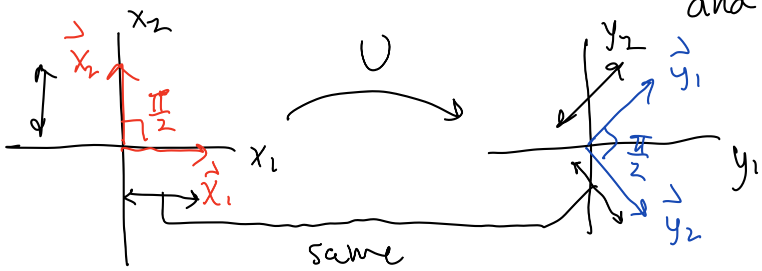
EECS16B DIS 9B remote

tall $\begin{bmatrix} \\ \\ \\ \end{bmatrix}$ square $\begin{bmatrix} \\ \end{bmatrix}$

Q: non-square orthonormal matrices?

Learning objectives

1. Tall orthonormal matrices, U, have $U^TU = I$
2. Orthonormal transforms preserve inner products (and therefore lengths and angles in $\mathbb{R}^n$)



$\langle \vec{x}, \vec{y} \rangle = \|\vec{x}\| \|\vec{y}\| \cos\theta$

3. Least squares problem and solution from a particular perspective☆: If some factorization of a matrix, $\underline{X}$, into $\underbrace{\underset{m\times m}{U}\,\underset{m\times n}{\Sigma}\,\underset{n\times n}{V^T}}$ exists, we can solve least squares really quickly. (U, V orthonormal, Σ diagonal)

☆ This is a preview. You'll get more in lecture on Thursday and next week.

Some dependencies/reference sheet

$\vec{x}, \vec{y} \in \mathbb{R}^n$

1 • $\langle \vec{x}, \vec{y} \rangle = \vec{x}^T\vec{y} = \vec{y}^T\vec{x}$ (inner product)

2 • $(AB)^T = B^TA^T$ (transpose of matrix product)

(to prove ↑ write columns of A, and rows of B and AB in terms of those vectors)

$$\begin{bmatrix} - \vec{x}_1^T V - \\ \vdots \\ - \vec{x}_m^T V - \end{bmatrix}$$

3 • $\|\vec{x}\|^2 = \langle \vec{x}, \vec{x} \rangle = \vec{x}^T\vec{x}$ (norm/magnitude/length in terms of inner product)

4 • $$A \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ A\vec{v}_1 & \cdots & A\vec{v}_n \\ | & & | \end{bmatrix}$$

$$\begin{bmatrix} - \vec{w}_1^T - \\ - \vec{w}_2^T - \\ \vdots \\ - \vec{w}_n^T - \end{bmatrix} A = \begin{bmatrix} - \vec{w}_1^T A - \\ - \vec{w}_2^T A - \\ \vdots \\ - \vec{w}_n^T A - \end{bmatrix}$$

(How matrix multiplication affects columns and rows to the right and left respectively)

5 • $A\vec{x} \approx \vec{b} \rightarrow \hat{\vec{x}} = (A^TA)^{-1}A^T\vec{b}$ (least squares solution)

$\hat{\vec{b}} = A(A^TA)^{-1}A^T\vec{b}$ (least squares projection)

EECS 16B Designing Information Devices and Systems II
Fall 2021 Discussion Worksheet **Discussion 9B**

The following notes are useful for this discussion: Note 14

1. Orthonormality and Least Squares

(a) Let U be an $m \times n$ matrix with orthonormal columns, with $m \geq n$. **Compute $U^\top U$. How does this change if $m < n$?**

$$U = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \\ | & | & & | \end{bmatrix}$$

$\vec{u}_i \in \mathbb{R}^m$

U is tall or square
(m>n) (m=n)

$U^T U = I$

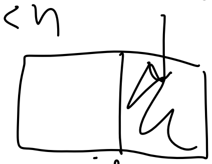
$(U^T U)_{ij} = \vec{u}_i^T \vec{u}_j = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$

$$U^T = \begin{bmatrix} - & \vec{u}_1^T & - \\ - & \vec{u}_2^T & - \\ & \vdots & \\ - & \vec{u}_n^T & - \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & & \\ 0 & 1 & \ddots & \\ & & 1 & \\ & & & \ddots \end{bmatrix}$$

$m < n$

these columns are repeats of first m or $\vec{0}$'s

$U =$ 

wide

If we have m orthonormal columns
→ forms a basis
No other directions in $\mathbb{R}^m$
$U^T U$ not! identity (m<n)

(b) Suppose you have a real, square, $n \times n$ orthonormal matrix U (the columns of U are unit norm and mutually orthogonal). You also have real vectors $\vec{x}_1, \vec{x}_2, \vec{y}_1, \vec{y}_2$ such that

$U\vec{x} \to \vec{y}$

$$\vec{y}_1 = U\vec{x}_1$$
$$\vec{y}_2 = U\vec{x}_2$$

$\vec{x}_1, \vec{x}_2, \vec{y}_1, \vec{y}_2 \in \mathbb{R}^n$

Calculate $\langle \vec{y}_1, \vec{y}_2 \rangle = \vec{y}_2^\top \vec{y}_1 = \vec{y}_1^\top \vec{y}_2$ in terms of $\langle \vec{x}_1, \vec{x}_2 \rangle = \vec{x}_2^\top \vec{x}_1 = \vec{x}_1^\top \vec{x}_2$.

$\langle \vec{x}_1, \vec{y}_2 \rangle = \vec{y}_2^T \vec{y}_1 = (U\vec{x}_2)^T (U\vec{x}_1)$

$= \vec{x}_2^T U^T U \vec{x}_1$
row 1xn, nxn, nxn, nx1

$= \vec{x}_2^T (I) \vec{x}_1$

$= \vec{x}_2^T \vec{x}_1 = \langle \vec{x}_2, \vec{x}_1 \rangle$

$= \langle \vec{x}_1, \vec{x}_2 \rangle$

rotation

If inner products tell us about length $\left(\sqrt{\langle \vec{v}, \vec{v} \rangle} = \|\vec{v}\|\right)$ ← length of vector $\vec{v}$

and angles between vectors
$\langle \vec{v}, \vec{w} \rangle = \|\vec{v}\| \|\vec{w}\| \cos\theta$ — angle between $\vec{v}$ & $\vec{w}$

Think: reflections or rotations

then orthonormal matrices / transformations preserve these quantities

(c) **Following the previous question, express $\|\vec{y}_1\|_2^2$ and $\|\vec{y}_2\|_2^2$ in terms of $\|\vec{x}_1\|_2^2$ and $\|\vec{x}_2\|_2^2$.**

$\langle \vec{y}_1, \vec{y}_1 \rangle = \|\vec{y}_1\|_2^2$

$\langle \vec{x}_1, \vec{x}_1 \rangle = \|\vec{x}_1\|_2^2$ &

$\|\vec{y}_2\|_2^2 = \|\vec{x}_2\|_2^2$

$\|\vec{y}_2\|_2 = \|\vec{x}_2\|_2$

$\langle \vec{x}, \vec{x} \rangle = \sum_{i=1}^{n} x_i^2$

(interpretation: U doesn't change lengths of vectors)

(d) Suppose you observe data coming from the model $y_i = \vec{a}^\top \vec{x}_i$, and you want to find the linear scale-parameters (each a_i). We are trying to learn the model $\vec{a}$. You have m data points $(\vec{x}_i, y_i)$, with each $\vec{x}_i \in \mathbb{R}^n$. Note that $\vec{x}_i$ refers to the i-th vector, not the i-th element of a single vector. Each $\vec{x}_i$ is a different input vector that you take the inner product of with $\vec{a}$, giving a scalar y_i.
Set up a least squares formulation for estimating $\vec{a}$, and find the solution to the least squares problem.

$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ want to find this

$\vec{a}^T = [a_1\ a_2 \cdots a_n]$

$\vec{x}_i = \mathbb{R}^n$

$(\vec{x}_1, y_1)$ scalar $\quad (\vec{x}_2, y_2) \cdots (\vec{x}_m, y_m)$

$y_1 = \vec{a}^T \vec{x}_1$

$y_2 = \vec{a}^T \vec{x}_2$

$y_3 = \vec{a}^T \vec{x}_3$

$\vdots$

$y_m = \vec{a}^T \vec{x}_m$

$\rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \vec{a}^T \vec{x}_1 \\ \vdots \\ \vec{a}^T \vec{x}_m \end{bmatrix} = \begin{bmatrix} \vec{x}_1^T \vec{a} \\ \vdots \\ \vec{x}_m^T \vec{a} \end{bmatrix}$

$\vec{y}$

? $\cancel{=}$ $\vec{a}^T \begin{bmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_m \\ | & | & & | \end{bmatrix} = [\ \cdots\ ]$

$= \begin{bmatrix} - \vec{x}_1^T - \\ - \vec{x}_2^T - \\ \vdots \\ - \vec{x}_m^T - \end{bmatrix} \vec{a}$ Call this X

$m \times n \qquad n \times 1$

Least squares problem:
Find $\vec{a}$ that minimizes the error norm squared:

$\| X\vec{a} - \vec{y} \|^2$

$\vec{y} = X\vec{a}$

(least square solution)

$\boxed{\hat{\vec{a}} = (X^T X)^{-1} X^T \vec{y}}$ ✓

(e) Now suppose V is an orthonormal square matrix, and rather than observing $\vec{a}^\top \vec{x}$ directly, we actually observe data points that result from our inputs being transformed by $V^\top$ as follows:

$$\vec{\tilde{x}} = V^\top \vec{x} \quad \rightarrow \quad \vec{\tilde{x}}^T = \vec{x}^T (V^T)^T = \vec{x}^T V \tag{1}$$

That is, our model acts on the modified input data $\vec{\tilde{x}}_i$, so the data points we collected are now $\left(\vec{\tilde{x}}_i, y_i\right)$. We must now consider the new model:

$$y_i = \vec{\tilde{a}}^\top \vec{\tilde{x}}_i \tag{2}$$

$$= \vec{\tilde{a}}^\top V^\top \vec{x}_i \quad (3)$$

Set up a least-squares formulation for $\hat{\vec{\tilde{a}}}$. How is $\hat{\vec{\tilde{a}}}$ related to $\hat{\vec{a}}$?

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \vec{\tilde{x}}_1^T \vec{\tilde{a}} \\ \vdots \\ \vec{\tilde{x}}_m^T \vec{\tilde{a}} \end{bmatrix} = \underbrace{\begin{bmatrix} \vec{\tilde{x}}_1^T \\ \vdots \\ \vec{\tilde{x}}_m^T \end{bmatrix}}_{\tilde{X}} \vec{\tilde{a}} = \begin{bmatrix} - \vec{x}_1^T V - \\ - \vec{x}_2^T V - \\ \vdots \\ - \vec{x}_m^T V - \end{bmatrix} \vec{\tilde{a}}$$

$$= \underbrace{\begin{bmatrix} - \vec{x}_1^T - \\ \vdots \\ - \vec{x}_m^T - \end{bmatrix}}_{X \text{ in (d)}} V \vec{\tilde{a}}$$

$\tilde{X} = XV \leftarrow$

$\hat{\vec{\tilde{a}}}$ (solution, minimizes $\|\tilde{X}\vec{\tilde{a}} - \vec{y}\|^2$)

$$\hat{\vec{\tilde{a}}} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\vec{y}$$

$$\hat{\vec{\tilde{a}}} = ((XV)^T(XV))^{-1}(XV)^T\vec{y}$$

V is a orthonormal square matrix

then V is invertible

$$= (V^TX^TXV)^{-1}V^TX^T\vec{y}$$

$$= V^{-1}\underbrace{(X^TX)^{-1}}(V^T)^{-1}V^TX^T\vec{y}$$

can compute when X's nullspace is only $\{\vec{0}\}$ or alt. X has linear independent columns

$(AB)^{-1} = B^{-1}A^{-1}$

square, invertible

$AB(B^{-1}A^{-1}) = A\underbrace{BB^{-1}}_{I}A^{-1} = I$

$V^TV = I$ (orthonormal V matrix)

$V^T = V^{-1}$

$$= V^{-1}(X^TX)^{-1}X^T\vec{y}$$

$$= V^T\underbrace{(X^TX)^{-1}X^T\vec{y}}_{\hat{\vec{a}}} = V^T\hat{\vec{a}}$$

Big picture:
When we transform our data ($\vec{x}$ vectors)
→ Least squares solution also ends up being transformed the same way

Compare $\vec{x}$'s, and $\hat{\vec{a}}$'s

$\vec{\tilde{x}} = V^T\vec{x}$

$\hat{\vec{\tilde{a}}} = V^T\hat{\vec{a}} \leftarrow$

$$V = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \ \sigma_n \\ \hline & 0_{m-n \times n} & \end{bmatrix}$$

$\vec{v}_1 \in \mathbb{R}^n$

(f) Now suppose that we have the matrix

$$U = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \\ | & | & & | \end{bmatrix} \quad \vec{u}_1 \in \mathbb{R}^m \quad \begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_m^\top \end{bmatrix} \triangleq X = U\Sigma V^\top. \tag{4}$$

$m \times m \ \cancel{m \times n}\ \cancel{n \times n} = m \times n$

where U is an $m \times m$ matrix, and V is an $n \times n$ matrix. Here, $\Sigma^{\top} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n & 0 & \cdots & 0 \end{bmatrix}$.

Here we assume that we have more data points than the dimension of our space (that is, $m > n$). Also, the transformation V in part e) is the same V in this factorized representation.

Set up a least squares formulation for estimating $\vec{\tilde{a}}$ and find the solution to the least squares. Is there anything interesting going on?

Note: Don't worry about how we would find U, Σ, $V^\top$ for now; assume that X has the given form and that U and V are orthonormal.

Hint: Start by substituting the factorized representation of X into the answer of the previous part.

$$\vec{\tilde{a}} = V^T \hat{\vec{a}} = V^T (X^T X)^{-1} X^T \vec{y} \qquad X = U\Sigma V^T$$

$$= V^T \left[(U\Sigma V^T)^T (U\Sigma V^T) \right]^{-1} (U\Sigma V^T)^T \vec{y}$$

$(V^T)^{-1} = V$

$$= V^T \left[V\Sigma^T \underbrace{U^T U}\Sigma V^T \right]^{-1} V\Sigma^T U^T \vec{y}$$

$V^{-1} = V^T$

(orthonormal square matrix)

$$= V^T (V\Sigma^T\Sigma V^T)^{-1} V\Sigma^T U^T \vec{y}$$

$$= \underbrace{V^T (V^T)^{-1}} (\Sigma^T\Sigma)^{-1} \underbrace{V^{-1} V} \Sigma^T U^T \vec{y}$$

$$= \underbrace{(\Sigma^T\Sigma)^{-1}\Sigma^T U^T} \vec{y}$$

Contributors:

- Neelesh Ramachandran.
- Kuan-Yun Lee.
- Anant Sahai.
- Kumar Krishna Agrawal.

$$(\Sigma^T\Sigma)^{-1} = \left(\left[\begin{array}{cccc|c}\sigma_1 & & & & \\ & \sigma_2 & & & 0_{n\times m-n} \\ & & \ddots & & \\ & & & \sigma_n & \end{array}\right]\left[\begin{array}{cccc}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ \hline & & 0 & \end{array}\right]\right)^{-1}$$

$$= \left(\begin{bmatrix}\sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2\end{bmatrix}\right)^{-1} = \begin{bmatrix}\frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n^2}\end{bmatrix}$$

assume $\sigma_i \neq 0$ all i's

$$\tilde{\tilde{a}} = \begin{bmatrix}\frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n^2}\end{bmatrix}\left[\begin{array}{cccc|c}\sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_n & \end{array}\right] U^T\vec{y}$$

$$= \left[\begin{array}{ccc|c}\frac{1}{\sigma_1} & & & \\ & \ddots & & 0 \\ & & \frac{1}{\sigma_n} & \end{array}\right] \underbrace{U^T\vec{y}} \rightarrow \begin{bmatrix}\vec{u}_1^T\vec{y} \\ \vdots \\ \vec{u}_m^T\vec{y}\end{bmatrix}$$

$$= \begin{bmatrix}\frac{1}{\sigma_1}\vec{u}_1^T\vec{y} \\ \vdots \\ \frac{1}{\sigma_n}\vec{u}_n^T\vec{y}\end{bmatrix}$$

Note! only inner products!

has the vibe of least squares with an orthonormal matrix A! ($A^TA = I$, $A\vec{x} \approx \vec{b}$, $\hat{\vec{x}} = A^T\vec{b}$)

But didn't assume X was orthonormal

If we have some factorization for X that gives us orthonormal matrices least squares is not computationally heavy

Q: $(\Sigma^T\Sigma)^{-1} \overset{?}{=} (\Sigma^{-1})(\Sigma^T)^{-1}$ A: No! Σ is not square!

Q: What is the name of the factorization?

$X = U \Sigma V^T$ ← SVD

$m \times n$ matrix

orthonormal

A: singular value decomposition

Q: $X^T X$ → if X or $X^T X$ has lin. dep. cols is $(X^T X)^{-1}$ computable?

A: No, need linearly independent columns of X.

$X^T X \vec{a} = X^T \vec{y}$ ← always has a solution.

BUT solution may not be unique. If so,

① To solve, find nullspace of $X^T X$

$\vec{a}_N \in N(X^T X) \rightarrow X^T X(\vec{a}_s + \vec{a}_n) = X^T X \vec{a}_s = \vec{y}$

($\vec{a}_s$ is a solution)

← find one solution

Alternatively:

② Take X, run GS on its columns / delete lin. dep. columns → get $\tilde{X}$ → Do least squares: $\tilde{\vec{a}}$

$\tilde{\vec{a}} \rightarrow \vec{a}$ by adding appropriate zeroes/transformations