* Formal Definition

Let \( A \in \mathbb{R}^{m \times n} \).

Then SVD of \( A \) is a factorization

\[
A = U \Sigma V^T
\]

where

- \( U \in \mathbb{R}^{m \times m} \) orthogonal
- \( \Sigma \in \mathbb{R}^{m \times n} \) diagonal
- \( V \in \mathbb{R}^{n \times n} \) orthogonal

\[
\text{diag}(\Sigma) = [\sigma_1, \sigma_2, \ldots, \sigma_p]^T
\]

\[
\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0.
\]

\[
p = \min(m, n)
\]

\[
\text{rank}(A) = r \leq p.
\]

\( A \) & \( \Sigma \) are the same shape.

Geometrically,

\[
\mathbb{R}^n \xrightarrow{\text{rotation or reflection in } \mathbb{R}^n} \mathbb{R}^n \xrightarrow{\Sigma} \mathbb{R}^m \xrightarrow{\text{rotation or reflection in } \mathbb{R}^m} \mathbb{R}^m
\]

So if we prove every \( A \in \mathbb{R}^{m \times n} \) has an SVD, then we shall have proved that \( A \) maps the unit sphere in \( \mathbb{R}^n \) to a hyperellipsoid in \( \mathbb{R}^m \).
**Existence & Uniqueness of SVD**

We can get peace of mind if we know that \( \exists ! \) SVD for any given matrix.

**Theorem** Every matrix \( A \in \mathbb{R}^{m \times n} \) has an SVD. Furthermore, the singular values \( \sigma_j \) are uniquely determined. If \( A \) is square and \( \sigma_j \)'s are distinct, then singular vectors \( \{ \mathbf{u}_j \}, \{ \mathbf{v}_j \} \) are uniquely determined up to signs (i.e., \( \pm 1 \) factor).

(Proof: Existence)

Let's check the largest action of \( A \) first, then do induction.

\[
\sigma_1 = \| A \|_2 = \sup_{\mathbf{v} \in \mathbb{S}} \| A \mathbf{v} \|_2
\]

This is often called the "compactness argument." Because we are dealing with vectors in \( \mathbb{R}^n \) (i.e., finite dimensional space), and \( \| A \cdot \|_2 \) is a continuous fcn, \( \exists \mathbf{u}_1 \in \mathbb{S} \subset \mathbb{R}^n \) s.t. \( \| A \mathbf{u}_1 \|_2 = \sigma_1 \) is attained.

Now set \( \tilde{\mathbf{u}}_1 = A \mathbf{u}_1 \in \mathbb{R}^m \), and consider orthogonal matrices \( V_1 = [ \mathbf{u}_1 \mathbf{u}_2 \ldots \mathbf{u}_n ] \in \mathbb{R}^{n \times n} \).
\[ U_1 = [u_1, u_2, \ldots, u_m] \in \mathbb{R}^{m \times m} \]

where \( u_1 = \frac{1}{\sigma_1} \tilde{u}_1 \).

Note \( \| u_1 \| = \frac{1}{\sigma_1} \| \tilde{u}_1 \| = \frac{1}{\sigma_1} \| A u_1 \| = \frac{1}{\sigma_1} \cdot \sigma_1 = 1 \quad \checkmark \)

Then, \( U_1^T A V_1 = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} A \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} \begin{bmatrix} A v_1 & \cdots & A v_n \end{bmatrix} \)

\[ u_1 = \sigma_1 u_1 \]

Let's call \( \Sigma_i \) \( 1 \times n \cdot 1 \)

where \( w^T = [u_1^T A v_2, \ldots, u_m^T A v_n] \in \mathbb{R}^T \)

\( B = \begin{bmatrix} u_2^T A v_2 & \cdots & u_n^T A v_n \\ \vdots & \ddots & \vdots \\ u_m^T A v_2 & \cdots & u_m^T A v_n \end{bmatrix} \in \mathbb{R}^{m-1 \times n-1} \)

\[ \| \begin{bmatrix} \sigma_i w^T \\ 0 & B \end{bmatrix} \| \geq \sigma_i^2 + w^T w \]

\[ \| \Sigma_i \| = \sqrt{\sigma_i^2 + \| w \|^2} \| \begin{bmatrix} \sigma_i \\ w \end{bmatrix} \| \]

\( \Rightarrow \| \Sigma_i \| \geq \sqrt{\sigma_i^2 + \| w \|^2} \quad \text{---} (1) \)

Since \( U_1, V_1 \) are orthogonal,

\( \| \Sigma_i \| = \| A \| = \sigma_i \quad \text{---} (2) \)
From ① & ②, we can conclude that \( \mathbf{w} = 0 \), i.e.,

\[
\mathbf{U}_1^T \mathbf{A} \mathbf{V}_1 = \begin{bmatrix}
\sigma_1 & 0 \\
0 & \mathbf{B}
\end{bmatrix}
\]

Hence if \( m = 1 \) or \( n = 1 \), we are done!

In general case, we can use the induction hypothesis:

Suppose an SVD exists for any \( m-1 \times n-1 \) matrix. Then the above matrix \( \mathbf{B} \) has its SVD: \( \mathbf{B} = \mathbf{U}_2 \Sigma_2 \mathbf{V}_2^T \)

Then \( \mathbf{A} = \mathbf{U}_1 \begin{bmatrix}
1 & 0 \\
0 & \mathbf{U}_2
\end{bmatrix} \begin{bmatrix}
\sigma_1 & 0 \\
0 & \Sigma_2
\end{bmatrix} \begin{bmatrix}
1 & 0 \\
0 & \mathbf{V}_2
\end{bmatrix}^T \mathbf{V}_1^T \)

\[
\mathbf{U} \quad \Sigma \quad \mathbf{V}^T
\]

This is an SVD of \( \mathbf{A} \)!

(Proof: Uniqueness)

Let \( \mathbf{u}_i \in S \subset \mathbb{R}^n \) s.t. \( \| \mathbf{A} \|_2 = \| \mathbf{u}_i \|_2 = \sigma_i \)

Suppose \( \exists \mathbf{w} \in S \), s.t., \( \mathbf{w} \perp \mathbf{u}_i \), \( \mathbf{w} \) is linearly independent from \( \mathbf{u}_i \), and \( \| \mathbf{A} \mathbf{w} \|_2 = \sigma_i \).

Let's define a unit vector \( \mathbf{u}_2 \in S \) by

\[
\mathbf{u}_2 := \frac{(I - \mathbf{P}_{\mathbf{u}_i})\mathbf{w}}{\| (I - \mathbf{P}_{\mathbf{u}_i})\mathbf{w} \|_2}
\]

\( \mathbf{u}_2 \perp \mathbf{u}_1 \)
Since \( \| A \|_2 = \sigma_1 \), by definition \( \| A \psi_2 \|_2 \leq \sigma_1 \) ---- (a)

We now claim \( \| A \psi_2 \|_2 = \sigma_1 \), why? Because \( \psi = P_{\psi_1} \psi + (I - P_{\psi_1}) \psi = c \psi_1 + s \psi_2 \)

Exercise: why \( c^2 + s^2 = 1 \)? where \( c, s \) : constants satisfying \( c^2 + s^2 = 1 \) ---- (b)

\[ \sigma_1^2 = \| A \psi \|_2^2 = \| c A \psi_1 + s A \psi_2 \|_2^2 \]
\[ = c^2 \| A \psi_1 \|_2^2 + 2cs (A \psi_1)^T A \psi_2 + s^2 \| A \psi_2 \|_2^2 \]
\[ = c^2 \sigma_1^2 + s^2 \| A \psi_2 \|_2^2 \leq c^2 \sigma_1^2 + s^2 \sigma_1^2 = \sigma_1^2 \] (a) (b)

This means that the inequality above must be an equality, and hence \( \| A \psi_2 \|_2 = \sigma_1 \).

Hence, what we have proved is:

if \( \psi_i \) is not unique, then the corresponding singular value \( \sigma_i \) is not simple (i.e., has some multiplicity).

After determining \( \sigma_1, u_1, \psi_1 \), we can use the induction argument.

In particular, for \( A \) : square, \( \{ \sigma_j \} \) are distinct (no multiple singular values), then it’s clear that \( \{ u_j \}, \{ \psi_j \} \) are uniquely determined up to signs.