Singular Value Decomposition (applications)

1) Determining the rank of a matrix

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

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

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

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

$$\underbrace{\underset{\sim}{u}}_{n\times 1}\; \underbrace{\underset{\sim}{v}^T}_{1\times n} \Rightarrow n\times n$$

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

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

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

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

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

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:

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

A: $m\times n$, U_R: $m\times k$, Σ_R: $k\times k$, V_R^T: $k\times n$

$k = \min(m, n)$

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

$$\boxed{\Sigma_R} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_k \end{bmatrix}$$

$\boxed{k\times k}$

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

Full rank matrix

In general

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

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

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

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

$$= \max_{\|\boxed{\boldsymbol{V}^T\boldsymbol{x}}\|_2=1} \left\|\boldsymbol{\Sigma}\,\boxed{\boldsymbol{V}^T\boldsymbol{x}}\right\|_2 = \boxed{\boxed{\max}_{\|\boldsymbol{y}\|_2=1} \|\boldsymbol{\Sigma}\,\boldsymbol{y}\|_2} \longrightarrow \text{largest diagonal entry value of } \Sigma$$

$\|V^T\| = I$ $\quad y \quad y$

Σ is a diagonal

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

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

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

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

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

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

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:

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

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

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

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

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

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: $rank(\boldsymbol{A}) < min(m, n) = r$

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

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:

Aapprox

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

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

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

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}

$$\begin{cases} \min\limits_{A_k} \; \|A - A_k\| \\ \text{such that } \; \boxed{\text{rank}(A_k) \le k.} \end{cases}$$

→ minimizing error

rank(A) = n = 10

$A_k \to A$ rank(A) ≤ 3

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

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

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

A_3

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

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

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

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

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

$Ax \cong b$

① $PLU = A$

② → For / Back x

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

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

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

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

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

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