Sps $A \in \mathbb{R}^{m \times n}$ (think of as linear map $\mathbb{R}^n \to \mathbb{R}^m$)

Key Fact 1: it's possible to maximize $\|A\vec{v}\|$ over the n-sphere,

i.e. let $\sigma_1 = \sup_{\|\vec{v}\|=1} \|A\vec{v}\|$

then: σ_1 is realized, i.e. $\exists \vec{v}_1 \in \mathbb{R}^n$ s.t. $\|A\vec{v}_1\| = \sigma_1$ with $\|\vec{v}_1\| = 1$

first "singular value" of A

why: the function

$\vec{v} \to \|A\vec{v}\|$

is continuous and the n-sphere $\{\vec{v} \in \mathbb{R}^n : \|\vec{v}\| = 1\}$ is compact

↳ cts functions realize their extrema on compact domains

---

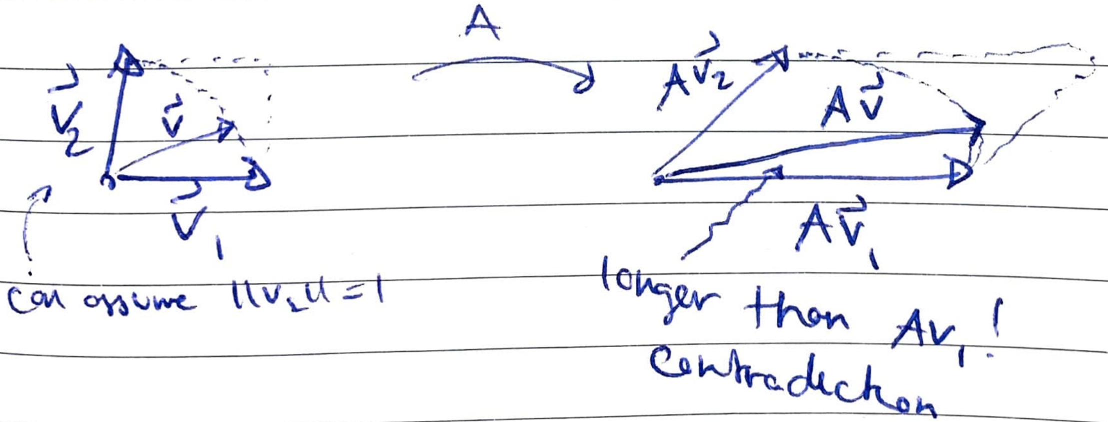
since $\|A\vec{v}\| = \sigma_1$ we can find $\vec{u} \in \mathbb{R}^m$ with $\|\vec{u}\| = 1$ s.t. $A\vec{v}_1 = \sigma_1 \vec{u}_1$

Key fact 2: if $\vec{v}_2 \perp \vec{v}_1$ (i.e. $\vec{v}_2 \in \{v_1\}^\perp$) then $A\vec{v}_2 \perp A\vec{v}_1$

!!! this is the crux

of course it's not in general true that if $\vec{u} \perp \vec{v}$ then $A\vec{u} \perp A\vec{v}$

It's true here because if $A\vec{v}_2$ were not orthog to $A\vec{v}_1$ would be possible to find a unit vector $\vec{v}$ ("leaning toward" $\vec{v}_2$) s.t. $\|A\vec{v}\| > \|A\vec{v}_1\|$, contradicting maximality of $\|A\vec{v}_1\|$ among unit vectors



(this is only a sketchy pictorial justification - but we'll take fact for granted)

~~[illegible]~~ Another way of saying this:

A maps the space $\{\vec{v}_1\}^\perp$ into $\{A\vec{v}_1\}^\perp (= \{\vec{u}_1\}^\perp)$

(of dim's $n-1$, $m-1$ resp.)

Now we can repeat this process <u>within</u> $\underline{\{v_1\}^\perp}$

i.e. we can find a $\vec{v}_2 \in \{\vec{v}_1\}^\perp$ with $\|\vec{v}_2\| = 1$ ~~[illegible]~~ so that $\|A\vec{v}_2\| = \sigma_2$ is max possible

$$\text{i.e. } \|Av_2\| = \sup_{\substack{\|\vec{v}\|=1 \\ \vec{v} \perp \vec{v}_1}} \|A\vec{v}\|$$

(possible to find such $\vec{v}_2$: again by compactness)

<u>the point</u>: by key fact 2

$A\vec{v}_2 \perp A\vec{v}_1 = \sigma_1\vec{u}_1 \longrightarrow \vec{u}_1 \perp \vec{u}_2$

can write $Av_2 = \sigma_2 \vec{u}_2$ w/ $\|\vec{u}_2\| = 1$

<u>notice</u> $\sigma_2 \leq \sigma_1$ since we maximized over smaller set.

and again: because $\vec{v}_2$ maximizes $\|A\vec{v}\|$ over $\{\vec{v}_1\}^\perp$,
if $\vec{v}_3 \in \{\vec{v}_1, \vec{v}_2\}^\perp$

$$\text{then } Av_3 \in \{A\vec{v}_1, A\vec{v}_2\}^\perp$$
$$= \{\vec{u}_1, \vec{u}_2\}^\perp$$

so continue...

---

summary: Given A $m \times n$

① Find $\vec{v}_1 \in \mathbb{R}^n$, $\|\vec{v}_1\| = 1$
s.t. $\|A\vec{v}_1\| = \sigma_1$ is max possible
for $\vec{v} \in \mathbb{R}^n$, $\|\vec{v}\| = 1$

can write $A\vec{v}_1 = \sigma_1 \vec{u}_1$ for some $u_1 \in \mathbb{R}^m$
$\|u_1\| = 1$
(possible by key fact 1)

② Find $\vec{v}_2 \perp \vec{v}_1$, $\|v_2\| = 1$
s.t. $\|A\vec{v}_2\| = \sigma_2$ is max poss.
for $\vec{v} \in \{v_1\}^\perp$, $\|\vec{v}\| = 1$

can write $A\vec{v}_2 = \sigma_2 u_2$ for some $\vec{u}_2 \in \mathbb{R}^m$

and we have (by key fact 2) $\|\vec{u}_2\| = 1$
$Av_1 \perp Av_2$ (so $u_1 \perp u_2$)

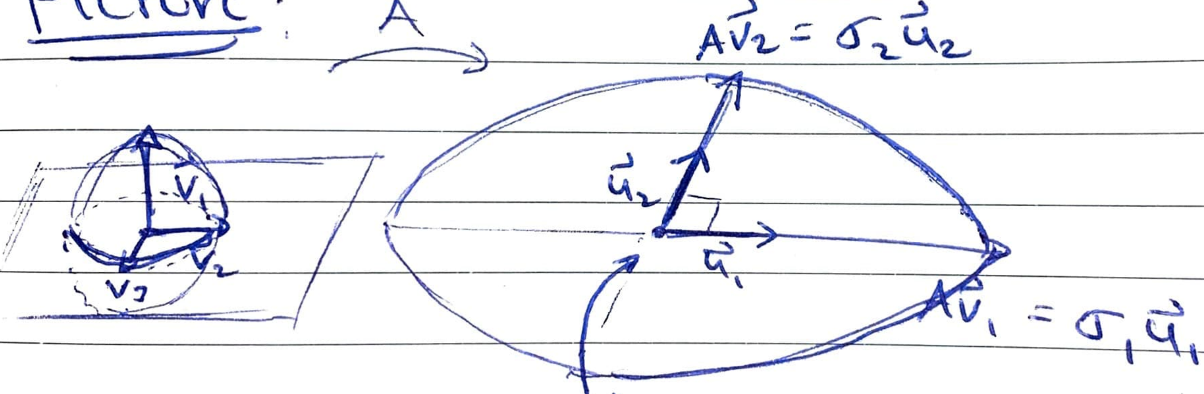
③ Find $\vec{v}_3 \perp \{\vec{v}_1, \vec{v}_2\}$ with $\|\vec{v}_3\| = 1$ s.t.
$\|A\vec{v}_3\| = \sigma_3$ is max poss.
for $\vec{v} \in \{\vec{v}_1, \vec{v}_2\}^\perp$ with $\|v\| = 1$

then $A\vec{v}_3 = \sigma_3 \vec{u}_3$ for some $\vec{u}_3$
$\|\vec{u}_3\| = 1$

<u>and</u> $\vec{u}_3 \in \{\vec{u}_1, \vec{u}_2\}^\perp$

⋮

<u>Picture</u>:



$A\vec{v}_3 = \sigma_3 \vec{v}_3 = 0\vec{v}_3 = \vec{0}$

* when we say "we can
find $\vec{u}_i$ w/ $\|u_i\| = 1$ s.t.
$A\vec{v}_i = \sigma_i \vec{u}_i$" we're assuming
$A\vec{v}_i$ is nonzero

- if $A\vec{v}_i = 0$ we either let $\vec{u}_i = \vec{0}$ (in case where we've already spanned $\mathbb{R}^m$ by $A\vec{v}_1, \dots A\vec{v}_{i-1}$)

or pick $\vec{u}_i$ to be some unit vector ($\|\vec{u}_i\| = 1$) orthonormal to $u_1, \dots, u_{i-1}$

In general: end up with orthonormal set $\vec{v}_1, \dots, \vec{v}_n \in \mathbb{R}^n$ and orthonormal set (up to some $\vec{0}$'s perhaps) $\vec{u}_1, \dots, \vec{u}_n$ in $\mathbb{R}^m$ s.t.

$$A\vec{v}_1 = \sigma_1 \vec{u}_1 \quad A\vec{v}_2 = \sigma_2 \vec{u}_2 \quad \dots$$

$$A\vec{v}_n = \sigma_n \vec{u}_n$$

and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$

In matrix form:

Let $V = \left[\vec{v}_1 \,\middle|\, \cdots \,\middle|\, \vec{v}_n \right]$ is $n \times n$

$\hat{U} = \left[\vec{u}_1 \,\middle|\, \cdots \,\middle|\, \vec{u}_n \right]$ $m \times n$

So: $AV = \left[A\vec{v}_1 \,\middle|\, \cdots \,\middle|\, A\vec{v}_n \right]$

$= \left[\sigma_1 \vec{u}_1 \,\middle|\, \cdots \,\middle|\, \sigma_n \vec{u}_n \right]$

$= \left[\vec{u}_1 \,\middle|\, \cdots \,\middle|\, \vec{u}_n \right] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$

$= \hat{U} \hat{\Sigma}$ $\quad$ $n \times n$

V is always orthogonal
(i.e. orthonormal columns + square)

$\hat{U}$ has orthonormal columns
(up to some $\vec{0}$'s)

to get full SVD we
"make $\hat{U}$ square" (keeping
columns orthonormal) and
adjust $\hat{\Sigma}$ accordingly

by adding some $\vec{0}$'s or extra
orthonormal columns.

<u>if $m \geq n$</u>: - add columns $\vec{u}_{n+1}, \ldots$
$\vec{u}_m$ to $\hat{U}$ so that $\{\vec{u}_1, \ldots, \vec{u}_n,$
$\vec{u}_{n+1}, \ldots, \vec{u}_m\}$ orthonormal (hence
a basis for $\mathbb{R}^m$)
- add rows of 0's to $\hat{\Sigma}$
to preserve eq'n

$$U = m\Big[\overbrace{\vec{u}_1 \,\Big|\, \cdots \,\Big|\, \vec{u}_n}^{\hat{U}} \,\Big|\, \vec{u}_{n+1} \,\Big|\, \cdots \,\Big|\, \vec{u}_m\Big]$$
$$\qquad\qquad m$$

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} \hat{\Sigma} \\ \\ \\ \end{matrix}$$