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.

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

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

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

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

◯ True
◯ False

(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

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

◯ True
◯ False

(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

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

◯ True
◯ False

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

◯ True
◯ False

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

◯ True
◯ False

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

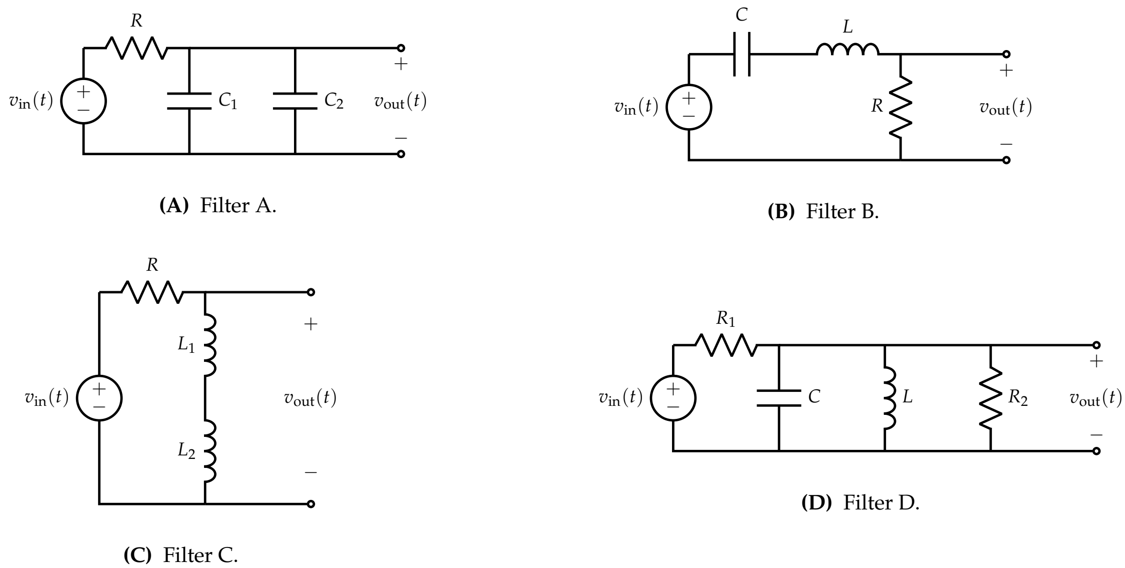
◯ True
◯ False

(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

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:

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

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

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

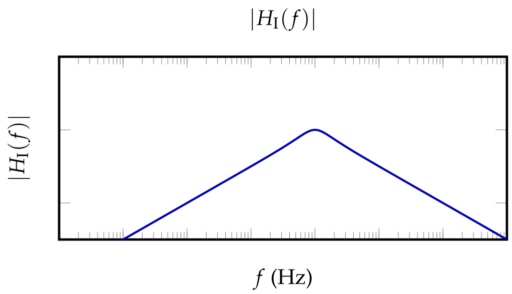
$$\tag{4}$$

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

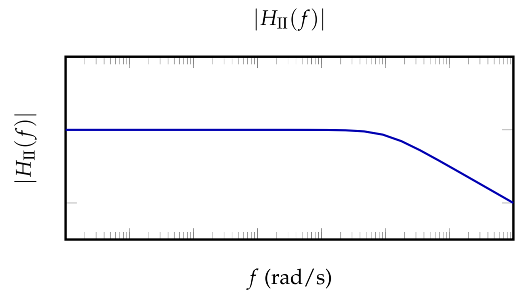
Filter Letter	TF I	TF II	TF III	None
A	○	○	○	○
B	○	○	○	○
C	○	○	○	○
D	○	○	○	○

[Extra page. If you want the work on this page to be graded, make sure you tell us on the problem's main page.]

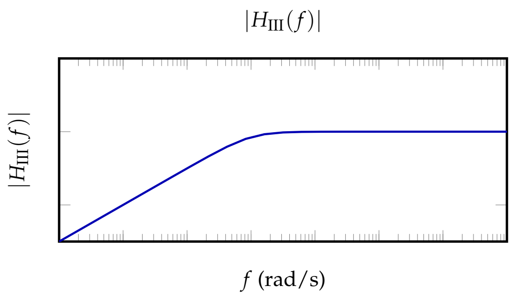
(b) (4 pts.) **Now, using the same 4 filter circuits (A, B, C, D) from 2, fill in the bubbles to match each filter to its corresponding magnitude Bode Plot out of choices I, II, III, IV.**



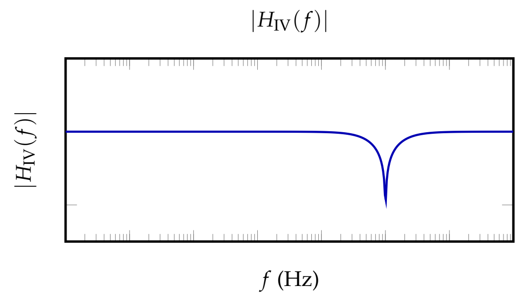
(a) Plot I.



(b) Plot II.



(c) Plot III.



(d) Plot IV.

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

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

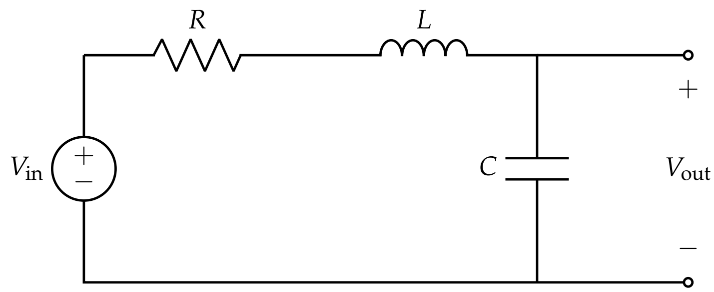


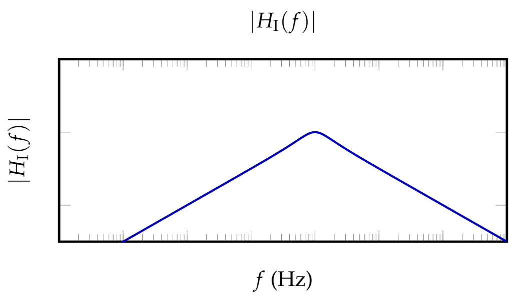
Figure 4: RLC Filter

(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**

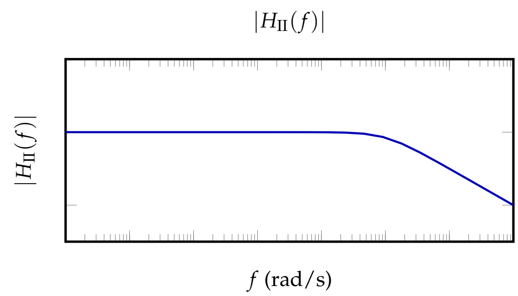
$$H(f) = \frac{-\mathrm{j}\alpha \frac{f_0}{f}}{1 + \mathrm{j}\beta\left(\frac{f}{f_0} - \frac{f_0}{f}\right)} \tag{5}$$

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.)*

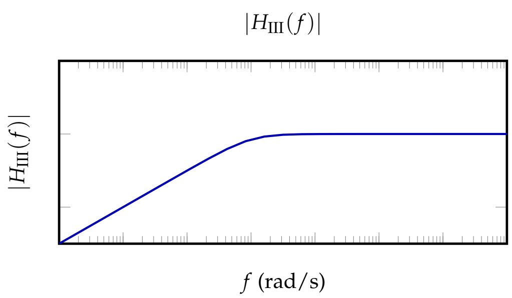
(d) (2 pts.) **Now, match the magnitude of the transfer function $H(f)$ from fig. 4 with the correct bode plot shape below.** You do not need to have successfully calculated values for α and β in item 5.c to solve this subpart.



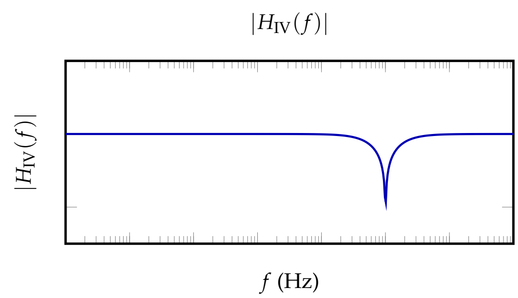
(a) Plot I.



(b) Plot II.



(c) Plot III.



(d) Plot IV.

◯	Plot I
◯	Plot II
◯	Plot III
◯	Plot IV

(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.

◯	$|H(f_0)|$ increase as α increases
◯	$|H(f_0)|$ decreases as α increases
◯	$|H(f_0)|$ stays the same as α increases

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

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{6}$$

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

◯ True
◯ False
◯ Not Enough Information

(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

(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

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

◯ True
◯ False
◯ Not Enough Information

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

◯ True
◯ False
◯ Not Enough Information

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

$$A = URU^{\top} \tag{7}$$

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

◯ True
◯ False
◯ Not Enough Information

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

◯ True
◯ False
◯ Not Enough Information

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

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{8}$$

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

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

Consider the following set of vectors:

$$\left\{\vec{d}_1, \vec{d}_2, \vec{d}_3\right\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \tag{9}$$

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

$$\operatorname{Span}\left(\vec{d}_1, \vec{d}_2, \vec{d}_3\right) = \operatorname{Span}(\vec{q}_1, \vec{q}_2, \vec{q}_3) \tag{10}$$

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

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

(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 $\vec{q}_1$.

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

(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 $\vec{q}_2$.

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

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

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \tag{11}$$

What is the QR decomposition of A?

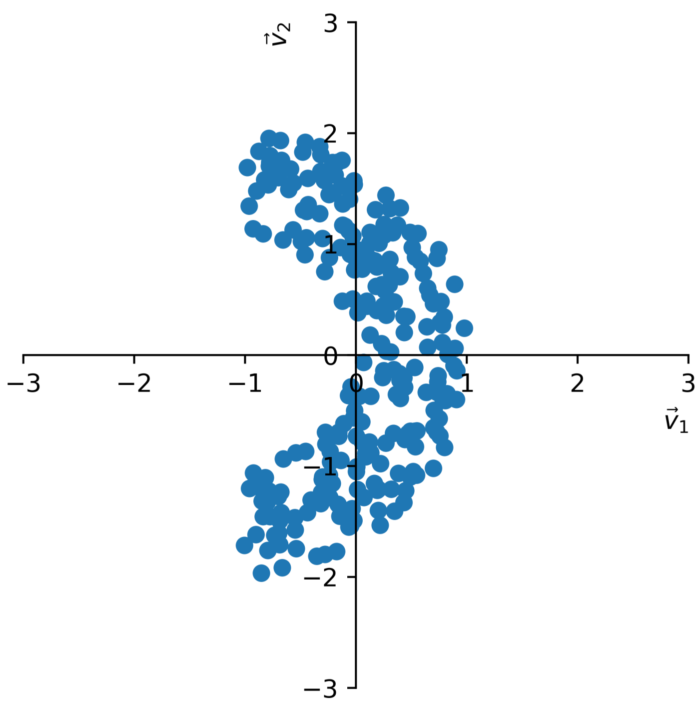
◯	$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}$	$R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2}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

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.

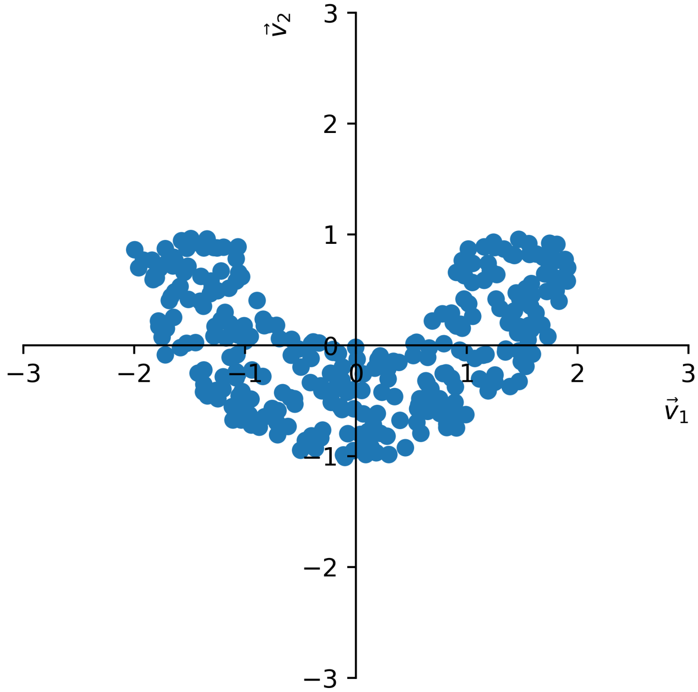
(HINT: Note that the mean or "center of mass" of the data points in all the plots are the origin, (0, 0).) (HINT: The procedure for this problem is very similar to the procedure in Discussion 12B, where you were similarly asked to judge the "validity" of scatterplots.)

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



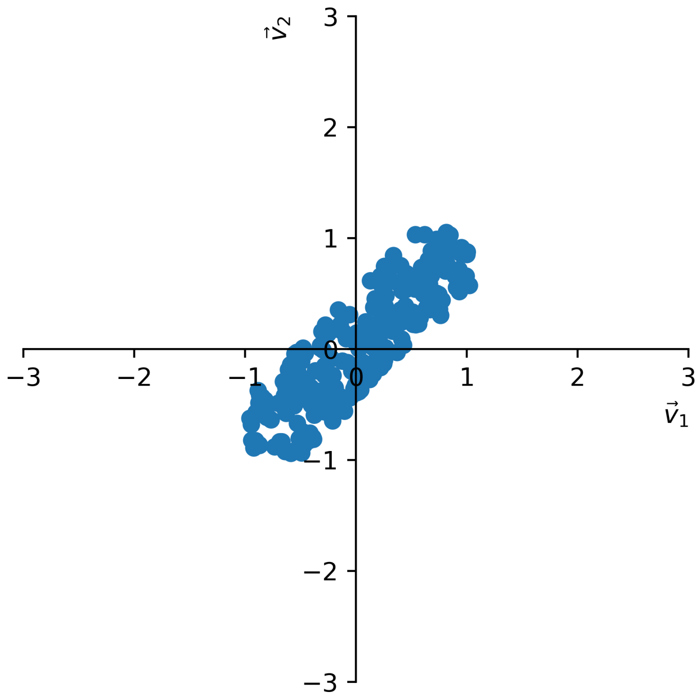
◯	Valid
◯	Invalid

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



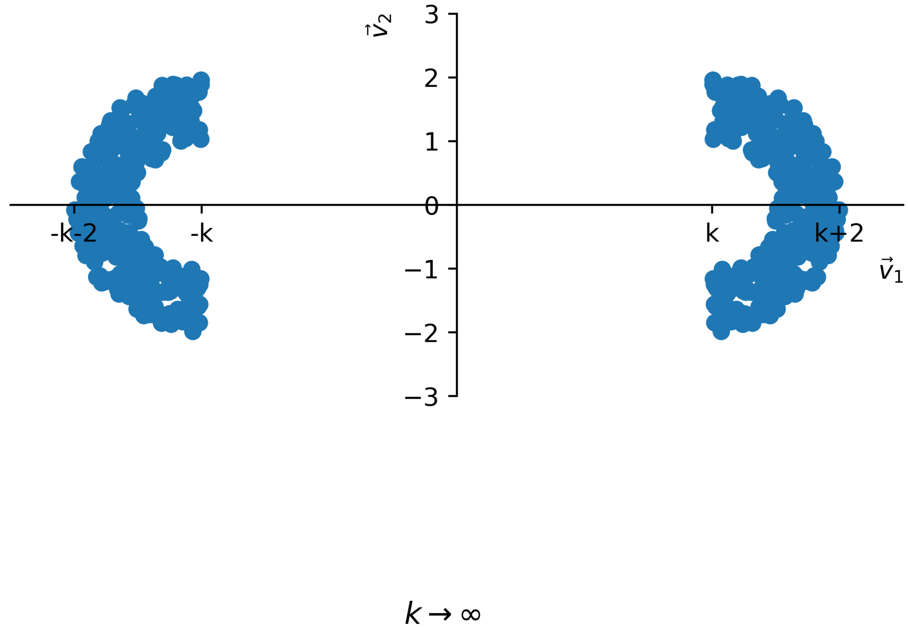
◯	Valid
◯	Invalid

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



◯	Valid
◯	Invalid

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



◯	Valid
◯	Invalid

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

Consider the discrete-time system given by

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{12}$$

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.)

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

$$\vec{y}[i] - C\sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j] = CA^{i}\vec{x}[0] \tag{13}$$

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

(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 $\mathcal{O}$?**

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

(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 $\vec{x}[0]$.)

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

Consider the optimization problem provided by least squares, namely

$$\underset{\vec{x}}{\operatorname{argmin}} \|\vec{y} - A\vec{x}\|^2 \tag{14}$$

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

$$\vec{x}[i+1] = \vec{x}[i] - \alpha\left(\nabla_{\vec{x}[i]}\mathcal{L}(\vec{x}[i])\right) \tag{15}$$

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

$$\nabla_{\vec{x}[i]}\mathcal{L}(\vec{x}[i]) = 2A^\top A\vec{x}[i] - 2A^\top\vec{y} \tag{16}$$

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. (15) and eq. (16), we may write

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top\vec{y} \tag{17}$$

Simplify eq. (17) 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. (17).)

(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 $\vec{v}$.**

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

(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 $\lambda_{\min}\left(A^\top A\right)$, the smallest eigenvalue of $A^\top A$, and $\lambda_{\max}\left(A^\top A\right)$, the largest eigenvalue of $A^\top A$. You may also assume that the eigenvalues of $A^\top A$ are nonnegative (i.e., $\lambda_{\min}\left(A^\top A\right) \geq 0$).

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

[Extra page. If you want the work on this page to be graded, make sure you tell us on the problem's main page.]

(d) (8 pts.) For your convenience, here is eq. (17) rewritten:

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top \vec{y} \tag{18}$$

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 $A^\top A$?)*

(e) (10 pts.) For your convenience, here is eq. (17) rewritten:

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top \vec{y} \tag{19}$$

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 $\vec{x}[i] \in \text{Col}\left(A^\top\right)$, regardless of the rank or dimensions of A.** To accomplish this, do the following steps:

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

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

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

(HINT: If a vector $\vec{b} \in \text{Col}(A)$, 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 $\text{Null}(A)$ are orthogonal to all vectors in $\text{Col}\left(A^\top\right)$.)

(f) (8 pts.) For your convenience, here is eq. (17) rewritten:

$$\vec{x}[i+1] = \vec{x}[i] - 2\alpha A^\top A\vec{x}[i] + 2\alpha A^\top \vec{y} \tag{20}$$

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 $AA^\top$?)*

11. Eckart-Young Proof (40 pts.)

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

$$\|A\|_2 = \max_{\vec{s}\in\mathbb{R}^n} \frac{\|A\vec{s}\|}{\|\vec{s}\|} \tag{21}$$

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

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

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

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

$$\left\| \Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix} \right\| = \sqrt{\sum_{i=1}^{k+1} (\sigma_i \alpha_i)^2} \tag{22}$$

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

$$\sqrt{\sum_{i=1}^{k+1} (\sigma_i \alpha_i)^2} \geq \sqrt{\sum_{i=1}^{k+1} \sigma_{k+1}^2 (\alpha_i)^2} \tag{23}$$

and hence

$$\left\| \Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0}_{n-k-1} \end{bmatrix} \right\| \geq \sigma_{k+1} \tag{24}$$

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 $\vec{\alpha}$?)*

(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.)

This concludes the proof of the Eckart-Young Theorem.

[Doodle page! Draw us something if you want or give us suggestions or complaints. You can also use this page to report anything suspicious that you might have noticed.

If needed, you can also use this space to work on problems. But if you want the work on this page to be graded, make sure you tell us on the problem's main page.]

[Extra page. If you want the work on this page to be graded, make sure you tell us on the problem's main page.]

