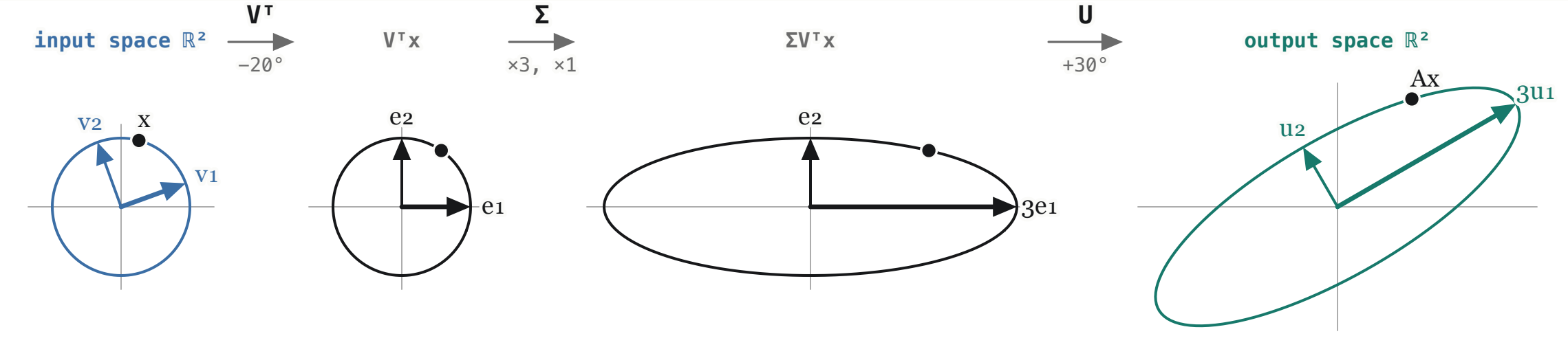


PICTURE

$$A \approx \begin{bmatrix} 2.61 & 0.42 \\ 1.11 & 1.33 \end{bmatrix}$$

Singular Value Decomposition

Orthogonal, Diagonal, Orthogonal



$$Ax = U\left(\Sigma\left(V^{\mathsf{T}}x\right)\right)$$

- V^{T} orthogonal: coordinates of x along v_1, v_2; here $R(-20^\circ)$
- Σ stretch coordinate i by σ_i; for $m \neq n$ it also drops or adds zeros
- U orthogonal: coordinate i goes along u_i; here $R(30^\circ)$
- $\det = \pm 1$ an orthogonal factor may reflect; here both rotate

VOTE

Singular Value Decomposition

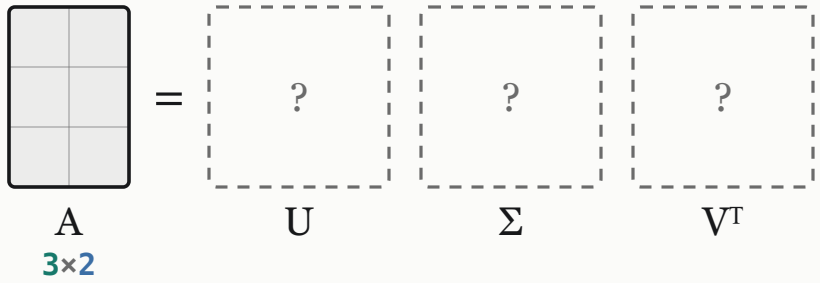
What Are the Shapes for a 3×2 Matrix?

A is 3×2, with full SVD $A = U\Sigma V^\mathsf{T}$. What are the shapes of $(U, \Sigma, V^\mathsf{T})$?

A $(3\times3, 3\times2, 2\times2)$

B $(3\times2, 2\times2, 2\times3)$

C $(2\times2, 2\times3, 3\times3)$



PICTURE

$$A \approx \begin{bmatrix} 2.61 & 0.42 \\ 1.11 & 1.33 \end{bmatrix}$$

Singular Value Decomposition

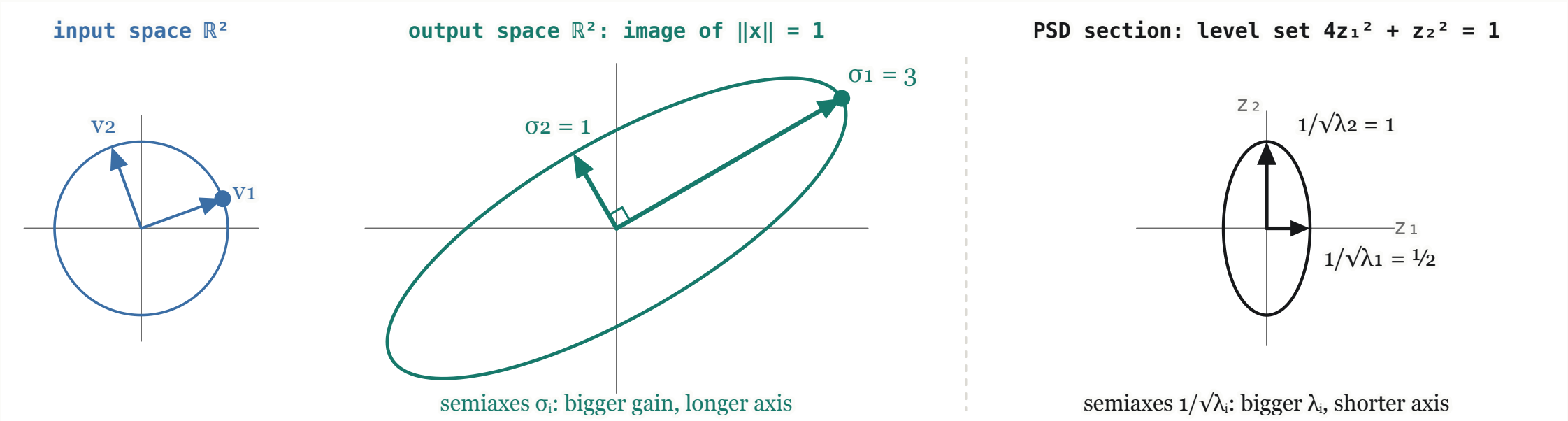
Singular Values Are the Ellipse Semiaxes

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0, \qquad p = \min(m, n)$$

σ_i singular values: sorted, never negative

p how many there are

$\{Ax : \|x\| = 1\}$ full-rank 2×2 A: an ellipse with semiaxes $\sigma_i u_i$



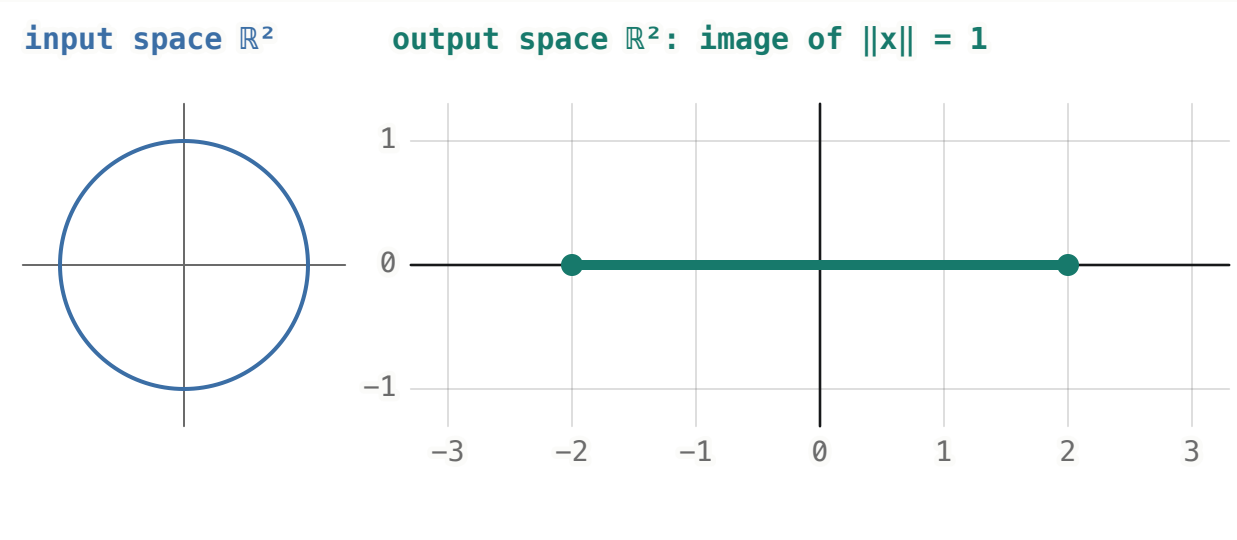
VOTE

Singular Value Decomposition

What Are the Singular Values of This Collapse?

A 2×2 matrix maps the unit circle onto the segment from $-2e_1$ to $2e_1$. What are (σ_1, σ_2)?

- **A** $(2, 1)$
- **B** $(2, 0)$
- **C** $(0, 0)$



DERIVATION

Singular Value Decomposition

Right Singular Vectors From the Gram Matrix

$$A^\top A v_i = \lambda_i v_i, \qquad \sigma_i = \sqrt{\lambda_i}$$

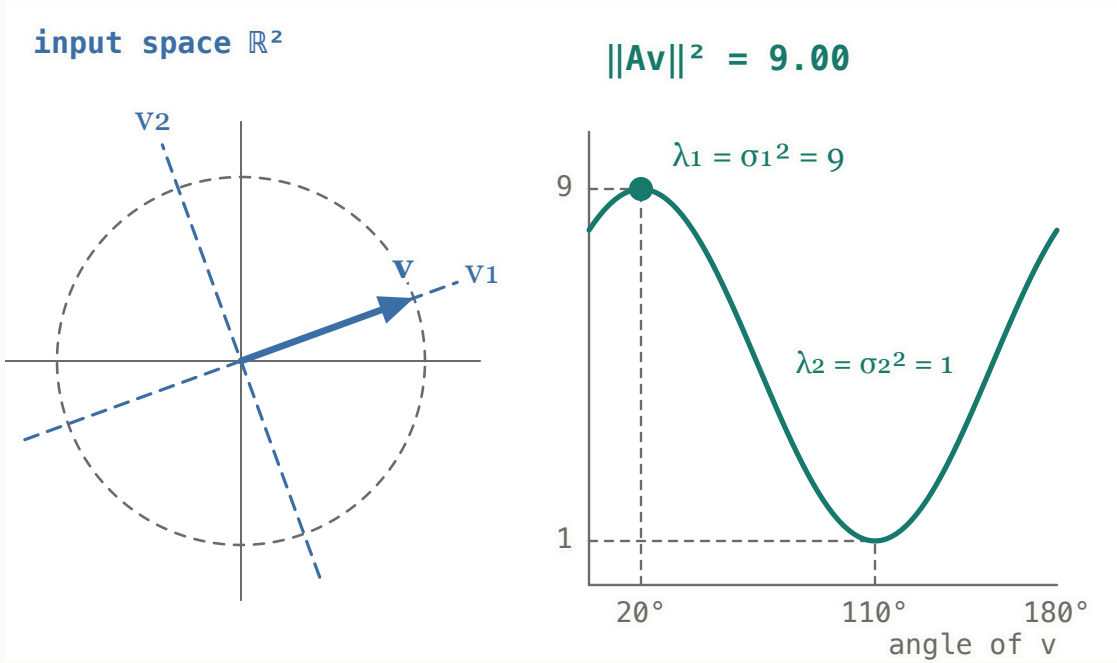
λ_i $= \|Av_i\|^2 \geq 0$: eigenvalues of PSD $A^\top A$

v_i orthonormal eigenvectors: input axes

WORKED EXAMPLE

$A^\top A \approx \begin{bmatrix} 8.06 & 2.57 \\ 2.57 & 1.94 \end{bmatrix}$ (rounded). Exact $A = R(30^\circ)\mathrm{diag}(3,1)R(-20^\circ)$: trace 10, det 9, $\lambda_i = 9, 1$, $\sigma_i = 3, 1$.

$A \approx \begin{bmatrix} 2.61 & 0.42 \\ 1.11 & 1.33 \end{bmatrix}$



$$A \approx \begin{bmatrix} 2.61 & 0.42 \\ 1.11 & 1.33 \end{bmatrix}$$

Singular Value Decomposition

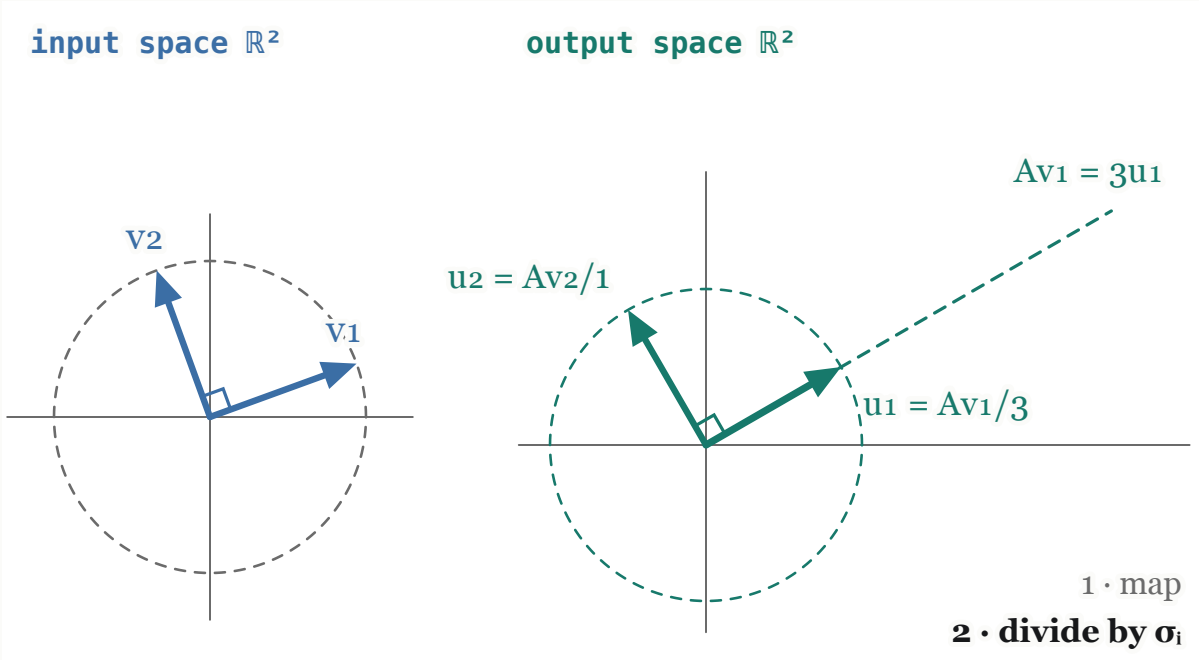
Left Singular Vectors: Map, Then Normalize

$$u_i = \frac{Av_i}{\sigma_i} \quad (\sigma_i > 0)$$

$\sigma_i = 0$ $Av_i = 0$: pick u_i to finish the basis

$$u_i^\top u_j = \frac{v_i^\top A^\top A v_j}{\sigma_i \sigma_j} = \frac{\sigma_j^2 v_i^\top v_j}{\sigma_i \sigma_j} = \delta_{ij}$$

δ_{ij} 1 if $i = j$, else 0: orthonormal u_i ($\sigma_i \sigma_j > 0$)



$$A \approx \begin{bmatrix} 2.61 & 0.42 \\ 1.11 & 1.33 \end{bmatrix}$$

VOTE

Singular Value Decomposition

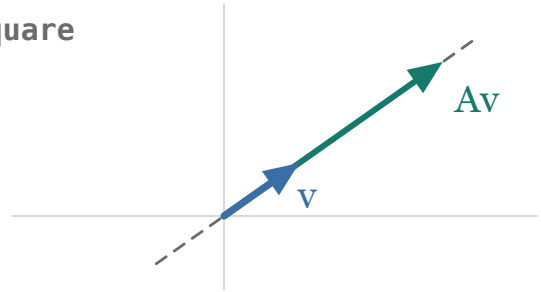
Eigenvectors or Singular Vectors?

Which statement about the SVD is true?

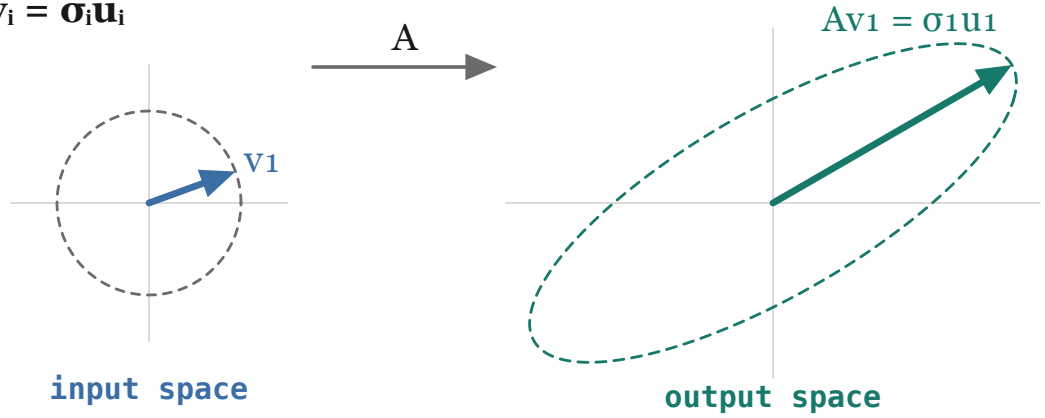
- (A) It exists only for square A
- (B) Input and output dimensions may differ
- (C) Singular values can be negative

$\mathbf{Av} = \lambda\mathbf{v}$

$\mathbb{R}^2 \to \mathbb{R}^2$, A square



$\mathbf{Av}_i = \sigma_i\mathbf{u}_i$



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

Singular Value Decomposition

Worked Example: A Symmetric Positive Definite Matrix

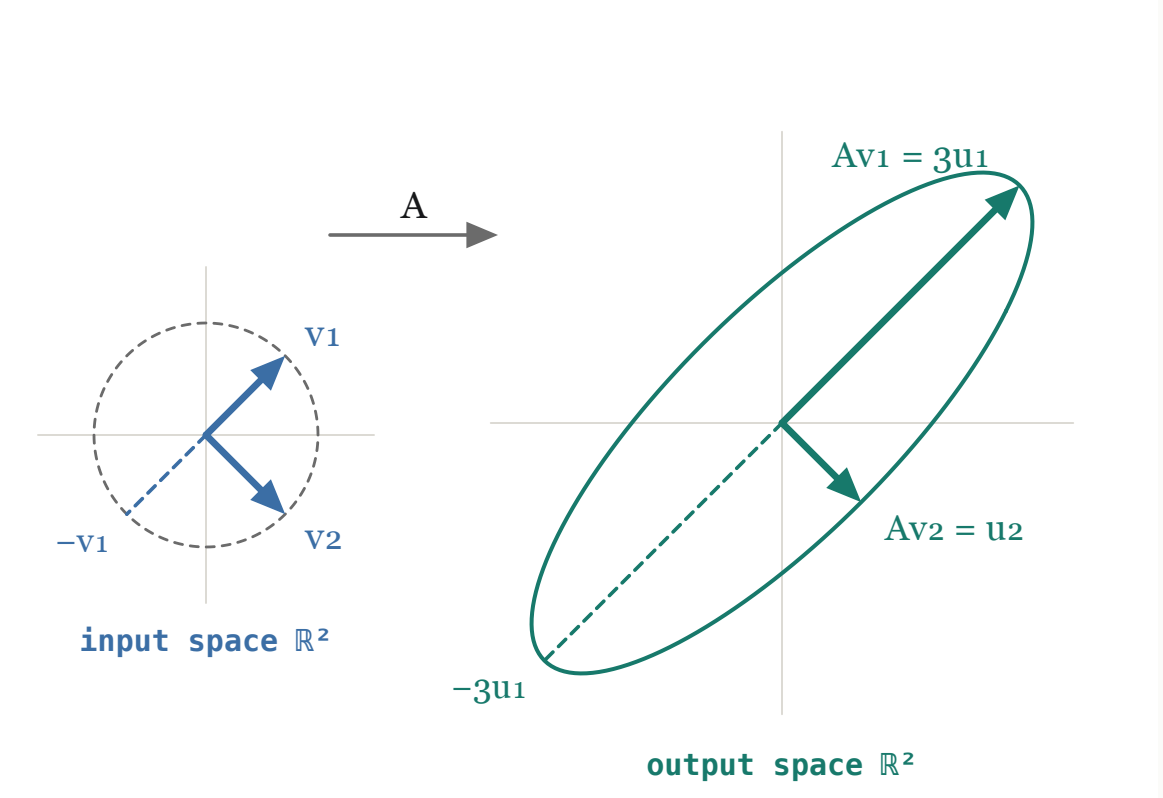
WORKED EXAMPLE

$v_{1,2} = (1, \pm 1)/\sqrt{2}$
$A(1,1) = (3,3)$, so $Av_1 = 3v_1$
$A(1,-1) = (1,-1)$, so $Av_2 = 1\, v_2$
$\sigma = (3,1)$, $u_i = Av_i/\sigma_i = v_i$

$$A = U\Sigma V^{\mathsf{T}} = (UD)\,\Sigma\,(VD)^{\mathsf{T}}$$

$U = V$ can be chosen exactly when A is symmetric PSD

D $\mathrm{diag}(\pm 1, \pm 1)$: flip u_i and v_i together



TRANSFER · CHOOSE A METHOD

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

Singular Value Decomposition

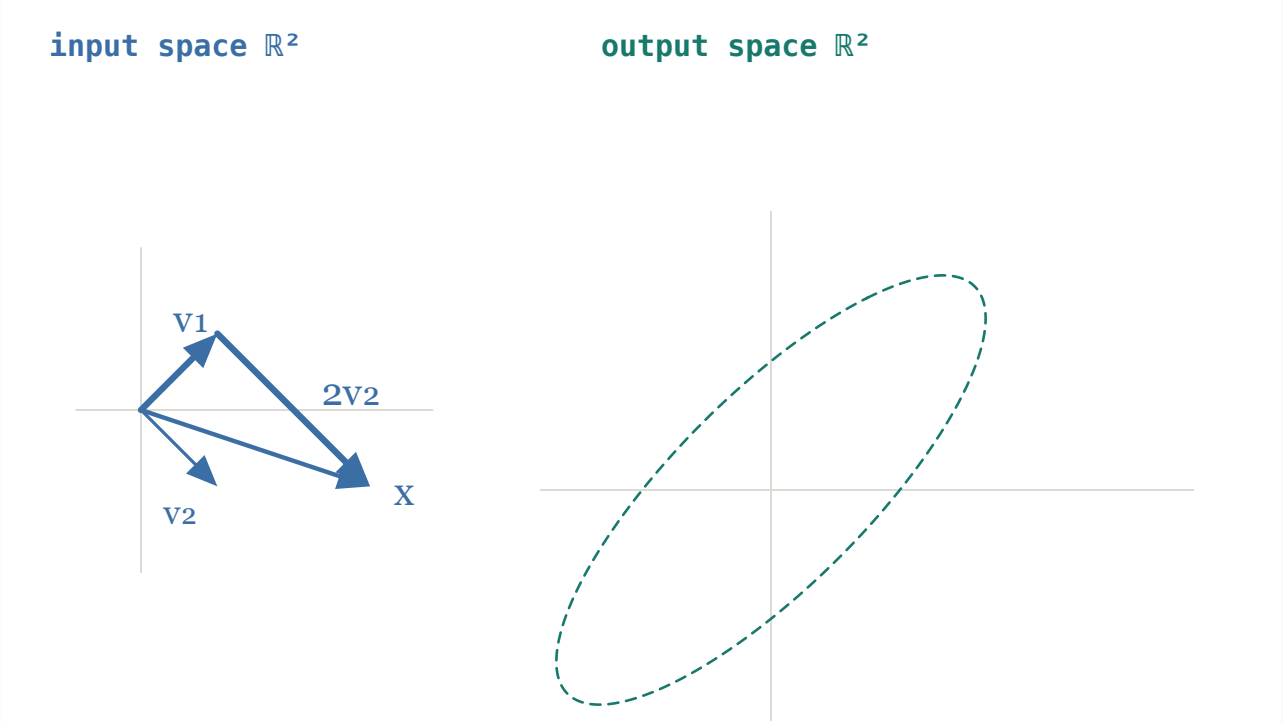
A Negative Eigenvalue

$v_{1,2} = (1, \pm 1)/\sqrt{2}$. For $x = v_1 + 2v_2$, find Ax, σ_i, u_i.

A $3v_1 - 2v_2$

B $3v_1 + 2v_2$

C $-3v_1 + 2v_2$



DEFINITION

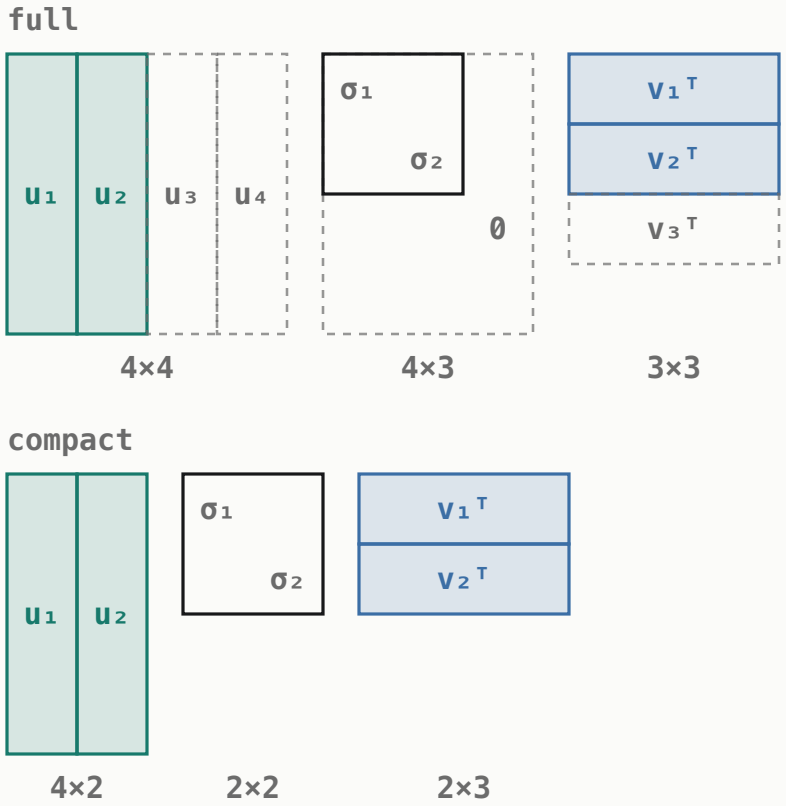
Singular Value Decomposition

The Compact SVD Keeps Only Rank r

$$A = U_r \Sigma_r V_r^{\mathsf{T}}, \quad U_r^{\mathsf{T}} U_r = V_r^{\mathsf{T}} V_r = I_r$$

r	$\operatorname{rank}(A)$: how many $\sigma_i > 0$
U_r, V_r	first r columns: $m \times r$, $n \times r$
Σ_r	$r \times r$ diagonal, all > 0
$U_r U_r^{\mathsf{T}}$	$m \times m$, $\neq I_m$ if $r < m$

Thin (economy): $\min(m, n)$ columns, zeros kept.



PICTURE

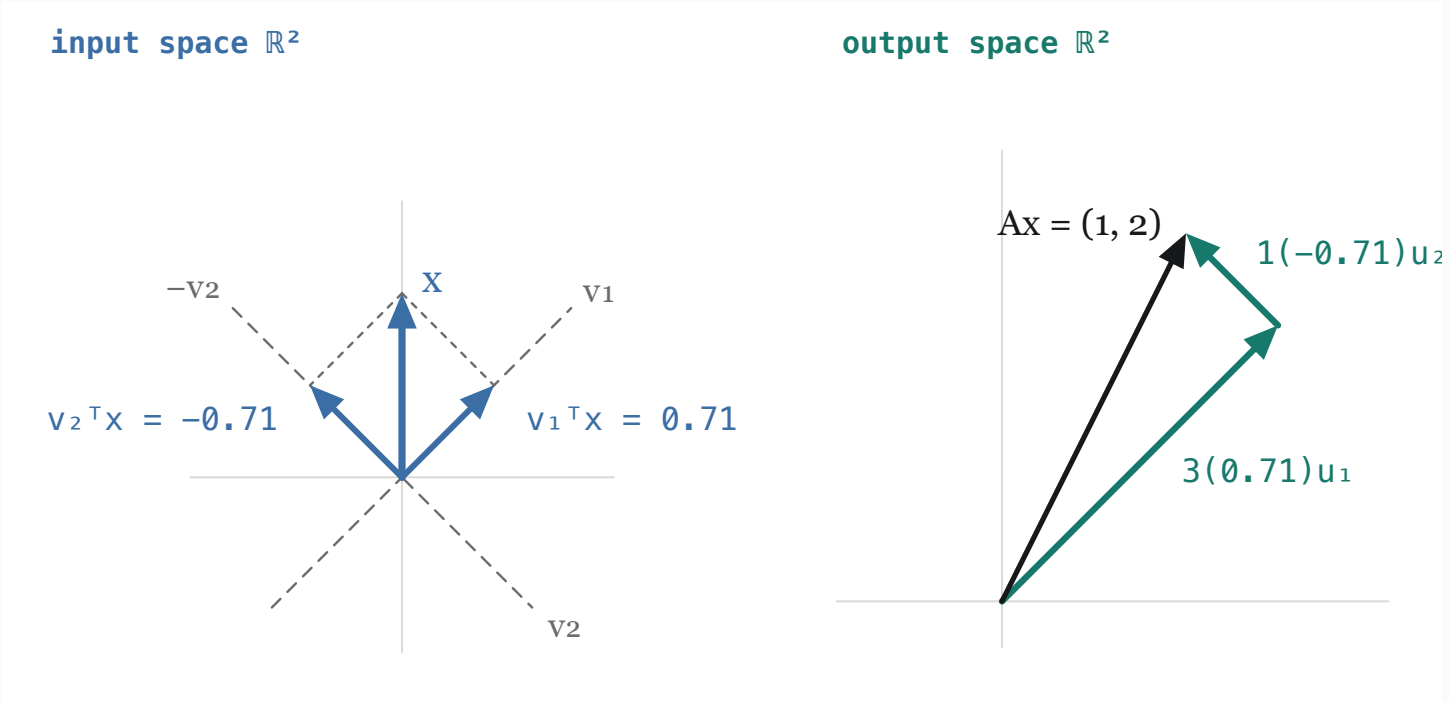
$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Singular Value Decomposition

A Sum of Rank-One Maps

$$A = \sum_{i=1}^{r} \sigma_i \, u_i v_i^{\mathsf{T}} \quad \Rightarrow \quad Ax = \sum_{i=1}^{r} \sigma_i \, (v_i^{\mathsf{T}} x) \, u_i$$

- $u_i v_i^{\mathsf{T}}$ — $m \times n$, rank one: reads along v_i, writes along u_i
- $v_i^{\mathsf{T}} x$ — a signed number, the coordinate of x along v_i
- r — the number of $\sigma_i > 0$



VOTE

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

Singular Value Decomposition

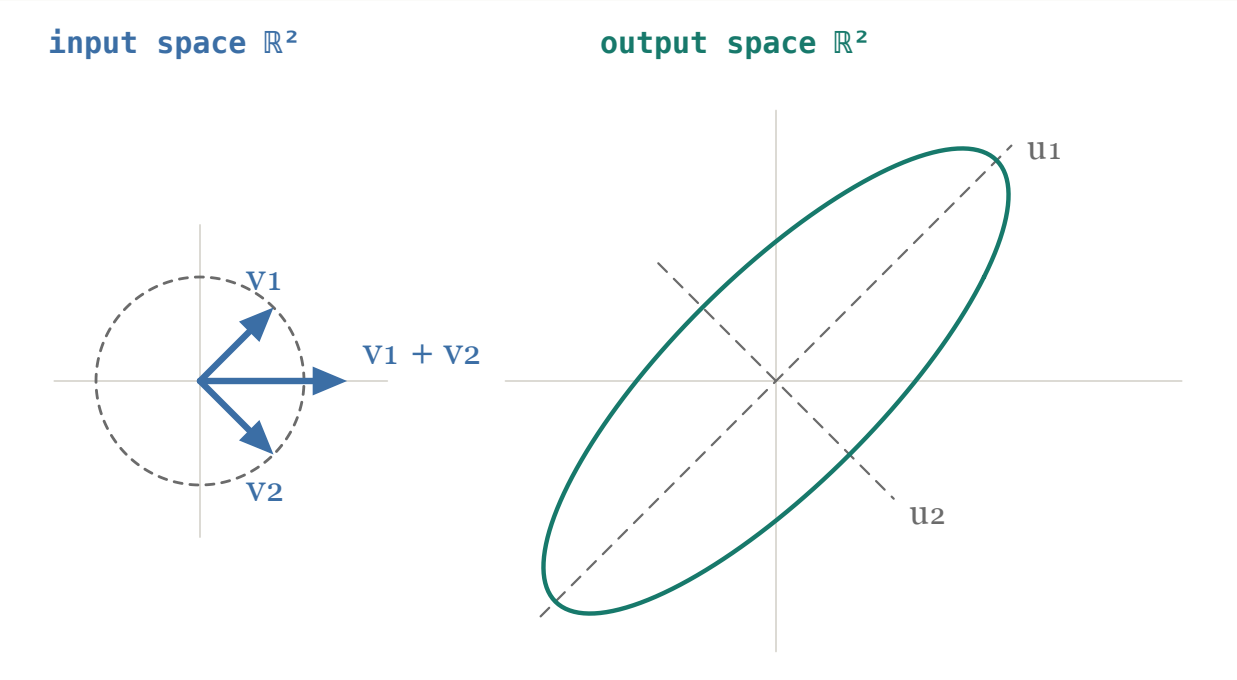
Which Input Does Truncation Erase?

Keep one term of $A = 3u_1v_1^\mathsf{T} + u_2v_2^\mathsf{T}$:
$A_1 = 3u_1v_1^\mathsf{T}$. Which input does A_1 send to 0?

A v_1

B v_2

C $v_1 + v_2$

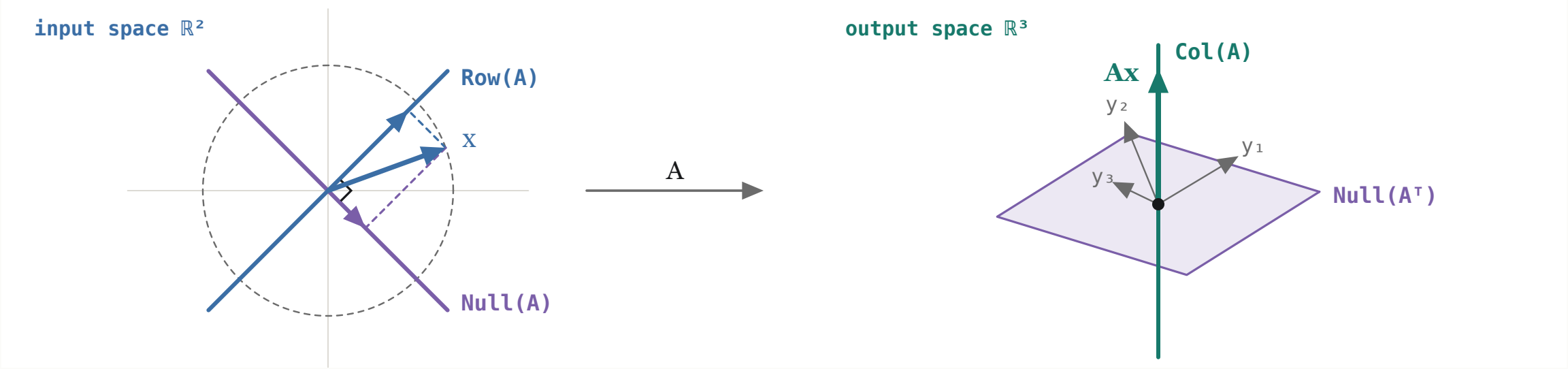


PICTURE

Singular Value Decomposition

The Four Subspaces From U and V

Full SVD: U is $m \times m$, V is $n \times n$. Example: the all-ones 3×2 matrix, rank $r = 1$.



$$\mathrm{Row}(A) = \mathrm{span}(v_1, \ldots, v_r), \dim r$$

$$\mathrm{Null}(A) = \mathrm{span}(v_{r+1}, \ldots, v_n), \dim n - r$$

$$\mathrm{Col}(A) = \mathrm{span}(u_1, \ldots, u_r), \dim r$$

$$\mathrm{Null}(A^\mathsf{T}) = \mathrm{span}(u_{r+1}, \ldots, u_m), \dim m - r$$

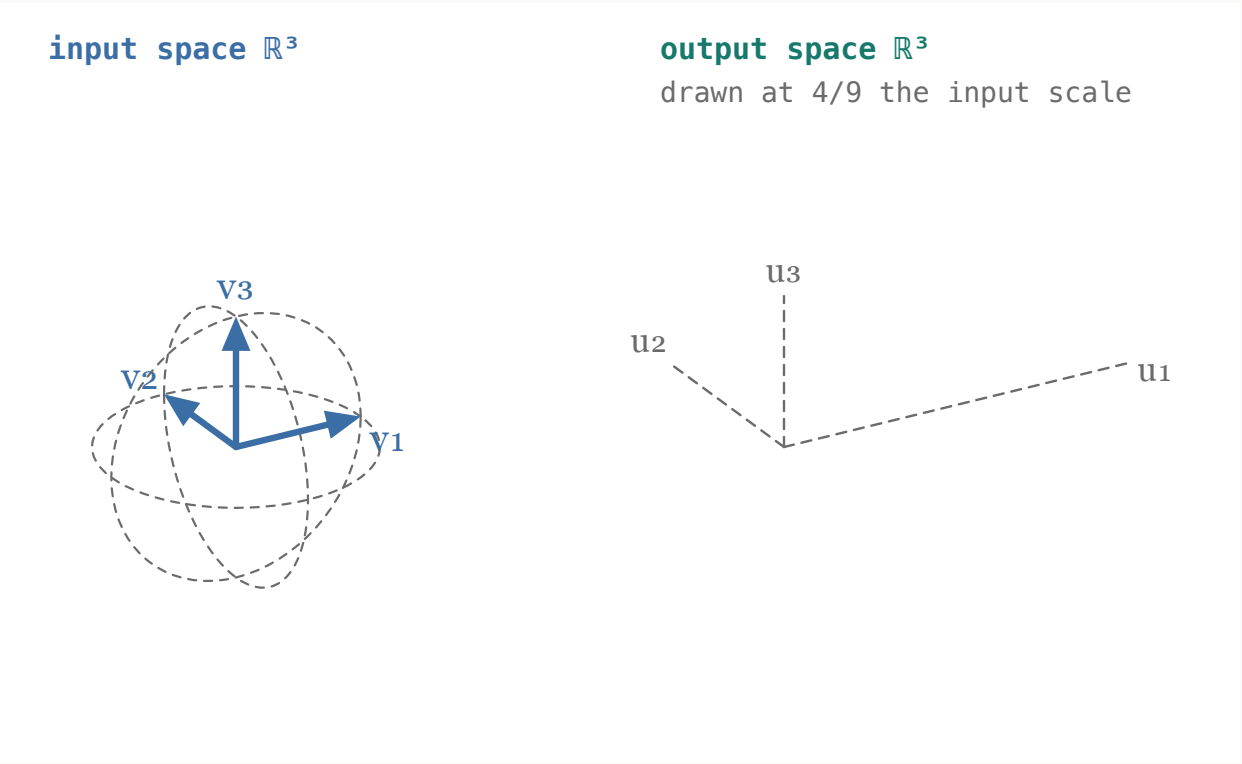
VOTE

Singular Value Decomposition

Rank and Nullity From Singular Values

A 3×3 matrix has singular values $5,\ 2,\ 0$. Which pair is (rank, nullity)?

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



THEOREM

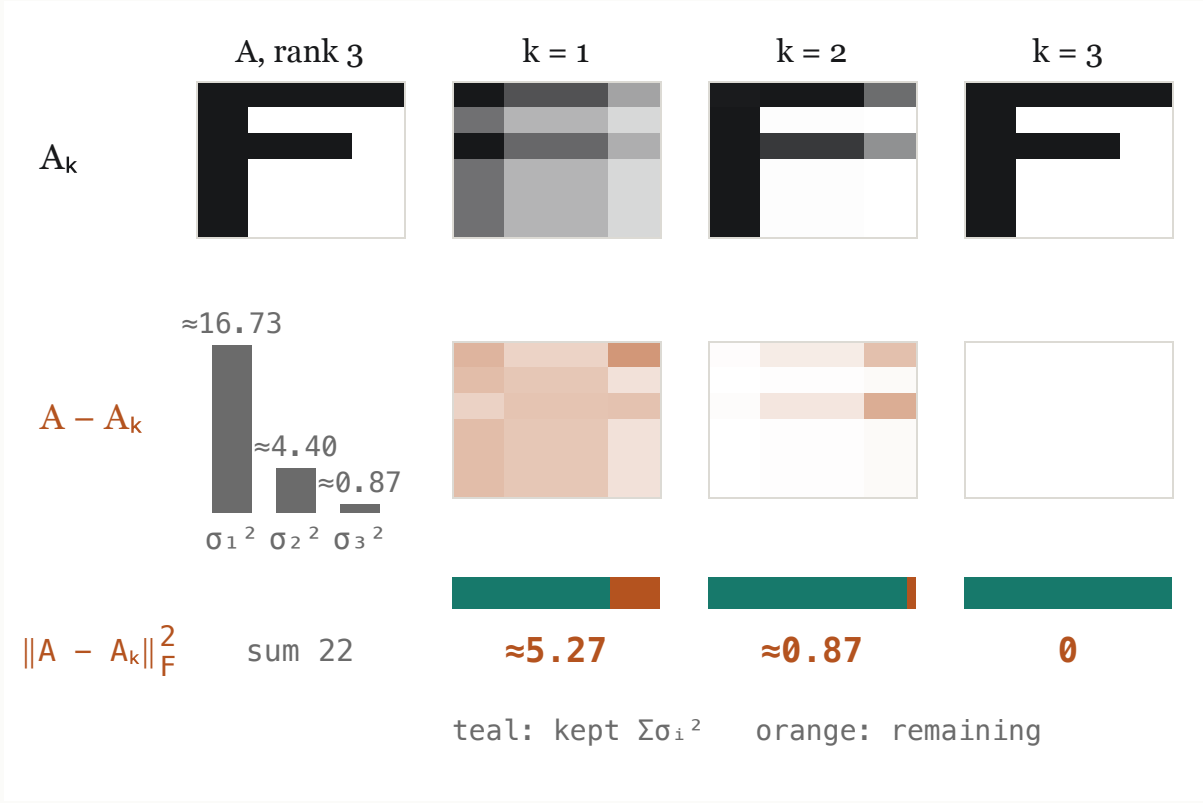
Singular Value Decomposition

The Best Rank-k Approximation

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^\mathsf{T} \in \arg\min_{\operatorname{rank} B \le k} \|A - B\|_F$$

$\|M\|_F^2$	$\sum_{i,j} M_{ij}^2$
$\|A - A_k\|_F^2$	$\sum_{i>k} \sigma_i^2$, the error
k	$0 \le k \le r$; $A_r = A$
B	any matrix with rank $\le k$
Eckart–Young	no proof; may not be unique if $\sigma_k = \sigma_{k+1} > 0$

Next: each row of A is one data sample.



DEFINITION

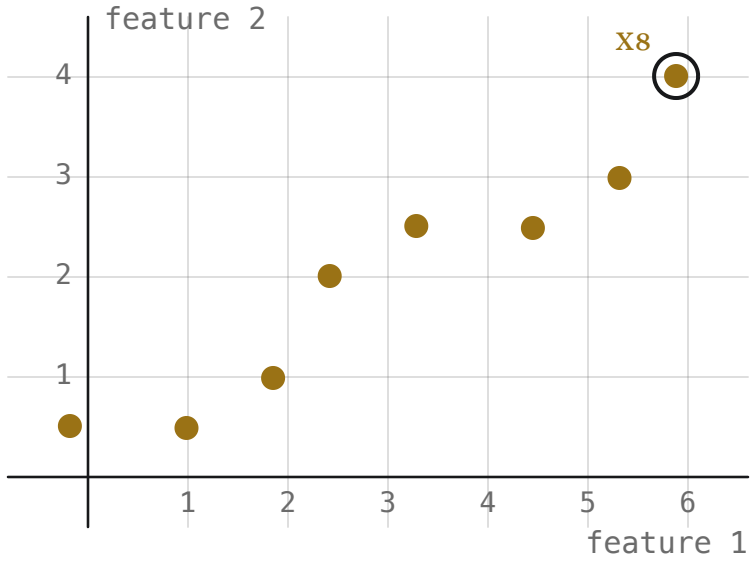
Principal Component Analysis

Data as a Matrix: One Row per Sample

data matrix X, 8 × 2

	feature 1	feature 2
$x_1^\top$	−0.18	0.51
$x_2^\top$	0.98	0.49
$x_3^\top$	1.85	0.99
$x_4^\top$	2.42	2.01
$x_5^\top$	3.28	2.51
$x_6^\top$	4.45	2.49
$x_7^\top$	5.32	2.99
$x_8^\top$	5.88	4.01

feature space $\mathbb{R}^2$



$$X = \begin{bmatrix} x_1^\top \\ \vdots \\ x_N^\top \end{bmatrix} \in \mathbb{R}^{N \times d}$$

$x_j \in \mathbb{R}^d$ sample j: one point in feature space

N rows = samples; here $N = 8$

d columns = features; here $d = 2$

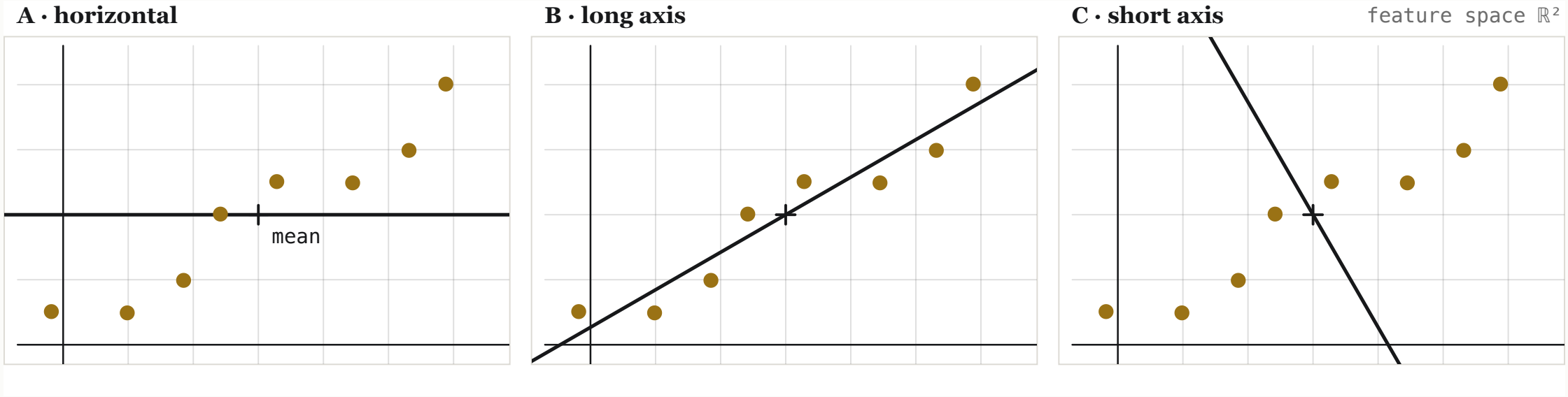
> Sixteen numbers, but a thin cloud. Could one number per sample nearly do?

VOTE

Principal Component Analysis

Which Line Best Summarizes the Cloud?

Which line through the mean gives the smallest sum of squared perpendicular distances?



A the horizontal line

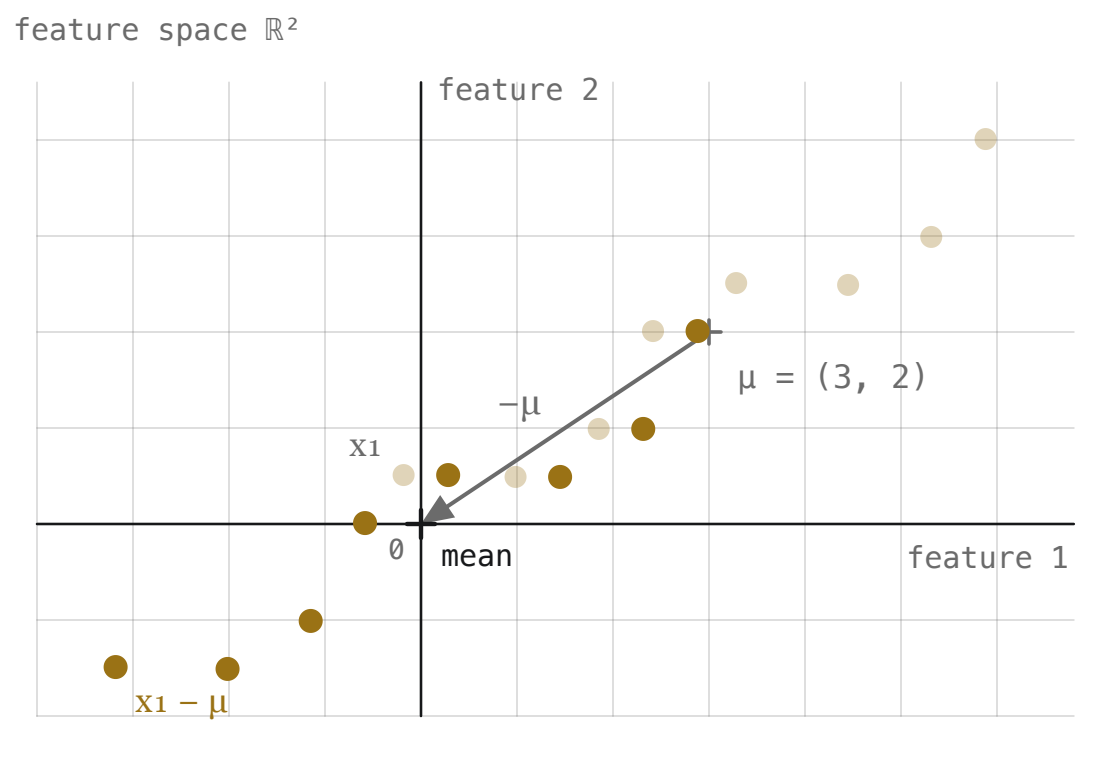
B the long-axis line

C the short-axis line

DEFINITION

Principal Component Analysis

Centering Moves the Mean to the Origin



$$\mu = \frac{1}{N}\sum_{j=1}^{N} x_j, \qquad X_c = X - \mathbf{1}\mu^{\mathsf{T}}$$

$\mu \in \mathbb{R}^d$ mean sample, in feature space

$\mathbf{1} \in \mathbb{R}^N$ all ones; $\mathbf{1}\mu^{\mathsf{T}}$ is $N \times d$, each row μ^{T}

X_c $N \times d$, row j is $(x_j - \mu)^{\mathsf{T}}$; columns sum to 0

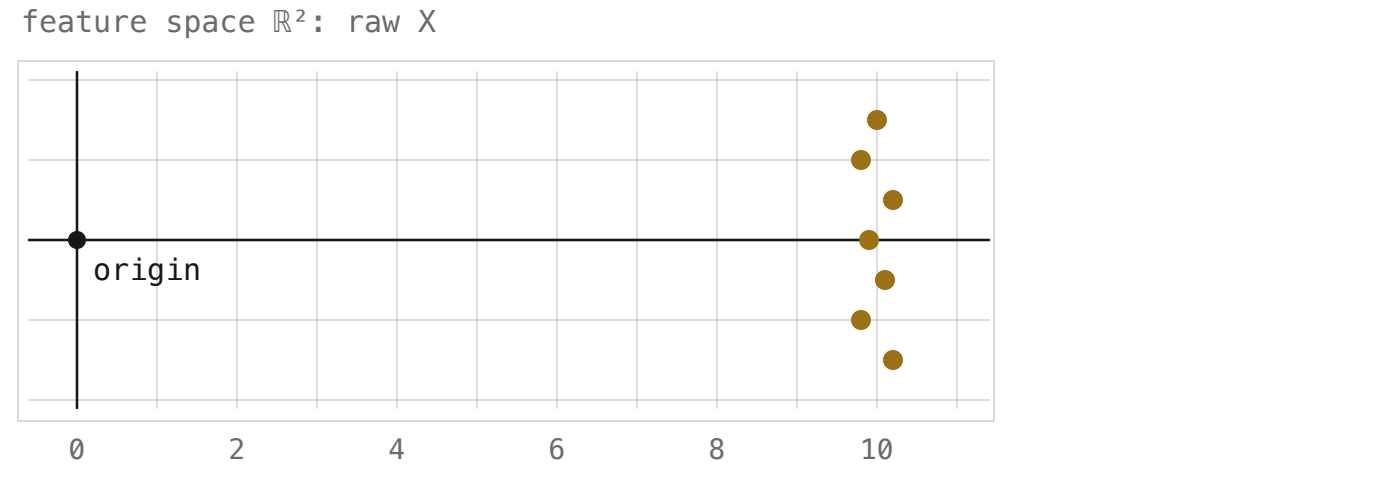
WORKED EXAMPLE

Row sums $(24, 16)$: $\mu = (3, 2)$; $x_1 \approx (-0.18, 0.51)$, $x_1 - \mu \approx (-3.18, -1.49)$

VOTE

Principal Component Analysis

What Does an Uncentered SVD Find?



Uncentered X: what does its first right singular vector v_1 mostly reflect?

A. the cloud's offset from the origin

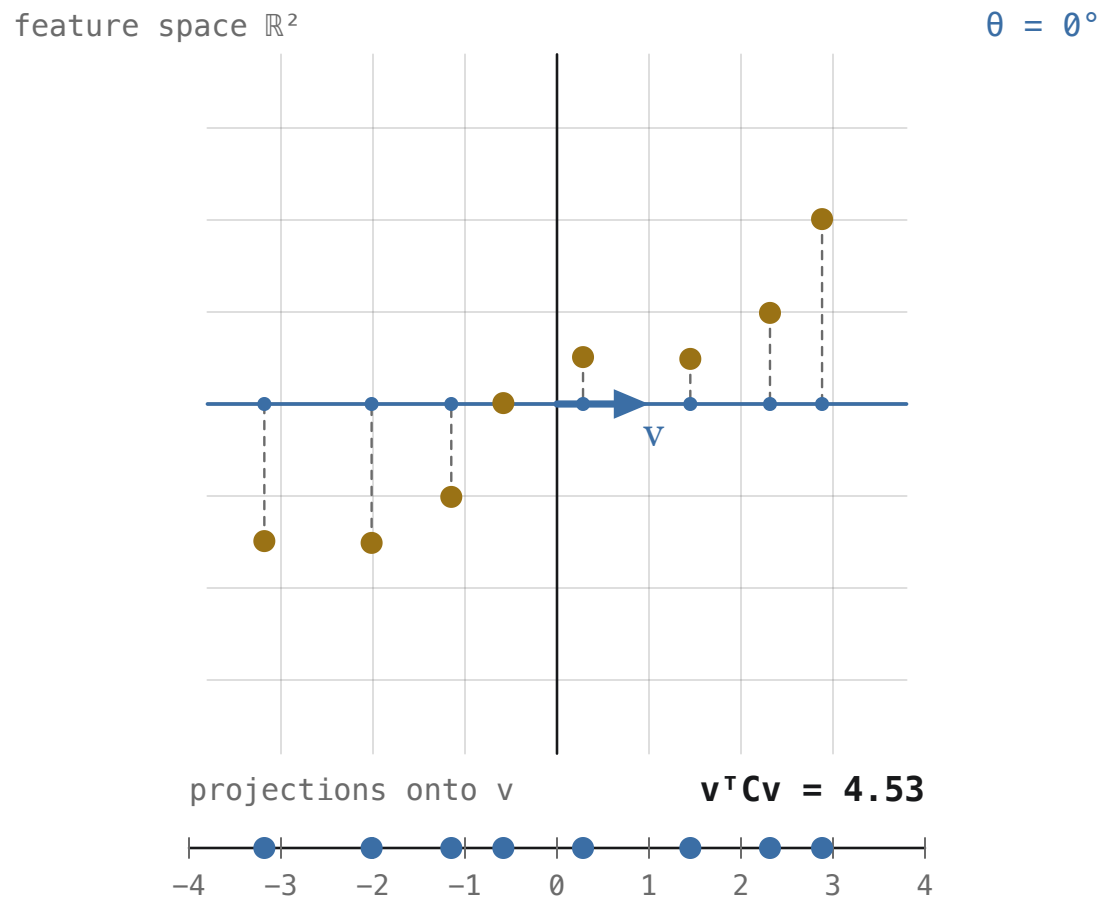
B. the largest variation around the mean, guaranteed

C. the number of samples N

DEFINITION

Principal Component Analysis

Sample Covariance and Directional Variance



$$C = \frac{X_c^\top X_c}{N-1}, \qquad \frac{1}{N-1}\sum_j (x_{c,j}^\top v)^2 = v^\top C v$$

$x_{c,j}^\top$	row j of X_c
$v \in \mathbb{R}^d$	unit direction
left side	variance of the N projections (mean 0): sum of squares over $N-1$
C	$d \times d$, symmetric, PSD; $N > 1$

WORKED EXAMPLE

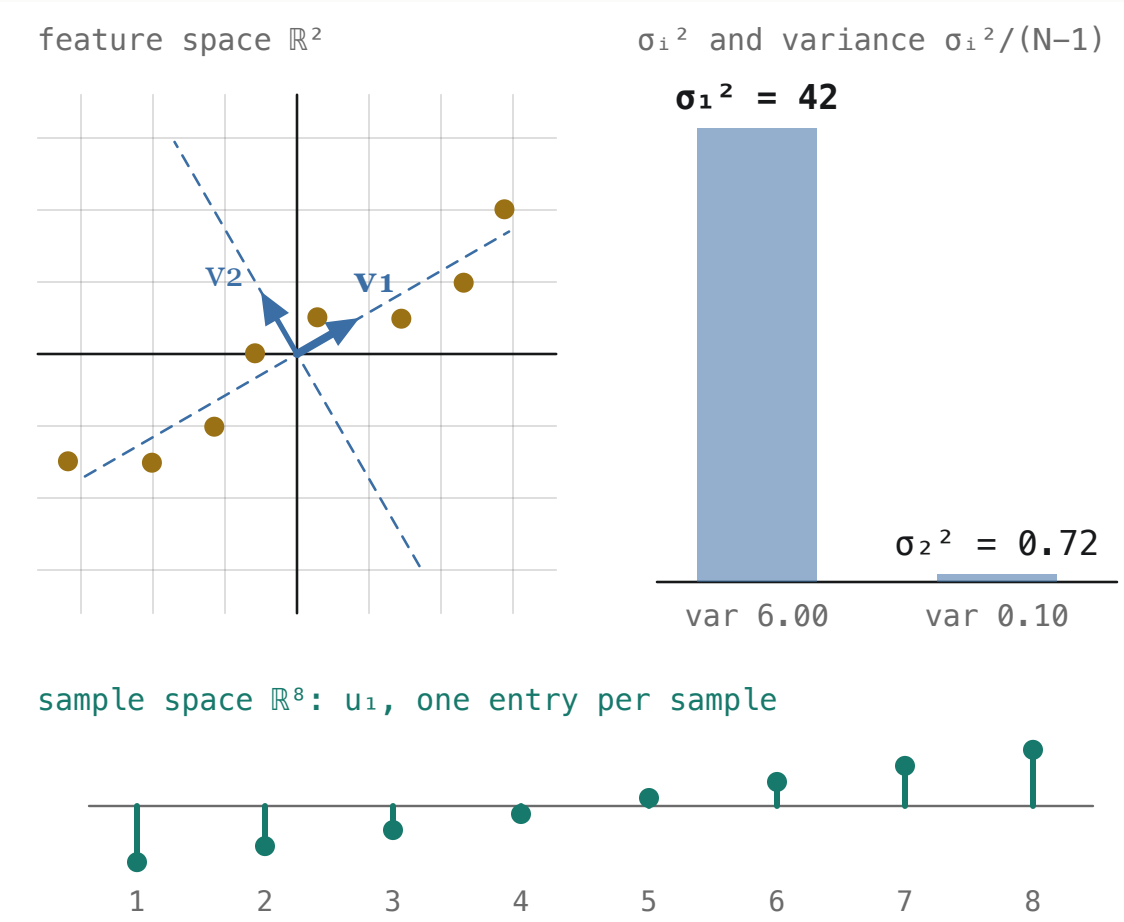
$v = e_1$: $v^\top C v = C_{11} = 31.68/7 \approx 4.53$, the variance of feature 1.

DERIVATION

Principal Component Analysis

The SVD of Centered Data Gives PCA

$\mu = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$



$$X_c = U\Sigma V^\top, \qquad Cv_i = \frac{\sigma_i^2}{N-1} v_i$$

why $X_c^\top X_c = V\Sigma^\top \Sigma V^\top$

$v_i \in \mathbb{R}^d$ feature space: principal directions; v_1 maximizes $v^\top C v$ over unit v

$u_i \in \mathbb{R}^N$ sample space: $u_i = X_c v_i / \sigma_i$ if $\sigma_i > 0$; one entry per sample

σ_i^2 $= \|X_c v_i\|^2$; variance along v_i is $\sigma_i^2/(N-1)$

VOTE

Principal Component Analysis

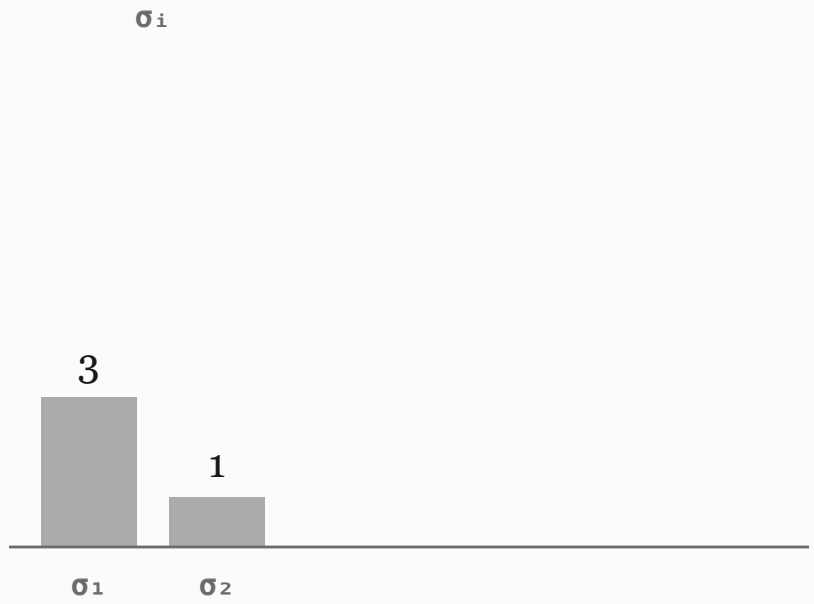
How Much Variance Does One Component Keep?

A new small example: centered data have singular values $\sigma_1 = 3$, $\sigma_2 = 1$. What share of the total variance does the first component keep?

A 75%

B 90%

C 50%

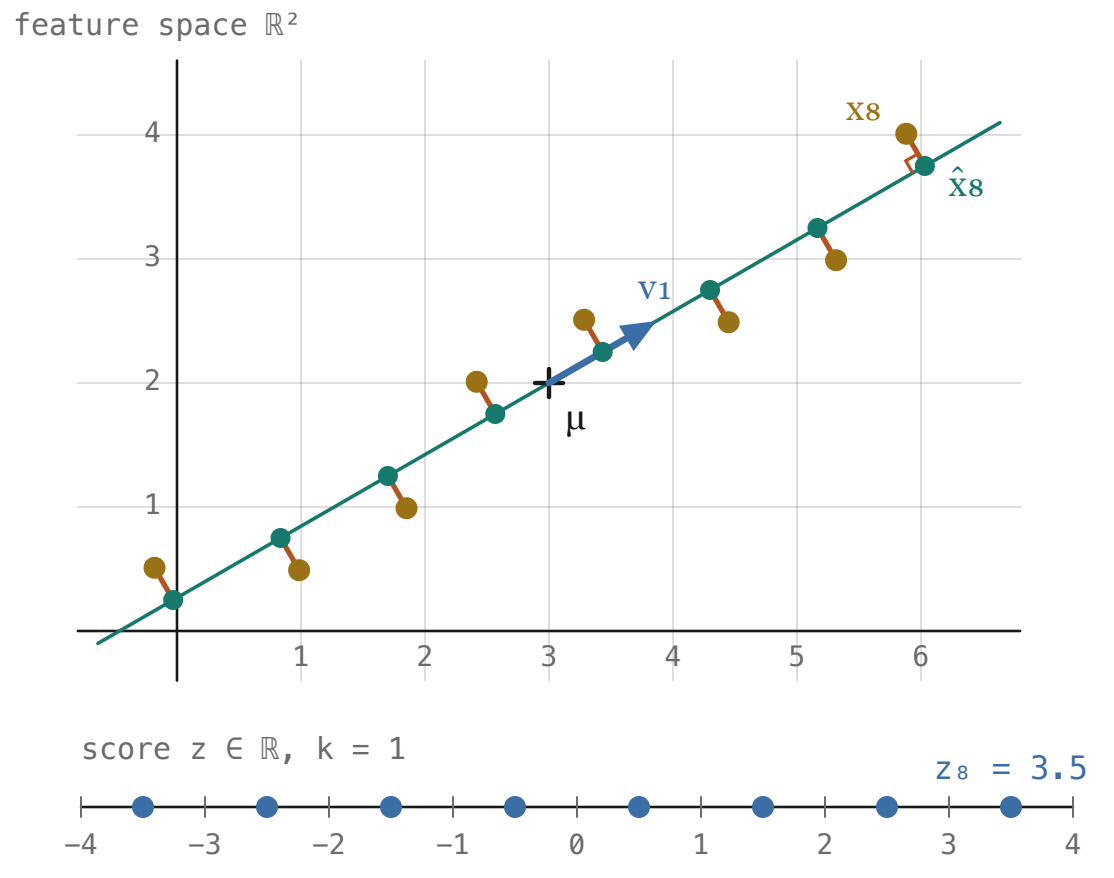


DEFINITION

Principal Component Analysis

Coordinates, Then Reconstruction

$\mu = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$



$$Z = X_c V_k, \quad \hat{X} = Z V_k^\top + \mathbf{1}\mu^\top$$

- V_k first k columns of V, $d \times k$, orthonormal
- Z $N \times k$ scores, one row per sample
- $\hat{X}$ $N \times d$; $Z V_k^\top = X_c V_k V_k^\top$: best rank-$k$ approximation of X_c

WORKED EXAMPLE

$x_8 \approx (5.88, 4.01)$: $z_8 = (x_8 - \mu)^\top v_1 \approx 3.5$, $\hat{x}_8 = \mu + 3.5 v_1 \approx (6.03, 3.75)$, error 0.3

TRANSFER · CENTER, ENCODE, RECONSTRUCT

Principal Component Analysis

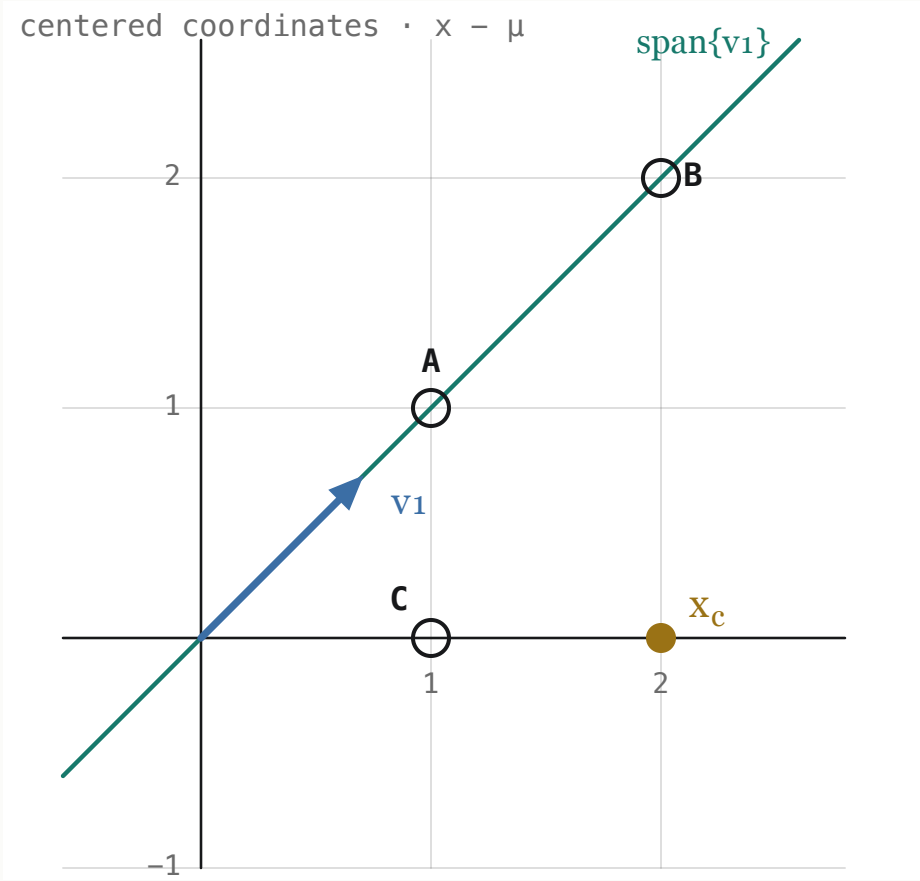
Center, Encode, Reconstruct

$\mu = (1, -1)$, $x = (3, -1)$, $v_1 = (1, 1)/\sqrt{2}$. Find z and $\hat{x}$ in original coordinates; diagram: centered.

A $z = \sqrt{2},\ \hat{x} = (2, 0)$

B $z = 2,\ \hat{x} = (3, 1)$

C $z = 1,\ \hat{x} = (2, -1)$



IN PRACTICE

Principal Component Analysis

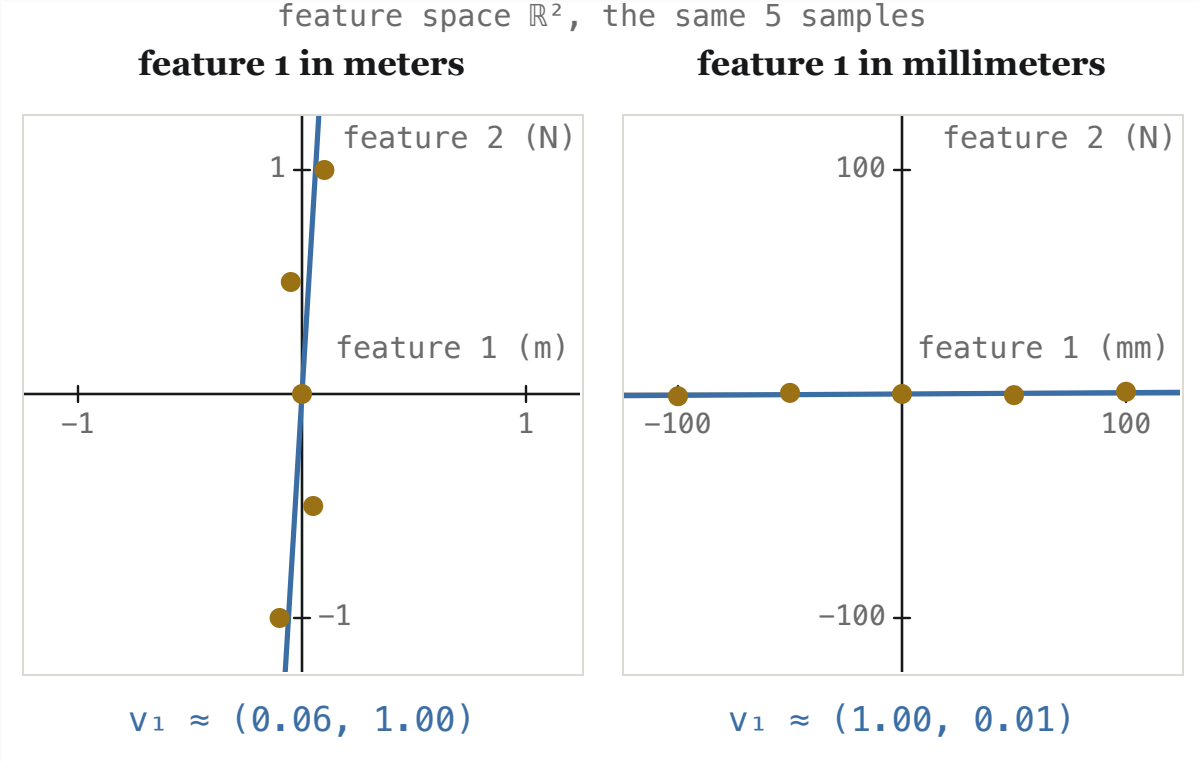
Units, Scale and What Variance Means

$$\text{kept}_k = \frac{\sum_{i \le k} \sigma_i^2}{\sum_i \sigma_i^2}$$

σ_i singular values of X_c

kept_k share in k components; undefined if $X_c = 0$

- Same samples, new units, new v_1.
- Center first. Standardizing is a modeling choice, not an automatic cure.
- Large variance need not mean task relevance.



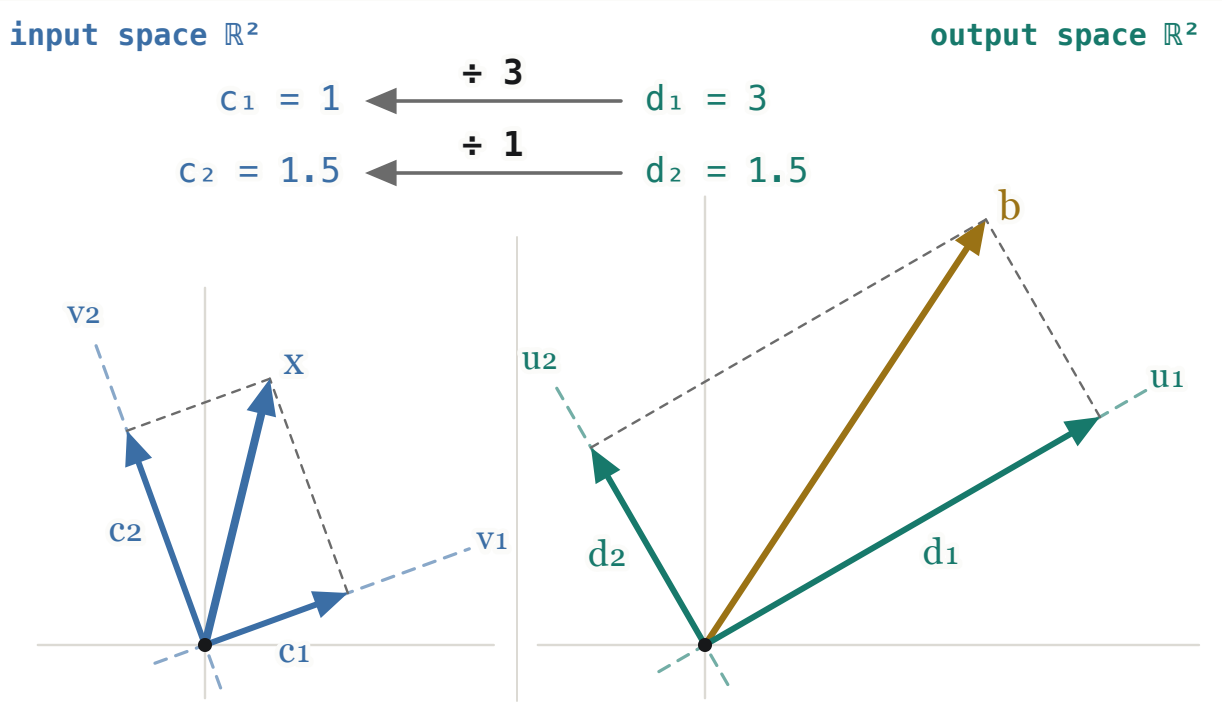
$$A \approx \begin{bmatrix} 2.61 & 0.42 \\ 1.11 & 1.33 \end{bmatrix}$$

Pseudoinverse and Conditioning

Reversing the SVD One Coordinate at a Time

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned} \iff \begin{aligned} \sigma_1 c_1 &= d_1 \\ \sigma_2 c_2 &= d_2 \end{aligned}$$

- $c = V^\top x$ — x along v_1, v_2
- $d = U^\top b$ — b along u_1, u_2
- $\sigma_i > 0$ — one per row: $c_i = d_i/\sigma_i$
- $\sigma_i = 0 \neq d_i$ — $0 = d_i$: no solution
- $\sigma_i = 0 = d_i$ — c_i free: take 0



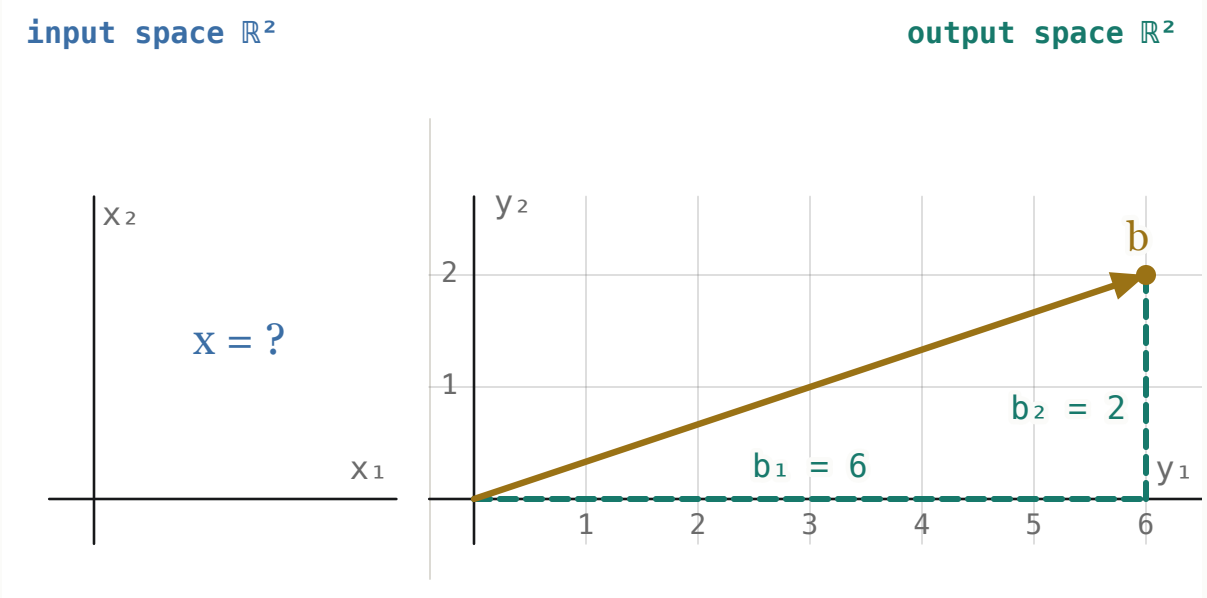
VOTE

Pseudoinverse and Conditioning

Which Input Produces This Output?

$A = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$, $b = (6, 2)$. Which x gives $Ax = b$?

- **A** $x = (18, 2)$
- **B** $x = (2, 2)$
- **C** $x = (2, 6)$



DEFINITION

Pseudoinverse and Conditioning

The Pseudoinverse

$$A^+ = V_r \Sigma_r^{-1} U_r^\mathsf{T} = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^\mathsf{T}$$

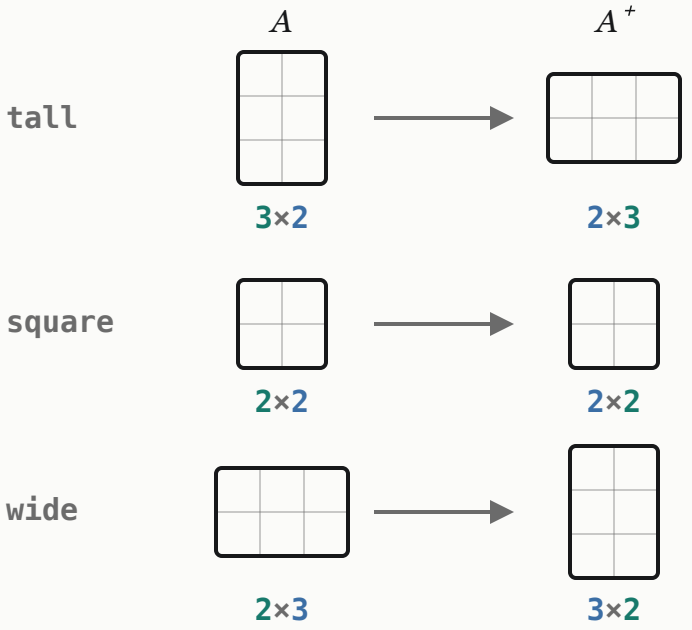
$A \in \mathbb{R}^{m \times n}$ $A^+ \in \mathbb{R}^{n \times m}$; rank r

$\sigma_i > 0$ inverted; a zero gain is not: what is lost stays lost

$r = 0$ $A^+ = 0$

WORKED EXAMPLE

$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 2 \end{bmatrix}$, $\sigma = (2, 1)$: $A^+ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$



VOTE

Pseudoinverse and Conditioning

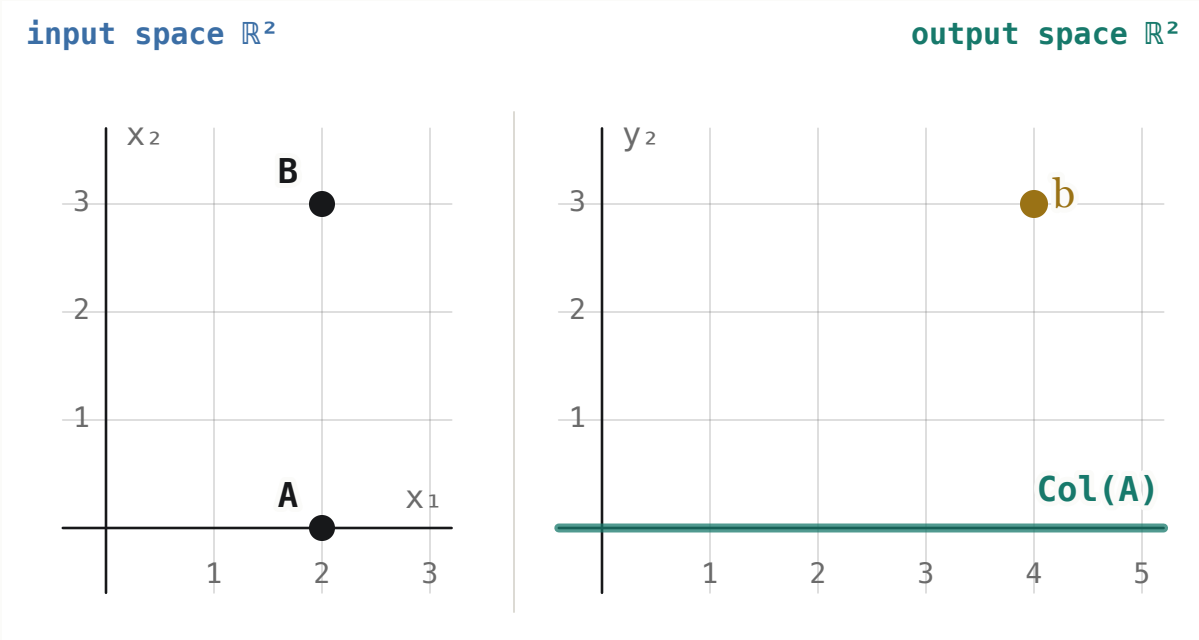
What Does the Pseudoinverse Return?

$A = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, $b = (4, 3)$. What is $A^{+}b$?

A $(2, 0)$

B $(2, 3)$

C none: no exact solution



PICTURE

Pseudoinverse and Conditioning

Two Projectors: Output Side and Input Side

$$AA^{+} = U_r U_r^{\mathsf{T}}, \qquad A^{+}A = V_r V_r^{\mathsf{T}}$$

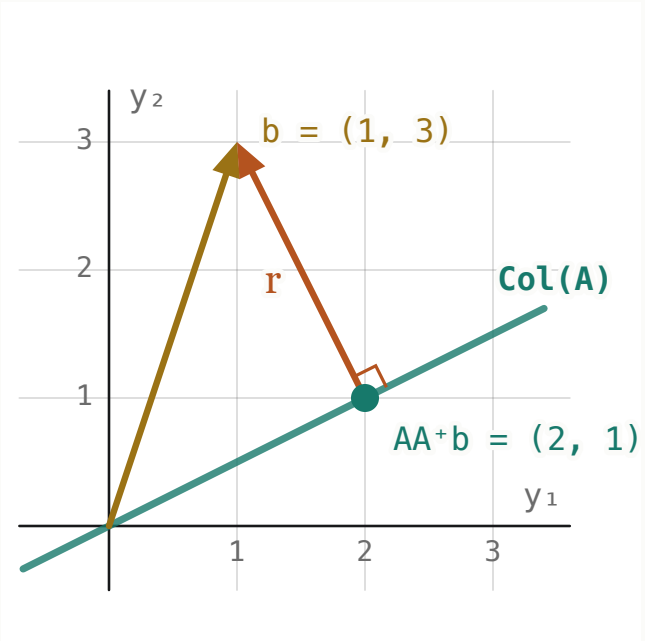
AA^{+} $m \times m$, projects onto $\mathrm{Col}(A)$

$A^{+}A$ $n \times n$, projects onto $\mathrm{Row}(A)$

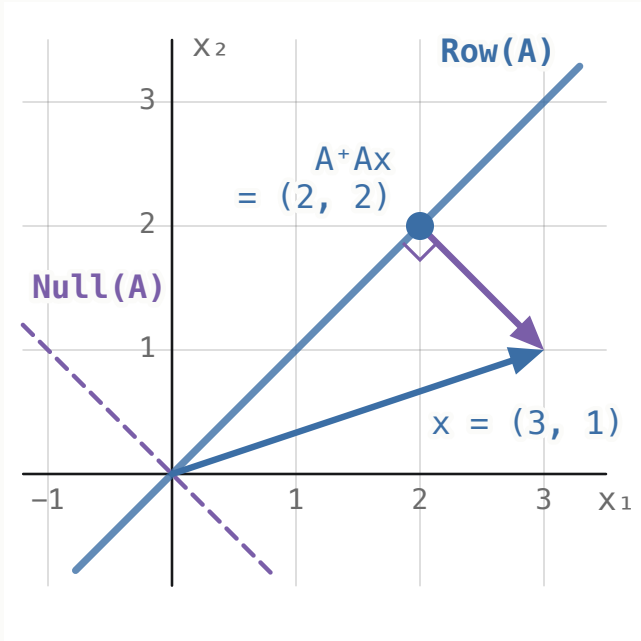
$\neq I$ unless $r = m$, resp. $r = n$

Here $A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$, rank 1.

output space $\mathbb{R}^m$, $m = 2$



input space $\mathbb{R}^n$, $n = 2$



DEFINITION

Pseudoinverse and Conditioning

Minimum-Norm Least Squares

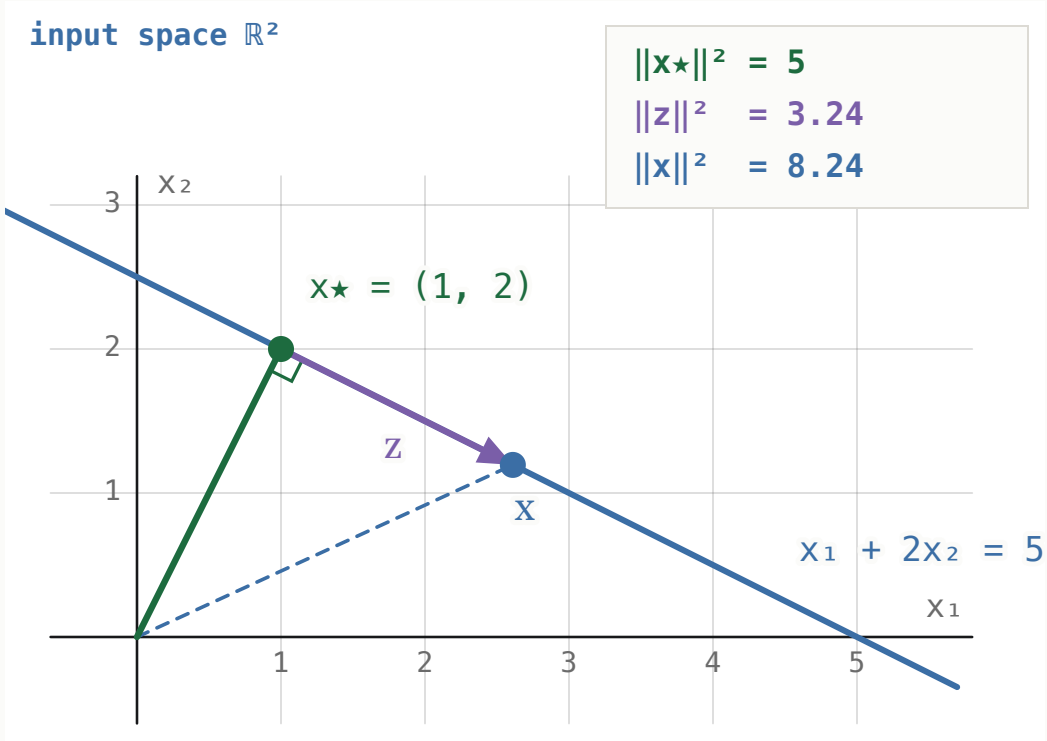
First minimize the residual $\|b - Ax\|$; then take the smallest norm among tied minimizers.

$$x = x^\star + z: \quad \|x\|^2 = \|x^\star\|^2 + \|z\|^2$$

- x any minimizer; $z \in \mathrm{Null}(A)$
- $x^\star = A^+ b$ in $\mathrm{Row}(A)$, so $x^\star \perp z$
- $\|x\|^2$ squared norm: least only at $z = 0$

WORKED EXAMPLE

$A = [1 \;\; 2]$, $b = 5$: $A^+ = (1, 2)/5$, $x^\star = (1, 2)$, $\|x^\star\|^2 = 5$



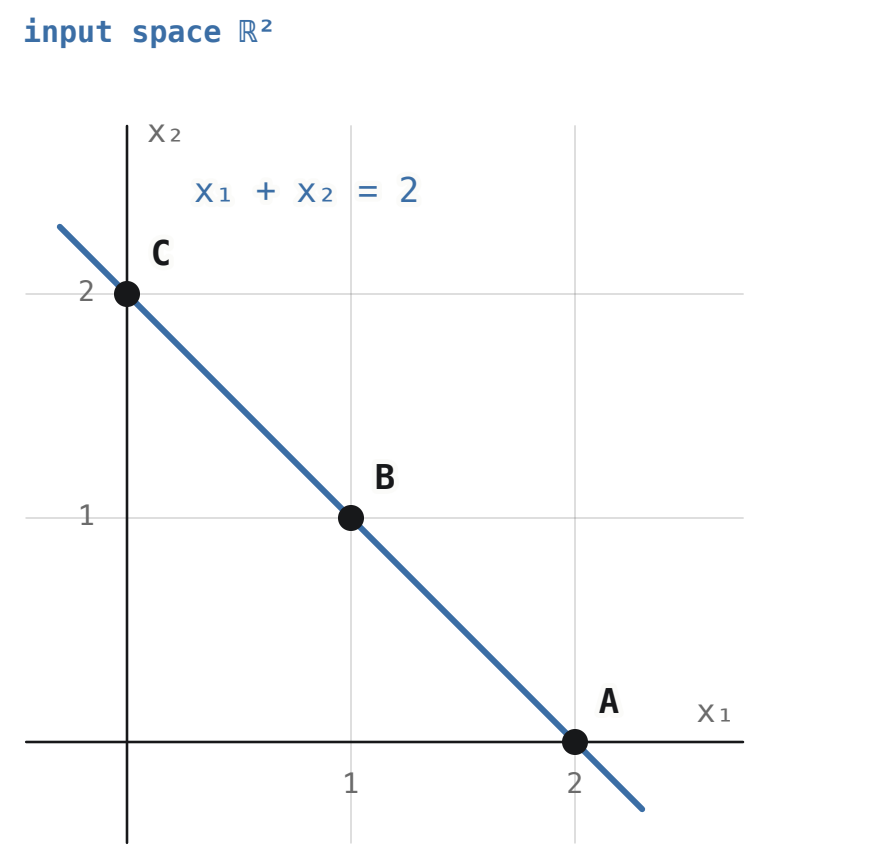
VOTE

Pseudoinverse and Conditioning

Which Solution Has the Smallest Norm?

All three solve $x_1 + x_2 = 2$. Which has the smallest norm $\|x\|$?

- **A** $(2, 0)$
- **B** $(1, 1)$
- **C** $(0, 2)$



PICTURE

Pseudoinverse and Conditioning

Small Gains Amplify Noise

$$\delta x = \sum_{\sigma_i > 0} \frac{u_i^{\mathsf{T}} \delta b}{\sigma_i} v_i$$

$\delta x \in \mathbb{R}^n$ input error (mm)

$\delta b \in \mathbb{R}^m$ output error (mm)

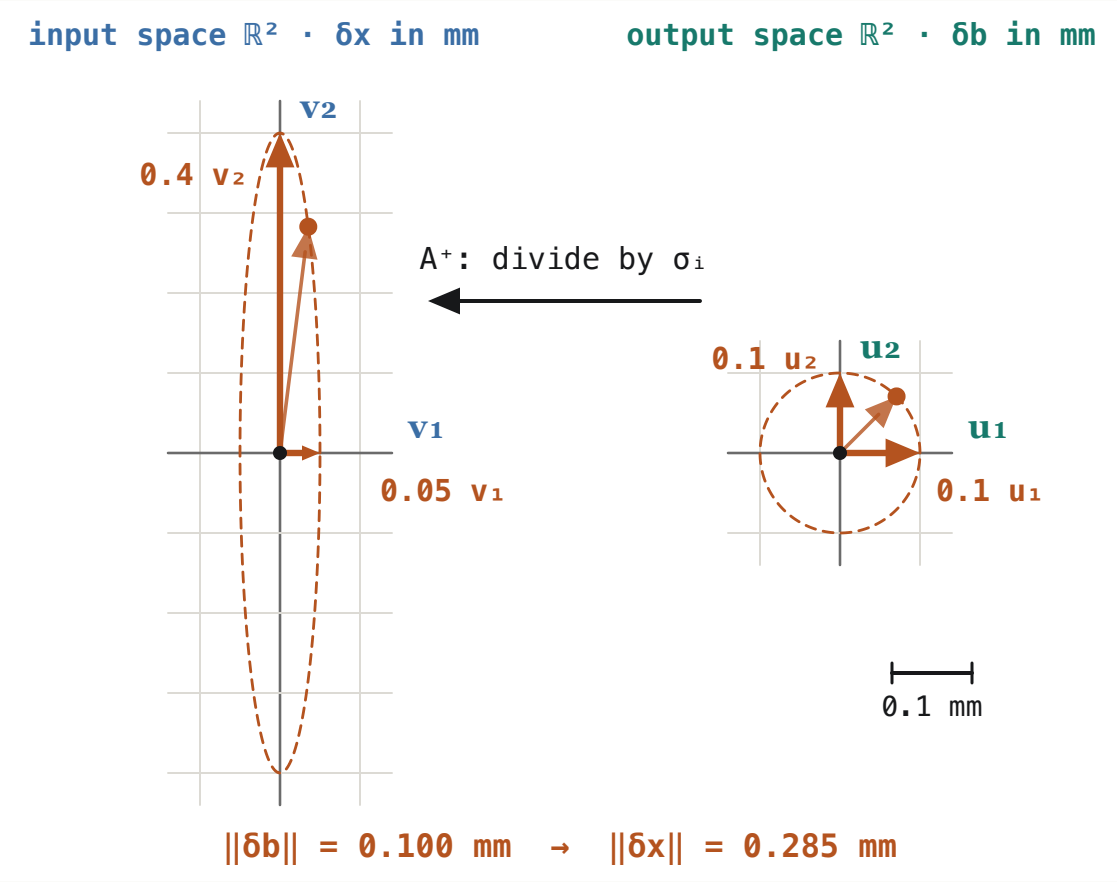
A fixed $x^\star + \delta x = A^+(b + \delta b)$: only b moves

WORKED EXAMPLE

$A = \mathrm{diag}(2, 0.25)$, $\|\delta b\| = 0.1$ mm.

Along u_1: $0.1/2 = 0.05$ mm along v_1.

Along u_2: $0.1/0.25 = 0.4$ mm along v_2: 8 times.



VOTE

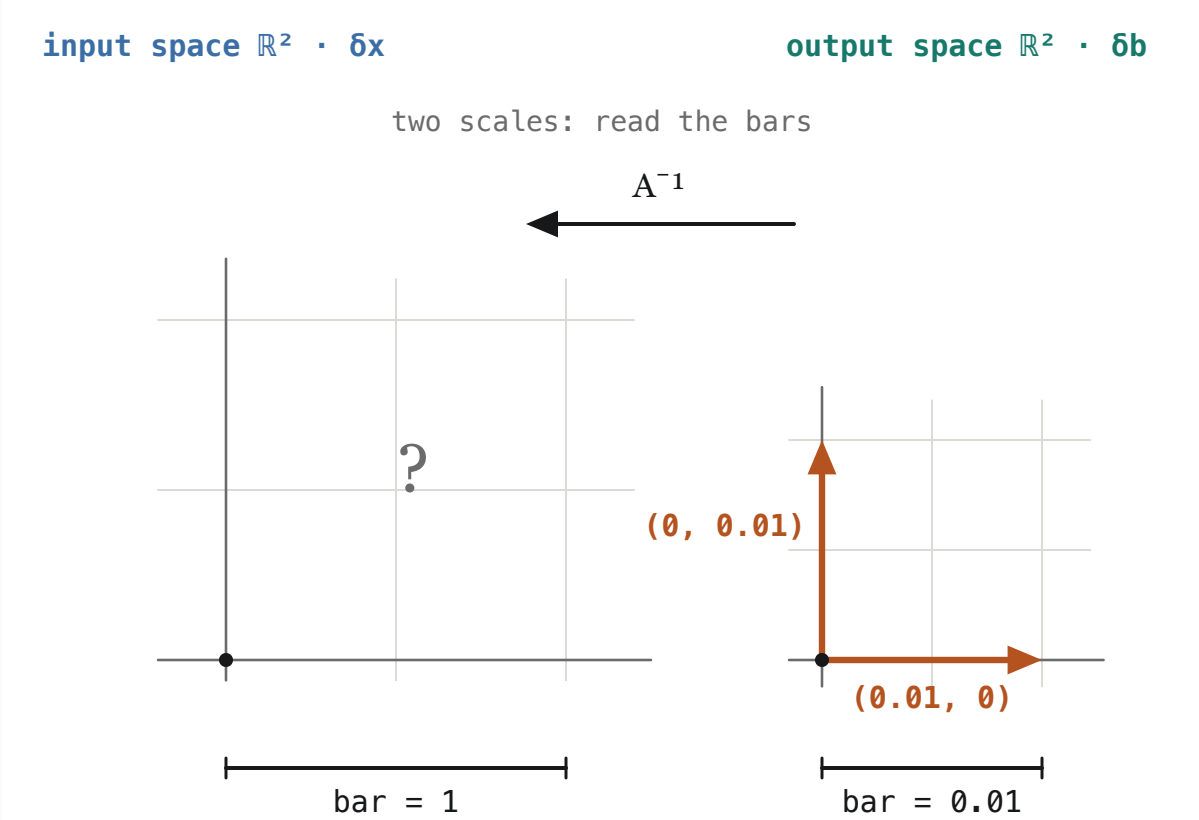
$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0.01 \end{bmatrix}$$

Pseudoinverse and Conditioning

Which Perturbation Moves the Input More?

Both output errors have length 0.01. Which gives the larger input error $\|\delta x\|$?

- (A) $\delta b = (0.01, 0)$
- (B) $\delta b = (0, 0.01)$
- (C) the same for both



DEFINITION

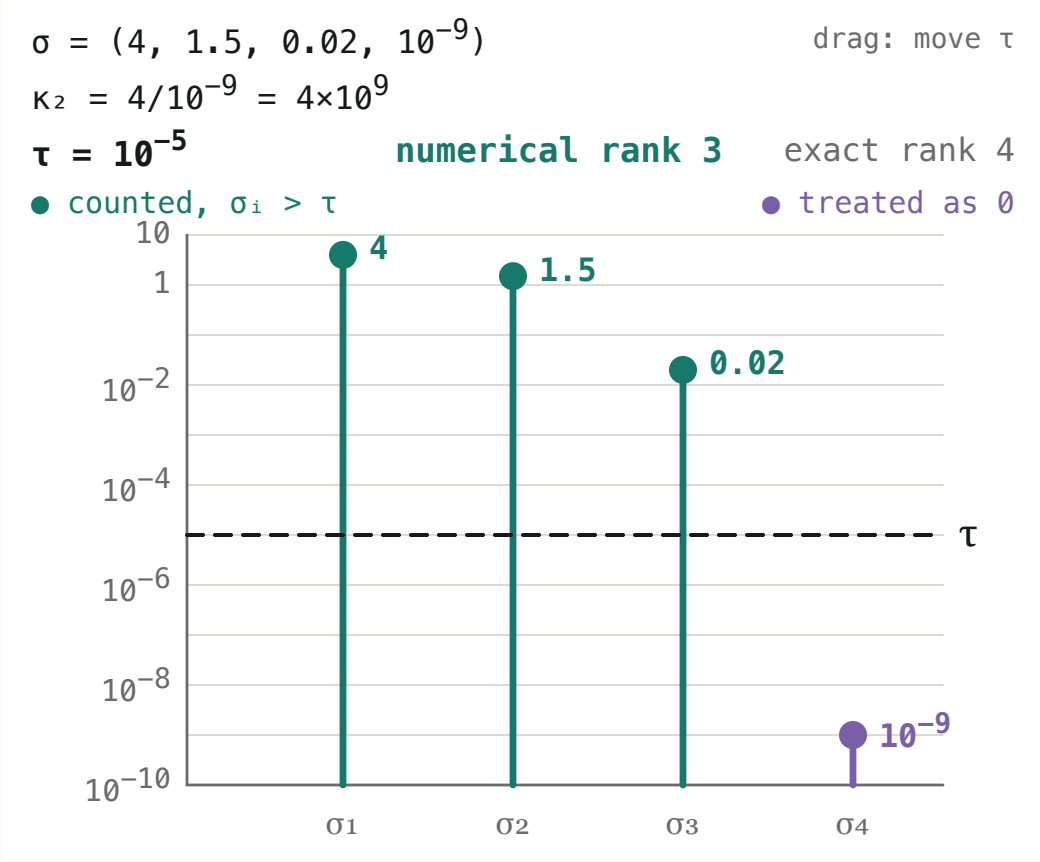
Pseudoinverse and Conditioning

Conditioning and Rank

$$\|A\|_2 = \sigma_{\max}, \quad \kappa_2(A) = \frac{\sigma_{\max}}{\sigma_{\min}}, \quad r_\tau = \#\{i : \sigma_i > \tau\}$$

$\|A\|_2, r_\tau$	any A, $\tau \geq 0$
κ_2	full rank; else ∞ by convention
$1/\sigma_{\min}$	absolute; κ_2 relative
scope	fixed invertible A; $Ax = b \neq 0$, only b moves

Conditioning is the problem's, stability the algorithm's.

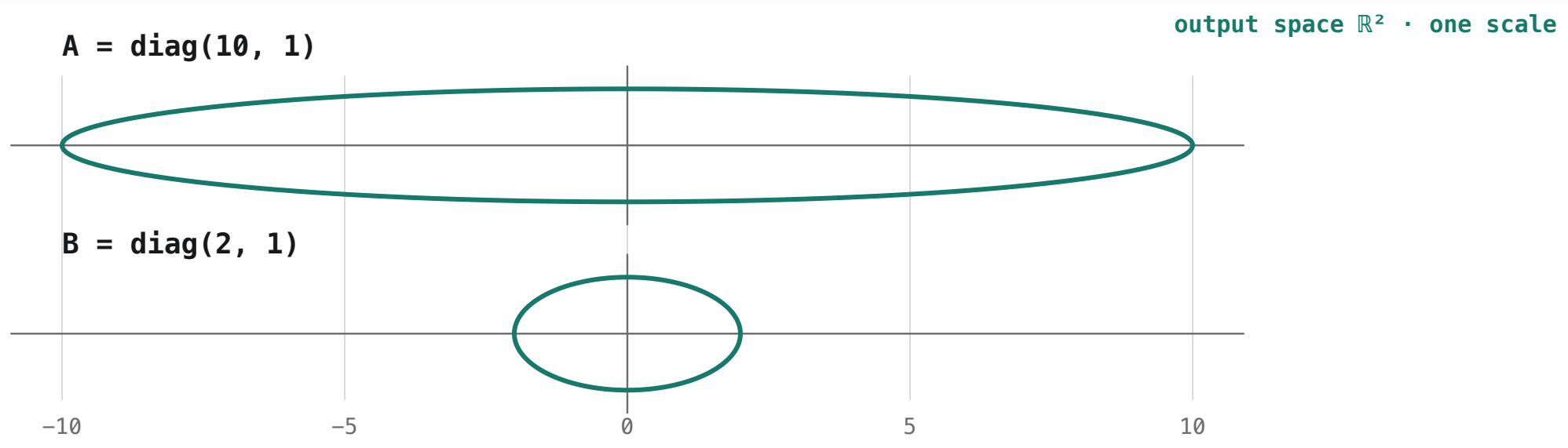


VOTE

Pseudoinverse and Conditioning

Which Matrix Is Worse Conditioned?

Image ellipses of A and B at one scale. Which is worse conditioned?



A $A = \mathrm{diag}(10, 1)$

B $B = \mathrm{diag}(2, 1)$

C equally conditioned

DEFINITION

Pseudoinverse and Conditioning

Damped Least Squares

Minimize $\|Ax - b\|^2 + \lambda^2 \|x\|^2$.

$c_i = v_i^\mathsf{T} x$, $d_i = u_i^\mathsf{T} b$, $\lambda > 0$.

$$\phi_i(c_i) = (\sigma_i c_i - d_i)^2 + \lambda^2 c_i^2$$

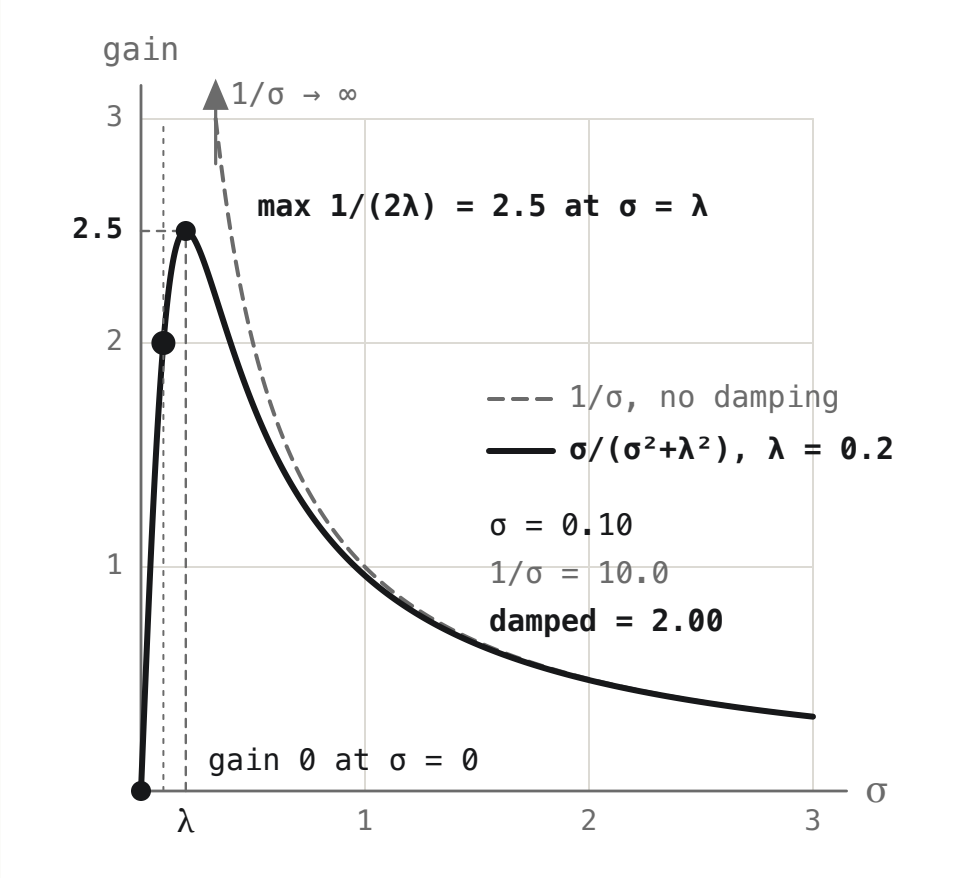
ϕ_i fit + penalty

Use $h_i = \sigma_i^2 + \lambda^2 > 0$. Complete square:

$\phi_i = h_i(c_i - \sigma_i d_i / h_i)^2 + \lambda^2 d_i^2 / h_i$.

> $c_i^\star = \sigma_i d_i / h_i$; $x_\lambda = \sum_{i=1}^{r} c_i^\star v_i$.

> Unique; gain $\le 1/(2\lambda)$.



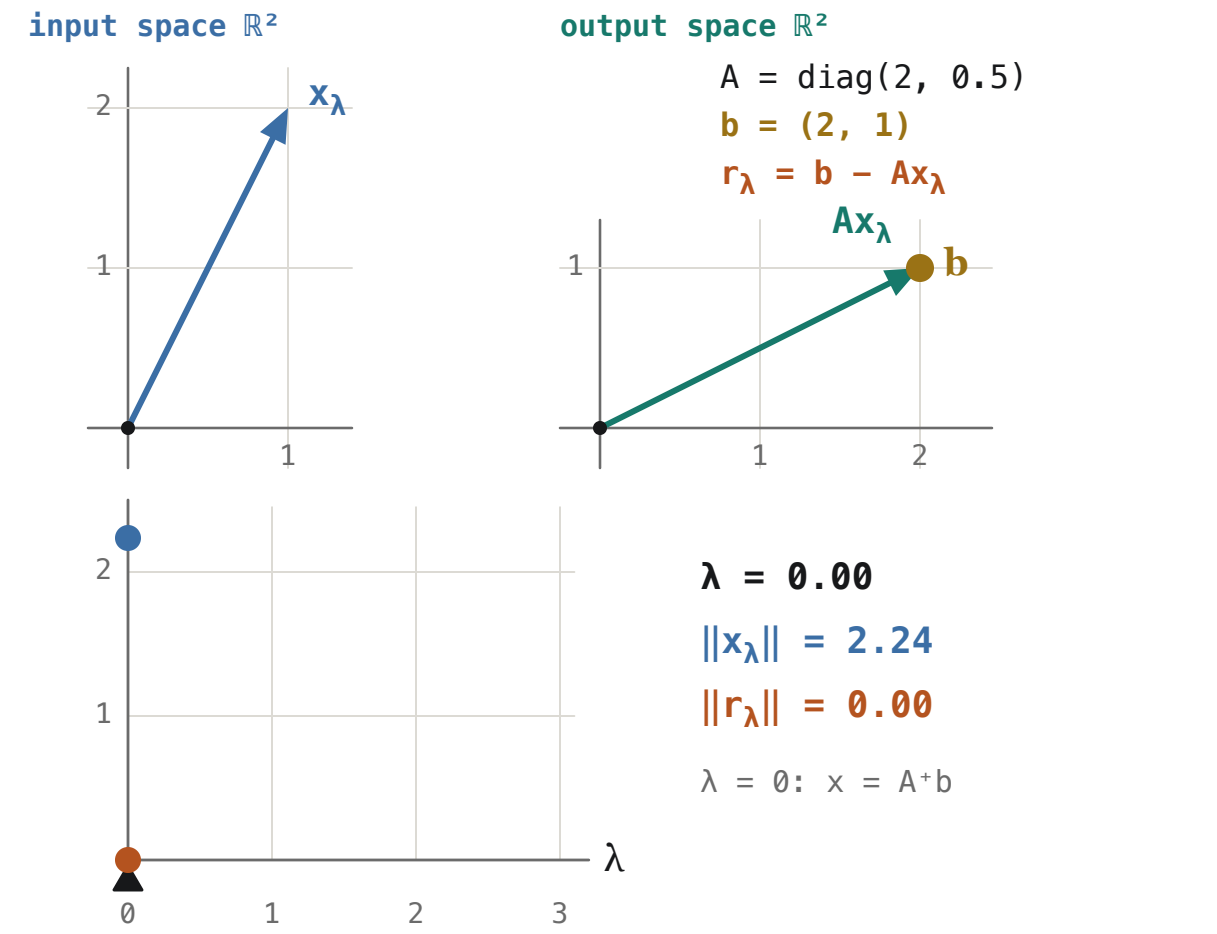
VOTE

Pseudoinverse and Conditioning

What Happens as Damping Grows?

Fix A, b. As λ grows $0 \to 3$, what do $\|x_\lambda\|$ and $\|r_\lambda\|$ do?

- (A) $\|x_\lambda\|$ nonincreasing, $\|r_\lambda\|$ nondecreasing
- (B) $\|x_\lambda\|$ increases, $\|r_\lambda\|$ always decreases
- (C) b stays exactly attained



IN PRACTICE

Pseudoinverse and Conditioning

Choosing a Solver

SOLVER	NEEDS	GIVES
thin QR, Lecture 3	full column rank	$R\hat{x} = Q^\top b$
SVD	any A	$A^+ b$, minimum norm
damped	$\lambda > 0$	x_λ (new objective), gains $\leq 1/(2\lambda)$

$$\kappa_2(A^\top A) = \kappa_2(A)^2$$

A, $r = n$ $\kappa_2 = \sigma_1/\sigma_n$

$A^\top A$ gains σ_i^2

A = diag(10, 1): image in output space



axes 10 : 1

κ₂ = 10

AᵀA = diag(100, 1): image in input space



axes 100 : 1

κ₂ = 100

both scaled to one long axis

In code, solve systems, never invert. QR exists for every A.

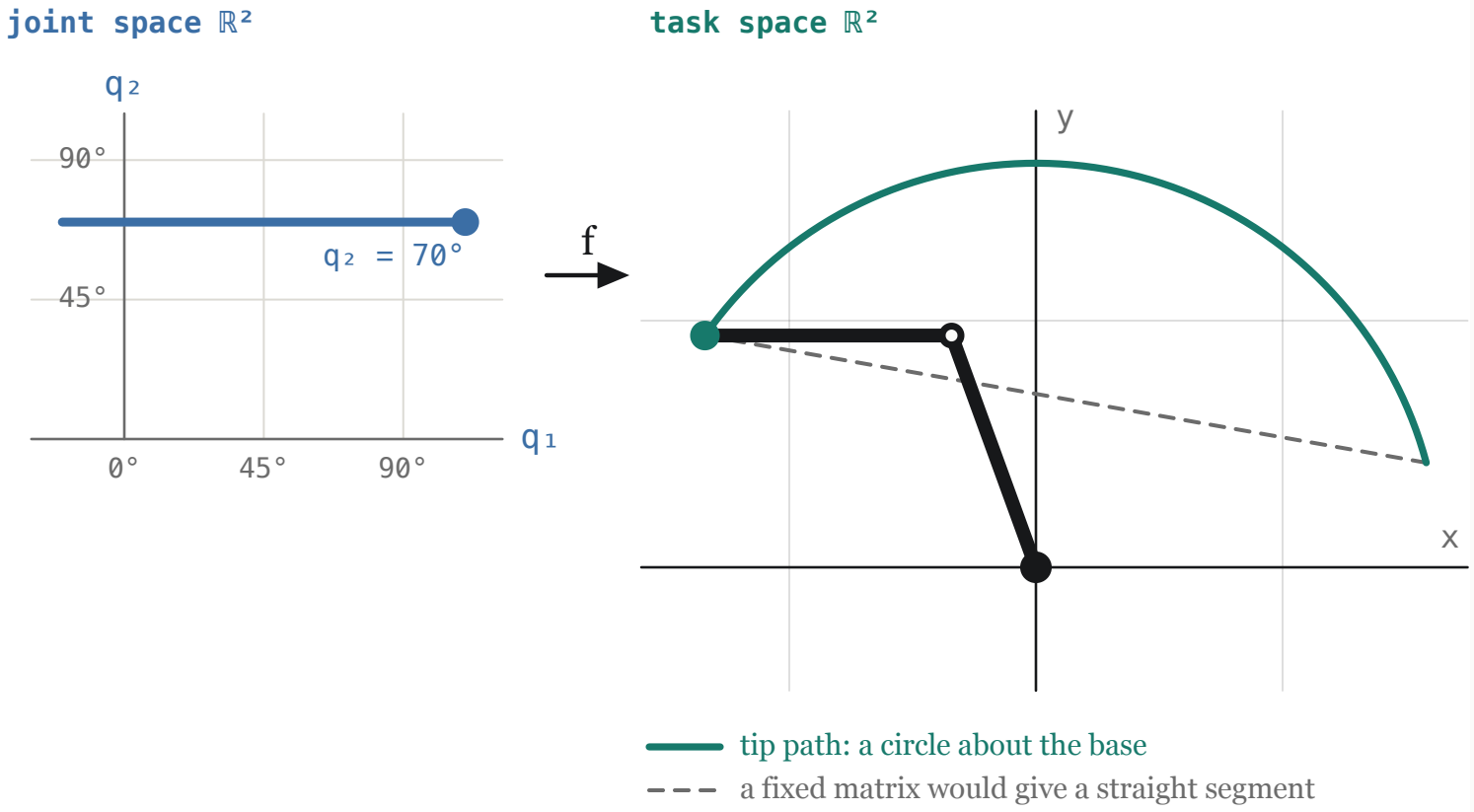
PICTURE

Jacobians and Robot Motion

Forward Kinematics Is Nonlinear

$$p = f(q), \qquad q \in \mathbb{R}^n, \;\; p \in \mathbb{R}^m$$

- q joint vector: n angles in radians here; a prismatic joint would contribute a length
- p tip position only: m task coordinates, no orientation
- f forward kinematics: sines and cosines of q, not a matrix



VOTE

Jacobians and Robot Motion

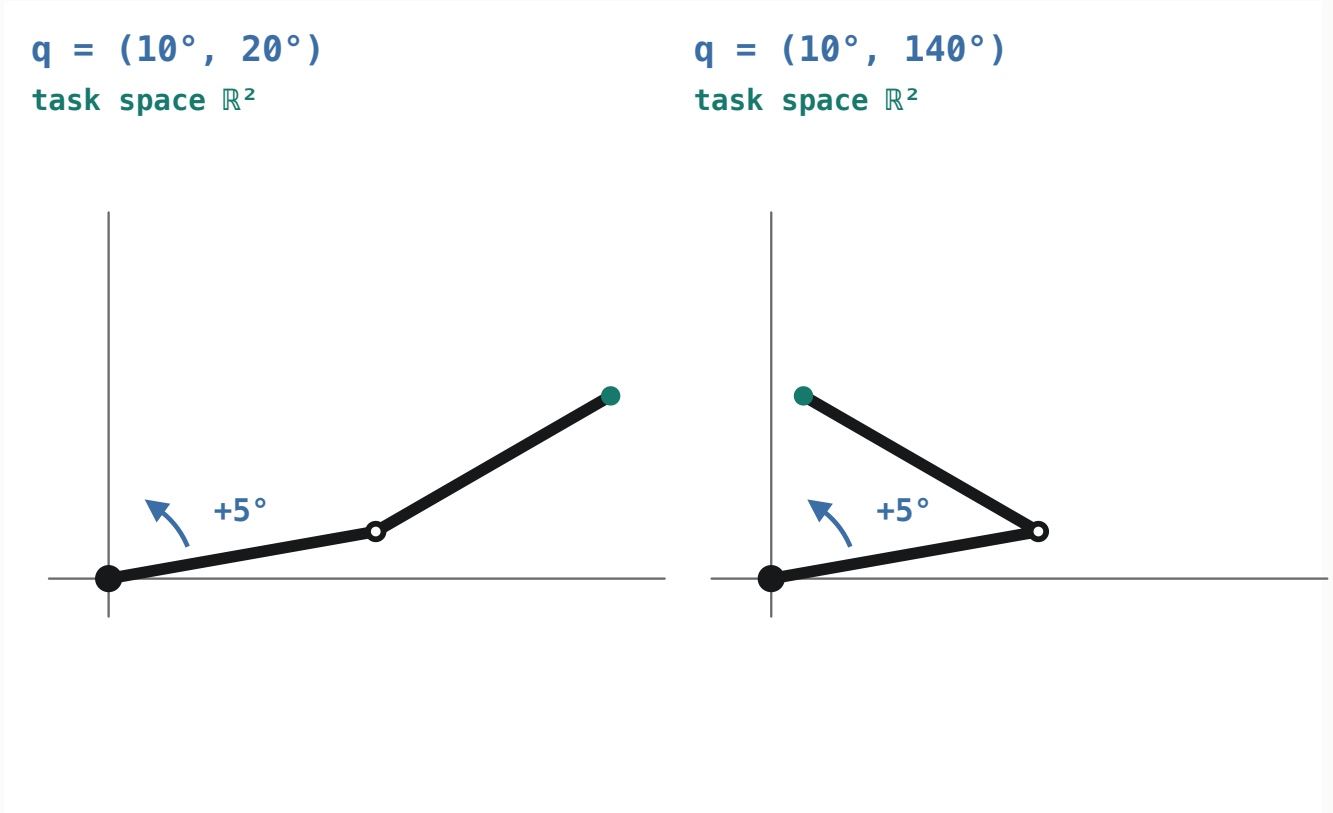
Does the Same Joint Change Move the Tip the Same Way?

Both shoulders turn $+5°$, elbows fixed. The tip changes Δp are…

A always identical

B generally pose-dependent

C determined only by joint count



RECALL

Jacobians and Robot Motion

The Derivative as a Local Gain

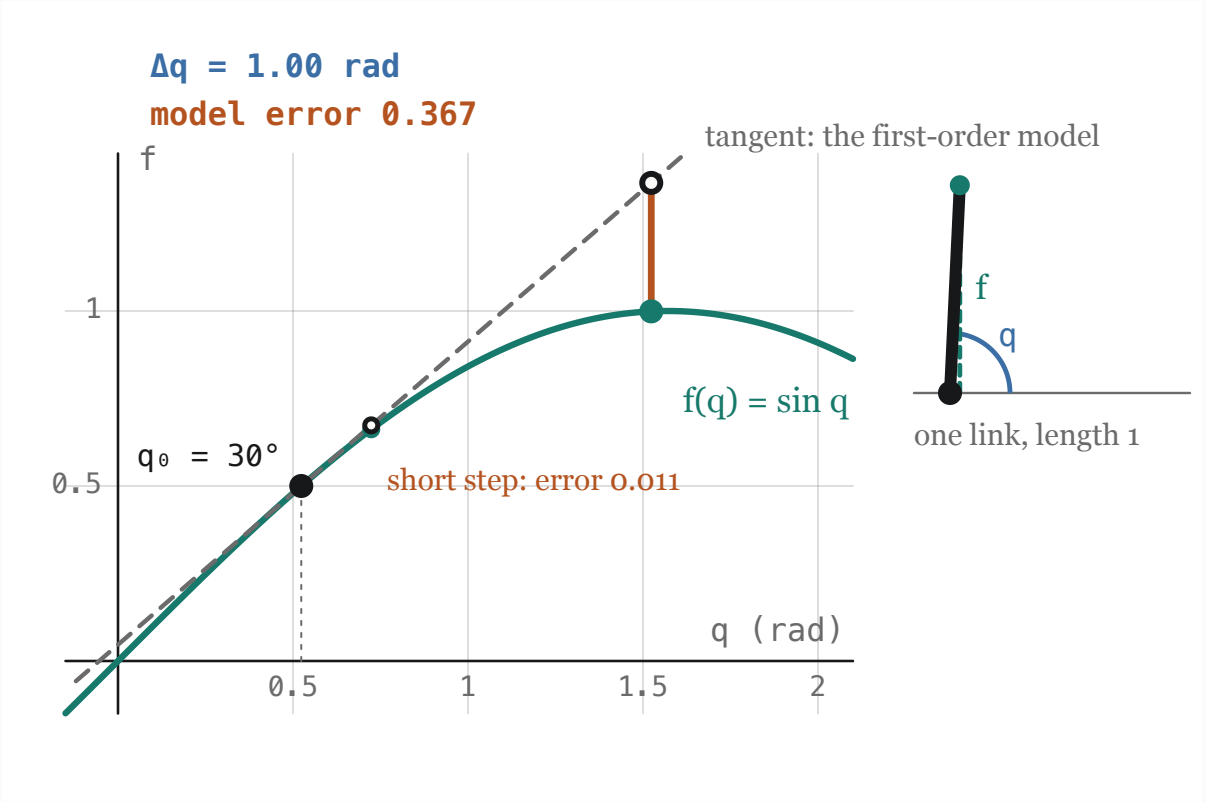
$$f(q + \Delta q) \approx f(q) + f'(q)\,\Delta q$$

$f'(q)$ local gain: output change per radian, at this q

Δq step in radians; for small steps the model error shrinks like Δq^2

WORKED EXAMPLE

$q = 30°$: $f = 0.5$, $f' = \cos 30° \approx 0.866$. Step $\Delta q = 0.2$ rad: model ≈ 0.673, actual ≈ 0.662 .



DEFINITION

Jacobians and Robot Motion

The Jacobian Matrix

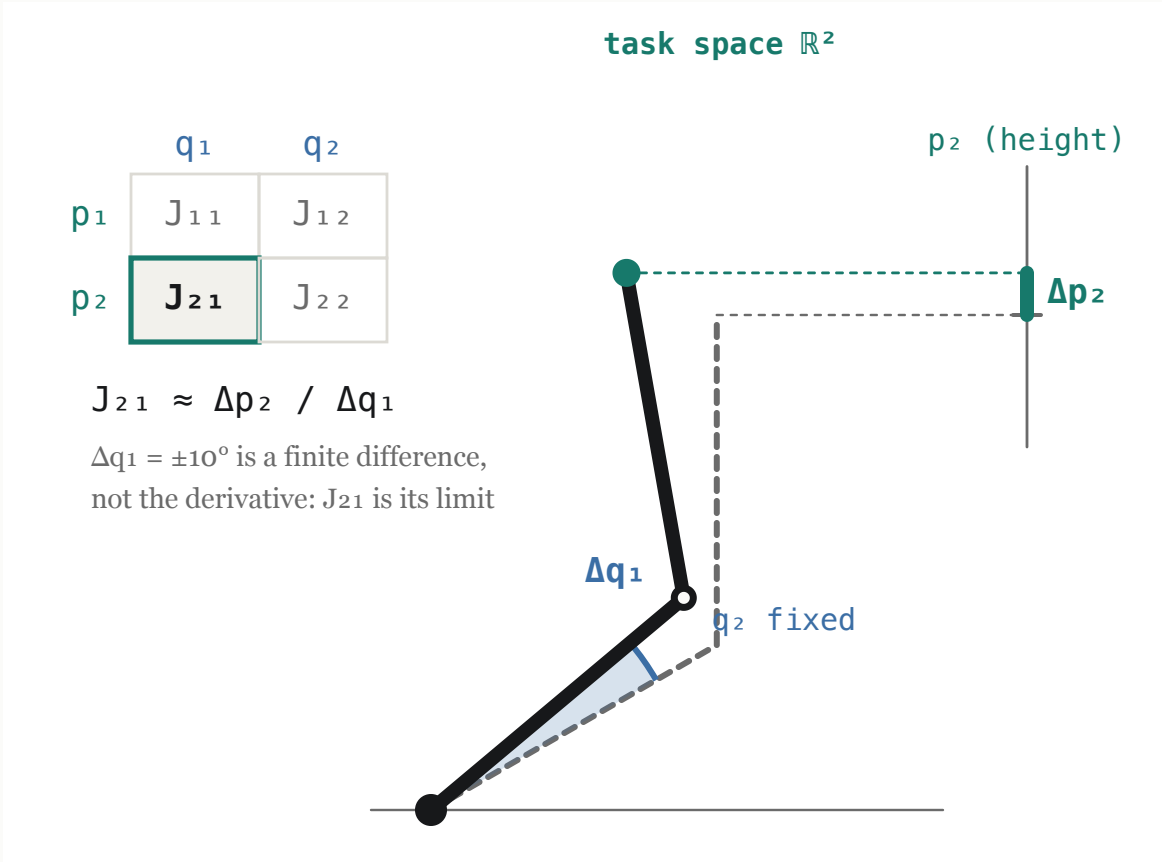
$$J_{ij}(q) = \frac{\partial f_i}{\partial q_j}(q), \qquad J(q) \in \mathbb{R}^{m \times n}$$

row i output $p_i = f_i(q), i = 1, \ldots, m$

column j joint $q_j, j = 1, \ldots, n$; the other joints held fixed

$$\Delta p \approx J(q)\, \Delta q$$

$\Delta q \in \mathbb{R}^n$ a small joint change, in radians; $\Delta p \in \mathbb{R}^m$



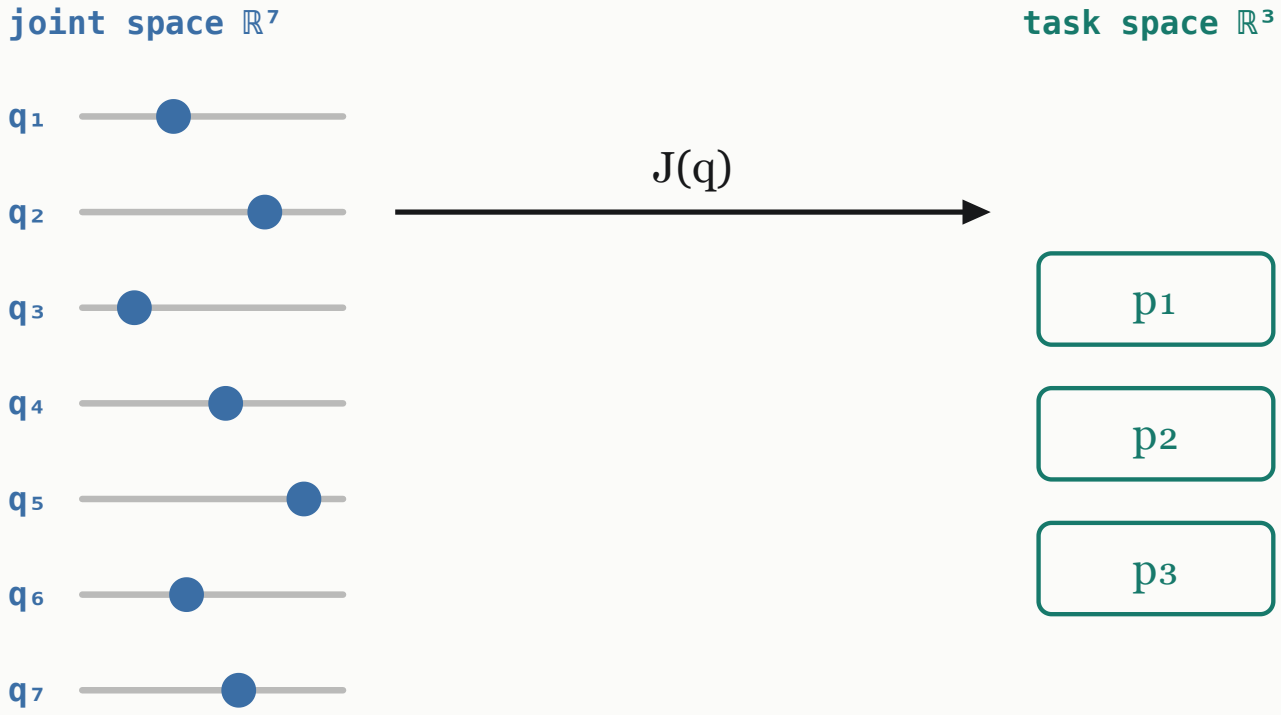
VOTE

Jacobians and Robot Motion

What Shape Is This Jacobian?

An arm has 7 joints. The task is only its 3D tip position. What shape is J?

- (A) 7×3
- (B) 3×7
- (C) 7×7



PICTURE

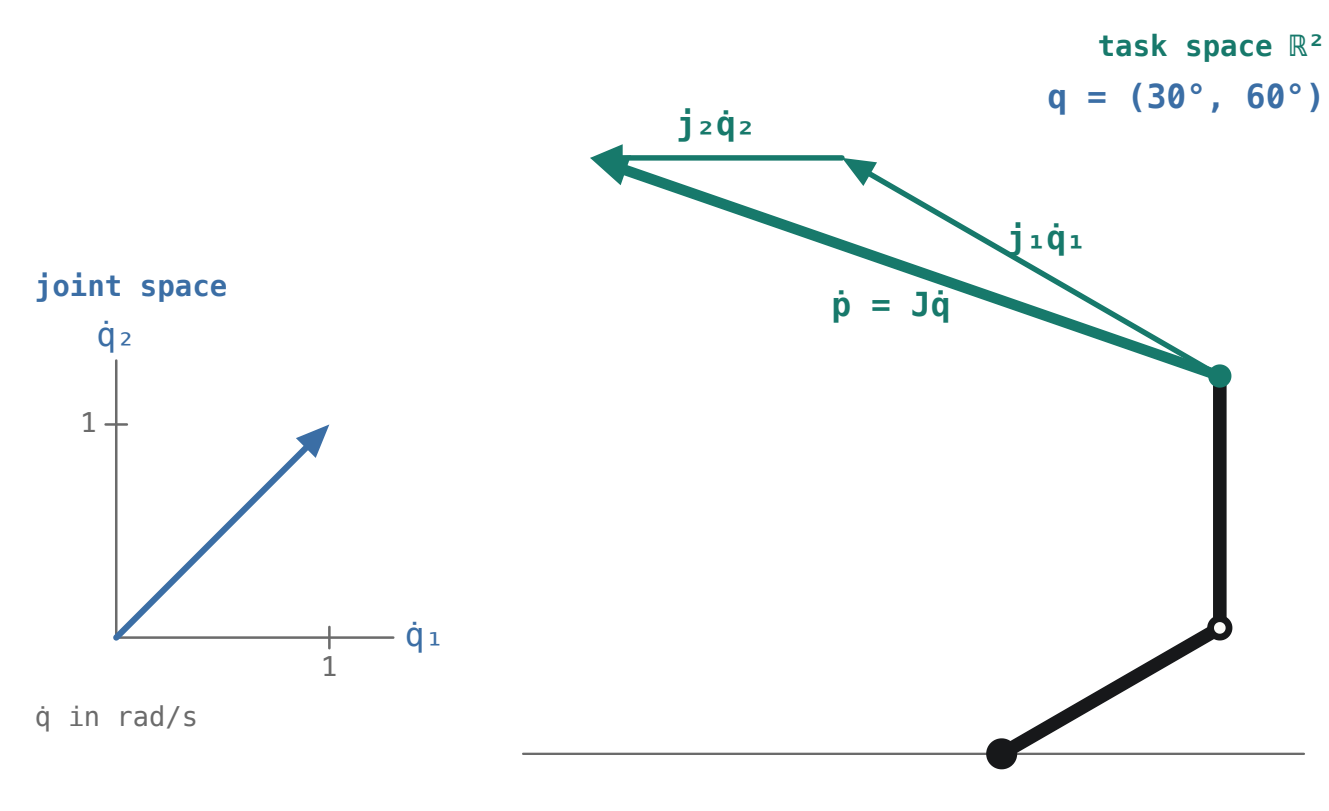
Jacobians and Robot Motion

Columns Are Joint Effects

$$J = \begin{bmatrix} j_1 & \cdots & j_n \end{bmatrix}, \quad J\Delta q = \sum_{j=1}^{n} j_j \, \Delta q_j$$

- $j_j \in \mathbb{R}^m$ — column j: tip length per radian of joint j alone
- Δq_j — how far joint j turns; the contributions add
- $j_j(q)$ — changes with the pose q

Drawn: $\dot{p} = J\dot{q}$ at $\dot{q} = (1, 1)$ rad/s, $q = (30°, 60°)$: velocities, not paths.



VOTE

Jacobians and Robot Motion

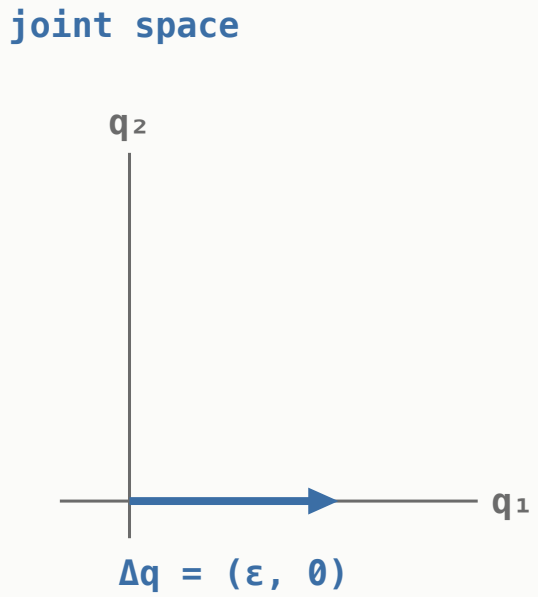
Which Way Does the Tip Move?

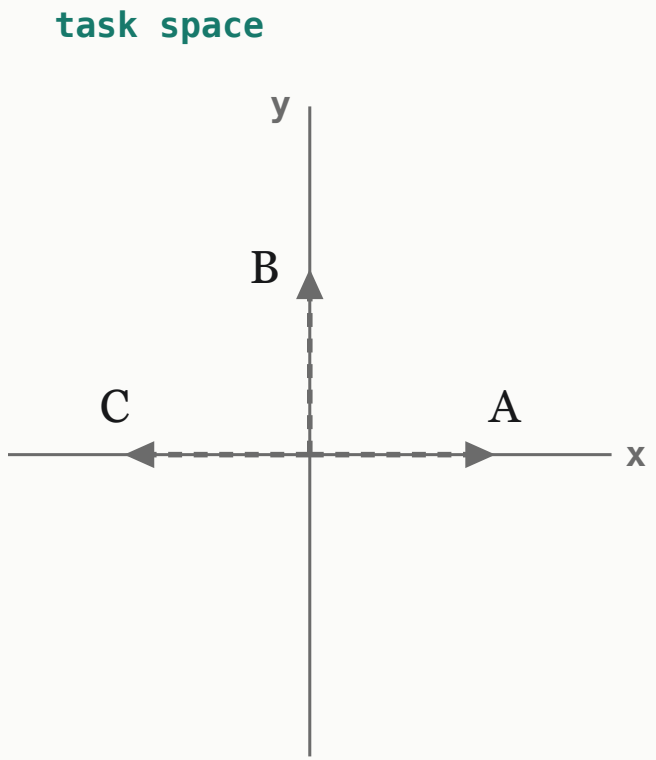
$$J = \begin{bmatrix} 0 & -1 \\ 2 & 0 \end{bmatrix}$$

J given; task axes x right, y up

Only joint 1 turns, by a small $\varepsilon > 0$: $\Delta q = (\varepsilon, 0)$. The tip mainly moves…

- A right
- B up
- C left





DERIVATION

Jacobians and Robot Motion

Worked Model: The Two-Link Arm

$$p = \begin{bmatrix} l_1 \cos q_1 + l_2 \cos q_{12} \\ l_1 \sin q_1 + l_2 \sin q_{12} \end{bmatrix}$$

q_{12} $q_1 + q_2$

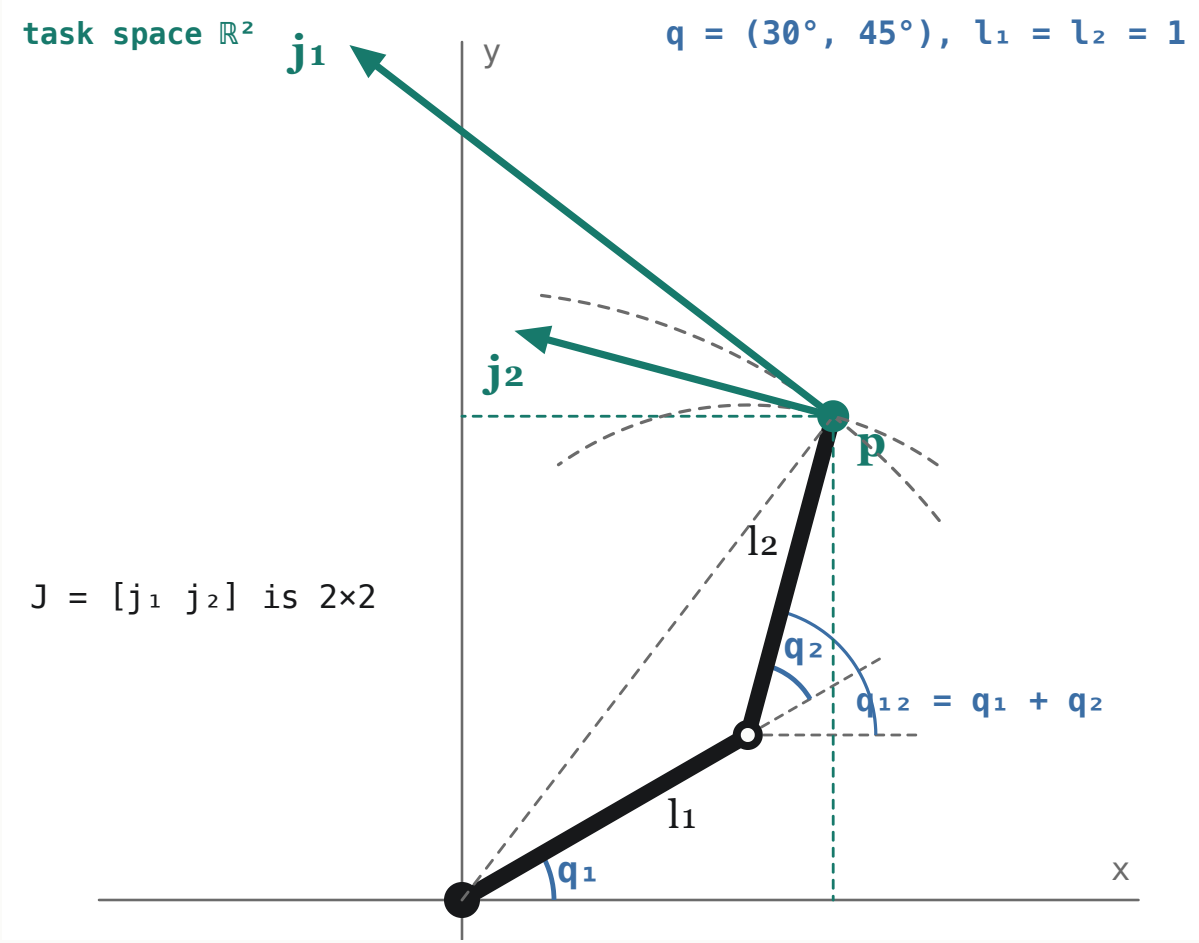
$l_1, l_2 > 0$ lengths; radians

$$j_1 = \begin{bmatrix} -l_1 \sin q_1 - l_2 \sin q_{12} \\ l_1 \cos q_1 + l_2 \cos q_{12} \end{bmatrix}$$

j_1 $\partial p / \partial q_1$

$$j_2 = \begin{bmatrix} -l_2 \sin q_{12} \\ l_2 \cos q_{12} \end{bmatrix}$$

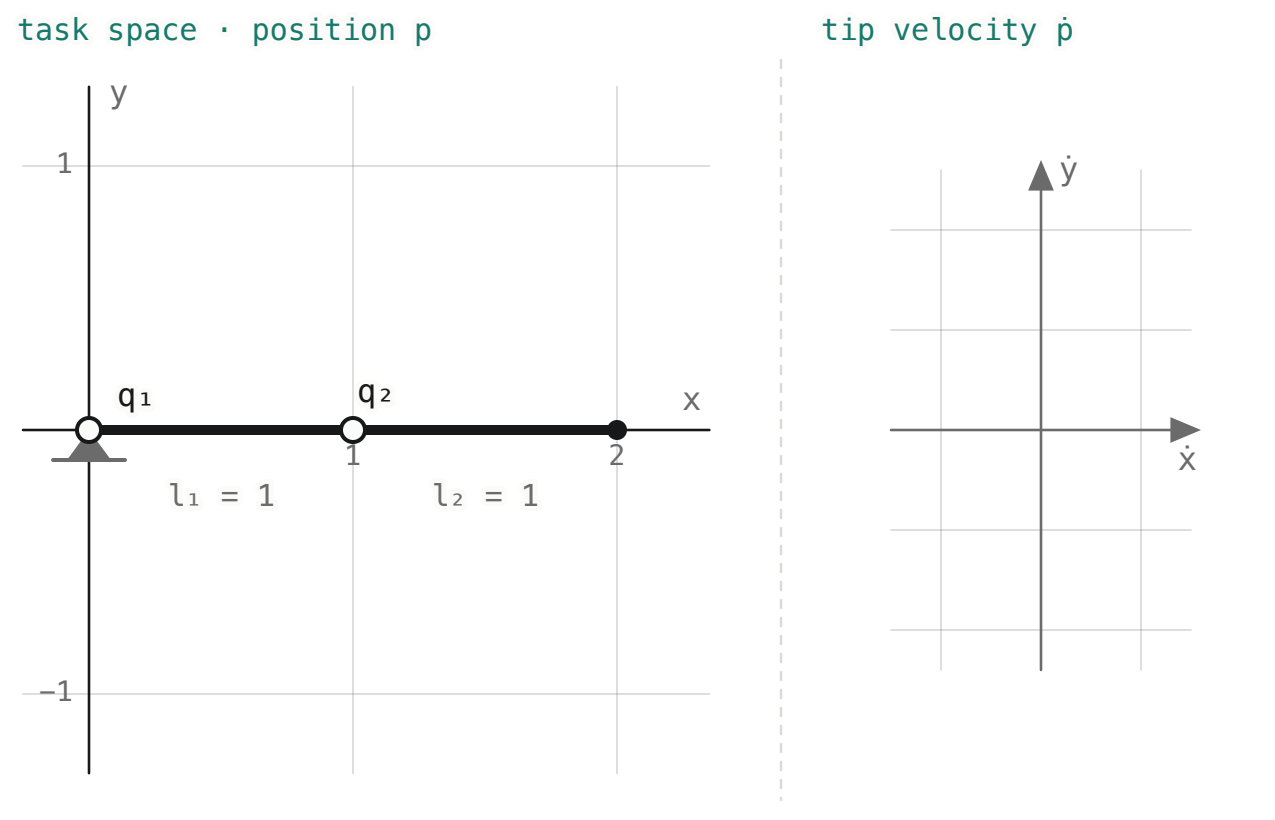
j_2 $\partial p / \partial q_2$



VOTE

Jacobians and Robot Motion

Which Tip Velocities Are Available?



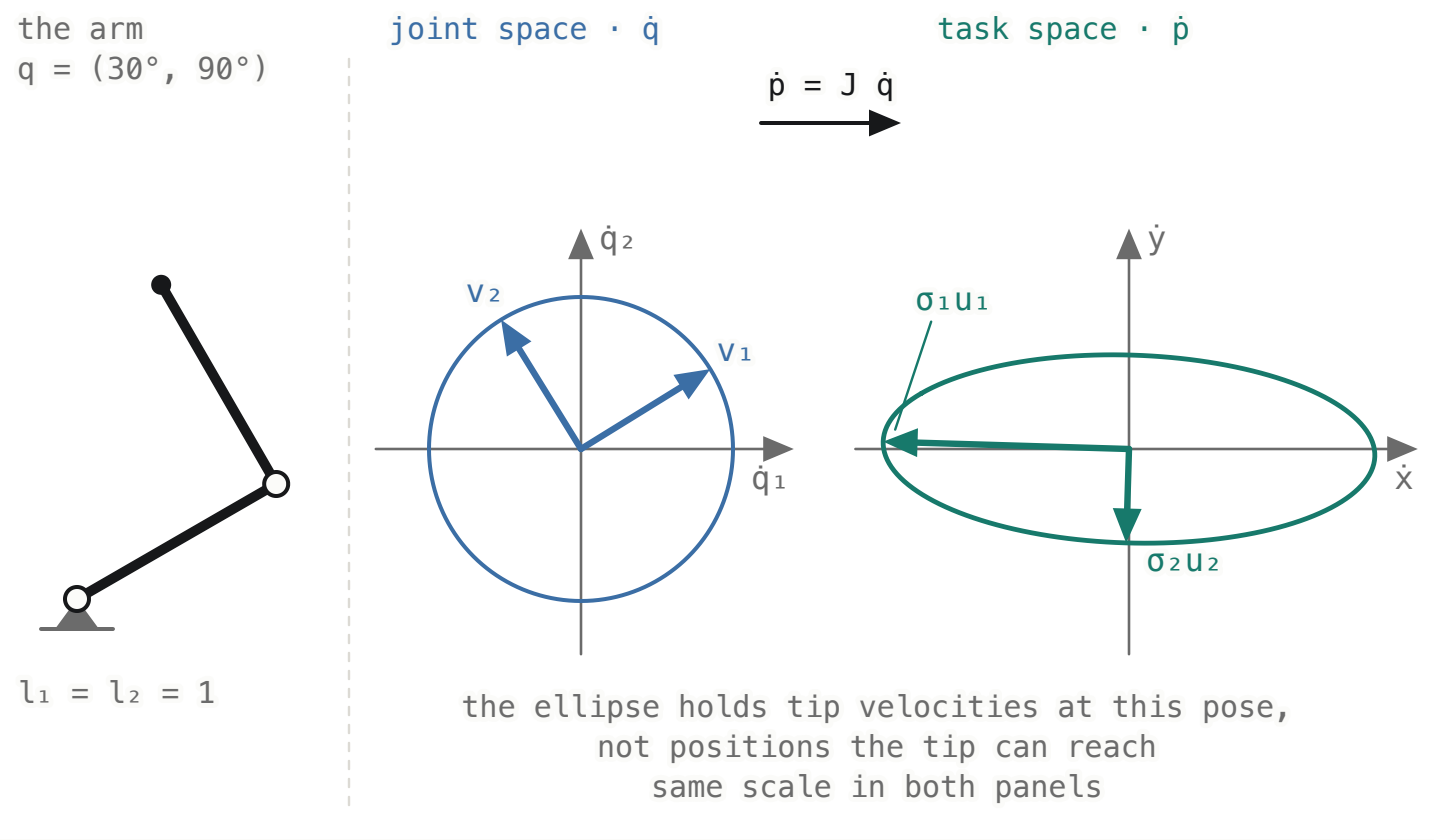
Straight arm, $q = (0, 0)$: $J = \begin{bmatrix} 0 & 0 \\ 2 & 1 \end{bmatrix}$. Which nonzero tip velocities $J\dot{q}$ exist?

- (A) only horizontal, along the arm
- (B) only vertical, across the arm
- (C) any direction in the plane

PICTURE

Jacobians and Robot Motion

The SVD of the Jacobian



$$J = U\Sigma V^{\mathsf{T}}, \qquad Jv_i = \sigma_i u_i$$

- v_i unit joint velocity: both joints move together
- u_i unit tip direction in task space
- σ_i local gain at this q: 1.62 and 0.62

$$\dot{p} = J(q)\,\dot{q}$$

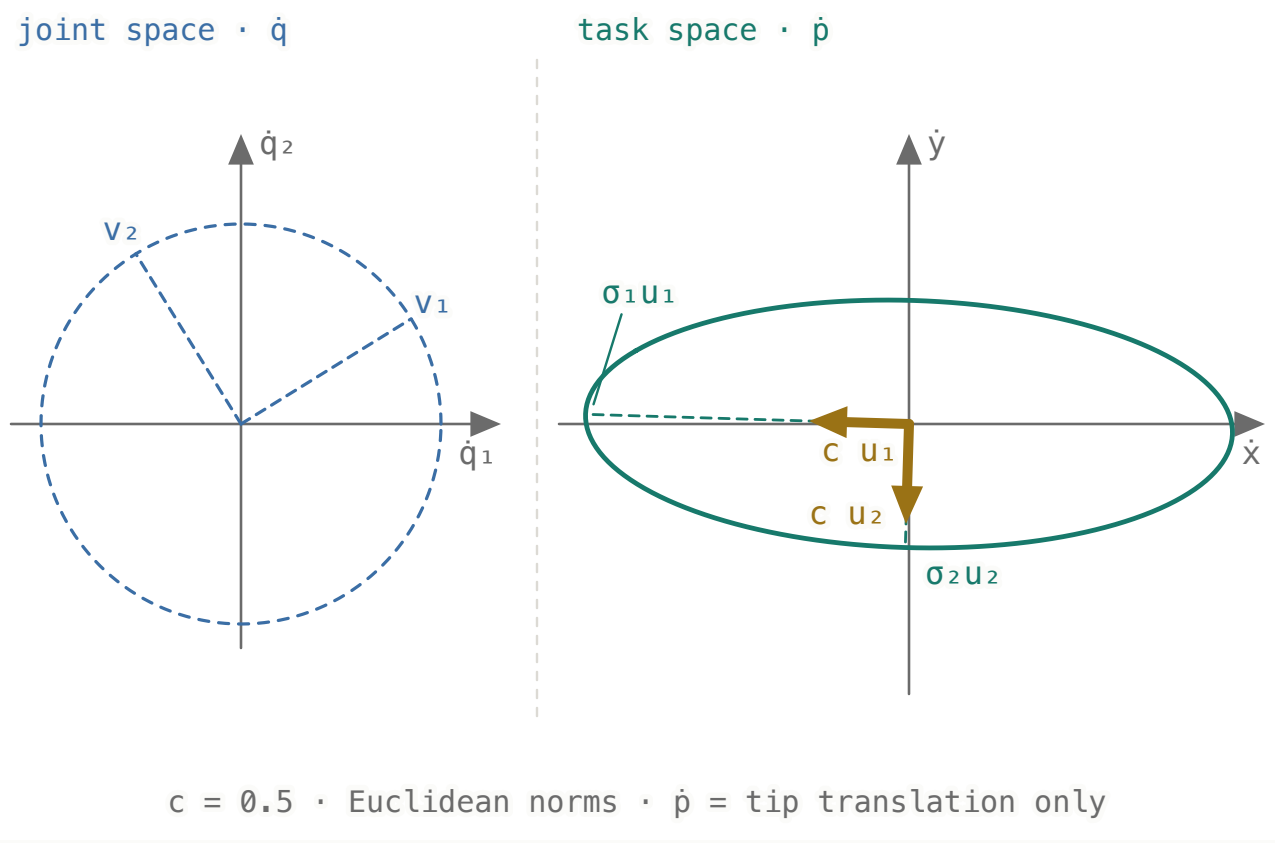
- $\dot{q}$, $\dot{p}$ joint and tip velocities; exact by the chain rule
- $\Delta p \approx J \Delta q$ finite steps: first order only

VOTE

$$J(30°,\ 90°) \approx \begin{bmatrix} -1.37 & -0.87 \\ 0.37 & -0.50 \end{bmatrix}$$

Jacobians and Robot Motion

Which Direction Needs Less Joint Speed?



The tip must move at the same speed c, along u_1 or u_2. Which needs the smaller joint speed $\|\dot{q}\|$?

A. along u_1, the long axis

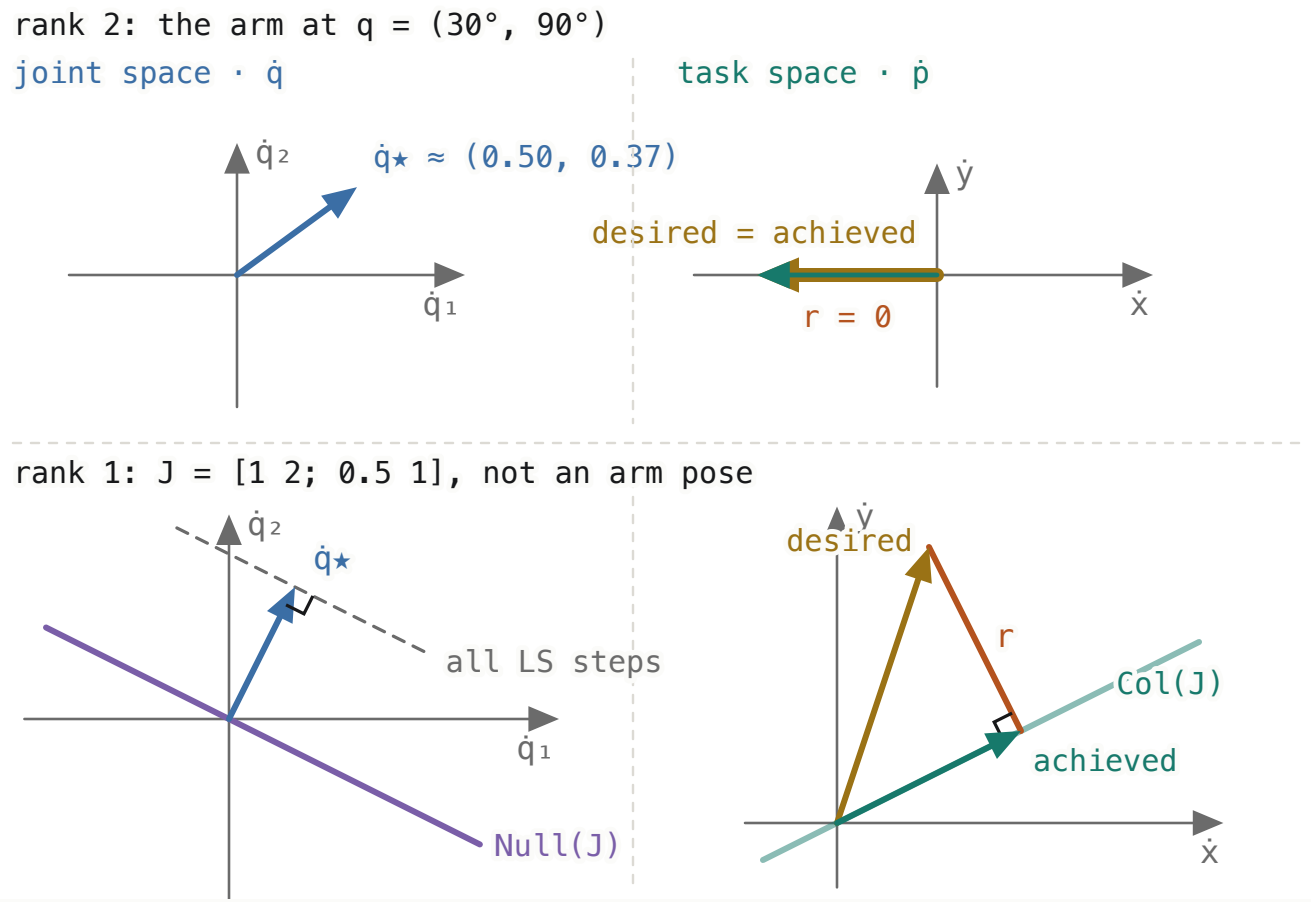
B. along u_2, the short axis

C. the same for both

J(30°,90°) ≈ $\begin{bmatrix} -1.37 & -0.87 \\ 0.37 & -0.50 \end{bmatrix}$

Jacobians and Robot Motion

Inverse Local Step



$$\dot{q}^\star = J^+ \dot{p}_d$$

$\dot{p}_d$ target

J^+ best fit, min $\|\dot{q}\|$

r $\dot{p}_d - J\dot{q}^\star$

EXAMPLE

Full rank: $(-1, 0) \mapsto (0.50, 0.37)$ approx..

Rank 1: $\dot{p}_d = (.5, 1.5)$, $a = (1, .5)$.

$s = \dot{q}_1 + 2\dot{q}_2$; $s^\star = (a^\mathsf{T}\dot{p}_d)/(a^\mathsf{T}a) = 1$.

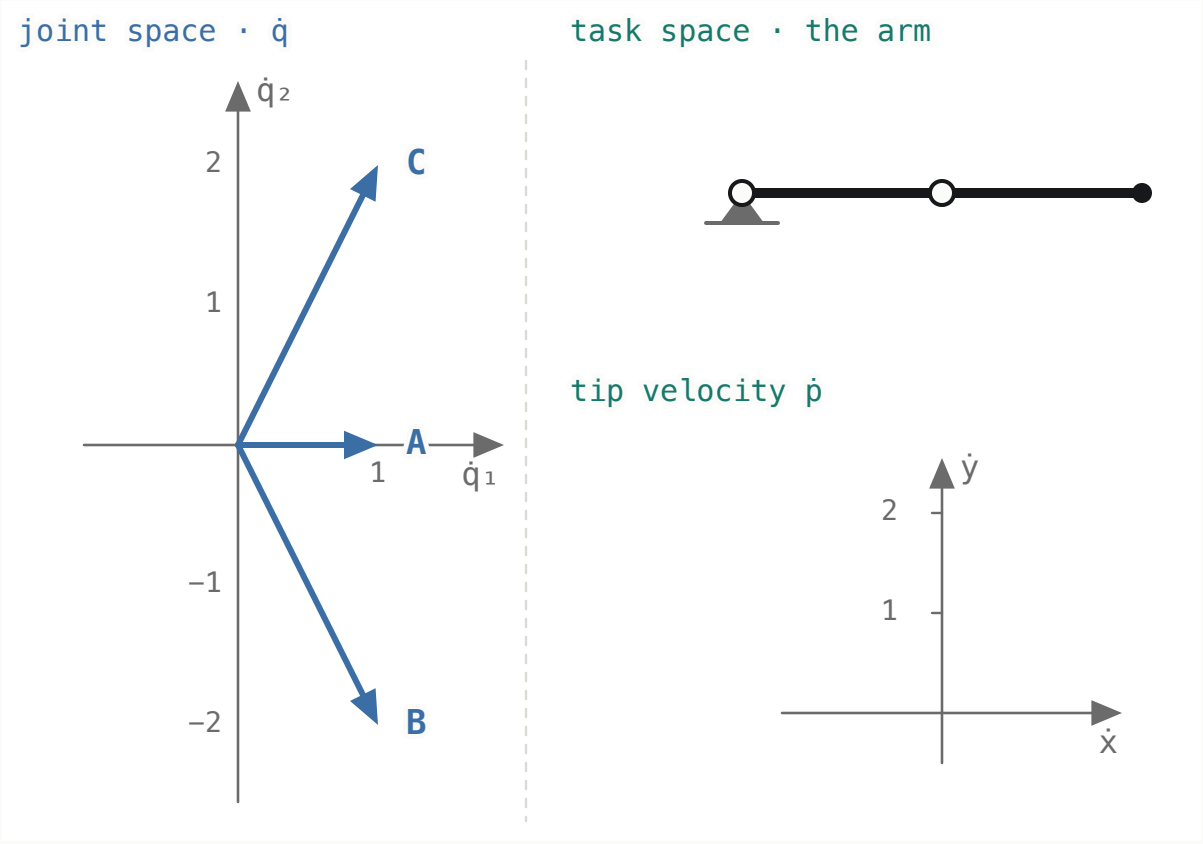
Min norm: $\dot{q}^\star = (1, 2)/5$.

$J\dot{q}^\star = (1, 0.5)$; $r = (-0.5, 1)$.

VOTE

Jacobians and Robot Motion

Which Joint Velocity Leaves the Tip Still?



The straight arm again. Which joint velocity gives $J\dot{q} = 0$?

A $\dot{q} = (1, 0)$

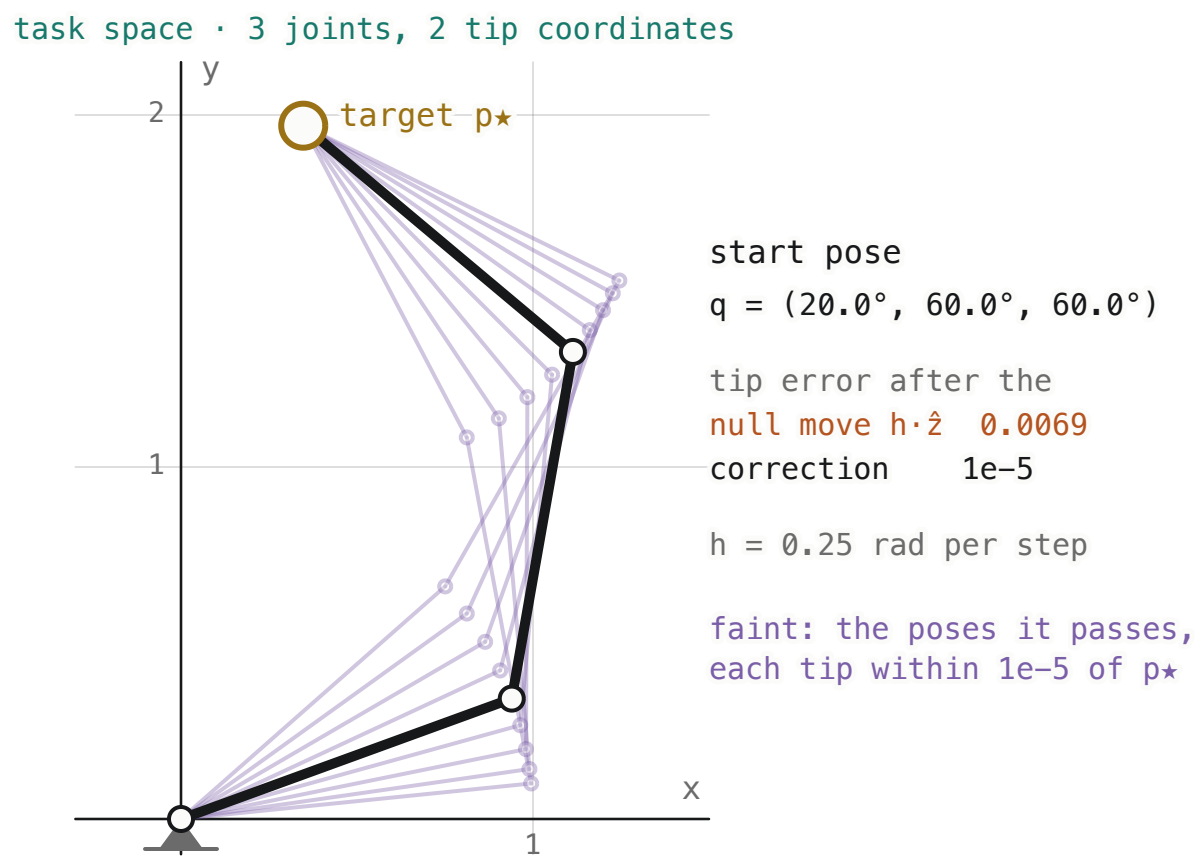
B $\dot{q} = (1, -2)$

C $\dot{q} = (1, 2)$

PICTURE

Jacobians and Robot Motion

Redundancy and Null Motion



$$\dot{q} = J^{+}\dot{p}_{\mathrm{d}} + z, \qquad Jz = 0$$

- J 2×3: two tip coordinates, three joints, rank 2
- z null motion, nullity $n - \operatorname{rank} J = 1$: z contributes exactly zero tip velocity

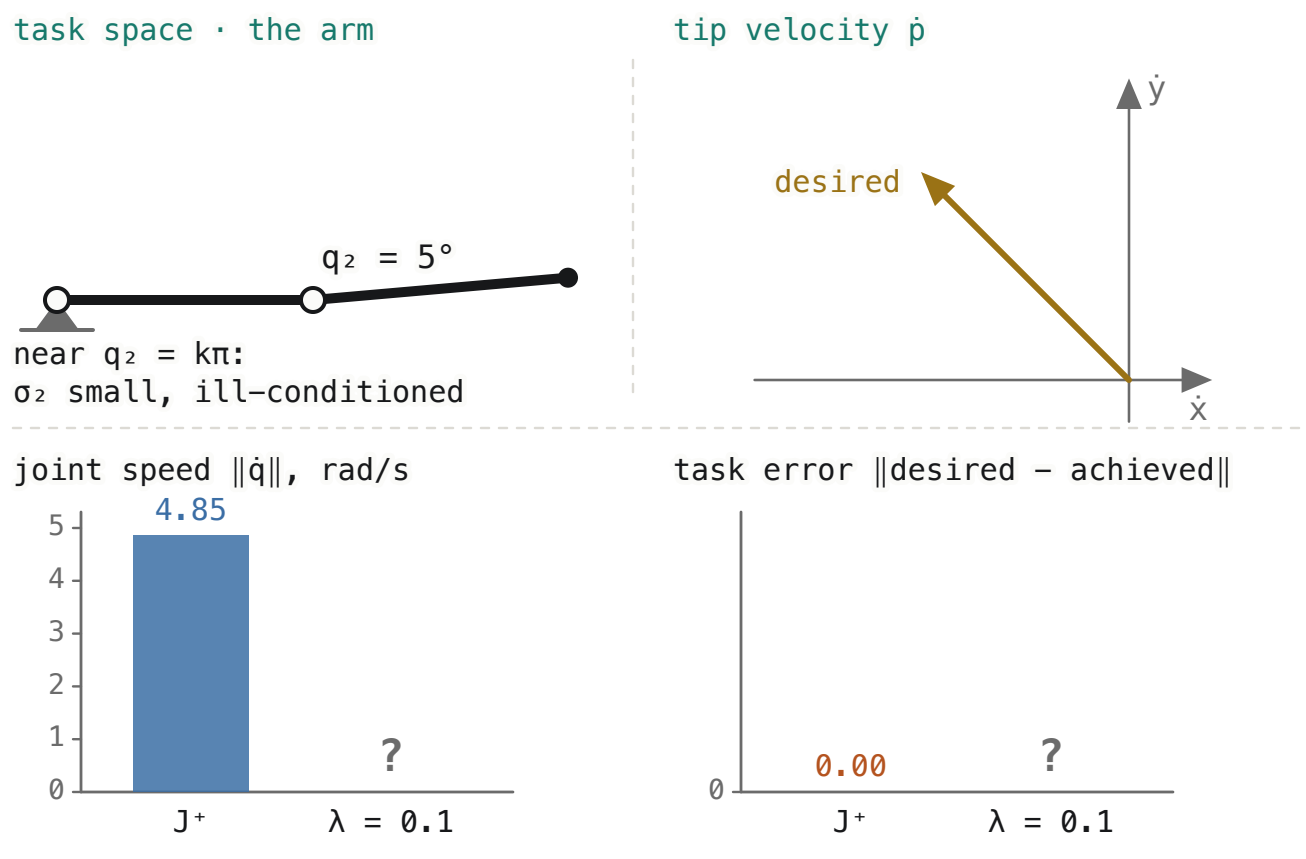
EACH FINITE STEP: FIRST ORDER ONLY

$q \leftarrow q + h\,\hat{z}(q)$, with $\hat{z}$ a unit vector in Null $J(q)$
then correct: $q \leftarrow q + J^{+}\big(p^{\star} - p(q)\big)$

VOTE

Jacobians and Robot Motion

What Does Damping Do Near a Singularity?



With $\lambda = 0.1$ the damped step minimises $\|J\dot{q} - \dot{p}_{\mathrm{d}}\|^2 + \lambda^2\|\dot{q}\|^2$. What does damping do?

A. hits every desired $\dot{p}_{\mathrm{d}}$ exactly

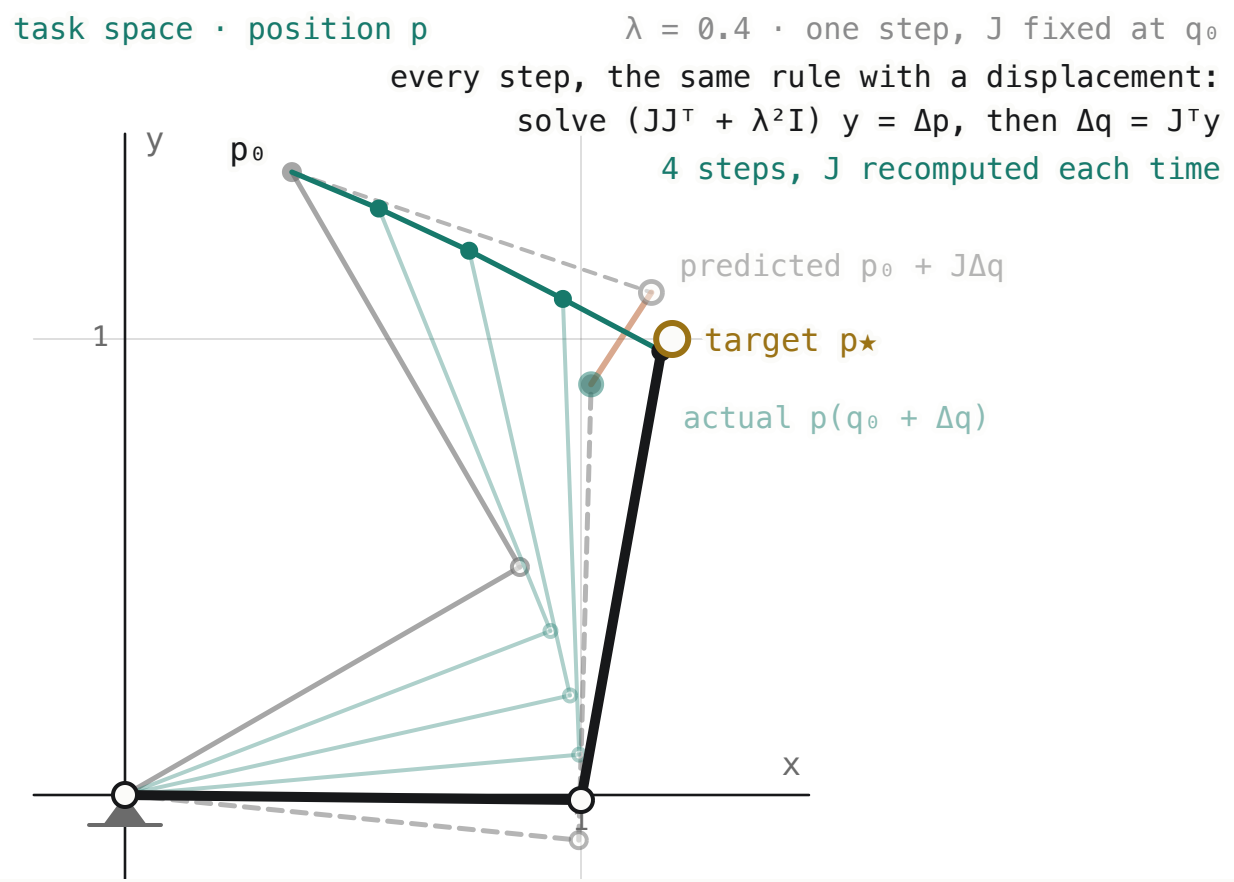
B. accepts task error to reduce joint speed

C. changes the arm's link lengths

IN PRACTICE

Jacobians and Robot Motion

Damped Steps



$$\dot{q}_\lambda = J^\mathsf{T}(JJ^\mathsf{T} + \lambda^2 I)^{-1}\dot{p}_\mathrm{d}, \quad \lambda > 0$$

J	$m \times n$ at this q
$JJ^\mathsf{T} + \lambda^2 I$	$m \times m$, positive definite
why	$(J^\mathsf{T}J + \lambda^2 I)J^\mathsf{T} = J^\mathsf{T}(JJ^\mathsf{T} + \lambda^2 I)$

STEPS	LARGEST MODEL ERROR	FINAL TARGET ERROR
1	0.24	0.20
4	0.024	0.036
8*	0.0064	0.018

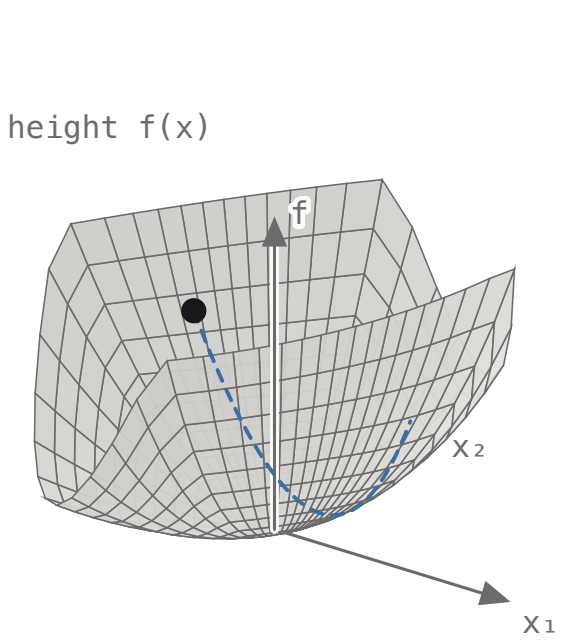
*computed. Smaller steps reduce model error faster than target error; limits separate.

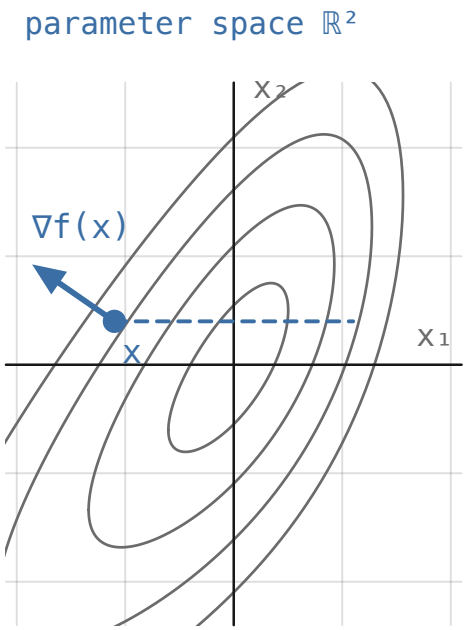
DEFINITION

Hessians and Local Curvature

From Vector Outputs to a Scalar Objective

$$f : \mathbb{R}^n \to \mathbb{R}, \qquad \nabla f(x) \in \mathbb{R}^n, \qquad H(x) = \nabla^2 f(x) \in \mathbb{R}^{n \times n}$$





- $f(x)$ one number, the objective at parameters x
- $\nabla f(x)$ $n \times 1$, the local slope, pointing uphill
- $H(x)$ $n \times n$, how ∇f changes as x moves; same letter as the quadratic forms, on purpose
- $J(q)$ vector output $p(q) \in \mathbb{R}^m$: $m \times n$

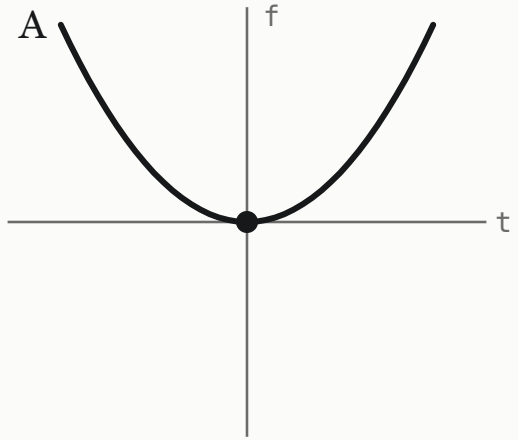
VOTE

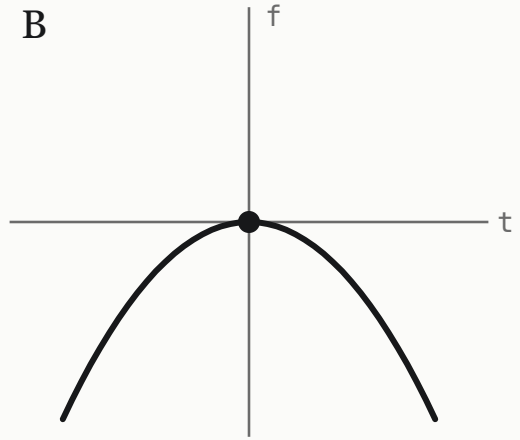
Hessians and Local Curvature

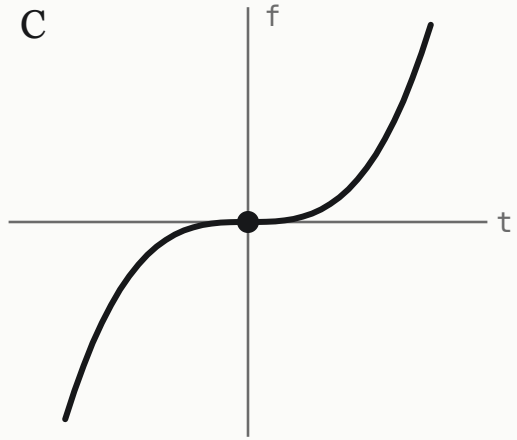
Which Curve Has a Local Minimum at 0?

All three have $f'(0) = 0$. Which one has a local minimum at $t = 0$?

`parameter t ∈ ℝ, height f(t)`







(A) $f(t) = t^2$

(B) $f(t) = -t^2$

(C) $f(t) = t^3$

DEFINITION

Hessians and Local Curvature

Hessian Entries

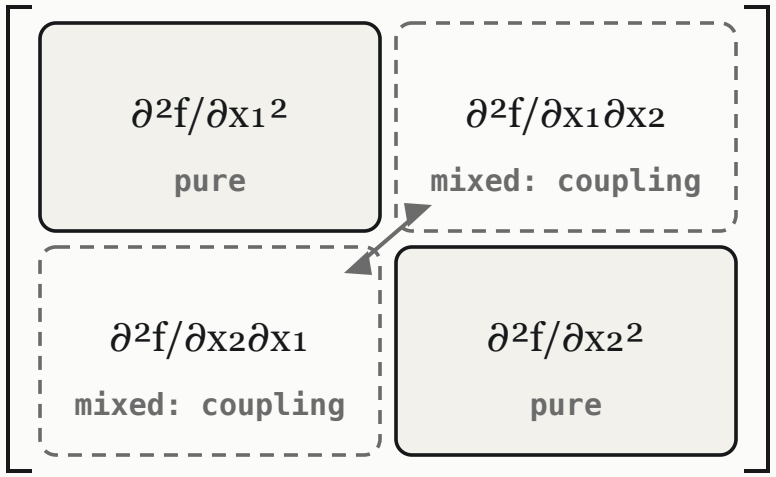
$$H_{ij}(x) = \partial_{x_j}(\partial_{x_i} f)(x), \qquad i, j = 1, \ldots, n$$

H_{ij} row i: slope; column j: input

$f \in C^2$ continuous second partials near x

then $H_{ij} = H_{ji}$: $H = H^{\mathsf{T}}$, symmetric $n \times n$

H(x) =



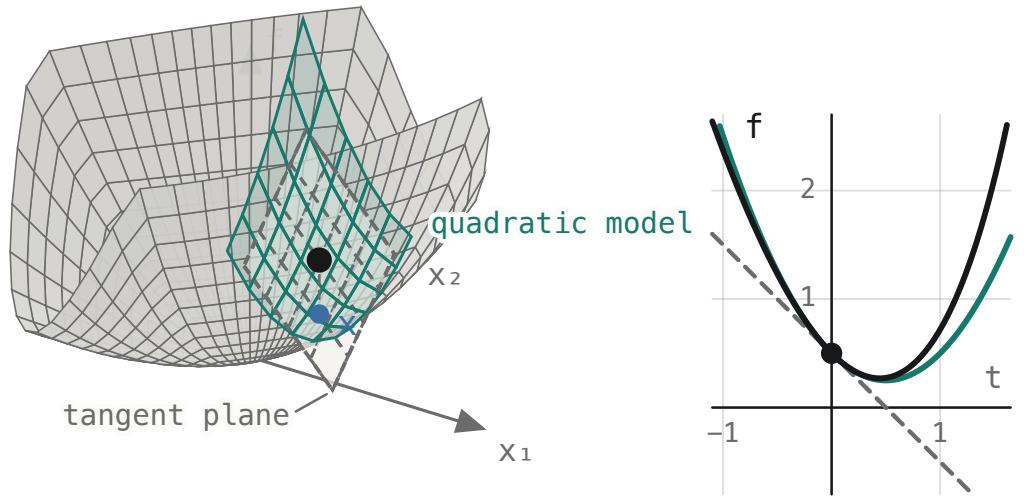
equal if f ∈ C²

MODEL

Hessians and Local Curvature

The Second-Order Approximation

$$f(x+\Delta x) \approx f(x) + \nabla f(x)^{\mathsf{T}} \Delta x + \frac{1}{2} \Delta x^{\mathsf{T}} H(x) \Delta x$$



- x fixed expansion point; $f \in C^2$
- Δx small step
- ∇f $n \times 1$, tilt: tangent plane
- H $n \times n$, bend: quadratic model

WORKED EXAMPLE

$f = e^{x_1} - 1 - x_1 + \frac{1}{2}(x_2 - x_1)^2$, $x = (0, 1)$:

$f(x) = \frac{1}{2}$, $\nabla f = (-1, 1)$, $H = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$

VOTE

Hessians and Local Curvature

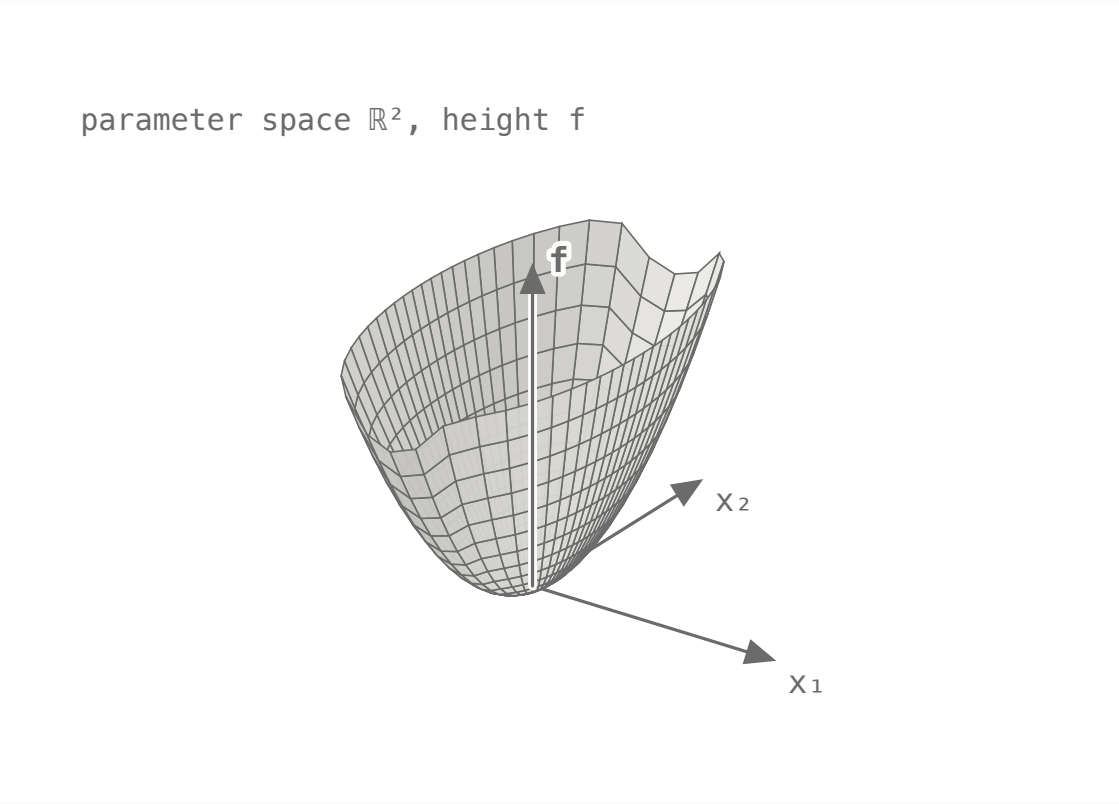
What Is the Hessian of This Bowl?

$f(x_1, x_2) = 2x_1^2 + \frac{1}{2}x_2^2$. Which matrix is its Hessian?

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

B $\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$

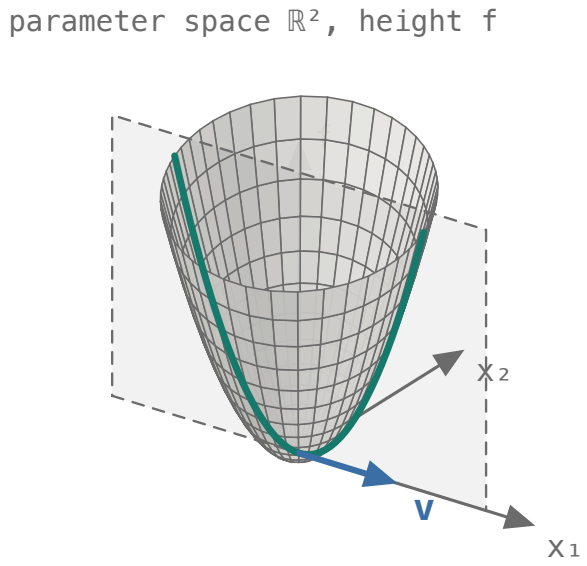
C $\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$

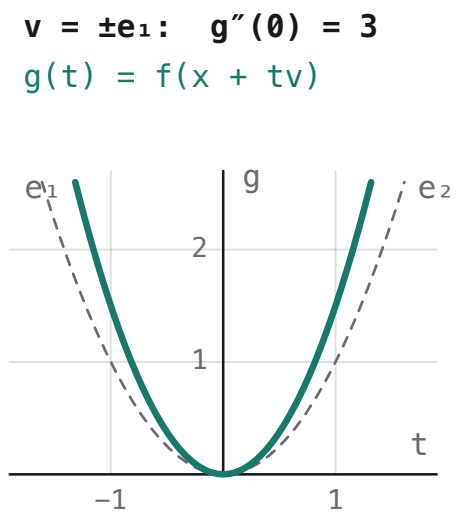


DEFINITION

Hessians and Local Curvature

Curvature Along a Direction





$$g(t) = f(x + tv), \qquad g''(0) = v^{\mathsf{T}} H(x)\, v$$

- v direction, unit: $\|v\| = 1$, so directions compare fairly
- g height of f along the line through x
- $g''(0)$ second derivative along that line, not the surface's geometric curvature
- why $g'(t) = \nabla f(x + tv)^{\mathsf{T}} v$; differentiate once more
- here $f = \frac{1}{2} x^{\mathsf{T}} H x$, $H = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$, $x = 0$

VOTE

$H = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$

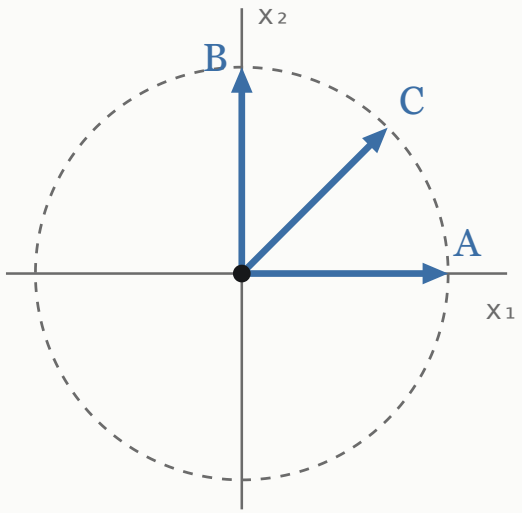
Hessians and Local Curvature

Which Direction Curves Up Most?

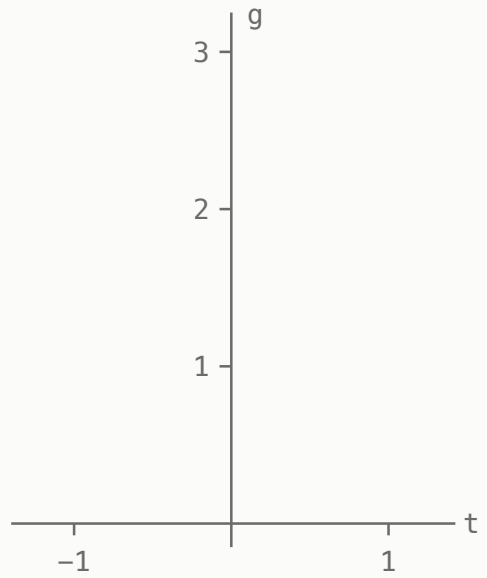
Along which unit v is $g''(0) = v^\top H v$ largest?

- **A** $v = e_1$
- **B** $v = e_2$
- **C** $v = \frac{1}{\sqrt{2}}(1, 1)$, at 45°

parameter space $\mathbb{R}^2$



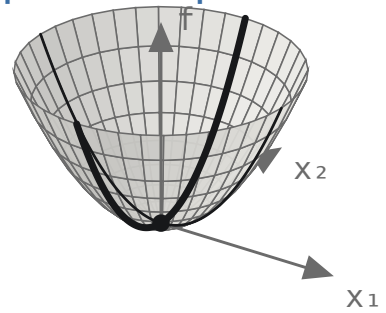
g(t) = f(x + tv)



Hessians and Local Curvature

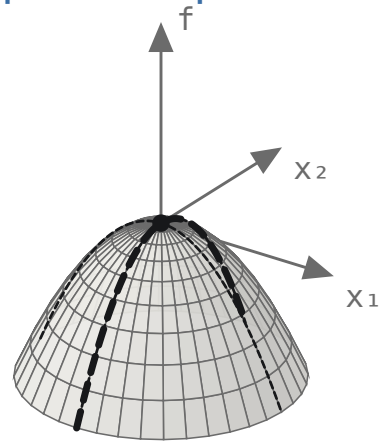
Classifying a Stationary Point

f = x₁² + x₂²
parameter space ℝ²



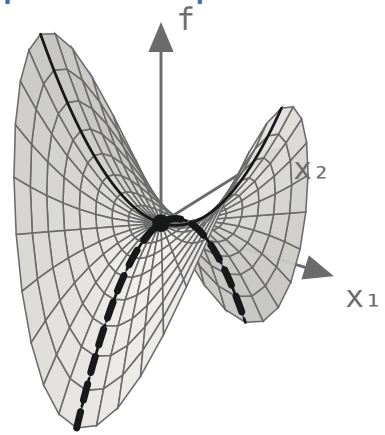
λ₁, λ₂ > 0 (PD): strict local min
vᵀHv = +2.00

f = −x₁² − x₂²
parameter space ℝ²



λ₁, λ₂ < 0 (ND): strict local max
vᵀHv = −2.00

f = x₁² − x₂²
parameter space ℝ²



λ₁ > 0 > λ₂ (indefinite): saddle
vᵀHv = −2.00

$$f(x^\star + \Delta x) \approx f(x^\star) + \tfrac{1}{2}\Delta x^\mathsf{T} H(x^\star)\Delta x$$

Local tests only. $f \in C^2$ on a convex open set: convex $\iff$ $H(x) \succeq 0$ at every x.

- $x^\star, \Delta x$ stationary point ($\nabla f = 0$), small step
- f C^2 near $x^\star$: H symmetric $n \times n$
- λ_i eigenvalues of $H(x^\star)$: their signs decide

VOTE

Hessians and Local Curvature

Does a PSD Hessian Decide?

$$f_{\pm}(x, y) = x^2 \pm y^4$$

(x, y) any point; both are C^2

at $(0, 0)$ $\nabla f_{\pm}(0, 0) = 0$, $H_{\pm}(0, 0) = \mathrm{diag}(2, 0)$

Can this Hessian alone classify both points?

A. Yes: both are minima

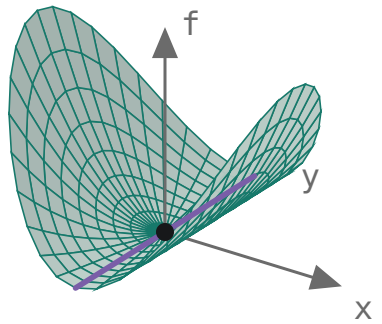
B. Yes: both are saddles

C. No: higher-order terms matter

parameter space $\mathbb{R}^2$

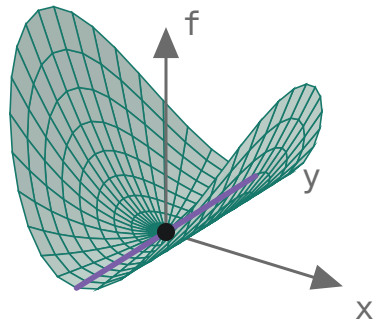
$f_+ = x^2 + y^4$

model at 0: x^2



$f_- = x^2 - y^4$

model at 0: x^2



DERIVATION

Hessians and Local Curvature

The Least-Squares Hessian

$$f(x) = \tfrac{1}{2}\|Ax - b\|^2, \qquad \nabla f(x) = A^\mathsf{T}(Ax - b), \qquad H = A^\mathsf{T}A$$

H Hessian

$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = A$, $b = (1, 0)$: best output, shortest input, curvature?

Project onto $Ax = (1, 2)s$.
$s = 1/5$, $\hat{b} = (1, 2)/5$, $r = (4, -2)/5$

$x_1 + 2x_2 = 1/5$; shortest $x = (1, 2)/25$

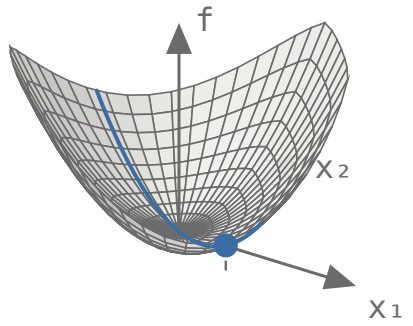
$\lambda(H) = 25, 0$
Flat direction $(2, -1)$; $\min f = 2/5$.

surfaces: b = 0

$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$ full column rank

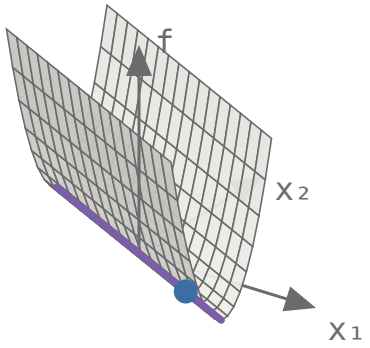
$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ rank 1

f = 0.50 height × 0.5



λ = 3, 1
one lowest point

f = 0.00 height × 0.5



λ = 25, 0
a line of lowest points

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

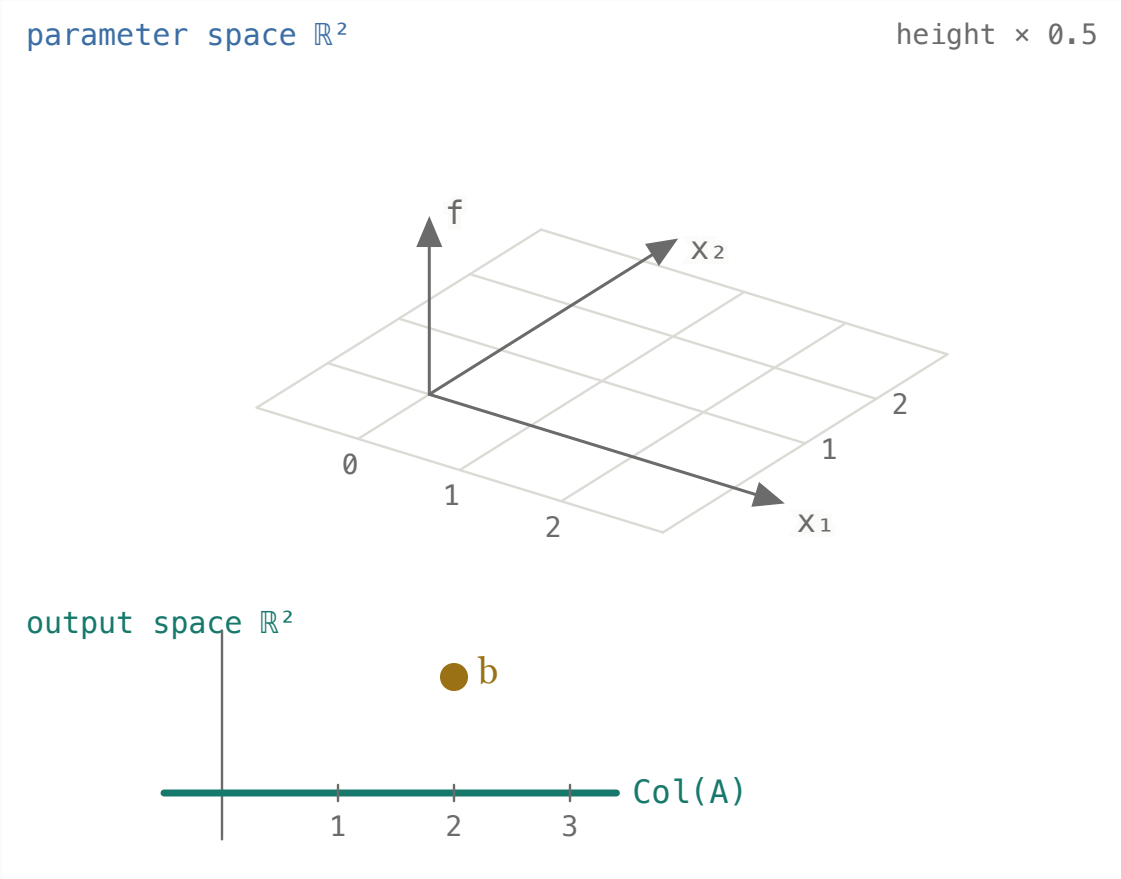
VOTE

Hessians and Local Curvature

What Shape Is This Least-Squares Objective?

$f = \frac{1}{2}\|Ax - b\|^2$. Choose a method; find $A\hat{x}$, r, A^+b, a flat direction and the graph shape.

- (A) A bowl with one lowest point
- (B) A trough, bottom: $x_1 + x_2 = 2$
- (C) A saddle



CONNECTION

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

Hessians and Local Curvature

Closing the Loop: Hessian, Gram Matrix, SVD

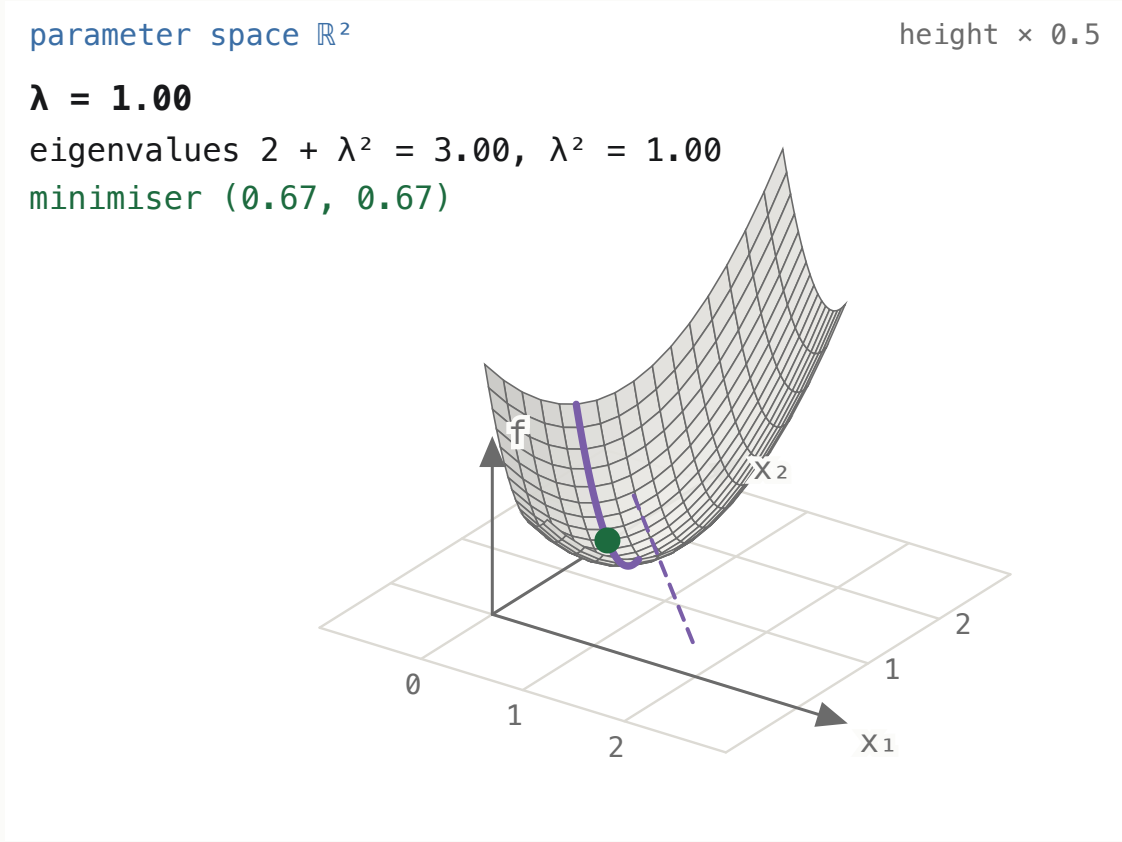
$$H = A^\mathsf{T} A = V \Sigma^\mathsf{T} \Sigma V^\mathsf{T}, \quad H_\lambda = A^\mathsf{T} A + \lambda^2 I \succ 0$$

$V, \Sigma^\mathsf{T}\Sigma$ SVD of A; diagonal σ_i^2, then zeros

H_λ Hessian of $\frac{1}{2}\|Ax - b\|^2 + \frac{\lambda^2}{2}\|x\|^2$, $\lambda > 0$

Damped LS as earlier, no halves: Hessian $2H_\lambda$, same minimiser.

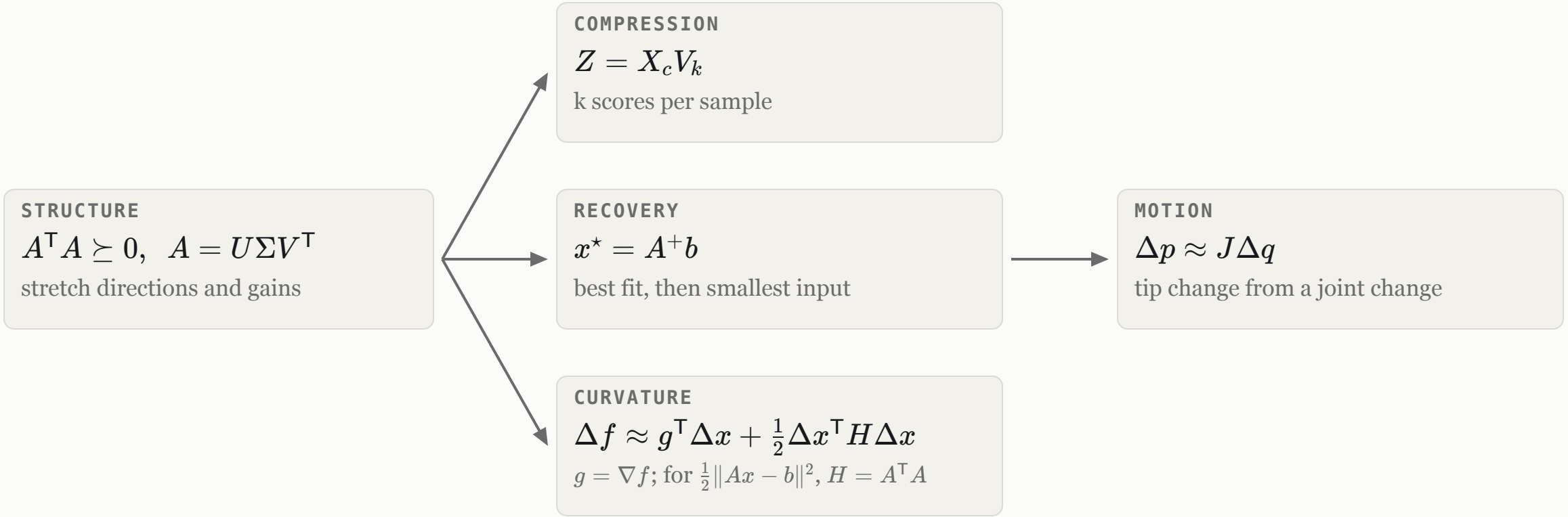
Next lecture: systematic derivatives.



CONCEPT MAP

Connecting the Concepts

One Matrix, Five Jobs



VOTE

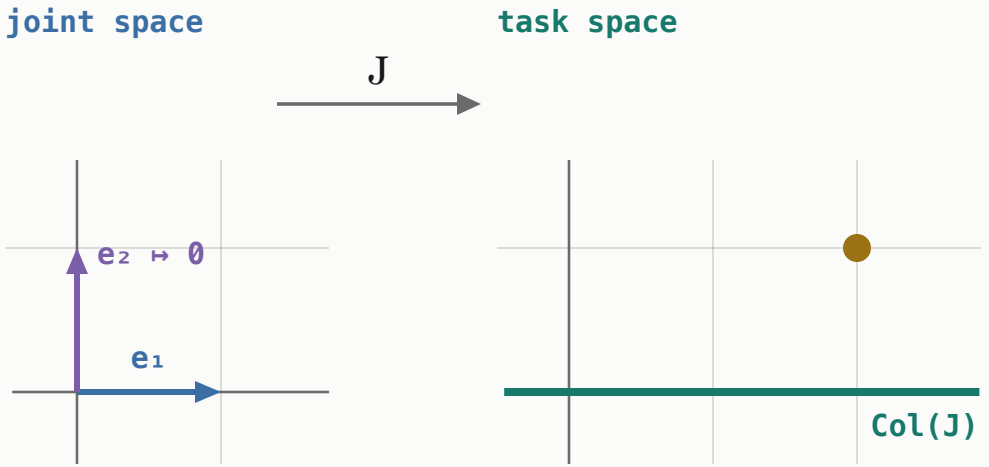
Connecting the Concepts

Exit Question: Rank, Step and Curvature

Linear model: $J = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, desired $\Delta p = (2, 1)$.

Which set is right?

- **A** rank 2, minimum-norm step $(1, 1)$, $J^\top J$ PD
- **B** rank 1, minimum-norm LS step $(1, 0)$, $J^\top J$ PSD but not PD
- **C** rank 1, minimum-norm step $(2, 0)$, target exactly reachable



SELECTED REFERENCES

References

1. Strang, G. *Linear Algebra and Learning from Data*, I.7–I.9, II.2. Wellesley, 2019.
2. Trefethen & Bau. *Numerical Linear Algebra*, Lec. 4–5, 11–12. SIAM, 1997.
3. Goodfellow et al. *Deep Learning*, 2.7–2.9, 2.12, 4.2–4.3. MIT Press, 2016.
4. Lynch & Park. *Modern Robotics*, 5.1, 5.3–5.4, 6.2–6.3. Cambridge, 2017.
5. Buss, S. R. *Introduction to Inverse Kinematics*. UCSD, 2004.

Course standards: *Notation and Conventions*, symbols, shapes and the two sign conventions; *Textbooks and Reading Guide*, what each source is for and where its notation differs.