Suppose $A \in \mathbb{R}^{m \times n}$ (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: $\sigma_1$ is realized, i.e.

$\exists \vec{v} \in \mathbb{R}^n$ s.t. $\|A\vec{v}\| = \sigma_1$

with $\|\vec{v}\|=1$

Why: The function $\vec{v} \mapsto \|A\vec{v}\|$ is continuous and the $n$-sphere $\{\vec{v} \in \mathbb{R}^n : \|\vec{v}\|=1\}$ is compact.

Let's functions realize their extrema on compact domain

Since $\|A\vec{v}\| = \sigma_1$, we can find $\vec{u} \in \mathbb{R}^m$ with $\|\vec{u}\|=1$

so that $A\vec{v} = \sigma_1 \vec{u}$.
Key fact 2: if \( \frac{\vec{V}}{||\vec{V}||} + \frac{\vec{V}_1}{||\vec{V}_1||} \) (i.e. \( \frac{\vec{V}}{||\vec{V}||} \vec{V}_1 \)) then \( A\vec{V}_2 \perp A\vec{V}_1 \)

!!! This is the crux

Of course it's not in general true that if \( A\vec{v} \perp \vec{V} \) then \( A\vec{v} \perp A\vec{V} \)

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

\( \vec{V} \) and \( \vec{V}_1 \) can assume \( ||\vec{V}|| = 1 \) longer than \( A\vec{v}_1 \), contradiction

(THis is only a sketchy pictorial justification - but we'll fill fck for grounded)
Another way of saying this:
A maps the space \( R^{n} \perp \) into \( R^{m} \perp \) (of dims \( n-1, m-1 \) resp.)

Now we can repeat this process within \( R^{n} \perp \)

i.e., we can find a \( v_{2} \in R^{n} \perp \) with \( ||v_{2}|| = 1 \)
so that \( ||A v_{2}|| = \sigma_{2} \)
is max possible

\[ 1 - \sigma_{2} = \sup_{||v|| = 1} ||A v|| \]

(possible to find \( \sigma_{2} \): again by

**The point:** by key fact 2
\[ A \hat{v}_{2} + A \hat{u}_{1} = \sigma_{1} \hat{u}_{1} \Rightarrow \hat{u}_{1} + \hat{u}_{2} \]

can write \( A \hat{v}_{2} = \sigma_{2} \hat{u}_{2} \) \( \Rightarrow \hat{v}_{2} ||\hat{v}_{2}|| \)

notice \( \sigma_{2} \leq \sigma_{1} \) since we maximized over smaller set.
end again: because $\mathbf{v}_2$

maximizes $\mathbf{u}^T \mathbf{A} \mathbf{u}$ over $\{\mathbf{u}^1, \mathbf{u}^2\}$

if $\mathbf{v}_2 \in \{\mathbf{u}^1, \mathbf{u}^2\}^+$

then $\mathbf{A} \mathbf{v}_2 \in \{\mathbf{A} \mathbf{u}^1, \mathbf{A} \mathbf{u}^2\}^+$

$\{\mathbf{u}^1, \mathbf{u}^2\}^+$

so continue...

summary: Given $\mathbf{A}$ an $m \times n$ matrix:

1. Find $\mathbf{v}_1 \in \mathbb{R}^n$, $\|\mathbf{v}_1\|_2 = 1$
   s.t. $\|\mathbf{A} \mathbf{v}_1\|_2 = \sigma_1$ is max possible
   for $\mathbf{v} \in \mathbb{R}^n$, $\|\mathbf{v}\|_2 = 1$
   
   Can write $\mathbf{A} \mathbf{v}_1 = \sigma_1 \mathbf{u}^1$ for some $\mathbf{u}^1 \in \mathbb{R}^m$
   (possible by key fact 1) $\|\mathbf{u}^1\|_2 = 1$

2. Find $\mathbf{v}_2 \perp \mathbf{v}_1$, $\|\mathbf{v}_2\|_2 = 1$
   s.t. $\|\mathbf{A} \mathbf{v}_2\|_2 = \sigma_2$ is max possible
   for $\mathbf{v} \in \{\mathbf{v}_1\}^+$, $\|\mathbf{v}\|_2 = 1$
   
   Can write $\mathbf{A} \mathbf{v}_2 = \sigma_2 \mathbf{u}_2$ for some $\mathbf{u}_2 \in \mathbb{R}^m$

   and we have (by key fact 2) $\|\mathbf{u}_2\|_2 = 1$

$\mathbf{A} \mathbf{v}_1 \perp \mathbf{A} \mathbf{v}_2$ (so $\mathbf{u}_1, \mathbf{u}_2$)
(3) Find \( \mathbf{v}_3 \perp \{ \mathbf{v}_1, \mathbf{v}_2 \} \) with \( \| \mathbf{v}_3 \| = 1 \)

\[ \| \mathbf{A} \mathbf{v}_3 \| = \sigma_3 \| \mathbf{v}_3 \| \text{ max pos.} \]

For \( \mathbf{v} \in \{ \mathbf{v}_1, \mathbf{v}_2 \}^\perp \) with \( \| \mathbf{v} \| = 1 \)

then \( \mathbf{A} \mathbf{v}_3 = \sigma_3 \mathbf{u}_3 \) for some \( \mathbf{u}_3 \) with \( \| \mathbf{u}_3 \| = 1 \)

and \( \mathbf{u}_3 \in \mathbb{F} \mathbf{v}_1, \mathbf{v}_2 \) \( \perp \)

\[ \mathbf{A} \mathbf{v}_2 = \sigma_2 \mathbf{u}_2 \]

\[ \mathbf{A} \mathbf{v}_1 = \sigma_1 \mathbf{u}_1 \]

\[ \mathbf{A} \mathbf{v}_3 = \sigma_3 \mathbf{v}_3 = 0 \mathbf{v}_3 = 0 \]

---

* when we say "we can find \( \mathbf{u}_3 \) of \( \| \mathbf{u}_3 \| = 1 \) s.t. \( \mathbf{A} \mathbf{u}_3 = \sigma_3 \mathbf{u}_3 \)" we're assuming \( \text{det} \mathbf{A} \neq 0 \) and \( \mathbf{A} \mathbf{u}_3 = \sigma_3 \mathbf{u}_3 \neq 0 \)
- If $A\mathbf{v}_i = 0$, we either
  let $\mathbf{v}_i = 0$ (in case where
  we've already spanned
  $\mathbb{R}^m$ by $A\mathbf{v}_1, \ldots, A\mathbf{v}_{i-1}$)

  or pick $\mathbf{v}_i \in \mathbb{R}^m$
  to be some unit vector
  ($\|\mathbf{v}_i\| = 1$) orthonormal to
  $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$.
In general, we end up with an orthonormal set \( \{ v_1, \ldots, v_n \} \) of \( \mathbb{R}^n \) and orthonormal set (up to some \( \sigma_i \)'s perhaps) \( \{ \hat{u}_1, \ldots, \hat{u}_n \} \) in \( \mathbb{R}^m \), s.t.

\[
\hat{A} \hat{v}_1 = \sigma_1 \hat{u}_1, \quad \hat{A} \hat{v}_2 = \sigma_2 \hat{u}_2, \quad \ldots \quad \hat{A} \hat{v}_n = \sigma_n \hat{u}_n
\]

and \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0 \)

**In matrix form:**

\[
\begin{align*}
\text{Let } V &= \begin{bmatrix} \hat{v}_1 & \cdots & \hat{v}_n \end{bmatrix} \in \mathbb{R}^{m \times n} \\
\hat{U} &= \begin{bmatrix} \hat{u}_1 & \cdots & \hat{u}_n \end{bmatrix} \in \mathbb{R}^{n \times n} \\
\text{So: } AV &= \begin{bmatrix} A\hat{v}_1 & \cdots & A\hat{v}_n \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1 \hat{u}_1 & \cdots & \sigma_n \hat{u}_n \end{bmatrix} \\
&= \hat{U} \Sigma \\
&= \hat{U} \Sigma \Theta
\end{align*}
\]
$V$ is always orthonormal (i.e., orthonormal columns + square)

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

to get full SVD we

"make $\tilde{U}$ square" (keeping columns orthonormal) and

adjust $\hat{c}$ accordingly

by adding some $\hat{c}$'s or extra orthonormal columns.

If $m \geq n$:

- add columns $\tilde{u}_{n+1}, \ldots, \tilde{u}_m$ to $\tilde{U}$ so that $\tilde{u}_1, \ldots, \tilde{u}_m$ orthonormal (hence a basis for IR$^m$)

- add rows of $\hat{c}$'s to $\hat{c}$

  to preserve $E \hat{c}$

\[
U = \begin{bmatrix} \tilde{u}_1 & \cdots & \tilde{u}_n & \tilde{u}_{n+1} & \cdots & \tilde{u}_m \end{bmatrix}
\]

\[
E = \begin{bmatrix} \hat{c}_1 & \cdots & \hat{c}_n \end{bmatrix}
\]