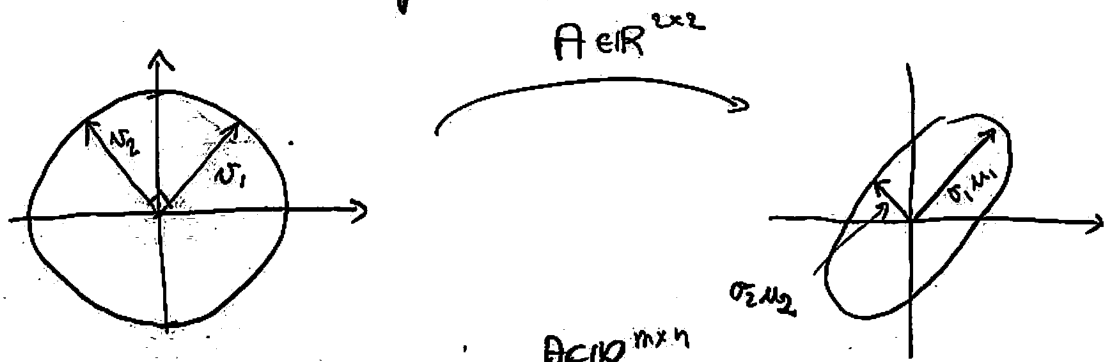
The Singular Value Decomposition (SVD)

<u>Idea</u>: A matrix $A \in \mathbb{R}^{m\times n}$ transforms unit sphere $\subset \mathbb{R}^n$ into a <u>hyperellipsoid</u> $\subset \mathbb{R}^m$.



mutually orthogonal vectors v_i in unit sphere $\xrightarrow{A \in \mathbb{R}^{m\times n}}$ $\sigma_1 u_1, \sigma_2 u_2, \dots, \sigma_n u_n$

where:

- σ_i = lengths of principal axis
- u_i = direction of principal axis.
  = mutually orthogonal and of unit length.

Usually the σ_i are ordered in decreasing order: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$.

(<u>Assumption</u> $m \geq n$: $A = {}_m\overset{n}{\square}$)

We get then:

$$
\begin{cases}
A v_i = \sigma_i u_i, \quad i = 1, \dots, n \\
v_i^T v_j = \delta_{ij} \\
u_i^T u_j = \delta_{ij}
\end{cases}
$$

In matrix form (always assuming $m \geq n$)

$$A[v_1\ v_2 \cdots v_n] = [u_1, u_2, \dots, u_n]\ \mathrm{diag}(\sigma_i)$$

$$A\ V^T = \tilde{U}\tilde{\Sigma}$$

where $\tilde{U} = [u_1\ u_2 \dots u_n]$

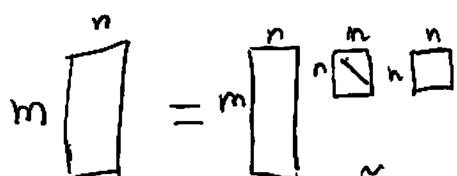
$$\tilde{\Sigma} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$$

u_i = left singular vectors
v_i = right singular vectors
σ_i = singular values

or more simply:

$$\boxed{A = \tilde{U}\tilde{\Sigma}V^T}, \text{ st } \tilde{U}^T\tilde{U} = I,\ V^TV = I.$$

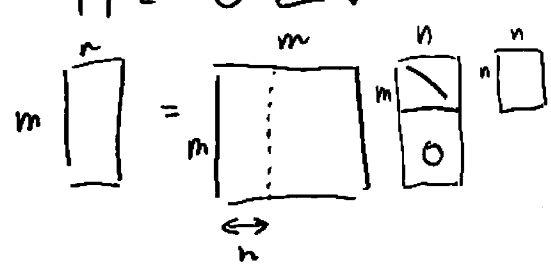
This is the "economy" sized SVD:



$$A = \tilde{U}\ \tilde{\Sigma}\ V^T$$

The full SVD is:

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



Here $U = [\tilde{U}, \hat{U}]$ s.t. $U^TU = \begin{bmatrix} \tilde{U}^T\tilde{U} & \tilde{U}^T\hat{U} \\ \hat{U}^T\tilde{U} & \hat{U}^T\hat{U} \end{bmatrix} =$

$$= \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = I$$

Also: $\Sigma = \begin{bmatrix} \tilde{\Sigma} \\ 0 \end{bmatrix}$ s.t. $U\Sigma = [\tilde{U}, \hat{U}] \begin{bmatrix} \tilde{\Sigma} \\ 0 \end{bmatrix} = \tilde{U}\tilde{\Sigma}$. ©

<u>Relation to eigendecomposition</u>

Let $A = \tilde{U}\tilde{\Sigma}V^T$ (economy sized, assuming $m \geq n$)

$$\boxed{A^TA = V\tilde{\Sigma}^T \underbrace{\tilde{U}^T\tilde{U}}_{I} \tilde{\Sigma} V^T = V\tilde{\Sigma}^2 V^T}$$

$\Rightarrow$ To compute SVD:

- Find eigendecomposition $V\Lambda V^T$ of A^TA, with eigenvalues sorted by decreasing order. ($\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$)
- Singular values are: $\sigma_i = \sqrt{\lambda_i}$.
- Right singular vectors are: V (matrix of eigenvectors)
- Left singular vectors are:

$$\tilde{U} = AV\tilde{\Sigma}^{-1}$$

(assuming $\sigma_i > 0$)

($u_i = \dfrac{Av_i}{\sigma_i}$ vector by vector).

<u>Note</u>: There are more efficient ways of computing SVD using the Golub-Kahan bidiagonalization. (matrix is reduced, through Householder reflectors, into a form where SVD is easier to compute)

When $A = A^T$ the SVD is related to eigendecomposition: D

$$A = U \Lambda U^T = \text{eigendecomposition}$$
$$= U |\Lambda| \, \mathrm{sign}(\Lambda) \, U^T = \text{SVD}$$

Indeed:
$$|\Lambda| = \Sigma$$
$$U = U$$
$$V = U \, \mathrm{sign}(\Lambda) \quad (\text{check } V^T V = I)$$

So for symmetric matrices:

- the singular values are the absolute values of eigenvalues
- the singular vectors are up to a sign the eigenvectors of A.

Here is another useful form of SVD: (assumes $m \geq n$)

$$\boxed{A = \sum_{i=1}^{n} \sigma_i u_i v_i^T} \qquad \left(\text{obtained } \Sigma = \sum_{i=1}^{n} \sigma_i e_i e_i^T \right)$$

$$e_i e_i^T = \begin{array}{c c} & i \\ i & \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix} \end{array}$$

If some of the σ_i are zero, say for example:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$$

Then sum collapses to r terms:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = U_r \Sigma_r V_r^T$$

Where:

$$U_r = \underset{m}{\overset{r}{[\;]}} , \quad \Sigma_r = r\overset{r}{[\diagdown]} \quad V_r = n\overset{r}{[\;]}$$

$= \mathrm{diag}(\sigma_1, \dots \sigma_r)$

= first r columns of U

= first r columns of V.

Going back to full SVD we partition:

$$U = [U_r, \bar{U}]$$
$$V = [V_r, \bar{V}]$$

$$A = U \Sigma V^T = [U_r \; \bar{U}] \left[\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} V_r^T \\ \bar{V}^T \end{bmatrix}$$

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

It is easy to show that:

Range $(A) =$ Range (U_r)
Null $(A) =$ Range $(\bar{V})$
Range $(A^T) =$ Range (V_r)
Null $(A^T) =$ Range $(\bar{U})$

So that:

Rank $(A) = \dim(\text{Range } U_r) = r.$
$= \#$ of non zero singular values

and also we can show Fundamental theorem of linear algebra immediately:

$$\boxed{\begin{aligned} \mathbb{R}^m &= \text{Range}(U_r) \oplus \text{Range}(\bar{U}) \\ &= \text{Range}(A) \oplus \text{Null}(A^T) \\ \mathbb{R}^n &= \text{Range}(V_r) \oplus \text{Range}(\bar{V}) \\ &= \text{Range}(A^T) \oplus \text{Null}(A) \end{aligned}}$$

- Another fundamental property of SVD is:

$$\sigma_1 = \|A\|_2 \quad \text{(induced matrix norm by } \|\cdot\|_2 = \text{Euclidean norm)}$$

The SVD provides a systematic way of approximating matrices with rank k matrices.

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

$\text{Rank}(A_k) = k \quad (\text{if } \sigma_k > 0)$

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

i.e. A_k = best rank k approx of B in induced 2-norm w/ error σ_{k+1}.

Similar result holds for Frobenius norm:

$$\|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} (a_{ij})^2 .$$

<u>Theorem</u>

$$\|A - A_k\|_F = \min_{\operatorname{rank}(B)=k} \|A - B\|_F = \sqrt{\sigma_{k+1}^2 + \dots + \sigma_n^2}$$