Linear Transformations and Reconstruction: Choosing a Decoder Basis

ECE 57000 — September 11, 2026

David I. Inouye

Wednesday made the SVD a way to read a linear transformation

- Orthogonal factors preserve lengths; singular values set directional scalings.
- The largest singular value bounds how much any input can stretch.
- Rank counts the nonzero singular values: how many independent directions survive.

Next question: When can we recover an input from its transformed output?

Rank counts the independent directions that survive

$$\operatorname{rank}(A) = \#\{j : \sigma_j > 0\} \le \min(m, d)$$

For example:

$$\Sigma = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Longrightarrow \qquad \operatorname{rank}(A) = 2$$

- Two independent input directions produce nonzero output directions.
- The third input direction disappears: $A\boldsymbol{v}_3 = \mathbf{0}$.
- Rotations and reflections change which directions these are, but not the rank.

A null direction makes different inputs indistinguishable

The null space contains all directions that map to zero:

$$\mathrm{null}(A) = \{\boldsymbol{h} : A\boldsymbol{h} = \boldsymbol{0}\}.$$

For any nonzero $\boldsymbol{h} \in \mathrm{null}(A)$:

$$\begin{aligned} A(\boldsymbol{x}+\boldsymbol{h}) &= A\boldsymbol{x} + A\boldsymbol{h} && \text{(Distributivity)} \\ &= A\boldsymbol{x} && \text{(Null direction)} \end{aligned}$$

A square operator is invertible when no direction collapses

In 1D, $y = ax$ can be undone by $x = y/a$ exactly when $a \neq 0$.

For square $A = U\Sigma V^T \in \mathbb{R}^{d\times d}$:

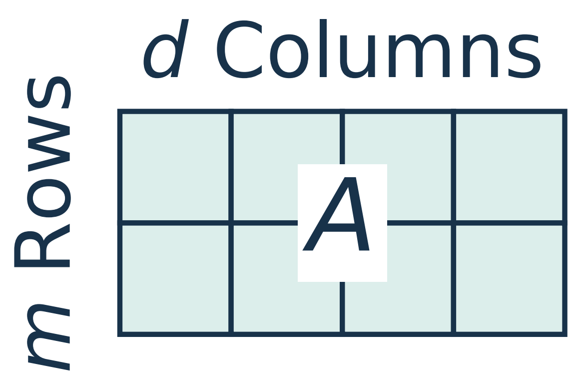
$$\text{all } \sigma_j > 0 \quad \Longrightarrow \quad A^{-1} = V\Sigma^{-1}U^T, \qquad \Sigma^{-1} = \operatorname{diag}(1/\sigma_1, \dots, 1/\sigma_d).$$

$$\begin{aligned}
A^{-1}A &= (V\Sigma^{-1}U^T)(U\Sigma V^T) && \text{(Substitute)}\\
&= V\Sigma^{-1}\Sigma V^T && (U^TU = I_d)\\
&= VV^T && (\Sigma^{-1}\Sigma = I_d)\\
&= I_d && (VV^T = I_d)
\end{aligned}$$

Recovery depends on preserving every input direction

$y = Ax$ maps d input coordinates to m output coordinates.

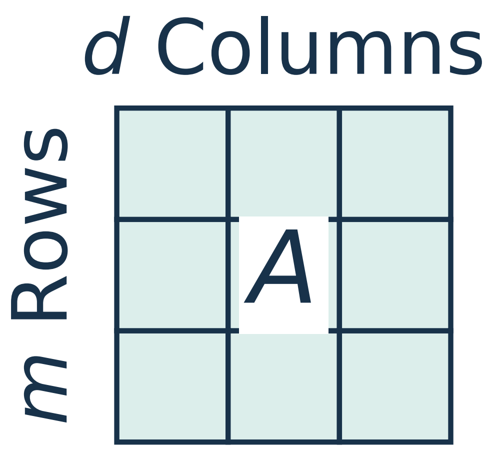
$m < d$



Recovery Impossible

Some Input Direction Is Lost

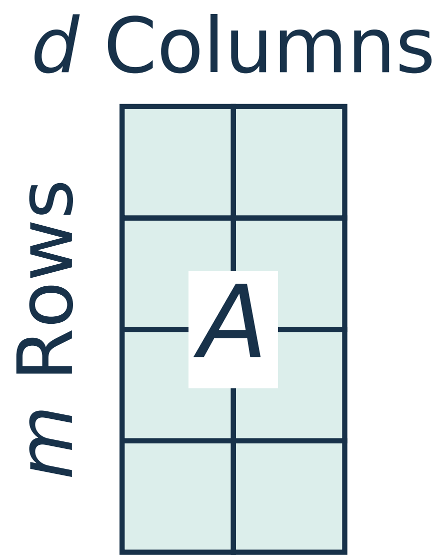
$m = d$



Recovery If $\mathrm{rank}(A) = d$

Every Input Direction Must Survive

$m > d$



Recovery If $\mathrm{rank}(A) = d$

Extra Outputs Can Be Redundant

Recovering **every** $x \in \mathbb{R}^d$ requires $\mathrm{rank}(A) = d$.

The pseudoinverse undoes every nonzero scaling

For **any** $A = U\Sigma V^T \in \mathbb{R}^{m\times d}$, define:

$$\underbrace{A^\dagger}_{d\times m} = \underbrace{V}_{d\times d}\,\underbrace{\Sigma^\dagger}_{d\times m}\,\underbrace{U^T}_{m\times m}.$$

$\Sigma^\dagger$ has the **transposed shape** of Σ, with zero off-diagonal entries:

$$(\Sigma^\dagger)_{jj} = \begin{cases} 1/\sigma_j, & \sigma_j > 0, \\ 0, & \sigma_j = 0. \end{cases}$$

- Reverse the output reorientation, undo nonzero scalings, then reorient back.
- A zero scaling destroyed information; leaving it zero cannot restore it.
- The **Moore–Penrose pseudoinverse** exists even when A has no inverse.

The pseudoinverse recovers inputs exactly when no input direction is lost

For $\boldsymbol{y} = A\boldsymbol{x}$, recovery of **every** input means:

$$\widehat{\boldsymbol{x}} = A^{\dagger}\boldsymbol{y} = \boldsymbol{x} \text{ for all } \boldsymbol{x} \in \mathbb{R}^d \quad \Longleftrightarrow \quad A^{\dagger}A = I_d.$$

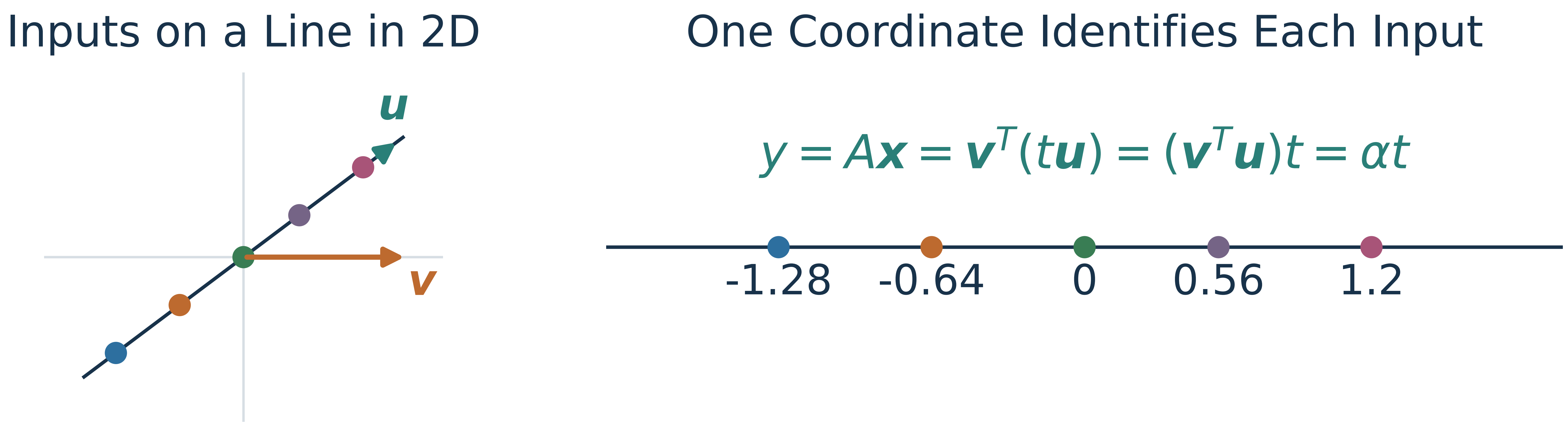
Let $r = \operatorname{rank}(A)$ and $V_r = [\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_r]$ contain its surviving input directions.

$$\begin{aligned} A^{\dagger}A &= V\Sigma^{\dagger}U^TU\Sigma V^T && \text{(Substitute)} \\ &= V(\Sigma^{\dagger}\Sigma)V^T && (U^TU = I_m) \\ &= V_rV_r^T && \text{(Keep nonzero directions)} \end{aligned}$$

$\Sigma^{\dagger}\Sigma$ has r diagonal ones and $d - r$ zeros. Thus $A^{\dagger}A = I_d$ **exactly when** $r = d$: full column rank, requiring $m \geq d$.

A restricted input subspace can be recovered exactly

Inputs: $\boldsymbol{x} = t\boldsymbol{u} \in \mathbb{R}^2$, $\|\boldsymbol{u}\|_2 = 1$. Encoder: $A = \boldsymbol{v}^T$, with $\alpha = \boldsymbol{v}^T\boldsymbol{u} \neq 0$.



Decode with $B = \boldsymbol{u}/\alpha$: knowing the line lets us undo its nonzero scaling.

$$\widehat{\boldsymbol{x}} = By = \frac{\boldsymbol{u}}{\alpha}(\alpha t) = t\boldsymbol{u} = \boldsymbol{x}.$$

If $\boldsymbol{v} = \boldsymbol{u}$, then $\alpha = 1$, $y = t$, and $B = A^\dagger = \boldsymbol{u}$. **In general, $B \neq A^\dagger$.**

Read what an operator does from its singular values

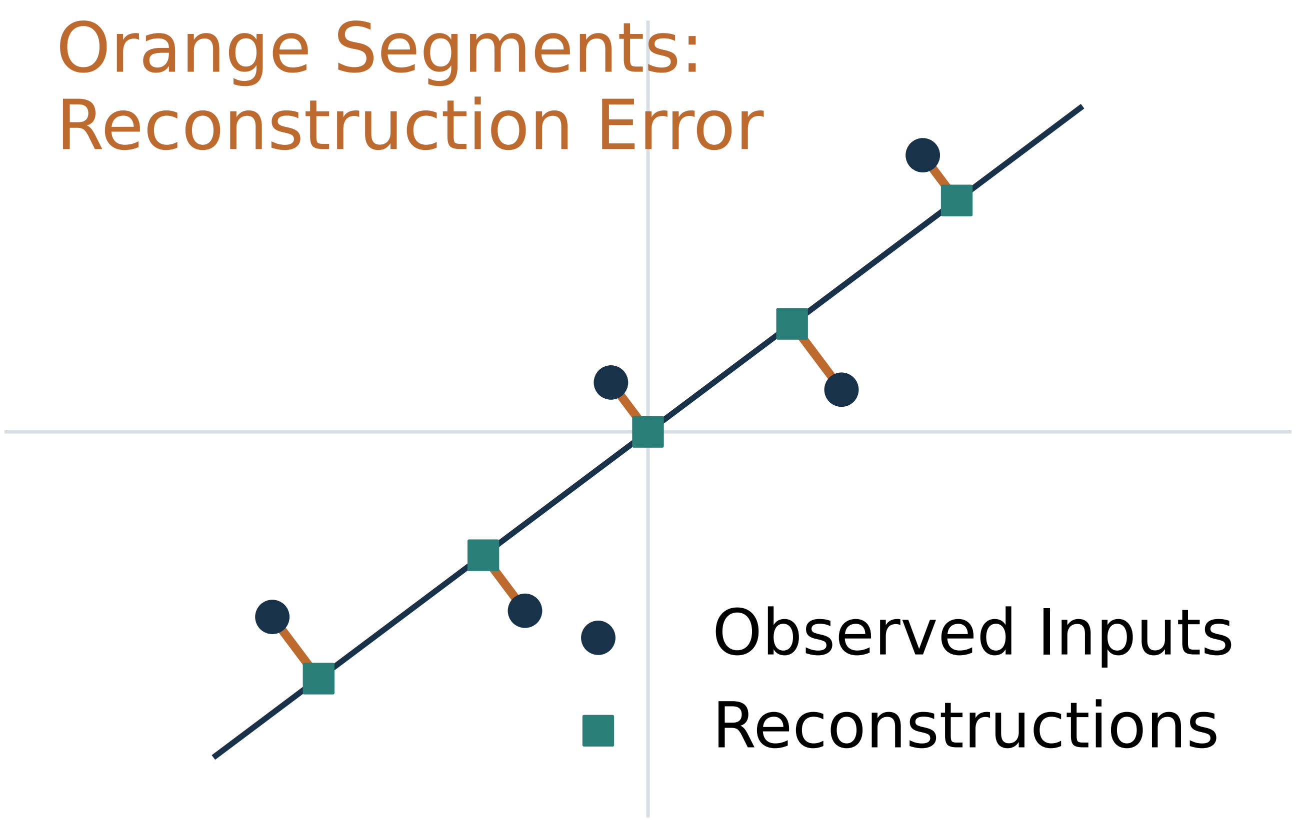
A sensor-processing system uses $A = U\,\mathrm{diag}(4, 1, 0)V^T$.

> **Can you recover every sensor input? What is the largest possible stretching factor? Could this operator preserve every pairwise distance? Explain which input direction is invisible to the system.**

Rank 2; recovery is impossible in general; maximum scaling is 4. The direction $\boldsymbol{v}_3$ disappears. All distances cannot be preserved.

Data near a subspace can have small reconstruction error

What if the inputs are **close to** a line, but do not lie exactly on it?



$$y = \boldsymbol{u}^T \boldsymbol{x}, \qquad \widehat{\boldsymbol{x}} = \boldsymbol{u} y, \qquad \text{error} = \|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2.$$

Which direction minimizes total reconstruction error?

Reconstruction and Principal Component Analysis

David I. Inouye

Dynamic Content · Last Updated: September 11, 2026

A useful reduction should reconstruct the observed data well

Return to the cells described by many measured features.

- Fewer coordinates cannot preserve every possible input exactly.
- The observed data may nevertheless lie near a lower-dimensional subspace.
- We will judge a representation by how well it reconstructs those observations.

> **What reconstruction error are we willing to accept in exchange for fewer features?**

Four steps lead to the PCA formulation

We will turn “reconstruct the data well” into a problem involving one matrix.

1. Define what a good reconstruction means — Next

Choose a representation size and measure reconstruction error.

2. Choose a convenient decoder basis.

An orthonormal basis can represent the same reconstruction subspace.

3. Find the best encoder for that decoder.

The closest point in the subspace determines the coordinates.

4. Formulate the subspace choice.

Combine these results into the PCA optimization problem.

Centering separates the mean from variation

For notation simplicity, we now call the **raw data** $\widetilde{X}$ and the **centered data** X.

$$\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^{n} \widetilde{\boldsymbol{x}}_i, \qquad X = \widetilde{X} - \mathbf{1}_n \boldsymbol{\mu}^T.$$

- Each row is $\boldsymbol{x}_i^T = (\widetilde{\boldsymbol{x}}_i - \boldsymbol{\mu})^T$.
- Every feature column of X has mean zero.
- **From here on, X and x_i always denote centered data.**

Add the mean back to reconstruct raw measurements: $\widehat{\widetilde{\boldsymbol{x}}}_i = \boldsymbol{\mu} + \widehat{\boldsymbol{x}}_i$.

A linear encoder and decoder make reconstruction explicit

For one centered observation, use the same operator convention as $\boldsymbol{y} = A\boldsymbol{x}$:

$$\boldsymbol{z}_i = \underbrace{E}_{k\times d}\,\boldsymbol{x}_i, \qquad \widehat{\boldsymbol{x}}_i = \underbrace{D}_{d\times k}\,\boldsymbol{z}_i.$$

Stacking observations as rows gives:

$$\underbrace{Z}_{n\times k} = \underbrace{X}_{n\times d}\,\underbrace{E^T}_{d\times k}, \qquad \underbrace{\widehat{X}}_{n\times d} = \underbrace{Z}_{n\times k}\,\underbrace{D^T}_{k\times d}.$$

The batch reconstruction is $\widehat{X} = XE^TD^T$, with rank at most k.

Reconstruction error formalizes one meaning of preservation

Choose a linear encoder and decoder to minimize squared reconstruction error:

$$\min_{E,D} \|X - XE^T D^T\|_F^2.$$

The **Frobenius norm** measures the error across every entry:

$$\|M\|_F^2 = \sum_{i,j} M_{ij}^2, \qquad \|X - \widehat{X}\|_F^2 = \sum_{i=1}^{n} \|\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_i\|_2^2.$$

Four steps lead to the PCA formulation

We have an error measure. Now consider which reconstructions the decoder can produce.

1. Define what a good reconstruction means — Established

Choose a representation size and measure reconstruction error.

2. Choose a convenient decoder basis — Next

An orthonormal basis can represent the same reconstruction subspace.

3. Find the best encoder for that decoder.

The closest point in the subspace determines the coordinates.

4. Formulate the subspace choice.

Combine these results into the PCA optimization problem.

Matrix–vector multiplication combines the matrix's columns

For a 3×2 decoder, the familiar row dot products can be regrouped by column:

$$\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ d_{31} & d_{32} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \underbrace{\begin{bmatrix} d_{11} z_1 + d_{12} z_2 \\ d_{21} z_1 + d_{22} z_2 \\ d_{31} z_1 + d_{32} z_2 \end{bmatrix}}_{\text{one dot product per row}} = z_1 \underbrace{\begin{bmatrix} d_{11} \\ d_{21} \\ d_{31} \end{bmatrix}}_{\boldsymbol{d}_1} + z_2 \underbrace{\begin{bmatrix} d_{12} \\ d_{22} \\ d_{32} \end{bmatrix}}_{\boldsymbol{d}_2}.$$

For any number of columns: $\widehat{\boldsymbol{x}} = D\boldsymbol{z} = \sum_{j=1}^{k} z_j \boldsymbol{d}_j$.

- **Let z be arbitrary for now.** Every reconstruction lies in $\mathcal{S} = \operatorname{span}\{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_k\}$.
- First choose how D describes that subspace; then choose the best coordinates.

Scaling a decoder direction only changes its coordinates

A line can use a unit vector $\boldsymbol{u}$ or the scaled vector $3\boldsymbol{u}$.

$$\underbrace{\boldsymbol{u}}_{\text{unit basis}}\ \underbrace{t}_{\text{coordinate}} = \underbrace{(3\boldsymbol{u})}_{\text{scaled basis}}\ \underbrace{(t/3)}_{\text{adjusted coordinate}}.$$

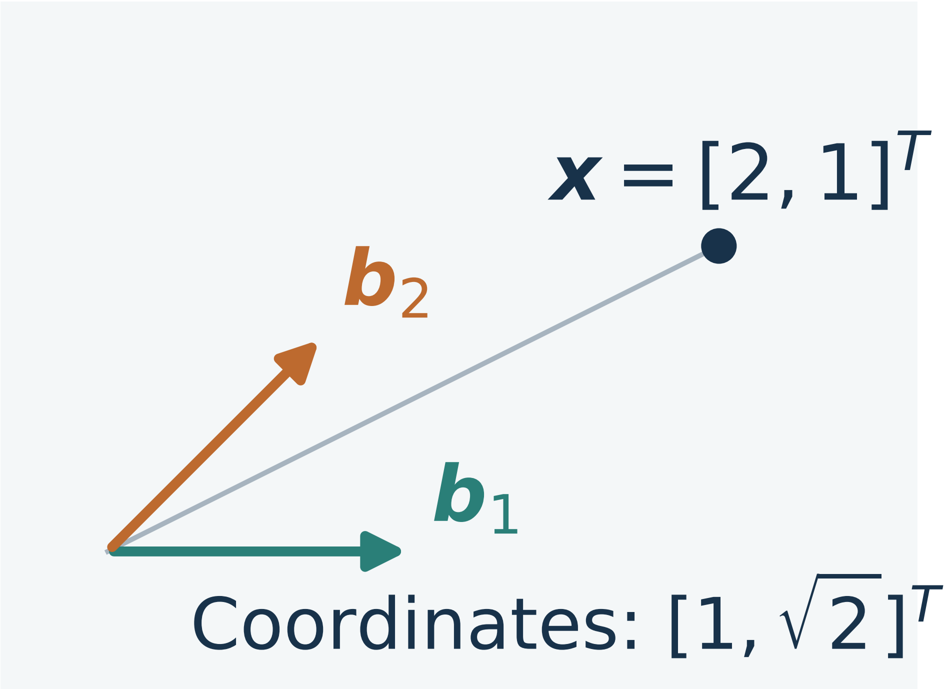
Both describe **the same point on the same line**.

- A nonzero column can be normalized if we rescale its coordinate to compensate.
- Unit columns remove an arbitrary scale choice; no reconstruction is lost.
- Next: can we choose perpendicular columns while keeping the same subspace?

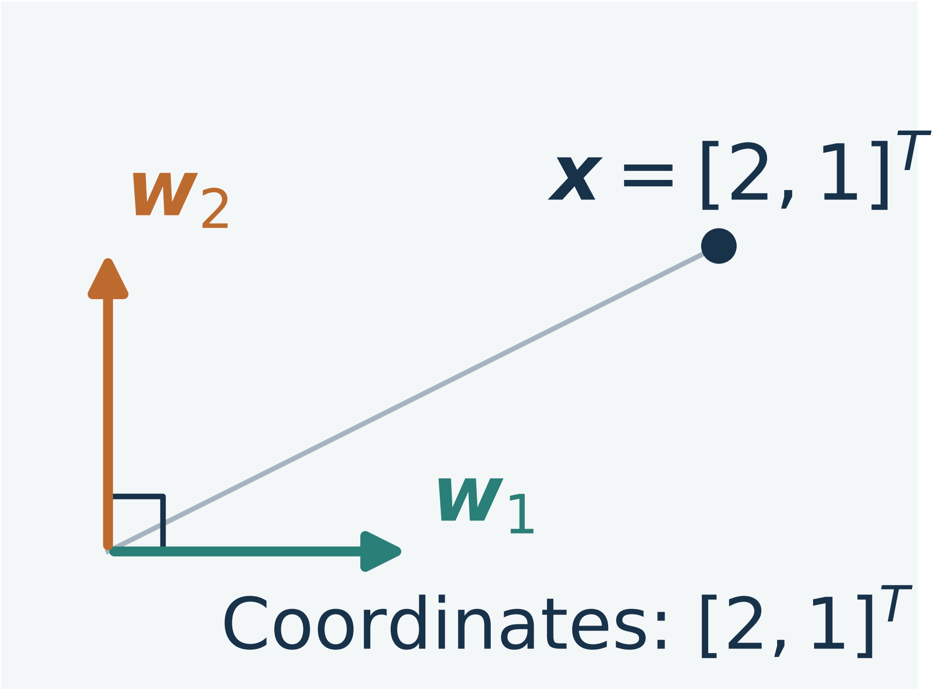
Perpendicular unit directions can describe the same subspace

Both unit-vector pairs span the plane: either represents **every** $\boldsymbol{x} = [x_1, x_2]^T$.

Unit Skew Directions



Orthonormal Directions



For $\boldsymbol{b}_1 = [1, 0]^T, \boldsymbol{b}_2 = [1, 1]^T/\sqrt{2}$:

$$\boldsymbol{x} = (x_1 - x_2)\boldsymbol{b}_1 + \sqrt{2}\, x_2 \boldsymbol{b}_2.$$

For $\boldsymbol{w}_1 = [1, 0]^T, \boldsymbol{w}_2 = [0, 1]^T$:

$$\boldsymbol{x} = (\boldsymbol{w}_1^T \boldsymbol{x})\boldsymbol{w}_1 + (\boldsymbol{w}_2^T \boldsymbol{x})\boldsymbol{w}_2.$$

Same plane; simpler coordinates: an orthonormal basis gives each coefficient by a dot product.

QR lets us choose an orthonormal decoder without changing reconstruction

Start with any encoder $E_0 \in \mathbb{R}^{k\times d}$ and decoder $D_0 \in \mathbb{R}^{d\times k}$, where $k \leq d$. Suppose D_0 does not have orthonormal columns.

The **thin QR decomposition** writes $D_0 = WR$: $W^T W = I_k$ and R is upper triangular. Here $W \in \mathbb{R}^{d\times k}$ and $R \in \mathbb{R}^{k\times k}$.

Can we choose $D = W$ and a new E to keep every reconstruction the same?

$$\begin{aligned} D_0 E_0 \boldsymbol{x} &= (WR)E_0\boldsymbol{x} && \text{(QR of the original decoder)} \\ &= W(RE_0)\boldsymbol{x} && \text{(Regroup the factors)} \\ &= DE\boldsymbol{x} && \text{(Choose } D = W, E = RE_0\text{)} \end{aligned}$$

Same reconstruction for every input: an orthonormal decoder is without loss of generality. Next, fix $D = W$ and find its best encoder.