Singular Value Decomposition
(matrix factorization)

Singular Value Decomposition

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

$m\times m$ $\quad$ $m\times n$

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

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

u_i (left singular vectors) $\quad$ singular values $\quad$ right singular vector

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots$$

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

$$\boldsymbol{A} = \underbrace{\begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \\ \vdots & \cdots & \vdots \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} \sigma_1 & & \phi \\ & \ddots & \\ \phi & & \sigma_n \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ \vdots & \cdots & \vdots \end{pmatrix}^T}_{V} \qquad \sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots$$

$$A = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} - & v_1^T & - \\ - & v_2^T & - \\ & \vdots & \\ - & v_n^T & - \end{pmatrix}$$

Reduced SVD

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

1) $\boldsymbol{m} > \boldsymbol{n}$

$$\underset{m \times n}{\boldsymbol{A}} = \underset{m\times m}{\boldsymbol{U}}\,\underset{m\times n}{\boldsymbol{\Sigma}}\,\boldsymbol{V}^T = \begin{pmatrix} \vdots & \cdots & \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n & \cdots & \boldsymbol{u}_m \\ \vdots & \cdots & \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \\ & \vdots & \\ & 0 & \end{pmatrix} \begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}$$

$m \times m$ $\qquad$ $m \times n$ $\qquad$ $n \times n$

$m > n$

$$\underset{m\times n}{\underline{\underline{A}}} = \underset{m\times n}{\underline{\underline{U}}_R}\,\underset{n\times n}{\underline{\underline{\Sigma}}_R}\,\underset{n\times n}{V^T}$$

$$\begin{pmatrix} \sigma_1 \mathbf{v}_1^T \\ \sigma_2 \mathbf{v}_2^T \\ \vdots \\ \sigma_n \mathbf{v}_n^T \\ 0 \end{pmatrix}$$

$m \times n$

Reduced SVD

2) $n > m$ $\quad \underset{m}{\overset{n}{\underline{\underline{A}}}}$

$$\underset{m\times n}{A} = \underset{m\times m}{U}\,\underset{m\times n}{\Sigma}\,\underset{n\times n}{V^T} = \underbrace{\begin{pmatrix} \vdots & \cdots & \vdots \\ u_1 & \cdots & u_m \\ \vdots & & \vdots \end{pmatrix}}_{m\times m} \underbrace{\begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_m & & & 0 \end{pmatrix}}_{m\times n} \underbrace{\begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_m^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}}_{n\times n}$$

$m\times m \qquad m\times m \qquad m\times n$

$m < n$

$$\underline{\underline{A}} = \underset{m\times m}{\underline{\underline{U}}}\;\underset{m\times m}{\underline{\underline{\Sigma_R}}}\;\underset{m\times n}{V_R^T}$$

General:

$$\underset{m\times n}{A} = \underline{\underline{U_R}}\;\underline{\underline{\Sigma_R}}\;\underline{\underline{V_R^T}}$$

	U_R	Σ_R	V_R^T
$m<n$	$m\times m$	$m\times m$	$m\times n$
$m>n$	$m\times n$	$n\times n$	$n\times n$

U_R : $m\times k$

Σ_R : $k\times k$

V_R^T : $k\times n$

$k = \min(m, n)$

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.

$m > n$

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

Σ_R^2

$n > m$

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

Σ_R^2

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

(x, λ) $\boxed{A^T A x = \lambda x}$

$(ABC)^T = C^T B^T A^T$

$$A^T A = (U \Sigma V^T)^T (U \Sigma V^T)$$

$$= (V^T)^T \Sigma^T \underbrace{U^T U} \Sigma V^T$$

$$= V \Sigma^T I \Sigma V^T$$

$$= V \Sigma^T \Sigma V^T$$

$\Sigma^2 = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}$

$\underline{\underline{U}}^T = \underline{\underline{U}}^{-1}$

$\underline{\underline{U}}^{-1} \underline{\underline{U}} = \underline{\underline{I}}$

$$\boxed{A^T A = \boxed{V}\,\boxed{\Sigma^2}\,V^T} \longleftrightarrow$$

Diagonalization:

$$\underline{\underline{B}} = \underline{\underline{X}}\,\underline{\underline{D}}\,\underline{\underline{X}}^{-1}$$

$\Rightarrow$ columns of $\underline{\underline{V}}$ are the eigenvectors of $\underline{\underline{A}}^T \underline{\underline{A}}$

$\Rightarrow$ diagonal entries of $\underline{\underline{\Sigma}}^2$ are the eigenvalues of $A^T A$

$x, \lambda = eig(A^T A)$ $\quad \boxed{\lambda_i = \sigma_i^2}$

In a similar way,

A^TA

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

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

Hence $AA^T = U\,\Sigma^2\,U^T$

$$\boxed{AA^T = U\Sigma^2 U^T}$$

$$B = XDX^{-1}$$

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

→ columns of U are the eigenvectors of AA^T

- The columns of U are the eigenvectors of the matrix 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 $\boldsymbol{A}^T\boldsymbol{A}$ la.eig(A^TA)
2. Make a matrix $\boldsymbol{V}$ from the normalized vectors $\mathbf{v}_i$. The columns are called "right singular vectors".

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

(*) note from after class: this is the correct equation. The video has incorrect one (sorry :-)

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

singular values $\sigma_i^2 = \lambda_i$

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

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

$\boxed{U = AV\Sigma^{-1}}$ (*)

True or False?

$\boldsymbol{A}$ has the singular value decomposition $\boxed{\boldsymbol{A} = \boldsymbol{U\,\Sigma\,V^T}}$.

- The matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are not singular True $U^{-1} = U^T$
- The matrix $\boldsymbol{\Sigma}$ can have zero diagonal entries True
- $\|\boldsymbol{U}\|_2 = 1$ True
- The SVD exists when the matrix $\boldsymbol{A}$ is singular True
- The algorithm to evaluate SVD will fail when taking the square root of a negative eigenvalue False $\lambda_i, x_i = \text{la.eig}(A^TA)$

$\sigma_i = \sqrt{\lambda_i} \rightarrow \lambda_i \geq 0$ Why?!

Singular values are always non-negative

A^TA

- A matrix is positive definite if $\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} > \mathbf{0}$ for $\forall \boldsymbol{x} \neq \mathbf{0}$
- A matrix is positive semi-definite if $\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} \geq \mathbf{0}$ for $\forall \boldsymbol{x} \neq \mathbf{0}$

$$x^T(A^TA)x = \underbrace{\underbrace{(Ax)^T}_{y}\underbrace{Ax}_{y}}_{y \cdot y = \|y\|_2^2} = \|Ax\|_2^2 \geq 0$$

A^TA is positive semi-definite

$$A^TAx = \lambda x \longrightarrow (x, \lambda)$$

$$x^TA^TAx = \lambda x^Tx = \lambda \underbrace{\|x\|_2^2}_{\geq 0}$$

$$\lambda = \frac{x^TA^TAx}{\|x\|_2^2} = \frac{\|Ax\|_2^2}{\|x\|^2} \geq 0$$

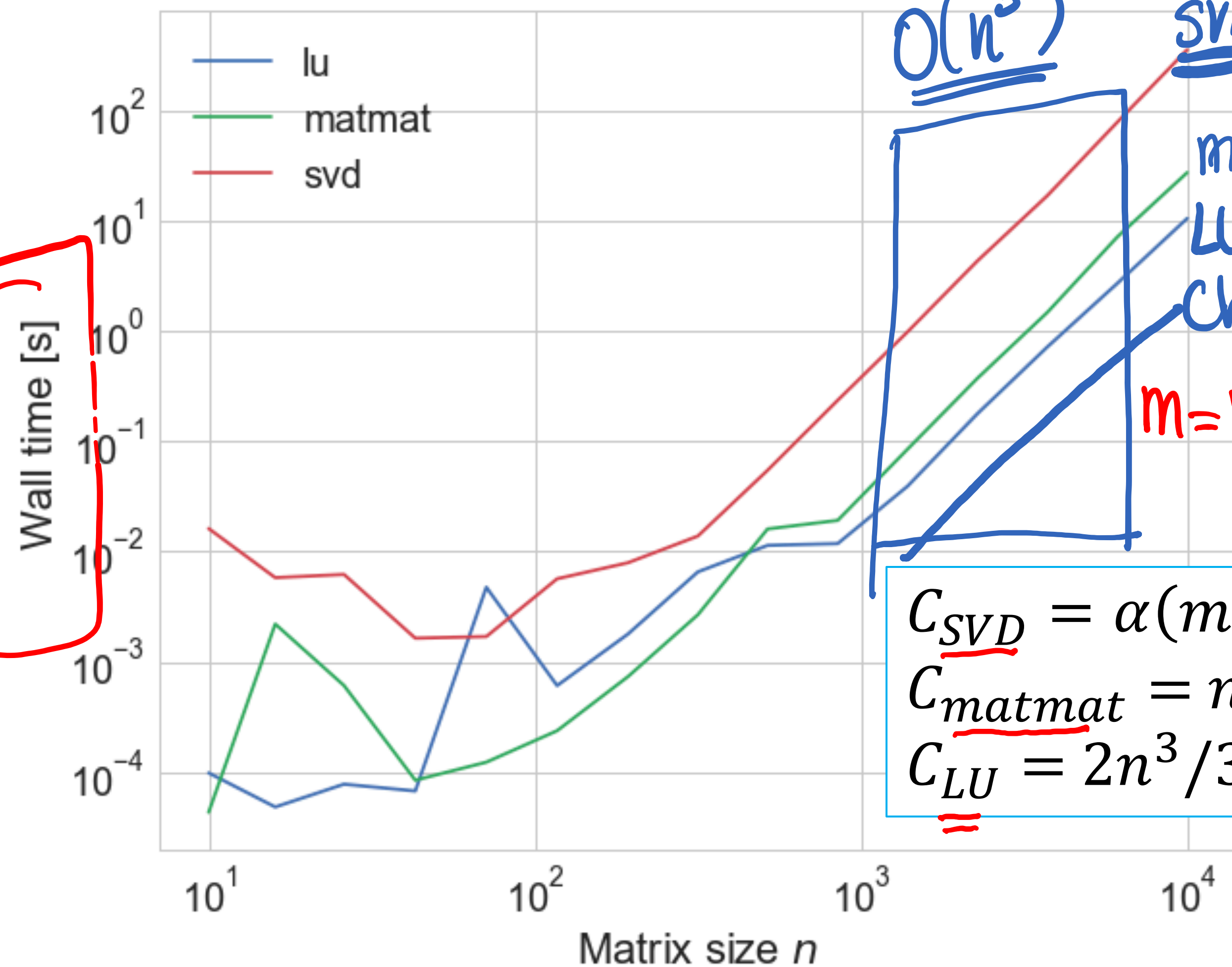
A^TA

$$\boxed{\lambda_i \geq 0}$$

Cost of SVD

$A_{m\times n}$

The cost of an SVD is proportional to $\boldsymbol{m\,n^2 + n^3}$ where the constant of proportionality constant ranging from 4 to 10 (or more) depending on the algorithm.



$O(n^3)$

SVD cost: $\propto (mn^2+n^3)$

mat.mat

LU

Chol $(4, 10)$

$m=n \rightarrow \text{cost} = \alpha\, n^3$

$$C_{SVD} = \alpha(m\,n^2 + n^3) = O(n^3)$$
$$C_{matmat} = n^3 = O(n^3)$$
$$C_{LU} = 2n^3/3 = O(n^3)$$

$\frac{2}{3}n^3$

$A_{n\times n}$ Chol. $\frac{1}{3}n^3$

SVD summary:

- 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.

- In reduced form: $\boldsymbol{A} = \boldsymbol{U}_{\boldsymbol{R}}\boldsymbol{\Sigma}_{\boldsymbol{R}}{\boldsymbol{V}_{\boldsymbol{R}}}^{\boldsymbol{T}}$, where $\boldsymbol{U}_{\boldsymbol{R}}$ is a $m \times k$ matrix, $\boldsymbol{\Sigma}_{\boldsymbol{R}}$ is a $k \times k$ matrix, and $\boldsymbol{V}_{\boldsymbol{R}}$ is a $n \times k$ matrix, and $k = \min(m, n)$.

- The columns of $\boldsymbol{V}$ are the eigenvectors of the matrix $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$, denoted the right singular vectors.

- The columns of $\boldsymbol{U}$ are the eigenvectors of the matrix $\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}}$, denoted the left singular vectors.

- The diagonal entries of $\boldsymbol{\Sigma}^{\boldsymbol{2}}$ are the eigenvalues of $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$. $\sigma_i = \sqrt{\lambda_i}$ are called the singular values.

- The singular values are always non-negative (since $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$ is a positive semi-definite matrix, the eigenvalues are always $\lambda \geq 0$)

Singular Value Decomposition (applications)

1) Determining the rank of a matrix

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

$$\boldsymbol{A} = \underbrace{\left(\begin{matrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \\ \vdots & \cdots & \vdots \end{matrix}\right.}_{\underset{=}{U_R}}\overset{\underset{=}{U}}{\left.\begin{matrix} \cdots & \vdots \\ \cdots & \boldsymbol{u}_m \\ \cdots & \vdots \end{matrix}\right)} \overset{\underset{=}{\Sigma}}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix}} \overset{\underset{=}{V}^T}{\begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}}$$

$\underset{\sim}{u}\,\underset{\sim}{v}^T$

$n\times 1$, $1\times n$ $\Rightarrow n\times n$

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

$$\boxed{\underset{=}{A} = \sum_{i=1}^{n} \sigma_i \underset{\sim}{u}_i \underset{\sim}{v}_i^T}$$

$$\underset{=}{A}_1 = \underbrace{\sigma_1 \underset{\sim}{u}_1 \underset{\sim}{v}_1^T}$$

$\text{rank}(A_1) = 1$

$$A_2 = \sigma_2 \underset{\sim}{u}_2 \underset{\sim}{v}_2^T + \sigma_1 \underset{\sim}{u}_1 \underset{\sim}{v}_1^T$$

$\text{rank}(A_2) = 2$

General

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

$\text{rank}(A_k) = k$

Rank of a matrix

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

$$\underbrace{\boldsymbol{A}}_{m\times n} = \underbrace{\boldsymbol{U_R}}_{m\times k}\underbrace{\boldsymbol{\Sigma_R}}_{k\times k}\underbrace{\boldsymbol{V_R}^{\boldsymbol{T}}}_{k\times n}$$

$k = \min(m,n)$

$$\underset{=}{A} = \sum_{i=1}^{k} \sigma_i \underset{\sim}{u_i} \underset{\sim}{v_i}^T$$

$$\underbrace{\boxed{\Sigma_R}}_{\boxed{k\times k}} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_k \end{bmatrix}$$

if $\underline{\underline{\sigma_i \neq 0}}$ $\forall i \Rightarrow \text{rank}(A) = k$

Full rank matrix

In general

$\text{rank}(A) = r \qquad r < k$

Matrix rank deficient

$$\Sigma_R = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}_{k\times k}$$

r is the # of non-zero singular values!

Rank of a matrix

- The rank of **A** equals the number of non-zero singular values which is the same as the number of non-zero diagonal elements in $\boldsymbol{\Sigma}$.
- 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 **A**.
- The left-singular vectors (columns of $\boldsymbol{U}$) corresponding to the non-zero singular values of **A** span the range of **A**.

2) Pseudo-inverse

$A = U \Sigma V^T$

$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$

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

Σ^{-1}

$\text{rank}(A) = r$

- **How to fix it:** Define the Pseudo Inverse

$A^{-1} = A^T$ orth.

- **Pseudo-Inverse of a diagonal matrix**:

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

$\Sigma^+ = \begin{bmatrix} 1/\sigma_1 & & & & & & \\ & 1/\sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & 1/\sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$

- **Pseudo-Inverse of a matrix *A***:

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

side note

$A = U \Sigma V^T$ (but A is invertible)

$A^{-1} = (U \Sigma V^T)^{-1} = (V^T)^{-1} \Sigma^{-1} U^{-1}$

$= V \Sigma^{-1} U^T$

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

3) Matrix norms

$(Ux)^T Ux = x^T \underbrace{U^T U}_{I} x = x^T x$

The Euclidean norm of an orthogonal matrix is equal to 1

$p=2$

$$\|\boldsymbol{U}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|\underbrace{\boldsymbol{U}\boldsymbol{x}}_{y}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \sqrt{(\boldsymbol{U}\boldsymbol{x})^T(\boldsymbol{U}\boldsymbol{x})} = \max_{\|\boldsymbol{x}\|_2=1} \sqrt{\boldsymbol{x}^T\boldsymbol{x}} = \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{x}\|_2 = 1$$

The Euclidean norm of a matrix is given by the largest singular value

$A = U\Sigma V^T$ $\qquad \|U\|_2 = 1$

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|\underbrace{\boldsymbol{A}\boldsymbol{x}}_{y}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \left\|\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^T\boldsymbol{x}\right\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \left\|\boldsymbol{\Sigma}\,\boldsymbol{V}^T\boldsymbol{x}\right\|_2$$

$$= \max_{\|\boldsymbol{V}^T\boldsymbol{x}\|_2=1} \left\|\boldsymbol{\Sigma}\,\boxed{\boldsymbol{V}^T\boldsymbol{x}}\right\|_2 = \boxed{\max_{\|\boldsymbol{y}\|_2=1} \|\boldsymbol{\Sigma}\,\boldsymbol{y}\|_2}$$

largest diagonal entry value of Σ

$\|V^T\| = 1$

Where we used the fact that $\|\boldsymbol{U}\|_2 = 1$, $\|\boldsymbol{V}\|_2 = 1$. Since $\boldsymbol{\Sigma}$ is diagonal we get:

Σ is a diagonal

$$\|\boldsymbol{A}\|_2 = \max(\sigma_i) = \sigma_{max}$$

σ_{max} is the largest singular value

$$\boxed{\|A\|_2 = \max \sigma_i} = \sigma_{max}$$

4) Norm for the inverse of a matrix

$\|A\|_2 = \sigma_{max} = \max \sigma_i$

The Euclidean norm of the inverse of a square-matrix is given by:

Assume here $\boldsymbol{A}$ is full rank, so that $\boldsymbol{A}^{-1}$ exists

$$\|\boldsymbol{A}^{-1}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \left\|(\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^T)^{-1}\boldsymbol{x}\right\|_2$$

$A = U\Sigma V^T$ $\quad$ $A^{-1} = (U\Sigma V^T)^{-1}$

$= V\Sigma^{-1}U^T$

$$\|\boldsymbol{A}^{-1}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \left\|\boldsymbol{V}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T\boldsymbol{x}\right\|_2$$

$$\Sigma = \begin{bmatrix} \sigma_{max} & & \\ & \ddots & \\ & & \sigma_{min} \end{bmatrix}$$

Since $\|\boldsymbol{U}\|_2 = 1$, $\|\boldsymbol{V}\|_2 = 1$ and $\boldsymbol{\Sigma}$ is diagonal then

$$\boxed{\|\boldsymbol{A}^{-1}\|_2 = \frac{1}{\sigma_{min}}}$$

σ_{min} is the smallest singular value

5) Norm of the pseudo-inverse matrix

The norm of the pseudo-inverse of a $\boxed{m \times n}$ matrix is:

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

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

$$\Sigma = \begin{bmatrix} \sigma_{max} & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \;\; 0 \end{bmatrix} \qquad \sigma_{min}$$

$$\boxed{\|\boldsymbol{A}^{+}\|_2 = \frac{1}{\sigma_r}}$$

where σ_r is the smallest **non-zero** singular value. This is valid for any matrix, regardless of the shape or rank.

Note that for a full rank square matrix, $\|\boldsymbol{A}^{+}\|_2$ is the same as $\|\boldsymbol{A}^{-1}\|_2$. $= \dfrac{1}{\sigma_{min}}$

Zero matrix: If $\boldsymbol{A}$ is a zero matrix, then $\boldsymbol{A}^{+}$ is also the zero matrix, and $\|\boldsymbol{A}^{+}\|_2 = 0$

6) Condition number of a matrix

The condition number of a matrix is given by

$$cond_2(\boldsymbol{A}) = \|\boldsymbol{A}\|_2 \, \|\boldsymbol{A}^+\|_2$$

If the matrix is full rank: $rank(\boldsymbol{A}) = min(m, n)$

$$\boxed{cond_2(\boldsymbol{A}) = \frac{\sigma_{max}}{\sigma_{min}}} = \|A\|_2 \|A^{-1}\|_2$$

where σ_{max} is the largest singular value and σ_{min} is the smallest singular value

If the matrix is rank deficient: $rank(\boldsymbol{A}) < min(m, n)$ = r

$$\boxed{cond_2(\boldsymbol{A}) = \infty}$$ set

7) Low-Rank Approximation

We will again use the SVD to write the matrix A as a sum of outer products (of left and right singular vectors) – here for $m > n$ without loss of generality:

$$\underline{A_{approx}}$$

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

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

$$\underline{\underline{A}} \boxed{= \sigma_1 \boldsymbol{u}_1 \mathbf{v}_1^T + \sigma_2 \boldsymbol{u}_2 \mathbf{v}_2^T + \cdots + \sigma_n \boldsymbol{u}_n \mathbf{v}_n^T}$$

rank(A) = n
Full rank matrix

7) Low-Rank Approximation (cont.)

The best **rank-*k*** approximation for a $m \times n$ matrix $\boldsymbol{A}$, (where $k \le min(m,n)$) is the one that minimizes the following problem:

$\|B\|_2$ = largest sing. value σ_{max}

$$\min_{A_k} \|A - A_k\| \quad \text{such that} \quad \boxed{\text{rank}(A_k) \le k.}$$

→ minimizing error

rank(A) = n = 10

$A_k \to A$ rank$(A_k) \le 3$

When using the induced 2-norm, the best **rank-*k*** approximation is given by:

$$\longrightarrow \boldsymbol{A}_k = \sigma_1 \boldsymbol{u}_1 \mathbf{v}_1^T + \sigma_2 \boldsymbol{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \boldsymbol{u}_k \mathbf{v}_k^T$$

A_3

$$\sigma_1 \ge \sigma_2 \ge \sigma_3 \ldots \ge 0$$

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

$$\boxed{A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T}$$

$k < n$

Note that $rank(\boldsymbol{A}) = n$ and $rank(\boldsymbol{A}_k) = k$ and the norm of the difference between the matrix and its approximation is:

$$\|A - A_k\|_2 = \|\sigma_{k+1} u_{k+1} v_{k+1}^T + \sigma_{k+2} u_{k+2} v_{k+2}^T + \cdots + \sigma_n u_n v_n^T\|_2 = \sigma_{k+1}$$

$$\boxed{\|A - A_k\| = \sigma_{k+1}}$$

Example: Image compression

$A_{500\times 1417} = U \Sigma V^T$

Σ_{500}

$k = \min(m,n) = 500$

1417

500

$\sum_{i=1}^{50} \sigma_i u_i v_i^T$

$\sum_{i=1}^{500} \sigma_i u_i v_i^T$

Image using rank-50 approximation

8) Using SVD to solve square system of linear equations

If $\boldsymbol{A}$ is a $n\times n$ square matrix and we want to solve $\boxed{\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}}$ we can use the SVD for $\boldsymbol{A}$ such that

$A\tilde{x} \cong b$

① $PLU = A$

② → For/Back $\tilde{x}$

① $A = U\Sigma V^T \longrightarrow \boxed{O(n^3)}$

$$Ax = b \longrightarrow U\Sigma V^T x = b$$

$$\Sigma \underbrace{V^T x}_{y} = U^T b \quad (U^{-1} = U^T)$$

② $\Sigma y = \underbrace{U^T b} \longrightarrow$ easy! Solve for y $\quad O(n)$

③ $V^T x = y \longrightarrow x = Vy \longrightarrow$ matrix vector mult. $O(n^2)$

$(V^T = V^{-1})$