

SVD

Note Title

LECTURE 14

★ Formal Definition

Let $A \in \mathbb{R}^{m \times n}$

Then **SVD** of A is a factorization
full SVD $\rightarrow A = U \Sigma V^T$

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

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

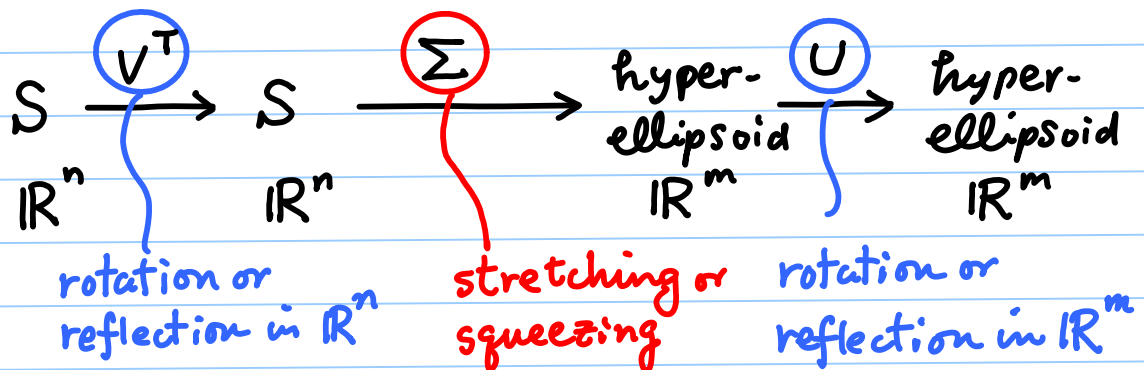
$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

$p = \min(m, n)$

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

A & Σ are the same shape.

Geometrically,



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 $\mathbb{R}^m$.

LECTURE 14

★ Existence & Uniqueness of SVD

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

Thm 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 σ_j 's are distinct, then singular vectors $\{u_j\}, \{v_j\}$ are uniquely determined up to signs (i.e., ± 1 factor).

(Proof : Existence)

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

$$\text{Set } \sigma_1 = \|A\|_2 \stackrel{\text{definition}}{=} \sup_{v \in S} \|Av\|_2$$

This is often called "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, $\exists v_1 \in S \subset \mathbb{R}^n$ s.t. $\|Av_1\|_2 = \sigma_1$ is attained.

Now set $\tilde{u}_1 = Av_1 \in \mathbb{R}^m$, and consider orthogonal matrices $V_1 = [v_1 \ v_2 \ \cdots \ v_n] \in \mathbb{R}^{n \times n}$,

LECTURE 14

$$U_1 = [u_1 \ u_2 \ \dots \ u_m] \in \mathbb{R}^{m \times m}$$

where $u_1 = \frac{1}{\sigma_1} \tilde{u}_1$

$$\begin{aligned} \text{Note } \|u_1\| &= \frac{1}{\sigma_1} \|\tilde{u}_1\| = \frac{1}{\sigma_1} \|Av_1\| \\ &= \frac{1}{\sigma_1} \cdot \sigma_1 = 1 \quad \checkmark \end{aligned}$$

$$\text{Then, } U_1^T A V_1 = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} A [v_1 \ \dots \ v_n]$$

$$= \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} [A v_1 \ \dots \ A v_n]$$

$\tilde{u}_1 = \sigma_1 u_1$

$$= \begin{bmatrix} \sigma_1 & w^T \\ \vdots & B \end{bmatrix}$$

$u_j^T u_1 = 0$ for $j \geq 2$.

let's call $= \Sigma_1$

$$\text{where } w^T = [u_1^T A v_2, \dots, u_1^T A v_n] \in \mathbb{R}^{1 \times n-1}$$

$$B = \begin{bmatrix} u_2^T A v_2 & \dots & u_2^T A v_n \\ \vdots & & \vdots \\ u_m^T A v_2 & \dots & u_m^T A v_n \end{bmatrix} \in \mathbb{R}^{m-1 \times n-1}$$

$$\begin{aligned} \left\| \underbrace{\begin{bmatrix} \sigma_1 & w^T \\ 0 & B \end{bmatrix}}_{\Sigma_1} \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 &\geq \sigma_1^2 + w^T w \\ &= \sqrt{\sigma_1^2 + \|w\|^2} \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 \end{aligned}$$

$$\Rightarrow \|\Sigma_1\|_2 \geq \sqrt{\sigma_1^2 + \|w\|^2} \quad \text{--- ①}$$

Since U_1, V_1 are orthogonal,

$$\|\Sigma_1\|_2 = \|A\|_2 = \sigma_1 \quad \text{--- ②}$$

LECTURE 14

From ① & ②, we can conclude that $w = 0$, i.e.,

$$U_1^T A V_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}$$

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 B has its SVD: $B = U_2 \Sigma_2 V_2^T$

$$\text{Then } A = \underbrace{U_1}_{U} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^T}_{V^T} V_1^T$$

This is an SVD of A ! ///

(Proof: Uniqueness)

Let $v_1 \in S \subset \mathbb{R}^n$ s.t.

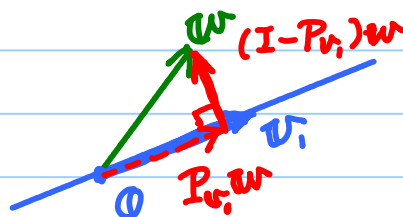
$$\|A\|_2 = \|\tilde{u}_1\|_2 = \|A v_1\|_2 = \sigma_1$$

Suppose $\exists w \in S$, s.t., $w \neq v_1$,
 w is linearly independent from v_1 ,
and $\|A w\|_2 = \sigma_1$.

Let's define a unit vector $v_2 \in S$ by

$$v_2 := \frac{(I - P_{v_1})w}{\|(I - P_{v_1})w\|_2}$$

$$v_2 \perp v_1$$



LECTURE 14

Since $\|A\|_2 = \sigma_1$, by definition

$$\|A v_2\|_2 \leq \sigma_1 \quad \text{--- (a)}$$

We now claim $\|A v_2\|_2 = \sigma_1$.

Exercise:

why?

$c^2 + s^2 = 1$?

Because $w = P_{v_1} w + (I - P_{v_1}) w$
 $= c v_1 + s v_2$

where c, s : constants satisfying $c^2 + s^2 = 1$ --- (b)

$$\sigma_1^2 = \|Aw\|_2^2 = \|c A v_1 + s A v_2\|_2^2$$

$$= c^2 \|A v_1\|_2^2 + 2cs (A v_1)^T A v_2 + s^2 \|A v_2\|_2^2$$

$$= c^2 \sigma_1^2 + s^2 \|A v_2\|_2^2 \stackrel{(a)}{\leq} c^2 \sigma_1^2 + s^2 \sigma_1^2 \stackrel{(b)}{=} \sigma_1^2$$

This means that the inequality above must be an equality, and hence $\|A v_2\|_2 = \sigma_1$ //

Hence, what we have proved is:

if v_1 is not unique, then the corresp. singular value σ_1 is not simple (i.e., has some multiplicity).

After determining σ_1, u_1, v_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 $\{u_j\}, \{v_j\}$ are uniquely determined up to signs. //