S V D

Note Title 5/3/2012

☆ Formal Definition

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

Then SVD of A is a factorization

full SVD → $$A = U \Sigma V^T$$

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

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

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

$$p = \min(m, n)$$

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

A & Σ are the same shape.

Geometrically,

$$\underset{\mathbb{R}^n}{S} \xrightarrow{V^T} \underset{\mathbb{R}^n}{S} \xrightarrow{\Sigma} \underset{\mathbb{R}^m}{\text{hyper-ellipse}} \xrightarrow{U} \underset{\mathbb{R}^m}{\text{hyper-ellipse}}$$

rotation or reflection in $\mathbb{R}^n$ — stretching or squeezing — rotation or reflection in $\mathbb{R}^m$

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 hyperellipse in $\mathbb{R}^m$.

☆ Existence & Uniqueness of SVD

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

<u>Thm</u> 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.

definition

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

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

$$V_1 = [v_1 \; v_2 \cdots v_n] \in \mathbb{R}^{n\times n},$$

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

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

Note $\|u_1\| = \frac{1}{\sigma_1}\|\tilde{u}_1\| = \frac{1}{\sigma_1}\|Av_1\|$

$$= \frac{1}{\sigma_1}\cdot\sigma_1 = 1 \ \checkmark$$

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

$$= \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} [\underbrace{Av_1}_{\tilde{u}_1 = \sigma_1 u_1} \cdots A v_n]$$

$$= \underbrace{\left[\begin{array}{c|c} \sigma_1 & w^T \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right]}_{\text{let's call } = \Sigma_1}$$

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

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

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

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

$$= \sqrt{\sigma_1^2 + \|w\|^2}\ \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2$$

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

Since U_1, V_1 are orthogonal,

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

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

$$U_1^T A V_1 = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & B \end{array}\right]$$

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: $B = U_2 \Sigma_2 V_2^T$

Then
$$A = U_1 \underbrace{\left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & U_2 \end{array}\right]}_{U} \underbrace{\left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & \Sigma_2 \end{array}\right]}_{\Sigma} \underbrace{\left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2 \end{array}\right]^T V_1^T}_{V^T}$$

This is an SVD of A! ///

(Proof: Uniqueness)

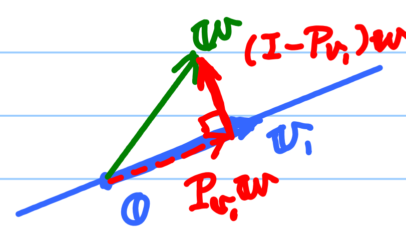
Let $v_1 \in S \subset \mathbb{R}^n$ s.t.

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

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

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



$v_2 \perp v_1$

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

$\|Av_2\|_2 \leq \sigma_1$ ---- (a)

We now claim $\|Av_2\|_2 = \sigma_1$.

Exercise: why $c^2+s^2=1$?

why? 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)

$$\sigma_1^2 = \|Aw\|_2^2 = \|cAv_1 + sAv_2\|_2^2$$

$$= c^2\|Av_1\|_2^2 + 2cs\,(Av_1)^T Av_2 + s^2\|Av_2\|_2^2$$

$$\leq c^2\sigma_1^2 + s^2\|Av_2\|_2^2 \underset{(a)}{\leq} c^2\sigma_1^2 + s^2\sigma_1^2 \underset{(b)}{=} \sigma_1^2$$

This means that the inequalities above must be equalities, and hence $\|Av_2\|_2 = \sigma_1$ //

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. ///