EECS16B DIS14B Remote

Learning Objectives

Goal/algorithm — Quantity/object — Property of solution

① Using the complex inner product to define Gram-Schmidt Orthonormalization for complex vectors

Input: set of vectors in a specific order

output: set of vectors that span the same subspace as our input set, but is also orthonormal

② Complex projections

③ Using the complex inner product to define Least Squares solution and least squares projection for complex vectors

"Input": $A, \vec{x}, \vec{b}$ (Solve $A\vec{x} \approx \vec{b}$)

A - complex matrix of knowns

$\vec{b}$ - complex matrix of known measurements

$\vec{x}$ - unknown variables in vector, complex valued

Output: $\hat{\vec{x}}$ - least Squares solution

$\hat{\vec{b}} = \text{proj}_{col(A)}\vec{b}$ - least squares projection (projection of $\vec{b}$ onto column-space of A)

Where $\hat{\vec{x}}$ and $\hat{\vec{b}}$ satisfy

$$\min_{\vec{x}} \|A\vec{x}-\vec{b}\|^2 = \|A\hat{\vec{x}}-\vec{b}\|^2 = \|\hat{\vec{b}}-\vec{b}\|^2$$

"The smallest value of the length of error $A\vec{x}-\vec{b}$ is" achieved by $\hat{\vec{x}}$

④ General Questions on complex inner products

Mnemonic: for convention

$\boxed{\square^* = (\overline{\square})^T}$

$\langle \vec{a}, \vec{b} \rangle = \vec{b}^* \vec{a}$ (reverse order)

OR

$\langle \vec{a}, \vec{b} \rangle = \vec{a}^T (\overline{\vec{b}})$ (conjugate the second argument)

Equal because:

$$\vec{b}^* \vec{a} = (\overline{\vec{b}})^T \vec{a} = \vec{a}^T (\overline{\vec{b}})$$

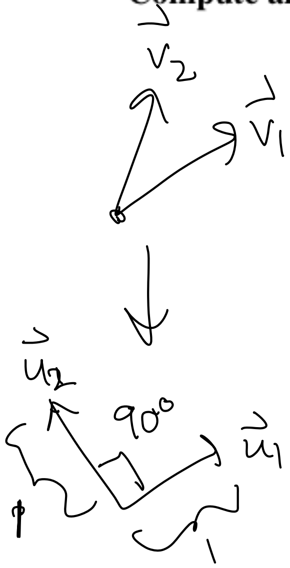
$$\vec{a}^T(\overline{\vec{b}}) = \sum_{i=1}^{n} a_i \overline{b_i}$$

1. Gram Schmidt on Complex Vectors

(a) Consider the three complex vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ j \\ 0 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \qquad \vec{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Compute an orthonormal basis from this list of vectors with Gram Schmidt.



① Normalize $\vec{v}_1 \to \vec{u}_1$

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \frac{\vec{v}_1}{\sqrt{\langle \vec{v}_1, \vec{v}_1 \rangle}}$$

norm relies on new complex inner product

$$\langle \vec{v}_1, \vec{v}_1 \rangle = \vec{v}_1^{\,*} \vec{v}_1$$

$$= \begin{bmatrix} \bar{1} \\ \bar{j} \\ \bar{0} \end{bmatrix}^T \begin{bmatrix} 1 \\ j \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ -j \\ 0 \end{pmatrix}^T \begin{pmatrix} 1 \\ j \\ 0 \end{pmatrix} = 1 + 1 = 2$$

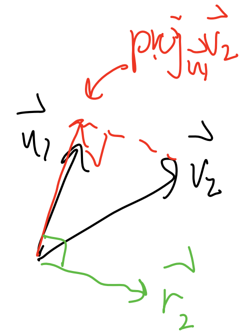
$$\vec{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ j \\ 0 \end{bmatrix}$$

② Project onto current set of orthonormal vectors and new vector ($\vec{v}_2$)
remove to get orthogonal vector.

$$\underline{\vec{r}_2 = \vec{v}_2 - \text{proj}_{\vec{u}_1} \vec{v}_2} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{j}{\sqrt{2}} \\ 0 \end{bmatrix} \underbrace{\begin{bmatrix} \frac{\bar{1}}{\sqrt{2}} & \frac{\bar{j}}{\sqrt{2}} & \bar{0} \end{bmatrix}}_{1} \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

$$\text{proj}_{\vec{u}_1} \vec{v}_2 = \frac{\vec{u}_1 \vec{u}_1^{\,*}}{\vec{u}_1^{\,*} \vec{u}_1} \vec{v}_2 = \frac{\langle \vec{v}_2, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \vec{u}_1$$

→ points along $\vec{u}_1$



Q: Why $\vec{r}_2$? A: So that $\vec{r}_2$ is $\perp$ $\vec{u}_1$ $\quad \langle \vec{r}_2, \vec{u}_1 \rangle$

$$\left\langle \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \vec{u}_1, \vec{u}_1 \right\rangle = \vec{u}_1^{\,*} \left(\vec{v}_2 - \frac{\langle \vec{v}_2, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \vec{u}_1 \right. = 0$$

$$\underline{\vec{r}_2 = \vec{v}_2 - \text{proj}_{\vec{u}_1}\vec{v}_2} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \frac{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}}{1} \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} - \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}}_{\text{projection matrix onto } \vec{u}_1} \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \left(\frac{1}{\sqrt{2}} + 0 + 0 \right)$$

$$\vec{r}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix}$$

$$\vec{u}_2 = \frac{\vec{r}_2}{\|\vec{r}_2\|} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \qquad \left\{ \vec{u}_1, \vec{u}_2 \right\} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \right\}$$

---

③ Repeat step 2: remove from new vector ($\vec{v}_3$) projections onto current orthonormal set ($\{\vec{u}_1, \vec{u}_2\}$)

$$\vec{r}_3 = \vec{v}_3 - \text{proj}_{\vec{u}_1}\vec{v}_3 - \text{proj}_{\vec{u}_2}\vec{v}_3$$

($\vec{r}_3$ is orthogonal to $\vec{u}_1$, $\vec{u}_2$)

$$\vec{u}_3 = \frac{\vec{r}_3}{\|\vec{r}_3\|} \qquad \vec{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \|\vec{u}_3\| = 1$$

(b) **Derive the least-squares solution for the case of a complex tall matrix of data and a tall matrix of values.** We want to find the best (complex) linear combination of the columns for predicting the observed values in a least-squares sense — we want to minimize the norm of the residual.

This can be formulated as having a feature matrix of data $D \in \mathbb{C}^{m \times n}$ where $m > n$ and measurements $\vec{y} \in \mathbb{C}^m$. In this case, feel free to assume that the columns of D are linearly independent even when we allow complex linear combinations. **First assume that the columns of D are orthonormal.** *Hint: You may find it useful to define a matrix* $\underline{U = \begin{bmatrix} D & D_+ \end{bmatrix}}$, *obtained via Gram-Schmidt.* Recall that the procedure here resembles that used in the SVD derivation (one portion of U contains the data/info we "care about", and the remainder is there via extension, to span the space.)

[Strategy: would like to separate quantity of interest into part we can minimize / control and error that will always exist.]

$$D = \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1n} \\ D_{21} & D_{22} & \cdots & D_{2n} \\ \vdots & & & \\ D_{m1} & \cdots & & D_{mn} \end{bmatrix} \rightarrow D_{ij} \in \mathbb{C}$$

(entries are complex #s)

(D is tall)

$D\vec{x} \simeq \vec{y}$ ← has complex #'s

has complex #'s

How to find $\vec{x}$ w/ complex values such that a solution $\vec{x}$ makes $\underbrace{\|D\vec{x} - \vec{y}\|}_{\text{error}}{}^2$ as small as possible

Real case:

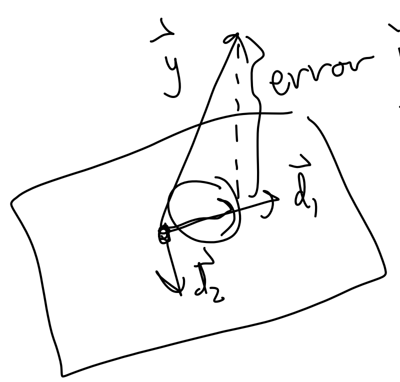
D has orthonormal cols: $D^T D = I$

Complex case:

D has orthonormal cols: $D^* D = I$ ← take complex inner products of cols of D

$$D^* D = \begin{bmatrix} - \vec{d}_1^{\,*} - \\ \vdots \\ - \vec{d}_n^{\,*} - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{d}_1 & \cdots & \vec{d}_n \\ | & & | \end{bmatrix} \rightarrow \vec{d}_i^{\,*} \vec{d}_j = \langle \vec{d}_j, \vec{d}_i \rangle$$

$$\rightarrow \underset{m \times m}{U^*} \underset{m \times m}{U} = I \qquad \underset{m \times m}{U} \underset{m \times m}{U^*} = I \qquad (DD^* \neq I \text{ because } D \text{ is not a full basis for } \mathbb{C}^m)$$



error is minimized w/ the choice that makes it orthogonal

$$\min_{\vec{x}} \|D\vec{x}-\vec{y}\|^2 \qquad U = \left[\, D \;\middle|\; D_\perp \,\right]$$

$$\|D\vec{x}-\vec{y}\|^2 = \langle D\vec{x}-\vec{y}, D\vec{x}-\vec{y}\rangle$$

$UU^* = U^*U = I$

$$\|U(\underbrace{D\vec{x}-\vec{y}}_{\vec{e}})\|^2 = \langle U\vec{e}, U\vec{e}\rangle \qquad (AB)^* = B^*A^*$$

$$= (U\vec{e})^* U\vec{e}$$

$$= \vec{e}^{\,*} \underbrace{U^* U}\vec{e}$$

Orthonormal matrix U doesn't change norm

$$= \vec{e}^{\,*}\vec{e}$$

$$= \langle \vec{e}, \vec{e}\rangle = \|\vec{e}\|^2 = \|D\vec{x}-\vec{y}\|^2$$

Q: Why U, U^* keep length the same

$$\|U^*(D\vec{x}-\vec{y})\|^2 = \left\| \begin{bmatrix} D^* \\ \hline D_\perp^* \end{bmatrix} (D\vec{x}-\vec{y}) \right\|^2$$

Assumed

$$= \left\| \begin{bmatrix} D^*D\vec{x} \\ D_\perp^* D\vec{x} \end{bmatrix} - \begin{bmatrix} D^*\vec{y} \\ D_\perp^*\vec{y} \end{bmatrix} \right\|^2$$

$D^*D = I$

$D_\perp \rightarrow$ comes from GS w/ D

$D_\perp^* D = 0$ (matrix) $\leftarrow$ helps span $\mathbb{C}^m$

$$\left\|\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\|^2 = 2^2 + 1^2$$

$$= \left\| \begin{bmatrix} \vec{x} \\ \vec{0} \end{bmatrix} - \begin{bmatrix} D^*\vec{y} \\ D_\perp^*\vec{y} \end{bmatrix} \right\|^2$$

$$\left\|\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}\right\|^2 = \left\|\begin{bmatrix} 2 \\ 2 \end{bmatrix}\right\|^2 + 1^2$$

9

$$= \left\| \begin{bmatrix} \vec{x} - D^*\vec{y} \\ -D_\perp^*\vec{y} \end{bmatrix} \right\|^2$$

top n entries affected by $\vec{x}$

bottom $m-n$ entries are not affected by $\vec{x}$

$$= \underbrace{\|\vec{x} - D^*\vec{y}\|^2 + \|-D_\perp^*\vec{y}\|^2}$$

$\Rightarrow$ Choose $\vec{x} = D^*\vec{y}$ to make above quantity as small as possible

$$\boxed{\hat{\vec{x}} = D^*\vec{y}} \leftarrow D \text{ was orthonormal}$$

(least squares solution)

$\rightarrow \hat{\vec{x}} = (D^TD)^{-1}D^T\vec{y}$ real case

(c) **Repeat the previous part without the assumption of orthonormality for the columns of D.** You can keep the assumption of linear independence.

☆ This is a different approach than in the solutions. Read both!

GS makes D_+ from D

$D = \begin{bmatrix} | & & | \\ d_1 & \cdots & d_n \\ | & & | \end{bmatrix}$

GS on $\{\vec{d}_1, \ldots, \vec{d}_n, \vec{e}_1, \ldots, \vec{e}_m\} \hookrightarrow [D \; D^+]$

$\vec{e}_i$ = vector w/ i-th entry 1

$$GS\left(\left[D \mid I\right]\right) \rightarrow \overset{m \times m}{\left[Q \mid D_+\right]}$$

$$D = QR$$

Q has the same column span as D but has orthonormal columns

$$Q^* Q = I$$

R is a matrix that encodes (complex) linear combinations used in GS

$$\min_{\vec{x}} \|D\vec{x} - \vec{y}\|^2$$

$$\|QR\vec{x} - \vec{y}\|^2$$

$$= \left\|\begin{bmatrix} Q^* \\ \hline D_+^* \end{bmatrix}(QR\vec{x} - \vec{y})\right\|^2$$

$$= \left\|\begin{bmatrix} R\vec{x} \\ \vec{0} \end{bmatrix} - \begin{bmatrix} Q^*\vec{y} \\ D_+^*\vec{y} \end{bmatrix}\right\|^2$$

$$= \|R\vec{x} - Q^*\vec{y}\|^2 + \|-D_+^*\vec{y}\|^2$$

minimize by choosing $R\vec{x} = Q^*\vec{y}$

$$\vec{x} = \underbrace{R^{-1}Q^*}\vec{y}$$

How to get this?

Check if $(D^*D)^{-1}D^*$ gives

$$(D^*D)^{-1}D^* = ((QR)^*QR)^{-1}(QR)^*$$

$$= (R^* \cancel{Q^*Q}^{\,I} R)^{-1} R^* Q^*$$

$$= (R^*R)^{-1} R^* Q^*$$

$$= R^{-1}(R^*)^{-1} R^* Q^*$$

$$= R^{-1} Q^* \checkmark$$

So: (when D has non orthonormal columns)

$$\boxed{\vec{x} = (D^*D)^{-1} D^* \vec{y}}$$

Questions from 14A, 14B

2. Q&A time! [≈ 20 minutes]

This time is here for you all to ask any questions from discussion 14A and 14B to the TAs to review the material on complex vectors. If there are no further questions, then feel free to discuss anything else related to the course content, as this is the last non-review discussion.

Q: Why is $U^* = \begin{bmatrix} D^* \\ \hline D_+^* \end{bmatrix}$

$$U = \left[\begin{array}{ccc|ccc} | & | & & | & | & \\ \vec{d}_1 & \vec{d}_2 \cdots & \vec{d}_n & \vec{d}_{n+1} & \cdots & \vec{d}_m \\ | & | & & | & | & \end{array}\right]$$

D D_+

$$U^* = \begin{bmatrix} - \bar{\vec{d}}_1^{\,T} - \\ - \bar{\vec{d}}_2^{\,T} - \\ \vdots \\ - \bar{\vec{d}}_m^{\,T} - \end{bmatrix} = \begin{bmatrix} D^* \\ \hline D_+^* \end{bmatrix} \begin{array}{l} \leftarrow n \times m \\ \leftarrow (m-n) \times m \end{array}$$

transpose + conjugate

Q: Is $UU^* = I$ if U is orthonormal in the complex case? and square

U^*U vs. UU^*

$$\begin{bmatrix} \langle \vec{u}_1, \vec{u}_1 \rangle & \cdots & \langle \vec{u}_m, \vec{u}_1 \rangle \\ \vdots & & \vdots \\ \langle \vec{u}_1, \vec{u}_m \rangle & \cdots & \langle \vec{u}_m, \vec{u}_m \rangle \end{bmatrix}$$

A: $U^*U = I$

$UU^*U = U \cdot I$

$\underbrace{(UU^*)}_{\to\, I \text{ has to be true}} U = U$

Q: If U orthonormal and tall is $UU^* = I$ and is $U^*U = I$?

A: From real matrices know for sure that $UU^* \neq I$

$U^*U = I$ ✓ boils down to UU^T when real

$UU^* \neq I$ when U tall orthonormal

When rows of wide U are orthonormal, $UU^* = I$ $U^*U \neq I$

Q: When is upper triangularization the same as SVD? (real matrices)

A: My guess: when A, a matrix is symmetric ($A = A^T$)

$A = V \Lambda V^T \Leftarrow$ upper triangularization for symmetric A is it's diagonalization

(not true)

SVD: $A^T A = V \Lambda^2 V^T \Rightarrow \sigma_i = \sqrt{\lambda_i^2} = |\lambda_i|$

left singular vectors being $\vec{v}_i$

$$U = \begin{bmatrix} | & & | \\ \text{sign}(\lambda_1)\vec{v}_1 & \cdots & \text{sign}(\lambda_n)\vec{v}_n \\ | & & | \end{bmatrix}$$

$\Rightarrow$ Guess

$\boxed{\text{Maybe a chance when } A = A^T \text{ and } A \text{ has non-negative eigenvalues}}$

Q: $\langle \vec{a}, \vec{b} \rangle \stackrel{?}{=} \langle \vec{b}, \vec{a} \rangle$

$\vec{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$

A: $\langle \vec{a}, \vec{b} \rangle = \vec{b}^* \vec{a} = \sum_{i=1}^{n} a_i \bar{b}_i$

$\langle \vec{b}, \vec{a} \rangle = \vec{a}^* \vec{b} = \sum_{i=1}^{n} \bar{a}_i b_i$

$$\overline{\langle \vec{a}, \vec{b} \rangle} = \overline{\left(\sum_{i=1}^{n} a_i \bar{b}_i\right)} = \sum_{i=1}^{n} \overline{a_i \bar{b}_i} = \sum_{i=1}^{n} \bar{a}_i \underbrace{\bar{\bar{b}}_i}_{b_i} = \sum_{i=1}^{n} \bar{a}_i b_i = \langle \vec{b}, \vec{a} \rangle$$

Q: When to use triangularization vs. SVD?

A: Upper triangularization $\rightarrow$ wanted to consider DE with a matrix that wasn't diagonalizable (still want benefit of simpler nested problems)

SVD $\rightarrow$ Minimum energy control / understanding subspaces of matrix A and possible solutions to $A\vec{x} = \vec{y}$