# Singular Value Decomposition

ARIZONA STATE UNIVERSITY
School of Mathematical & Statistical Sciences

## Definition/Theorem (SVD)

The factorization

$$\underset{\substack{m\times n\\ \boxed{m\geq n}}}{A} = \underset{\substack{U \textbf{ orthogonal}\\ m\times m\\ U^{-1}=U^T}}{\begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}} \underset{\substack{\Sigma \textbf{ diagonal}\\ m\times n\\ \sigma_1\geq\cdots\geq\sigma_n\geq 0}}{\begin{bmatrix} \sigma_1 & & 0\\ & \ddots & \\ & & \sigma_n \\ 0 & & \end{bmatrix}} \underset{\substack{V \textbf{ orthogonal}\\ n\times n\\ V^{-1}=V^T}}{\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}} = U\Sigma V^T$$

or

$$\underset{\substack{m\times n\\ \boxed{m\leq n}}}{A} = \underset{\substack{U \textbf{ orthogonal}\\ m\times m\\ U^{-1}=U^T}}{\begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}} \underset{\substack{\Sigma \textbf{ diagonal}\\ m\times n\\ \sigma_1\geq\cdots\geq\sigma_m\geq 0}}{\begin{bmatrix} \sigma_1 & & 0 & \\ & \ddots & & \\ 0 & & \sigma_m & \end{bmatrix}} \underset{\substack{V \textbf{ orthogonal}\\ n\times n\\ V^{-1}=V^T}}{\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}} = U\Sigma V^T$$

exists for any $A$ and is called **singular value decomposition** (SVD) of $A$.

$\mathbf{u}_1, \mathbf{u}_2, \ldots =$ left singular vectors

$\sigma_1, \sigma_2, \ldots =$ singular values

$\mathbf{v}_1, \mathbf{v}_2, \ldots =$ right singular vectors

**Theorem**

*rank(A) = number of $\neq 0$ singular values (counting possibly multiple values).*

**Proof.** (Assume $m > n$ for convenience). Let $\sigma_1, \ldots, \sigma_r$ be the nonzero singular values of $A$

$$A = U\Sigma V^T = \left[\mathbf{u}_1 \cdots \mathbf{u}_r \cdots\cdots \mathbf{u}_m\right] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

$$= \left[\mathbf{u}_1 \cdots \mathbf{u}_r\right] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} = U_r \Sigma_r V_r^T \text{ with } \begin{cases} U_r^T U_r = I_r \\ \Sigma_r \text{ square diagonal, nonsingular} \\ V_r^T V_r = I_r \end{cases}$$

$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

Each of the matrices $\mathbf{u}_i \mathbf{v}_i^T$ has rank 1 and since the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r$ are linearly independent, the sum has rank $r$.

## Algebraic determination of SVD ($A$ is $m \times n$ with $\text{rank}(A) = r$)

$\boxed{(A^T A)V_r = V_r \Sigma_r^T U_r^T U_r \Sigma_r V_r^T V_r = V_r \Sigma_r^2} \Rightarrow$

❶ The singular values $\sigma_j$ are the square roots of the eigenvalues $\lambda_j$ of $A^T A$

❷ The columns $\mathbf{v}_1, \ldots, \mathbf{v}_r$ of $V_r$ are orthonormal eigenvectors of $A^T A$:

$$A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_j = \sigma_j^2 \mathbf{v}_j, \quad j = 1, \ldots, r$$

$A^T A$ symmetric
$\Rightarrow \perp$ eigenvectors

$\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an ON basis of $\mathrm{R}(A^T)$ (row space of $A$).

$\boxed{U_r = A V_r \Sigma_r^{-1}} \Rightarrow$ ❸ $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ given by $\mathbf{u}_j = \dfrac{1}{\sigma_j} A \mathbf{v}_j, \quad j = 1, \ldots, r$

$\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ ON basis of $\mathrm{R}(A)$ (column space of $A$)

$\boxed{A^T \mathbf{u}_{r+1} = \ldots = A^T \mathbf{u}_m = \mathbf{0}} \Rightarrow$ ❹ [if needed] $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ ON basis of $\mathrm{N}(A^T)$

$$A^T \xrightarrow{\text{RREF}} \text{basis of } \mathrm{N}(A^T) \xrightarrow[\text{Schmidt}]{\text{Gram}} \{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\} \text{ ON basis of } \mathrm{N}(A^T)$$

$\boxed{A \mathbf{v}_{r+1} = \ldots = A \mathbf{v}_n = \mathbf{0}} \Rightarrow$ ❺ [if needed] $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ ON basis of $\mathrm{N}(A)$

$$A \xrightarrow{\text{RREF}} \text{basis of } \mathrm{N}(A) \xrightarrow[\text{Schmidt}]{\text{Gram}} \{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\} \text{ ON basis of } \mathrm{N}(A)$$

**Example 1. Determine the SVD of $A = \begin{bmatrix} 2 & 2 \\ -1 & 2 \end{bmatrix}$**

❶ $A^TA = \begin{bmatrix} 5 & 2 \\ 2 & 8 \end{bmatrix} \Rightarrow$

$$\begin{aligned} \det(A^TA - \lambda I) &= (5-\lambda)(8-\lambda) - 2^2 \\ &= \lambda^2 - 13\lambda + 36 \\ &= (\lambda - 9)(\lambda - 4) = 0 \end{aligned} \quad \Rightarrow \quad \begin{cases} \lambda_1 = 9 \Rightarrow \sigma_1 = \sqrt{\lambda_1} = 3 \\ \lambda_2 = 4 \Rightarrow \sigma_2 = \sqrt{\lambda_2} = 2 \end{cases}$$

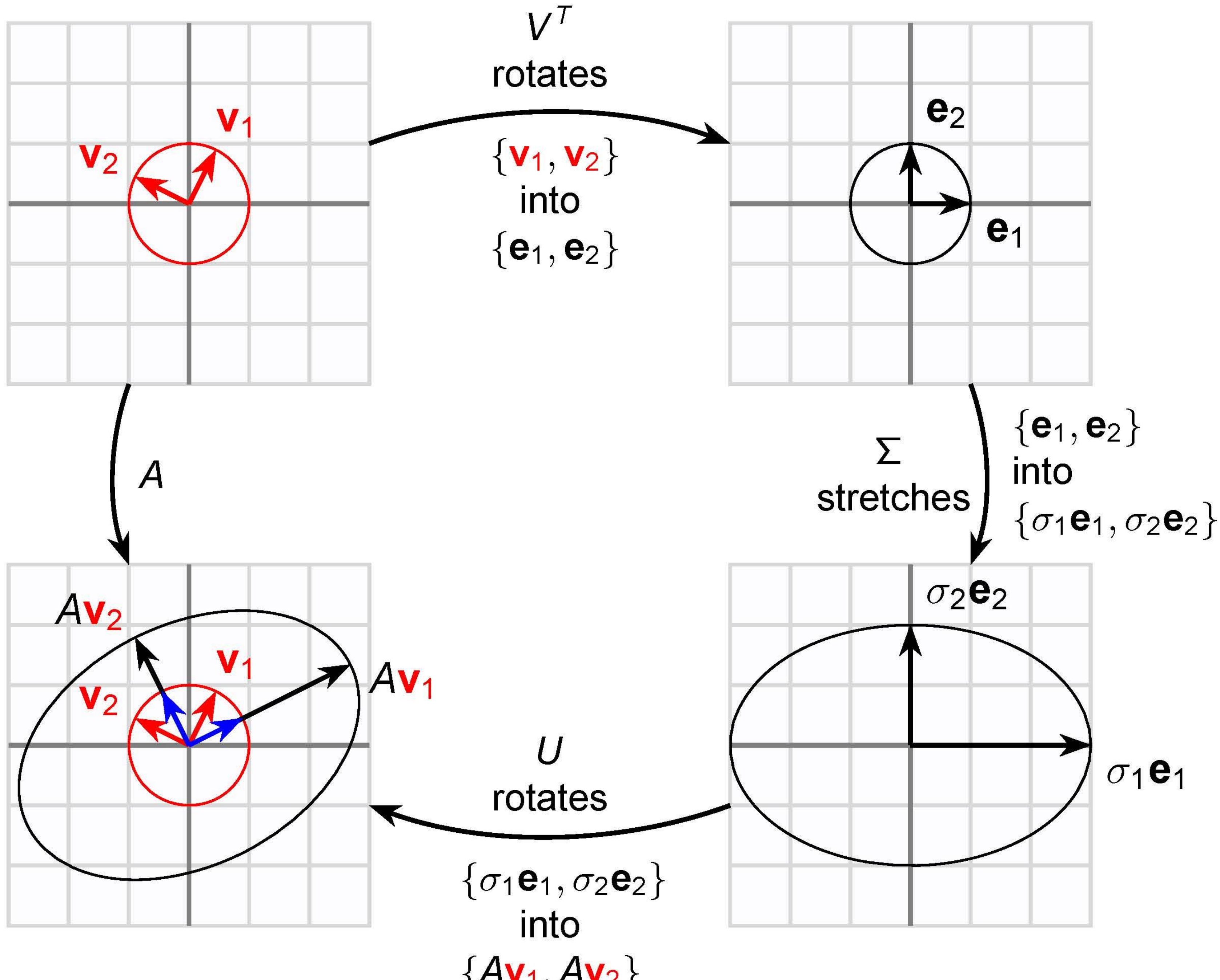
❷ $A^TA - \lambda_1 I = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{bmatrix} \Rightarrow \mathbf{v} = t\begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \xrightarrow{\text{normalize}} \mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$

$A^TA - \lambda_2 I = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \Rightarrow \mathbf{v} = t\begin{bmatrix} -2 \\ 1 \end{bmatrix} \xrightarrow{\text{normalize}} \mathbf{v}_2 = \begin{bmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$

❸ $\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{1}{3}\begin{bmatrix} 2 & 2 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$, $\mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = \frac{1}{2}\begin{bmatrix} 2 & 2 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}^T$$

# SVD $=$ rotation + scaling + rotation



**Example 2.** **Determine the SVD of** $A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}$

❶ $A^TA = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow \det(A^TA - \lambda I) = (2-\lambda)^2 - 1 = 0 \Rightarrow \begin{cases} \lambda_1 = 3 \Rightarrow \sigma_1 = \sqrt{\lambda_1} = \sqrt{3} \\ \lambda_2 = 1 \Rightarrow \sigma_2 = \sqrt{\lambda_2} = 1 \end{cases}$

❷ $A^TA - \lambda_1 I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \Rightarrow \mathbf{v} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix} \xrightarrow{\text{normalize}} \mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

$A^TA - \lambda_2 I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow \mathbf{v} = t \begin{bmatrix} -1 \\ 1 \end{bmatrix} \xrightarrow{\text{normalize}} \mathbf{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

❸ $\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}$, $\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{1} \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$

❹ $A^T \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \Rightarrow \mathbf{u} = t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \xrightarrow{\text{normalize}} \mathbf{u}_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T = \begin{bmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T$$

## Properties of SVD

**Property 1** The singular values measure the stretching/compression of vectors by $A$:

$$\sigma_{\min(m,n)} \le \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \le \sigma_1$$

(equalities hold when $\mathbf{x} = \mathbf{v}_{\min(m,n)}$ and $\mathbf{x} = \mathbf{v}_1$)

**Property 2** $A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$, $1 \le k < r$, is the matrix of rank $k$ closest to $A$ when distance is measured in the Frobenius norm. The distance from $A$ to $A_k$ is

$$\min_{\text{any } B} \|A - B\|_F = \|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$$

In particular, if $r = n$ ($\Rightarrow m \ge n$) then $\sigma_n$ is the distance to the nearest rank-deficient ($m > n$) or singular ($m = n$) matrix.

**Property 3** For square matrices, the ratio $0 \le \dfrac{\sigma_n}{\sigma_1} \le 1$ is a better measure of proximity to a singular matrix than $\det(A)$.

**Example 3**. **Let** $A = U\Sigma V^T = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} & 0 \\ \frac{4}{5} & \frac{3}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 30 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}^T$.

**Determine the rank 1 and 2 matrices** $A_1$ **and** $A_2$ **closest to** $A$ **in the Frobenius norm and evaluate their distance to** $A$

- $\text{rank}(A) = 3$ since $\sigma_3 = 3 > 0$
- Closest (w.r.t. Frobenius norm) matrix $A_1$ of rank 1 (use Property 2):

$$A_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = 30 \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} = \boxed{\begin{bmatrix} 6 & 12 & 12 \\ 8 & 16 & 16 \\ 0 & 0 & 0 \end{bmatrix}}$$

$$\|A - A_1\|_F = \sqrt{\sigma_2^2 + \sigma_3^2} = \boxed{\sqrt{234}} \approx 15.29$$

- Closest (w.r.t Frobenius norm) matrix $A_2$ of rank 2 (use Property 2):

$$A_2 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T = A_1 + 15 \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix} = \boxed{\begin{bmatrix} -2 & 8 & 20 \\ 8 & 16 & 16 \\ 0 & 0 & 0 \end{bmatrix}}$$

$$\|A - A_2\|_F = \sqrt{\sigma_3^2} = \sigma_3 = \boxed{3}$$

## Theorem (SVD and Least Squares)

*If $m \geq n = r = rank(A)$ the solution of the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$ is*

$$\mathbf{x} = V\Sigma_n^{-1}U_n^T\mathbf{b} = \frac{\mathbf{u}_1^T b}{\sigma_1}\mathbf{v}_1 + \ldots + \frac{\mathbf{u}_n^T b}{\sigma_n}\mathbf{v}_n$$

**Proof**: $$A = U\Sigma V^T = \begin{bmatrix} \mathbf{u}_1 \cdots \mathbf{u}_n \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} = U_n\Sigma_n V^T$$

$$\begin{aligned}
A^T A\mathbf{x} = A^T\mathbf{b} \quad & \Leftrightarrow \quad (U_n\Sigma_n V^T)^T(U_n\Sigma_n V^T)\mathbf{x} = (U_n\Sigma_n V^T)^T\mathbf{b} \\
(\Sigma_n \text{ diagonal}, \Sigma_n^T = \Sigma_n) \quad & \Leftrightarrow \quad V\Sigma_n(U_n^T U_n)\Sigma_n V^T\mathbf{x} = V\Sigma_n U_n^T\mathbf{b} \\
(U_n^T U_n = I,\ V^T V = I) \quad & \Leftrightarrow \quad \Sigma_n V^T\mathbf{x} = U_n^T\mathbf{b} \\
(\Sigma_n\ n \times n \text{ nonsingular}) \quad & \Leftrightarrow \quad V^T\mathbf{x} = \Sigma_n^{-1}U_n^T\mathbf{b} \\
(V^T = V^{-1}) \quad & \Leftrightarrow \quad \mathbf{x} = V\Sigma_n^{-1}U_n^T\mathbf{b}
\end{aligned}$$

$$\mathbf{x} = V\begin{bmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_n \end{bmatrix}\begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}\mathbf{b} = V\begin{bmatrix} \mathbf{u}_1^T b/\sigma_1 \\ \vdots \\ \mathbf{u}_n^T b/\sigma_n \end{bmatrix} = \frac{\mathbf{u}_1^T b}{\sigma_1}\mathbf{v}_1 + \ldots + \frac{\mathbf{u}_n^T b}{\sigma_n}\mathbf{v}_n$$

**Example 4.** A matrix $A$ has SVD $A = \begin{bmatrix} .5 & .5 & .5 & .5 \\ .5 & -.5 & -.5 & .5 \\ .5 & .5 & -.5 & -.5 \\ .5 & -.5 & .5 & -.5 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 5 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} .6 & .8 \\ -.8 & .6 \end{bmatrix}$.

**Solve the LS problem** $A\mathbf{x} = \begin{bmatrix} 4 \\ 4 \\ 1 \\ -4 \end{bmatrix} = \mathbf{b}$.

$$\mathbf{x} = \frac{\mathbf{u}_1^T\mathbf{b}}{\sigma_1}\mathbf{v}_1 + \frac{\mathbf{u}_2^T\mathbf{b}}{\sigma_2}\mathbf{v}_2 = \frac{2.5}{10}\begin{bmatrix} .6 \\ .8 \end{bmatrix} + \frac{2.5}{5}\begin{bmatrix} -.8 \\ .6 \end{bmatrix} = \boxed{\begin{bmatrix} -.25 \\ .5 \end{bmatrix}}$$

Equivalently,

$$U\Sigma V^T\mathbf{x} = \begin{bmatrix} 4 \\ 4 \\ 1 \\ -4 \end{bmatrix} \Leftrightarrow U_2\Sigma_2 V^T\mathbf{x} = \begin{bmatrix} .5 & .5 \\ .5 & -.5 \\ .5 & .5 \\ .5 & -.5 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} .6 & .8 \\ -.8 & .6 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 4 \\ 4 \\ 1 \\ -4 \end{bmatrix}$$

$$\Leftrightarrow \mathbf{x} = V\Sigma_2^{-1}U_2^T \begin{bmatrix} 4 \\ 4 \\ 1 \\ -4 \end{bmatrix} = \begin{bmatrix} .6 & -.8 \\ .8 & .6 \end{bmatrix} \begin{bmatrix} \frac{1}{10} & 0 \\ 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} .5 & .5 & .5 & .5 \\ .5 & -.5 & .5 & -.5 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \\ 1 \\ -4 \end{bmatrix} = \boxed{\begin{bmatrix} -.25 \\ .5 \end{bmatrix}}$$