EECS 16B Designing Information Systems and Devices II UC Berkeley Fall 2022

Final

Exam Location: Draft

PRINT your student ID: ______________________________

PRINT AND SIGN your name: ______________, ______________ ______________
(last) (first) (sign)

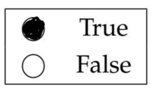
PRINT your discussion sections and (u)GSIs (the ones you attend): ______________

Row Number: ______________ Seat Number: ______________

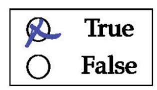
Name and SID of the person to your left: ______________________________

Name and SID of the person to your right: ______________________________

When answering multiple choice questions, completely fill in the bubble next to your answer choice. See Figure 1 for examples of correctly filled and incorrectly filled bubbles.



(a) Correctly filled bubble



(b) Incorrectly filled bubble

Figure 1: Multiple Choice Bubbles

1. **Honor Code (0 pts.)**

   Please copy the following statement in the space provided below and sign your name.

   As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. I will follow the rules and do this exam on my own.

   Note that if you do not copy the honor code and sign your name, you will get a 0 on the exam.

   Solution: Any attempt to copy the honor code and sign should get full points.

2. **What are you planning to do during your winter break? (2 pts.)**

   Solution: Any answer is sufficient.

3. **What is the most underrated fruit? (2 pts.)**

   Solution: Any answer is sufficient.

Do not turn this page until the proctor tells you to do so.
You can work on the above problems before time starts.

4. Circuits True/False (20 pts.)

Indicate True/False for each of the following statements.

(a) (2 pts.) The current of a capacitor can change instantaneously.

○ True
○ False

Solution:

● True
○ False

The current across a capacitor can change instantaneously. It is the the voltage across a capacitor that cannot change instantaneously due to $i_C = C\frac{dv_C}{dt}$.

(b) (2 pts.) The current of an inductor can change instantaneously.

○ True
○ False

Solution:

○ True
● False

The current across an inductor cannot change instantaneously since that would require infinite voltage $v_L = L\frac{di_L}{dt}$.

(c) (2 pts.) When all voltages are turned OFF, the voltage across a capacitor in a series R-C circuit linearly goes down to zero.

○ True
○ False

Solution:

○ True
● False

The voltage across a capacitor in a discharging RC circuit will decay exponentially, not linearly.

(d) (2 pts.) For a faster charging of a capacitor, one should increase the capacitor area.

○ True
○ False

Solution:

○ True
● False

$C = \frac{\epsilon A}{d}$, so increasing the area of a capacitor means that it requires more charge to reach the same voltage. Therefore, greater area will cause the capacitor to charge slower.

(e) (2 pts.) The frequency of the voltage across a capacitor in a RLC circuit is independent of R, where $R, L, C \neq 0$.

◯	True
◯	False

Solution:

•	True
◯	False

The frequency of the voltage across the capacitor in an RLC circuit only depends on the frequency of the input voltage or current source.

(f) (2 pts.) The steady state voltage of a capacitor in a series RLC circuit is always smaller than the supply voltage.

◯	True
◯	False

Solution:

◯	True
•	False

At steady state, the voltage across the capacitor will be equal to the supply voltage.

(g) (2 pts.) Voltage across the capacitor in a series RC circuit leads the current by $90°$.

◯	True
◯	False

Solution:

◯	True
•	False

Capacitances cause current to lead voltage by $90°$. If V_C is the phasor voltage across the capacitor, then the phasor current is $I_C = \frac{V_C}{\frac{1}{j\omega C}} = j\omega C V_C$ which means that I_C leads V_C by j, or $90°$.

(h) (2 pts.) Current through the inductor in a series RL circuit leads the voltage by $90°$.

◯	True
◯	False

Solution:

◯	True
•	False

Inductances cause current to lag behind voltage by $90°$. If V_L is the phasor voltage across the inductor, then the phasor current is $I_L = \frac{V_L}{j\omega L} = -j\frac{V_L}{\omega L}$ which means I_L lags V_L by j, or $90°$.

(i) (2 pts.) The power dissipation in a series LC circuit is zero.

◯	True
◯	False

Solution:

•	True
◯	False

An LC circuit is purely reactive and therefore will not dissipate any power.

(j) (2 pts.) It is possible to have the same circuit behave as a high pass or low pass filter simply by changing the two ports from where output is taken.

◯ True
◯ False

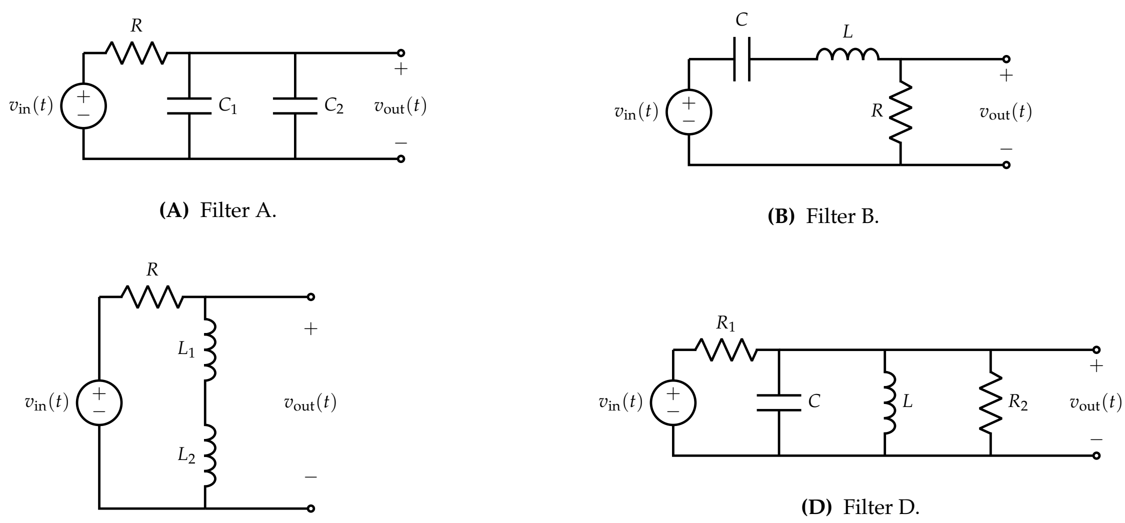
Solution:

● True
◯ False

Take a standard RC circuit as an example. If your output voltage is v_R, then you will have a highpass filter. If your output voltage is v_C, you will get a lowpass filter.

5. Transfer Functions Again... (20 pts.)

(a) (4 pts.) Below, you are given 4 filter circuits A, B, C, D. **Fill in the bubbles to match each filter to its corresponding transfer function form (I, II, III, or None).**



(A) Filter A.

(B) Filter B.

(C) Filter C.

(D) Filter D.

Figure 2: Various Filter Circuits

Transfer function forms:

$$\textbf{TF I:}\ H(f) = \frac{1}{1 + \mathrm{j}\frac{f}{f_0}} \tag{1}$$

$$\textbf{TF II:}\ H(f) = \frac{\mathrm{j}\frac{f}{f_0}}{1 + \mathrm{j}\frac{f}{f_0}} \tag{2}$$

$$\textbf{TF III:}\ H(f) = \frac{1}{k_1 + \mathrm{j}k_2\left(\frac{f}{f_0} - \frac{f_0}{f}\right)} \tag{3}$$

$$\tag{4}$$

Assume k_1 and k_2 are constants that depend on the values of the circuit components.

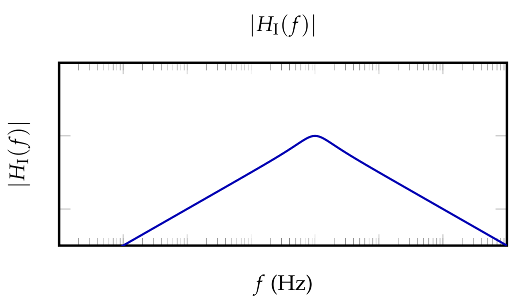
Filter Letter	TF I	TF II	TF III	None
A	○	○	○	○
B	○	○	○	○
C	○	○	○	○
D	○	○	○	○

Solution:

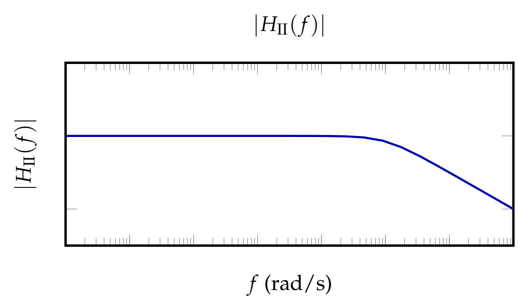
Filter Letter	TF I	TF II	TF III	None
A	●	○	○	○
B	○	○	●	○
C	○	●	○	○
D	○	○	●	○

Circuit A is our standard RC lowpass filter if you combine C_1 and C_2 into an equivalent capacitance, which will take on the form of TF I.
Circuit B is a series resonance circuit so will take on the form of TF III.
Circuit C is a standard RL highpass filter if you combine L_1 and L_2 into an equivalent inductance, which has the form of TF II.
Circuit D is the same parallel resonance circuit from the midterm, which we saw had the transfer function of $H(f) = \frac{1}{k_1 + jk_2\left(\frac{f}{f_0} - \frac{f_0}{f}\right)}$ which is equivalent to TF III.

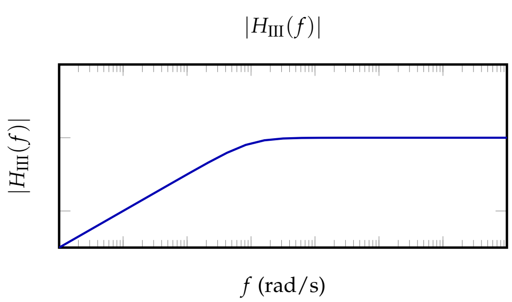
(b) (4 pts.) **Now, using the same 4 filter circuits (A, B, C, D) from 2, fill in the bubbles to match each filter to its corresponding magnitude Bode Plot out of choices I, II, III, IV.**



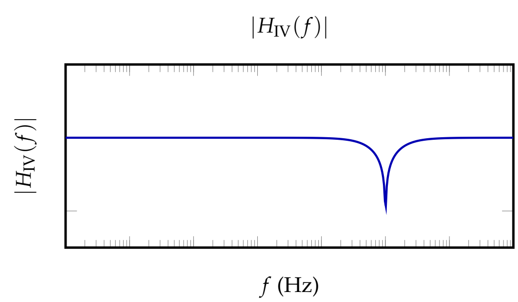
(a) Plot I.



(b) Plot II.



(c) Plot III.



(d) Plot IV.

Filter Letter	Plot I	Plot II	Plot III	Plot IV
A	○	○	○	○
B	○	○	○	○
C	○	○	○	○
D	○	○	○	○

Solution:

Filter Letter	Plot I	Plot II	Plot III	Plot IV
A	○	●	○	○
B	●	○	○	○
C	○	○	●	○
D	●	○	○	○

Circuit A is a lowpass filter and Circuit C is a highpass. The series and parallel resonance circuits both act as bandpass filters and will take on a shape similar to Plot I.

For the remaining parts of the problem, suppose you are given the following circuit in 4.

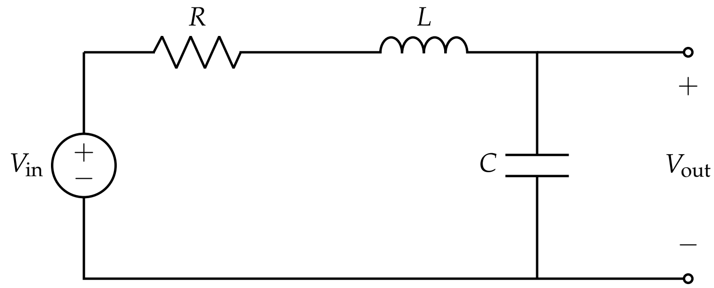


Figure 4: RLC Filter

(c) (8 pts.) **Solve for the transfer function $H(f) = \frac{V_{\text{out}}}{V_{\text{in}}}$ of the circuit from fig. 4 in the form of**

$$H(f) = \frac{-\mathrm{j}\alpha\frac{f_0}{f}}{1 + \mathrm{j}\beta\left(\frac{f}{f_0} - \frac{f_0}{f}\right)} \tag{5}$$

where α and β are positive constants that may be in terms of R, L, C, f_0, Q_s and/or other constants. *(HINT: Recall that for a series resonant circuit, we have $Q_s = \frac{2\pi f_0 L}{R} = \frac{1}{2\pi f_0 RC}$ and that $f_0 = \frac{1}{2\pi\sqrt{LC}}$ where Q_s is the quality factor and f_0 is the resonant frequency.)*

Solution: As stated in the hint, we have a series resonance circuit, so the total impedance can be calculated from:

$$\begin{aligned} Z_s &= Z_R + Z_L + Z_C & (6)\\ &= R + \mathrm{j}2\pi f L + \frac{1}{\mathrm{j}2\pi f C} & (7)\\ &= R + \mathrm{j}\left((2\pi L)f - \left(\frac{1}{2\pi C}\right)\frac{1}{f}\right) & (8) \end{aligned}$$

Using the provided quality factors for a series resonant circuit, we have:

$$\begin{aligned} Q_s &= \frac{2\pi f_0 L}{R} \longrightarrow 2\pi L = \frac{Q_s R}{f_0} & (9)\\ Q_s &= \frac{1}{2\pi f_0 RC} \longrightarrow \frac{1}{2\pi C} = Q_s f_0 R & (10) \end{aligned}$$

Plugging these back into our impedance equation, we get:

$$\begin{aligned} Z_s &= R + \mathrm{j}\left((2\pi L)f - \left(\frac{1}{2\pi C}\right)\frac{1}{f}\right) & (11)\\ &= R + \mathrm{j}\left(\left(\frac{Q_s R}{f_0}\right)f - (Q_s f_0 R)\frac{1}{f}\right) & (12)\\ &= R + \mathrm{j}Q_s R\left(\frac{f}{f_0} - \frac{f_0}{f}\right) & (13)\\ Z_s &= R\left(1 + \mathrm{j}Q_s\left(\frac{f}{f_0} - \frac{f_0}{f}\right)\right) & (14) \end{aligned}$$

You could also directly use the equivalent impedance of a series resonant circuit without rederiving it from scratch. Now, we can apply the voltage divider principle to calculate our transfer function.

$$H(f) = \frac{Z_C}{Z_s} \tag{15}$$

$$= \frac{\frac{1}{\mathrm{j}2\pi f C}}{R\left(1 + \mathrm{j}Q_s\left(\frac{f}{f_0} - \frac{f_0}{f}\right)\right)} \tag{16}$$

$$= \frac{-\mathrm{j}\frac{1}{2\pi RC}\frac{1}{f}}{1 + \mathrm{j}Q_s\left(\frac{f}{f_0} - \frac{f_0}{f}\right)} \tag{17}$$

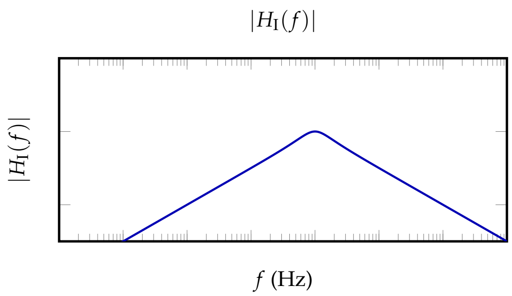
$$= \frac{-\mathrm{j}\frac{1}{2\pi f_0 RC}\frac{f_0}{f}}{1 + \mathrm{j}Q_s\left(\frac{f}{f_0} - \frac{f_0}{f}\right)} \tag{18}$$

Matching this to the form provided, we have

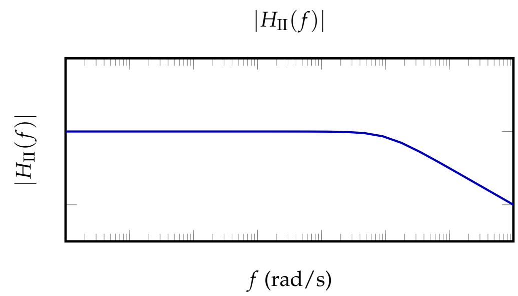
$$\alpha = \frac{1}{2\pi f_0 RC} = Q_s = \frac{2\pi f_0 L}{R} \tag{19}$$

$$\beta = Q_s = \frac{1}{2\pi f_0 RC} = \frac{2\pi f_0 L}{R} \tag{20}$$

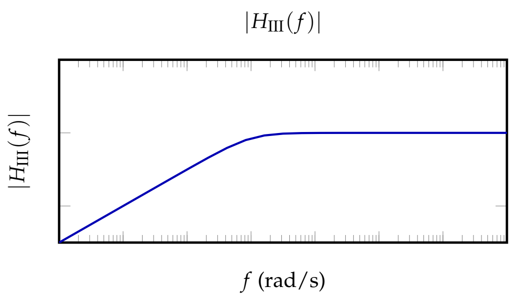
(d) (2 pts.) **Now, match the magnitude of the transfer function $H(f)$ from fig. 4 with the correct bode plot shape below.** You do not need to have successfully calculated values for α and β in item 5.c to solve this subpart.



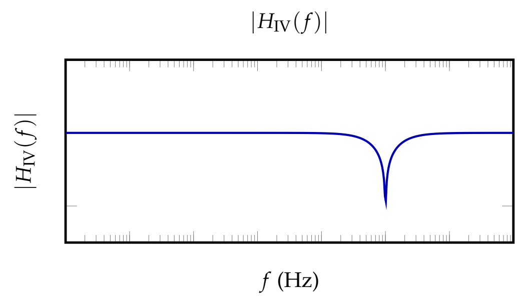
(a) Plot I.



(b) Plot II.



(c) Plot III.



(d) Plot IV.

○	Plot I
○	Plot II
○	Plot III
○	Plot IV

Solution:

○	Plot I
●	Plot II
○	Plot III
○	Plot IV

We can take limits on the transfer function to see the behavior of the magnitude of the transfer function at high and low frequencies.

$$\lim_{f\to 0} H(f) = \lim_{f\to 0} \frac{-\mathrm{j}\alpha\frac{f_0}{f}}{1+\mathrm{j}\beta\left(\frac{f}{f_0}-\frac{f_0}{f}\right)} \tag{21}$$

$$\approx \lim_{f\to 0} \frac{-\mathrm{j}\alpha\frac{f_0}{f}}{\mathrm{j}\beta\left(0-\frac{f_0}{f}\right)} \tag{22}$$

$$= \lim_{f\to 0} \frac{-\mathrm{j}\alpha\frac{f_0}{f}}{-\mathrm{j}\beta\frac{f_0}{f}} \tag{23}$$

$$= \frac{\alpha}{\beta} \lim_{f\to 0} |H(f)| = \left|\frac{\alpha}{\beta}\right| \tag{24}$$

As $f \to 0$ we see that the magnitude approaches some non-zero constant.

$$\lim_{f\to\infty} H(f) = \lim_{f\to\infty} \frac{-\mathrm{j}\alpha \frac{f_0}{f}}{1+\mathrm{j}\beta\left(\frac{f}{f_0}-\frac{f_0}{f}\right)} \tag{25}$$

$$\approx \frac{0}{1+\mathrm{j}\beta(\infty-0)} \tag{26}$$

$$= 0 \tag{27}$$

$$\lim_{f\to\infty} |H(f)| = 0 \tag{28}$$

We see that at high frequencies the magnitude our transfer function goes to 0, while at low frequencies it approaches some non-zero constant. Therefore, we have some zero low-pass filter which matches Plot IV.

(e) (2 pts.) Suppose we could choose what the values of α and β are by choosing specific values of R, L, and C. **How would a larger α impact $|H(f_0)|$, the magnitude of $H(f)$ at resonance.** Once again, you do not need to have successfully calculated values for α and β in item 5.c to solve this subpart.

◯	$\lvert H(f_0)\rvert$ increase as α increases
◯	$\lvert H(f_0)\rvert$ decreases as α increases
◯	$\lvert H(f_0)\rvert$ stays the same as α increases

Solution:

●	$\lvert H(f_0)\rvert$ increase as α increases
◯	$\lvert H(f_0)\rvert$ decreases as α increases
◯	$\lvert H(f_0)\rvert$ stays the same as α increases

Notice that when $f = f_0$, we have:

$$H(f_0) = \frac{-\mathrm{j}\alpha \frac{f_0}{f_0}}{1+\mathrm{j}\beta\left(\frac{f_0}{f_0}-\frac{f_0}{f_0}\right)} \tag{29}$$

$$= \frac{-\mathrm{j}\alpha}{1+\mathrm{j}\beta(1-1)} \tag{30}$$

$$= -\mathrm{j}\alpha \tag{31}$$

$$|H(f_0)| = |-\mathrm{j}\alpha| \tag{32}$$

$$= |\alpha| \tag{33}$$

Therefore, we see that $|H(f_0)|$ is directly proportional to the magnitude of α.

6. Systems and Linear Algebra True/False (16 pts.)

Indicate True/False for each of the following statements.

(a) (2 pts.) If the matrix A in the discrete-time linear system

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{34}$$

has eigenvalues with real parts strictly negative, then the system is always stable.

◯ True
◯ False
◯ Not Enough Information

Solution:

◯ True
● False
◯ Not Enough Information

The stability criterion for discrete-time linear systems is that $|\lambda_A| < 1$, where λ_A represents the eigenvalues of A. However, we can have an eigenvalue with $|\lambda| > 1$ and $\text{Re}\{\lambda\} < 0$, such as $\lambda = -10$. Clearly, the system is not always stable.

(b) (2 pts.) If a matrix $A \in \mathbb{R}^{n\times n}$ has all distinct eigenvalues, then it is diagonalizable.

◯ True
◯ False
◯ Not Enough Information

Solution:

● True
◯ False
◯ Not Enough Information

If a matrix has distinct eigenvalues, all of the eigenvectors must be linearly independent.

(c) (2 pts.) If a matrix $A \in \mathbb{R}^{n\times n}$ has some non-distinct eigenvalues, then it is diagonalizable.

◯ True
◯ False
◯ Not Enough Information

Solution:

◯ True
◯ False
● Not Enough Information

The identity matrix is clearly diagonalizable, but it has 1 as a repeated eigenvalue. However, the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is not diagonalizable despite also having 1 as a repeated eigenvalue.

(d) (2 pts.) If a linear system can be transformed to controllable canonical form, then it is controllable.

◯	True
◯	False
◯	Not Enough Information

Solution:

•	True
◯	False
◯	Not Enough Information

We define our transformation matrix as $T = \mathcal{C}\mathcal{C}_{\text{CCF}}^{-1}$, and if T needs to be full rank (i.e., invertible), then $\mathcal{C}$ has to be full rank.

(e) (2 pts.) Any square matrix can be upper-triangularized.

◯	True
◯	False
◯	Not Enough Information

Solution:

•	True
◯	False
◯	Not Enough Information

The proof for the Schur Decomposition did not make any assumptions about the matrix that you are upper triangularizing.

(f) (2 pts.) If $A \in \mathbb{R}^{n\times n}$ and

$$A = URU^\top \tag{35}$$

is a valid upper triangularization of A from the Schur Decomposition algorithm, then U is an orthonormal matrix.

◯	True
◯	False
◯	Not Enough Information

Solution:

•	True
◯	False
◯	Not Enough Information

By the construction of the Schur Decomposition, U is an orthonormal matrix and R is an upper triangular matrix.

(g) (2 pts.) The largest singular value of any matrix is always strictly positive.

◯	True
◯	False
◯	Not Enough Information

Solution:

◯ True
● False
◯ Not Enough Information

Consider a matrix of all zeros – this matrix does not have any singular values that are strictly positive.

(h) (2 pts.) For the discrete-time linear system

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{36}$$

the controllability matrix will not change rank if you append $A^n B$, i.e., $\text{rank}(\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}) = \text{rank}(\begin{bmatrix} B & AB & \cdots & A^{n-1}B & A^n B \end{bmatrix})$. You may assume that $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$.

◯ True
◯ False
◯ Not Enough Information

Solution:

● True
◯ False
◯ Not Enough Information

This holds from the Cayley-Hamilton Theorem.

7. Gram-Schmidt and QR Decomposition (20 pts.)

Consider the following set of vectors:

$$\left\{\vec{d}_1, \vec{d}_2, \vec{d}_3\right\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \tag{37}$$

Perform Gram-Schmidt orthonormalization to obtain an orthonormal set of vectors $\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}$ such that

$$\text{Span}\left(\vec{d}_1, \vec{d}_2, \vec{d}_3\right) = \text{Span}(\vec{q}_1, \vec{q}_2, \vec{q}_3) \tag{38}$$

(a) (4 pts.) Find $\vec{q}_1$ such that $\text{Span}(\vec{q}_1) = \text{Span}\left(\vec{d}_1\right)$ and $\|\vec{q}_1\| = 1$.

○	$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$
○	$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$
○	$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$
○	$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

Solution:

○	$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$
●	$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$
○	$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$
○	$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

We have that

$$\vec{q}_1 = \frac{\vec{d}_1}{\left\|\vec{d}_1\right\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \tag{39}$$

(b) (6 pts.) Find $\vec{q}_2$ such that $\text{Span}(\vec{q}_1, \vec{q}_2) = \text{Span}\left(\vec{d}_1, \vec{d}_2\right)$ and $\vec{q}_2$ is orthonormal with respect to $\vec{q}_1$.

○ $\frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix}$
○ $\frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix}$
○ $\begin{bmatrix}0\\0\\0\end{bmatrix}$
○ $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}$

Solution:

○ $\frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix}$
● $\frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix}$
○ $\begin{bmatrix}0\\0\\0\end{bmatrix}$
○ $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}$

We have that

$$\vec{z}_2 = \vec{d}_2 - \left(\vec{d}_2^\top \vec{q}_1\right)\vec{q}_1 \tag{40}$$

$$= \begin{bmatrix}1\\1\\0\end{bmatrix} - \left(\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right)^\top \left(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}\right)\right)\left(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}\right) \tag{41}$$

$$= \begin{bmatrix}\frac{1}{2}\\1\\\frac{1}{2}\end{bmatrix} \tag{42}$$

Normalizing, we get

$$\vec{q}_2 = \frac{\vec{z}_2}{\|\vec{z}_2\|} = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix} \tag{43}$$

(c) (6 pts.) Find $\vec{q}_3$ such that $\text{Span}(\vec{q}_1, \vec{q}_2, \vec{q}_3) = \text{Span}\left(\vec{d}_1, \vec{d}_2, \vec{d}_3\right)$ and $\vec{q}_3$ is orthonormal with respect to $\vec{q}_1$ and $\vec{q}_2$.

○ $\frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix}$
○ $\frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix}$
○ $\begin{bmatrix}0\\0\\0\end{bmatrix}$
○ $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}$

Solution:

● $\frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix}$
○ $\frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix}$
○ $\begin{bmatrix}0\\0\\0\end{bmatrix}$
○ $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}$

We have that

$$\vec{z}_3 = \vec{d}_3 - \left(\vec{d}_3^\top \vec{q}_1\right)\vec{q}_1 - \left(\vec{d}_3^\top \vec{q}_2\right)\vec{q}_2 \tag{44}$$

$$= \begin{bmatrix}0\\0\\1\end{bmatrix} - \left(\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right)^\top \left(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}\right)\right)\left(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}\right) - \left(\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right)^\top \left(\frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix}\right)\right)\left(\frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix}\right) \tag{45}$$

$$= \begin{bmatrix}\frac{1}{3}\\-\frac{1}{3}\\\frac{1}{3}\end{bmatrix} \tag{46}$$

Normalizing, we get

$$\vec{q}_3 = \frac{\vec{z}_3}{\|\vec{z}_3\|} = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix} \tag{47}$$

(d) (4 pts.) Suppose you are given the following matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \tag{48}$$

What is the QR decomposition of A?

○ $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}$ $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$

○ $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$

○ $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$

○ $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix}$

Solution:

○ $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}$ $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$

● $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$

○ $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$

○ $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix}$

From the previous parts, we have

$$Q = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \vec{q}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \tag{49}$$

and if we compute $R_{ij} = \vec{q}_i^\top \vec{d}_j$, we will get

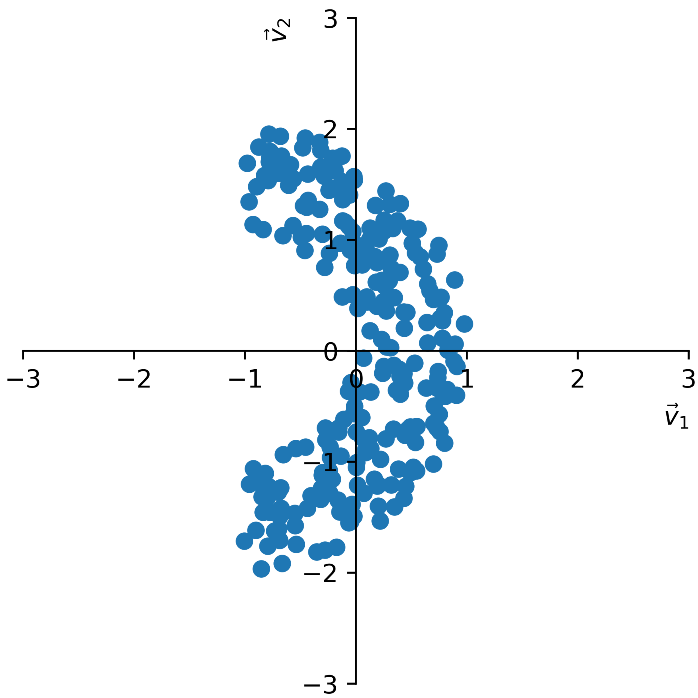
$$R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \tag{50}$$

8. PCA Plots (20 pts.)

In each plot below, some d-dimensional data is projected onto two unit vectors. The x-coordinate is the projection onto the first vector (written as $\vec{v}_1$), and the y-coordinate is the projection onto the econd vector (written as $\vec{v}_2$). Mathematically, we can say that we have some data matrix $D = \begin{bmatrix} \vec{d}_1 & \vec{d}_2 & \cdots & \vec{d}_n \end{bmatrix}$ where each $\vec{d}_i \in \mathbb{R}^d$. We then project each $\vec{d}_i$ onto the column space of the matrix $V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$, and we plot the projection coefficients below. We say that a plot is "valid" if $\vec{v}_1$ could be the first principal component.

(HINT: Note that the mean or "center of mass" of the data points in all the plots are the origin, (0, 0).) (HINT: The procedure for this problem is very similar to the procedure in Discussion 12B, where you were similarly asked to judge the "validity" of scatterplots.)

(a) (5 pts.) **Is the following plot valid?**



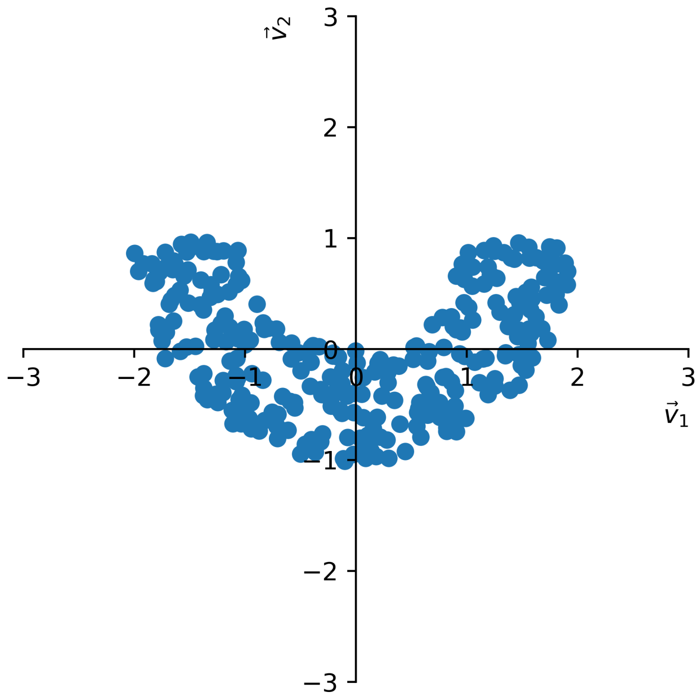
◯	Valid
◯	Invalid

Solution:

◯	Valid
•	Invalid

In this case, the coordinates on $\vec{v}_2$ axis seems to have a larger sum of squares than the $\vec{v}_1$ axis, since the coordinates on the $\vec{v}_2$ axis range from -2 to 2.

(b) (5 pts.) **Is the following plot valid?**



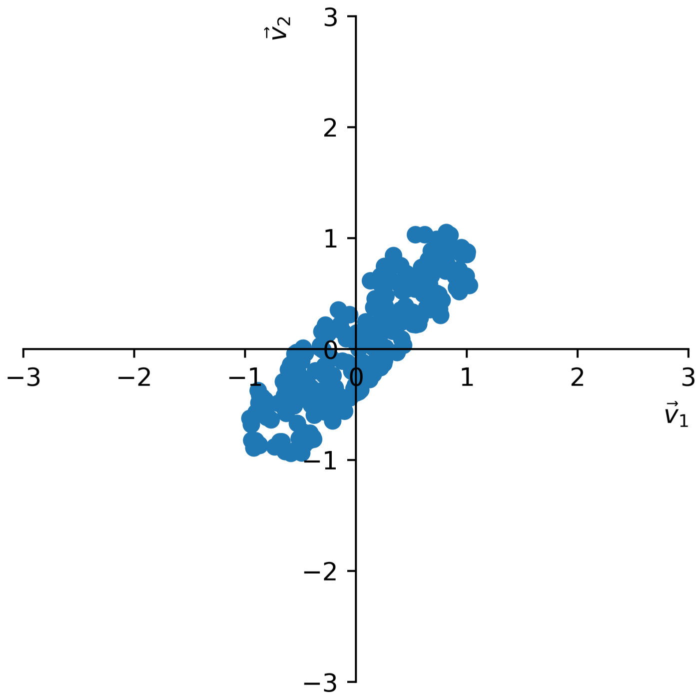
◯	Valid
◯	Invalid

Solution:

●	Valid
◯	Invalid

In this case, the coordinates on the $\vec{v}_1$ axis seem to maximize the sum of squares (i.e., there is no other axis that we can construct which will have a larger sum of squares of coordinates). Another way to approach this problem would be to notice that the data is centered around the origin, so the first principal component will be the direction of greatest "spread" in the data, which is the $\vec{v}_1$ direction.

(c) (5 pts.) **Is the following plot valid?**



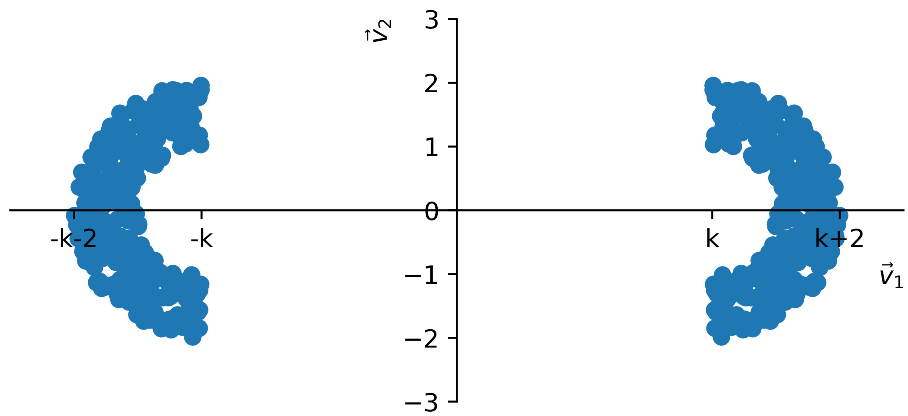
◯	Valid
◯	Invalid

Solution:

◯	Valid
●	Invalid

Since our data appears to be centered (the data is symmetric about the x- and y-axes), we can find the direction that maximizes the spread in our data. In this case, the direction appears to be given by the axis $\frac{1}{\sqrt{2}}\vec{v}_1 + \frac{1}{\sqrt{2}}\vec{v}_2$.

(d) (5 pts.) **Is the following plot valid?** Remember, we are assuming that $k \to \infty$.



$k \to \infty$

◯	Valid
◯	Invalid

Solution:

●	Valid
◯	Invalid

Any axis we choose can be represented as $\alpha\vec{v}_1 + \beta\vec{v}_2$, a linear combination of the axes shown in the plot. We require $\alpha^2 + \beta^2 = 1$ since the axes have to be normal, so as long as $\alpha \neq 0$, we will be maximizing the sum of squares of coordinates. Specifically, if we choose an axis with $\alpha \neq 0$, then the sum of squares of coordinates will be ∞ in the limit, and $\vec{v}_1$ is an axis with $\alpha \neq 0$ so the plot is valid.

9. Observability of Discrete-Time Systems (20 pts.)

Consider the discrete-time system given by

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{51}$$

Suppose we cannot directly observe $\vec{x}[i]$, but we observe $\vec{y}[i] = C\vec{x}[i]$ instead (where C need not be invertible, or even square). You may assume that you can observe $\vec{u}[i]$, since this is the input you provide to the system. Assume that A, B, and C are known matrices.

(a) (6 pts.) **Argue that, if you know $\vec{x}[0]$, then you can determine every $\vec{x}[i]$ for $i \geq 1$.**

(HINT: Express $\vec{x}[i]$ in terms of $\vec{x}[0]$ and other known quantities.)

Solution: From homework, we know that we can write

$$\vec{x}[i] = A^i\vec{x}[0] + \sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j] \tag{52}$$

So if we know $\vec{x}[0]$, we can calculate $\vec{x}[i]$ since all other quantities in that equation are known.

(b) (8 pts.) **Show that**

$$\vec{y}[i] - C\sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j] = CA^i\vec{x}[0] \tag{53}$$

for $i \geq 1$ and that $\vec{y}[0] = C\vec{x}[0]$.

Solution: From the solution to the previous part, we have that

$$\vec{x}[i] = A^i\vec{x}[0] + \sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j] \tag{54}$$

and from the problem statement, we know $\vec{y}[i] = C\vec{x}[i]$. Combining these two, we have

$$\vec{y}[i] = C\vec{x}[i] \tag{55}$$

$$= C\left(A^i\vec{x}[0] + \sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j]\right) \tag{56}$$

$$= CA^i\vec{x}[0] + C\sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j] \tag{57}$$

Rearranging this yields

$$\vec{y}[i] - C\sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j] = CA^i\vec{x}[0] \tag{58}$$

The expression $\vec{y}[0] = C\vec{x}[0]$ follows from the definition provided in the problem statement.

(c) (4 pts.) Use the result of the previous part and plug in $i = 0, 1, \ldots, n-1$ to obtain n different equations. Combine all of these equations into a single matrix-vector equation, of the form $\vec{z} = \mathcal{O}\vec{x}[0]$. **What are $\vec{z}$ and $\mathcal{O}$?**

◯	$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$	$\vec{z} = \begin{bmatrix} \vec{y}[0] \\ \vec{y}[1] - CB\vec{u}[0] \\ \vec{y}[2] - CB\vec{u}[1] - CAB\vec{u}[0] \\ \vdots \\ \vec{y}[n-1] - C\sum_{j=0}^{n-2} A^{n-2-j}B\vec{u}[j] \end{bmatrix}$
◯	$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$	$\vec{z} = \begin{bmatrix} \vec{y}[0] \\ \vec{y}[1] \\ \vec{y}[2] - CAB\vec{u}[0] \\ \vdots \\ \vec{y}[n-1] - C\sum_{j=0}^{n-2} A^{n-2-j}B\vec{u}[j] \end{bmatrix}$
◯	$\mathcal{O} = \begin{bmatrix} I \\ A \\ A^2 \\ \vdots \\ A^{n-1} \end{bmatrix}$	$\vec{z} = \begin{bmatrix} \vec{y}[0] \\ \vec{y}[1] - CAB\vec{u}[0] \\ \vec{y}[2] - CAB\vec{u}[1] - CA^2\vec{u}[0] \\ \vdots \\ \vec{y}[n-1] - C\sum_{j=0}^{n-2} A^{n-2-j}B\vec{u}[j] \end{bmatrix}$
◯	$\mathcal{O} = \begin{bmatrix} I \\ A \\ A^2 \\ \vdots \\ A^{n-1} \end{bmatrix}$	$\vec{z} = \begin{bmatrix} \vec{y}[0] \\ \vec{y}[1] \\ \vec{y}[2] - CAB\vec{u}[0] \\ \vdots \\ \vec{y}[n-1] - C\sum_{j=0}^{n-2} A^{n-2-j}B\vec{u}[j] \end{bmatrix}$

Solution:

●	$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$	$\vec{z} = \begin{bmatrix} \vec{y}[0] \\ \vec{y}[1] - CB\vec{u}[0] \\ \vec{y}[2] - CB\vec{u}[1] - CAB\vec{u}[0] \\ \vdots \\ \vec{y}[n-1] - C\sum_{j=0}^{n-2} A^{n-2-j}B\vec{u}[j] \end{bmatrix}$
○	$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$	$\vec{z} = \begin{bmatrix} \vec{y}[0] \\ \vec{y}[1] \\ \vec{y}[2] - CAB\vec{u}[0] \\ \vdots \\ \vec{y}[n-1] - C\sum_{j=0}^{n-2} A^{n-2-j}B\vec{u}[j] \end{bmatrix}$
○	$\mathcal{O} = \begin{bmatrix} I \\ A \\ A^2 \\ \vdots \\ A^{n-1} \end{bmatrix}$	$\vec{z} = \begin{bmatrix} \vec{y}[0] \\ \vec{y}[1] - CAB\vec{u}[0] \\ \vec{y}[2] - CAB\vec{u}[1] - CA^2B\vec{u}[0] \\ \vdots \\ \vec{y}[n-1] - C\sum_{j=0}^{n-2} A^{n-2-j}B\vec{u}[j] \end{bmatrix}$
○	$\mathcal{O} = \begin{bmatrix} I \\ A \\ A^2 \\ \vdots \\ A^{n-1} \end{bmatrix}$	$\vec{z} = \begin{bmatrix} \vec{y}[0] \\ \vec{y}[1] \\ \vec{y}[2] - CAB\vec{u}[0] \\ \vdots \\ \vec{y}[n-1] - C\sum_{j=0}^{n-2} A^{n-2-j}B\vec{u}[j] \end{bmatrix}$

From the previous part, we have that the kth row of this matrix-vector equation would be

$$\begin{cases} \vec{y}[k] - C\sum_{j=0}^{k-1} A^{k-1-j}B\vec{u}[j] = CA^k\vec{x}[0] & k \geq 1 \\ \vec{y}[k] = C\vec{x}[k] & k = 0 \end{cases} \tag{59}$$

so we can say that the kth block matrix "row" of $\mathcal{O}$ would be

$$\mathcal{O}_k = CA^k \tag{60}$$

and the kth element of $\vec{z}$ would be

$$z_k = \begin{cases} \vec{y}[k] - C\sum_{j=0}^{k-1} A^{k-1-j}B\vec{u}[j] & k \geq 1 \\ \vec{y}[k] & k = 0 \end{cases} \tag{61}$$

(d) (2 pts.) **What is the loosest (least strict) condition on $\mathcal{O}$ from the prevous part, for you to be able to uniquely estimate $\vec{x}[0]$ in the matrix-vector equation from the previous part?**

◯	$\mathcal{O}$ has to be full row rank
◯	$\mathcal{O}$ has to be full column rank
◯	$\mathcal{O}$ has to be square and invertible
◯	We can uniquely solve for $\vec{x}[0]$ for any $\mathcal{O}$

Note: This condition is the "observability" condition for linear,discrete-time systems.

(HINT: Would $\mathcal{O}$ be square for all possible matrices C? Notice that the problems says that we are trying to estimate $\vec{x}[0]$.)

Solution:

◯	$\mathcal{O}$ has to be full row rank
●	$\mathcal{O}$ has to be full column rank
◯	$\mathcal{O}$ has to be square and invertible
◯	We can uniquely solve for $\vec{x}[0]$ for any $\mathcal{O}$

We need $\mathcal{O}$ to be full column rank, so that we can apply our least-squares solution, i.e.,

$$\vec{x}[0] = \left(\mathcal{O}^\top \mathcal{O}\right)^{-1} \mathcal{O}^\top \vec{z} \tag{62}$$

In other words, we need $\mathcal{O}^\top \mathcal{O}$ to be invertible, which is true if and only if $\mathcal{O}$ is full column rank.

10. Gradient Descent and Discrete-Time Systems (40 pts.)

Consider the optimization problem provided by least squares, namely

$$\underset{\vec{x}}{\operatorname{argmin}} \|\vec{y} - A\vec{x}\|^2 \tag{63}$$

where $A \in \mathbb{R}^{n \times m}$, $\vec{x} \in \mathbb{R}^m$, and $\vec{y} \in \mathbb{R}^n$. We can define a "loss function" based on this optimization problem, which we will denote $\mathcal{L}(\vec{x}) := \|\vec{y} - A\vec{x}\|^2$. When trying to solve for an optimal $\vec{x}$, we can perform "gradient descent" which is written as

$$\vec{x}[i+1] = \vec{x}[i] - \alpha\left(\nabla_{\vec{x}[i]}\mathcal{L}(\vec{x}[i])\right) \tag{64}$$

where $\vec{x}[i]$ is the proposed, optimal value of $\vec{x}$ on the ith iteration of the algorith, $\nabla_{\vec{x}[i]}\mathcal{L}(\vec{x}[i])$ is the gradient of the loss function, and α is a "learning rate" parameter that the user defines. For now, you may assume that

$$\nabla_{\vec{x}[i]}\mathcal{L}(\vec{x}[i]) = 2A^\top A\vec{x}[i] - 2A^\top\vec{y} \tag{65}$$

Note: parts (d), (e), and (f) are designed to be independent of the rest of the subparts and independent amongst themselves, so you may solve these subparts in any order that you wish.

(a) (4 pts.) By combining eq. (64) and eq. (65), we may write

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top\vec{y} \tag{66}$$

Simplify eq. (66) to resemble the form $\vec{x}[i+1] = M\vec{x}[i] + \vec{z}$ for an appropriately defined matrix M and vector $\vec{z}$.

◯	$M = I - 2\alpha A^\top A$	$\vec{z} = A^\top\vec{y}$
◯	$M = -2\alpha A^\top A$	$\vec{z} = A^\top\vec{y}$
◯	$M = I - 2\alpha A^\top A$	$\vec{z} = 2\alpha A^\top\vec{y}$
◯	$M = -2\alpha A^\top A$	$\vec{z} = 2\alpha A^\top\vec{y}$

(HINT: Remember that $\vec{x}[i] = I\vec{x}[i]$, and try factoring eq. (66).)

Solution:

◯	$M = I - 2\alpha A^\top A$	$\vec{z} = A^\top\vec{y}$
◯	$M = -2\alpha A^\top A$	$\vec{z} = A^\top\vec{y}$
●	$M = I - 2\alpha A^\top A$	$\vec{z} = 2\alpha A^\top\vec{y}$
◯	$M = -2\alpha A^\top A$	$\vec{z} = 2\alpha A^\top\vec{y}$

We can simplify eq. (66) as follows:

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top\vec{y} \tag{67}$$

$$= \underbrace{\left(I - 2\alpha A^\top A\right)}_{M}\vec{x}[i] + \underbrace{2\alpha A^\top\vec{y}}_{\vec{z}} \tag{68}$$

(b) (4 pts.) Let $\vec{v}$ be an eigenvector of $A^\top A$, with corresponding eigenvalue $\lambda_{\vec{v}}$. **Using the fact that any vector is an eigenvector of the identity matrix, find an expression for the eigenvalue of M that corresponds to the eigenvector $\vec{v}$.**

○	$1-2\alpha\lambda_{\vec{v}}$
○	$-2\alpha\lambda_{\vec{v}}$
○	$1-\lambda_{\vec{v}}$
○	$\lambda_{\vec{v}}$

Solution:

•	$1-2\alpha\lambda_{\vec{v}}$
○	$-2\alpha\lambda_{\vec{v}}$
○	$1-\lambda_{\vec{v}}$
○	$\lambda_{\vec{v}}$

Let $\vec{v}$ be an eigenvector of $A^\top A$. We have that

$$\left(I-2\alpha A^\top A\right)\vec{v} = \vec{v}-2\alpha A^\top A\vec{v} \tag{69}$$
$$= \vec{v}-2\alpha\lambda_{\vec{v}}\vec{v} \tag{70}$$
$$= (1-2\alpha\lambda_{\vec{v}})\vec{v} \tag{71}$$

so the corresponding eigenvalue is $1-2\alpha\lambda_{\vec{v}}$.

(c) (6 pts.) **Find a suitable range of values of α so that the discrete system from part (a) is stable.** You may write your answer in terms of $\lambda_{\min}(A^\top A)$, the smallest eigenvalue of $A^\top A$, and $\lambda_{\max}(A^\top A)$, the largest eigenvalue of $A^\top A$. You may also assume that the eigenvalues of $A^\top A$ are nonnegative (i.e., $\lambda_{\min}(A^\top A) \geq 0$).

○	$0<\alpha<\lambda_{\max}(A^\top A)$
○	$\lambda_{\min}(A^\top A)<\alpha<\lambda_{\max}(A^\top A)$
○	$0<\alpha<\frac{1}{\lambda_{\max}(A^\top A)}$
○	$0<\alpha<\frac{1}{\lambda_{\min}(A^\top A)}$

Solution:

○	$0<\alpha<\lambda_{\max}(A^\top A)$
○	$\lambda_{\min}(A^\top A)<\alpha<\lambda_{\max}(A^\top A)$
•	$0<\alpha<\frac{1}{\lambda_{\max}(A^\top A)}$
○	$0<\alpha<\frac{1}{\lambda_{\min}(A^\top A)}$

We need $|1-2\alpha\lambda|<1$, where λ is an arbitrary eigenvalue of $A^\top A$. In other words, we need $1-2\alpha\lambda<1$ and $1-2\alpha\lambda>-1$. We know that $1-2\alpha\lambda \leq 1-2\alpha\lambda_{\min}(A^\top A)$ since $\lambda \geq \lambda_{\min}(A^\top A)$, so we can find a bound on α such that

$$1-2\alpha\lambda_{\min}\left(A^\top A\right)<1 \tag{72}$$
$$\implies \alpha>0 \tag{73}$$

We can repeat this similarly for the upper bound on α. That is, $1-2\alpha\lambda \geq 1-2\alpha\lambda_{\max}(A^\top A)$, since $\lambda \leq \lambda_{\max}(A^\top A)$. Hence, we can find a bound on α such that

$$1-2\alpha\lambda_{\max}\left(A^\top A\right)>-1 \tag{74}$$

$$-2\alpha\lambda_{\max}\left(A^\top A\right) > -2 \tag{75}$$

$$\alpha < \frac{1}{\lambda_{\max}\left(A^\top A\right)} \tag{76}$$

Altogether, $0 < \alpha < \frac{1}{\lambda_{\max}\left(A^\top A\right)}$.

(d) (8 pts.) For your convenience, here is eq. (66) rewritten:

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top \vec{y} \tag{77}$$

Let us assume that A is full column rank (i.e., we are solving an overdetermined system). If our system has stabilized and reach steady state, then $\vec{x}[i+1] = \vec{x}[i] = \vec{x}^\star$. **Show that, at steady state, $\vec{x}^\star = \left(A^\top A\right)^{-1} A^\top \vec{y}$, your least-squares solution.** *(HINT: If A is full column rank, what does this mean about $A^\top A$?)*

Solution: Let us denote $\vec{x}[i] = \vec{x}[i+1] = \vec{x}^\star$ at steady state. From eq. (66), we have

$$\vec{x}^\star = \vec{x}^\star - 2\alpha A^\top A\vec{x}^\star + 2\alpha A^\top \vec{y} \tag{78}$$

$$2\alpha A^\top A\vec{x}^\star = 2\alpha A^\top \vec{y} \tag{79}$$

$$A^\top A\vec{x}^\star = A^\top \vec{y} \tag{80}$$

$$\vec{x}^\star = \left(A^\top A\right)^{-1} A^\top \vec{y} \tag{81}$$

(e) (10 pts.) For your convenience, here is eq. (66) rewritten:

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top \vec{y} \tag{82}$$

Now, let us consider the case where A need not be full column rank (e.g., it is full row rank and an underdetermined system). **Show that, if $\vec{x}[0] = \vec{0}$, then $\vec{x}[i] \in \mathrm{Col}\left(A^\top\right)$, regardless of the rank or dimensions of A.** To accomplish this, do the following steps:

i. Show that $\vec{x}[1] \in \mathrm{Col}\left(A^\top\right)$.

ii. Assume $\vec{x}[i] \in \mathrm{Col}\left(A^\top\right)$. Show that $\vec{x}[i+1] \in \mathrm{Col}\left(A^\top\right)$.

Conclude that $\vec{x}[i]$ is orthogonal to any vector in $\mathrm{Null}(A)$ and the optimal solution found by the algorithm will be the min-norm solution.

(HINT: If a vector $\vec{b} \in \mathrm{Col}(A)$, there exists a $\vec{x}$ such that $A\vec{x} = \vec{b}$.) (HINT: Recall that, from properties of the SVD, we can derive that all vectors in $\mathrm{Null}(A)$ are orthogonal to all vectors in $\mathrm{Col}\left(A^\top\right)$.)

Solution: Following the steps, we can compute

$$\vec{x}[1] = \left(I - 2\alpha A^\top A\right)\vec{x}[0] + 2\alpha A^\top \vec{y} \tag{83}$$

$$= 2\alpha A^\top \vec{y} \tag{84}$$

From the hint, we have that $\vec{x}[1] \in \mathrm{Col}\left(A^\top\right)$. Now, let's assume that $\vec{x}[i] \in \mathrm{Col}\left(A^\top\right)$. Using the hint again, we can say that $\vec{x}[i] = A^\top\vec{b}$ for some $\vec{b}$. Thus,

$$\vec{x}[i+1] = \left(I - 2\alpha A^\top A\right)\vec{x}[i] + 2\alpha A^\top \vec{y} \tag{85}$$

$$= \left(I - 2\alpha A^\top A\right)A^\top\vec{b} + 2\alpha A^\top \vec{y} \tag{86}$$

$$= A^\top\vec{b} - 2\alpha A^\top A A^\top \vec{b} + 2\alpha A^\top \vec{y} \tag{87}$$

$$= A^\top\left(\vec{b} - 2\alpha A A^\top \vec{b} + 2\alpha \vec{y}\right) \tag{88}$$

which, using the hint, shows us that $\vec{x}[i+1] \in \mathrm{Col}\left(A^\top\right)$. From the second hint, we have that $\vec{x}[i]$ is orthogonal to all vectors in $\mathrm{Null}(A)$. From the min-norm solution derivation, we chose the solution so that it is completely orthogonal to any vector in $\mathrm{Null}(A)$, so the solution found by the algorithm will always be the min-norm solution.

(f) (8 pts.) For your convenience, here is eq. (66) rewritten:

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top \vec{y} \tag{89}$$

From the previous part, we have $\vec{x}[i] \in \mathrm{Col}\left(A^\top\right)$ for every $i \geq 1$, when $\vec{x}[0] = \vec{0}$. **If A is full row rank (we are solving an underdetermined system), show that, at steady state, $\vec{x}^\star = A^\top\left(AA^\top\right)^{-1}\vec{y}$.** Here, we are using the same definition of steady state as in part (d). *(HINT: You may want to borrow some of your work from part (d), and now apply the fact that $\vec{x}^\star \in \mathrm{Col}\left(A^\top\right)$. The hint from part (e) might help you with this.) (HINT: Remember that $A^\top$ is full column rank, so if $A^\top\vec{s} = A^\top\vec{p}$, this would only be true if $\vec{s} = \vec{p}$.) (HINT: If A is full row rank, what does this mean about $AA^\top$?)*

Solution: Let us denote $\vec{x}[i+1] = \vec{x}[i] = \vec{x}^\star$. From eq. (66), we have

$$\vec{x}^\star = \vec{x}^\star - 2\alpha A^\top A\vec{x}^\star + 2\alpha A^\top\vec{y} \tag{90}$$

$$2\alpha A^\top A\vec{x}^\star = 2\alpha A^\top\vec{y} \tag{91}$$

$$A^\top A\vec{x}^\star = A^\top\vec{y} \tag{92}$$

From the previous part's hint, we have $\vec{x}^\star = A^\top\vec{b}^\star$ for some vector $\vec{b}^\star$. Plugging this in, we have

$$A^\top AA^\top\vec{b}^\star = A^\top\vec{y} \tag{93}$$

Since $A^\top$ is full column rank (because A is full row rank), it does not have a nontrivial nullspace. Hence, the equality above holds true if and only if

$$AA^\top\vec{b}^\star = \vec{y} \tag{94}$$

We may use the fact that A is full row rank to write

$$\vec{b}^\star = \left(AA^\top\right)^{-1}\vec{y} \tag{95}$$

Plugging this in for $\vec{x}^\star$, we get

$$\vec{x}^\star = A^\top\vec{b}^\star = A^\top\left(AA^\top\right)^{-1}\vec{y} \tag{96}$$

which is exactly the min-norm solution.

11. Eckart-Young Proof (40 pts.)

We can define the 2-norm of a matrix $A \in \mathbb{R}^{m \times n}$ as

$$\|A\|_2 = \max_{\vec{s} \in \mathbb{R}^n} \frac{\|A\vec{s}\|}{\|\vec{s}\|} \tag{97}$$

It is the case that $\|A\|_2 = \sigma_1(A)$, where $\sigma_1(A)$ is the largest singular value of A. Using this definition of the 2-norm of a matrix, the Eckart-Young Theorem can be restated as follows:

Let $A \in \mathbb{R}^{m \times n}$ have $\operatorname{rank}(A) = r$, and let A have an SVD $A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top$. Define $A_k := \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^\top$, where $k < r$. Notice that $\operatorname{rank}(A_k) = k$. For any matrix $B \in \mathbb{R}^{m \times n}$ where $\operatorname{rank}(B) = k$, it is the case that $\|A - B\|_2 \geq \|A - A_k\|_2$.

In this problem, we will prove the restated Eckart-Young Theorem. *Note: the parts are designed to be independent, so you may solve the subparts in any order that you wish.*

(a) (9 pts.) First, we will choose a unit vector $\vec{x} \in \mathbb{R}^n$ such that $\vec{x} \in \operatorname{Null}(B)$ and $\vec{x} \in \operatorname{Col}(V_{k+1})$ (where $V_{k+1} := \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{k+1} \end{bmatrix}$). Let us further choose $\vec{x}$ so that $\|\vec{x}\| = 1$. This selection is possible because of the pigeonhole principle – since $\dim \operatorname{Null}(B) + \dim \operatorname{Col}(V_{k+1}) = (n-k) + (k+1) = n+1$, there must be some nontrivial intersection *(you don't need to show this)*. **Argue that we can write $\vec{x}$ as $\vec{x} = V \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix}$ where $\vec{\alpha} \in \mathbb{R}^{k+1}$, and argue that $\vec{\alpha}$ has norm 1.**

(HINT: If $\vec{x} \in \operatorname{Col}(V_{k+1})$, we can write $\vec{x} = V_{k+1}\vec{\alpha}$, for some vector $\vec{\alpha} \in \mathbb{R}^{k+1}$. Use this to write $\vec{x} = V \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix}$. Then, argue that $\|\vec{\alpha}\| = 1$.)

Solution: From the hint, we have that $\vec{x} = V_{k+1}\vec{\alpha}$. We can write this more verbosely, in terms of $V_{n-k-1} := \begin{bmatrix} \vec{v}_{k+2} & \cdots & \vec{v}_n \end{bmatrix}$ as

$$\vec{x} = V_{k+1}\vec{\alpha} \tag{98}$$
$$= V_{k+1}\vec{\alpha} + V_{n-k-1}\vec{0}_{n-k-1} \tag{99}$$
$$= \begin{bmatrix} V_{k+1} & V_{n-k-1} \end{bmatrix} \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix} \tag{100}$$
$$= V \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix} \tag{101}$$

Next, we can show that $\|\vec{\alpha}\| = 1$. We have that

$$\|\vec{x}\|^2 = \|V_{k+1}\vec{\alpha}\|^2 \tag{102}$$
$$1 = \vec{\alpha}^\top \underbrace{V_{k+1}^\top V_{k+1}}_{I_{k+1}} \vec{\alpha} \tag{103}$$
$$= \vec{\alpha}^\top \vec{\alpha} \tag{104}$$
$$= \|\vec{\alpha}\|^2 \tag{105}$$

so $\|\vec{\alpha}\| = 1$.

(b) (9 pts.) **Using the definition of the 2-norm for matrices, argue that $\|A-B\|_2 \geq \|(A-B)\vec{x}\|$, where $\|\vec{x}\| = 1$. Then, using the specific $\vec{x}$ we defined in part (a), argue that $\|(A-B)\vec{x}\| = \|A\vec{x}\|$.**

(HINT: Use eq. (97) to show the first part. We are choosing a specific vector $\vec{x}$ here, so $\frac{\|(A-B)\vec{x}\|}{\|\vec{x}\|} \leq \max_{\vec{s}\in\mathbb{R}^n} \frac{\|(A-B)\vec{s}\|}{\|\vec{s}\|}$.)

Solution: From the hint, we have that

$$\|A-B\|_2 = \max_{\vec{s}\in\mathbb{R}^n} \frac{\|(A-B)\vec{s}\|}{\|\vec{s}\|} \tag{106}$$

$$\geq \frac{\|(A-B)\vec{x}\|}{\|\vec{x}\|} \tag{107}$$

$$= \|(A-B)\vec{x}\| \tag{108}$$

where in the last step we notice that $\|\vec{x}\| = 1$. Then, from the definition of $\vec{x}$, we have that $\vec{x} \in \mathrm{Null}(B)$ so $B\vec{x} = \vec{0}_m$, and thus, $\|(A-B)\vec{x}\| = \|A\vec{x}\|$.

(c) (9 pts.) **Using the SVD of A and the result of part (a), argue that $\|A\vec{x}\| = \left\|\Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix}\right\|$.**

Solution: We can write that $\|A\vec{x}\| = \left\|U\Sigma V^\top \vec{x}\right\|$. From homework and discussion, we know that $\left\|U\Sigma V^\top \vec{x}\right\| = \left\|\Sigma V^\top \vec{x}\right\|$. Then, from part (a), we know that $\vec{x} = V \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix}$. Plugging this in, we have

$$\|A\vec{x}\| = \left\|\Sigma V^\top \vec{x}\right\| \tag{109}$$

$$= \left\|\Sigma V^\top V \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix}\right\| \tag{110}$$

$$= \left\|\Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix}\right\| \tag{111}$$

(d) (9 pts.) From the previous part, we may write

$$\left\| \Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix} \right\| = \sqrt{\sum_{i=1}^{k+1} (\sigma_i \alpha_i)^2} \tag{112}$$

where α_i is the ith element of $\vec{\alpha}$. **Argue that**

$$\sqrt{\sum_{i=1}^{k+1} (\sigma_i \alpha_i)^2} \geq \sqrt{\sum_{i=1}^{k+1} \sigma_{k+1}^2 (\alpha_i)^2} \tag{113}$$

and hence

$$\left\| \Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix} \right\| \geq \sigma_{k+1} \tag{114}$$

From this, we have $\|A - B\|_2 \geq \sigma_{k+1}$. *(HINT: Think about the definition of $\|\vec{\alpha}\|$. How can you write it as a function of the entries of $\vec{\alpha}$?)*

Solution: Since $\sigma_{k+1} \leq \sigma_i$ for $i \in \{1, 2, \ldots, k+1\}$, we have that

$$\sqrt{\sum_{i=1}^{k+1} (\sigma_i \alpha_i)^2} \geq \sqrt{\sum_{i=1}^{k+1} (\sigma_{k+1} \alpha_i)^2} \tag{115}$$

$$= \sqrt{\sigma_{k+1}^2 \sum_{i=1}^{k+1} \alpha_i^2} \tag{116}$$

$$= \sigma_{k+1} \sqrt{\sum_{i=1}^{k+1} \alpha_i^2} \tag{117}$$

$$= \sigma_{k+1} \|\vec{\alpha}\| \tag{118}$$

$$= \sigma_{k+1} \tag{119}$$

where we notice that $\sqrt{\sum_{i=1}^{k+1} \alpha_i^2} = \|\vec{\alpha}\| = 1$. Hence, from the problem statement, we have

$$\left\| \Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix} \right\| \geq \sigma_{k+1} \tag{120}$$

and combining all the previous parts, we have $\|A - B\|_2 \geq \sigma_{k+1}$.

(e) (4 pts.) **What is $\|A - A_k\|_2$?** In other words, what is the largest singular value of $A - A_k$?

○	σ_k
○	σ_{k+1}
○	σ_{k-1}
○	σ_{k+2}
○	σ_1

(HINT: Use the SVD outer product form of A and A_k.)

Solution:

○	σ_k
●	σ_{k+1}
○	σ_{k-1}
○	σ_{k+2}
○	σ_1

Following the hint, we have

$$A - A_k = \left(\sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top\right) - \left(\sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^\top\right) \tag{121}$$

$$= \sum_{i=k+1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top \tag{122}$$

This outer product representation of $A - A_k$ is a valid SVD, since the σ_i's are positive and nonincreasing for as i increases, $\vec{u}_i$'s are orthonormal, and $\vec{v}_i$'s are orthonormal. Hence, $\|A - A_k\|_2 = \sigma_{k+1}$, since it is the largest singular value according to the outer product form.

This concludes the proof of the Eckart-Young Theorem.