THE SINGULAR VALUE DECOMPOSITION

- The SVD – existence - properties.
- Pseudo-inverses and the SVD
- Use of SVD for least-squares problems
- Applications of the SVD

The Singular Value Decomposition (SVD)

Theorem For any matrix $A \in \mathbb{R}^{m\times n}$ there exist unitary matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ such that

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

where Σ is a diagonal matrix with entries $\sigma_{ii} \geq 0$.

$$\sigma_{11} \geq \sigma_{22} \geq \cdots \sigma_{pp} \geq 0 \text{ with } p = \min(n, m)$$

- The σ_{ii}'s are the singular values. Notation change $\sigma_{ii} \longrightarrow \sigma_i$

Proof: Let $\sigma_1 = \|A\|_2 = \max_{x,\|x\|_2=1} \|Ax\|_2$. There exists a pair of unit vectors v_1, u_1 such that

$$Av_1 = \sigma_1 u_1$$

➤ Complete v_1 into an orthonormal basis of $\mathbb{R}^n$

$$V \equiv [v_1, V_2] = n \times n \text{ unitary}$$

➤ Complete u_1 into an orthonormal basis of $\mathbb{R}^m$

$$U \equiv [u_1, U_2] = m \times m \text{ unitary}$$

✍ Define U, V as single Householder reflectors.

➤ Then, it is easy to show that

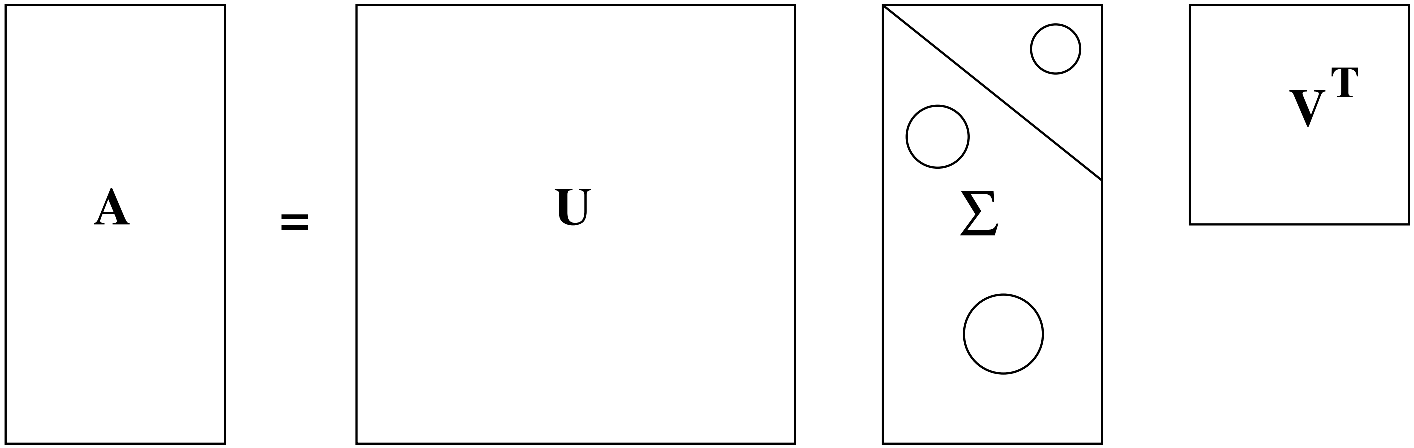
$$AV = U \times \begin{pmatrix} \sigma_1 & w^T \\ 0 & B \end{pmatrix} \rightarrow U^T AV = \begin{pmatrix} \sigma_1 & w^T \\ 0 & B \end{pmatrix} \equiv A_1$$

- Observe that

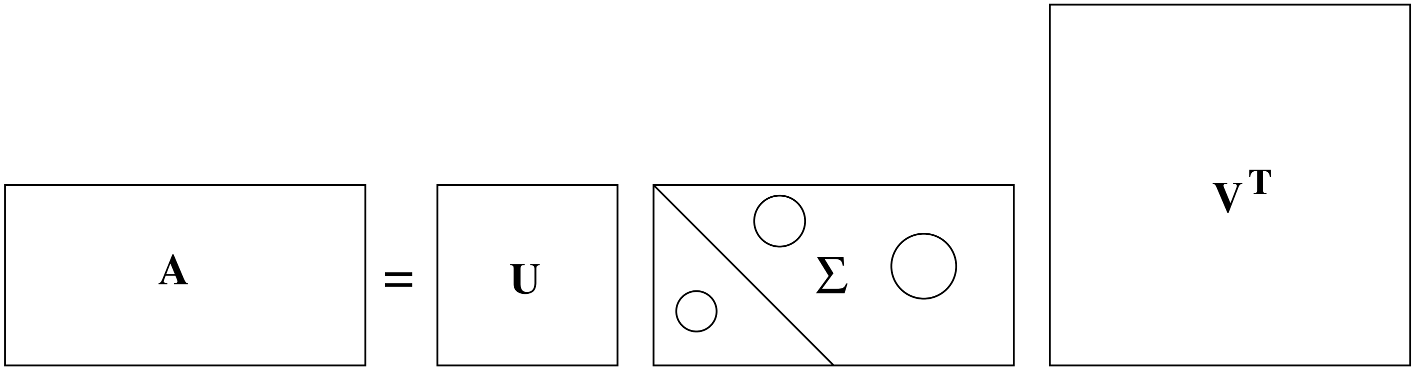
$$\left\|A_1\begin{pmatrix}\sigma_1\\ w\end{pmatrix}\right\|_2 \geq \sigma_1^2 + \|w\|^2 = \sqrt{\sigma_1^2 + \|w\|^2}\left\|\begin{pmatrix}\sigma_1\\ w\end{pmatrix}\right\|_2$$

- This shows that w must be zero [why?]
- Complete the proof by an induction argument. ■

Case 1:



Case 2:



The "thin" SVD

➤ Consider the Case-1. It can be rewritten as

$$A = [U_1 U_2] \begin{pmatrix} \Sigma_1 \\ 0 \end{pmatrix} V^T$$

Which gives:

$$A = U_1 \Sigma_1 V^T$$

where U_1 is $m \times n$ (same shape as A), and Σ_1 and V are $n \times n$

➤ Referred to as the "thin" SVD. Important in practice.

✍ How can you obtain the thin SVD from the QR factorization of A and the SVD of an $n \times n$ matrix?

A few properties. Assume that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \text{ and } \sigma_{r+1} = \cdots = \sigma_p = 0$$

Then:

- $\mathbf{rank}(A) = r$ = number of nonzero singular values.
- $\mathbf{Ran}(A) = \mathbf{span}\{u_1, u_2, \ldots, u_r\}$
- $\mathbf{Null}(A^T) = \mathbf{span}\{u_{r+1}, u_{r+2}, \ldots, u_m\}$
- $\mathbf{Ran}(A^T) = \mathbf{span}\{v_1, v_2, \ldots, v_r\}$
- $\mathbf{Null}(A) = \mathbf{span}\{v_{r+1}, v_{r+2}, \ldots, v_n\}$

Properties of the SVD (continued)

- The matrix $\boldsymbol{A}$ admits the SVD expansion:

$$\boldsymbol{A} = \sum_{i=1}^{r} \boldsymbol{\sigma}_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

- $\|\boldsymbol{A}\|_2 = \boldsymbol{\sigma}_1 =$ largest singular value
- $\|\boldsymbol{A}\|_F = \left(\sum_{i=1}^{r} \boldsymbol{\sigma}_i^2\right)^{1/2}$
- When $\boldsymbol{A}$ is an $\boldsymbol{n} \times \boldsymbol{n}$ nonsingular matrix then $\|\boldsymbol{A}^{-1}\|_2 = 1/\boldsymbol{\sigma}_n$

Theorem Let $k < r$ and

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

then

$$\min_{rank(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

Proof: First: $\|A - B\|_2 \geq \sigma_{k+1}$, for any rank-$k$ matrix B.

Consider $\mathcal{X} = \text{span}\{v_1, v_2, \cdots, v_{k+1}\}$. Note:

$$dim(Null(B)) = n - k \rightarrow Null(B) \cap \mathcal{X} \neq \{0\}$$

[Why?]

Let $x_0 \in Null(B) \cap \mathcal{X}, \; x_0 \neq 0$. Write $x_0 = Vy$. Then

$$\|(A - B)x_0\|_2 = \|Ax_0\|_2 = \|U\Sigma V^T V y\|_2 = \|\Sigma y\|_2$$

But $\|\Sigma y\|_2 \geq \sigma_{k+1}\|x_0\|_2$ (Show this). $\rightarrow \|A - B\|_2 \geq \sigma_{k+1}$

Second: take $B = A_k$. Achieves the min. ■

Right and Left Singular vectors:

$$Av_i = \sigma_i u_i$$
$$A^T u_j = \sigma_j v_j$$

- Consequence $A^T A v_i = \sigma_i^2 v_i$ and $A A^T u_i = \sigma_i^2 u_i$
- Right singular vectors (v_i's) are eigenvectors of $A^T A$
- Left singular vectors (u_i's) are eigenvectors of $A A^T$
- Possible to get the SVD from eigenvectors of AA^T and $A^T A$ – but: difficulties due to non-uniqueness of the SVD

Define the $r \times r$ matrix

$$\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_r)$$

- Let $A \in \mathbb{R}^{m\times n}$ and consider $A^T A$ ($\in \mathbb{R}^{n\times n}$):

$$A^T A = V \Sigma^T \Sigma V^T \;\rightarrow\; A^T A = V \underbrace{\begin{pmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix}}_{n\times n} V^T$$

- This gives the spectral decomposition of $A^T A$.

- Similarly, U gives the eigenvectors of AA^T.

$$AA^T = U \underbrace{\begin{pmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix}}_{m \times m} U^T$$

Important:

$A^TA = VD_1V^T$ and $AA^T = UD_2U^T$ give the SVD factors U, V up to signs!

Pseudo-inverse of an arbitrary matrix

- Let $A = U\Sigma V^T$ which we rewrite as

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 \Sigma_1 V_1^T$$

Then the pseudo inverse of A is

$$\boxed{A^\dagger = V_1 \Sigma_1^{-1} U_1^T = \sum_{j=1}^{r} \frac{1}{\sigma_j} v_j u_j^T}$$

- The pseudo-inverse of A is the mapping from a vector b to the solution $\min_x \|Ax - b\|_2^2$ that has minimal norm (to be shown)

- In the full-rank overdetermined case, the normal equations yield $x = \underbrace{(A^T A)^{-1} A^T}_{A^\dagger} b$

Least-squares problem via the SVD

Pb: $\min \|b - Ax\|_2$ in general case. Consider SVD of A:

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = \sum_{i=1}^{r} \sigma_i v_i u_i^T$$

Then left multiply by U^T to get

$$\|Ax - b\|_2^2 = \left\| \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix} b \right\|_2^2$$

$$\text{with} \quad \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} x$$

✍ What are **all** least-squares solutions to the system? Among these which one has minimum norm?

Answer: From above, must have $y_1 = \Sigma_1^{-1} U_1^T b$ and $y_2 =$ anything (free).

➤ Recall that $x = Vy$ and write

$$\begin{aligned} x &= [V_1, V_2] \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = V_1 y_1 + V_2 y_2 \\ &= V_1 \Sigma_1^{-1} U_1^T b + V_2 y_2 \\ &= A^\dagger b + V_2 y_2 \end{aligned}$$

➤ Note: $A^\dagger b \ \in \ \mathrm{Ran}(A)$ and $V_2 y_2 \ \in \ \mathrm{Null}(A)$.

➤ Therefore: least-squares solutions are of the form $A^\dagger b + w$ where $w \ \in \ \mathrm{Null}(A)$.

➤ Smallest norm when $y_2 = 0$.

➤ Minimum norm solution to $\min_x \|Ax - b\|_2^2$ satisfies $\Sigma_1 y_1 = U_1^T b$, $y_2 = 0$. It is:

$$x_{LS} = V_1 \Sigma_1^{-1} U_1^T b = A^\dagger b$$

✍ If $A \in \mathbb{R}^{m\times n}$ what are the dimensions of $A^\dagger$?, $A^\dagger A$?, $AA^\dagger$?

✍ Show that $A^\dagger A$ is an orthogonal projector. What are its range and null-space?

✍ Same questions for $AA^\dagger$.

Moore-Penrose Inverse

The pseudo-inverse of A is given by

$$A^{\dagger} = V \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T = \sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i}$$

Moore-Penrose conditions:

The pseudo inverse of a matrix is uniquely determined by these four conditions:

(1) $AXA = A$ (2) $XAX = X$
(3) $(AX)^H = AX$ (4) $(XA)^H = XA$

➤ In the full-rank overdetermined case, $A^{\dagger} = (A^T A)^{-1} A^T$

Least-squares problems and the SVD

➤ SVD can give much information about solving overdetermined and underdetermined linear systems.

Let A be an $m \times n$ matrix and $A = U\Sigma V^T$ its SVD with $r = \mathbf{rank}(A)$, $V = [v_1, \ldots, v_n]$ $U = [u_1, \ldots, u_m]$. Then

$$x_{LS} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

minimizes $\|b - Ax\|_2$ and has the smallest 2-norm among all possible minimizers. In addition,

$$\rho_{LS} \equiv \|b - Ax_{LS}\|_2 = \|z\|_2 \text{ with } z = [u_{r+1}, \ldots, u_m]^T b$$

Least-squares problems and pseudo-inverses

➤ A restatement of the first part of the previous result:

Consider the general linear least-squares problem

$$\min_{x \in S} \|x\|_2, \quad S = \{x \in \mathbb{R}^n \mid \|b - Ax\|_2 \min\}.$$

This problem always has a unique solution given by

$$x = A^{\dagger} b$$

✍ Consider the matrix:

$$A = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 0 & -2 & 1 \end{pmatrix}$$

- Compute the singular value decomposition of A
- Find the matrix B of rank 1 which is the closest to the above matrix in the 2-norm sense.
- What is the pseudo-inverse of A?
- What is the pseudo-inverse of B?
- Find the vector x of smallest norm which minimizes $\|b - Ax\|_2$ with $b = (1, 1)^T$
- Find the vector x of smallest norm which minimizes $\|b - Bx\|_2$ with $b = (1, 1)^T$

Ill-conditioned systems and the SVD

- Let A be $m \times m$ and $A = U\Sigma V^T$ its SVD
- Solution of $Ax = b$ is $x = A^{-1}b = \sum_{i=1}^{m} \frac{u_i^T b}{\sigma_i} v_i$
- When A is very ill-conditioned, it has many small singular values. The division by these small σ_i's will amplify any noise in the data. If $\tilde{b} = b + \epsilon$ then

$$A^{-1}\tilde{b} = \sum_{i=1}^{m} \frac{u_i^T b}{\sigma_i} v_i + \underbrace{\sum_{i=1}^{m} \frac{u_i^T \epsilon}{\sigma_i} v_i}_{Error}$$

- Result: solution could be completely meaningless.

Remedy: SVD regularization

Truncate the SVD by only keeping the $\sigma_i's$ that are $\geq \tau$, where τ is a threshold

➤ Gives the Truncated SVD solution (TSVD solution:)

$$x_{TSVD} = \sum_{\sigma_i \geq \tau} \frac{u_i^T b}{\sigma_i} v_i$$

➤ Many applications [e.g., Image and signal processing,..]

Numerical rank and the SVD

➤ Assuming the original matrix A is exactly of rank k the computed SVD of A will be the SVD of a nearby matrix $A + E$ – Can show: $|\hat{\sigma}_i - \sigma_i| \leq \alpha \; \sigma_1 \underline{\mathbf{u}}$

➤ Result: zero singular values will yield small computed singular values and r larger sing. values.

➤ Reverse problem: *numerical rank* – The ϵ-rank of A :

$$r_\epsilon = \min\{rank(B) : B \in \mathbb{R}^{m\times n}, \|A - B\|_2 \leq \epsilon\},$$

✍ Show that r_ϵ equals the number sing. values that are $>\epsilon$

✍ Show: r_ϵ equals the number of columns of A that are linearly independent for any perturbation of A with norm $\leq \epsilon$.

➤ Practical problem : How to set ϵ?

Pseudo inverses of full-rank matrices

Case 1: $m > n$ Then $A^{\dagger} = (A^T A)^{-1} A^T$

- Thin SVD is $A = U_1 \Sigma_1 V_1^T$ and V_1, Σ_1 are $n \times n$. Then:

$$\begin{aligned} (A^T A)^{-1} A^T &= (V_1 \Sigma_1^2 V_1^T)^{-1} V_1 \Sigma_1 U_1^T \\ &= V_1 \Sigma_1^{-2} V_1^T V_1 \Sigma_1 U_1^T \\ &= V_1 \Sigma_1^{-1} U_1^T \\ &= A^{\dagger} \end{aligned}$$

Example: Pseudo-inverse of $\begin{pmatrix} 0 & 1 \\ 1 & 2 \\ 2 & -1 \\ 0 & 1 \end{pmatrix}$ is?

Case 2: $m < n$ Then $A^\dagger = A^T(AA^T)^{-1}$

➤ Thin SVD is $A = U_1\Sigma_1 V_1^T$. Now U_1, Σ_1 are $m \times m$ and:

$$\begin{aligned} A^T(AA^T)^{-1} &= V_1\Sigma_1 U_1^T[U_1\Sigma_1^2 U_1^T]^{-1} \\ &= V_1\Sigma_1 U_1^T U_1 \Sigma_1^{-2} U_1^T \\ &= V_1\Sigma_1\Sigma_1^{-2} U_1^T \\ &= V_1\Sigma_1^{-1} U_1^T \\ &= A^\dagger \end{aligned}$$

Example: Pseudo-inverse of $\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & -1 & 1 \end{pmatrix}$ is?

➤ Mnemonic: The pseudo inverse of A is A^T completed by the inverse of the smallest of $(A^TA)^{-1}$ or $(AA^T)^{-1}$ where it fits (i.e., left or right)