SVD: Singular Value Decomposition.

- The SVD is a matrix factorization used in many algorithms & applications.

- The SVD is motivated by the following geometric fact:

  The image of the unit sphere under any $m \times n$ matrix is a $\underbrace{\text{hyperellipse}}_{\substack{m\text{-dimensional}\\ \text{generalization of}\\ \text{an ellipse.}}}$.

Why is that?

Consider the unit circle in $\mathbb{R}^2$ ($\mathbb{R}^n$):

$$S = \{ x \in \mathbb{R}^2, \ \|x\|_2 = 1 \} \quad \left(S = \{ x \in \mathbb{R}^n, \ \|x\|_2 = 1 \} \right)$$

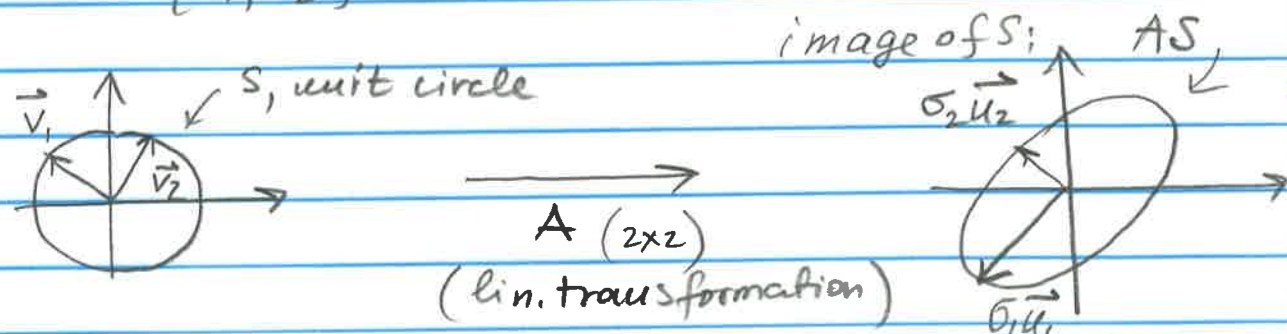
Q: What is AS, where A is an $m \times n$ matrix?

Hyperellipse (or ellipse) in $\mathbb{R}^2$

$$AS = \{ Ax, \ x \in \mathbb{R}^2 \mid \|x\|_2 = 1 \}$$

$$\left(\text{in } \mathbb{R}^n: \ AS = \{ \underset{m \times n}{A} x, \ x \in \mathbb{R}^n \mid \|x\|_2 = 1 \} \right.$$

Let $\{ \vec{v}_1, \vec{v}_2 \}$ be an ON basis in $\mathbb{R}^2$



σ_1, σ_2 are lengths of principal semiaxes, $\{ \vec{u}_1, \vec{u}_2 \}$ - ON basis in $\mathbb{R}^2$

σ_1, σ_2 are also called singular values of A

$A\vec{v}_1 = \sigma_1 \vec{u}_1$ & $A\vec{v}_2 = \sigma_2 \vec{u}_2$ or

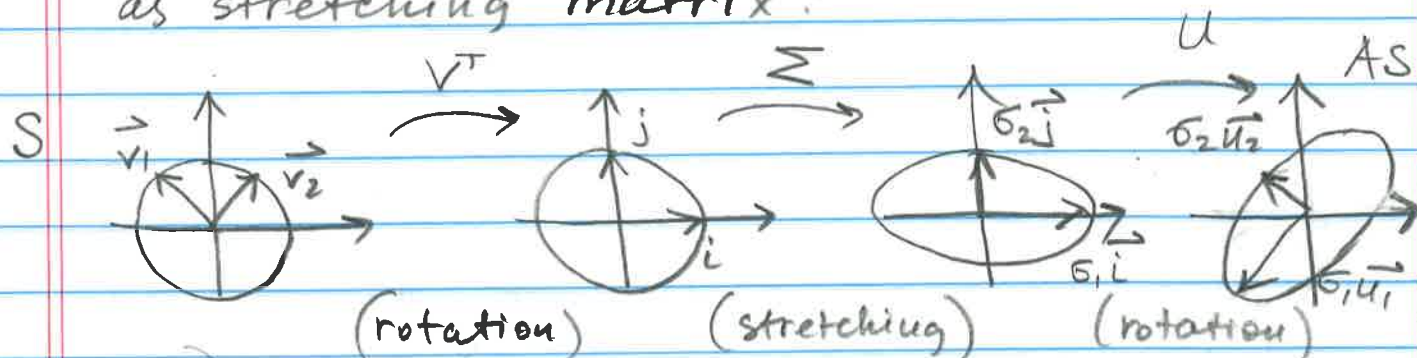
$$A \underbrace{[\vec{v}_1, \vec{v}_2]}_{V} = [\sigma_1 \vec{u}_1, \sigma_2 \vec{u}_2] = \underbrace{[\vec{u}_1, \vec{u}_2]}_{U} \underbrace{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}}_{\Sigma}$$

$$AV = U\Sigma$$

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

Note: $V^TV = VV^T = I$ & $U^TU = UU^T = I$

- Think of U & V as rotations and Σ as stretching matrix:



* for rotations: angle & length are preserved.

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

Generally: $\underset{m\times n}{A} = U\Sigma V^T$ (or $AV = U\Sigma$)

$m \geq n$

$$\underset{m\times n}{A} \cdot \underbrace{\left[\vec{v}_1 \,|\, \vec{v}_2 \,|\, \dots \,|\, \vec{v}_n\right]}_{V_{n\times n}} = \underbrace{\left[\vec{u}_1 \,|\, \vec{u}_2 \,|\, \dots \,|\, \vec{u}_n\right]}_{U_{m\times n}} \cdot \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}}_{\underset{n\times n}{\Sigma}}$$

$$A\vec{v}_j = \sigma_j \vec{u}_j \;, \quad j = 1, \dots, n$$

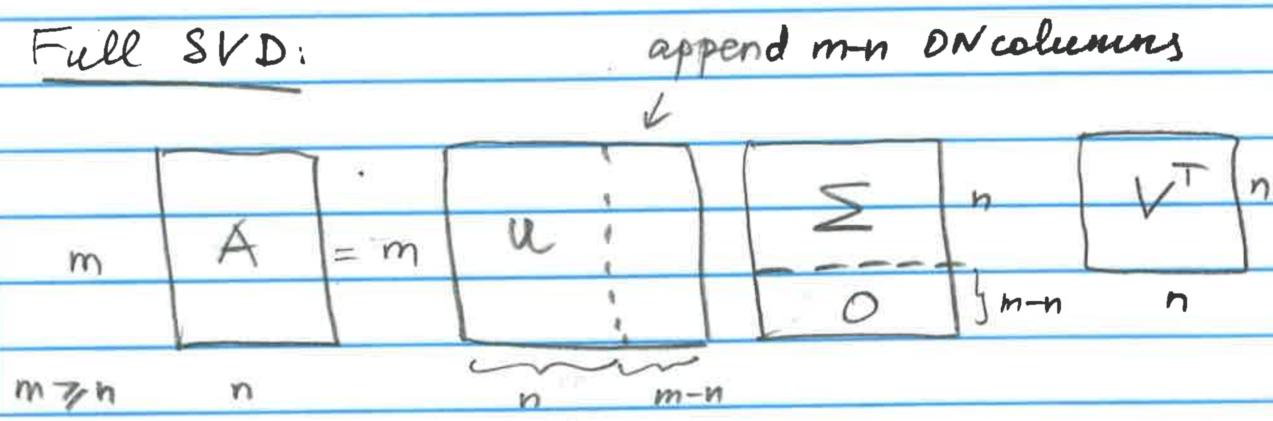
<u>Reduced SVD</u>

$$m \underset{n}{\boxed{A}} = \underset{n}{\boxed{U}} \; \underset{n}{\boxed{\Sigma}} \; \underset{n}{\boxed{V^T}}\, n$$

- $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$ are singular values
- $\vec{u}_1, \dots, \vec{u}_n$ are left singular vectors (in the dir. of principal semiaxes of AS), ON basis for range(A)
- $\vec{v}_1, \dots, \vec{v}_n$ are right singular vectors (preimages of the principal semiaxes of AS)

Reduced SVD used in many applications.

Full SVD:



(standard way)

Summary: for any $m \times n$ matrix A (with m & n arbitrary), with rank of A, $rk(A) = r \leq \min(m,n)$ ($rk(A)$ is the number of linearly independ. col's or rows of A), $A = U \Sigma V^T$ with U & V orthonormal and Σ diagonal with diagonal entries $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ ($\sigma_{r+1} = \dots = \sigma_n = 0$)

(Proof is by construction)

Note 1: singular values σ_i are unique.

Note 2: U & V are not uniquely determined: you can flip signs in U and V.

We can say, SVD is almost unique.

The (reduced) SVD can be written as

$$A = U\Sigma V^T = \begin{bmatrix} u_1 \dots u_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$$= \sigma_1 \underbrace{u_1 v_1^T} + \sigma_2 \underbrace{u_2 v_2^T} + \dots + \sigma_n \underbrace{u_n v_n^T} = \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

outer products $\rightarrow$ 1-rank matrices

$\left(\text{if } rk(A) < \min(m,n) \Rightarrow \sigma_{r+1}, \dots, \sigma_n = 0 \right)$

- Some facts about SVD:

① $A = U\Sigma V^T \Rightarrow A^T A = (U\Sigma V^T)^T (U\Sigma V^T)$

$= V\Sigma U^T U \Sigma V^T = V\Sigma^2 V^T$ where

$$\Sigma^2 = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix}, \quad \Rightarrow (A^T A) V = V\Sigma^2$$

$$AA^T = U\Sigma^2 U^T \Rightarrow (AA^T) U = U\Sigma^2$$

Thus, $\sigma_1^2, \dots, \sigma_n^2$ are eigen values of $A^T A$ & AA^T ; $\sigma_i = \sqrt{\lambda_i(A^T A)} = \sqrt{\lambda_i(AA^T)}$

$\vec{v}_1, \ldots \vec{v}_n$ (right singular vectors) are eigenvectors of A^TA and
$\vec{u}_1, \ldots, \vec{u}_n$ (left singular vectors) are eigenvectors of AA^T

2 for a symmetric A, $\sigma_i = |\lambda_i|$

3 for a square A, $|\det A| = \sigma_1 \sigma_2 \cdot \ldots \cdot \sigma_r$
$r = rk(A)$

4 $col(A) \overset{\text{or}}{=} range(A) = \langle u_1, \ldots, u_r \rangle$
$null(A) = \langle v_{r+1}, \ldots, v_n \rangle$
$\quad \| $
$\{x : Ax = 0\}$

$rk(A) = \dim(col(A)) = r$
$\dim(null(A)) = n - r$

5 $\|A\|_2 = \sqrt{\lambda_{max}(A^TA)} = \sigma_1$
$\|A\|_F = \sqrt{\sum_i \sigma_i^2}$

6 SVD vs. EVD (eigenvalue decomposition)
$(A = U\Sigma V^T)$ $\quad$ $(A = V\Lambda V^T)$

SVD	EVD
- uses 2 bases	- uses 1 basis
- bases are ON	- basis is not necessarily ON
- holds for any matrix	- not for all matrices (e.g., non-square A does not have EVD)

7 Work for SVD: $\sim O(mn^2)$ flops
MATLAB: $[U, S, V] = svd(A) \rightarrow$ full
$[U, S, V] = svd(A, 0) \rightarrow$ reduced

SVD is used for:

1) Calculating a pseudo-inverse matrix
2) Low-rank matrix approximations
3) Least squares (LS)
4) Principal component analysis

1) Pseudo-inverse.

If A is square and full-rank $\Rightarrow \exists A^{-1}$

$$A^{-1} = \underset{(\text{reduced})}{(U \Sigma V^T)^{-1}} = (V^T)^{-1} \Sigma^{-1} U^{-1} = V \Sigma^{-1} U^T.$$

If A is not square and/or is rank deficient $\Rightarrow \nexists A^{-1}$, but we can still use $V\Sigma^{-1}U^T$ to compute a pseudo-inverse:

$A^{+} = V \Sigma^{-1} U^T$, also called the Moore (1920) – Penrose (1951) inverse.

$\Rightarrow$ $Ax = b$ and A is singular $\Rightarrow$ x can be found as $x = A^{+} b$

(MATLAB: A^{+} = pinv(A))

2) Low-rank matrix approximations:

$$A = U \Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T$$

$r = rk(A)$

$r \le \min(m,n)$

$\underbrace{u_i v_i^T}_{\text{rank 1 matrices}}$

Take $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$, $k<r$ then we can approximate A with A_k:

$$A \approx A_k = U \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} V^T$$

3) Least squares via SVD:

$$Ax = b \quad w/ \quad m << n$$

Find x minimizing $\| b - Ax \|_2^2$ via SVD

(1) Compute $A = U \Sigma V^T$ (reduced)

(2) Compute $U^T b$

(3) Solve $\Sigma y = U^T b$ for y (diagonal system)

(4) Set $x = Vy$

4) PCA (principal component analysis):

if $X_{m \times n}$ is a data matrix (mean-centered, i.e., the mean of each column is 0), and if $X = U \Sigma V^T$, then the right singular vec's v_i are principal component directions of X. The vector $z_1 = Xv_1 = \sigma_1 u_1$ has the largest variance; $Var(z_1) = \sigma_1^2/m$; v_1 is the vector of max variance; $u_1 = \frac{1}{\sigma_1}(Xv_1)$ is called the first principal component of X. One can find second principal component (PC) and so on. These PC's are used to reduce dimensionality of data X while retaining the variation present.

SVD Example:

$$A = \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix} \Rightarrow A^T A = \begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$$

eig $(A^T A)$ are singular values of A

$\det(A^T A - \lambda I) = 0$

$$\begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} = (5-\lambda)^2 - 9 = 25 - 10\lambda + \lambda^2 - 9 = 0$$

$\Rightarrow \lambda^2 - 10\lambda + 16 = 0$

$\lambda_1 = 8, \quad \lambda_2 = 2 \quad \Rightarrow \quad \sigma_1 = \sqrt{8} = 2\sqrt{2}, \quad \sigma_2 = \sqrt{2}$

$(\sigma_1 > \sigma_2)$

Right singular vectors of A are eigenvectors of $A^T A$:

$$\lambda_1 = 8 \Rightarrow \begin{pmatrix} -3 & 3 \\ 3 & -3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \tilde{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\Rightarrow v_1 = \frac{\tilde{v}_1}{\|\tilde{v}_1\|} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

$$\lambda_2 = 2 \Rightarrow \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \tilde{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\Rightarrow v_2 = \frac{\tilde{v}_2}{\|\tilde{v}_2\|} = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Since $A v_1 = \sigma_1 u_1 \Rightarrow \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = 2\sqrt{2}\, u_1$

$\Rightarrow u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Similarly, from $A v_2 = \sigma_2 u_2$

$\Rightarrow u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Thus,

$$A = U \Sigma V^T = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}}_{V^T}$$

Note that we can compute u_1 & u_2 directly from AA^T:

u_1 & u_2 are eigenvectors of AA^T.

Remark: u_i & v_j are not unique
→ you can flip signs (but you should do it in u_i & v_j simultaneously).
σ_k's are uniquely determined.