EECS 16B Fall 2019 — Designing Information Devices and Systems II, UC Berkeley

Midterm 2

1. Stability (14pts)

Consider the complex plane below, which is broken into non-overlapping regions A through H. The circle drawn on the figure is the unit circle $|\lambda| = 1$.

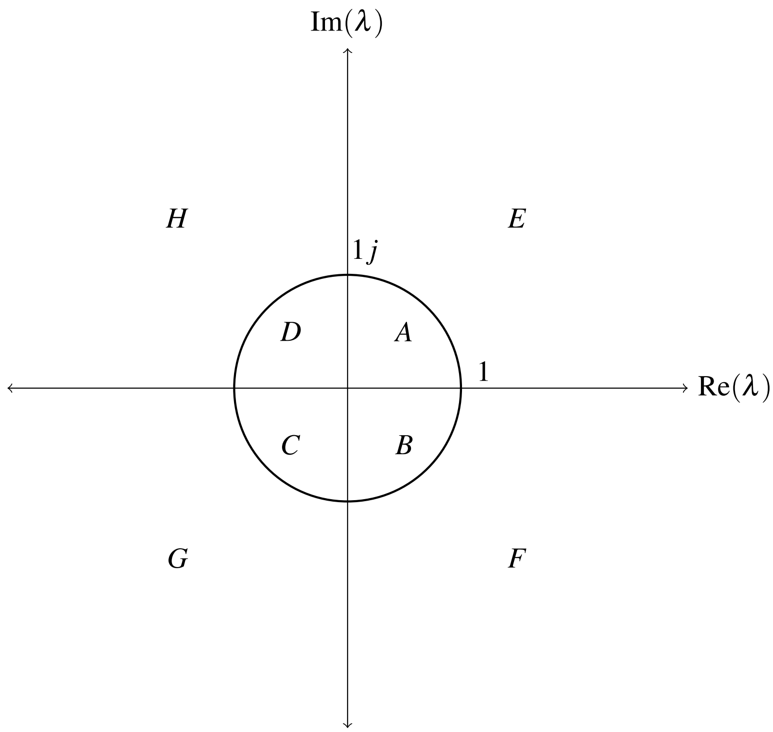

Figure 1: Complex plane divided into regions.

(a) (4pts) Consider the continuous-time system $\frac{d}{dt}x(t) = \lambda x(t) + v(t)$ and the discrete-time system $y(t+1) = \lambda y(t) + w(t)$.
In which regions can the eigenvalue λ be for a *stable* system? Fill out the table below to indicate *stable* regions. Assume that the eigenvalue λ does not fall directly on the boundary between two regions.

	A	B	C	D	E	F	G	H
Continuous Time System $x(t)$	○	○	○	○	○	○	○	○
Discrete Time System $y(t)$	○	○	○	○	○	○	○	○

Solution: For the continuous time system to be stable, we need the real part of λ to be less than zero. Hence, C,D,G,H satisfy this condition.
On the other hand, for the discrete time system to be stable, we need the norm of λ to be less than one. Hence, A,B,C,D satisfy this condition.

(b) (10pts) Consider the continuous time system

$$\frac{d}{dt}x(t) = \lambda x(t) + u(t)$$

where λ is real and $\lambda < 0$.

Assume that $x(0) = 0$ and that $|u(t)| < \varepsilon$ for all $t \geq 0$.

Prove that the solution $x(t)$ will be bounded (i.e. $\exists k$ so that $|x(t)| \leq k\varepsilon$ for all time $t \geq 0$.).

(Hint: Recall that the solution to such a first-order scalar differential equation is:

$$x(t) = x_0 e^{\lambda t} + \int_0^t u(\tau) e^{\lambda(t-\tau)} d\tau$$

You may use this fact without proof.)

Solution: Start by taking the abolute value of both sides.

$$\begin{aligned}
|x(t)| &= \left| x_0 e^{\lambda t} + \int_0^t u(\tau) e^{\lambda(t-\tau)} d\tau \right| && (1) \\
&= \left| \int_0^t u(\tau) e^{\lambda(t-\tau)} d\tau \right| && (2) \\
&\leq \int_0^t \left| u(\tau) e^{\lambda(t-\tau)} \right| d\tau && (3) \\
&= \int_0^t |u(\tau)| \left| e^{\lambda(t-\tau)} \right| d\tau && (4) \\
&< \int_0^t \varepsilon \left| e^{\lambda(t-\tau)} \right| d\tau && (5) \\
&= \varepsilon \int_0^t \left| e^{\lambda(t-\tau)} \right| d\tau && (6) \\
&= \varepsilon \int_0^t e^{\lambda(t-\tau)} d\tau && (7) \\
&= \varepsilon \cdot \frac{e^{\lambda t} - 1}{\lambda} && (8) \\
&= \varepsilon \cdot \frac{1 - e^{\lambda t} - 1}{-\lambda} && (9) \\
&\leq \varepsilon \cdot \frac{1}{-\lambda}. && (10)
\end{aligned}$$

We have written things out in gory detail, you didn't need to call out each of these steps.

Hence the solution $x(t)$ will be bounded. In the above, we used the following integration:

$$\int_0^t e^{\lambda(t-\tau)} d\tau = e^{\lambda t} \cdot \int_0^t e^{-\lambda \tau} d\tau = e^{\lambda t} \cdot \left(-\frac{1}{\lambda} e^{-\lambda \tau} \right)_0^t = e^{\lambda t} \cdot \left(\frac{1}{\lambda} - \frac{1}{\lambda} e^{-\lambda t} \right) = \frac{e^{\lambda t} - 1}{\lambda}.$$

2. Computing the SVD (10pts)

Consider the matrix

$$A = \begin{bmatrix} 4 & -3 & 0 \\ 3 & 4 & 0 \end{bmatrix}.$$

Write out a singular value decomposition of the matrix A in the form $U\Sigma V^T$ where U is a 2×2 orthonormal matrix, Σ is a diagonal rectangular matrix, and V is a 3×3 orthonormal matrix.

Solution: We can calculate the singular values of A by finding the eigenvalues of

$$AA^\top = \begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix}$$

and we see that the eigenvalues of $AA^\top$ are both 25, and hence

$$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix}.$$

Then, we can find U by finding the set of normalized eigenvectors of $AA^\top$. Here, we see that

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Finally, to find V, we can calculate

$$U\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix}.$$

Recall that $A = U\Sigma V^\top$. We observe $V^\top$ is

$$V^\top = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} & 0 \\ \frac{3}{5} & \frac{4}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which gives us

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

This completes one possible way of doing a singular value decomposition of A:

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

Another way to do this problem is to start from

$$A^\top A = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

with corresponding eigenspace

$$V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

To find U, we can simply calculate

$$U = AV\Sigma^{-1} = \begin{bmatrix} 4 & -3 & 0 \\ 3 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{5} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}.$$

This completes another possible way of doing a singular value decomposition of A:

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

There are essentially infinite possible ways of getting a singular value decomposition, and as we saw in homework, the SVD is not unique. However, your solution of the SVD must satisfy the requirements that $U^\top U = I$, $V^\top V = I$, $\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix}$, and $A = U\Sigma V^\top$.

3. Outlier Removal (14pts)

Suppose we have a system where we believe that a 2-dimensional vector input $\vec{x}$ leads to scalar outputs in a linear way $\vec{p}^T\vec{x}$. However, the parameters $\vec{p}$ are unknown and must be learned from data. Our data collection process is imperfect and instead of directly seeing $\vec{p}^T\vec{x}$, we get observations $y = \vec{p}^T\vec{x} + w$ where the w is some kind of disturbance or noise that nature introduces.

To allow us to learn the parameters, we have 4 experimentally obtained data points: input-output pairs $(\vec{x}_i, y_i)$ where the index $i = 1, \ldots, 4$.

$$X = \begin{bmatrix} 1 & 2 \\ 3 & 2 \\ 5 & 1 \\ 4 & 0 \end{bmatrix} \text{ and } \vec{y} = \begin{bmatrix} 5 \\ 7 \\ 5000 \\ 300 \end{bmatrix}$$

Then we can express the approximate system of equations that we want to solve as $X\vec{p} \approx \vec{y}$.

(a) (6pts) Suppose we know that the third data point $\left(\begin{bmatrix} 5 \\ 1 \end{bmatrix}, 5000\right)$ may have been corrupted, and we wish to effectively remove it from the data set. We decide to do so by augmenting our approximate system of equations to be

$$[X, \vec{a}] \begin{bmatrix} \vec{p} \\ f \end{bmatrix} \approx \vec{y}. \tag{11}$$

Mark all of the following choices for $\vec{a}$ which have the effect of eliminating the third data point from the data set if we run least-squares on (11) **to estimate $\vec{p}$.** Fill in the circles corresponding to your selections.

○ $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, ○ $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, ○ $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, ○ $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, ○ $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$, ○ $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$,

○ $\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, ○ $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$, ○ $\begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, ○ $\begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}$, ○ $\begin{bmatrix} 0 \\ 0 \\ 3 \\ 0 \end{bmatrix}$, ○ $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix}$

Solution: We need to choose a vector that spans the space of the third data point. Hence, the vectors to choose are

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ 3 \\ 0 \end{bmatrix}.$$

(b) (8pts) Now suppose that we know that both the third data point $\left(\begin{bmatrix}5\\1\end{bmatrix}, 5000\right)$ and the fourth data point $\left(\begin{bmatrix}4\\0\end{bmatrix}, 300\right)$ have been corrupted, and we wish to effectively remove both of them from the data set. We decide to do so by augmenting our approximate system of equations to be

$$[X, \vec{a}, \vec{b}] \begin{bmatrix}\vec{p}\\f_a\\f_b\end{bmatrix} \approx \vec{y}. \tag{12}$$

Mark all of the following choices for $\vec{a}, \vec{b}$ pairs which have the effect of eliminating the third and fourth data points from the data set if we run least-squares on (12) **to estimate $\vec{p}$.** Fill in the circles corresponding to your selections.

○ $\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$; ○ $\begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$; ○ $\begin{bmatrix}0\\0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$;

○ $\begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\end{bmatrix}$; ○ $\begin{bmatrix}3\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\4\\0\\0\end{bmatrix}$; ○ $\begin{bmatrix}0\\0\\3\\4\end{bmatrix}, \begin{bmatrix}0\\0\\3\\4\end{bmatrix}$

Solution: We need to choose a vector that spans the space of the third and fourth data points. Hence, the vector pairs to choose are

$$\begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}0\\0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix}.$$

4. Control (18 pts)

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

$$\vec{x}(t+1) = \begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{7}{6} \end{bmatrix} \vec{x}(t) + \vec{w}(t).$$

(a) (2pts) **Is the system stable? Why or why not?**

Here, we give you that

$$A = \begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{7}{6} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -\frac{4}{3} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

and that the characteristic polynomial $\det(\lambda I - A) = \lambda^2 + \frac{11}{6}\lambda + \frac{2}{3}$.

Solution: The eigenvalues are $\lambda_1 = -\frac{4}{3}$ and $\lambda_2 = -\frac{1}{2}$. The system is unstable because we have an eigenvector with absolute value greater than 1.

(b) (6pts) Suppose that there is no disturbance and we can now influence the system using a scalar input $u(t)$ to our system:

$$\vec{x}(t+1) = \begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{7}{6} \end{bmatrix} \vec{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t).$$

Is the system controllable?

Solution: Yes, and we can see this by writing out the controllability matrix

$$\mathscr{C} = \begin{bmatrix} \vec{b} & A\vec{b} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}.$$

The columns are linearly independent, so $\mathscr{C}$ is full rank and the system is controllable.

(c) (10pts) We want to set the closed-loop eigenvalues of the system

$$\vec{x}(t+1) = \begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{7}{6} \end{bmatrix} \vec{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t)$$

to be $\lambda_1 = -\frac{5}{6}, \lambda_2 = \frac{5}{6}$ using state feedback

$$u(t) = -\begin{bmatrix} k_1 & k_2 \end{bmatrix} \vec{x}(t).$$

What specific numeric values of k_1 and k_2 should we use?

Solution: The new system matrix is

$$\tilde{A} = \begin{bmatrix} -\frac{2}{3} - k_1 & \frac{1}{3} - k_2 \\ \frac{1}{3} & -\frac{7}{6} \end{bmatrix}$$

and we find the characteristic polynomial

$$\left(\lambda + \frac{7}{6}\right) \cdot \left(\lambda + \frac{2}{3} + k_1\right) - \frac{1}{9} + \frac{1}{3}k_2 = \lambda^2 + \left(\frac{11}{6} + k_1\right)\lambda + \frac{2}{3} + \frac{7}{6}k_1 + \frac{1}{3}k_2.$$

Since we want $\lambda_1 = -\frac{5}{6}, \lambda_2 = \frac{5}{6}$, we would want the characteristic polynomial to be of the form

$$\lambda^2 - \frac{25}{36} = 0.$$

Comparing the terms, we want

$$\frac{2}{3} + \frac{7}{6}k_1 + \frac{1}{3}k_2 = -\frac{25}{36}$$
$$\frac{11}{6} + k_1 = 0$$

and we solve

$$k_1 = -\frac{11}{6}, \quad k_2 = \frac{7}{3}.$$

5. Upper Triangularization (12 pts)

In this problem, you need to upper-triangularize the matrix

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 3 & -1 & 6 \\ -2 & 2 & -2 \end{bmatrix}$$

The eigenvalues of this matrix A are $\lambda_1 = \lambda_2 = 2$ and $\lambda_3 = -4$. We want to express A as

$$A = \begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & a & b \\ 0 & \lambda_2 & c \\ 0 & 0 & \lambda_3 \end{bmatrix} \cdot \begin{bmatrix} \vec{x}^\top \\ \vec{y}^\top \\ \vec{z}^\top \end{bmatrix}$$

where the $\vec{x}, \vec{y}, \vec{z}$ are orthonormal. Your goal in this problem is to compute $\vec{x}, \vec{y}, \vec{z}$ so that they satisfy the above relationship for some constants a, b, c.

Here are some potentially useful facts that we have gathered to save you some computations, you'll have to grind out the rest yourself.

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

We also know that

$$\left\{ \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is an orthonormal basis, and

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & -1 & 2 \\ 3 & -1 & 6 \\ -2 & 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -2\sqrt{2} \\ -2\sqrt{2} & -2 \end{bmatrix}.$$

We also know that $\begin{bmatrix} 0 & -2\sqrt{2} \\ -2\sqrt{2} & -2 \end{bmatrix}$ has eigenvalues 2 and -4. The normalized eigenvector corresponding to $\lambda = 2$ is $\begin{bmatrix} -\frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix}$ and $\begin{bmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} \end{bmatrix}$ is a vector that is orthogonal to that and also has norm 1.

Based on the above information, compute $\vec{x}, \vec{y}, \vec{z}$. Show your work.

You don't have to compute the constants a, b, c in the interests of time.

Solution: Let $V = [\vec{v}_0, R]$ where $R = [\vec{v}_1, \vec{v}_2]$. We can write

$$A = \begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & ? \\ \vec{0} & Q \end{bmatrix} \cdot \begin{bmatrix} \vec{x}^\top \\ \vec{y}^\top \\ \vec{z}^\top \end{bmatrix}$$

where $Q = R^\top AR$. Here, $\vec{x} = \vec{v}_0 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}$. Continuing above, we know that

$$Q = R^\top AR = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & -1 & 2 \\ 3 & -1 & 6 \\ -2 & 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -2\sqrt{2} \\ -2\sqrt{2} & -2 \end{bmatrix}.$$

Q has eigenvalues 2 and -4. The normalized eigenvector corresponding to $\lambda = 2$ is $\vec{u}_0 = \begin{bmatrix} -\frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix}$. Then, we choose $\vec{u}_1 = \begin{bmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} \end{bmatrix}$ so that $\vec{u}_0$ and $\vec{u}_1$ form an orthogonal basis.

Then,

$$\vec{y} = R\vec{u}_0 = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -\frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix}.$$

$$\vec{z} = R\vec{u}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{6} \\ -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{3} \end{bmatrix}.$$

Finally, we can find the values of a, b, c by

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{3} \end{bmatrix} \begin{bmatrix} 3 & -1 & 2 \\ 3 & -1 & 6 \\ -2 & 2 & -2 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{6} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} & -\frac{\sqrt{6}}{6} \\ 0 & \frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 4\sqrt{3} \\ 0 & 2 & 0 \\ 0 & 0 & -4 \end{bmatrix}.$$

Hence, we have

$$a = 0$$

$$b = 4\sqrt{3}$$

$$c = 0$$

which completes our upper triangularization.

6. Minimum Norm Variants (52 pts)

In lecture and HW, you saw how to solve minimum norm problems in which we have a wide matrix A and solve $A\vec{x} = \vec{y}$ such that $\vec{x}$ is a minimum norm solution: $||\vec{x}|| \le ||\vec{z}||$ for all $\vec{z}$ such that $A\vec{z} = \vec{y}$.

We also saw in the HW how we can solve some variants in which we were interested in minimizing the norm $||C\vec{x}||$ instead. You have solved the case where C is invertible and square or a tall matrix. This question asks you about the case when C is a wide matrix. The key issue is that wide matrices have nontrivial nullspaces — that means that there are "free" directions in which we can vary $\vec{x}$ while not having to pay anything. How do we best take advantage of these "free" directions?

Parts (a-b) are connected; parts (c-d) are another group that can be done independently of (a-b); and parts (e-g) are another group that can be started independently of either (a-b) or (c-d). If you can't do a part, move on. **During debugging, many TAs found it easier to start with parts (c-g), and coming back to (a-b) at the end.**

(a) (8pts) Given a wide matrix A (with m columns and n rows) and a wide matrix C (with m columns and r rows), we want to solve:

$$\min_{\vec{x} \text{ such that } A\vec{x}=\vec{y}} ||C\vec{x}|| \tag{13}$$

As mentioned above, the key new issue is to isolate the "free" directions in which we can vary $\vec{x}$ so that they might be properly exploited. Consider the full SVD of $C = U\Sigma_c V^\top = \sum_{i=1}^{\ell} \sigma_{c,i}\vec{u}_i\vec{v}_i^\top$. Here, we write $V = [V_c, V_f]$ so that $V_c = [\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_\ell]$ all correspond to singular values $\sigma_{c,i} > 0$ of C, and $V_f = [\vec{v}_{\ell+1}, \cdots, \vec{v}_m]$ form an orthonormal basis for the nullspace of C.

Change variables in the problem to be in terms of $\vec{\tilde{x}} = \begin{bmatrix} \vec{\tilde{x}}_c \\ \vec{\tilde{x}}_f \end{bmatrix}$ where the ℓ-dimensional $\vec{\tilde{x}}_c$ has components $\tilde{x}_c[i] = \alpha_i \vec{v}_i^\top \vec{x}$, and the $(m-\ell)$-dimensional $\vec{\tilde{x}}_f$ has components $\tilde{x}_f[i] = \vec{v}_{\ell+i}^\top \vec{x}$. In vector/matrix form, $\vec{\tilde{x}}_f = V_f^\top \vec{x}$ and $\vec{\tilde{x}}_c = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_\ell \end{bmatrix} V_c^\top \vec{x}$. Or directly, $\vec{\tilde{x}} = \begin{bmatrix} \vec{\tilde{x}}_c \\ \vec{\tilde{x}}_f \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_\ell \end{bmatrix} V_c^\top \\ V_f^\top \end{bmatrix} \vec{x}$.

Express $\vec{x}$ in terms of $\vec{\tilde{x}}_f$ and $\vec{\tilde{x}}_c$. Assume the $\alpha_i \neq 0$ so the relevant matrix is invertible.

What is $||C\vec{x}||$ in terms of $\vec{\tilde{x}}_f$ and $\vec{\tilde{x}}_c$? Simplify as much as you can for full credit.

(HINT: If you get stuck on how to express $\vec{x}$ in terms of the new variables, think about the special case when $\ell = 1$ and $\alpha_1 = \frac{1}{2}$. How is this different from when $\alpha_1 = 1$? The SVD of C might be useful when looking at $||C\vec{x}||$.)

Solution: Let us give a name to the matrix

$$B = \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \alpha_l \end{bmatrix}$$

Note that we can write

$$\vec{\tilde{x}}_c = BV_c^\top \vec{x}$$

and

$$\vec{\tilde{x}}_f = V_f^\top \vec{x}.$$

Since V_c and V_f are orthogonal to each other and form a basis for the space, it follows that we can write $\vec{x} = V_c\vec{w}_1 + V_f\vec{w}_2$. Then, the above two equations directly imply (since $V_c^\top V_f = 0$ and $V_f^\top V_c = 0$, and $V_c^\top V_C = I$ along with $V_f^\top V_f = I$) that

$$\vec{w}_1 = B^{-1}\vec{\tilde{x}}_c \qquad \vec{w}_2 = \vec{\tilde{x}}_f.$$

This implies that we have

$$\vec{x} = \begin{bmatrix} V_cB^{-1} & V_f \end{bmatrix} \begin{bmatrix} \vec{\tilde{x}}_c \\ \vec{\tilde{x}}_f \end{bmatrix} = V_cB^{-1}\vec{\tilde{x}}_c + V_f\vec{\tilde{x}}_f$$

The above is written in more detail than we expected, but is done so that those who didn't understand can better understand.

Then, we see that

$$\begin{aligned}
\|C\vec{x}\|^2 &= \vec{x}^\top C^\top C\vec{x} \\
&= \vec{x}^\top V\Sigma_c^\top U^\top U\Sigma_c V^\top \vec{x} \\
&= (V_cB^{-1}\vec{\tilde{x}}_c + V_f\vec{\tilde{x}}_f)^\top V\Sigma_c^\top \Sigma_c V^\top (V_cB^{-1}\vec{\tilde{x}}_c + V_f\vec{\tilde{x}}_f) \\
&= \vec{\tilde{x}}_c^\top B^{-1} \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\ell^2 \end{bmatrix} B^{-1}\vec{\tilde{x}}_c + \vec{\tilde{x}}_f^\top \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \vec{\tilde{x}}_f \\
&= \vec{\tilde{x}}_c^\top \begin{bmatrix} \frac{\sigma_1^2}{\alpha_1^2} & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2^2}{\alpha_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_\ell^2}{\alpha_\ell^2} \end{bmatrix} \vec{\tilde{x}}_c \\
&= \sum_{i=1}^{\ell} (\vec{\tilde{x}}_c[i])^2 \cdot \left(\frac{\sigma_i^2}{\alpha_i^2}\right).
\end{aligned} \tag{14}$$

Above, we relied crucially on the fact that $V_c^\top V_f$ would give 0s while $V_c^\top V_c$ gives an identity and so does $V_f^\top V_f$. This is what allows us to effectively break the huge $\Sigma_c^\top \Sigma_c$ matrix into two relevant pieces — one containing the diagonal square matrix with the squared nonzero singular values and the other filled with zeros.

Hence, the norm $\|C\vec{x}\|$ is

$$\|C\vec{x}\| = \sqrt{\sum_{i=1}^{\ell} (\vec{\tilde{x}}_c[i])^2 \cdot \left(\frac{\sigma_i^2}{\alpha_i^2}\right)}.$$

(b) (12pts) Continuing the previous part, **give appropriate values for the α_i so that the problem** (13) **becomes**

$$\min_{\vec{\tilde{x}}=\begin{bmatrix}\vec{\tilde{x}}_c\\ \vec{\tilde{x}}_f\end{bmatrix} \text{ such that } [A_c,A_f]\begin{bmatrix}\vec{\tilde{x}}_c\\ \vec{\tilde{x}}_f\end{bmatrix}=\vec{y}} \|\vec{\tilde{x}}_c\| \tag{15}$$

Give explicit expressions for A_c and A_f in terms of the original A and terms arising from the SVD of C. Because you have picked values for the α_i, there should be no α_i in your final expressions for full credit.

(HINT: How do the singular values $\sigma_{c,i}$ interact with the α_i? Then apply the appropriate substitution to (13) *to get* (15)*.)*

Solution: From the equation we derived in the previous solution we see that in order to have $\|C\vec{x}\| = \|\vec{\tilde{x}}_c\|$ we should choose $\alpha_1 = \sigma_1, \alpha_2 = \sigma_2, \ldots, \alpha_l = \sigma_l$. Then, by substituting

$$\vec{x} = \begin{bmatrix} V_c B^{-1} & V_f \end{bmatrix} \begin{bmatrix}\vec{\tilde{x}}_c\\ \vec{\tilde{x}}_f\end{bmatrix}$$

which we derived from the previous part into $A\vec{x} = \vec{y}$ to get the desired

$$\begin{bmatrix} A_c, A_f \end{bmatrix} \begin{bmatrix}\vec{\tilde{x}}_c\\ \vec{\tilde{x}}_f\end{bmatrix} = \vec{y},$$

we immediately see that we should choose $A_c = AV_cB^{-1} = AV_c \cdot \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma_2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \frac{1}{\sigma_l} \end{bmatrix}$ and $A_f = AV_f$.

(c) (5pts) Let us focus on a simple case. *(You can do this even if you didn't get the previous parts.)* Suppose that $A = [A_c, A_f]$ where the columns of A_f are orthonormal, as well as orthogonal to the columns of A_c. The columns of A together span the entire n-dimensional space. We directly write $\vec{x} = \begin{bmatrix}\vec{x}_c\\ \vec{x}_f\end{bmatrix}$ so that $A\vec{x} = A_c\vec{x}_c + A_f\vec{x}_f$. Now suppose that we want to solve $A\vec{x} = \vec{y}$ and only care about minimizing $\|\vec{x}_c\|$. We don't care about the length of $\vec{x}_f$ — it can be as big or small as necessary.

In other words, we want to:

$$\min_{\vec{x}=\begin{bmatrix}\vec{x}_c\\ \vec{x}_f\end{bmatrix} \text{ such that } [A_c,A_f]\begin{bmatrix}\vec{x}_c\\ \vec{x}_f\end{bmatrix}=\vec{y}} \|\vec{x}_c\| \tag{16}$$

Show that the optimal solution has $\vec{x}_f = A_f^\top \vec{y}$.

(HINT: Multiplying both sides of something by $A_f^\top$ might be helpful.)

Solution: The optimal solution satisfies what all solutions must satisfy:

$$\vec{y} = A\vec{x} = A_c\vec{x}_c + A_f\vec{x}_f$$

and hence we can multiply both sides by $A_f^\top$ on the left side to get

$$A_f^\top\vec{y} = A_f^\top A_c\vec{x}_c + A_f^\top A_f\vec{x}_f = \vec{0} + I\cdot\vec{x}_f = \vec{x}_f.$$

(d) (8pts) Continuing the previous part, **compute the optimal $\vec{x}_c$.** Show your work.

(HINT: What is the work that $\vec{x}_c$ needs to do? $\vec{y} - A_f A_f^\top \vec{y}$ might play a useful role, as will the SVD of $A_c = \sum_i \sigma_i \vec{t}_i \vec{w}_i^\top$.)

Solution: From the above question, we have

$$\vec{y} = A_c \vec{x}_c + A_f \vec{x}_f = A_c \vec{x}_c + A_f A_f^\top \vec{y}$$

which can be rewritten as

$$A_c \vec{x}_c = (I - A_f A_f^\top)\vec{y}$$

and since we want to minimize $\|\vec{x}_c\|$, the solution will be

$$\hat{\vec{x}}_c = A_c^\dagger (I - A_f A_f^\top)\vec{y},$$

where recall that $A_c^\dagger$ is the Moore-Penrose pseudoinverse of A_c. To calculate the pseudoinverse, let

$$A_c = \sum_i \sigma_i \vec{t}_i \vec{w}_i^\top = T \begin{bmatrix} \Sigma_c & \mathbf{0} \end{bmatrix} W^\top.$$

Here we want Σ_c to be the square matrix of nonzero singular values. Then, the pseudoinverse of A_c is given by

$$A_c^\dagger = W \begin{bmatrix} \Sigma_c^{-1} \\ \mathbf{0} \end{bmatrix} T^\top.$$

Finally, the solution we want is

$$\hat{\vec{x}}_c = W \begin{bmatrix} \Sigma_c^{-1} \\ \mathbf{0} \end{bmatrix} T^\top (I - A_f A_f^\top)\vec{y} = \sum_i \vec{w}_i (\frac{1}{\sigma_i}) \vec{t}_i^\top (I - A_f A_f^\top)\vec{y}.$$

Here, we've kept the fuller form of the SVD while computing the pseudo-inverse to show the shapes involved, but the above could also have been done easily using the compact form of the SVD, or as we have written, the outer-product form.

However, it is interesting to ask the question whether this special case can be further simplified. And indeed, it can be. Notice that the columns $\vec{t}_i$ span the columns of A_c and in this special case, we have said that the columns of A_c are orthogonal to the columns of A_f. This means that $\vec{t}_i^\top A_f = \vec{0}^\top$. Consequently,

$$\hat{\vec{x}}_c = \sum_i \vec{w}_i (\frac{1}{\sigma_i}) \vec{t}_i^\top (I - A_f A_f^\top)\vec{y} = \sum_i \vec{w}_i (\frac{1}{\sigma_i}) \vec{t}_i^\top \vec{y} = W \begin{bmatrix} \Sigma_c^{-1} \\ \mathbf{0} \end{bmatrix} T^\top \vec{y} = A_c^\dagger \vec{y}$$

This is also a fully correct answer for this special case. The psuedo-inverse for a matrix that doesn't have a full column rank automatically contains an implicit projection-type aspect to it. This is something that you could have suspected from the MIMO HW problem's exploration of the connection between least-squares and minimum-norm solutions.

(e) (5pts) Now suppose that A_c did not necessarily have its columns orthogonal to A_f. Continue to assume that A_f has orthonormal columns. *(You can do this part even if you didn't get any of the previous parts.)* Write the matrix $A_c = A_{c\perp} + A_{cf}$ where the columns of A_{cf} are all in the column span of A_f and the columns of $A_{c\perp}$ are all orthogonal to the columns of A_f. **Give an expression for A_{cf} in terms of A_c and A_f.**

(HINT: What does this have to do with projection and least squares?)

Solution: We can see that A_{cf} is the projection of A_c onto A_f, and hence we can directly get $A_{cf} = A_f A_f^\top A_c$ since the A_f has orthonormal columns. This is the style of projection we did when we speeded up OMP, and it is also the combined mode of projection (a whole matrix at a time) we used when we did system identification using least squares.

To expand the argument, we can also do it the following way. Note that

$$A_f^\top A_c = A_f^\top A_{c\perp} + A_f^\top A_{cf} = A_f^\top A_{cf}.$$

Since A_{cf} is in the span of A_f, we can write $A_{cf} = A_f W$ for some W. Then,

$$A_f^\top A_c = A_f^\top A_f W \implies W = (A_f^\top A_f)^{-1} A_f^\top A_c = A_f^\top A_c.$$

In this case, we didn't need to write out the SVD of A_f to get this inverse term. This is because we assumed A_f is orthonormal and hence $A_f^\top A_f = I$ is known.

Finally, recall that we defined $A_{cf} = A_f W$. Then, we conclude

$$A_{cf} = A_f A_f^\top A_c.$$

(f) (8pts) Continuing the previous part, **compute the optimal** $\vec{x}_c$ that solves (16): (copied below)

$$\min_{\vec{x}=\begin{bmatrix}\vec{x}_c\\ \vec{x}_f\end{bmatrix} \text{ such that } [A_c, A_f]\begin{bmatrix}\vec{x}_c\\ \vec{x}_f\end{bmatrix}=\vec{y}} \|\vec{x}_c\|$$

Show your work. Feel free to call the SVD as a black box as a part of your computation.

(HINT: What is the work that $\vec{x}_c$ needs to do? The SVD of $A_{c\perp}$ might be useful.)

Solution: Note that $\vec{y}$ can be written as $\vec{y}_{c\perp} + \vec{y}_f$ such that $\vec{y}_{c\perp}$ is in the span of $A_{c\perp}$, and $\vec{y}_f$ is in the span of A_f. By projection, we have $\vec{y}_f = A_f A_f^\top \vec{y}$ and $\vec{y}_{c\perp} = y - A_f A_f^\top \vec{y} = (I - A_f A_f^\top)\vec{y}$. On the other hand, the amount $\vec{x}_c$ contributes to $\vec{y}_{c\perp}$ is $A_{c\perp}\vec{x}_c$. Hence, all solutions must satisfy:

$$A_{c\perp}\vec{x}_c = \left(I - A_f A_f^\top\right)\vec{y}.$$

Notice that we couldn't have just used A_c here instead of $A_{c\perp}$ because we don't really know what $A_{c\perp}\vec{x}_c$ must be — it is allowed to have some components in the subspace spanned by A_f. However, $A_{c\perp}\vec{x}_c$ cannot have any components in that direction by construction, since every column in $A_{c\perp}$ is orthogonal to the entire subspace spanned by A_f.

Anyway, because we want to minimize $\|\vec{x}_c\|$, we solve this as

$$\vec{x}_c = A_{c\perp}^\dagger \left(I - A_f A_f^\top\right)\vec{y}.$$

To calculate this pseudoinverse, let

$$A_{c\perp} = U_{c\perp}\begin{bmatrix}\Sigma_{c\perp} & \mathbf{0}\end{bmatrix} V_{c\perp}^\top.$$

Then, the pseudoinverse of $A_{c\perp}$ is given by

$$A_{c\perp}^\dagger = V_{c\perp}\begin{bmatrix}\Sigma_{c\perp}^{-1}\\ \mathbf{0}\end{bmatrix} U_{c\perp}^\top.$$

Hence, the solution we want is

$$\vec{x}_c = V_{c\perp} \begin{bmatrix} \Sigma_{c\perp}^{-1} \\ \mathbf{0} \end{bmatrix} U_{c\perp}^\top \left(I - A_f A_f^\top\right) \vec{y}.$$

Once again, we can notice that because the columns of $A_{c\perp}$ are orthogonal to the columns of A_f by construction, that this also carries over to the relevant columns of $U_{c\perp}$ (the ones that correspond to nonzero singular values). Consequently, the above can be simplified further to:

$$\vec{x}_c = V_{c\perp} \begin{bmatrix} \Sigma_{c\perp}^{-1} \\ \mathbf{0} \end{bmatrix} U_{c\perp}^\top \vec{y}$$

by again implicitly leveraging the connection between minimum-norm solutions and least squares. We didn't expect any students to necessarily notice this fact that the pseudo-inverse of $A_{c\perp}$ effectively can do all the work itself.

(g) (6pts) Continuing the previous part, **compute the optimal $\vec{x}_f$.** Show your work.

You can use the optimal $\vec{x}_c$ in your expression just assuming that you did the previous part correctly, even if you didn't. You can also assume a decomposition $A_c = A_{c\perp} + A_{cf}$ from further above in part (e) without having to write what these are, just assume that you did them correctly, even if you didn't do them at all.

(HINT: What is the work that $\vec{x}_f$ needs to do? How is A_{cf} relevant here?)

Solution: We have

$$A_c\vec{x}_c + A_f\vec{x}_f = \vec{y}. \tag{17}$$

The simplest aproach is just to collect the terms and solve for $\vec{x}_f$ by noticing $A_f\vec{x}_f = \vec{y} - A_c\vec{x}_c$ and so $\vec{x}_f = A_f^\top\vec{y} - A_f^\top A_c\vec{x}_c$. You could just stop here for full credit.

Replacing $A_c = A_{c\perp} + A_{cf}$ and then using $A_{cf} = A_f A_f^\top A_c$ as derived in part (e), we have

$$A_{c\perp}\vec{x}_c + A_f A_f^\top A_c\vec{x}_c + A_f\vec{x}_f = \vec{y}. \tag{18}$$

First, we multiply $A_f^\top$ on the left for both sides and recalling $A_f^\top A_{c\perp} = 0$, we get

$$A_f^\top A_c\vec{x}_c + \vec{x}_f = A_f^\top\vec{y}.$$

This again implies that

$$\vec{x}_f = A_f^\top\vec{y} - A_f^\top A_c\vec{x}_c. \tag{19}$$

So doing this substitution doesn't really buy us anything more.

Although you have worked out the problem here for the case of A_f having orthonormal columns, you should hopefully see that you actually could use what you know to solve the fully general case. But those parts will have to wait for the final exam.

Contributors:

- Sidney Buchbinder.
- Pavan Bhargava.
- Anant Sahai.
- Regina Eckert.
- Kuan-Yun Lee.

