SVD

"yet another" matrix factorization... but
a very important one for many algorithms!

- Key step in many algorithms.
- Gives insight into many algebraic problems.

Observe: Given $\underline{A} \in \mathbb{R}^{m \times n}$, the image of
the hypersphere under $\underline{A}$

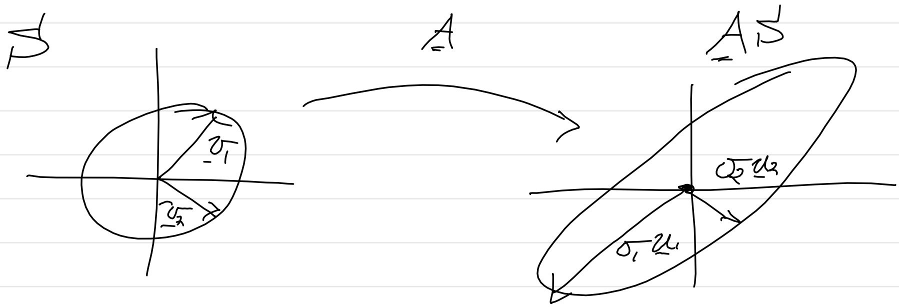
$$\underline{A}S = \{ \underline{A}\underline{x} \in \mathbb{R}^m \mid \|\underline{x}\|_2 = 1, \underline{x} \in \mathbb{R}^n \}$$

is a hyper-ellipse

Q: What is a hyper-ellipse?

A: Generalization of an ellipse. Given orthogonal unit vectors $\underline{u}_1, \dots, \underline{u}_m$ and scale factors $\sigma_1, \dots, \sigma_m$, a hyper-ellipse is obtained by stretching the unit sphere along each $\underline{u}_i$ by a factor σ_i.

We call the $\{ \sigma_i \underline{u}_i \in \mathbb{R}^m \}$ the principal semi-axes of the ellipse.



We will take this observation as a fact now, and will prove it later.

For now, let $\underline{\underline{A}} \in \mathbb{R}^{m \times n}$ $(m \geq n)$ have full rank.

Def: The n singular values of the matrix $\underline{\underline{A}} \in \mathbb{R}^{m \times n}$ are the lengths $\sigma_i \in \mathbb{R}_{\geq 0}$ of the principal semi-axes.

We typically order the singular values so that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$.

Def: The n left singular vectors are the unit vectors $\underline{u}_i \in \mathbb{R}^m$, $i = 1, \dots, n$, oriented along the axes of $\underline{\underline{A}}\underline{S}$. Hence, $\sigma_i \cdot \underline{u}_i$ is the i^{th}-largest principal semi-axis of $\underline{\underline{A}}\underline{S}$!

Def: The n right singular vectors of $\underline{\underline{A}}$ are the unit vectors $\underline{v}_i \in \mathbb{R}^n$ such that $\underline{\underline{A}}\,\underline{v}_i = \sigma_i \underline{u}_i$. i.e., they are the pre-images of the left singular vectors.

Reduced SVD

We have the algebraic relation

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

$$\underline{\underline{A}} \begin{pmatrix} | & | & & | \\ \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \underline{u}_1 & \underline{u}_2 & \cdots & \underline{u}_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{pmatrix}$$

i.e., $\underline{\underline{A}}\,\underline{\underline{\overline{V}}} = \underline{\underline{\hat{U}}}\,\underline{\underline{\overline{\Sigma}}}$ $\qquad \underline{\underline{A}} \in \mathbb{R}^{m\times n} \quad \underline{\underline{\hat{U}}} \in \mathbb{R}^{m\times n}$

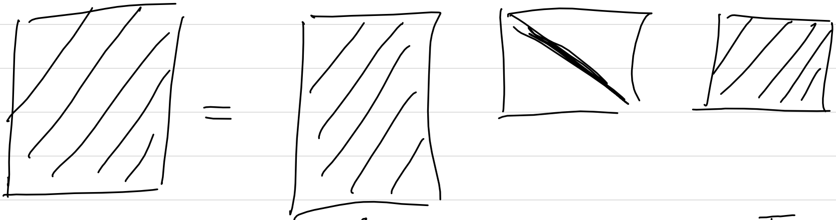
$\qquad \underline{\underline{\overline{V}}} \in \mathbb{R}^{n\times n} \quad \underline{\underline{\overline{\Sigma}}} \in \mathbb{R}^{n\times n}$

Because the $\{\underline{v}_i\}$ are <u>orthonormal</u>,

$$\underline{\underline{A}} = \underline{\underline{\hat{U}}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T \quad \text{(reduced SVD)}$$

left sv's are columns

right sv's are rows



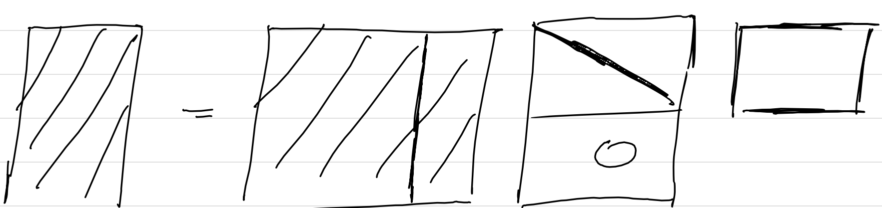
$A \quad = \quad \hat{U} \quad \Sigma \quad V^T$

"Full" SVD

Similar to "full" and "reduced" QR, there is also a full SVD.

Because $\hat{U} \in \mathbb{R}^{m \times n}$ with $m \geq n$ and the columns orthonormal, $\hat{U}$ "almost" orthogonal!

We can add $m-n$ arbitrary orthonormal columns to $\hat{U}$ to make it into an orthogonal matrix, and pad Σ with zeros.



$A \in \mathbb{R}^{m \times n} \quad = \quad U \in \mathbb{R}^{m \times m}$ (new) $\quad \Sigma \in \mathbb{R}^{m \times n} \quad V^T \in \mathbb{R}^{n \times n}$

<u>Note</u>: If $\underline{\underline{A}}$ has rank $r < n$, we can still construct the full QR factorization by appending $m-r$ rather than $m-n$ arbitrary columns to $\hat{\underline{\underline{U}}}$.

<u>Def</u>: Given $\underline{\underline{A}} \in \mathbb{C}^{m \times n}$, the SVD of $\underline{\underline{A}}$ is a factorization

$$\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^*$$

$U \in \mathbb{C}^{m \times m}$ unitary
$V \in \mathbb{C}^{n \times n}$ unitary
$\Sigma \in \mathbb{R}^{m \times n}$ diagonal

The elements σ_i of $\operatorname{diag}(\Sigma)$ are ≥ 0 and sorted in non-decreasing order.

<u>Note</u>: $\underline{\underline{V}}^*$ preserves the sphere

$\underline{\underline{\Sigma}}$ stretches to a hyperellipse aligned with canonical basis

$\underline{\underline{U}}$ rotates the hyperellipse

<u>Thm</u> (<u>existence of SVD</u>):

Every $\underline{\underline{A}} \in \mathbb{R}^{m \times n}$ has an SVD. The σ_i are uniquely determined. Moreover, if $\underline{\underline{A}}$

is square and the σ_i are distinct, the $\{\underline{u}_i\}$ and the $\{\underline{v}_i\}$ are determined up to complex sign ($|z| = 1$).

<u>Pf</u>: A bit out of scope. See Trefethen and Bau, Chapter I.

We will only cover existence.

<u>Induction</u>: Set $\sigma_1 = \|\underline{\underline{A}}\|_2$.

Let $\underline{u}_1$ be such that $\underline{\underline{A}}\,\underline{v}_1 = \sigma_1 \underline{u}_1$ with $\|\underline{u}_1\|_2 = \|\underline{v}_1\|_2 = 1$. This exists by compactness of the sphere.

<u>Extend</u> $\underline{u}_1$ to $\underline{\underline{U}}_1$, $\underline{v}_1$ to $\underline{\underline{V}}_1$ where the cols of $\underline{\underline{U}}_1$ and $\underline{\underline{V}}_1$ are orthonormal bases of $\mathbb{C}^m$, $\mathbb{C}^n$, respectively. Then,

$$\underline{\underline{U}}_1^* \underline{\underline{A}}\, \underline{\underline{V}}_1 = \underbrace{\begin{pmatrix} \sigma_1 & \underline{w}^* \\ \underline{0} & \underline{\underline{B}} \end{pmatrix}}_{\triangleq S} \qquad \begin{array}{l} \underline{w}^* \in \mathbb{R}^{1 \times (n-1)} \\ \underline{0} \in \mathbb{R}^{m-1} \\ \underline{\underline{B}} \in \mathbb{R}^{(m-1) \times (n-1)} \end{array}$$

Then,

$$\underbrace{\begin{pmatrix} \sigma_1 & \underline{w}^* \\ \underline{0} & \underline{\underline{B}} \end{pmatrix}}_{\triangleq\, \underline{\underline{S}}} \begin{pmatrix} \sigma_1 \\ \underline{w} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 + \|\underline{w}\|_2^2 \\ \underline{\underline{B}}\,\underline{w} \end{pmatrix} \triangleq \underline{b}$$

So that $\|\underline{b}\|_2^2 = \sigma_1^2 + \|\underline{w}\|_2^2 + \underbrace{\|\underline{\underline{B}}\,\underline{w}\|_2^2}_{\geq 0}$

$$\geq \sigma_1^2 + \|\underline{w}\|_2^2$$

$$= \left(\sigma_1^2 + \|\underline{w}\|_2^2\right)^{1/2} \cdot \left\| \begin{pmatrix} \sigma_1 \\ \underline{w} \end{pmatrix} \right\|_2$$

so that

$$\|\underline{\underline{S}}\|_2 = \max_{\underline{x}} \frac{\|\underline{\underline{S}}\,\underline{x}\|}{\|\underline{x}\|} \geq \left(\sigma_1^2 + \|\underline{w}\|_2^2\right)^{1/2}$$

But $\|\underline{\underline{S}}\|_2 = \|\underline{\underline{A}}\|_2 = \sigma_1$ because $\underline{\underline{U}}_1$ and $\underline{\underline{V}}_1$ are unitary, so $\underline{w} = \underline{0}$.

Hence,

$$\underline{\underline{A}} = \underline{\underline{U}}_1 \begin{pmatrix} \sigma_1 & \underline{0}^T \\ \underline{0} & \underline{\underline{B}} \end{pmatrix} \underline{\underline{V}}_1^*$$

By induction, $\underline{\underline{B}}$ has SVD $\underline{\underline{B}} = \underline{\underline{U}}_2 \underline{\underline{\Sigma}}_2 \underline{\underline{V}}_2^*$

Then,

$$\underline{\underline{A}} = \underline{\underline{U}}_1 \begin{pmatrix} 1 & \underline{0}^T \\ \underline{0} & \underline{\underline{U}}_2 \end{pmatrix} \begin{pmatrix} \sigma_1 & \underline{0}^T \\ \underline{0} & \underline{\underline{\Sigma}}_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \underline{\underline{V}}_2 \end{pmatrix}^* \underline{\underline{V}}_1^* \quad \square$$

Properties of the SVD

<u>Note</u>: SVD tells us that with the "right basis" for the range and the domain, every matrix is diagonal!

$\underline{\underline{U}}^T$ and $\underline{\underline{V}}^T$ project onto the left and right singular vectors, respectively. So if we define for $\underline{b} \in \mathbb{R}^m$, $\underline{x} \in \mathbb{R}^n$

$$\underline{b}' = \underline{\underline{U}}^T \underline{b}, \quad \underline{x}' = \underline{\underline{V}}^T \underline{x}$$

then:

$$\underline{b} = \underline{\underline{A}}\,\underline{x} \iff \underline{b}' = \underline{\underline{U}}^T \underline{\underline{A}}\,\underline{x}$$

$$= \underline{\underline{U}}^T (\underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T)\,\underline{x}$$

$$\Rightarrow \underline{b}' = \underbrace{\underline{\underline{\Sigma}}}_{\text{diagonal}}\,\underline{x}' \;!$$

<u>Thm</u>: The rank of $\underline{\underline{A}}$ is the number of nonzero singular values.

Pf: $\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T$, and $\underline{\underline{U}}$ and $\underline{\underline{V}}$ have full rank. The rank of $\underline{\underline{A}}$ is then the rank of $\underline{\underline{\Sigma}}$. But $\underline{\underline{\Sigma}}$ is diagonal, so its rank is the number of nonzero diagonals.

Thm: With $r = \text{Rank}(\underline{\underline{A}})$,

$$\text{range}(\underline{\underline{A}}) = \text{span}(\{u_1, \dots, u_r\})$$

$$\text{null}(\underline{\underline{A}}) = \text{span}(\{\underline{v}_{r+1}, \dots, \underline{v}_n\}) \qquad \square$$

Pf: Simply because $\underline{\underline{\Sigma}}$ diagonal,

$$\text{range}(\underline{\underline{\Sigma}}) = \text{span}(\{e_1, \dots, e_r\})$$

$$\text{Null}(\underline{\underline{\Sigma}}) = \text{span}(\{e_{r+1}, \dots, e_n\}) \qquad \square$$

Thm: The nonzero singular values of $\underline{\underline{A}}$ are the square roots of the nonzero eigenvalues of $\underline{\underline{A}}^*\underline{\underline{A}}$ and $\underline{\underline{A}}\,\underline{\underline{A}}^*$ (these have the same eigenvalues).

Pf: Note that

$$A^*A = (U\Sigma V^*)^*(U\Sigma V^*)$$

$$= (V\Sigma^* U^* U \Sigma V^*)$$

$$= \underbrace{V(\Sigma^*\Sigma)V^*}_{\text{Similarity transformation!}}$$

Eigenvalues of A^*A match $\Sigma^*\Sigma$.
$\Sigma^*\Sigma$ diagonal $\Rightarrow$ eigenvalues are $\sigma_1^2, \dots, \sigma_p^2, 0, \dots, 0$ with $n-p$ zeros if $n>p$.
Similar calculation for AA^*. $\square$

<u>Thm</u>: For $A \in \mathbb{C}^{m \times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$.

<u>Pf</u>: $|\det(A)| = |\det(U\Sigma V^*)|$

$$= \underbrace{|\det(U)|}_{=1}\,|\det(\Sigma)|\,\underbrace{|\det(V^*)|}_{=1}$$

$$= |\det(\Sigma)| = \prod_{i=1}^{m} \sigma_i. \quad \square$$

Thm: $\|\underline{\underline{A}}\|_2 = \sigma_1$, $\|\underline{\underline{A}}\|_F^2 = \|\underline{\sigma}\|_2^2$

Pf: we already saw that $\|A\|_2 = \sigma_1$ in the construction proof.

Now, note that:

$$\|A\|_F^2 = \sum_{ij} |a_{ij}|^2 = \sum_{ij} a_{ij}^* \cdot a_{ij}$$

$$= \mathrm{Tr}(\underline{\underline{A}}^* \underline{\underline{A}})$$

So that

$$\|A\|_F^2 = \mathrm{Tr}(V \Sigma^* U^* U \Sigma V^*)$$

$$= \mathrm{Tr}(V \Sigma^* \Sigma V^*)$$

$$= \mathrm{Tr}(V^* V \Sigma^* \Sigma) = \mathrm{Tr}(\Sigma^* \Sigma)$$

$$= \|\underline{\sigma}\|_2^2 \quad \square$$

Thm: If $\underline{\underline{A}} = \underline{\underline{A}}^*$, the singular values of $\underline{\underline{A}}$ are the absolute values of the e.v.'s.

Pf: $\underline{\underline{A}}$ symmetric $\Rightarrow$ set of orthogonal eigenvectors w/ real eigenvalues.

Then,

$$\underline{\underline{A}} = \underline{\underline{Q}}\, \underline{\underline{\Lambda}}\, \underline{\underline{Q}}^* = \underline{\underline{Q}}\, |\underline{\underline{\Lambda}}|\, \mathrm{sign}(\underline{\underline{\Lambda}})\, \underline{\underline{Q}}^*$$

With $|\underline{\underline{\Lambda}}|_{ij} = |\Lambda_{ij}|$,

$\underline{\underline{\mathrm{sign}}}(\underline{\underline{\Lambda}})_{ij} = \mathrm{sign}(\Lambda_{ij})$

Note that $\mathrm{sign}(\underline{\underline{\Lambda}})\,\underline{\underline{Q}}^*$ is unitary because $\underline{\underline{Q}}^*$ is, so that this is an SVD of $\underline{\underline{A}}$! $\square$

Q: But what even is this thing?

Note: we can write

$$\underline{\underline{A}} = \sum_{j=1}^{r} \sigma_j \underline{u}_j \underline{v}_j^*$$

as a sum of rank one matrices. There are many ways to do this, but SVD is the "best" one.

Thm: Let $0 \leq \nu \leq r$, and define

$$\underline{\underline{A}}_\nu = \sum_{j=1}^{\nu} \sigma_j \underline{u}_j \underline{v}_j^*$$

If $\nu = \min(m,n)$, define $\sigma_{\nu+1} = 0$. Then,

$$\|\underline{\underline{A}} - \underline{\underline{A}}_\nu\| = \underbrace{\min_{\substack{B \in \mathbb{C}^{m \times n} \\ \operatorname{rank}(\underline{\underline{B}}) \leq \nu}} \|A - B\|_2}_{\text{"best low-rank approx"}} = \sigma_{\nu+1}$$

Pf: See T+B, Ch. 1. □

Note: Also true w/ $\|\cdot\|_F$ replacing $\|\cdot\|_2$.

Geometrically, best approximation to a m-dimensional hyper-ellipse by a ν-dimensional one is the ν-largest principle semi-axes.