Singular Values and Singular Vectors

CE40282-1: Linear Algebra

Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

Singular Value

- $S_{m \times n}$

  $$\sigma_i = \sqrt{\lambda_i} \qquad \lambda_i \in \sigma(S^T S),\ i = 1, \dots, n$$

- $$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{m-1} \geq \sigma_m$$

- Example

$$S = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

$$S^T S = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix} \implies \sigma(S^T S) = \{360, 90, 0\}$$

$$\implies \begin{cases} \sigma_1 = \sqrt{360} = 6\sqrt{10} \\ \sigma_2 = \sqrt{90} = 3\sqrt{10} \\ \sigma_3 = 0 \end{cases}$$

Singular value and eigenvalue

- **Lemma**

  $\{v_1, \ldots, v_n\}$ are orthonormal eigenvectors of matrix $S^T S$ then singular values of matrix S are norm of Sv_i vectors:

$$||Sv_i|| = \sigma_i$$

  - Proof?

Example:

$$S = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \qquad S^T S = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

$$\sigma_1 = \sqrt{360},\ \sigma_2 = \sqrt{90}, \sigma_3 = 0$$

$$v_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, v_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, v_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix} \quad :$$

$$Sv_1 = \begin{bmatrix} 18 \\ 6 \end{bmatrix} \Longrightarrow \|Sv_1\| = \sqrt{18^2 + 6^2} = \sqrt{360} = \sigma_1$$

$$Sv_2 = \begin{bmatrix} 3 \\ -9 \end{bmatrix} \Longrightarrow \|Sv_2\| = \sqrt{3^2 + (-9)^2} = \sqrt{90} = \sigma_2$$

$$Sv_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longrightarrow \|Sv_3\| = 0 = \sigma_3$$

Singular value and Rank

- **Lemma**

  $\{v_1, \ldots, v_n\}$ are orthonormal eigenvectors of matrix $S^T S$ and S has r non-zero singular value:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0, \quad \sigma_{r+1} = \cdots = \sigma_n = 0$$

  - $\{Sv_1, \ldots, Sv_r\}$ is a orthogonal basis for range of S
  - $rank(S) = r$
    - Proof?
  - $\{u_1, \ldots, u_r\}$ is a orthogonal basis for range of S

Rank of Matrix = Number of nonzero singular values

Introduction

- Generalization of the spectral decomposition that applies to all matrices, rather than just normal matrices.
- Applications:
  - Compute the size of a matrix (in a way that typically makes more sense than norm)
  - Provide a new geometric interpretation of linear transformations
  - Solve optimization problems
  - Construct an “almost inverse” for matrices that do not have an inverse.

Singular Value Decomposition (SVD)

- Given any $m \times n$ matrix **A**, algorithm to find matrices **U**, **V**, and $\sum$ such that (**always exists**)

$$A = U \Sigma V^T \qquad\qquad A = U \Sigma V^*$$

U is $m \times m$ and orthonormal (always real)
$\sum$ is $m \times n$ and diagonal with non-negative (always real) called **singular values**.
V is $n \times n$ and orthonormal (always real)

Columns of **U** are the eigenvectors of $\mathbf{AA}^T$ (called the left singular vectors).
Columns of **V** are the eigenvectors of $\mathbf{A}^T\mathbf{A}$ (called the right singular vectors).
The non-zero singular values are the positive square roots of non-zero eigenvalues of $\mathbf{AA}^T$ or $\mathbf{A}^T\mathbf{A}$.

SVD Comparison

SVD	Diagonalization	Spectral decomposition	Schur triangularization
applies to every single matrix (even rectangular ones).	only applies to matrices with a basis of eigenvectors	only applies to normal matrices	only applies to square matrices
matrix $\sum$ in the middle of the SVD is diagonal (and even has real non-negative entries)	do not guarantee an entrywise non-negative matrix	do not guarantee an entrywise non-negative matrix	only results in an upper triangular middle piece
It requires two unitary matrices U and V	only required one invertible matrix	only required one unitary matrix	only required one unitary matrix

SVD

- The $\sum_i$ are called the singular values of **A**
- If **A** is singular, some of the $\sum_i$ will be 0
- In general *rank*(**A**) = number of nonzero $\sum_i$
- SVD is mostly unique (up to permutation of singular values, or if some $\sum_i$ are equal)

SVD for Square Matrix

The SVD is a factorization of a $m\times n$ matrix into

$$\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}$$

where $\boldsymbol{U}$ is a $m\times m$ orthogonal matrix, $\boldsymbol{V}^{\boldsymbol{T}}$ is a $n\times n$ orthogonal matrix and $\boldsymbol{\Sigma}$ is a $m\times n$ diagonal matrix.

For a square matrix ($\boldsymbol{m} = \boldsymbol{n}$):

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \dots$$

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \dots & \vdots \\ \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \\ \vdots & \dots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} \dots & \mathbf{v}_1^T & \dots \\ \vdots & \vdots & \vdots \\ \dots & \mathbf{v}_n^T & \dots \end{pmatrix}$$

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \dots & \vdots \\ \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \\ \vdots & \dots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} \vdots & \dots & \vdots \\ \boldsymbol{v}_1 & \dots & \boldsymbol{v}_n \\ \vdots & \dots & \vdots \end{pmatrix}^T$$

Reduced SVD

- $$[Sv_1 \quad \cdots \quad Sv_r \quad 0 \quad \cdots \quad 0]_{m\times n} = [\sigma_1 u_1 \quad \cdots \quad \sigma_r u_r \quad 0 \quad \cdots \quad 0]_{m\times n}$$

$$[Sv_1 \quad \cdots \quad Sv_r \quad Sv_{r+1} \quad \cdots \quad Sv_n] = [\sigma_1 u_1 \quad \cdots \quad \sigma_r u_r \quad 0 \quad \cdots \quad 0]$$

$$S\,[v_1 \quad \cdots \quad v_n] = [u_1 \quad \cdots \quad u_m] \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & & \vdots & 0 \\ 0 & \cdots & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$

$$S_{m\times n} V_{n\times n} = U_{m\times m} \Sigma_{m\times n}$$

$$S = U\Sigma V^T$$

Reduced SVD

What happens when $\boldsymbol{A}$ is not a square matrix?

- $\boldsymbol{n > m}$

$$\boldsymbol{A = U\,\Sigma\,V^T} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \\ \vdots & \cdots & \vdots \end{pmatrix} \left(\boxed{\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{matrix}} \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \right) \begin{pmatrix} \boxed{\begin{matrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_m^T & \cdots \end{matrix}} \\ \begin{matrix} \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{matrix} \end{pmatrix}$$

$m{\times}m$ $\quad$ $m{\times}n$ $\quad$ $n{\times}n$

We can instead re-write the above as:

$$\boldsymbol{A = U\,\Sigma_R V_R{}^T}$$

Now U and V are not orthogonal.
But their columns are orthonormal.

where $\boldsymbol{V_R}$ is a $n{\times}m$ matrix and $\boldsymbol{\Sigma_R}$ is a $m{\times}m$ matrix

In general:

$$\boldsymbol{A = U_R\Sigma_R V_R{}^T}$$

$\boldsymbol{U_R}$ is a $m{\times}k$ matrix
$\boldsymbol{\Sigma_R}$ is a $k\times k$ matrix
$\boldsymbol{V_R}$ is a $n{\times}k$ matrix

$$k = \min(m,n)$$

Reduced SVD

- $m > n$

$$A = U\,\Sigma\,V^T = \begin{pmatrix} \boxed{\begin{matrix} \vdots & \cdots & \vdots \\ u_1 & \cdots & u_n \\ \vdots & \cdots & \vdots \end{matrix}} & \begin{matrix} \cdots & \vdots \\ \cdots & u_m \\ \cdots & \vdots \end{matrix} \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{matrix}} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}$$

$m{\times}m$ $\quad$ $m{\times}n$ $\quad$ $n{\times}n$

We can instead re-write the above as:

$$A = U_R\,\Sigma_R V^T$$

Now U and V are not orthogonal.
But their columns are orthonormal.

Where U_R is a $m{\times}n$ matrix and Σ_R is a $n{\times}n$ matrix

Reduced SVD

Let's take a look at the product $\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ has the singular values of a $\boldsymbol{A}$, a $m{\times}n$ matrix.

$$\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} = \underset{n\times m}{\begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_n & & & 0 \end{pmatrix}} \underset{m\times n}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix}} = \underset{n\times n}{\boxed{\begin{pmatrix} {\sigma_1}^2 & & \\ & \ddots & \\ & & {\sigma_n}^2 \end{pmatrix}}}$$

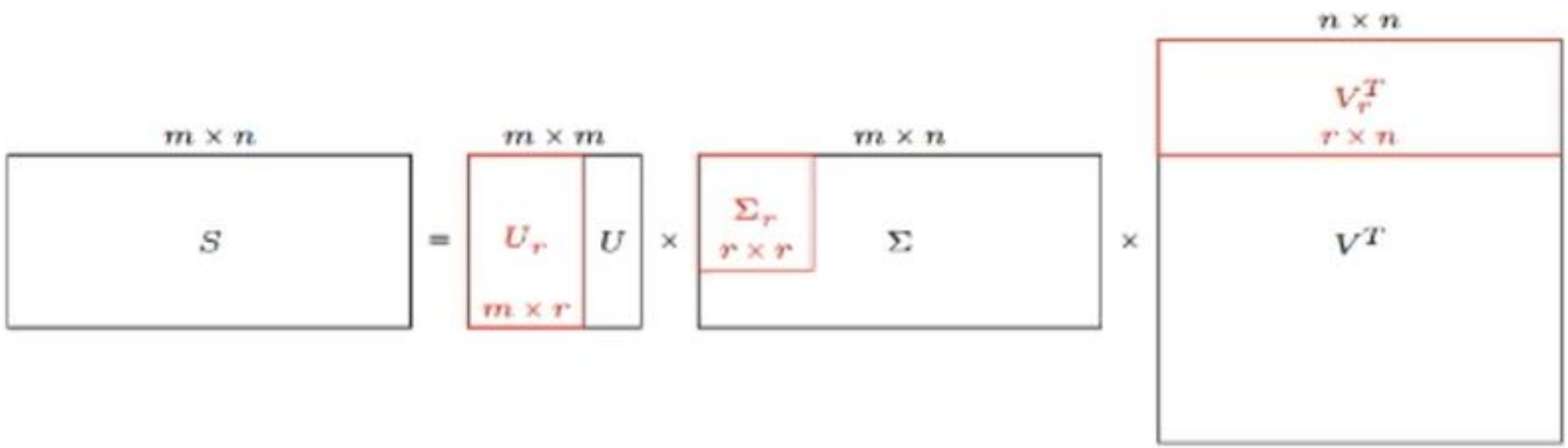
$m > n$

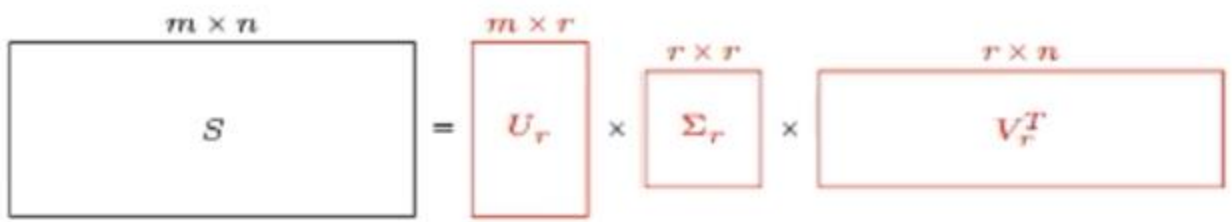
$$\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} = \underset{n\times m}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix}} \underset{m\times n}{\begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_m & & & 0 \end{pmatrix}} = \underset{n\times n}{\begin{pmatrix} \boxed{\begin{matrix} {\sigma_1}^2 & & \\ & \ddots & \\ & & {\sigma_m}^2 \end{matrix}} & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \\ \begin{matrix} 0 & & 0 \\ & \ddots & \\ & & 0 \end{matrix} & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{pmatrix}}$$

$n > m$

Reduced SVD

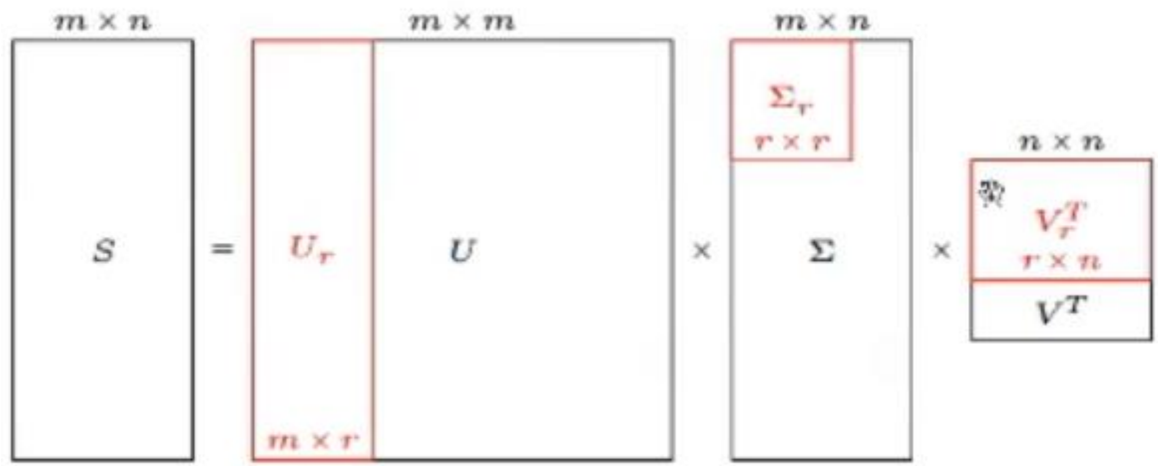
- Wide Matrix

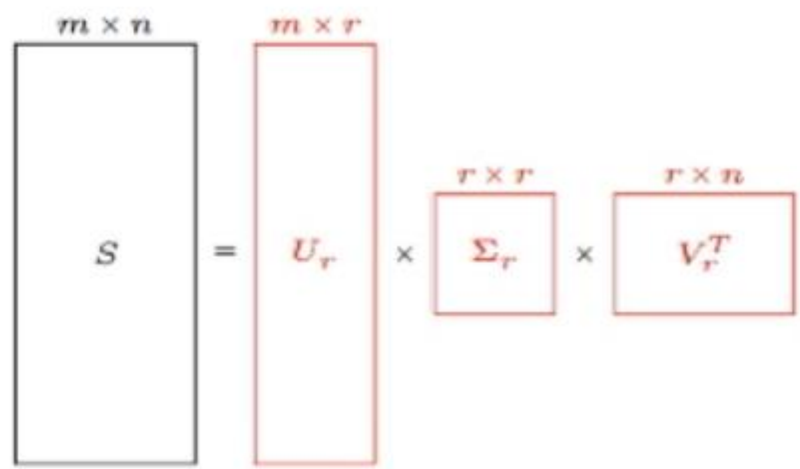




Reduced SVD

- Tall Matrix





How can we compute an SVD of a matrix A?

Assume $\boldsymbol{A}$ with the singular value decomposition $\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}$. Let's take a look at the eigenpairs corresponding to $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$:

$$\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} = \left(\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}\right)^{\boldsymbol{T}}\left(\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}\right)$$

$$\left(\boldsymbol{V}^{\boldsymbol{T}}\right)^{\boldsymbol{T}}(\,\boldsymbol{\Sigma}\,)^{\boldsymbol{T}}\boldsymbol{U}^{\boldsymbol{T}}\left(\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}\right) = \boldsymbol{V}\boldsymbol{\Sigma}^{\boldsymbol{T}}\boldsymbol{U}^{\boldsymbol{T}}\,\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}} = \boldsymbol{V}\,\boldsymbol{\Sigma}^{\boldsymbol{T}}\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}$$

Hence $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} = \boldsymbol{V}\,\boldsymbol{\Sigma}^{\boldsymbol{2}}\,\boldsymbol{V}^{\boldsymbol{T}}$

Recall that columns of $\boldsymbol{V}$ are all linear independent (orthogonal matrix), then from diagonalization ($\boldsymbol{B} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-\boldsymbol{1}}$), we get:

- the columns of $\boldsymbol{V}$ are the eigenvectors of the matrix $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$
- The diagonal entries of $\boldsymbol{\Sigma}^{\boldsymbol{2}}$ are the eigenvalues of $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$

Let's call λ the eigenvalues of $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$, then $\sigma_i{}^2 = \lambda_i$

How can we compute an SVD of a matrix A?

- In a similar way,

$$\boldsymbol{AA^T} = (\boldsymbol{U\,\Sigma\,V^T})\,(\boldsymbol{U\,\Sigma\,V^T})^{\boldsymbol{T}}$$

$$(\boldsymbol{U\,\Sigma\,V^T})(\boldsymbol{V^T})^{\boldsymbol{T}}(\,\boldsymbol{\Sigma}\,)^{\boldsymbol{T}}\boldsymbol{U^T} = \boldsymbol{U\,\Sigma\,V^TV\Sigma^TU^T} = \boldsymbol{U\Sigma\,\Sigma^TU^T}$$

Hence $\boldsymbol{AA^T} = \boldsymbol{U\,\Sigma^2\,U^T}$

Recall that columns of $\boldsymbol{U}$ are all linear independent (orthogonal matrices), then from diagonalization ($\boldsymbol{B} = \boldsymbol{XDX^{-1}}$), we get:

- The columns of $\boldsymbol{U}$ are the eigenvectors of the matrix $\boldsymbol{AA^T}$

How can we compute an SVD of a matrix A?

1. Evaluate the n eigenvectors $\mathbf{v}_i$ and eigenvalues λ_i of $\boldsymbol{A}^T\boldsymbol{A}$
2. Make a matrix $\boldsymbol{V}$ from the normalized vectors $\mathbf{v}_i$. The columns are called "right singular vectors".

$$\boldsymbol{V} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \vdots & \cdots & \vdots \end{pmatrix}$$

3. Make a diagonal matrix from the square roots of the eigenvalues.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \quad \sigma_i = \sqrt{\lambda_i} \quad \text{and} \quad \sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots$$

4. Find $\boldsymbol{U}$: $\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^T \Longrightarrow \boldsymbol{U}\,\boldsymbol{\Sigma} = \boldsymbol{A}\,\boldsymbol{V} \Longrightarrow \boldsymbol{U} = \boldsymbol{A}\,\boldsymbol{V}\,\boldsymbol{\Sigma}^{-1}$. The columns are called the "left singular vectors".

How can we compute an SVD of a matrix A?

- Example

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

$$S^T S = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix} \qquad \text{rank}(S) = 1$$

$$\Delta(\lambda) = \lambda^2 - 18\lambda = 0 \Rightarrow \sigma_1 = \sqrt{18},\ \sigma_2 = 0 \Rightarrow \boxed{\Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}$$

$$v_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, v_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \Longrightarrow \boxed{V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}$$

$$Sv_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix} \Longrightarrow u_1 = \frac{1}{\sigma_1} Sv_1 = \frac{1}{3\sqrt{2}} \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

$$u_2 = \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix}, u_3 = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix} \Longrightarrow \boxed{U = \frac{1}{3} \begin{bmatrix} 1 & 2 & -2 \\ -2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}}$$

$$S = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 2 & -2 \\ -2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = U\Sigma V^T$$

Lemma

- Unitary Freedom of PSD Decompositions

Suppose $B, C \in \mathcal{M}_{m,n}(\mathbb{F})$. The following are equivalent:

a) There exists a unitary matrix $U \in \mathcal{M}_m(\mathbb{F})$ such that $C = UB$,

b) $B^*B = C^*C$,

c) $(B\mathbf{v}) \cdot (B\mathbf{w}) = (C\mathbf{v}) \cdot (C\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{F}^n$, and

d) $\|B\mathbf{v}\| = \|C\mathbf{v}\|$ for all $\mathbf{v} \in \mathbb{F}^n$.

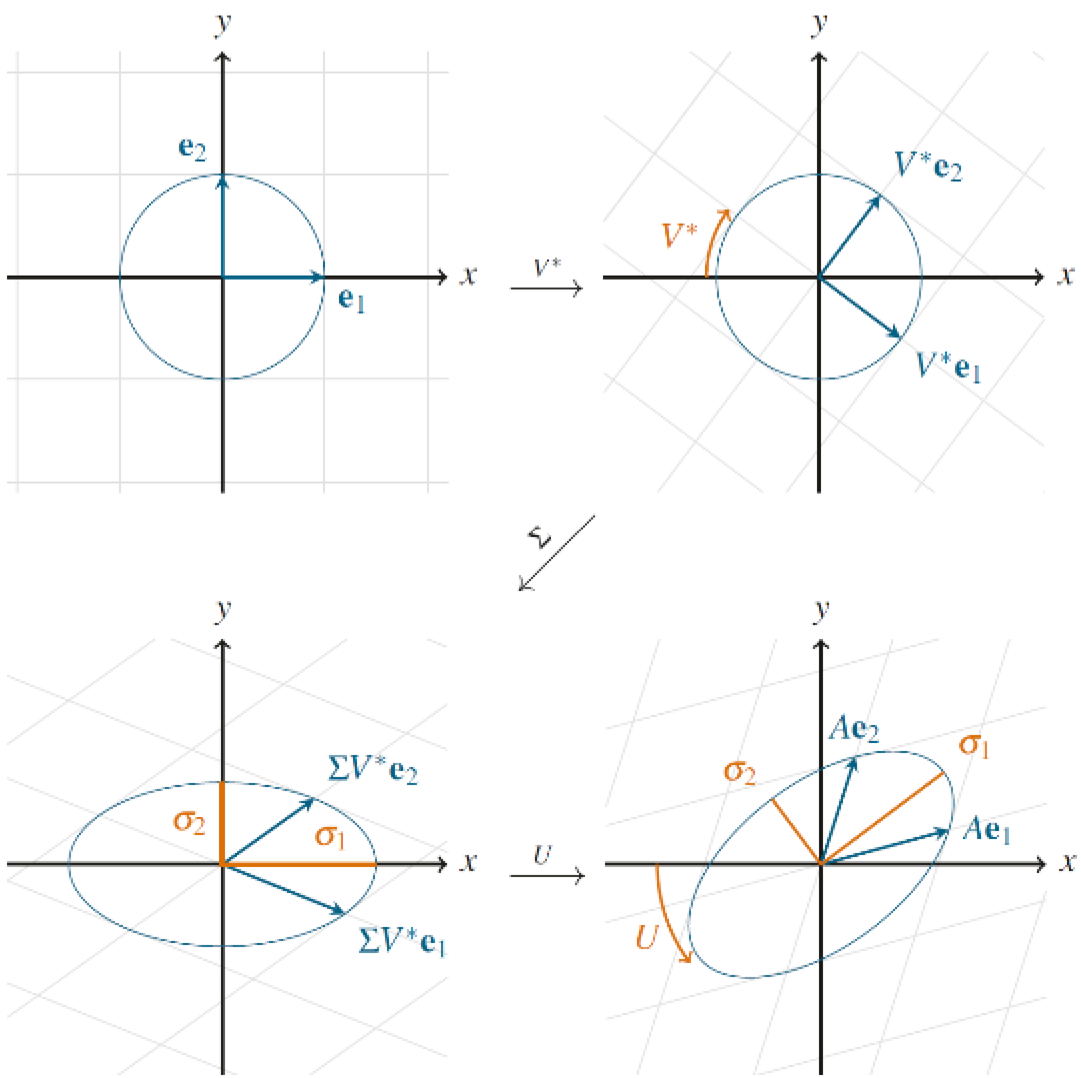
SVD Proof

- If $m \neq n$ then A^*A, AA^* have different sizes, but they still have essentially the same eigenvalues—whichever one is larger just has some extra 0 eigenvalues.
- The same is actually true of AB and BA for any A and B.

- Proof SVD:

Geometric Interpretation and the Fundamental Subspaces

the product of a matrix's singular values equals the absolute value of its determinant



Determining the rank of a matrix

Suppose $\boldsymbol{A}$ is a $m \times n$ rectangular matrix where $m > n$:

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \cdots & \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n & \cdots & \boldsymbol{u}_m \\ \vdots & \cdots & \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix} \begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}$$

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \cdots & \sigma_1\,\mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \sigma_n\,\mathbf{v}_n^T & \cdots \end{pmatrix} = \sigma_1\boldsymbol{u}_1\mathbf{v}_1^T + \sigma_2\boldsymbol{u}_2\mathbf{v}_2^T + \cdots + \sigma_n\boldsymbol{u}_n\mathbf{v}_n^T$$

$$\boldsymbol{A} = \sum_{i=1}^{n} \sigma_i \boldsymbol{u}_i \mathbf{v}_i^T$$

$\boldsymbol{A}_1 = \sigma_1\boldsymbol{u}_1\mathbf{v}_1^T$ what is $\text{rank}(\boldsymbol{A}_1) = ?$

In general, $\text{rank}(\boldsymbol{A}_k) = k$

Rank of a matrix

For general rectangular matrix $\boldsymbol{A}$ with dimensions $m \times n$, the reduced SVD is:

- $$\underbrace{\boldsymbol{A}}_{m \times n} = \underbrace{\boldsymbol{U}_{\boldsymbol{R}}}_{m \times k} \underbrace{\boldsymbol{\Sigma}_{\boldsymbol{R}}}_{k \times k} \underbrace{\boldsymbol{V}_{\boldsymbol{R}}{}^{\boldsymbol{T}}}_{k \times n}$$

$$k = \min(m, n)$$

$$\boldsymbol{A} = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \mathbf{v}_i^T$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \\ 0 & & 0 \\ & \ddots & \vdots \\ & & 0 \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_k & 0 & \dots & 0 \end{pmatrix}$$

If $\sigma_i \neq 0 \; \forall i$, then $\text{rank}(\boldsymbol{A}) = k$ (Full rank matrix)

In general, $\text{rank}(\boldsymbol{A}) = \boldsymbol{r}$, where $\boldsymbol{r}$ is the number of non-zero singular values σ_i

$r < k$ (Rank deficient)

Rank of a matrix

- The rank of $\mathbf{A}$ equals the number of non-zero singular values which is the same as the number of non-zero diagonal elements in $\mathbf{\Sigma}$.
- Rounding errors may lead to small but non-zero singular values in a rank deficient matrix, hence the rank of a matrix determined by the number of non-zero singular values is sometimes called "effective rank".
- The right-singular vectors (columns of $\boldsymbol{V}$) corresponding to vanishing singular values span the null space of $\mathbf{A}$.
- The left-singular vectors (columns of $\boldsymbol{U}$) corresponding to the non-zero singular values of $\mathbf{A}$ span the range of $\mathbf{A}$.

Conclusion

- Let $A \in \mathcal{M}_{m,n}$ be a matrix with $\operatorname{rank}(A) = r$ and singular value decomposition $A = U\Sigma V^*$, where

$$U = \left[\, \mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_m \,\right] \quad \text{and} \quad V = \left[\, \mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n \,\right].$$

Then

a) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ is an orthonormal basis of $\operatorname{range}(A)$,

b) $\{\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \ldots, \mathbf{u}_m\}$ is an orthonormal basis of $\operatorname{null}(A^*)$,

c) $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthonormal basis of $\operatorname{range}(A^*)$, and

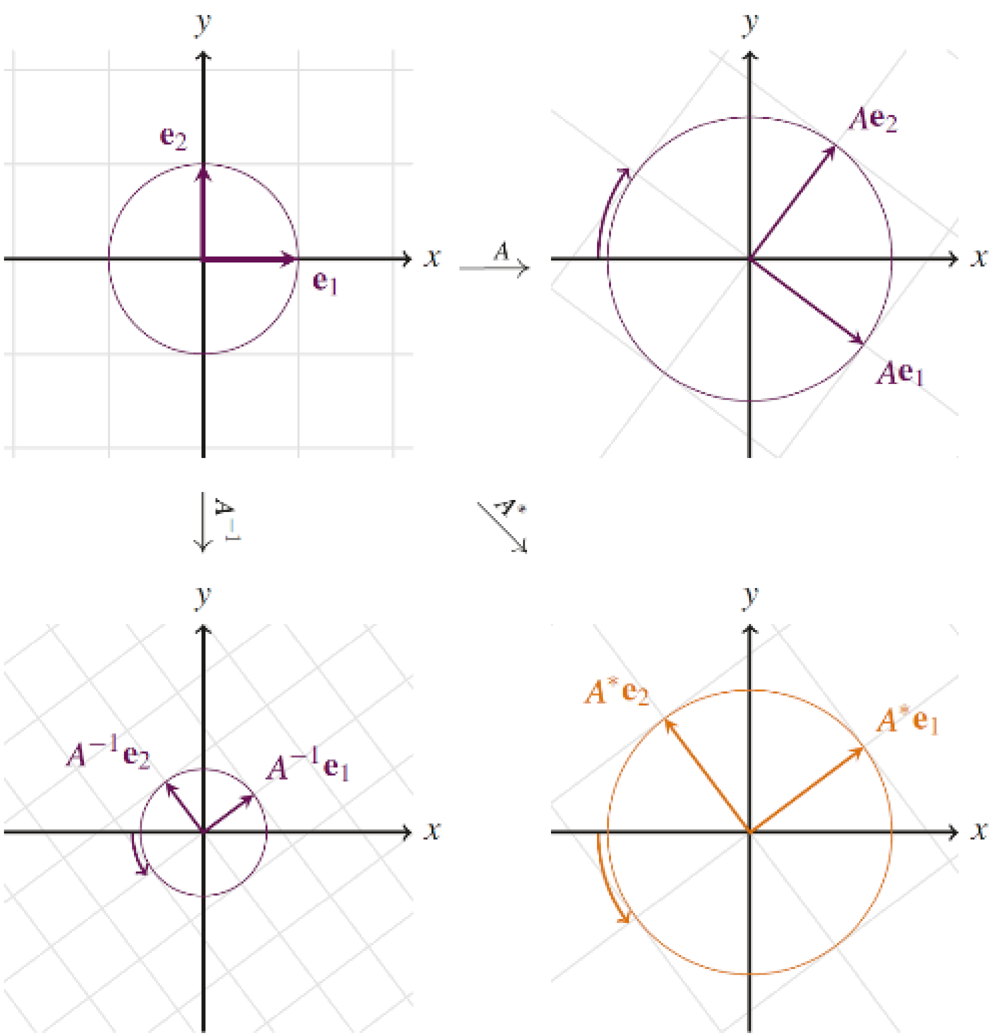
d) $\{\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of $\operatorname{null}(A)$.

A Geometric Interpretation

$$A = U\Sigma V^*$$

$$A^* = V\Sigma^* U^*$$

$$A^{-1} = V\Sigma^{-1} U^*$$



SVD and Inverses

- Why is SVD so useful?
- Application #1: inverses
- $A^{-1}=(V^T)^{-1}\Sigma^{-1}U^{-1} = V\Sigma^{-1}U^T$
  - Using fact that inverse = transpose for orthogonal matrices
  - Since Σ is diagonal, Σ^{-1} also diagonal with reciprocals of entries of Σ

SVD and Inverses

- $A^{-1}=(V^T)^{-1}\Sigma^{-1}U^{-1} = V\Sigma^{-1}U^T$
- This fails when some Σ_i are 0
  - It's *supposed* to fail – singular matrix
- Pseudoinverse: if $\Sigma_i=0$, set $1/\Sigma_i$ to 0 (!)
  - "Closest" matrix to inverse
  - Defined for all (even non-square, singular, etc.) matrices
  - Equal to $(A^TA)^{-1}A^T$ if A^TA invertible

Pseudo-inverse

- **Problem:** if A is rank-deficient, **Σ** is not be invertible

  How to fix it: Define the Pseudo Inverse

  Pseudo-Inverse of a diagonal matrix:

$$(\boldsymbol{\Sigma}^{+})_i = \begin{cases} \frac{1}{\sigma_i}, & \text{if } \sigma_i \neq 0 \\ 0, & \text{if } \sigma_i = 0 \end{cases}$$

  Pseudo-Inverse of a matrix $\boldsymbol{A}$**:**

$$\boldsymbol{A}^{+} = \boldsymbol{V}\boldsymbol{\Sigma}^{+}\boldsymbol{U}^{T}$$

Pseudo-inverse

If a matrix A has the singular value decomposition

$$A = UWV^T$$

then the pseudo-inverse or Moore-Penrose inverse of A is

$$A^+ = V^T W^{-1} U$$

If A is 'tall' (m>n) and has full rank

$$A^+ = (A^T A)^{-1} A^T$$ (it gives the least-squares solution $x_{lsq} = A^+ b$)

If A is 'short' (m<n) and has full rank

$$A^+ = A^T (AA^T)^{-1}$$ (it gives the least-norm solution $x_{l\text{-}n} = A^+ b$)

In general, $x_{pinv} = A^+ b$ is the minimum-norm, least-squares solution.

SVD and Eigenvectors

- Let **A=U ∑V**^{T}, and let x_i be i^{th} column of **V**
- Consider **A**^{T}**A** x_i:

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}x_i = \mathbf{V}\Sigma^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}x_i = \mathbf{V}\Sigma^2\mathbf{V}^{\mathrm{T}}x_i = \mathbf{V}\Sigma^2\begin{pmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{pmatrix} = \mathbf{V}\begin{pmatrix}0\\ \vdots\\ {\Sigma_i}^2\\ \vdots\\ 0\end{pmatrix}$$

$$= {\Sigma_i}^2 x_i$$

- So elements of ∑ are sqrt(eigenvalues) and columns of **V** are eigenvectors of **A**^{T}**A**
  - What we wanted for robust least squares fitting!

SVD and Matrix Similarity

- One common definition for the norm of a matrix is the Frobenius norm:

$$\|\mathbf{A}\|_{\mathrm{F}} = \sum_i \sum_j a_{ij}^{\ 2}$$

- Frobenius norm can be computed from SVD

$$\|\mathbf{A}\|_{\mathrm{F}} = \sum_i \Sigma_i^{\ 2}$$

- So changes to a matrix can be evaluated by looking at changes to singular values

SVD and Matrix Similarity

- Suppose you want to find best rank-*k* approximation to **A**
  - Answer: set all but the largest *k* singular values to zero
- Can form compact representation by eliminating columns of **U** and **V** corresponding to zeroed Σ_i