From last time:

Thm (SVD, outer product form):
Let A be an $m\times n$ matrix of rank r. Then

$$\boxed{A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T}$$

where

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$
- $\{u_1, \dots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$
- $\{v_1, \dots, v_r\}$ is an orthonormal set in $\mathbb{R}^n$.

To compute:

- $\lambda_1 \geq \cdots \geq \lambda_r > 0$ are the nonzero eigenvalues of A^TA
- $\{v_1, \dots, v_r\}$ are orthonormal eigenvectors, $A^TAv_i = \lambda_i v_i$
- $\sigma_i = \sqrt{\lambda_i}$, $u_i = \frac{1}{\sigma_i} A v_i$

SVD of A^T:

$$A^T = \sigma_1 v_1 u_1^T + \sigma_2 v_2 u_2^T + \cdots + \sigma_r v_r u_r^T$$

NB: If A is a wide matrix $(m<n)$ then

A^TA : $n\times n$ $\quad$ AA^T : $m\times m$ is smaller

So it's easier to compute eigenvalues & eigenvectors of AA^T!

$$\boxed{\text{If } A \text{ is wide, compute the SVD of } A^T.}$$

Eg: $A = \begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix}$

$$A^TA = \begin{pmatrix} 200 & -50 & 200 & -50 \\ -50 & 125 & -50 & 125 \\ 200 & -50 & 200 & -50 \\ -50 & 125 & -50 & 125 \end{pmatrix}$$

yikes!

Let's compute the SVD of A^T instead.

$AA^T = \begin{pmatrix} 400 & -100 \\ -100 & 250 \end{pmatrix}$ $\quad p(\lambda) = (\lambda - 450)(\lambda - 200)$

$\lambda_1 = 450 \Rightarrow \sigma_1 = \sqrt{450} = 15\sqrt{2}$ $\quad u_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ $\quad v_1 = \frac{1}{\sigma_1} A^T v_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} -2 \\ 1 \\ -2 \\ 1 \end{pmatrix}$

$\lambda_2 = 200 \Rightarrow \sigma_2 = \sqrt{200} = 10\sqrt{2}$ $\quad u_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $\quad v_2 = \frac{1}{\sigma_2} A^T v_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$

$\Rightarrow A^T = 15\sqrt{2}\, v_1 u_1^T + 10\sqrt{2}\, v_2 u_2^T$

$\Rightarrow A = 15\sqrt{2}\, u_1 v_1^T + 10\sqrt{2}\, u_2 v_2^T$

u_i are right-singular vectors of A^T $\rightarrow$ left-singular vectors of A

SVD: Matrix Form

Let

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

be the SVD of A in outer product form.

Facts:

(1) $\{u_1, \dots, u_r\}$ is an orthonormal basis for $\text{Col}(A)$

(2) $\{v_1, \dots, v_r\}$ is an orthonormal basis for $\text{Row}(A)$

Proof:
- $u_i = A \cdot \frac{1}{\sigma_i} v_i \implies u_i \in \text{Col}(A)$
- $\{u_1, \dots, u_r\}$ is orthonormal $\implies$ linearly independent
- $\dim(\text{Col}(A)) = \text{rank} = r$
- Basis theorem $\implies \{u_1, \dots, u_r\}$ is a basis for $\text{Col}(A)$.

Fact (2) follows by replacing A by A^T.

✓

Let $\{u_{r+1}, \dots, u_m\}$ be an orthonormal basis for $\mathrm{Nul}(A^T)$.

$\rightsquigarrow$ $\{u_1, \dots, u_r, u_{r+1}, \dots, u_m\}$ is an orthonormal basis of $\mathbb{R}^m$

$\rightsquigarrow$ it's also an orthonormal eigenbasis for AA^T

$(\mathrm{Nul}(A^T) = \mathrm{Nul}(AA^T) = 0\text{-eigenspace of } AA^T)$

Let $\{v_{r+1}, \dots, v_n\}$ be an orthonormal basis for $\mathrm{Nul}(A)$.

$\rightsquigarrow$ $\{v_1, \dots, v_r, v_{r+1}, \dots, v_n\}$ is an orthonormal basis of $\mathbb{R}^n$

$\rightsquigarrow$ it's also an orthonormal eigenbasis for A^TA

$(\mathrm{Nul}(A) = \mathrm{Nul}(A^TA) = 0\text{-eigenspace of } A^TA)$

Thm (SVD, Matrix Form):

Let A be an $m \times n$ matrix of rank r. Then

$$A = U \Sigma V^T \quad \text{where:}$$

- $U = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix}$ is an $m \times m$ orthogonal matrix
- $V = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$ is an $n \times n$ orthogonal matrix
- $\Sigma = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & 0 & \\ & & \sigma_r & & \\ & 0 & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix}$ is an $m \times n$ diagonal matrix.

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values

Procedure to Compute $A = U\Sigma V^T$:

(1) Compute the singular values and singular vectors

$\{v_1, \dots, v_r\}$ $\{u_1, \dots, u_r\}$ $\sigma_1, \dots, \sigma_r$

as before

(2) Find orthonormal bases

$\{u_{r+1}, \dots, u_m\}$ for $\mathrm{Nul}(A^T)$

$\{v_{r+1}, \dots, v_n\}$ for $\mathrm{Nul}(A)$

(3) $U = \begin{pmatrix} | & & | & | & & | \\ u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \\ | & & | & | & & | \end{pmatrix}$ $V = \begin{pmatrix} | & & | & | & & | \\ v_1 & \cdots & v_r & v_{r+1} & \cdots & v_n \\ | & & | & | & & | \end{pmatrix}$

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ & & & 0 \\ 0 & & & & \ddots \\ & & & & & 0 \end{pmatrix}$$ (same size as A)

Proof: Use the outer product version of matrix mult:

$$U\Sigma V^T = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix} \begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} - & \sigma_1 v_1 & - \\ & \vdots & \\ - & \sigma_r v_r & - \\ - & 0 & - \\ & \vdots & \\ - & 0 & - \end{pmatrix}$$

$$= \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T + 0 + \cdots + 0$$ ✓

Eg: $A = \begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix}$

(1) $A = 15\sqrt{2}\, u_1 v_1^T + 10\sqrt{2}\, u_2 v_2^T$ for

$$u_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix} \qquad v_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} -2 \\ 1 \\ -2 \\ 1 \end{pmatrix}$$

$$u_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad v_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$$

(2) $\mathrm{Nul}(A^T) = \{0\}$ because $r = m$

$\mathrm{Nul}(A)$: $\begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix} \xrightarrow{\text{rref}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$

$$\xrightarrow{\text{PVF}} \mathrm{Span}\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

[already orthogonal — usually need Gram-Schmidt] $\xrightarrow{\text{normalize}}$ $v_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ $\quad v_4 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$

(3) So $A = U \Sigma V^T$ for

$$U = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 15\sqrt{2} & 0 & 0 & 0 \\ 0 & 10\sqrt{2} & 0 & 0 \end{pmatrix}$$

$$V = \begin{pmatrix} -2/\sqrt{10} & 1/\sqrt{10} & -1/\sqrt{2} & 0 \\ 1/\sqrt{10} & 2/\sqrt{10} & 0 & -1/\sqrt{2} \\ -2/\sqrt{10} & 1/\sqrt{10} & 1/\sqrt{2} & 0 \\ 1/\sqrt{10} & 2/\sqrt{10} & 0 & 1/\sqrt{2} \end{pmatrix}$$

NB: $A = U\Sigma V^T$ contains full orthogonal diagonalizations of A^TA and of AA^T:

$$A^TA = V\begin{pmatrix} \sigma_1^2 & & & 0 \\ & \ddots & & \\ & & \sigma_r^2 & \\ 0 & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}V^T \qquad AA^T = U\begin{pmatrix} \sigma_1^2 & & & 0 \\ & \ddots & & \\ & & \sigma_r^2 & \\ 0 & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}U^T$$

It also contains orthonormal bases for all four subspaces:

$$U = \begin{pmatrix} \overbrace{\begin{matrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{matrix}}^{\text{o.n. basis for } \mathrm{Col}(A)} & \overbrace{\begin{matrix} | & & | \\ u_{r+1} & \cdots & u_m \\ | & & | \end{matrix}}^{\text{o.n. basis for } \mathrm{Nul}(A^T)} \end{pmatrix}$$

$i \le r$ $\boxed{Av_i = \sigma_i u_i \uparrow \; \downarrow A^Tu_i = \sigma_i v_i}$ $\boxed{Av_i = 0 \uparrow \; \downarrow A^Tu_i = 0}$ $i > r$

$$V = \begin{pmatrix} \underbrace{\begin{matrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{matrix}}_{\text{o.n. basis for } \mathrm{Row}(A)} & \underbrace{\begin{matrix} | & & | \\ v_{r+1} & \cdots & v_n \\ | & & | \end{matrix}}_{\text{o.n. basis for } \mathrm{Nul}(A)} \end{pmatrix}$$

The Pseudo-Inverse

This is a matrix A^+ that is the "best possible" substitute for A^{-1} when A is not invertible.

- Works for non-square matrices:
  if A is $m \times n$ then A^+ is $n \times m$
- $A^+ b$ is the shortest least-squares solution of $Ax=b$.

First let's do diagonal matrices.

Def: If Σ is an $m \times n$ diagonal matrix with nonzero diagonal entries $\sigma_1, \dots, \sigma_r$, its pseudo-inverse Σ^+ is the $n \times m$ diagonal matrix with diagonal entries $\sigma_1^{-1}, \dots, \sigma_r^{-1}$.

$$\Sigma = \underbrace{\begin{pmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}}_{4 \times 5} \rightsquigarrow \Sigma^+ = \underbrace{\begin{pmatrix} 1/3 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{5 \times 4}$$

NB: If Σ is invertible (hence square) then $\Sigma^+ = \Sigma^{-1}$:

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Now let's do general matrices.

Def: Let A be an $m \times n$ matrix with SVD

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T \qquad A = U \Sigma V^T$$

The pseudo-inverse of A is the $n \times m$ matrix

$$A^+ = \frac{1}{\sigma_1} v_1 u_1^T + \cdots + \frac{1}{\sigma_r} v_r u_r^T \qquad A^+ = V \Sigma^+ U^T$$

This has the same singular vectors (switch right & left) and reciprocal singular values.

Eg: $A = \begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix} = 15\sqrt{2}\, u_1 v_1^T + 10\sqrt{2}\, u_2 v_2^T$

for $\quad u_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix} \qquad v_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} -2 \\ 1 \\ -2 \\ 1 \end{pmatrix}$

$\quad u_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad v_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$

$$\rightsquigarrow A^+ = \frac{1}{15\sqrt{2}} v_1 u_1^T + \frac{1}{10\sqrt{2}} v_2 u_2^T$$

$$= \frac{1}{15\sqrt{2}} \cdot \frac{1}{\sqrt{10}}\begin{pmatrix} -2 \\ 1 \\ -2 \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & -1 \end{pmatrix} + \frac{1}{10\sqrt{2}} \cdot \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix} \cdot \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \end{pmatrix}$$

$$= \frac{1}{150}\begin{pmatrix} -4 & 2 \\ 2 & -1 \\ -4 & 2 \\ 2 & -1 \end{pmatrix} + \frac{1}{100}\begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 1 & 2 \\ 2 & 4 \end{pmatrix} = \frac{1}{300}\begin{pmatrix} -5 & 10 \\ 10 & 10 \\ -5 & 10 \\ 10 & 10 \end{pmatrix}$$

NB: If A is invertible then $r=m=n$ and Σ is invertible, so $\Sigma^+ = \Sigma^{-1}$ and

$$AA^+ = (U\Sigma V^T)(V\Sigma^+ U^T)$$

$$= U\Sigma(\underbrace{V^T V}_{I})\Sigma^{-1}U^T = U\underbrace{\Sigma\Sigma^{-1}}_{I}U^T = \underbrace{UU^T}_{I} = I_n$$

$$\boxed{A \text{ is invertible} \iff A^{-1} = A^+}$$

In general, for $i \leq r$ we have

same singular vectors
reciprocal singular values

$$A^+Av_i = A^+(\sigma_i u_i) = \sigma_i A^+ u_i = \sigma_i \cdot \frac{1}{\sigma_i} v_i = v_i$$

$$AA^+u_i = A\left(\frac{1}{\sigma_i}v_i\right) = \frac{1}{\sigma_i}Av_i = \frac{1}{\sigma_i}\cdot\sigma_i u_i = u_i$$

But for $i > r$ we have

$$A^+Av_i = A^+\cdot 0 = 0 \qquad (v_i \in \mathrm{Nul}(A))$$

$$AA^+u_i = A\cdot 0 = 0 \qquad (u_i \in \mathrm{Nul}(A^T) = \mathrm{Nul}(A^+))$$

$i \leq r$

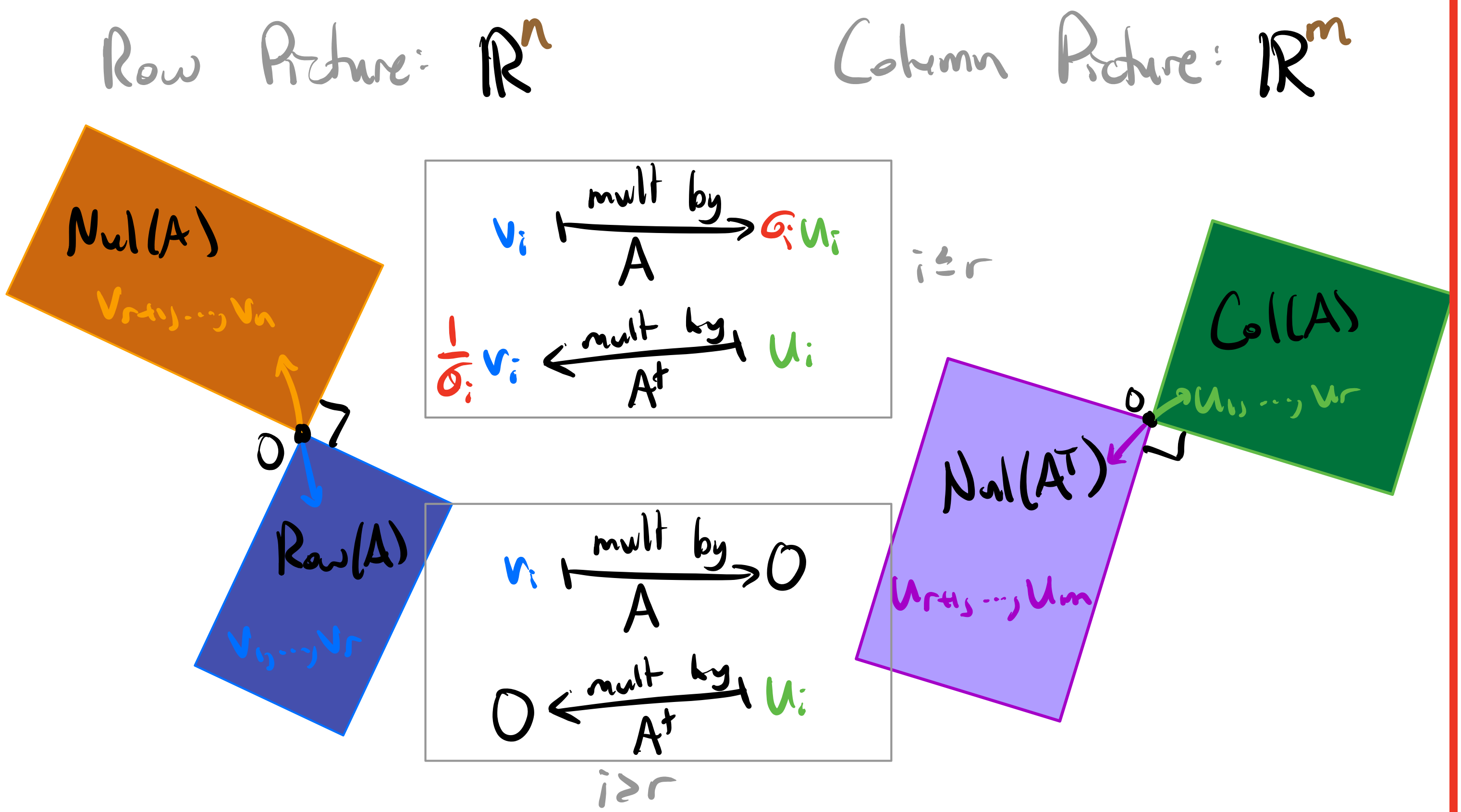
$$\boxed{\begin{aligned} A^+Av_i &= v_i \\ AA^+u_i &= u_i \end{aligned}}$$

$i > r$

$$\boxed{\begin{aligned} A^+Av_i &= 0 \\ AA^+u_i &= 0 \end{aligned}}$$

The Big Picture Revisited

for an $m \times n$ matrix A of rank r



This tells us exactly what A^+A and AA^+ are when A is not invertible:

Prop: $A^+A =$ projection onto $\text{Row}(A)$
$AA^+ =$ projection onto $\text{Col}(A)$

Proof: If $P =$ projection matrix for $\text{Row}(A)$ then

$$Pv_i = v_i = A^+Av_i \quad \text{for } i \leq r \quad (v_i \in \text{Row}(A))$$
$$Pv_i = 0 = A^+Av_i \quad \text{for } i > r \quad (v_i \in \text{Nul}(A) = \text{Row}(A)^\perp)$$

Now take linear combinations. Same for AA^+. ✓

Recall: A projection matrix P_V is the identity matrix

$\Longrightarrow V$ is all of $\mathbb{R}^n$

Consequence:

- $A^+A = I_n \Longleftrightarrow A$ has full column rank

  $\left(\mathrm{Row}(A) = \mathrm{Nul}(A)^\perp = \{0\}^\perp = \mathbb{R}^n\right)$

- $AA^+ = I_m \Longleftrightarrow A$ has full row rank

  $\left(\mathrm{Col}(A) = \mathbb{R}^m\right)$

Eg:

$$A = \begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix} \qquad A^+ = \frac{1}{300}\begin{pmatrix} -5 & 10 \\ 10 & 10 \\ -5 & 10 \\ 10 & 10 \end{pmatrix}$$

$$A^+A = \frac{1}{300}\begin{pmatrix} -5 & 10 \\ 10 & 10 \\ -5 & 10 \\ 10 & 10 \end{pmatrix}\begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix} = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}$$

$$AA^+ = \frac{1}{300}\begin{pmatrix} -10 & 10 & -10 & 10 \\ 10 & 5 & 10 & 5 \end{pmatrix}\begin{pmatrix} -5 & 10 \\ 10 & 10 \\ -5 & 10 \\ 10 & 10 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$\left(\mathrm{Col}(A) = \mathbb{R}^2 \Rightarrow \text{projection is } I_2\right)$

Prop: For any $b \in \mathbb{R}^m$, $\hat{x} = A^+b$ is the shortest least-squares solution of $Ax = b$.

Proof: First note $A\hat{x} = AA^{+}b =$ projection of b onto $\mathrm{Col}(A)$

$\Longrightarrow \hat{x}$ is a least-squares solution.

Note $\hat{x} = \frac{1}{\sigma_1} v_1 u_1^T b + \cdots + \frac{1}{\sigma_r} v_r u_r^T b$

$$= \frac{1}{\sigma_1}(u_1 \cdot b) v_1 + \cdots + \frac{1}{\sigma_r}(u_r \cdot b) v_r$$

$$\in \mathrm{Span}\{v_1, \dots, v_r\} = \mathrm{Row}(A). \qquad \rightsquigarrow \hat{x} \in \mathrm{Row}(A)$$

Any other solution $\hat{x}'$ has the form $\hat{x}' = \hat{x} + y$ for $y \in \mathrm{Nul}(A)$. Note $y \in \mathrm{Row}(A)^{\perp} \Longrightarrow \hat{x} \cdot y = 0$

$$\|\hat{x} + y\|^2 = (\hat{x} + y) \cdot (\hat{x} + y) = \hat{x} \cdot \hat{x} + \cancelto{0}{2\hat{x} \cdot y} + y \cdot y$$

$$\|\hat{x} + y\|^2 = \|\hat{x}\|^2 + \|y\|^2 \geq \|\hat{x}\|^2$$

$\Longrightarrow \hat{x}$ is the shortest ✓

Eg: $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 2 \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \end{pmatrix}$

$A^{+} = \frac{1}{2} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

The shortest least-squares solution of $Ax = b = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ is $\hat{x} = A^{+}b = \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

All other least-squares solutions differ by $\mathrm{Nul}(A) = \mathrm{Span}\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}$.