16B Prof. Sahai

Today: SVD (by Finishing "Wide Equations" & Minimum Norm Story)

$$\underset{\vec{u}\in\mathbb{R}^{\ell}}{\arg\min}\ \|\vec{u}\|^2 \quad \text{s.t. } C\vec{u}=\vec{d}$$

$$n\left\{\overbrace{\begin{bmatrix} \quad C \quad \end{bmatrix}}^{\ell}\right. \begin{bmatrix} \vec{u} \end{bmatrix} = \begin{bmatrix} \vec{d} \end{bmatrix}\Big\}n$$

ℓ-long.

Imagine if we had such an orthonormal basis V.

First n columns of V are <u>NOT</u> in the nullspace of C

Last $\ell-n$ " " " are in the nullspace of C.

$$\ell\Big\{\underbrace{V}_{\ell} = \begin{bmatrix} \underbrace{V_1}_{n} & \underbrace{V_2}_{\ell-n} \end{bmatrix}\Big\}\ell$$

V_2 is an orthonormal basis for the nullspace of C.

$V_1 \perp V_2$

Work in these nicer coordinates,

$$\vec{\tilde{u}} = V^T\vec{u} \implies \vec{u} = V\vec{\tilde{u}}$$

Coordinates in nice basis

original basis.

$$C\vec{u} = CV\vec{\tilde{u}}$$

$$= \begin{bmatrix} \quad C \quad \end{bmatrix}\begin{bmatrix} V_1 & V_2 \end{bmatrix}\begin{bmatrix} \vec{\tilde{u}} \end{bmatrix}$$

ℓ

$$n\left\{\left[\underbrace{C V_1}_{n} \;\middle|\; \underbrace{C V_2}_{l-n}\right]\right. \left[\vec{\tilde{u}}\right]$$

Recall $\|\tilde{u}\| = \|\vec{u}\|$ by orthonormality of V.

$$= \left[\overbrace{C V_1}^{n} \;\middle|\; \overbrace{\;0\;}^{l-n}\right] \begin{bmatrix} \tilde{u}_{top} \\ \tilde{u}_{bot} \end{bmatrix} \begin{matrix} \}\, n \\ \}\, l-n \end{matrix} = \left[\vec{d}\right]$$

$$\|\tilde{u}\|^2 = \|\tilde{u}_{top}\|^2 + \|\tilde{u}_{bot}\|^2$$

$\Longrightarrow$ Set $\tilde{u}_{bot} = \vec{0}$ and solve for $\tilde{u}_{top}$.

$$n\left\{\underbrace{\left[C V_1\right]}_{n}\right. \vec{\tilde{u}}_{top} = \vec{d}$$

Invert CV_1 to solve.

We can invoke the spectral theorem for real sym. matrices on the matrix $S = C^T C$ ← has the same nullspace as C

V = the orthonormal basis of the eigenvectors of S.

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

$$S\vec{v}_i = \lambda_i \vec{v}_i \quad \Longrightarrow \quad C^T C \vec{v}_i = \lambda_i \vec{v}_i$$

So $\vec{v}_i^{\,T} C^T C \vec{v}_i = \lambda_i \vec{v}_i^{\,T} \vec{v}_i$

$$(C\vec{v}_i)^T (C\vec{v}_i) = \lambda_i \cdot \|\vec{v}_i\|^2 = \lambda_i$$

$$\|C\vec{v}_i\|^2$$

$$\Longrightarrow \lambda_i = \|C\vec{v}_i\|^2 \geq 0$$

Learned Matrices of the form C^TC have non-negative eigenvalues.

Arrange $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_\ell$ so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0 = \underbrace{\lambda_{n+1} = \lambda_{n+2} = \cdots = \lambda_\ell}$$

Let $V_1 = \underbrace{[\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n]}_{\text{First } n \text{ eigenvectors of } C^TC}$ So $\underbrace{[\vec{v}_{n+1}, \ldots, \vec{v}_\ell]}_{V_2}$ is a basis for nullspace of C.

$$CV_1 = [C\vec{v}_1 \;\; C\vec{v}_2 \;\; \cdots \;\; C\vec{v}_n]$$

I know $\|C\vec{v}_i\|^2 = \lambda_i$ Let $\vec{\omega}_i = \dfrac{C\vec{v}_i}{\sqrt{\lambda_i}}$ So $\|\vec{\omega}_i\| = 1$

Then $CV_1 = [\sqrt{\lambda_1}\cdot\vec{\omega}_1 \;\; \sqrt{\lambda_2}\cdot\vec{\omega}_2 \;\; \cdots \;\; \sqrt{\lambda_n}\cdot\vec{\omega}_n]$

For convenience: $\boxed{\sigma_i = \sqrt{\lambda_i}}$ So $CV_1 = [\sigma_1\vec{\omega}_1 \;\; \sigma_2\vec{\omega}_2 \;\; \cdots \;\; \sigma_n\vec{\omega}_n]$

Out of wild & unrealistic hope, we compute $\langle \vec{\omega}_i, \vec{\omega}_j \rangle$
(also laziness: can we avoid inverting a matrix?)

$$\langle \vec{\omega}_i, \vec{\omega}_j \rangle = \vec{\omega}_j^T \vec{\omega}_i = \left(\frac{C\vec{v}_j}{\sigma_j}\right)^T \frac{C\vec{v}_i}{\sigma_i}$$

$$= \frac{\vec{v}_j^T C^T C \vec{v}_i}{\sigma_j \cdot \sigma_i}$$

$$= \frac{\vec{v}_j^T \cdot \overset{\sigma_i^2}{\lambda_i} \vec{v}_i}{\sigma_j \sigma_i} = \frac{\sigma_i}{\sigma_j} \vec{v}_j^T \vec{v}_i$$

$$= 0$$

<u>Amazing Luck!</u>
They are orthogonal to each other!

This means $C V_1 = [\sigma_1 \vec{w}_1 \; \cdots \; \sigma_n \vec{w}_n]$

$$= W \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 \ddots & \\ 0 & & \sigma_n \end{bmatrix} \quad \text{Call } \Sigma_c = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}$$

Where $W = [\vec{w}_1 \; \vec{w}_2 \; \cdots \; \vec{w}_n]$

is orthonormal and square

$$(CV_1)^{-1} = (W \Sigma_c)^{-1} = \Sigma_c^{-1} W^{-1}$$

$$= \Sigma_c^{-1} W^T$$

$$= \begin{bmatrix} 1/\sigma_1 & & 0 \\ & 1/\sigma_2 \ddots & \\ 0 & & \frac{1}{\sigma_n} \end{bmatrix} W^T$$

So $\vec{\tilde{u}}_{top} = \Sigma_c^{-1} W^T \vec{d}$ So $\tilde{u}_i = \frac{1}{\sigma_i} \vec{w}_i^T \vec{d}$

Recall $\tilde{u}_{bottom} = \vec{0}$

So $\vec{u} = V_1 \vec{\tilde{u}}_{top} = \sum_{i=1}^{n} \vec{v}_i \cdot \frac{1}{\sigma_i} \vec{w}_i^T \vec{d}$

We have found the minimum norm solution!

Step back and reflect.

We know $\boxed{CV_1}_{n \times n} = \overset{n\times n}{W} \, \overset{n \times n}{\Sigma_c}$

What about $C V = C[V_1 \; V_2]$

$$= [CV_1 \; CV_2]$$

$$= [W\Sigma_c \;\; 0]$$

$$= W[\Sigma_c \;\; 0]$$

Let $\Sigma = \underbrace{\begin{bmatrix} \Sigma_c & 0 \end{bmatrix}}_{\ell} \Big\} n$ $\quad$ Σ is the same shape as C

$CV = W\Sigma$ $\quad$ But V is invertible.

So $\boxed{C = W \Sigma V^T}$

$$= n\underbrace{\begin{bmatrix} W \end{bmatrix}}_{n} \; n\underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & 0 & \\ 0 & & \ddots & \\ & & & \sigma_n \end{bmatrix}\begin{matrix} \\ 0 \\ \\ \end{matrix}}_{\ell} \; \overbrace{\begin{bmatrix} \vec{v}_1^{\,T} \\ \vec{v}_2^{\,T} \\ \vdots \\ \vec{v}_\ell^{\,T} \end{bmatrix}}^{\ell} \Big\} \ell$$

σ_i is the ith singular value.

This is the Full SVD.
Singular Value Decomposition

To understand better, consider a generic $\vec{u} = \sum_{i=1}^{\ell} \alpha_i \vec{v}_i$

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

$$C\vec{u} = \sum_{i=1}^{\ell} \alpha_i C \vec{v}_i = \sum_{i=1}^{n} \alpha_i \sigma_i \vec{w}_i + \sum_{i=n+1}^{\ell} \alpha_i \cancelto{\vec{0}}{C\vec{v}_i}$$

$$= \sum_{i=1}^{n} \sigma_i \vec{w}_i \alpha_i$$

$$= \sum_{i=1}^{n} \sigma_i \vec{w}_i \vec{v}_i^{\,T} \vec{u}$$

$$= \left(\sum_{i=1}^{n} \sigma_i \vec{w}_i \vec{v}_i^{\,T} \right) \vec{u}$$

$$\Longrightarrow \boxed{C = \sum_{i=1}^{n} \sigma_i \left(\vec{w}_i \vec{v}_i^{\,T} \right)}$$

Matrices.

$\vec{w}_i \vec{v}_i^{\,T}$ is the outer product.
Each is same shape as C.

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.

$\Longrightarrow$ C has a range (i.e. Col Space) that was n-dimensional

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

We would still have $l \times l$ V orthonormal.

(Spectral Thm on $C^T C$)

We would only get r nonzero eigenvalues $\lambda_1, \dots, \lambda_r$

$\Longrightarrow$ r nonzero σ_i

Only Consequence: $$\boxed{C = \sum_{i=1}^{r} \sigma_i (\vec{w}_i \vec{v}_i^T)}$$ $\longleftarrow$ Completely General for real matrices

Outer-Product Form of SVD

Challenge: Can we get a full SVD?

Obstacle: We only have r orthonormal $\vec{w}_1, \dots, \vec{w}_r$.

Need n orthonormal vectors to form W.

$\Longrightarrow$ Can use G.S. trick to get n of them.

So we have $W = [\vec{w}_1, \dots, \vec{w}_r, \vec{w}_{r+1}, \dots, \vec{w}_n]$

Can write $$C = W \begin{bmatrix} \begin{matrix} \sigma_1 & & & \\ & \sigma_2 & & 0 \\ & & \ddots & \\ 0 & & & \sigma_r \\ & & & & 0 \\ & & & & & 0 \\ & & & & & & \ddots \\ & & & & & & & 0 \end{matrix} & 0 \end{bmatrix} V^T$$

The Full SVD is general for real matrices.

(Can extend to complex later)

This has an interpretation.

V^T is an orthonormal matrix = It is like a rotation/reflection.

W is an orthonormal matrix = " " "

Σ : 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 σ_i are like the magnitude & there are two "rotations"

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

$$C = n\Bigg\{\overbrace{\begin{bmatrix} \vec{w}_1 & \cdots & \vec{w}_r \end{bmatrix}}^{r} \overbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}}^{r}\Bigg\} r \;\underbrace{\begin{bmatrix} \vec{v}_1^{\,T} \\ \vec{v}_2^{\,T} \\ \vdots \\ \vec{v}_r^{\,T} \end{bmatrix}}_{\ell}\Bigg\} r$$

This is the compact form of the SVD,

Why is this true? $$C\vec{u} = \begin{bmatrix} \vec{w}_1, & \cdots, & \vec{w}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \vec{v}_1^{\,T}\vec{u} \\ \vdots \\ \vec{v}_r^{\,T}\vec{u} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1\vec{w}_1 & \cdots & \sigma_r\vec{w}_r \end{bmatrix} \begin{bmatrix} \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

$$= \sum_{i=1}^{r} \sigma_i \vec{w}_i \, \vec{v}_i^T \vec{u}$$

Outer Product form.

This shows the validity of compact SVD.

What about tall matrices?

$$\underbrace{\begin{bmatrix} A \\ \\ \\ \\ \end{bmatrix}}_{m} \Bigg\} \ell \qquad \text{when } \ell \geq m$$

This is now wide. And has an SVD.

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

$$= (\underbrace{W \Sigma V^T}_{\text{has same shape as } A^T})^T = V \Sigma^T W^T$$

Σ^T has same shape as A

Full SVD

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

Traditional Notation

Numpy uses this.

$$A = U \Sigma V^T \qquad \text{Full SVD}$$

$$= \sum_{i=1} \sigma_i \vec{u}_i \vec{v}_i^T \qquad \text{outer product}$$

$$= U_c \underbrace{\Sigma_c}_{r \times r} V_c^T \qquad \text{Compact form}$$

<u>Comment on computation</u> How big of a matrix do we need to compute eigenvectors for?

If $\quad AA^T = U \Sigma V^T V \Sigma^T U^T$

$$= U \underbrace{\Sigma \Sigma^T}_{\text{diagonal}} U^T$$

If you take eigenvectors of AA^T, we can get a set of $\vec{u}_i$

$\Longrightarrow$ We can get $\vec{v}_i$ from these. (Think transpose)

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

What does the SVD reveal about matrices?

$$A : m\Big\{\overbrace{\begin{bmatrix} A \end{bmatrix}}^{n} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

Let r be the rank of A.

Since $A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T$

$r = \dim \text{ColSpan}(A)$

In fact $\vec{u}_1, \ldots, \vec{u}_r$ is a basis for ColSpan(A)

$$A^T = \sum_{i=1}^{r} \sigma_i \vec{v}_i \vec{u}_i^T$$

$r = \dim \text{ColSpan}(A^T)$
$= \dim \text{RowSpan}(A)$

We know $V_2 = [\vec{v}_{r+1}, \ldots, \vec{v}_n]$ is a basis for Nullspace of A

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

$\dim(\text{Nullspace}(A^T)) = m - n$

$V_1 = [\vec{v}_1, \ldots, \vec{v}_r]$ spans the Col Space (A^T)

i.e. a basis for the rows of A.

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

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