

I.8] Singular Values and Singular Vectors in the SVD

$$A \in \mathbb{R}^{m \times n} \quad r = \text{rank}(A)$$

$$AV_1 = \sigma_1 U_1, \dots, AV_r = \sigma_r U_r$$

Singular
values

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$$

$$AV_{r+1} = 0, \dots, AV_n = 0$$

Singular
Vectors

$$V_1, \dots, V_n \in \mathbb{R}^n \text{ orthonormal}$$

$U_1, \dots, U_m \in \mathbb{R}^m$ orthonormal

$\{U_1, \dots, U_r\}$ is a basis for $\text{col}(A)$

$\{V_{r+1}, \dots, V_n\}$ is a basis for $\text{null}(A)$

Then $V = [V_1 \cdots V_n]$ and $U = [U_1 \cdots U_m]$

are orthogonal matrices ($V^T = V^{-1}$ and $U^T = U^{-1}$).

$$AV = U\Sigma$$

$$A \begin{bmatrix} V_1 & \cdots & V_r & \cdots & V_n \end{bmatrix} = \begin{bmatrix} U_1 & \cdots & U_r & \cdots & U_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ \hline & & & 0 & \\ & & & & 0 \end{bmatrix}$$

m x n n x n m x m m x n

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

U_1, \dots, U_m are the left singular vectors

V_1, \dots, V_n are the right singular vectors

$$A = \sigma_1 U_1 V_1^T + \cdots + \sigma_r U_r V_r^T$$

Reduced Form of the SVD

$$V_r = [V_1 \cdots V_r] \quad n \times r \text{ orthonormal cols}$$

$$U_r = [U_1 \cdots U_r] \quad m \times r \text{ orthonormal cols}$$

$$\Sigma_r = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} \quad r \times r \text{ positive diagonal entries}$$

$$AV_r = U_r \Sigma_r$$

Even though $V_r V_r^T \neq I$, we still have

$$A = U_r \Sigma_r V_r^T$$

Since $A = \sigma_1 U_1 V_1^T + \dots + \sigma_r U_r V_r^T$.

Note that Σ_r is invertible:

$$\Sigma_r^{-1} = \begin{bmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \end{bmatrix}.$$

The pseudo-inverse of A is

$$A^+ = V_r \Sigma_r^{-1} U_r^T$$

and satisfies:

$$(1) \quad AA^+A = A$$

$$(3) \quad (A^+A)^T = A^+A$$

$$(2) \quad A^+AA^+ = A^+$$

$$(4) \quad (AA^+)^T = AA^+$$

These are called the Moore-Penrose conditions and are uniquely satisfied by A^+ .

Pf of (1):

$$\text{Note: } A^T A = V_r \sum_r^2 V_r^T \Rightarrow A^T A V_i = \sigma_i^2 V_i$$

$$A A^T = U_r \sum_r^2 U_r^T \Rightarrow A A^T U_i = \sigma_i^2 U_i$$

$\therefore \sigma_1^2, \dots, \sigma_r^2$ are the nonzero eigenvalues of $A^T A$ and $A A^T$; V_1, \dots, V_r are eigenvectors of $A^T A$ and U_1, \dots, U_r are eigenvectors of $A A^T$.

- If A has linearly independent columns, then $A^T A$ is invertible and $A^+ = (A^T A)^{-1} A^T$.
- If A has linearly independent rows, then $A A^T$ is invertible and $A^+ = A^T (A A^T)^{-1}$.

- If $b \in \mathbb{R}^m$, then $x = A^+b$ is the minimizer of $\|Ax - b\|_2$ having the smallest $\|x\|_2$.
-

Best Rank k Approximation of A

Let $A_k = \sigma_1 u_1 v_1^T + \dots + \sigma_k u_k v_k^T$.

The Eckart-Young Theorem states that A_k is the best rank k approximation of A :

[If $\text{rank}(B) = k$, then $\|A - A_k\| \leq \|A - B\|$.]

The norm here can be the Frobenius norm,

$$\|X\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n x_{ij}^2 \right)^{\frac{1}{2}} = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}.$$

or the spectral norm,

$$\|X\|_2 = \max_{v \neq 0} \frac{\|Xv\|_2}{\|v\|_2} = \sigma_1.$$

Principal component analysis (PCA) is based on the Eckart-Young Theorem and is used to perform dimensionality reduction on a data set.

A nice way to visualize Eckart-Young is to use SVD for image compression.

(numerical demonstration)

SVD and Symmetric Matrices

- ① We have already seen that $\sigma_1^2, \dots, \sigma_r^2$ are the nonzero eigenvalues of the symmetric positive semidefinite matrices $A^T A$ and $A A^T$. However, the standard numerical method for computing the SVD is based on another symmetric matrix.

First note that

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

Thus A^T has the same singular values as A , but left singular vectors become right singular vectors and vice-versa.

$$A^T u_1 = \sigma_1 v_1, \dots, A^T u_r = \sigma_r v_r$$

$$A^T u_{r+1} = 0, \dots, A^T u_m = 0$$

Thus the $(m+n) \times (m+n)$ symmetric matrix

$$\beta = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$$

has eigenvectors

$$\begin{bmatrix} u_1 \\ v_1 \end{bmatrix}, \dots, \begin{bmatrix} u_r \\ v_r \end{bmatrix}, \begin{bmatrix} -u_1 \\ v_1 \end{bmatrix}, \dots, \begin{bmatrix} -u_r \\ v_r \end{bmatrix}$$

having eigenvalues $\sigma_1, \dots, \sigma_r, -\sigma_1, \dots, -\sigma_r$,
and eigenvectors

$$\begin{bmatrix} u_{r+1} \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} u_m \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ v_{r+1} \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ v_n \end{bmatrix}$$

having eigenvalue 0. This gives us

$$r + r + (m-r) + (n-r) = m+n$$

orthogonal eigenvectors, so we have
a complete set of eigenvectors
of B and the $m+n$ eigenvalues of
 B are

$$\underbrace{\sigma_1, \dots, \sigma_r}_r, \underbrace{-\sigma_1, \dots, -\sigma_r}_r, \underbrace{0, \dots, 0}_{(m-r)+(n-r)}$$

② If $A \in \mathbb{R}^{n \times n}$ is symmetric and positive semidefinite, then the eigen decomposition $A = Q \Lambda Q^T$ gives us the SVD of A :

$$U = Q, \Sigma = \Lambda, V = Q.$$

③ If $A \in \mathbb{R}^{n \times n}$ is symmetric but not positive semidefinite, then

$$A = Q \Lambda Q^T$$

and if $s = (\text{sign}(\lambda_1), \dots, \text{sign}(\lambda_n))$

(eg eigenvalues $-2, 0, 3 \Rightarrow s = (-1, 1, 1)$)

then the SVD of A is given by

$$U = Q, \Sigma = \Lambda \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{bmatrix} = |\Lambda|,$$

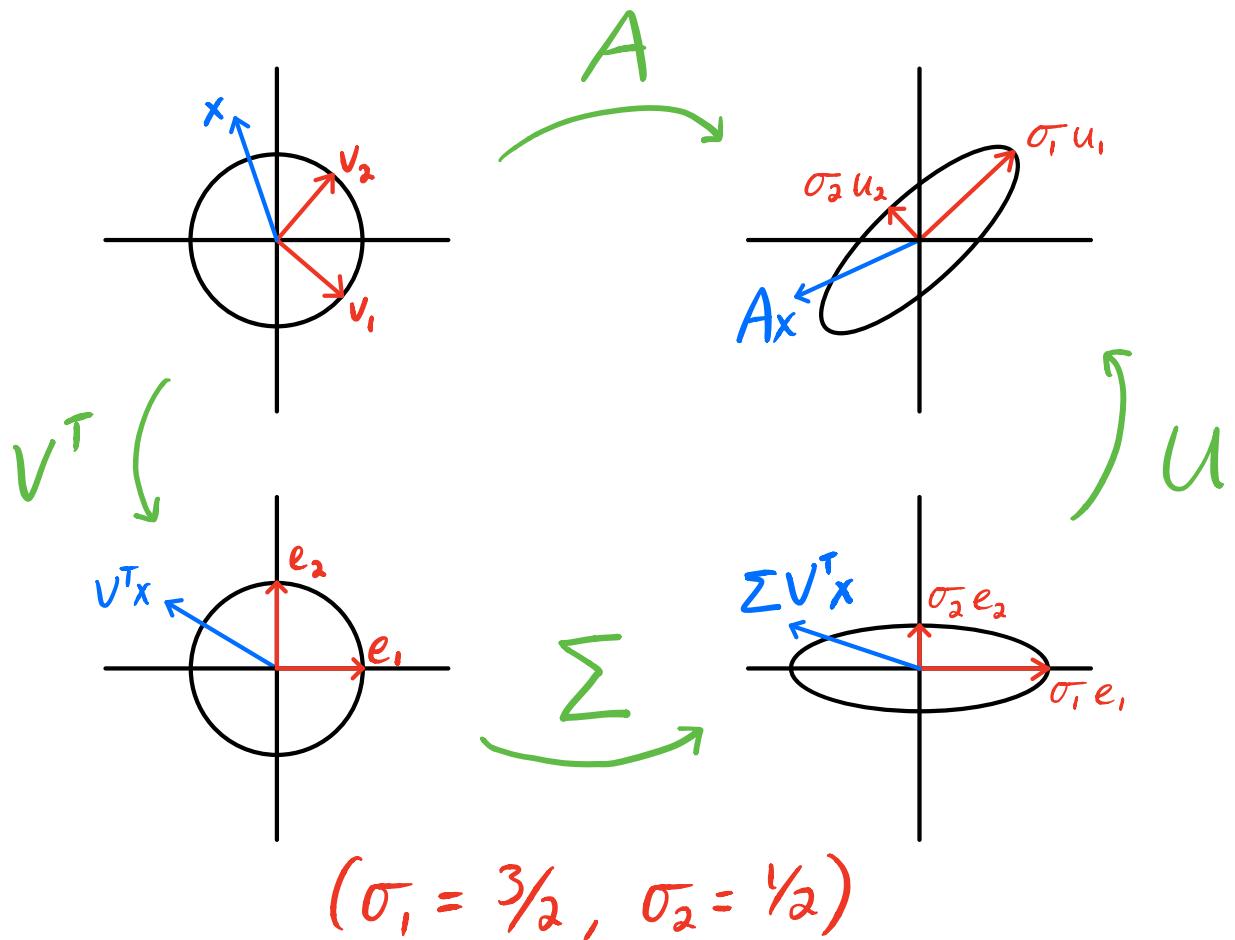
$$V = Q \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{bmatrix}.$$

The Geometry of the SVD

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

$$Ax = U \left(\sum (V^T_x) \right)$$

- ① rotation
- ② scaling
- ③ rotation



Finding Singular Vectors via Optimization

$$\max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \equiv \max_{\|x\|_2=1} \|Ax\|_2$$

We can find the optimal x^* by solving

$$\left[\max \|Ax\|_2^2 \text{ s.t. } \|x\|^2 = 1 \right].$$

We use Lagrange multipliers to solve constrained optimization problems.

The Lagrangian is

$$\begin{aligned} L(x, \lambda) &= \|Ax\|_2^2 + \lambda(1 - \|x\|_2^2) \\ &= (Ax)^T(Ax) + \lambda(1 - x^T x) \\ &= x^T A^T A x + \lambda(1 - x^T x) \end{aligned}$$

We want to find x^* and λ^* such that $\nabla L(x^*, \lambda^*) = 0$.

$$\left[\begin{array}{l} \nabla(x^T x) = 2x \\ \nabla(x^T A^T A x) = 2A^T A x \end{array} \right]$$

Thus,

$$\left[\begin{array}{l} \nabla_x L(x, \lambda) = 2A^T A x - 2\lambda x, \\ \nabla_\lambda L(x, \lambda) = 1 - x^T x. \end{array} \right]$$

$\therefore \nabla L(x^*, \lambda^*) = 0$ implies

$$A^T A x^* = \lambda^* x^*, \quad \|x^*\|_2^2 = 1.$$

This means that λ^* is an eigenvalue of $A^T A$ and x^* is the associated eigenvector.

Which unit eigenvector x of $A^T A$ maximizes $\|Ax\|_2^2 = x^T A^T A x = \lambda$?

It is the eigenvector x_1 corresponding to the largest eigenvalue λ_1 .

i.e. $x^* = x_1$ and $\lambda^* = \lambda_1$.

$$\therefore \max_{\|x\|_2=1} \|Ax\|_2 = \|Ax_1\|_2 = \sqrt{\lambda_1} = \sigma_1.$$

This proves that the spectral norm is

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1.$$
