Today: SVD (by finishing "Wide Equation" & Minim Norm Strg)

\[
\begin{align*}
\arg\min_{\tilde{u} \in \mathbb{R}^k} & \| C \tilde{u} \|_2^2 \\
\text{s.t.} & \ C \tilde{u} = \tilde{d} \\
\end{align*}
\]

Imagine if we had such an orthonormal basis \( V \).
First \( n \) columns of \( V \) are NOT in the nullspace of \( C \).
Last \( l-n \) \( \tilde{u} \)'s are in the nullspace of \( C \).

\[
\left\{ V = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \right\} \text{ for the nullsp of } C
\]

\( V_1 \perp V_2 \)

Work in these nicer coordinates,
\[
\tilde{u} = V^T \tilde{w} \implies \tilde{w} = V \tilde{u}
\]

\[
C \tilde{w} = C V \tilde{u}
\]
We can invoke the spectral theorem for real symmetric matrices:

or the matrix $S = C^t C$ has the same nullspace as $V = \text{the orthonormal basis of the eigenvectors of } S$.

Consider $\vec{v}_i \in V = [\vec{v}_1 \; \vec{v}_2 \; \cdots \; \vec{v}_n]$

$S \vec{v}_i = \lambda_i \vec{v}_i \quad \Rightarrow \quad C^t C \; \vec{v}_i = \lambda_i \vec{v}_i$

$S = \frac{1}{n-1} \left[ \begin{array}{c|c} C U_1 & C U_n \end{array} \right] \vec{w} = \left[ \begin{array}{c} \vec{w} \end{array} \right] \begin{array}{c} \sum_{i=1}^{n-1} \alpha_i \vec{w}_i \\ \sum_{i=1}^{n-1} \beta_i \vec{w}_i \end{array} \right] = \left[ \begin{array}{c} \vec{d} \end{array} \right]$

Recall $\| \vec{w} \| = \| \vec{w}_1 \| + \| \vec{w}_2 \|$

by orthogonality of $V$. 

$= \left[ \begin{array}{c|c} C U_1 & 0 \end{array} \right] \begin{array}{c} \vec{w}_1 \\ \vec{w}_2 \end{array} \begin{array}{c} \sum_{i=1}^{n-1} \alpha_i \vec{w}_i \\ \sum_{i=1}^{n-1} \beta_i \vec{w}_i \end{array} = \left[ \begin{array}{c} \vec{d} \end{array} \right]$

$\vec{w} = \frac{1}{n-1} \left[ \begin{array}{c} \vec{w}_1 \\ \vec{w}_2 \end{array} \right] \left[ \begin{array}{c} \sum_{i=1}^{n-1} \alpha_i \vec{w}_i \\ \sum_{i=1}^{n-1} \beta_i \vec{w}_i \end{array} \right] = \left[ \begin{array}{c} \vec{d} \end{array} \right]$

$\| \vec{w} \|^2 = \| \vec{w}_1 \|^2 + \| \vec{w}_2 \|^2$

$\Rightarrow$ Set $\vec{w}_1 = \vec{0}$ and solve for $\vec{w}_2$.

$\begin{array}{c|c} \sum_{i=1}^{n-1} \alpha_i \vec{w}_i \\ \sum_{i=1}^{n-1} \beta_i \vec{w}_i \end{array} = \left[ \begin{array}{c} \vec{d} \end{array} \right]$

Invert $C U_1$ to solve.
Learned Matrices of the form $C^T C$ have non-negative eigenvalues.

Arrange $v^1, v^2, \ldots, v^n$ so that
$$
\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0 = \lambda_{n+1} = \lambda_{n+2} = \ldots $$

Let $V_1 = [v^1, v^2, \ldots, v^n]$ and $S = [v^1, \ldots, v^n]$ is a basis for multiples of $C$.

First $n$ eigenvalues of $C^T C$ are

$$CV_i = [Cv^i, Cv^i, \ldots, Cv^i]$$

$\|Cv^i\|^2 = \lambda_i$ (let $w_i = \frac{Cv^i}{\sqrt{\lambda_i}}$ so $\|w_i\|^2 = 1$.

Thus $CV_i = [\sqrt{\lambda_1} w_i, \sqrt{\lambda_2} w_i, \ldots, \sqrt{\lambda_n} w_i]$.

For convenience: $\sigma_i = \sqrt{\lambda_i}$, so $CV_i = [\sigma_1 w_i, \sigma_2 w_i, \ldots, \sigma_n w_i]$.

Out of wild & unrealistic hopes, we compute $\langle \bar{w}_i, \bar{w}_j \rangle$.

(also laziness: can we avoid inverting a matrix?)

$$
\langle \bar{w}_i, \bar{w}_j \rangle = \bar{w}_i^T \bar{w}_j = \left( \frac{Cv_i}{\sigma_i} \right)^T \frac{Cv_j}{\sigma_j} \frac{Cv_i}{\sigma_i} \frac{Cv_j}{\sigma_j}
$$

$$
= \bar{v}_i^T C^T C \bar{v}_j
$$

$$
= \frac{\sigma_j}{\sigma_i} \cdot \frac{\sigma_i}{\sigma_i}
$$

$$
= \sigma_i \sigma_j
$$

$$
= 0 \quad \text{Amazing luck!}
$$

They are orthogonal to each other!
This means \( \mathbf{CV}_1 = [\sigma_1 \tilde{\mathbf{w}}, \ldots, \sigma_n \tilde{\mathbf{w}}_n] \)
\[
= \mathbf{W} [\sigma_1 \mathbf{0}, \ldots, \sigma_n \mathbf{0}] \quad \text{Call } \mathbf{E}_c = [\mathbf{0}, \mathbf{0}]
\]

When \( \mathbf{W} = [\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2, \ldots, \tilde{\mathbf{w}}_n] \)
is orthonormal and square

\[
(\mathbf{CV}_1)^T = (\mathbf{W} \mathbf{E}_c)^T = \mathbf{E}_c^{-1} \mathbf{W}^{-1}
\]
\[
= \mathbf{E}_c^{-1} \mathbf{W}^T
\]
\[
= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \ldots & \frac{1}{\sigma_n}
\end{bmatrix} \mathbf{W}^T
\]

So \( \tilde{\mathbf{w}}_{top} = \mathbf{E}_c^{-1} \mathbf{W}^T \mathbf{d} \)

So \( \tilde{\mathbf{w}}_i = \frac{1}{\sigma_i} \mathbf{w}_i \mathbf{d} \)

Recall \( \mathbf{w}_0 \mathbf{d} = 0 \)

So \( \tilde{\mathbf{w}} = \mathbf{V}_1 \tilde{\mathbf{w}}_{top} = \sum_{i=1}^{n} \mathbf{v}_i \mathbf{d} \)

We have found the minimum norm solution!

Step back and reflect.

We know \( \mathbf{CV}_1 = \mathbf{W} \mathbf{E}_c \)

What about \( \mathbf{CV} = \mathbf{C}[\mathbf{V}_1, \mathbf{V}_2] \)
\[
= [\mathbf{CV}_1, \mathbf{CV}_2]
\]
\[
= [\mathbf{W} \mathbf{E}_c, \mathbf{0}]
\]
\[
= \mathbf{W} [\mathbf{E}_c, \mathbf{0}]
\]
Let $E = [E_{c_1}, 0] \{n \}$. $E'$ is the same shape as $E$.

$CV = W \Sigma$ but $V$ is invertible.

So $C = W \Sigma V^T$

$\sigma_i$ is the $i$th singular value.

This is the Full SVD (Singular Value Decomposition).

To understand better, consider a general $\hat{u} = \sum_{i=1}^k \alpha_i \hat{v}_i$.

where $\alpha_i = \hat{v}_i^T \hat{u}$

$C \hat{u} = \sum_{i=1}^k \alpha_i \hat{v}_i$

$C \hat{v}_i = \sum_{i=1}^k \alpha_i \sigma_i \hat{w}_i + \sum_{i \neq i} \alpha_i \sigma_i \hat{w}_i \hat{v}_i^T \hat{u}$

$C = \sum_{i=1}^k \sigma_i (\hat{w}_i \hat{v}_i^T)$

This is called the outer product form of the SVD.
What were our assumptions? Do we actually need them?

We assumed controllability in our original problem.

\[ \implies C \] has a range (i.e. Col Space) that is n-dimensional.

What if this wasn't true. Suppose it was only r-dimensional.

We would still have \( V \) orthonormal.

(Specifically: \( V = C^* C \))

We would only get \( r \) nonzero singular values \( \lambda_1, \ldots, \lambda_r \).

\[ \implies r \text{ nonzero } \sigma_i \]

Only consequence:

\[ C = \sum_{i=1}^{r} \sigma_i \langle w_i, v_i \rangle \]

\( \text{Outer product form of SVD} \)

\text{Completely general for real matrices.}

Challenge: Can we get a full SVD?

Obstacle: We only have \( r \) orthonormal \( \tilde{w}_1, \ldots, \tilde{w}_r \).

Need \( n \) orthonormal vectors \( \{\tilde{w}_i\} \in \mathbb{W} \).

\[ \implies \text{Can use Givens Trick to set } n \text{ at them.} \]

\( S \), we have

\[ \mathbb{W} = [\tilde{w}_1, \ldots, \tilde{w}_r, \tilde{w}_{r+1}, \ldots, \tilde{w}_n] \]

Can write

\[ C = \mathbb{W} \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \mathbb{V}^T \]

The Full SVD is general for real matrices.

(Can extend to complex later.)
This has an interpretation:

\[ V \] is an orthogonal matrix = it is like a rotation/reflection.

\[ W \] is an orthogonal matrix = ... ...

\[ \Sigma \] scales axes & drops some of them at the end.

Every matrix \( C \) is just:
1) Rotate/Reflect
2) Scale along axes
3) Rotate/Reflect.

Like “Polar Coordinates” for matrices. The \( \sigma_i \) are like the magnitude & the \( \omega_i \) are two “rotation”.

Compact Form of SVD. (Like outer-product form but written using matrices)

\[ C = \{ [\overline{\omega}_1, \ldots, \overline{\omega}_r] [\sigma_1 \ldots \sigma_r] \} \rightarrow \begin{bmatrix} \overline{\omega}_1^T \\ \vdots \\ \overline{\omega}_r^T \end{bmatrix} \]

This is the compact form of the SVD.

Why is this true? \( C \overline{\omega} = \begin{bmatrix} [\overline{\omega}_1, \ldots, \overline{\omega}_r] [\sigma_1 \ldots \sigma_r] \end{bmatrix} \begin{bmatrix} \overline{\omega}_1^T \\ \vdots \\ \overline{\omega}_r^T \end{bmatrix} \]

\[ = [\sigma_1 \overline{\omega}_1 \ldots \sigma_r \overline{\omega}_r] \]
If $\sum_{i=1}^{\infty} \sigma_i \tilde{w}_i v_i u_i = \sum_{i=1}^{\infty} \sigma_i \tilde{w}_i v_i u_i$ is the outer product form.

This shows the validity of compact SVD.

What about tall matrices?

$$A = \begin{bmatrix} A \\ \vdots \\ A \end{bmatrix} \text{ when } l \geq m$$

$$= A = (A^T)^T$$

$$= (\Sigma \Sigma^T) = \Sigma \Sigma^T$$

$$= \begin{bmatrix} V \Sigma^T \\ \vdots \\ \Sigma^T \end{bmatrix} \text{ is full SVD}$$

Comment on notation. We wanted to avoid confusion with controls.

Traditional Notation

$$A = U \Sigma V^T$$ full SVD

Numpy uses this

$$= \begin{bmatrix} \Sigma \Sigma^T \\ \vdots \\ \Sigma^T \end{bmatrix} \text{ outer product}$$

$$= \begin{bmatrix} \Sigma \Sigma^T \\ \vdots \\ \Sigma^T \end{bmatrix} \text{ compact form}$$

Comment on computation. How big of a matrix do we need to compute eigenvectors for?

$$A A^T = U \Sigma V^T V \Sigma^T U^T$$

$$= U \Sigma \Sigma^T \text{ dimensions}$$

If you take eigenvectors of $A A^T$, we can set a set of $\tilde{w}_i$.
We can get $\bar{v}_i$ from these. (Think transpose)

So we can choose the smaller of $A^T A$ or $AA^T$. See discussion.

What does the SVD reveal about matrices?

\[
A : m \times n \quad : \mathbb{R}^n \rightarrow \mathbb{R}^m
\]

Let $r$ be the rank of $A$.

Since $A = \sum_{i=1}^{\min(m,n)} s_i \bar{v}_i \bar{u}_i^T$

\[
A^T = \sum_{i=1}^{\min(m,n)} s_i \bar{u}_i \bar{v}_i^T
\]

\[
r = \dim \text{Colspan}(A)
\]

In fact, $\bar{v}_1, \ldots, \bar{v}_r$ is a basis for $\text{Colspan}(A)$

\[
r = \dim \text{Colspan}(A^T)
\]

= \dim \text{Rowspan}(A)

We know $V_r = [\bar{v}_1, \ldots, \bar{v}_r]$ is a basis for $\text{Nullspace}(A)$

\[
\dim (\text{Nullspace}(A)) = n - r
\]

\[
\dim (\text{Nullspace}(A^T)) = m - r
\]

\[
V_r^T = [\bar{u}_1, \ldots, \bar{u}_r] \quad \text{spans the Colspace}(A^T)
\]

is a basis for the row of $A$.

$\text{Nullspace}(A) \perp \text{Colspace}(A^T)$

$\text{Nullspace}(A^T) \perp \text{Colspace}(A)$