EECS 16B Designing Information Devices and Systems II
Fall 2019 UC Berkeley Final Exam

Exam location: Proctoring Office

PRINT your student ID: ____

PRINT AND SIGN your name: ____, ________________________
(first and last) (signature)

PRINT your discussion section and GSI(s) (the one(s) you attend): ____

Row Number (front row is 1): __________ Seat Number (left most is 1): __________

Name and SID of the person to your left: ____

Name and SID of the person to your right: ____

Name and SID of the person in front of you: ____

Name and SID of the person behind you: ____

Section 0: Pre-exam questions (4 points)

1. Tell us about something you are proud of this semester. (2 pt)

2. What are you looking forward to in winter break? Describe how you will feel. (2 pts)

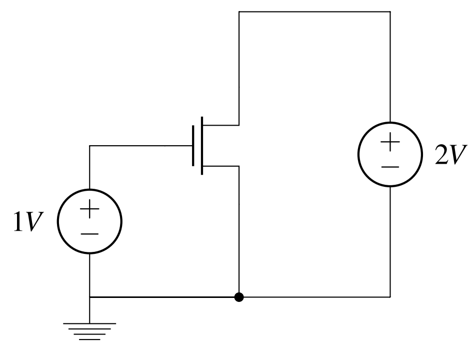
Do not turn this page until the proctor tells you to do so. You can work on Section 0 above before time starts.

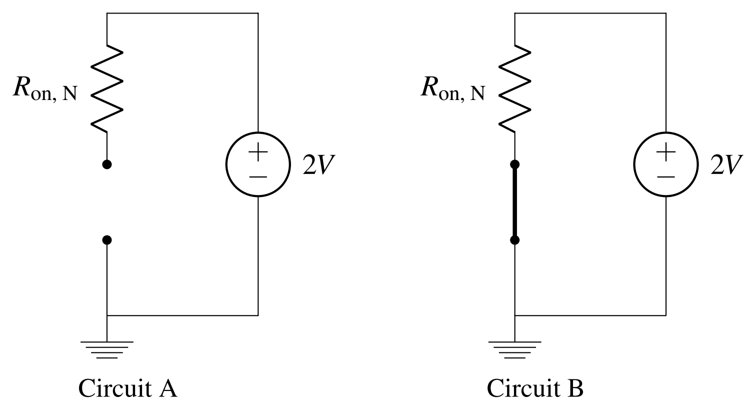
PRINT your name and student ID: ________________________________

3. Transistor Behavior (12 pts)

For all NMOS devices in this problem, $V_{tn} = 0.5\text{V}$. For all PMOS devices in this problem, $|V_{tp}| = 0.6\text{V}$.

(a) **(4 pts)** Which is the equivalent circuit for the right-hand side of the circuit? **Fill in the correct bubble.**

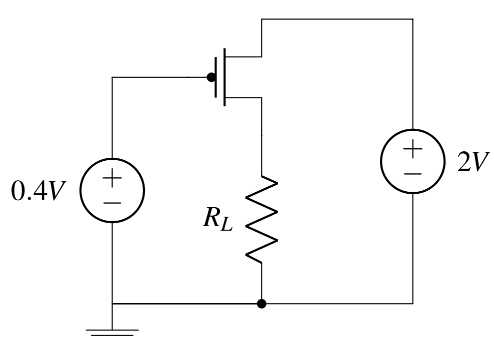


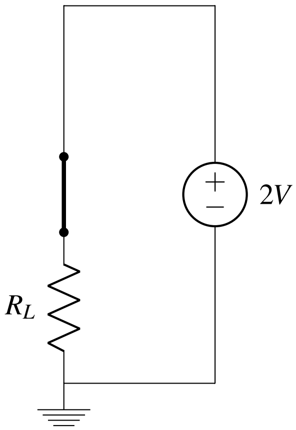


	A	B
Equivalent Circuit	◯	◯

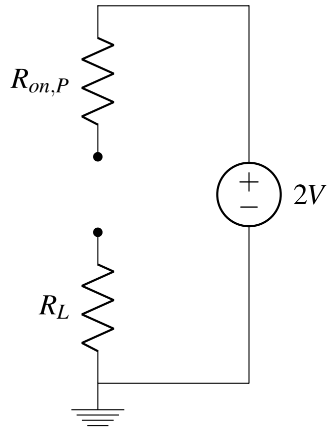
Solution: For the NMOS, $V_{GS} = 1V > V_{tn} = 0.5V$, so the NMOS transistor is on. Thus circuit B is equivalent.

(b) **(4 pts)** Which is the equivalent circuit for the right-hand side of the circuit? **Fill in the correct bubble.**

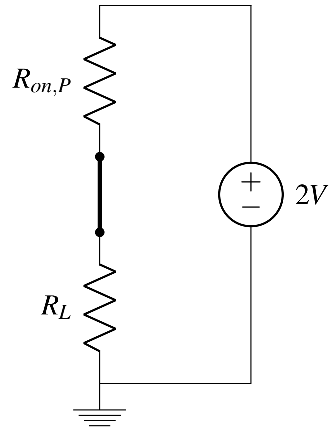




Circuit A



Circuit B

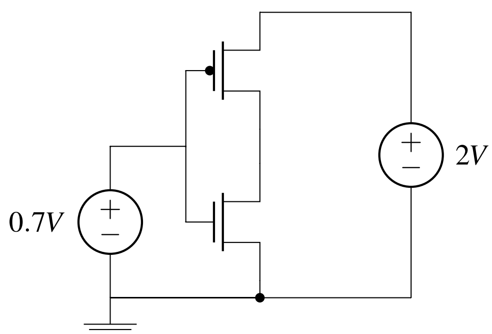


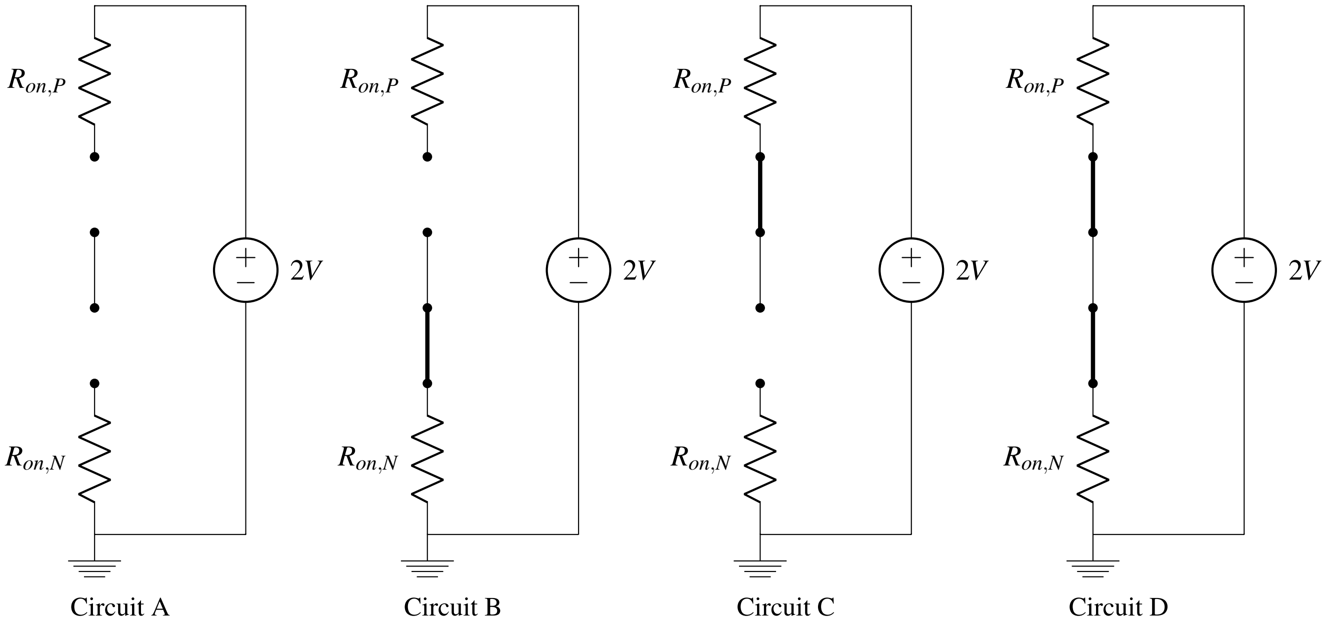
Circuit C

	A	B	C
Equivalent Circuit	◯	◯	◯

Solution: For the PMOS transistor, $|V_{GS}| = 1.6V > |V_{tp}| = 0.6V$, so the PMOS transistor is on. Thus circuit C is equivalent.

(c) **(4 pts)** Which is the equivalent circuit for the right-hand side of the circuit? **Fill in the correct bubble.**





	A	B	C	D
Equivalent Circuit	◯	◯	◯	◯

Solution: For the PMOS transistor, $|V_{GS}| = 1.3V > |V_{tp}| = 0.6V$, so the PMOS transistor is on. For the NMOS transistor, $V_{GS} = 0.7V > V_{tn} = 0.5V$, so the NMOS transistor is on.

Note that in this case, both transistors are on.

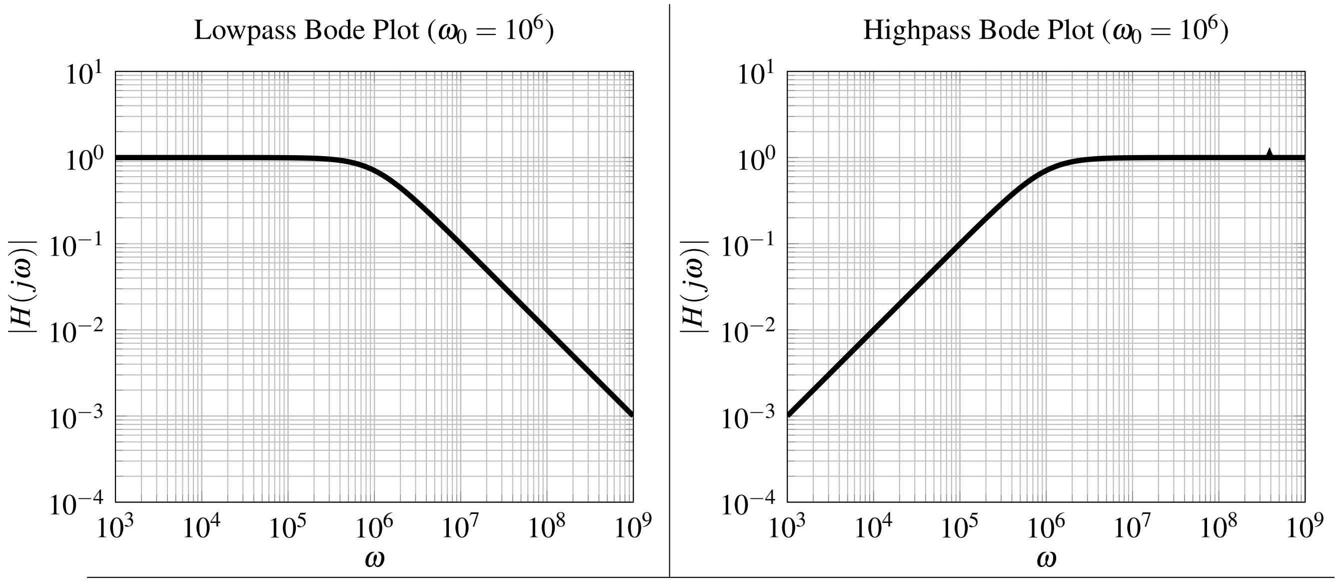
Thus circuit D is equivalent.

Aside: In digital logic, it is usually undesirable to have this state in your system for several reasons. First, the output voltage of the inverter (the voltage at the shared drain of the NMOS and PMOS) will not be either 0 or VDD, which means the output voltage is not at 'true' binary value. In addition, we now have a direct current path through the NMOS and PMOS transistors from VDD to ground. This will burn a lot of power! In reality, all inverters briefly transition through this state where both NMOS and PMOS are on when the inputs change from 1 to 0 or 0 to 1.

PRINT your name and student ID: ______________________________

4. Filter Design and Bode Plots (28 pts)

On the Bode plots below, we have plotted the magnitude responses of first-order low pass filters and high pass filters using the example of cutoff frequency $\omega_0 = 10^6$.



Recall that the transfer functions for such simple low pass filters and high pass filters are:

$$H_{lowpass}(j\omega) = \frac{1}{1+\frac{j\omega}{\omega_0}}; \qquad H_{highpass}(j\omega) = \frac{\frac{j\omega}{\omega_0}}{1+\frac{j\omega}{\omega_0}}.$$

(a) (6 pts) We want to design a bandpass filter that can pass through a 2.4GHz WiFi signal while blocking other interfering signals — FM radio at 100MHz and WiGig at 60GHz. (Recall: Mega = 10^6 and Giga = 10^9.) We will achieve this by cascading lowpass and highpass filters, using ideal op-amp buffers in between to prevent any loading effects.

Unfortunately, when we look in the lab, we only see inductors, 1kΩ resistors, and op-amps.

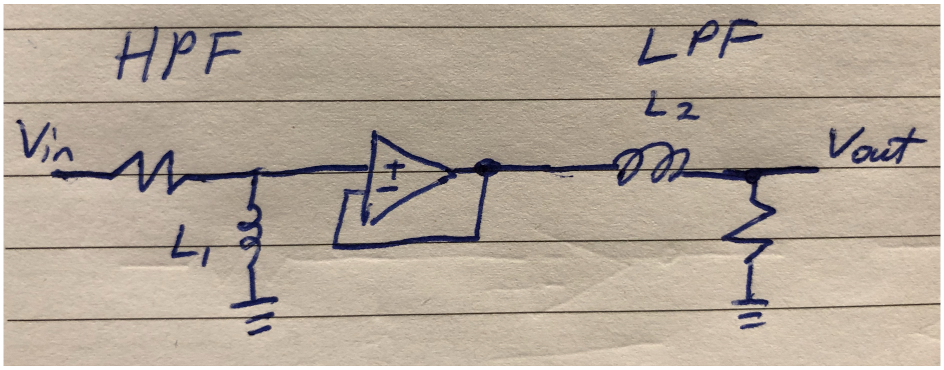
We will start by cascading a single highpass filter followed by a single lowpass filter, with an op-amp buffer in between. **Using only op-amps, two inductors, and resistors (as many as needed), draw the full band-pass filter.** Label V_{in} and V_{out} and label the two inductors with L_1 and L_2. Do not worry about picking the values for L_1 and L_2 in this part.

Solution:

The band-pass filter is drawn below.

Important notes:

- All resistors are $1k\Omega$
- An op-amp in unity-gain configuration (input to the $+$ port, output fed back to the $-$ port) separates the HPF from the LPF



(b) (8 pts) One interfering signal that we want to block is the WiGig signal at 60GHz. If we want to attenuate/reduce the magnitude of the WiGig signal by a factor of about $\sqrt{101} \approx 10$, **What is a candidate 'cutoff frequency' (in** Hz) **desired for this lowpass filter?**

What inductance value should we use for the lowpass filter? Recall that we only have resistors with $1\text{k}\Omega$ resistance. It is fine to give your inductance as a formula — you don't have to simplify it.

For your convenience, here are some calculations that may or may not be relevant:

$\frac{60\times10^9}{2\pi} = 9.549\times10^9$	$\frac{2.4\times10^9}{2\pi} = 382\times10^6$	$\frac{100\times10^6}{2\pi} = 15.9\times10^6$
$60\times10^9\times2\pi = 377\times10^9$	$2.4\times10^9\times2\pi = 15.08\times10^9$	$100\times10^6\times2\pi = 628\times10^6$

(HINT: Look at the relevant Bode plot and read off how far away in frequency from ω_0 you need to be to reduce the magnitude by the desired factor of around 10.*)*

Solution: From the Bode plot, we see that to attenuate a particular frequency by 10x, the cutoff frequency must be 10x lower.

Because we are interested in attenuating $f_{WiGig} = 60\text{GHz}$ by 10x, we find that the cutoff frequency (in Hz) must be $f_0 = \frac{1}{10} f_{WiGig} = 6\text{GHz}$.

Recall that the transfer function for a LR lowpass filter (derived using the phasor-domain voltage divider) is

$$H(j\omega) = \frac{1}{1+\frac{j\omega}{\omega_0}} = \frac{R}{R+j\omega L} = \frac{1}{1+\frac{j\omega}{\frac{R}{L}}}$$

This indicates that $\omega_0 = \frac{R}{L}$. Thus

$$L = \frac{R}{\omega_0} = \frac{1\text{k}\Omega}{2\pi\cdot 6\times10^9\text{Hz}} = \frac{1}{12\pi\times10^6} \approx \frac{1}{377\times10^5} \approx 26.5\text{nH}$$

(c) (14 pts) Another interfering signal that we want to block is FM radio at 100MHz and we want to reduce its magnitude by a factor of around 100. We decide to use multiple highpass filters in a row (separated by ideal op-amp buffers) to attenuate the FM radio signal more strongly. We design the system with the highpass filter cutoff frequencies all at 1GHz. In this case, **what inductor value should each of the highpass filters use?** Recall that we only have resistors with $1\text{k}\Omega$ resistance. It is fine to give your inductance as a formula — you don't have to simplify it.

For your convenience, here are some calculations that may or may not be relevant:

$\frac{60\times10^9}{2\pi} = 9.549\times10^9$	$\frac{2.4\times10^9}{2\pi} = 382\times10^6$	$\frac{100\times10^6}{2\pi} = 15.9\times10^6$
$60\times10^9\times2\pi = 377\times10^9$	$2.4\times10^9\times2\pi = 15.08\times10^9$	$100\times10^6\times2\pi = 628\times10^6$

How many highpass filters must we cascade in order to attenuate the FM signal at 100MHz by a factor of around 100?

Draw the full circuit for your complete filter including op-amp buffers, the lowpass filter, and the highpass filters.

Solution:

We are given a cutoff frequency of $f_0 = 1\text{GHz}$. We can therefore solve for the inductor value:

$$L = \frac{R}{\omega_0} = \frac{R}{2\pi f_0} = \frac{1\text{k}\Omega}{2\pi 1\times10^9 Hz} = \frac{1}{2\pi\times10^6}\text{henry} \approx 159\text{nH}$$

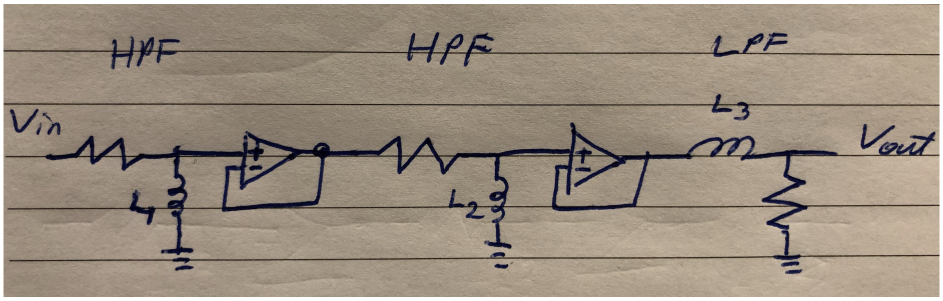
Because the filter cutoff frequency is 10x higher than the FM radio signal, the FM signal is attenuated by a factor of roughly 10x by a single filter.

Therefore, to get a factor of 100 attenuation, we must use 2 filters in cascade (each one providing 0x attenuation, for a total of 100x).

The full band-pass filter is drawn below.

Important notes:

- All resistors are $1k\Omega$
- An op-amp in unity-gain configuration (input to the $+$ port, output fed back to the $-$ port) separates the HPF from the LPF
- $L_1 = L_2 = 159\text{nH}$ for the highpass filters
- $L_3 = 26.5\text{nH}$ for the lowpass filters



PRINT your name and student ID: ____

5. Controllable Canonical Form (26 pts)

Suppose that we have a two-dimensional continuous time system governed by:

$$\frac{d}{dt}\vec{x}(t) = \begin{bmatrix} 1 & -1 \\ 0 & -4 \end{bmatrix} \vec{x}(t) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t)$$

We would like to put this system into Controllable Canonical Form (CCF) to use state feedback to place the eigenvalues at desired locations. For your convenience, the characteristic polynomial $\det\left(\lambda I - \begin{bmatrix} 1 & -1 \\ 0 & -4 \end{bmatrix}\right) = (\lambda - 1)(\lambda + 4) = \lambda^2 + 3\lambda - 4$.

(a) (4 pts) **Is the system stable? Why or why not?**

Solution: The system is not stable because it has an eigenvalue $\lambda = +1$ that has a non-negative real part. The system will tend to explode along that direction.
The $\lambda = -4$ eigenvalue is perfectly stable by contrast.

(b) (16 pts) Recall that our original system is:

$$\frac{d}{dt}\vec{x}(t) = \begin{bmatrix} 1 & -1 \\ 0 & -4 \end{bmatrix} \vec{x}(t) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t) \quad \text{with } \det\left(\lambda I - \begin{bmatrix} 1 & -1 \\ 0 & -4 \end{bmatrix}\right) = \lambda^2 + 3\lambda - 4.$$

We would like to change coordinates to bring the system into CCF:

$$\frac{d}{dt}\vec{z}(t) = \begin{bmatrix} 0 & 1 \\ a_0 & a_1 \end{bmatrix} \vec{z}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$$

Compute the T basis such that $\vec{z}(t) = T^{-1}\vec{x}(t)$ or equivalently (if you want), give us T^{-1} directly.
To help you along, here are some calculations already done for you:

$$\begin{bmatrix} 1 & 0 \\ 1 & -4 \end{bmatrix}^{-1} = \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 \\ 1 & a \end{bmatrix}^{-1} = \begin{bmatrix} -a & 1 \\ 1 & 0 \end{bmatrix}$$

What is a_0? What is a_1?

Solution:

The characteristic polynomial of $\begin{bmatrix} 0 & 1 \\ a_0 & a_1 \end{bmatrix}$ is $\lambda^2 - a_1\lambda - a_0$ and since changing coordinates doesn't change the characteristic polynomial, we know that $a_0 = 4$ and $a_1 = -3$.

Now, we can compute the controllability matrix for the original system $C = \begin{bmatrix} 1 & 0 \\ 1 & -4 \end{bmatrix}$ as well as for the system that is already in CCF: $C_z = \begin{bmatrix} 0 & 1 \\ 1 & -3 \end{bmatrix}$.

At this point, we know that the original system's A matrix in the C basis looks like the transpose of A_z. Since we also know that C_z has the same basic relationship, we know that $T = CC_z^{-1}$ is the basis in which the original system's matrix will look like A_z. Computing this:

$$T = CC_z^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & -3 \end{bmatrix}^{-1} \tag{1}$$

$$= \begin{bmatrix} 1 & 0 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} 3 & 1 \\ -1 & 1 \end{bmatrix}. \tag{3}$$

We could also compute T^{-1} if we wanted by the same essential reasoning:

$$T^{-1} = C_z C^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & -4 \end{bmatrix}^{-1} \tag{4}$$

$$= \begin{bmatrix} 0 & 1 \\ 1 & -3 \end{bmatrix} \frac{1}{4} \begin{bmatrix} 4 & 0 \\ 1 & -1 \end{bmatrix} \tag{5}$$

$$= \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix}. \tag{6}$$

If we wanted, we could check our work by computing

$$T^{-1} \begin{bmatrix} 1 & -1 \\ 0 & -4 \end{bmatrix} T = \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 0 & -4 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ -1 & 1 \end{bmatrix} \tag{7}$$

$$= \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 \\ 4 & -4 \end{bmatrix} \tag{8}$$

$$= \begin{bmatrix} 0 & 1 \\ 4 & -3 \end{bmatrix}. \tag{9}$$

And so, this checks out as far as the system matrix goes. We can also see what happens with $T^{-1}\vec{b} = \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ which also checks out.

Finally, it is important to note that we can also compute the a_0, a_1 by computing

$$C^{-1}A^2\vec{b} = \frac{1}{4} \begin{bmatrix} 4 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 16 \end{bmatrix} \tag{10}$$

$$= \begin{bmatrix} 4 \\ -3 \end{bmatrix}. \tag{11}$$

That also works.

(c) (6 pts) Using state feedback

$$u(t) = \begin{bmatrix} \widetilde{k}_0 & \widetilde{k}_1 \end{bmatrix} \vec{z}(t)$$

place the closed-loop eigenvalues at $\lambda_1 = -1$, $\lambda_2 = -2$. **What is $\widetilde{k}_0$? What is $\widetilde{k}_1$?**

(Notice that we are asking for the feedback gains in terms of $\vec{z}(t)$ not the original $\vec{x}(t)$. If you have time, feel free to check your work by the original $\vec{x}(t)$.)

Solution:

The key here is to realize what we want the new characteristic polynomial to be. $(\lambda-\lambda_1)(\lambda-\lambda_2) = (\lambda+1)(\lambda+2) = \lambda^2+3\lambda+2$. In CCF form, the coefficients of the characteristic polynomial are visible on the bottom row. So we need $4+\widetilde{k}_0 = -2$ which means that $\widetilde{k}_0 = -6$. We also need $-3+\widetilde{k}_1 = -3$ which means that $\widetilde{k}_1 = 0$.

To check our work, we realize that $[-6,0]T^{-1} = [-6,0]\frac{1}{4}\begin{bmatrix}1 & -1\\ 1 & 3\end{bmatrix} = [-\frac{3}{2},\frac{3}{2}]$ are the claimed gains in the original coordinates. This makes the original coordinate closed-loop matrix be $\begin{bmatrix}1 & -1\\ 0 & -4\end{bmatrix} + \begin{bmatrix}1\\ 1\end{bmatrix}[-\frac{3}{2},\frac{3}{2}] = \begin{bmatrix}1-\frac{3}{2} & -1+\frac{3}{2}\\ 0-\frac{3}{2} & -4+\frac{3}{2}\end{bmatrix} = \begin{bmatrix}-\frac{1}{2} & +\frac{1}{2}\\ -\frac{3}{2} & -\frac{5}{2}\end{bmatrix}$ which has eigenvalues at -2 and -1 as desired since $(\lambda+\frac{1}{2})(\lambda+\frac{5}{2})+\frac{3}{4} = \lambda^2+3\lambda+\frac{5}{4}+\frac{3}{4} = \lambda^2+3\lambda+2$.

PRINT your name and student ID: ____

6. You knew this was coming (16 pts)

In the last midterm, you began solving a minimum norm problem in which we had a wide matrix A and wanted to solve $A\vec{x} = \vec{y}$ while minimizing the norm $\|C\vec{x}\|$ where C was also a wide matrix. You then focused on some special cases where certain matrices were assumed to have orthonormal columns. The nonorthonormal case wasn't resolved. This question is about the heart of what remains to be resolved.

Suppose that we have a matrix A_f, but we don't know whether or not the columns of A_f are linearly independent or orthonormal. How can you use the results from the SVD $A_f = U_f \Sigma_f V_f^\top = \sum_{i=1}^r \sigma_{f,i} \vec{u}_{f,i} \vec{v}_{f,i}^\top$ to help project a vector $\vec{s}$ onto the subspace spanned by the columns of A_f? Here, the $\sigma_{f,1}, \ldots, \sigma_{f,r}$ are all strictly positive. **Give an expression for a matrix P_f such that $P_f\vec{s}$ is the projection of $\vec{s}$ in the subspace spanned by the columns of A_f.**

Also give an expression for an $\vec{x}_f$ so that $A_f\vec{x}_f = P_f\vec{s}$.

Solution: There are multiple ways to approach this problem. If we start with the first part, it is easiest to see that the columns of $U_f = [\vec{u}_{f,1}, \vec{u}_{f,2}, \cdots, \vec{u}_{f,r}]$ form an orthonormal basis for the span of the columns of A_f. This is what the SVD tells us, if we think of the SVD being given to us in the compact form. (The question didn't specify explicitly which form. However, pointing out explicitly what you wanted U_f to be is a good idea.) This means that we can project onto this subspace by just using $P_f = U_f U_f^\top$ since this is an orthonormal basis.

The second part also has multiple answers that are valid. One approach is to say that $P_f\vec{s}$ is clearly in the column-span of A_f by construction. This means that Gaussian elimination can solve $A_f\vec{x}_f = P_f\vec{s}$ and we can let it pick any solution that it wants to. This is a valid answer, but it doesn't really give us an expression strictly speaking.

The approach that most naturally gives us an expression is to look for the minimum norm solution. Here, we can use the Moore-Penrose Pseudoinverse $A_f^\dagger = V_f \Sigma_f^\dagger U_f^\top$. If you assume that the SVD is given in compact form, then Σ_f is a square $r \times r$ matrix and $\Sigma_f^\dagger = \Sigma_f^{-1}$. The natural thing to write is $\vec{x}_f = A_f^\dagger P_f \vec{s} = A_f^\dagger U_f U f^\top \vec{s} = V_f \Sigma_f^\dagger U_f^\top U_f U_f^\top \vec{s} = V_f \Sigma_f^\dagger U_f^\top \vec{s} = V_f \Sigma_f^{-1} U_f^\top \vec{s}$ if we assume the compact form of the SVD for the last step. Here, we are effectively defining $V_f = [\vec{v}_{f,1}, \vec{v}_{f,2}, \cdots, \vec{v}_{f,r}]$.

Notice that this again verifies the connection between least-squares and the pseudoinverse — the pseudoinverse does least-squares when no true solution exists, and among the least-squares solutions, it picks the minimum norm one.

One can also go from the Pseudoinverse solution to the projection itself by saying that $P_f = A_f A_f^\dagger$ which, when fully substituted, simplifies to just $U_f U_f^\top$ assuming the compact form of the SVD because everything else cancels out.

PRINT your name and student ID: ________________________________

7. Movie Ratings and PCA (36 pts)

Recall from the lecture on PCA that we can think of movie ratings as a structured set of data. For every person i and movie j, we have that person's rating $R_{i,j}$ (thought of as a real number).

Suppose that there are m movies and n people. Let's think about arranging this data into a big $n \times m$ matrix R with people corresponding to rows and movies corresponding to columns. So the entry in the i-th row and j-th column should be $R_{i,j}$ above.

(a) (6 pts) Suppose we believe that there is actually an underlying pattern to this rating data and that a separate study has revealed that every movie is characterized by a set of characteristics: say action and comedy. This means that every movie j has a pair of numbers $a[j], c[j]$ that define it. At the same time, every person i has a pair of sensitivities $s_a[i], s_c[i]$ that essentially define that person's preferences. A person i will rate the movie j as $R_{i,j} = s_a[i]a[j] + s_c[i]c[j]$.

If we arrange the sensitivities into a pair of n-dimensional vectors $\vec{s_a}, \vec{s_c}$, and the movie characteristics into a pair of m-dimensional vectors $\vec{a}, \vec{c}$, **use outer products to express the rating matrix R in terms of these vectors $\vec{s_a}, \vec{s_c}, \vec{a}, \vec{c}$.**

Recall that the outer product of a real vector $\vec{u}$ with a real vector $\vec{v}$ is $\vec{u}\vec{v}^\top$ — the reverse order as compared to the usual inner product.

Solution:

Here $R = \vec{s_a}\vec{a}^\top + \vec{s_c}\vec{c}^\top$. The $\vec{s_a}$ has one entry for every person, and the people are arranged as rows So to get the sizes to work out, it has to be that we do it in this order. If it helps, imagine that there is exactly 1 movie. In that case, R is a tall matrix that is only 1 wide. The only tall things we have are $\vec{s_a}$ and $\vec{s_b}$.

The outer product $\vec{s_a}\vec{a}^\top$ is therefore a matrix of the same size as R that is filled with $s_a[i]a[j]$ in the i-th row and j-th column. Similarly $\vec{s_c}\vec{c}^\top$ is a matrix of the same size as R that is filled with $s_c[i]c[j]$ in the i-th row and j-th column. Adding them together gives us a matrix whose i-th row and j-th column contains $R_{i,j} = s_a[i]a[j] + s_c[i]c[j]$ as desired.

(b) (6 pts) Now suppose that we want to discover the underlying nature of movies from the data R itself.

Suppose for this part, that you have four observed rating data vectors (corresponding to four different movies being rated by six individuals).

All of the movie data vectors just happened to be multiples of the following 6-dimensional vector $\vec{v} = \begin{bmatrix} 2 \\ -2 \\ 3 \\ -4 \\ 4 \\ 0 \end{bmatrix}$. (For your convenience, note that $\|\vec{v}\| = 7$.)

You arrange the movie data vectors as the columns of a matrix R given by:

$$R = \begin{bmatrix} | & | & | & | \\ -\vec{v} & -2\vec{v} & 2\vec{v} & 4\vec{v} \\ | & | & | & | \end{bmatrix} \tag{12}$$

You want to perform PCA (for movies) using the SVD of the matrix R to better understand the pattern in your data.

The first "principal component vector" is a *unit vector* that tells which direction we would want to project the columns of R onto to get the best rank-1 approximation for R.

Find this first principal component vector of R to explain the nature of your movie data points.

(HINT: You don't need to actually compute any SVDs in this simple case. Also, be sure to think about what size vector you want as the answer. Don't forget that you want a unit vector!)

Solution:

Principal component analysis is in general about understanding how best to approximate our (potentially) high-dimensional data (like recordings from a microphone, or in this case, a movie's ratings by lots of people) with its lower-dimensional essence. The first principal component is about seeing which one-dimensional line best approximates the data points — i.e. which is the line for which projecting the data points onto it results in "estimates" that are as close as possible to the data points.

In the case of this problem, every point is explicitly given as a multiple of a single vector $\vec{v}$ and so the data already lies on such a straight line going through the origin. So, the first principal component is just along the direction of $\vec{v}$. Because a principal component represents a direction, it is conventional to normalize the vector to have unit length. In this case, we are told that the vector $\vec{v}$ has length 7, and so the answer is $\frac{\vec{v}}{7}$.

(Because the line is all that matters, you could also have used the negative of this $-\frac{\vec{v}}{7}$.)

We can also do this using the SVD.

The singular value decomposition of a matrix R is a way of decomposing R into a sum of rank 1 matrices. In this sum the i^{th} rank 1 matrix is formed from taking the outer product of normalized column vectors $\vec{u}_i$ and normalized row vectors $\vec{v}_i^\top$, scaled by their respective singular values σ_i.

(Note that the $\vec{v}_i^\top$ row vectors in the SVD decomposition $A = U\Sigma V^\top$ are completely unrelated to the $\vec{v}$ column vector that we have defined for our data matrix A above.)

Looking at our given R, we can see that the matrix itself is rank 1 as the columns are all multiples of the same vector: $\vec{v}$. Seeing this we realize we can rewrite the matrix R as the following outer product:

$$R = \begin{bmatrix} | \\ \vec{v} \\ | \end{bmatrix} \begin{bmatrix} -1 & -2 & 2 & 4 \end{bmatrix} \tag{13}$$

However the SVD requires we normalize the vectors $\vec{u}_1$ and $\vec{v_1^\top}$. In order to reconstruct A properly we must scale back with the norms that we divided out to normalize.

$\|[-1,-2,2,4]\| = \sqrt{25} = 5$ and $\|\vec{v}\| = 7$. Consequently, when we pull that out, we get $\sigma_1 = 35$ as the singular value that corresponds to the first (and only) principal component.

Thus we can write the SVD of R as:

$$R = \begin{bmatrix} | \\ \frac{\vec{v}}{7} \\ | \end{bmatrix} 35 \begin{bmatrix} \frac{-1}{5} & \frac{-2}{5} & \frac{2}{5} & \frac{4}{5} \end{bmatrix} \tag{14}$$

Now we just have to pick which normalized vector to deem the principal component. Since our data (the movie ratings) are collected as columns we choose $\frac{\vec{v}}{7}$ as the principal component.

(c) (12 pts) Suppose that now, we have two more data points (corresponding to two more movies being rated by the same set of six people) that are multiples of a different vector $\vec{p}$ where:

$\vec{p} = \begin{bmatrix} 6 \\ 3 \\ -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. (For your convenience, note that $\|\vec{p}\| = 7$ and that $\vec{p}^\top \vec{v} = 0$.)

We augment our ratings data matrix with these two new data points to get:

$$R = \begin{bmatrix} | & | & | & | & | & | \\ -\vec{v} & -2\vec{v} & 2\vec{v} & 4\vec{v} & -3\vec{p} & 3\vec{p} \\ | & | & | & | & | & | \end{bmatrix} \tag{15}$$

Find the first two principal components corresponding to the nonzero singular values of R. This is what we would use to best project the movie data points onto a two-dimensional subspace.

What is the first principal component vector? What is the second principal component vector? Justify your answer.

(Hint: Think about the inner product of $\vec{v}$ and $\vec{p}$ and what that implies for being able to appropriately decompose R. Again, very little computation is required here.)

Solution: The solution to the previous part tells you what we need to do. We need to find the best two-dimensional subspace that best represents our data.

We start by taking the SVD of R.

The columns of R are all multiples of two vectors: $\vec{v}$ and $\vec{p}$. Each of these can be used to create a rank 1 matrix, and these can be summed together to form R.

Since $\vec{v}$ and $\vec{p}$ are orthogonal to one another, our life is easier. This problem's R matrix is made especially nice by seeing that a data point is either purely in the $\vec{v}$ direction, or the $\vec{p}$ direction.

Using this knowledge we rewrite R as:

$$R = \begin{bmatrix} | \\ \vec{v} \\ | \end{bmatrix} \begin{bmatrix} -1 & -2 & 2 & 4 & 0 & 0 \end{bmatrix} + \begin{bmatrix} | \\ \vec{p} \\ | \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & -3 & 3 \end{bmatrix}.$$

The orthogonality relationships demanded by the SVD are clearly satisfied since the row-vectors involved above have disjoint support (i.e. when one is nonzero, the other is zero) and the columns are orthogonal since we've been told so.

However for the SVD the vectors: $\vec{u}_1$, $\vec{v}_1^\top$, $\vec{u}_2$ and $\vec{v}_2^\top$ must be normalized and each rank 1 matrix must be scaled by the appropriate σ_i to allow the sum to properly reconstruct R. We also need to figure out which σ_i is bigger so we can order them properly. In the previous part, we have already done the calculations for $\vec{v}$'s part in this story. So what remains is the $\vec{p}$ part. Clearly the norm of the relevant row is $3\sqrt{2}$ which the norm of the relevant column is 7. So the singular value in question is $21\sqrt{2}$.

Using this we can rewrite R as:

$$= \begin{bmatrix} | \\ \frac{\vec{v}}{7} \\ | \end{bmatrix} 35 \begin{bmatrix} \frac{-1}{5} & \frac{-2}{5} & \frac{2}{5} & \frac{4}{5} & 0 & 0 \end{bmatrix} + \begin{bmatrix} | \\ \frac{\vec{p}}{7} \\ | \end{bmatrix} 21\sqrt{2} \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{-3}{3\sqrt{2}} & \frac{3}{3\sqrt{2}} \end{bmatrix}.$$

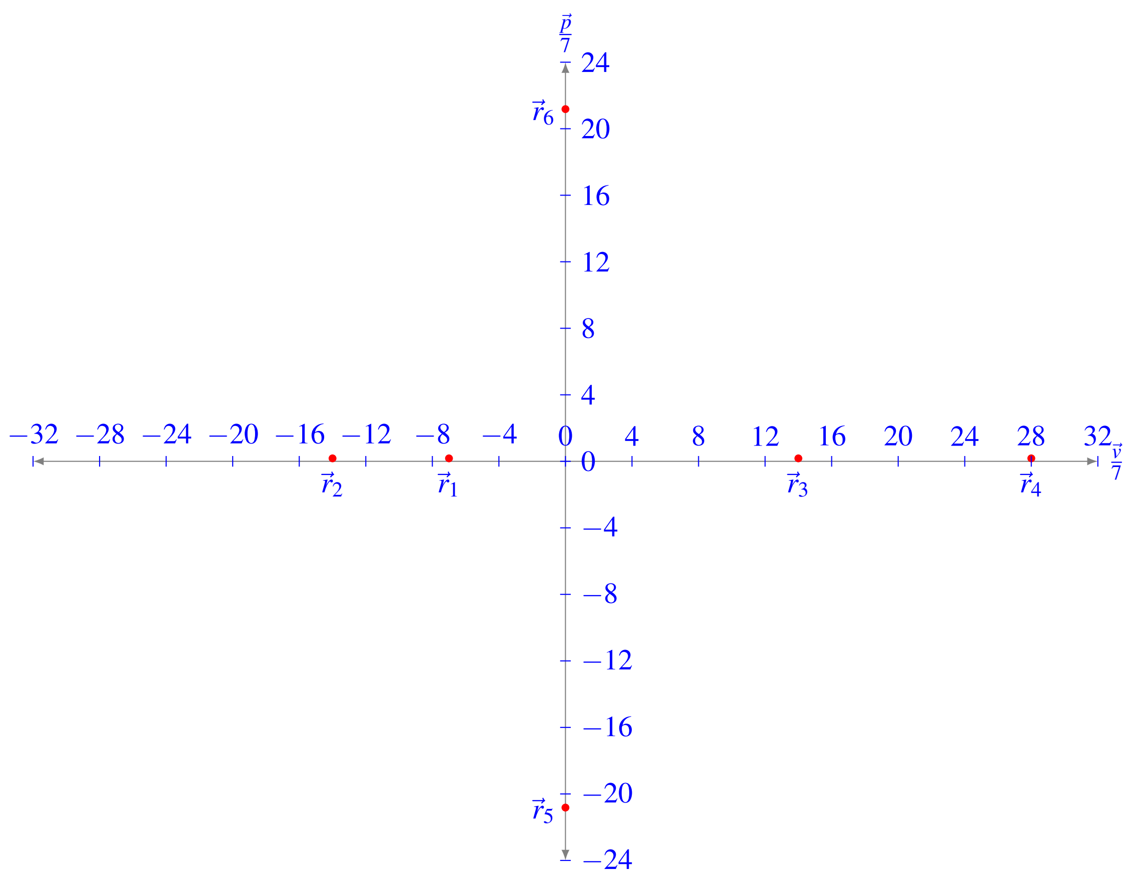
So, what is bigger 35 or $21\sqrt{2}$? This is the same as asking which is bigger 5 or $3\sqrt{2}$. Squaring both tells us that this is the same as asking which is bigger, 25 or 18. Clearly, the first term is larger.
From this we see that our singular values are $\sigma_1 = 35$ and $\sigma_2 = 21\sqrt{2}$ since $35 > 21\sqrt{2}$. Thus $\frac{\vec{v}}{7}$ which corresponds to σ_1 is still the first principal component vector and $\frac{\vec{p}}{7}$ which corresponds to σ_2 is the second principal component vector.

(d) (12 pts) In the previous part, you had

$$R = \begin{bmatrix} | & | & | & | & | & | \\ -\vec{v} & -2\vec{v} & 2\vec{v} & 4\vec{v} & -3\vec{p} & 3\vec{p} \\ | & | & | & | & | & | \end{bmatrix}$$

with $\|\vec{v}\| = 7$ and $\|\vec{p}\| = 7$, satisfying $\vec{p}^\top \vec{v} = 0$.
If we use $\vec{r}_i$ to denote the i-th column of R, **plot the movie data points $\vec{r}_i$ projected onto the first and second principal component vectors.** The coordinate along the first principal component should be represented by horizontal axis and the coordinate along the second principal component should be the vertical axis. **Label each point**.

Solution:
Once we know what the principal components are, we know that the first four data points are just multiples of the first principal component and the last two data points are just multiples of the second principal component. What multiples? For the first four, the multiples are clearly $-7, -14, 14, 28$ since the norm of $\vec{v}$ is 7. For the final two, the multiples are clearly $-21, +21$ since the norm of $\vec{p}$ is also 7.
Plotting:



PRINT your name and student ID: ________________________________

8. Latch (46 pts)

The circuit below is a type of latch, which is one of the fundamental components of memory in many digital systems. The latch is a bistable circuit, which means that there are two possible stable states: one representing a stored '1' bit and the other a stored '0' bit.

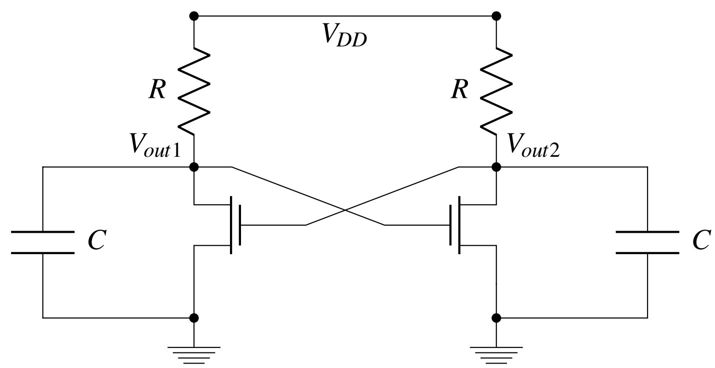


Figure 7: Simplified latch: the gate capacitances have been pulled out explicitly.

(a) (6 pts) To get a basic understanding of the stable operating points for the latch, consider the following simplified circuit using the pure switch model for MOSFETs (and a threshold voltage of $\frac{V_{DD}}{2}$).

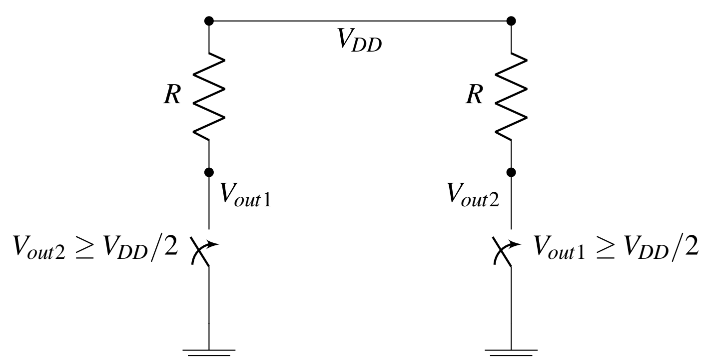


Figure 8: Pure switch model for a latch

First assume that $V_{out1} = 0$. What is V_{out2}? Are the left and right switches open or closed?

	Open or V_{DD}	**Closed or 0**
Left Switch	◯	◯
Right Switch	◯	◯
V_{out2}	◯	◯

Solution: If $V_{out1} = 0$, then the switch on the right is open because that transistor is clearly off. (You can look at the switch model or think about the threshold of $V_{DD}/2$, either way.) At that point, there can be no current flowing through the resistor above it, and $V_{out2} = V_{DD}$. Now, we can look at the switch to the left. It is clearly closed since that transistor is clearly on.

So, the first set of answers should be:

	Open or V_{DD}	**Closed or 0**
Left Switch	○	●
Right Switch	●	○
V_{out2}	●	○

Notice how in this, this is self-fulfilling. Having the left transistor closed means that there is current in the corresponding resistor and V_{out1} is indeed pulled down to 0 since the transistor is being modeled with no output resistance. So the initial assumption is self-consistent.

Suppose that $V_{out1} = V_{DD}$. What is V_{out2}? Are the left and right switches open or closed?

	Open or V_{DD}	**Closed or 0**
Left Switch	○	○
Right Switch	○	○
V_{out2}	○	○

Solution: If $V_{out1} = V_{DD}$, then the switch on the right is closed because that transistor is clearly on. (You can look at the switch model or think about the threshold of $V_{DD}/2$, either way.) At that point, there is current flowing through the resistor above it, and $V_{out2} = 0$. This is because we are assuming that there is no resistance in our pure switch model of the transistor. Now, we can look at the switch to the left. It is clearly open since that transistor is clearly off because V_{out2} is low.

So, the second set of answers should be:

	Open or V_{DD}	**Closed or 0**
Left Switch	●	○
Right Switch	○	●
V_{out2}	○	●

Notice how in this case also, this is self-fulfilling. Having the left transistor open means that there is no current in the corresponding resistor and V_{out1} is indeed V_{DD} for consistency.

(b) (10 pts) To get an understanding of latch dynamics, we will now break it down into smaller pieces. Below is one half of the latch circuit.

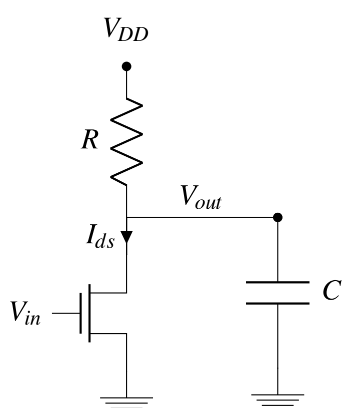


Figure 9: Latch half-circuit

Write a differential equation for the voltage V_{out} in terms of the drain to source current I_{ds}. Treat I_{ds} as some specified input signal and treat the transistor as a current source connected to ground. (i.e. There is no dependence on V_{in} in this part. In this part, treat the I_{ds} as a constant that you are given.)

Solution: Using KCL at the output node, let's equate the currents in (from the resistor) to the currents out (to the capacitor and through the transistor).

$$\frac{V_{DD} - V_{out}}{R} = I_{ds} + C\frac{d}{dt}(V_{out})$$

Rearranging terms, we get:

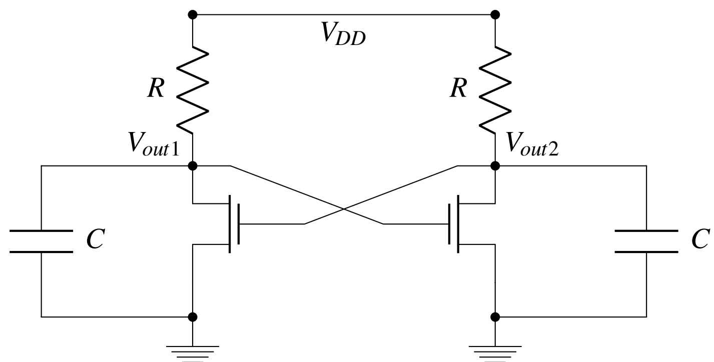
$$\frac{dV_{out}}{dt} = \frac{1}{C}(\frac{V_{DD}}{R} - \frac{V_{out}}{R} - I_{ds})$$

Notice that this can be made into a more familiar form as:

$$\frac{d}{dt}V_{out}(t) = \frac{-1}{RC}V_{out}(t) + \frac{V_{DD}}{RC} - \frac{I_{ds}}{C}$$

Here, we see that $\frac{V_{DD}}{RC} - \frac{I_{ds}}{C}$ is playing the role of the "input" in our standard view of a first-order linear scalar differential equation.

(c) (10 pts)



For this circuit, we care about more detailed analog characteristics of the MOSFETs, so we will model their behavior more accurately as current sources that are controlled by their gate voltages V_{in} with the following equation:

$$I_{ds} = g(V_{in})$$

Where $g(V_{in})$ is a some nonlinear function. **Using this I_{ds} expression together with the result from the previous part, write down a system of differential equations for V_{out1} and V_{out2} in vector form:**

$$\frac{d}{dt}\begin{bmatrix} V_{out1}(t) \\ V_{out2}(t) \end{bmatrix} = \vec{f}\left(\begin{bmatrix} V_{out1}(t) \\ V_{out2}(t) \end{bmatrix}\right).$$

(Hint: Notice that the latch above can be constructed by taking two of the circuit in Figure 9, and connecting the V_{out} of one to the gate V_{in} of the other and vice-versa.)

Solution:

We observe that V_{in} for the half-circuit on the left is V_{out2}. That means that the I_{ds} for the differential equation corresponding to V_{out1} is given by $g(V_{out2})$ and similarly, the I_{ds} for the differential equation corresponding to V_{out2} is given by $g(V_{out1})$.

Writing those two equations with that substitution gives:

$$\frac{d}{dt}V_{out1}(t) = \frac{-1}{RC}V_{out1}(t) + \frac{V_{DD}}{RC} - \frac{g(V_{out2}(t))}{C}$$

$$\frac{d}{dt}V_{out2}(t) = \frac{-1}{RC}V_{out2}(t) + \frac{V_{DD}}{RC} - \frac{g(V_{out1}(t))}{C}$$

Putting this into vector form yields:

$$\frac{d}{dt}\begin{bmatrix} V_{out1}(t) \\ V_{out2}(t) \end{bmatrix} = \vec{f}\left(\begin{bmatrix} V_{out1}(t) \\ V_{out2}(t) \end{bmatrix}\right) = \begin{bmatrix} \frac{-1}{RC}V_{out1}(t) + \frac{V_{DD}}{RC} - \frac{g(V_{out2}(t))}{C} \\ \frac{-1}{RC}V_{out2}(t) + \frac{V_{DD}}{RC} - \frac{g(V_{out1}(t))}{C} \end{bmatrix}.$$

(d) (20 pts) For the rest of this problem, assume that your analysis yields the following system of nonlinear differential equations:

$$\begin{bmatrix} \frac{dV_{out1}}{dt} \\ \frac{dV_{out2}}{dt} \end{bmatrix} = \begin{bmatrix} 1 - V_{out_1} - g(V_{out2}) \\ 1 - V_{out_2} - g(V_{out1}) \end{bmatrix}$$

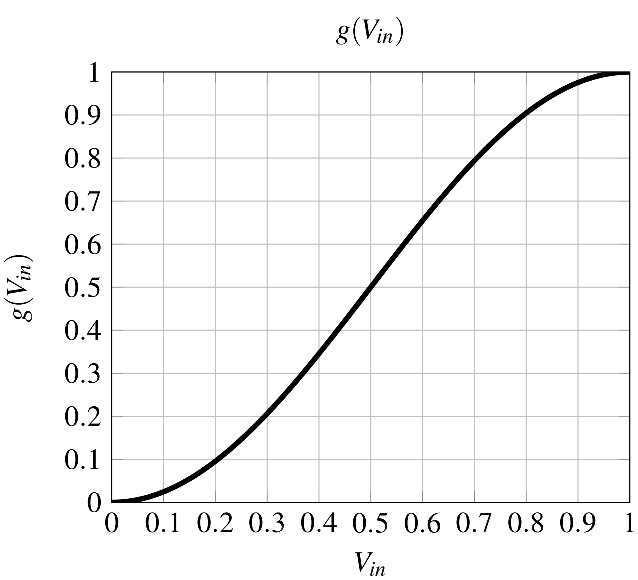
Suppose that you put this latch into an ideal circuit simulator, and measure $g(V_{in})$ and $\frac{dg}{dt}(V_{in})$. The results from these measurements are shown in the graphs below. From your simulations, you also can see that for the following initial conditions, the latch voltages do not change over time:

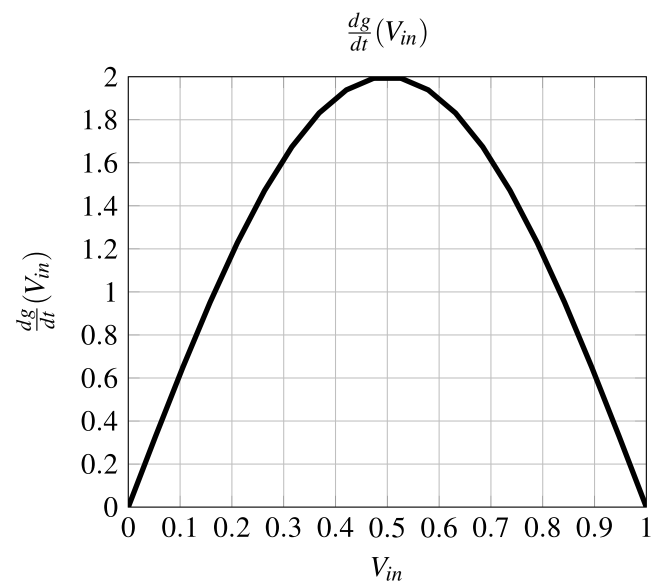
$$\begin{bmatrix} V^*_{out1} \\ V^*_{out2} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

Use the graphs below to linearize the differential equations around the three operating points.

Write a linearized system of differential equations around each of those operating points $\begin{bmatrix} V^*_{out1} \\ V^*_{out2} \end{bmatrix}$.

For which of the provided $\begin{bmatrix} V^*_{out1} \\ V^*_{out2} \end{bmatrix}$ points is the latch locally stable? For which of the provided points is the latch locally unstable? Why?





Solution: To linearize these differential equations, we need to take the derivative of $\vec{f}$ with respect to $\begin{bmatrix} V_{out1} \\ V_{out2} \end{bmatrix}$. This will give a matrix. The derivative of the first component of $\vec{f}$ with respect to V_{out1}

is just -1 since that dependence is linear. Similarly for the derivative of the second component of $\vec{f}$ with respect to V_{out2} being -1. The off-diagonal derivatives are clearly $-g'(V_{out2})$ and $-g'(V_{out1})$ respectively, where $g'(V_{in}) = \frac{dg}{dV_{in}}(V_{in})$. This means that we get:

$$\vec{f}\left(\begin{bmatrix} V_{out1} \\ V_{out2} \end{bmatrix}\right) = \vec{f}\left(\begin{bmatrix} V^*_{out1} + \delta V_1 \\ V^*_{out2} + \delta V_2 \end{bmatrix}\right) \tag{16}$$

$$= \vec{f}\left(\begin{bmatrix} V^*_{out1} \\ V^*_{out2} \end{bmatrix}\right) + \frac{d}{d\vec{v}} f(\vec{v})\Big|_{\vec{v}=\begin{bmatrix} V^*_{out1} \\ V^*_{out2} \end{bmatrix}} \begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix} + \vec{w} \tag{17}$$

$$= \begin{bmatrix} 1 - V^*_{out1} - g(V^*_{out2}) \\ 1 - V^*_{out2} - g(V^*_{out1}) \end{bmatrix} + \begin{bmatrix} -1 & -g'(V^*_{out2}) \\ -g'(V^*_{out1}) & -1 \end{bmatrix} \begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix} + \vec{w} \tag{18}$$

where the disturbance term $\vec{w}$ captures all the approximation errors that are coming from linearization. That is what lets us write equalities above instead of $\approx$.

Using this in the differential equation, since the operating point is not changing with time, we get:

$$\frac{d}{dt}\begin{bmatrix} V_{out1}(t) \\ V_{out2}(t) \end{bmatrix} = \frac{d}{dt}\begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) = \begin{bmatrix} -1 & -g'(V^*_{out2}) \\ -g'(V^*_{out1}) & -1 \end{bmatrix} \begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) + \begin{bmatrix} 1 - V^*_{out1} - g(V^*_{out2}) \\ 1 - V^*_{out2} - g(V^*_{out1}) \end{bmatrix} + \vec{w}(t)$$

.

With the general form established, we need to actually write this out for the neighborhoods of the provided operating points.

For $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, we know from the provided graphs that $g(0) = 0$, $g(1) = 1$, and $g'(0) = 0, g'(1) = 0$. Here, we view $\begin{bmatrix} \delta V_1(t) \\ \delta V_2(t) \end{bmatrix}$ as the local deviation from the given operating point of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. In other words, $\begin{bmatrix} V_1(t) \\ V_2(t) \end{bmatrix} = \begin{bmatrix} \delta V_1(t) \\ \delta V_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Substituting everything in, we get:

$$\frac{d}{dt}\begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) = \begin{bmatrix} 1-1-0 \\ 1-0-1 \end{bmatrix} + \begin{bmatrix} -1 & -0 \\ -0 & -1 \end{bmatrix} \begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) + \vec{w}(t) \tag{19}$$

$$= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) + \vec{w}(t). \tag{20}$$

This clearly has two eigenvalues of -1 and so is locally stable.

For $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we know from the provided graphs that $g(0) = 0$, $g(1) = 1$, and $g'(0) = 0, g'(1) = 0$. Here, we view $\begin{bmatrix} \delta V_1(t) \\ \delta V_2(t) \end{bmatrix}$ as the local deviation from the given operating point of $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. In other words,

$\begin{bmatrix} V_1(t) \\ V_2(t) \end{bmatrix} = \begin{bmatrix} \delta V_1(t) \\ \delta V_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Substituting everything in, we get:

$$\frac{d}{dt}\begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) = \begin{bmatrix} 1-1-0 \\ 1-0-1 \end{bmatrix} + \begin{bmatrix} -1 & -0 \\ -0 & -1 \end{bmatrix}\begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) + \vec{w}(t) \tag{21}$$

$$= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) + \vec{w}(t). \tag{22}$$

This also clearly has two eigenvalues of -1 and so is locally stable.

So both of these operating points for the latch are stable. Small disturbances will be rejected by the local dynamics and the memory will regenerate its value. This is important in digital circuits.

For $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$, we know from the provided graphs that $g(0.5) = 0.5$, $g(0.5) = 0.5$, and $g'(0.5) = 2$. Here, we view $\begin{bmatrix} \delta V_1(t) \\ \delta V_2(t) \end{bmatrix}$ as the local deviation from the given operating point of $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$. In other words, $\begin{bmatrix} V_1(t) \\ V_2(t) \end{bmatrix} = \begin{bmatrix} \delta V_1(t) \\ \delta V_2(t) \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$. Substituting everything in, we get:

$$\frac{d}{dt}\begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) = \begin{bmatrix} 1-0.5-0.5 \\ 1-0.5-0.5 \end{bmatrix} + \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}\begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) + \vec{w}(t) \tag{23}$$

$$= \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}\begin{bmatrix} \delta V_1 \\ \delta V_2 \end{bmatrix}(t) + \vec{w}(t). \tag{24}$$

Notice how the constant terms went away (which makes sense since this is an operating point) but now we need to calculate some eigenvalues to understand if this is locally stable or not.

$$\det\left(\lambda I - \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}\right) = (\lambda+1)^2 - 4 = \lambda^2 + 2\lambda - 3 = (\lambda+3)(\lambda-1).$$

This has one stable eigenvalue of -3 and an unstable eigenvalue of $+1$. This means that this operating point is unstable overall.

You will learn in more advanced digital circuits courses that this instability is actually desirable for this particular operating point because that is what lets us more effectively use this circuit as a latch. To set the value of the latch, we just have to drive it past the unstable point and then the unstable point will push the state to the region of attraction of the desired stable operating point, which will then hold that value. If the local gain g' at the middle operating point was too low, then this would also be a stable operating point and the circuit could end up remembering a useless intermediate voltage that is not cleanly interpretable as a binary 0 or binary 1.

PRINT your name and student ID: ______________________________

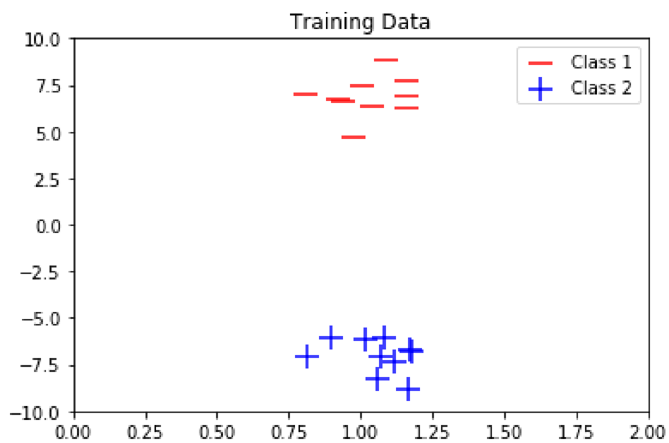
9. Choosing cost functions for learning classification (12 pts)

In this problem, we want to design classifiers for data that comes from two classes "+" and "-". We have labeled data $\{(\vec{x}_i, \ell_i)\}$ from which we want to learn a vector $\vec{w}$ so that we can use the sign of $\vec{w}^\top \vec{x}$ to classify $\vec{x}$. To do this, we are going to be trying to minimize a sum $c_{total}(\vec{w}) = \sum_i c^{\ell_i}(\vec{x}_i^\top \vec{w})$.

The cost functions we are considering include:

- Squared loss: $c^+(p) = (p-1)^2$ and $c^-(p) = (p-(-1))^2$
- Exponential loss: $c^+(p) = e^{-p}$ and $c^-(p) = e^{+p}$
- Logistic loss: $c^+(p) = \ln\left(1+e^{-p}\right)$ and $c^-(p) = \ln\left(1+e^{+p}\right)$

For the plotted data, which cost functions will give reasonable answers? Select all that apply.

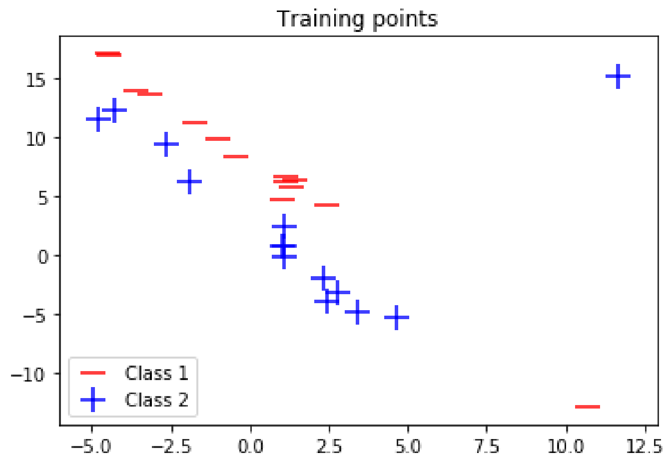


	Squared	**Exponential**	**Logistic**
Reasonable Choice	◯	◯	◯

Solution: The solution is:

	Squared	**Exponential**	**Logistic**
Reasonable Choice	●	●	●

Why? This data is pretty balanced and so any of these approaches will work. There are no outliers of any kind to confuse least-squares.

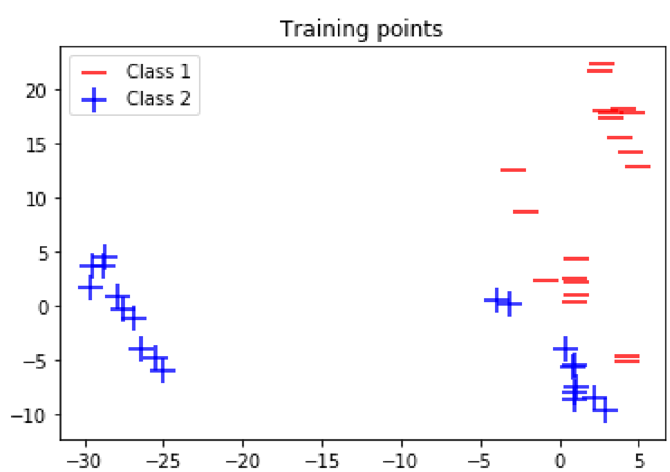


	Squared	Exponential	Logistic
Reasonable Choice	○	○	○

Solution:

This data clearly has a “wrongly classified” outlier that is a “+” point deep in what should be “-” territory. Both the squared loss and (especially) exponential loss will be quite sensitive to this while logistic loss will end up largely ignoring this point. In such cases, we really need to be using logistic loss of the choices offered here.

	Squared	Exponential	Logistic
Reasonable Choice	○	○	●

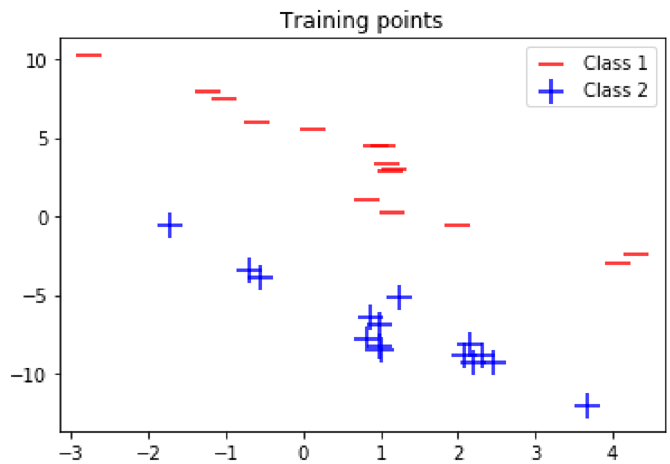


	Squared	Exponential	Logistic
Reasonable Choice	◯	◯	◯

Solution:

In this case, we clearly have two masses of points in each category. The + category is particularly striking because the second mass of points is deep within what we would consider "+" category. These type of points will confuse squared loss quite a bit while both exponential and logistic loss don't care about points that are deeply within their own proper territories. They focus on things closer to the boundaries.

	Squared	Exponential	Logistic
Reasonable Choice	◯	●	●



	Squared	Exponential	Logistic
Reasonable Choice	◯	◯	◯

Solution:

This is actually an example with thin parallel categories that have no outliers. For this kind of data, even squared-loss will work fine because of the balanced nature of the training data. However, in this kind of example, because the mass distribution doesn't necessarily reflect the parallel structure of the true boundary, mean-based classification won't work as well. However, that is not a choice that you are offered here.

	Squared	Exponential	Logistic
Reasonable Choice	●	●	●

PRINT your name and student ID: ______________________________

10. DFT and Polynomials (34 pts)

Consider the polynomial-style DFT basis that comes from evaluating the monomials x^0, x^1, x^2, x^3 on the 4-th roots of unity $x_0 = e^{j\frac{2\pi}{4}0} = 1$, $x_1 = e^{j\frac{2\pi}{4}1} = j$, $x_2 = e^{j\frac{2\pi}{4}2} = -1$, $x_3 = e^{j\frac{2\pi}{4}3} = -j$.

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & e^{j\frac{2\pi}{4}} & e^{j2\frac{2\pi}{4}} & e^{j3\frac{2\pi}{4}} \\ 1 & e^{j2\frac{2\pi}{4}} & e^{j4\frac{2\pi}{4}} & e^{j6\frac{2\pi}{4}} \\ 1 & e^{j3\frac{2\pi}{4}} & e^{j6\frac{2\pi}{4}} & e^{j9\frac{2\pi}{4}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix}.$$

(a) (4 pts) The DFT coefficients $\vec{F}$ are related to a vector of samples $\vec{f}$ by the relationship $\vec{f} = B\vec{F}$. In other words, $\vec{F}$ represents $\vec{f}$ in the basis given by the columns of B. Similarly for the DFT coefficients $\vec{G}$ and a vector of samples $\vec{g}$ — they too satisfy the relationship $\vec{g} = B\vec{G}$.
What are the DFT coefficients $\vec{H}$ for $\vec{h} = \alpha\vec{f} + \beta\vec{g}$ in terms of $\vec{F}$ and $\vec{G}$. Here, α and β are constant real numbers.

Solution: The DFT coefficients of $\vec{H}$ are

$$\vec{H} = B^{-1}\vec{h} = B^{-1}(\alpha\vec{f} + \beta\vec{g}) = \alpha \cdot B^{-1}\vec{f} + \beta \cdot B^{-1}\vec{g} = \alpha\vec{F} + \beta\vec{G}.$$

(b) (8 pts) **Explicitly find the DFT coefficients $\vec{F}$ of the vector** $\vec{f} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}$ i.e. we want $\vec{F}$ so that $\vec{f} = B\vec{F}$.

Solution:

$$\vec{F} = B^{-1}\vec{f} = \frac{1}{4}B^*\vec{f} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{j}{2} \\ 0 \\ \frac{j}{2} \end{bmatrix}.$$

You could also have noticed that the samples look like they came from $\sin(\theta)$ for $\theta = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ and since $\sin(\theta) = \frac{1}{2j}e^{j\theta} - \frac{1}{2j}e^{-j\theta}$, we can get those same coefficients directly by using an interpolation perspective.

(c) (10 pts) **Show that if $\vec{f}$ is a real vector, then:**

- $F[0]$ **is always real and so is** $F[2]$**.**
- $F[1] = \overline{F}[3]$**.**

(HINT: What do you know about B^{-1}?)

Solution: Let

$$\vec{f} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

where a,b,c,d are real numbers. Then,

$$\vec{F} = B^{-1}\vec{f} = \frac{1}{4}B^*\vec{f} = \frac{1}{4}\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & -j & -1 & j\\ 1 & -1 & 1 & -1\\ 1 & j & -1 & -j\end{bmatrix}\begin{bmatrix}a\\ b\\ c\\ d\end{bmatrix} = \frac{1}{4}\begin{bmatrix}a+b+c+d\\ a-bj-c+dj\\ a-b+c-d\\ a+bj-c-dj\end{bmatrix}.$$

Hence, we see that $F[0]$ and $F[2]$ are always real since a,b,c,d are real, while $F[1] = \overline{F}[3]$ since everything multiplied by j in one is multiplied by $-j$ in the other and vice versa.

This is a direct demonstration of the desired result.

Alternatively, you could have noticed that if we consider $B = [\vec{b}_0, \vec{b}_1, \vec{b}_2, \vec{b}_3]$, then $\vec{b}_0$ and $\vec{b}_2$ are real while $\vec{b}_1$ and $\vec{b}_3$ are complex conjugates by inspection. The formula $\vec{F} = \frac{1}{4}B^*\vec{f}$ means that $F[i] = \frac{1}{4}\vec{b}_i^*\vec{f}$. Now, $\vec{b}_0$ and $\vec{b}_2$ are real and that immediately gives us what we want since the conjugate transpose of a real vector is just the transpose. Similarly, $F[1] = \overline{F[3]}$ since $\overline{F[3]} = \overline{\frac{1}{4}\vec{b}_3^*\vec{f}} = \frac{1}{4}\overline{\vec{b}_3^*}\vec{f} = \frac{1}{4}\vec{b}_1^*\vec{f} = F[1]$.

Either approach works.

(d) (12 pts) We notice that when we use the standard polynomial interpolation approach $\hat{f}(e^{j\theta}) = F[0] + F[1]e^{j\theta} + F[2]e^{j2\theta} + F[3]e^{j3\theta}$, we often get nonzero imaginary parts even though our samples $\vec{f}$ were purely real. Here $\vec{F}$ are the DFT coefficients corresponding to $\vec{f}$ so that $\vec{f} = B\vec{F}$.

Give new constants k_1, k_3 so that

$$\hat{f}(e^{j\theta}) = F[0] + F[1]e^{jk_1\theta} + F[2]\frac{e^{j2\theta}+e^{-j2\theta}}{2} + F[3]e^{jk_3\theta}$$

will always return a real number for each $\theta \in [0, 2\pi]$, and actually interpolates the data points — i.e. $f[i] = \hat{f}(e^{j\frac{2\pi i}{4}})$ for $i = 0,1,2,3$.

Prove that this $\hat{f}(e^{j\theta})$ is always real for $\theta \in [0, 2\pi]$ no matter what real values the vector $\vec{f}$ has. *(HINT: For proving realness, use what is already established by the previous part (even if you didn't get the previous part!) and write out $\hat{f}(e^{j\theta})$ in terms of constants, sines, and/or cosines. Using a polar form for $F[1]$ might be helpful.)*

Solution: Here, we use $k_1 = 1$ and $k_3 = -1$. This is the natural choice. From the previous part, we know that $F[1] = \overline{F}[3]$. This implies that $F[1]e^{j\theta}$ and $F[3]e^{-j\theta}$ are complex conjugates, and hence,

$$F[1]e^{j\theta} + F[3]e^{-j\theta} \quad \text{is real.}$$

If you wanted, you could also have written $F[1] = |F[1]|e^{j\angle F[1]}$ and so $F[3] = |F[1]|e^{-j\angle F[1]}$ by complex conjugacy. Therefore $F[1]e^{j\theta} + F[3]e^{-j\theta} = |F[1]|e^{j\angle F[1]}e^{j\theta} + |F[1]|e^{-j\angle F[1]}e^{-j\theta} = |F[1]|(e^{j(\angle F[1]+\theta)} + e^{-j(\angle F[1]+\theta)}) = 2|F[1]|\cos(\angle F[1]+\theta)$ which is clearly real.

Next, from the previous part, we know that $F[0]$ and $F[2]$ are both real. Since $\frac{e^{j2\theta}+e^{-j2\theta}}{2} = \cos(2\theta)$ is real, it follows that

$$F[0] + F[2]\frac{e^{j2\theta}+e^{-j2\theta}}{2} \quad \text{is real.}$$

Therefore, it follows that $\hat{f}(e^{j\theta})$ is always real for $\theta \in [0, 2\pi]$ no matter what real values the vector $\vec{f}$ has.

In actuality, the constraint that the function actually interpolates the data points just requires $k_1 = 1 + q4$ for any integer q. So $k_1 = 1$ is the natural choice, but $-3, 5, 9$ all could work. For realness, we just need $k_3 = -k_1$, we don't actually need it to be -1 per se — just properly paired with k_1.

The reason for this is that we simply need $e^{jk_1\theta}$ to agree with the first column $\vec{b}_1$ when evaluated at $\theta = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ and any of the k_1 values above will do that.

Contributors:

- Sidney Buchbinder.
- Anant Sahai.
- Kuan-Yun Lee.
- Nikhil Shinde.
- Nathan Lambert.
- Pavan Bhargava.