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

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

$$\boxed{Av_1 = \sigma_1 u_1, \ \ldots, \ Av_r = \sigma_r u_r}$$

Singular values $\qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

$$\boxed{Av_{r+1} = 0, \ \ldots, \ Av_n = 0}$$

Singular vectors

$v_1, \ldots, v_n \in \mathbb{R}^n$ orthonormal

$u_1, \ldots, u_m \in \mathbb{R}^m$ orthonormal

$\{u_1, \ldots, u_r\}$ is a basis for col(A)
$\{v_{r+1}, \ldots, v_n\}$ is a basis for null(A)

Then $V = [v_1 \cdots v_n]$ and $U = [u_1 \cdots u_m]$

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

$$\boxed{AV = U\Sigma}$$

$$\underset{m\times n}{A}\ \underset{n\times n}{\begin{bmatrix} v_1 & \cdots & v_r & \cdots & v_n \end{bmatrix}} = \underset{m\times m}{\begin{bmatrix} u_1 & \cdots & u_r & \cdots & u_m \end{bmatrix}}\ \underset{m\times n}{\left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]}$$

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

$u_1, \dots, u_m$ are the <u>left</u> singular vectors

$v_1, \dots, v_n$ are the <u>right</u> singular vectors

$$\boxed{A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T}$$

<u>Reduced Form of the SVD</u>

$V_r = \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix}$ $n \times r$ orthonormal cols

$U_r = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}$ $m \times r$ orthonormal cols

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

$$\boxed{AV_r = U_r\Sigma_r}$$

$V_r^TV_r = I$

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

$$\boxed{A = U_r\Sigma_rV_r^T}$$

since $A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$.

Note that Σ_r is invertible:

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

The <u>pseudo-inverse</u> of A is

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

and satisfies:

(1) $AA^+A = A$

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

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

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

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

[<u>Pf of (1):</u>

$$AA^+A = (U_r\Sigma_r\cancel{V_r^T})(\cancel{V_r}\cancel{\Sigma_r^{-1}}\cancel{U_r^T})(\cancel{U_r}\cancel{\Sigma_r}V_r^T)$$
$$= U_r\Sigma_r V_r^T = A$$
]

<u>Note:</u> $A^TA = V_r\Sigma_r^2V_r^T \implies A^TAv_i = \sigma_i^2 v_i$

$AA^T = U_r\Sigma_r^2U_r^T \implies AA^Tu_i = \sigma_i^2 u_i$

$\therefore$ $\sigma_1^2, \ldots, \sigma_r^2$ are the <u>nonzero eigenvalues</u> of A^TA and AA^T; $v_1, \ldots, v_r$ are eigenvectors of A^TA and $u_1, \ldots, u_r$ are eigenvectors of AA^T.

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

- 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 $A_k = \sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$. rank$(A_k) = k$ $k \le r$

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

[If rank$(B) = k$, then $\|A - A_k\| \le \|A - B\|$.]

The norm here can be the Frobenius norm,

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

or the spectral norm, $\|Xv\|_2 \le \|X\|_2 \|v\|_2$

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

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 $\sigma_1^2, \ldots, \sigma_r^2$ are the nonzero eigenvalues of the symmetric positive semidefinite matrices $\underset{n \times n}{A^TA}$ and $\underset{m \times m}{AA^T}$. 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.

$$\boxed{A^T u_1 = \sigma_1 v_1, \ \ldots, \ A^T u_r = \sigma_r v_r}$$

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

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

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

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}\begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = \begin{bmatrix} Av_1 \\ A^T u_1 \end{bmatrix} = \sigma_1 \begin{bmatrix} u_1 \\ v_1 \end{bmatrix}$$

has eigenvectors

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

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 eigendecomposition $A = Q \Lambda Q^T$ gives us the SVD of A:

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

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

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

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

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

then the SVD of A is given by

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

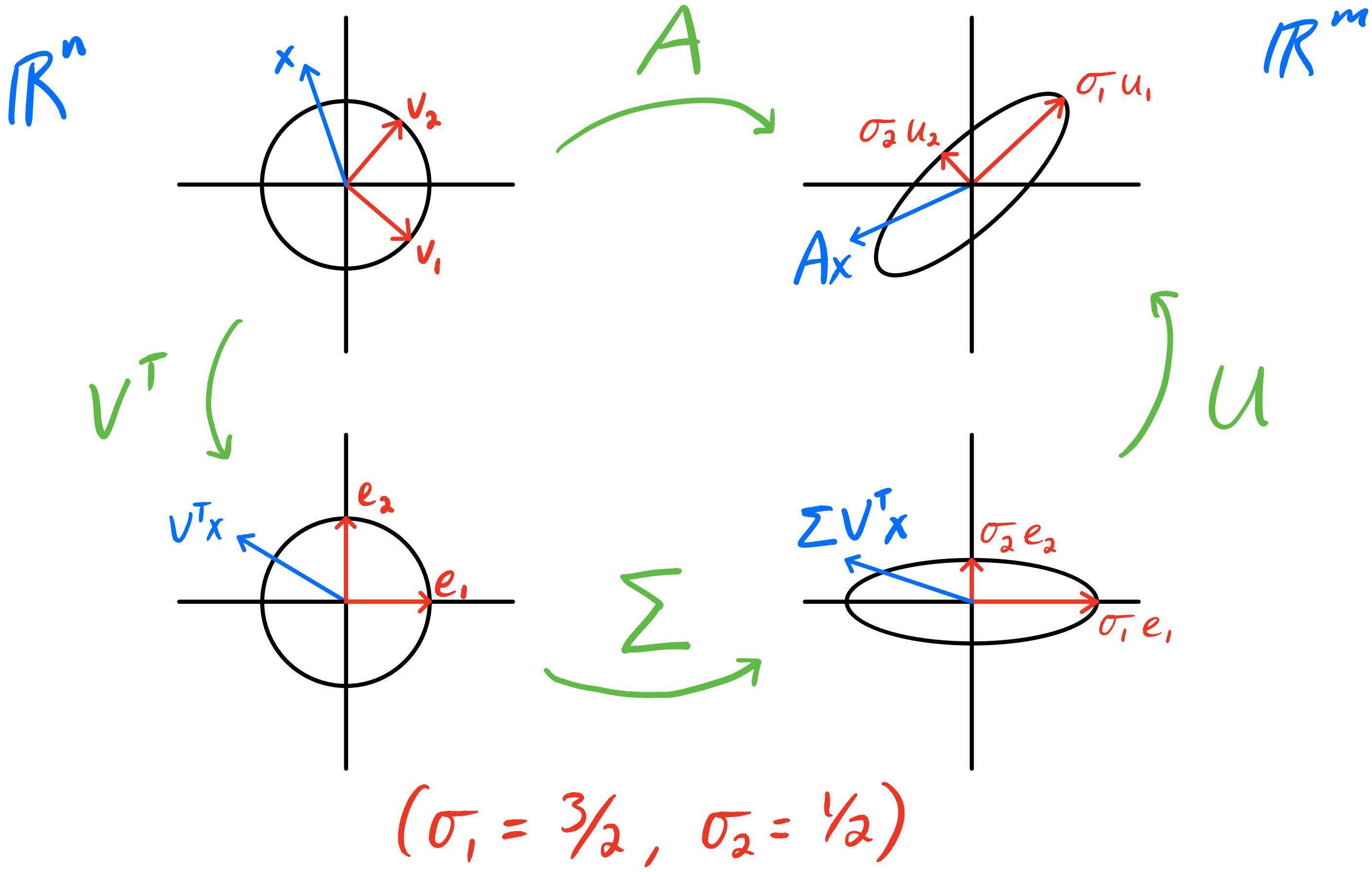
$$V = Q \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{bmatrix}.$$

The Geometry of the SVD

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

$$Ax = \underset{③}{U}\Big(\underset{②}{\Sigma}\big(\underset{①}{V^T}x\big)\Big)$$

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

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

We use <u>Lagrange multipliers</u> to solve constrained optimization problems.

The <u>Lagrangian</u> is

$$\mathcal{L}(x,\lambda) = \|Ax\|_2^2 + \lambda(1 - \|x\|_2^2)$$
$$= (Ax)^T(Ax) + \lambda(1 - x^Tx)$$
$$= x^TA^TAx + \lambda(1 - x^Tx)$$

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

$$\sum_{i=1}^{n} x_i^2$$

$$\begin{bmatrix} \nabla(x^Tx) = 2x \\ \nabla(x^TA^TAx) = 2A^TAx \end{bmatrix}$$

Thus,

$$\begin{bmatrix} \nabla_x\mathcal{L}(x,\lambda) = 2A^TAx - 2\lambda x, \\ \nabla_\lambda\mathcal{L}(x,\lambda) = 1 - x^Tx. \end{bmatrix}$$

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

$$\boxed{A^TAx^* = \lambda^*x^*, \quad \|x^*\|_2^2 = 1}.$$

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

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

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

<u>ie</u> $x^* = x_1$ and $\lambda^* = \lambda_1$.

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

This proves that the <u>spectral norm</u> is

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