EECS 16B Designing Information Devices and Systems II
Fall 2021 Final Exam — Final Exam

Exam location: <examloc>

PRINT your student ID: ____

PRINT AND SIGN your name: _________________, _________________ _____________________
(last) (first) (signature)

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. **Honor Code: 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. **How are you hoping to relax during winter break? (2 pts)**

3. **Think about something that you know how to do well and enjoy doing. What is it? (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: ______________________________

4. SVD Calculation (11 points)

(a) (7pts) Let $A = \begin{bmatrix} 0 & 2 \\ \sqrt{2} & 0 \\ 0 & 1 \end{bmatrix}$. The eigenvalues of $AA^\top = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix}$ are $5, 2, 0$ with corresponding unnormalized eigenvectors $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$.

In addition, we know that:

$$A^\top \begin{bmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 5 & 0 & 0 \end{bmatrix} \tag{1}$$

Write out the singular value decomposition (SVD) of A in any form you choose (outer product form, compact, or full). (No need to show work.)

(b) (4pts) **What is the best rank 1 approximation of A (i.e., what is the rank 1 matrix B that minimizes $\|A - B\|_F$)? Write your answer as a 3×2 dimensional matrix.** (No need to show work)

PRINT your name and student ID: ______________________________

5. Gram-Schmidt with Complex Vectors (11 points)

Suppose you are given two complex vectors $\vec{v}_1 = \begin{bmatrix} \mathrm{j} \\ 1 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. In addition, the complex inner products are given as:

$$\langle \vec{v}_1, \vec{v}_1 \rangle = \vec{v}_1^* \vec{v}_1 = 2 \tag{2}$$
$$\langle \vec{v}_2, \vec{v}_2 \rangle = \vec{v}_2^* \vec{v}_2 = 1 \tag{3}$$
$$\langle \vec{v}_1, \vec{v}_2 \rangle = \vec{v}_2^* \vec{v}_1 = +\mathrm{j} \tag{4}$$
$$\langle \vec{v}_2, \vec{v}_1 \rangle = \vec{v}_1^* \vec{v}_2 = -\mathrm{j} \tag{5}$$

Use the Gram-Schmidt algorithm to generate an orthonormal sequence of vectors $(\vec{u}_1, \vec{u}_2)$ from the list of vectors $(\vec{v}_1, \vec{v}_2)$, starting with $\vec{v}_1$. **If you start the Gram-Schmidt algorithm with $\vec{v}_2$, you will receive zero credit for this entire problem.**

(a) (3pts) **What is the first vector $\vec{u}_1$?** (No need to show work)

(b) (8pts) **What is the second vector $\vec{u}_2$?** (Show work)

PRINT your name and student ID: ________________________________

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

PRINT your name and student ID: ____

6. Choosing cost functions for learning classification (9 points)

We have labeled data $\{(\vec{x}_i, y_i)\}$ where the labels y_i are either '+' or '-'. 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 will minimize a sum $c_{\text{total}}(\vec{w}) = \sum_i c^{y_i}(\vec{x}_i^\top \vec{w})$. We consider cost functions:

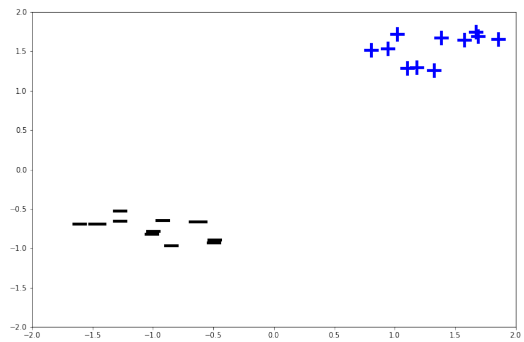
(a) Squared loss: $c^+(p) = (p-1)^2$ and $c^-(p) = (p-(-1))^2$

(b) Exponential loss: $c^+(p) = e^{-p}$ and $c^-(p) = e^{+p}$

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

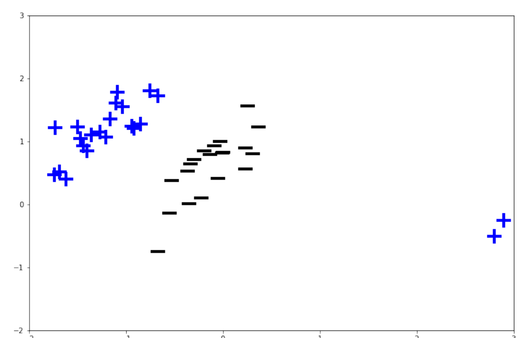
For the plotted data, which cost functions will result in learning reasonable classifiers? (Multiple options might be correct and should be marked for full credit, but every plot has at least one correct option.)

Cost Function	
Squared	☐
Exponential	☐
Logistic	☐

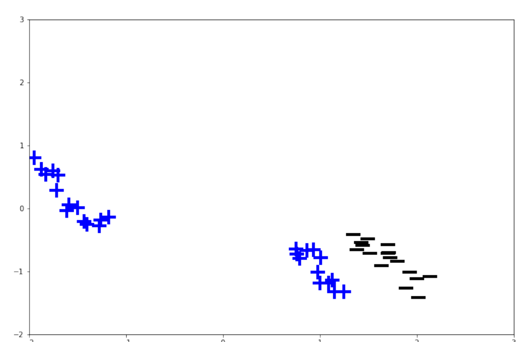


Cost Function	
Squared	☐
Exponential	☐
Logistic	☐

Note: You want to make sure that the two right-most '+' points don't influence your learned classifier by too much.



Cost Function	
Squared	☐
Exponential	☐
Logistic	☐



PRINT your name and student ID: ______________________________

7. Nonlinear Feedback Control (14 points)

Consider the following differential equation model for a control system:

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{x}(t) = \vec{f}(\vec{x}(t)) + \vec{g}(\vec{x}(t))u(t) \tag{6}$$

Assume that our state $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is two-dimensional and our input $u \in \mathbb{R}$ is a scalar. Let our dynamics functions be

$$\vec{f}(\vec{x}) = \begin{bmatrix} x_1^3 \\ x_2^2 \end{bmatrix},\ \vec{g}(\vec{x}) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} \tag{7}$$

(a) (4pts) **Show that $x_1^* = -1, x_2^* = 1, u^* = 1$ is an equilibrium point.** (i.e. if we find ourselves exactly in this state with exactly this input and no disturbances, we will not move from here.)

PRINT your name and student ID: ____________________

(b) (8pts) Recall that our nonlinear controlled differential equation for the state $\vec{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$ is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{x}(t) = \begin{bmatrix} x_1^3(t) + x_2(t)u(t) \\ x_2^2(t) + x_1(t)u(t) \end{bmatrix}.$$

The equilibrium of interest has $\vec{x}^* = \begin{bmatrix} -1 \\ +1 \end{bmatrix}$ and $u^* = +1$.

We define the variables $\vec{\widetilde{x}} = \vec{x} - \vec{x}^*$ and $\widetilde{u} = u - u^*$ as the deviations from the equilibrium point $(\vec{x}^*, u^*)$.

What should the matrix A and vector $\vec{b}$ be in our locally linearized model $\frac{\mathrm{d}}{\mathrm{d}t}\vec{\widetilde{x}}(t) \approx A\vec{\widetilde{x}}(t) + \vec{b}\widetilde{u}(t)$? (Show work and give specific numbers for A and $\vec{b}$.)

(c) (2pts) **Is the resulting linearized system from the above part controllable?**

PRINT your name and student ID: ________________________________

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

PRINT your name and student ID: ____

8. Bode Plot Analysis of Band Stop and Notch Filters (35 points)

Consider a system with an input that consists of three signals, i.e. $v_i(t) = s_1(t) + s_2(t) + n(t)$, where $s_1(t) = 2\cos\left(10^2 t\right)$ and $s_2(t) = 2\cos\left(10^7 t\right)$ are the two desired signals, and $n(t) = 10\cos\left(10^4 t\right)$ is an interference signal. We wish to construct a filter so that the amplitude of the interference signal is attenuated to be at most $\frac{1}{10}$ times the amplitude of the desired signals after filtering.

We will attempt two different ways of achieving this behavior. In parts (a), (b), and (c), we will analyze the first approach, which is to use a band stop filter that attenuates the interference signal and passes the desired signals. In part (d), we will analyze the second approach, which is to create a notch at the interference signal frequency.

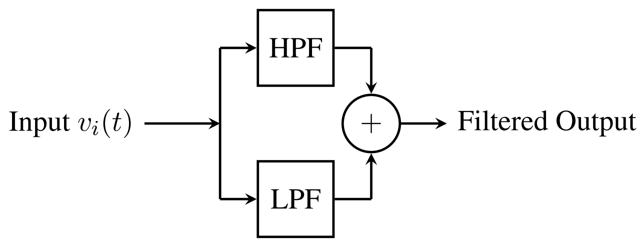


Figure 1: Block diagram of a Band Stop Filter. The HPF and LPF blocks represent a high-pass filter and a low-pass filter respectively. *NOTE*: Remember, you can always turn block diagrams into circuits by using the corresponding circuit for the transfer function in the block diagram, and isolating transfer functions with buffers.

PRINT your name and student ID: ______________________________

(a) (10pts) Let's start by analyzing the HPF path in fig. 1, which should pass $s_2(t) = 2\cos\left(10^7 t\right)$ and attenuate $n(t) = 10\cos\left(10^4 t\right)$. Assume that the high-pass filter transfer function is $H_{\mathrm{HPF}}(\mathrm{j}\omega) = \frac{\mathrm{j}\frac{\omega}{10^6}}{1+\mathrm{j}\frac{\omega}{10^6}}$.

i. **Draw the magnitude and phase Bode plots of the high-pass filter $H_{\mathrm{HPF}}(\mathrm{j}\omega) = \frac{\mathrm{j}\frac{\omega}{10^6}}{1+\mathrm{j}\frac{\omega}{10^6}}$.**

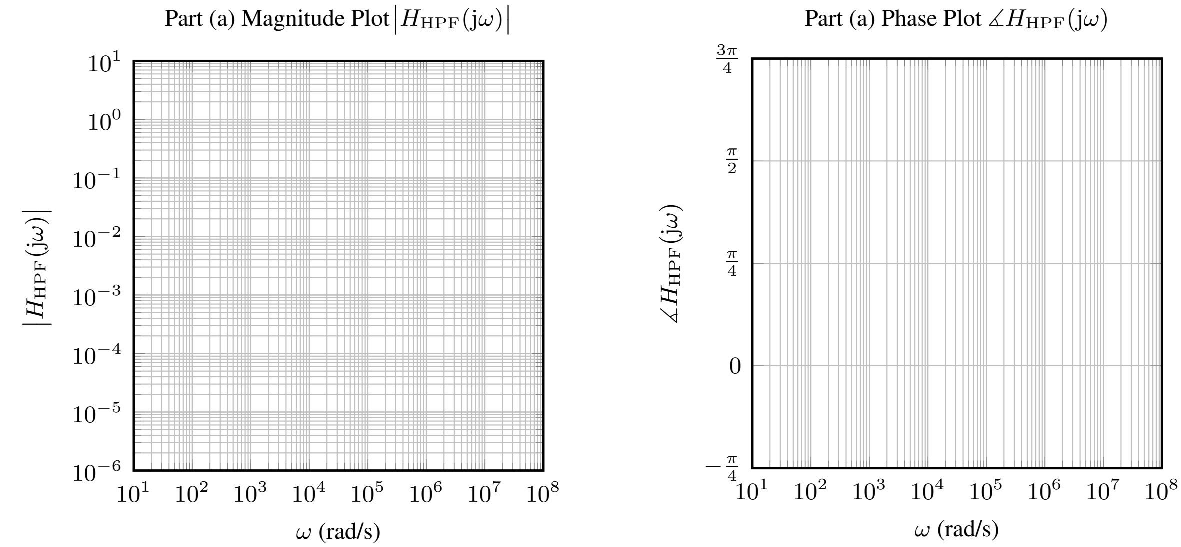


Figure 2: Part (a) Magnitude and Phase Bode Plots for the transfer function $H_{\mathrm{HPF}}(\mathrm{j}\omega)$.

ii. **By reading the corresponding values from the Bode plots, write down the approximate output signal expressions corresponding to $s_1(t) = 2\cos\left(10^2 t\right)$, $s_2(t) = 2\cos\left(10^7 t\right)$ and $n(t) = 10\cos\left(10^4 t\right)$ after high-pass filtering.** *NOTE*: You do not need to compute exact magnitude and phase values using the transfer function. Just use the Bode approximation by reading from the plots.

PRINT your name and student ID: ____

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

PRINT your name and student ID: ____

(b) (17pts) Now let's analyze the LPF path in fig. 1, which should pass $s_1(t) = 2\cos\left(10^2 t\right)$ and attenuate $n(t) = 10\cos\left(10^4 t\right)$. Assume that you are allowed to use multiple copies of a low-pass filter given by $H_{\text{LPF}}(\text{j}\omega) = \frac{1}{1+\text{j}\frac{\omega}{10^3}}$ and a unity gain buffer given by $H_{\text{buf}}(\text{j}\omega) = 1$.

i. Remember that we wish to attenuate the amplitude of $n(t) = 10\cos\left(10^4 t\right)$ to be at most $\frac{1}{10}$ times the amplitude of $s_1(t) = 2\cos\left(10^2 t\right)$ after filtering. **At least how many copies of $H_{\text{LPF}}(\text{j}\omega)$ and $H_{\text{buf}}(\text{j}\omega)$ do we need to cascade to achieve this goal? What is the overall transfer function $H_1(\text{j}\omega)$ of this cascaded low-pass filter?** *NOTE*: You do not have to simplify the expression.

ii. **Draw the magnitude and phase Bode plots of the overall low-pass filter $H_1(\text{j}\omega)$.**

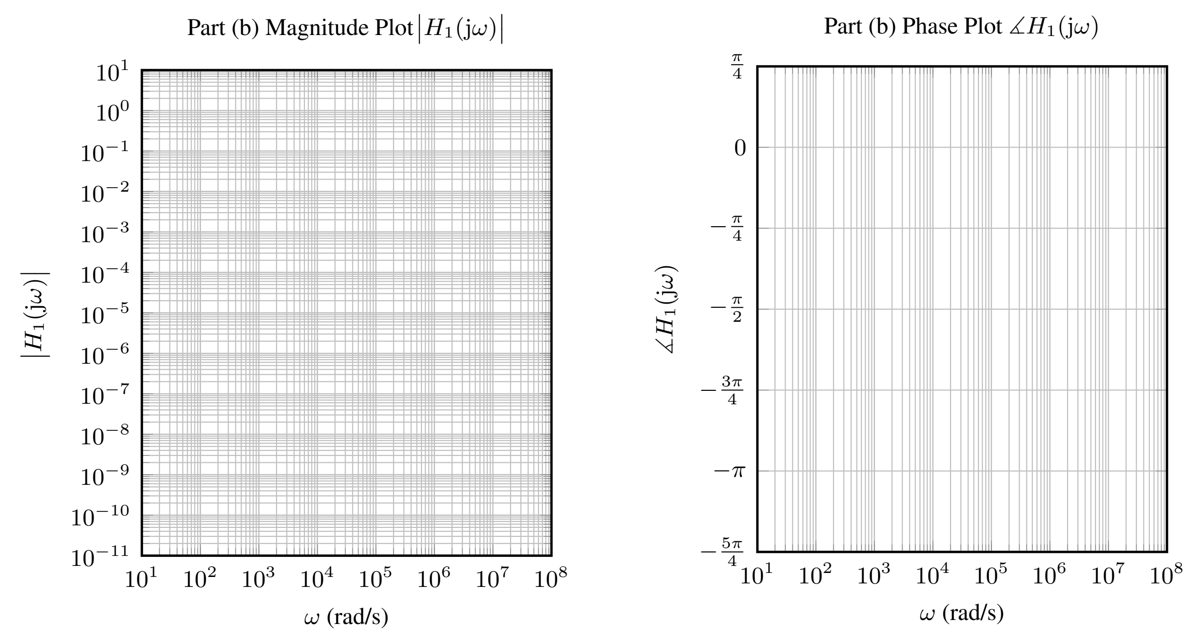


Figure 3: Part (b) Magnitude and Phase Bode Plots for the transfer function $H_1(\text{j}\omega)$.

iii. **By reading the corresponding values from the Bode plots, write down the approximate output signal expressions corresponding to $s_1(t) = 2\cos\left(10^2 t\right)$, $s_2(t) = 2\cos\left(10^7 t\right)$ and $n(t) = 10\cos\left(10^4 t\right)$ after the overall low-pass filtering using $H_1(\text{j}\omega)$.** *NOTE*: You do not need to compute exact magnitude and phase values using the transfer function. Just use the Bode approximation by reading from the plots.

PRINT your name and student ID: ____

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

PRINT your name and student ID: ______________________________

We have redrawn fig. 1 here for your convenience with the cutoff/corner frequencies labeled.

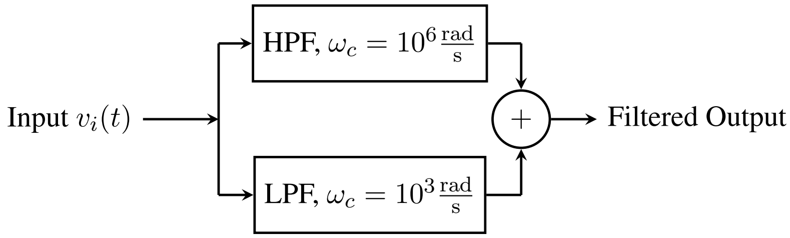


Figure 4: Block diagram of a Band Stop Filter.

(c) (3pts) **Explain in words what is the effect of the complete filter above on the three signals** $s_1(t) = 2\cos\left(10^2 t\right)$**,** $s_2(t) = 2\cos\left(10^7 t\right)$ **and** $n(t) = 10\cos\left(10^4 t\right)$**?** You do not have to provide numerical answers for this part.

PRINT your name and student ID: ____________________

(d) (5pts) Finally, let's consider the second approach. Since we know the exact frequency of the interference signal $n(t)$, we can use a Notch filter, as shown in fig. 5, to create a notch at that frequency, instead of using the previous band stop filter. This should completely attenuate $n(t) = 10\cos\left(10^4 t\right)$ and pass signals of all other frequencies, including $s_1(t)$ and $s_2(t)$. **If $C = 10\,\mu\mathrm{F}$ in the Notch filter in fig. 5, calculate the inductance L needed to completely block signals at $10^4\,\frac{\mathrm{rad}}{\mathrm{s}}$.** *NOTE*: You do not need to know the value of R to solve this question.

(*HINT: At $\omega = 10^4\,\frac{\mathrm{rad}}{\mathrm{s}}$, what do you want the impedance of the series connection of L and C to be?*)

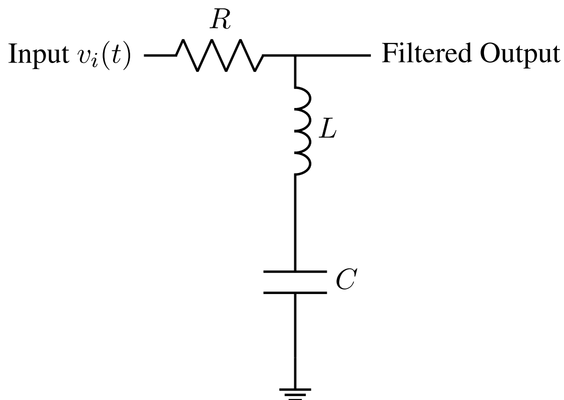


Figure 5: LC Notch filter

PRINT your name and student ID: ____

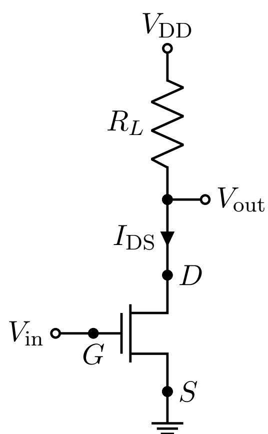
9. Using a Nonlinear NMOS Transistor for Amplification (35 points)

Consider the following schematic where $V_{\text{DD}} = 1.5\,\text{V}$, $R_L = 400\,\Omega$ and the NMOS transistor has threshold voltage $V_{\text{th}} = 0.2\,\text{V}$. We are interested in analyzing the response of this circuit to input voltages of the form $V_{\text{in}}(t) = V_{\text{in,DC}} + v_{\text{in,AC}}(t)$, where $V_{\text{in,DC}}$ is some constant voltage and $v_{\text{in,AC}}(t) = 0.001\cos(\omega t)\text{V}$ is a sinusoidal signal whose magnitude is much smaller than $V_{\text{in,DC}}$.

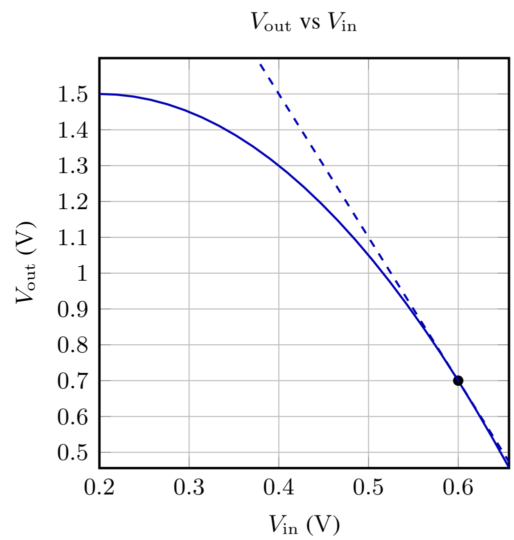
The I-V relationship of an NMOS can be modeled as non-linear functions over different regions of operation. For simplicity, let's just focus on the case when $0 \le V_{\text{GS}} - V_{\text{th}} < V_{\text{DS}}$. In this regime of interest, the relevant I-V relationship is given by

$$I_{\text{DS}}(V_{\text{GS}}) = \frac{K}{2}(V_{\text{GS}} - V_{\text{th}})^2 \tag{8}$$

where K is a constant that depends on the NMOS transistor size and properties.



(a) NMOS Transistor circuit



(b) V_{out} vs V_{in} in the regime of interest. Tangent is drawn at the operating point $V_{\text{in,DC}} = 0.6\,\text{V}$, $V_{\text{out,DC}} = 0.7\,\text{V}$

Figure 6: NMOS figures.

From Ohm's law and KCL, we know that

$$V_{\text{out}}(t) = V_{\text{DD}} - R_L I_{\text{DS}}(t). \tag{9}$$

Note from fig. 6a that $V_{\text{in}} = V_{\text{GS}}$ and $V_{\text{out}} = V_{\text{DS}}$. In fig. 6b, we can see the curve of V_{out} vs V_{in} in the transistor operating regime of interest.

(a) (4 pts) **Using eq. (8) and eq. (9), express $V_{\text{out}}(t)$ as a function of $V_{\text{in}}(t)$ symbolically.** (You can use $V_{\text{DD}}, R_L, V_{\text{in}}, K, V_{\text{th}}$ in your answer.)

PRINT your name and student ID: ____

The input $V_{\text{in}}(t) = V_{\text{in,DC}} + v_{\text{in,AC}}(t)$ results in an output of the form $V_{\text{out}}(t) = V_{\text{out,DC}} + v_{\text{out,AC}}(t)$. Since V_{DD} is constant, we can linearize $V_{\text{out}}(t)$ around $V_{\text{out,DC}}$ from eq. (9), as illustrated in fig. 6b.

$$A_v(V_{\text{in}}, V_{\text{out}})\bigg|_{V_{\text{in}}^\star = V_{\text{in,DC}}} = \frac{\mathrm{d}V_{\text{out}}}{\mathrm{d}V_{\text{in}}}\bigg|_{V_{\text{in}}^\star = V_{\text{in,DC}}} = -R_L \frac{\mathrm{d}I_{\text{DS}}}{\mathrm{d}V_{\text{in}}}\bigg|_{V_{\text{in}}^\star = V_{\text{in,DC}}} \tag{10}$$

$$= -R_L \frac{\mathrm{d}I_{\text{DS}}}{\mathrm{d}V_{\text{GS}}}\bigg|_{V_{\text{GS}}^\star = V_{\text{in,DC}}} \tag{11}$$

$$= -R_L g_m(V_{\text{GS}})\bigg|_{V_{\text{GS}}^\star = V_{\text{in,DC}}} \tag{12}$$

Here A_v is defined as the linearized voltage gain, which is illustrated by the slope of the tangent to the V_{out} vs V_{in} curve in fig. 6b at the point $\left(V_{\text{in,DC}},\ V_{\text{out,DC}}\right)$, and $g_m = \frac{\mathrm{d}I_{\text{DS}}}{\mathrm{d}V_{\text{GS}}}$ is the transistor transconductance linearized around the point $V_{\text{GS}}^\star = V_{\text{in,DC}}$.

(b) (10pts) To linearize the whole circuit around the operating point $V_{\text{in}}^\star = V_{\text{in,DC}}$, as shown in eq. (10), eq. (11), eq. (12), we need to linearize the non-linear transistor I-V curve, given by $I_{\text{DS}}(V_{\text{GS}}) = \frac{K}{2}(V_{\text{GS}} - V_{\text{th}})^2$ in eq. (8) to find the linearized transconductance gain $g_m = \frac{\mathrm{d}I_{\text{DS}}}{\mathrm{d}V_{\text{GS}}}$.

i. **Using eq. (8), derive the linearized transconductance gain $g_m = \frac{\mathrm{d}I_{\text{DS}}}{\mathrm{d}V_{\text{GS}}}$ symbolically.**
NOTE: Please simplify your answer.

ii. **From the following options, choose which circuit element can be used to represent the transistor in a linearized circuit with $\Delta I_{\text{DS}} = g_m \Delta V_{\text{GS}}$**, where ΔI_{DS} and ΔV_{GS} are small deviations around I_{DS} and V_{GS} respectively.

Select one	Choices
◯	resistor between G and S terminals
◯	resistor between D and S terminals
◯	voltage controlled current source between D and S terminals

iii. Using $K = 0.025\,\frac{1}{\Omega\,\text{V}}$, and $V_{\text{th}} = 0.2\,\text{V}$ and $R_L = 400\,\Omega$, **calculate the numerical values of the following quantities at the operating point $\left(V_{\text{GS}}^\star = 0.6\,\text{V},\ V_{\text{DS}}^\star = 0.7\,\text{V}\right)$:**

- **linearized transconductance gain g_m**
- **linearized voltage gain $A_v = -R_L g_m$ from eq. (12)**

NOTE: Please simplify the numerical answers — this will also help you check your answer graphically against fig. 6b if you want.

PRINT your name and student ID: ________________________________

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

PRINT your name and student ID: ______________________________

(c) (5pts) The circuit in fig. 7 below is a linearized model for the transistor circuit in fig. 6a, according to eq. (10), eq. (11), eq. (12). **Find the transfer function $H_1(\mathrm{j}\omega) = \frac{\widetilde{v}_{\mathrm{out,AC}}(\mathrm{j}\omega)}{\widetilde{v}_{\mathrm{in,AC}}(\mathrm{j}\omega)}$ in terms of g_m and R_L.**

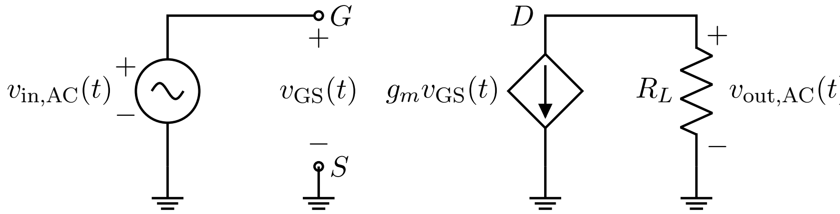


Figure 7: Small signal model for NMOS circuit in fig. 6a, according to eq. (10), eq. (11), eq. (12).

(d) (8pts) Consider the following modified model of the transistor circuit in fig. 6a where the transistor has a drain capacitance C_D as shown in fig. 8 below. **Find the transfer function $H_2(\mathrm{j}\omega) = \frac{\widetilde{v}_{\mathrm{out,AC}}(\mathrm{j}\omega)}{\widetilde{v}_{\mathrm{in,AC}}(\mathrm{j}\omega)}$ in terms of g_m, R_L, C_D and $\mathrm{j}\omega$. What type of filter is implemented by this circuit model?**

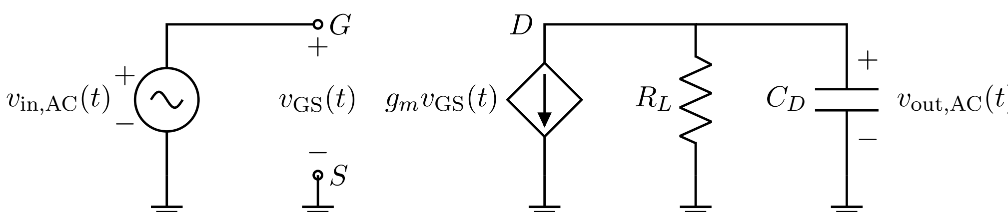


Figure 8: Small signal model for NMOS circuit in fig. 6a with drain capacitance.

PRINT your name and student ID: ________________________________

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

PRINT your name and student ID: ______________________________

(e) (8pts) Consider the following modified model of the transistor circuit in fig. 6a where V_{in} is not ideal and has some resistance in series, R_{in}, and the transistor has a gate capacitance C_{GS} as shown in fig. 9 below. **Find the transfer function** $H_3(\mathrm{j}\omega) = \frac{\widetilde{v}_{\text{out,AC}}(\mathrm{j}\omega)}{\widetilde{v}_{\text{in,AC}}(\mathrm{j}\omega)}$ **in terms of** $R_{\text{in}}, C_{\text{GS}}, g_m, R_L, C_D$ **and** $\mathrm{j}\omega$**.** (*HINT: First analyze* $v_{\text{GS}}(t)$ *in phasor domain. Then try to re-use the result from the previous part.*)

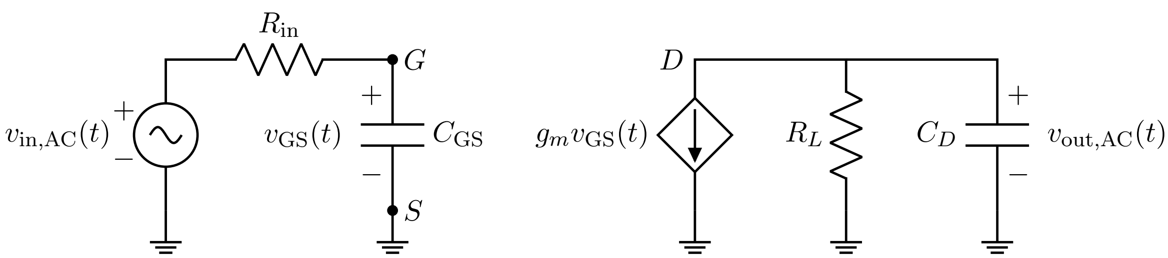


Figure 9: Small signal model for NMOS circuit in fig. 6a with non-ideal source and gate capacitance, in addition to drain capacitance.

PRINT your name and student ID: ____________________

10. Minimum Norm Solutions for Circuits involving Resistors (32 points)

Consider a current i_s flowing into a network of two parallel resistors R_1 and R_2, as shown in fig. 10 below.

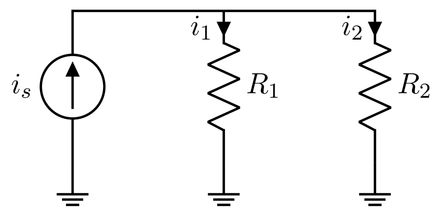


Figure 10: Current i_s dividing into i_1 and i_2.

From EECS 16A, we know that we can equate the voltage drops across the parallel resistors to derive $i_1 = \frac{R_2}{R_1+R_2} i_s$ and $i_2 = \frac{R_1}{R_1+R_2} i_s$. In this problem, we will try to derive the same current division result using the concept of minimum norm instead of voltage analysis.

It turns out that the current i_s will divide into two parts i_1 and i_2 in such a way that minimizes the total power dissipation $P = i_1^2 R_1 + i_2^2 R_2$ in the resistors.

(a) (8pts) **Argue that the current division result given by $i_1 = \frac{R_2}{R_1+R_2} i_s$ and $i_2 = \frac{R_1}{R_1+R_2} i_s$ minimizes the total power dissipation $P = i_1^2 R_1 + i_2^2 R_2$ using calculus.** Use the fact that KCL gives $i_2 = i_s - i_1$ to express P as a function of i_1 only. (*HINT: Once you solve for the optimal i_1, you don't have to do calculus again for i_2. Just use KCL.*)

PRINT your name and student ID: ____________________

(b) (12pts) Instead of using calculus to minimize the total power dissipation P, we can represent the current division problem as a minimum norm problem. Consider the vector $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ where $x_1 = i_1\sqrt{R_1}$ and $x_2 = i_2\sqrt{R_2}$. Notice that $P = i_1^2 R_1 + i_2^2 R_2 = x_1^2 + x_2^2 = \|\vec{x}\|^2$.

i. **Find the row vector A so that the KCL constraint $i_1 + i_2 = i_s$ can be written as $A\vec{x} = i_s$.**

ii. Using the A matrix you found above, **what is the minimum norm solution to $A\vec{x} = i_s$?** Show your work.
To help you save computation, the compact SVD of a general 1×2 row vector is given by

$$\begin{bmatrix} a & b \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{a^2+b^2} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} & \frac{b}{\sqrt{a^2+b^2}} \end{bmatrix}}_{V^\top} \tag{13}$$

iii. **Transform the minimum norm solution of $A\vec{x} = i_s$ to the original variables i_1 and i_2, and confirm that the result is $i_1 = \frac{R_2}{R_1+R_2} i_s$ and $i_2 = \frac{R_1}{R_1+R_2} i_s$ as the current-divider formula predicts.** Show your work.

PRINT your name and student ID: ____

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

PRINT your name and student ID: ________________________________

(c) (12 pts) We can solve any arbitrarily complicated circuit network using KCL and norm minimization, following the same technique that we used for the simple network in fig. 10. Consider a resistor network which has n resistor branches, with currents $i_1, i_2, \ldots, i_n$ across the branch resistances R_1, $R_2, \ldots, R_n$ respectively, and m total nodes each with current sources $i_{s_1}, i_{s_2}, \ldots, i_{s_m}$, which may be positive, negative or zero, as shown in fig. 11. Then the m KCL equations at the m nodes can be written as $K\vec{i} = \vec{i}_s$, where $K \in \mathbb{R}^{m\times n}$, $\vec{i} = \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_n \end{bmatrix} \in \mathbb{R}^n$, and $\vec{i}_s = \begin{bmatrix} i_{s_1} \\ i_{s_2} \\ \vdots \\ i_{s_m} \end{bmatrix} \in \mathbb{R}^m$. This KCL constraint $K\vec{i} = \vec{i}_s$ completely captures what is visualized in fig. 11, so you don't have to write any additional KCL. Note that fig. 10 is a simple example of fig. 11 with $n = 2$ and $m = 1$.

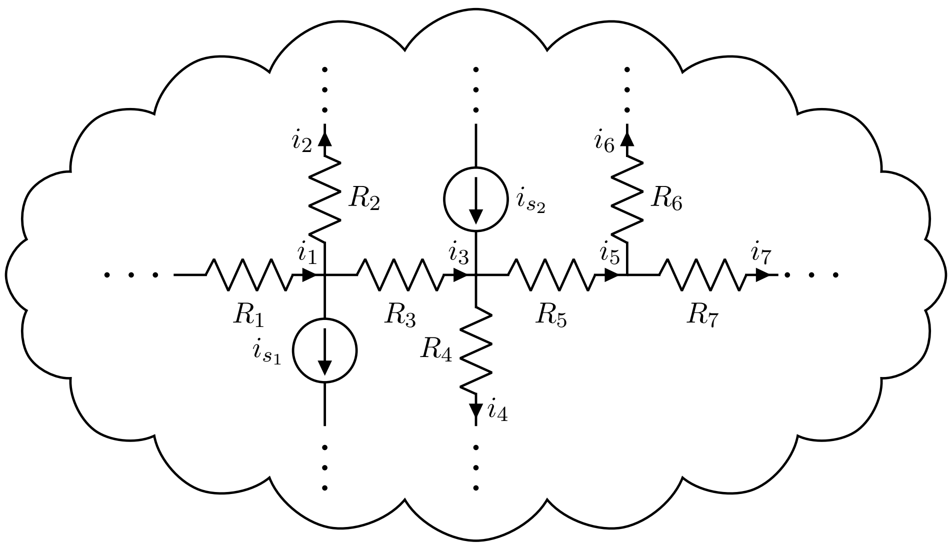


Figure 11: A section of an arbitrarily complicated network with n branches and m nodes.

i. We can change variables to $\vec{x} = D\vec{i}$ to represent the KCL constraint $K\vec{i} = \vec{i}_s$ as $A\vec{x} = \vec{i}_s$, and so the minimization of dissipated power $P = \sum_{j=1}^{n} i_j^2 R_j$ is just the minimization of $\sum_{j=1}^{n} x_j^2 = \|\vec{x}\|^2$.
Find the diagonal matrix D, and then find the matrix A in terms of D and K.
(*HINT: Look at how $\vec{x}$ was defined in the previous part.*)

PRINT your name and student ID: ____________________

ii. Assume the compact SVD of A is given by $A = U\Sigma V^\top$. **Use the minimum norm solution to $A\vec{x} = \vec{i}_s$ to solve for $\vec{i}$.** Recall from the previous part that $\vec{x} = D\vec{i}$. Your final answer for $\vec{i}$ can only use U, Σ, V, D, $\vec{i}_s$ as well as standard matrix operations like inverses, etc.

PRINT your name and student ID: ____

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

PRINT your name and student ID: ______________________________

11. A Proof in the Complex Case (12 points)

Suppose M is a generic $m \times n$ complex matrix with rank r and SVD $M = U\Sigma V^* = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^*$, where the matrices $U = [\vec{u}_1 \cdots \vec{u}_m]$ and $V = [\vec{v}_1 \cdots \vec{v}_n]$ have orthonormal columns according to the complex inner product (i.e. the U and V matrices are unitary) and Σ is diagonal with real non-negative diagonal entries sorted in non-ascending order.

Recall that we defined the Frobenius norm as

$$\|M\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \left|M_{ij}\right|^2} \tag{14}$$

and note that the same definition works for complex matrices.

Suppose $k < r$. **Prove that the best rank-at-most-k approximation to M in the Frobenius norm is given by $\sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^*$.**

In other words, prove that no matter what the collection of vectors $\{\vec{p}_i\}$ and $\{\vec{q}_i\}$ may be,

$$\left\| M - \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^* \right\|_F \leq \left\| M - \sum_{i=1}^{k} \vec{p}_i \vec{q}_i^* \right\|_F . \tag{15}$$

You may use the following facts without proof:

- If U and V are square unitary matrices, then $\|UA\|_F = \|A\|_F = \|AV\|_F$.
- The best rank-at-most-k approximation to a diagonal $m \times n$ real matrix Σ with the diagonal consisting of real non-negative values σ_i sorted in non-ascending order is given by $\sum_{i=1}^{k} \sigma_i \vec{e}_{m,i} \vec{e}_{n,i}^*$, where $\vec{e}_{k,i}$ is the i^{th} column of a $k \times k$ identity matrix.

  In math: No matter what the collection of vectors $\{\vec{s}_i\}$ and $\{\vec{w}_i\}$ may be,

$$\left\| \Sigma - \sum_{i=1}^{k} \sigma_i \vec{e}_{m,i} \vec{e}_{n,i}^* \right\|_F \leq \left\| \Sigma - \sum_{i=1}^{k} \vec{s}_i \vec{w}_i^* \right\|_F . \tag{16}$$

- If U and V are square unitary matrices, then the rank of a matrix A is the same as the rank of UA and the rank of AV.
- The inverse of a square unitary matrix U is given by its conjugate transpose $U^{-1} = U^*$, which is also unitary.

PRINT your name and student ID: ____________________

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

PRINT your name and student ID: ____

12. System ID for Continuous Systems (16 points)

So far we have seen system identification only done for discrete-time systems. But what if we really want to identify some underlying continuous-time model instead? We will explore how to do so in this problem.

(a) (8pts) Suppose we believed that our system was of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \lambda x(t) + bu(t) \tag{17}$$

where $x(t) \in \mathbb{R}$ and $u(t) \in \mathbb{R}$ is a scalar input.

Given an initial condition $x(t_0)$, and that $u(t)$ is some constant input $\bar{u}$ over the interval $[t_0, t_f)$, then for all $t \in [t_0, t_f)$, we know that this differential equation eq. (17) has the unique solution

$$x(t) = x(t_0)e^{\lambda(t-t_0)} + \frac{e^{\lambda(t-t_0)} - 1}{\lambda} b\bar{u}. \tag{18}$$

Assume that we record the state at known times $\tau_0, \tau_1, \ldots, \tau_n$ as having corresponding state values $x_0 = x(\tau_0), x_1 = x(\tau_1), \ldots, x_n = x(\tau_n)$. The continuous-time input is known to be piecewise constant $u(t) = u_i$ for $t \in [\tau_i, \tau_{i+1})$, where we know the sequence of inputs $u_0, u_1, \ldots, u_{n-1}$.

We now want to formulate this as a system ID question by relating the unknown parameters λ, b to the data we have. However, the relationship between the parameters and the data we collected is now non-linear. **For the data point x_{i+1}, use eq. (18) to write out how x_{i+1} should be related to λ and b in the form**

$$x_{i+1} = f(\lambda, x_i, \tau_i, \tau_{i+1}) + bg(\lambda, x_i, u_i, \tau_i, \tau_{i+1}). \tag{19}$$

What are the functions f and g?

PRINT your name and student ID: ________________________________

(b) (8pts) The previous part gave rise to a sequence of n equations of the form eq. (19). Because of observation noise and imperfection in our model, we are going to assume that these equations hold only approximately and hope to find values for the two parameters λ, b that minimize the cost function:

$$c(\lambda, b) = \sum_{i=0}^{n-1} \ell\left(x_{i+1}, f(\lambda, x_i, \tau_i, \tau_{i+1}) + bg(\lambda, x_i, u_i, \tau_i, \tau_{i+1})\right) \tag{20}$$

where $f(\lambda, x, \sigma, \tau)$ and $g(\lambda, x, u, \sigma, \tau)$ are given nonlinear scalar functions, and $\ell(s, p)$ is a loss function that penalizes how far the prediction p is from the measured state s. For example, you could use squared loss $\ell(s, p) = (s - p)^2$.

To find the best possible λ_*, b_*, you observe that you want to solve the nonlinear system of equations:

$$\frac{\partial}{\partial \lambda} c(\lambda, b)\Big|_{\lambda_*, b_*} = 0 \tag{21}$$

$$\frac{\partial}{\partial b} c(\lambda, b)\Big|_{\lambda_*, b_*} = 0 \tag{22}$$

and decide to do so using Newton's method starting with an initial guess λ_0, b_0 and linearizing the system of equations eq. (21) and eq. (22) to get a system of linear equations to solve at each step.

The system of linear equations at each iteration $j + 1$ can be expressed in vector form as:

$$A \begin{bmatrix} \lambda - \lambda_j \\ b - b_j \end{bmatrix} = \vec{y}. \tag{23}$$

What are the entries of the matrix A and the vector $\vec{y}$ in terms of the appropriate partial derivatives of $c(\lambda, b)$ evaluated at the appropriate arguments?

Assume that you can use PyTorch to compute whatever derivatives of $c(\lambda, b)$ that you want — all given functions are sufficiently differentiable. You don't have to take the derivatives by hand. You just need to tell us what derivatives and what arguments to evaluate them at.

PRINT your name and student ID: ______________________________

13. Movie Ratings with Missing Entries (23 points)

In a matrix R, you have users' movie ratings. However, not all users watched all the movies.

$$R = \left[\begin{array}{cccc|cc} 0.50 & 0.00 & 0.50 & 0.50 & 0.20 & 1.0 \\ 0.60 & 0.20 & 0.40 & 0.50 & ? & ? \\ 0.50 & 0.50 & 0.00 & 0.25 & 0.60 & 1.0 \\ 0.60 & 0.10 & 0.50 & 0.55 & ? & ? \\ \hline 1.00 & 0.40 & 0.60 & ? & ? & ? \end{array}\right] \tag{24}$$

where the element at the ith row and jth column indicates the rating of movie i by user j. A "?" means that there's no rating for that movie.

Our goal is to predict ratings for the missing entries, so we can recommend movies to users. In order to do this, you want to find the hidden goodness vectors for the movies, and the hidden sensitivity vectors of the users. However, due to missing entries, it is not possible to run an SVD on the entire rating matrix R.

It turns out that we have a submatrix R' in R that does not have any missing entries.

$$R' = \begin{bmatrix} 0.50 & 0.00 & 0.50 & 0.50 \\ 0.60 & 0.20 & 0.40 & 0.50 \\ 0.50 & 0.50 & 0.00 & 0.25 \\ 0.60 & 0.10 & 0.50 & 0.55 \end{bmatrix} \tag{25}$$

We provide a decomposition of this matrix:

$$R' = \underbrace{\begin{bmatrix} 0.5 & 0.0 \\ 0.4 & 0.2 \\ 0.0 & 0.5 \\ 0.5 & 0.1 \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 2.0 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 0.25 & 0.0 & 0.25 & 0.25 \\ 0.5 & 0.5 & 0.0 & 0.25 \end{bmatrix}}_{S} \tag{26}$$

where the ith row of the matrix G represents the goodness row vector $\vec{g}_i^\top$ of the movie i, the jth column of the matrix S represents the sensitivity vector $\vec{s}_j$ of user j, and each diagonal entry of the matrix D shows the weight each attribute has in determining the rating of a movie by a user.

NOTE: This decomposition in eq. (26) is not an SVD; G and S do not have orthonormal vectors.

(a) (4pts) Suppose $\vec{s}_6$ (the sensitivity vector of user 6) is:

$$\vec{s}_6 = \begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix} \tag{27}$$

Use this to **estimate the rating of movie 2 as rated by user 6.** (Show work that uses $\vec{s}_6$. Unjustified answers will get no credit.)

PRINT your name and student ID: ______________________________

(b) (6pts) For the 5th movie, we have three ratings and want to find two parameters of goodness. **Formulate a least squares problem $A\vec{g}_5 \approx \vec{b}$ to estimate $\vec{g}_5$ (goodness vector of movie 5).** You need to tell us A explicitly as a matrix with numerical entries. We give you that $\vec{b} = \begin{bmatrix} 1.00 \\ 0.40 \\ 0.60 \end{bmatrix}$ since those are the three ratings we know for this movie.

(c) (3pts) We now consider a ratings matrix R without missing entries (that is different from the previous R) where the matrix is partitioned into four blocks $R_{11}, R_{12}, R_{21}, R_{22}$ as below.

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \tag{28}$$

In order to find the optimal number of principal components, we compute a PCA model from the SVD of R_{11} with k principal components, with $k = 2, 3, 4, 5$. We then use the chosen components and the singular values of R_{11} together with the information in R_{12} and R_{21} to create an estimate $\widehat{R}_{22}$ for the held-out ratings in R_{22}. We can also use the first k terms of the SVD of R_{11} to reconstruct $\widehat{R}_{11}$ as the best rank-k approximation to R_{11}.
The training errors $\|R_{11} - \widehat{R}_{11}\|_F^2$ and validation errors $\|R_{22} - \widehat{R}_{22}\|_F^2$ for each candidate choice for k are given in the table below.

Select	k	Training error	Validation error
◯	1	1.428	3.104
◯	2	0.414	2.494
◯	3	0.093	0.462
◯	4	0.017	0.090
◯	5	0.011	0.132
◯	6	0.006	0.161

Find the optimal number of principal components k we should use and fill in the appropriate bubble. (No need to give any justification.)

PRINT your name and student ID: ____

(d) (10pts) Suppose that we want to approximate R with a rank-k matrix $\widehat{R} = CXL^\top$ where C is known (e.g. it has a specific k columns selected from R), and so is $L^\top$ (e.g. it has a specific k rows selected from R). The only freedom is in choosing the k by k matrix X.

We want to find X that minimizes the Frobenius norm error between R and $\widehat{R}$:

$$\operatorname*{argmin}_X \|R - \underbrace{CXL^\top}_{\widehat{R}}\|_F^2 \tag{29}$$

This is a least-squares problem since $CXL^\top$ is linear in the entries of the matrix X and minimizing the Frobenius norm squared is just minimizing a sum of squared errors. Suppose we further know that C has linearly independent columns and that $L^\top$ has linearly independent rows. It turns out that the optimal $X = \left((C^\top C)^{-1}C^\top\right) R \left(L(L^\top L)^{-1}\right)$.

Suppose that we know the full SVDs of C and $L^\top$:

$$C = \begin{bmatrix} U_C & U_{C,n} \end{bmatrix} \begin{bmatrix} \Sigma_C \\ 0 \end{bmatrix} V_C^\top, \qquad L^\top = U_L \begin{bmatrix} \Sigma_L & 0 \end{bmatrix} \begin{bmatrix} V_L^\top \\ V_{L,n}^\top \end{bmatrix}. \tag{30}$$

Using these SVDs and remembering how they simplify projections, we notice that:

$$\widehat{R} = CXL^\top = \left(C(C^\top C)^{-1}C^\top\right)R\left(L(L^\top L)^{-1}L^\top\right) = U_C U_C^\top R V_L V_L^\top. \tag{31}$$

This suggests that the orthonormal bases $U = [U_C \quad U_{C,n}]$ and $V = [V_L \quad V_{L,n}]$ are interesting to consider, so we notice that

$$\begin{align}
R &= UU^\top RVV^\top \tag{32}\\
&= \begin{bmatrix} U_C & U_{C,n} \end{bmatrix} \begin{bmatrix} U_C^\top \\ U_{C,n}^\top \end{bmatrix} R \begin{bmatrix} V_L & V_{L,n} \end{bmatrix} \begin{bmatrix} V_L^\top \\ V_{L,n}^\top \end{bmatrix} \tag{33}\\
&= (U_C U_C^\top + U_{C,n} U_{C,n}^\top) R (V_L V_L^\top + V_{L,n} V_{L,n}^\top) \tag{34}\\
&= \underbrace{U_C U_C^\top R V_L V_L^\top}_{\widehat{R}} + U_C U_C^\top R V_{L,n} V_{L,n}^\top + U_{C,n} U_{C,n}^\top R V_L V_L^\top + U_{C,n} U_{C,n}^\top R V_{L,n} V_{L,n}^\top. \tag{35}
\end{align}$$

Use eq. (35) together with eq. (31) to establish that this X satisfies the key condition of least-squares optimality: **show that the residual $R - \widehat{R}$ is orthogonal to the estimate $\widehat{R}$** when you use the inner product corresponding to the Frobenius norm — which basically treats a matrix as a big vector. $\langle A, B\rangle_F = \text{trace}(A^\top B) = \text{trace}(B^\top A) = \text{trace}(AB^\top) = \text{trace}(BA^\top)$.

(HINT 1: Given that the Frobenius inner-product between two matrices of the same size can be interpreted as either the sum of the inner-products of all the rows in A with their B counterparts (trace($AB^\top$)) or all the columns in A with their B counterparts (trace($A^\top B$)), why do you think that the first term in eq. (35) must be orthogonal to each of the final three terms in eq. (35)?)

(HINT 2: What subspace do the rows of the second term in eq. (35) live in? What subspace do the columns of the third term in eq. (35) live in? .... What subspaces do the rows and columns of $\widehat{R}$ live in according to eq. (31)?)

PRINT your name and student ID: ____________________

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

PRINT your name and student ID: ____

[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.]

PRINT your name and student ID: ____

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

