EE263 Final Exam, August 2023

- This is a 12-hour take-home exam. Please turn it in on Gradescope.

  Be aware that you *must* turn it in within 12 hours of downloading it. After that, Gradescope will not let you turn it in.
- You may use any books, notes, or computer programs, including searching online, *except for ChatGPT or similar tools.* You may not discuss the exam or course material with others, or work in a group.
- Do not discuss the exam until 8/23, after everyone has taken it.
- If you have a question, please email the staff. We have done our best to make the exam unambiguous. So unless there is a mistake, we are unlikely to say much.
- Please check your email during the exam, in case we need to send a clarification or announcement.
- We expect your solutions to be legible, neat, and clear. Do not hand in your rough notes. Please try to simplify your solutions as much as you can. We will deduct points from solutions that are technically correct, but much more complicated than they need to be.
- Start each solution on a new page. Individual parts can be on same page.
- When a problem involves some computation, we expect a clear discussion and justification of what you did as well as the final numerical result.
- Because this is an exam, **you must turn in your code**. Include the code in your pdf submission. We will deduct points for missing code.
- Good luck!

6.2320. The Sinewave Scalewing. The Sinewave Scalewing is a rare mathematically-inclined sea serpent which can only fly in patterns of the form

$$x(t) = \sum_{k=1}^{N} a_k^{(1)} \sin\left(\frac{k}{30}\pi t\right) + a_k^{(2)} \cos\left(\frac{k}{30}\pi t\right)$$
$$y(t) = \sum_{k=1}^{N} b_k^{(1)} \sin\left(\frac{k}{30}\pi t\right) + b_k^{(2)} \cos\left(\frac{k}{30}\pi t\right)$$

where $a^{(1)}, a^{(2)}, b^{(1)}, b^{(2)} \in \mathbb{R}^N$. For simplicity, define $a = \begin{bmatrix} a^{(1)} \\ a^{(2)} \end{bmatrix}$ and $b = \begin{bmatrix} b^{(1)} \\ b^{(2)} \end{bmatrix}$.

The scalewing is a solitary creature and regularly surveys its territory for intruders. To do so, it plans to start at its home base $(0,0)$, fly over some islands with (x, y) coordinates $(s_1^{(x)}, s_1^{(y)}), \ldots, (s_5^{(x)}, s_5^{(y)})$ and return home (a total of $T = 7$ waypoints). It plans to depart at time t_0, reach island i at time t_i and return at time t_6. You can assume $0 \leq t_i < t_{i+1}$ for all $i = 0, \ldots, T-1$.
You may use $T = 7$, $N = 6$ and the following waypoints and times:

$$s^{(x)} = \begin{bmatrix} 0 & 17 & 25 & 21 & 26 & 8 & 0 \end{bmatrix}^\top$$
$$s^{(y)} = \begin{bmatrix} 0 & -5 & -3 & 5 & 11 & 7 & 0 \end{bmatrix}^\top$$
$$t = \begin{bmatrix} 0 & 4 & 7 & 19 & 25 & 30 & 34 \end{bmatrix}^\top.$$

a) Define two matrices $U \in \mathbb{R}^{T \times N}$ and $V \in \mathbb{R}^{T \times N}$ such that

$$\begin{bmatrix} x(t_0) \\ \vdots \\ x(t_{T-1}) \end{bmatrix} = \begin{bmatrix} U & V \end{bmatrix} a, \qquad \begin{bmatrix} y(t_0) \\ \vdots \\ x(t_{T-1}) \end{bmatrix} = \begin{bmatrix} U & V \end{bmatrix} b.$$

b) Propose a method to plan a trajectory that reaches each island at the correct time while minimizing the loss function
$$\|a\|_2^2 + \|b\|_2^2 \tag{1}$$
which represents the energy the scalewing expends while flying. Clearly describe any matrices you construct. You are encouraged to use block matrix notation and the definitions of U and V from part (a).

c) Use your method to compute a trajectory for the scalewing. Report the minimum energy defined in (**??**) the scalewing must expend to visit all 5 islands. Plot the islands and $\big(x(\hat{t}), y(\hat{t})\big)$ using a smaller timestep; for example, `t_hat = t[1]:0.5:t[end]`. You should generate a 2D plot showing the trajectory of the scalewing and the locations of the islands.

d) Due to its unusual flight pattern, the scalewing has difficulty controlling its altitude $z(t)$, which we can model as
$$z(t) = \frac{1}{10}x(t) + \frac{1}{10}y(t).$$

Suppose the scalewing wants to make a more precise flyover reaching specific (x, y, z) coordinates at each waypoint. Given the coordinates

$$s^{(z)} = \begin{bmatrix} 0 & 5 & 3 & 6 & 2 & 1 & 0 \end{bmatrix}^\top,$$

are the three-dimensional waypoints reachable? Why or why not? Provide a short explanation. Your answer may use facts about the matrices you constructed earlier.

e) Propose a method to minimize the sum of errors

$$J = \sum_{i=0}^{T-1} \left(x(t_i) - s_i^{(x)}\right)^2 + \left(y(t_i) - s_i^{(y)}\right)^2 + \left(z(t_i) - s_i^{(z)}\right)^2.$$

Provide a short justification for your proposed method (not a formal proof).

f) Use your method to compute a three-dimensional trajectory for the scalewing. Display your trajectory, along with the locations of the islands, on a three-dimensional plot using the same smaller timestep as in part (c).

Submit your plot, the value of J, and the energy required for this trajectory.

Hint: In Julia you can import `Plots.jl`, then use `plot3d(x, y, z)` and `scatter3d(x,y,z)`.

Solution. Here is the solution.

a)

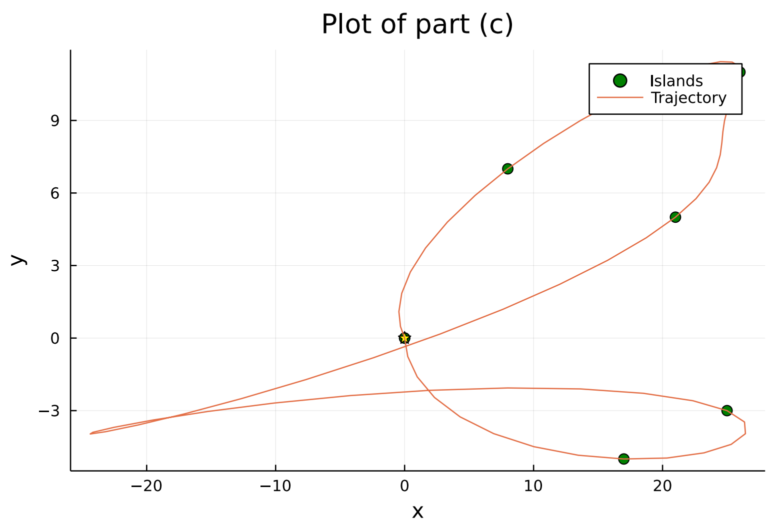
$$U = \begin{bmatrix} \sin\left(\frac{1}{30}\pi t_0\right) & \dots & \sin\left(\frac{N}{30}\pi t_0\right) \\ \vdots & \ddots & \vdots \\ \sin\left(\frac{1}{30}\pi t_{T-1}\right) & \dots & \sin\left(\frac{N}{30}\pi t_{T-1}\right) \end{bmatrix}, \quad V = \begin{bmatrix} \cos\left(\frac{1}{30}\pi t_0\right) & \dots & \cos\left(\frac{N}{30}\pi t_0\right) \\ \vdots & \ddots & \vdots \\ \cos\left(\frac{1}{30}\pi t_{T-1}\right) & \dots & \cos\left(\frac{N}{30}\pi t_{T-1}\right) \end{bmatrix}.$$

b) We can construct the least-norm problem to solve

$$\begin{bmatrix} U & V & 0 & 0 \\ 0 & 0 & U & V \end{bmatrix} \begin{bmatrix} a^{(1)} \\ a^{(2)} \\ b^{(1)} \\ b^{(2)} \end{bmatrix} \approx \begin{bmatrix} s^{(x)} \\ s^{(y)} \end{bmatrix}.$$

Since the matrix is fat, this will return the least-norm solution for $a^{(1)}, a^{(2)}, b^{(1)}, b^{(2)}$.

c) (Julia code at end of solution.)



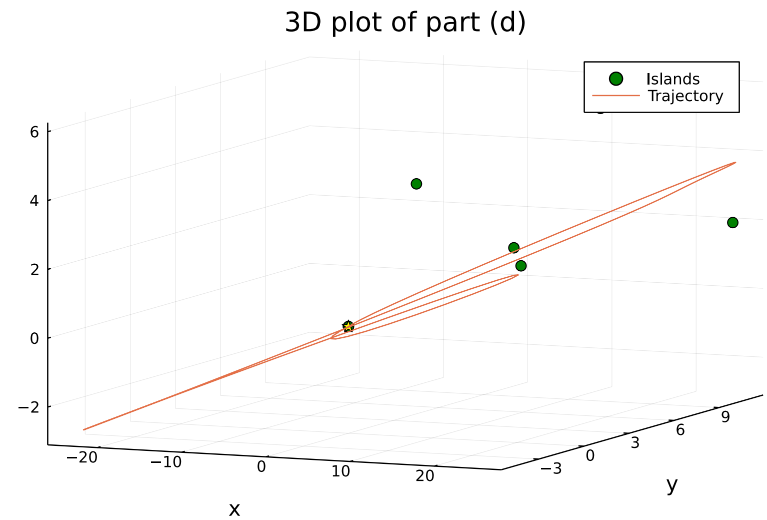
The energy used is 30.827.

d) We can build a new least-squares problem:

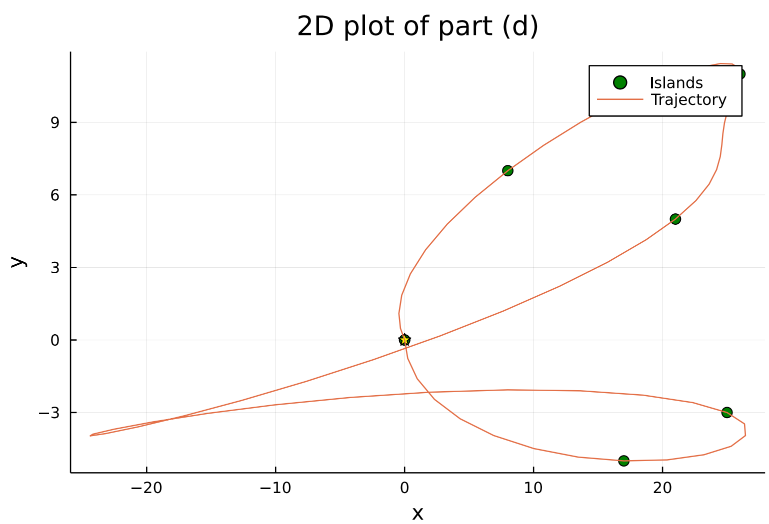
$$\begin{bmatrix} U & V & 0 & 0 \\ 0 & 0 & U & V \\ \frac{1}{10}U & \frac{1}{10}V & \frac{1}{10}U & \frac{1}{10}V \end{bmatrix} \begin{bmatrix} a^{(1)} \\ a^{(2)} \\ b^{(1)} \\ b^{(2)} \end{bmatrix} \approx \begin{bmatrix} s^{(x)} \\ s^{(y)} \end{bmatrix}.$$

However, by inspection (or by Julia code check) the new matrix is not full rank so there will be some unreachable positions. In fact our desired set of waypoints is one of these unreachable positions.

e) (Julia code at end of solution.)



Here's a 2d plot as well.



The energy used is still 30.827.

Here is the Julia code.

18.2310. Particle in magnetic chamber. You have just been employed by Magneto Inc. to work on their Small Particle Collider (SPC). The SPC generates magnetic fields in order to apply forces to particles inside of it. These forces can be summed to produce a resultant force in any direction in 3D space. Your assigned task is to develop a sequence of input forces $u(0), \dots, u(t)$, generated by the SPC walls, that will move particles in specific ways. We will use the vector $x(t)$ to represent the position and velocity of a particle in the x, y, and z directions. Namely,

$$x(t) = \begin{bmatrix} p(t) \\ \dot{p}(t) \end{bmatrix} \in \mathbb{R}^6,$$

where $p(t)$ and $\dot{p}(t)$ represent the particle's position and velocity, respectively, at time t. Note that $u(t) \in \mathbb{R}^3$.

a) We know that the motion dynamics for a particle with mass m can be expressed as $u = m\ddot{p}$. Use the forward Euler approximation, given by $x(t+1) = x(t) + h\dot{x}(t)$, to express this system as a discrete-time linear dynamical system. Explicitly define your matrices A, B, C, and D in terms of m and h. For our vector of outputs, $y(t)$, we are interested in just the particle's position for now, *i.e.*, $y(t) = p(t)$.

b) Write a general expression for

$$\tilde{y} = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(t) \end{bmatrix} \in \mathbb{R}^{3(t+1)}$$

given the initial position and velocity $x(0)$. Your answer may involve $x(0)$, A, B, C, D, and the inputs $u(0), \cdots, u(t)$.

c) We would like to apply forces to the particle such that it returns back to its initial position after some time T has passed. We also want the particle to pass by a detector at time $T/2$ (assuming that T is even).
We would like to find

$$\tilde{u} = \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(T) \end{bmatrix} \in \mathbb{R}^{3(t+1)}$$

such that $y(T) = p_0$ and $y(T/2) = p_{T/2}$, where $p_0, p_{T/2} \in \mathbb{R}^3$ are the initial position and the position of the detector respectively. Also, assume that the initial velocity $\dot{p}_0 \in \mathbb{R}^3$ is known. Describe a method to find the forces that satisfy these constraints while minimizing $\|\tilde{u}\|^2$.

d) Apply your method to the following numerical values and plot the x, y, and z coordinates of the particle from time $t = 0$ to $t = T$. Make sure to plot the $T/2$ way-point as well. You should have one figure with 3 plots, one for each of the axes directions.

$$p_0 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \dot{p}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad p_{T/2} = \begin{bmatrix} 2 \\ -1.5 \\ -1 \end{bmatrix}, \quad T = 200, \quad m = 5, \quad h = 1$$

e) The lead engineer has a new goal; they would like the particle to not only pass through the desired positions as explained in the previous parts, but also to do so with precise velocities denoted by $\dot{p}_0$ and $\dot{p}_{T/2}$. Adjust your definitions from part (a) for x, y, A, B, C, or D to suit this new task if necessary.
Describe a method that achieves the constraints

$$\begin{bmatrix} p(T/2) \\ \dot{p}(T/2) \end{bmatrix} = \begin{bmatrix} p_{T/2} \\ \dot{p}_{T/2} \end{bmatrix}, \quad \begin{bmatrix} p(T) \\ \dot{p}(T) \end{bmatrix} = \begin{bmatrix} p_0 \\ \dot{p}_0 \end{bmatrix}$$

while minimizing $\|\tilde{u}\|^2$. Similar to part (d), run your method on the given data and plot the x, y, and z coordinates of the particle from time $t = 0$ to $t = T$. Use the same numerical values as in part (d), along with

$$\dot{p}_{T/2} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

f) Due to the limitations of the machine, we want to define an objective that represents the rate of change of the force produced. Define

$$J_2 = \sum_{k=1}^{n-1} \|u(k+1) - u(k-1)\|^2$$

as the central difference approximation for the force rate of change. Also define $J_1 = \|\tilde{u}\|^2$. Devise a method that minimizes $J_1 + \gamma J_2$ while adhering to the position and velocity constraints defined in part (e).

g) Plot the new x, y, and z coordinates of the particle over time for $\gamma = 10^p$ for $p \in \{-2, 0, 2, 4, 6\}$. What do you see as γ is increased?

Solution. Here is the solution.

a) If $p(t+1) = p(t) + h\dot{p}(t)$ and $\dot{p}(t+1) = \dot{p}(t) + h\ddot{p}(t)$, and $\ddot{p} = \frac{1}{m}u$, then

$$x(t+1) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t) + Du(t)$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 & h & 0 & 0 \\ 0 & 1 & 0 & 0 & h & 0 \\ 0 & 0 & 1 & 0 & 0 & h \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1/m & 0 & 0 \\ 0 & 1/m & 0 \\ 0 & 0 & 1/m \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$D = 0$$

b) Since this problem is formulated as a discrete-time LDS with inputs and outputs, we already have a closed form solution for $\tilde{y}$:

$$\tilde{y} = \tilde{A}\tilde{u} + \tilde{b}x(0)$$

where

$$\tilde{A} = \begin{bmatrix} D & 0 & \cdots & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ CA^{t-1}B & CA^{t-2}B & \dots & CB & D \end{bmatrix}$$

$$\tilde{u} = \begin{bmatrix} u(0) \\ \vdots \\ u(t) \end{bmatrix}$$

$$\tilde{b} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^t \end{bmatrix}$$

c) Let $\tilde{a}_i^T$ represent the i^{th} row of the block Toeplitz matrix $\tilde{A}$. Then, to satisfy our position constraints at times $t = T/2$ and $t = T$, we only care about the rows of $\tilde{A}$ corresponding to $i = T+1$ and $i = \frac{T+1}{2}$. Thus, we require

$$\hat{y} = \hat{A}\tilde{u} + \hat{b}x(0)$$

where

$$\hat{y} = \begin{bmatrix} y(T/2) \\ y(T) \end{bmatrix}$$

$$\hat{A} = \begin{bmatrix} \tilde{a}^T_{(T+1)/2} \\ \tilde{a}^T_{T+1} \end{bmatrix}$$

$$\hat{b} = \begin{bmatrix} CA^{T/2} \\ CA^T \end{bmatrix}$$

We can then formulate our problem as a least-norm solution with constraints:

$$\begin{gathered} \text{minimize } \|\tilde{u}\|^2 \\ \text{s.t. } \hat{y} - \hat{b}x(0) = \hat{A}\tilde{u} \end{gathered}$$

Using the lecture notes, the optimal solution for $\tilde{u}$ is

$$\tilde{u} = \hat{A}^T(\hat{A}\hat{A}^T)^{-1}(\hat{y} - \hat{b}x(0))$$

d) Insert plots here

e) Redefine $y(t)$ and C as follows:

$$y(t) = x(t)$$

$$C = I$$

Then, repeat the same process as in part (c) and

$$\tilde{u} = \hat{A}^T(\hat{A}\hat{A}^T)^{-1}(\hat{y} - \hat{b}x(0))$$

f) First, we will define a matrix W such that $J_2 = \|W\tilde{u}\|^2$. The W that achieves this is

$$W = \begin{bmatrix} -I & 0 & I & 0 & 0 & \cdots & 0 \\ 0 & -I & 0 & I & 0 & \cdots & 0 \\ \vdots & & & \vdots & & & \vdots \\ 0 & \dots & \dots & 0 & -I & 0 & I \end{bmatrix}$$

Then, our task becomes

$$\text{minimize } J_1 + \gamma J_2 \text{ s.t. } \hat{y} - \hat{b}x(0) = \hat{A}\tilde{u}$$

We can then rewrite

$$J_1 + \gamma J_2 = \|\tilde{u}\|^2 + \gamma\|W\tilde{u}\|^2 = \| \begin{bmatrix} I \\ \sqrt{\gamma}W \end{bmatrix} \tilde{u}\|^2$$

From the lecture notes, we know that the solution to this optimization using Lagrange multipliers is

$$\begin{bmatrix} \tilde{u} \\ \lambda \end{bmatrix} = \begin{bmatrix} I + \gamma W^T W & \hat{A}^T \\ \hat{A} & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \hat{y} - \hat{b}x(0) \end{bmatrix}$$

g) Insert plots here

19.2330. Multi-Armed bandits. A multi-armed bandit is a concept often used in probability theory and decision-making. Imagine you're in a casino facing multiple slot machines (the "arms" of the bandit). Each machine has a different unknown probability of giving you a reward when you pull its lever. You have a limited number of pulls to make and you want to maximize your total reward.

The challenge is to decide which machines to pull and in what order, to strike a balance between exploring (trying out different machines to learn their probabilities) and exploiting (focusing on the machine that seems to give the best rewards based on your current knowledge).

In real-world scenarios beyond casinos, multi-armed bandits can be used to model situations where you have limited resources (like time, money, or experiments) and need to make decisions to optimize outcomes while dealing with uncertainty. This concept finds applications in various fields, such as online advertising, clinical trials, and recommendation systems.

An extension of this setup, often referred to as a stochastic linear bandit, can be formulated as follows: Each potential decision, denoted as an arm of the bandit, is represented by a vector $x \in \mathbb{R}^d$. It is assumed that there exists an unknown parameter vector $\theta \in \mathbb{R}^d$ that influences the observed rewards. To put it more formally, at each time step t, the player is presented with a decision set $\mathcal{D}_t \subseteq \mathbb{R}^d$ from which they must select an action x_t. Consequently, they observe a reward $y_t = x_t^\top \theta + v_t$, where v_t is a zero-mean random noise.

a) Suppose the value of θ is known and $\mathcal{D}_t = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$. Propose an optimal choice for x_t to maximize the reward and briefly explain why it is a good choice.

b) In practice, θ is unknown and we try to estimate it. Suppose we have T pairs of observations (x_i, y_i) for $i = 1, \cdots, T$ from the past decisions and rewards. We aim to find an estimate of θ, namely $\hat{\theta}$, to minimize the cost

$$J = \|X_T\hat{\theta} - y\|^2,$$

where $X_T = \begin{bmatrix} x_1^\top \\ \vdots \\ x_T^\top \end{bmatrix} \in \mathbb{R}^{T\times d}$ and $y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} \in \mathbb{R}^T$.

Which of the following statements most accurately describes the problem of minimizing J when $T < d$? Briefly justify your answer.

i. There are infinitely many solutions that achieve $J = 0$.

ii. There are two possible cases: one unique solution, or infinitely many solutions that achieve $J = 0$.

iii. There are two possible cases: no solution, or infinitely many solutions that achieve $J = 0$. In the case of no solution with $J = 0$, the optimal $\hat{\theta}$ with $J > 0$ can be unique.

iv. There are two possible cases: no solution, or infinitely many solutions that achieve $J = 0$. In the case of no solution with $J = 0$, the optimal $\hat{\theta}$ with $J > 0$ is not unique.

v. There are three possible cases: No solutions, one unique solution, or infinitely many solutions that achieve $J = 0$.

c) It is mostly common to solve the regularized version of the problem, *i.e.*, to minimize the objective

$$J_\mu = \|X_T\hat{\theta} - y\|^2 + \mu\|\hat{\theta}\|^2,$$

where $\mu > 0$ is the regularization parameter.

Give a closed form solution for the optimal $\hat{\theta}$ that minimizes J_μ. Under what circumstances is this solution unique?

d) In some applications, we might have some prior knowledge about θ. A simple case is to have a prior guess, namely θ_0, and we would like our estimated $\hat{\theta}$ to be close to θ_0. In this case, we can minimize the cost

$$J_{\mu,\lambda} = \|X_T\hat{\theta} - y\|^2 + \mu\|\hat{\theta}\|^2 + \lambda\|\hat{\theta} - \theta_0\|^2,$$

where $\mu, \lambda > 0$ are the regularization parameters.

Give a closed form solution for the optimal $\hat{\theta}$ that minimizes $J_{\mu,\lambda}$. Under what circumstances is this solution unique?

e) A practical approach for selecting an action x_t at time t involves estimating θ using $\hat{\theta}_{t-1}$ based on all past observations (x_τ, y_τ) for $\tau = 1, \cdots, t-1$. This estimation is performed by minimizing J_μ. Subsequently, x_t is chosen to maximize the reward under the assumption that $\theta = \hat{\theta}_{t-1}$. Afterward, we observe the reward y_t and incorporate the new pair (x_t, y_t) into the set of observations. This procedure can be repeated for as many steps as needed.

It has been demonstrated that if we continue this process for T steps, there exists a probability of at least 0.99 that the following inequality holds:

$$\|\theta - \hat{\theta}_T\|_{V_T} \leq R,$$

where R is a linear function of $\sqrt{d}$, $\sqrt{\log T}$, and $\sqrt{\mu}$. This means that while we do not know the exact value of θ, there exists a region within which we can assert, with high probability, that θ resides. Here, $\|\theta - \hat{\theta}_T\|_{V_T}$ represents the norm of the difference between θ and $\hat{\theta}_T$ with respect to the matrix V_T, defined as $V_T = X_T^\top X_T + \mu I$. The norm of a vector x with respect to a matrix A is commonly denoted as $\|x\|_A = \sqrt{x^\top A x}$.

What conditions are required for a matrix A to be valid for defining a norm? In simpler terms, what condition must matrix A fulfill to ensure that $\|x\|_A$ is defined for every $x \in \mathbb{R}^d$ and is zero only if $x = 0$? Does V_T meet this condition?

f) According to the previous part, we know that $\|\theta - \hat{\theta}_T\|_{V_T} \leq R$ holds with a high probability. In the worst case scenario, the inequality holds with equality, *i.e.*, $\|\theta - \hat{\theta}_T\|_{V_T} = R$. In such a situation, what is the smallest and largest possible values of $\|\theta - \hat{\theta}_T\|$? Here, the norm denotes the standard Euclidean norm. Your answers should be in terms of R, μ, and possibly the eigenvalues of $X_T^\top X_T$.

Solution.

a) We should choose the input that maximizes the inner product $x^\top\theta$, so

$$x = \frac{\theta}{\|\theta\|}.$$

b) When $T < d$, X_T is a fat matrix, so it has a non-empty null space. If $y \in \text{range}(X_T)$, then at least one exact solution exists—meaning a solution where $J = 0$. However, since the null space is non-empty, there will be infinitely many solutions. On the other hand, if $y \notin \text{range}(X_T)$, there won't be an exact solution. Yet, once we identify an approximate solution, an infinite array of additional solutions emerges since we can generate new solutions by adding a member of the null space to a previous solution. So, the statement (iv) is correct.

c) From the lecture slides, we have that

$$\hat\theta = \left(X_T^\top X_T + \mu I\right)^{-1} X_T^\top y.$$

This is the unique solution of the problem and it holds for any choice of X_T without any further conditions since $X_T^\top X_T + \mu I$ is positive definite.

d) We can write $J_{\mu,\lambda}$ as

$$J_{\mu,\lambda} = \left\| \begin{bmatrix} X_T \\ \sqrt{\mu} I \\ \sqrt{\lambda} I \end{bmatrix} \hat\theta - \begin{bmatrix} y \\ 0 \\ \sqrt{\lambda}\theta_0 \end{bmatrix} \right\|^2 = \|\tilde A \hat\theta - \tilde y\|^2.$$

So, the optimal $\hat\theta$ is

$$\hat\theta = \left(\tilde A^\top \tilde A\right)^{-1} \tilde A^\top \tilde y = \left(X_T^\top X_T + (\mu+\lambda) I\right)^{-1} \left(X_T^\top y + \lambda\theta_0\right).$$

This is the unique solution and holds for any choice of X_T since $X_T^\top X_T + (\mu+\lambda)I$ is positive definite.

e) A has to be positive definite so that $x^\top A x > 0$ for any $x \neq 0$. This holds for V_T as well since $x^\top V_T x = \|X_T x\|^2 + \mu\|x\|^2$.

f) $\|\theta - \hat\theta_T\|_{V_T} \le R$ is an ellipsoid centered at $\hat\theta_T$ associated with the matrix $\frac{1}{\sqrt{R}} V_T$. We are looking for the points on the boundary of the ellipsoid that are closest to or furthest from the center. Thus, the minimum and maximum possible values of $\|\theta - \hat\theta_T\|$ are the lengths of the smallest and largest semi-axes of the ellipsoid, given by $\sqrt{R}\lambda_{\min}(V_T)^{-1/2}$ and $\sqrt{R}\lambda_{\max}(V_T)^{-1/2}$. Note that $\lambda_i(V_T) = \lambda_i(X_T^\top X_T + \mu I) = \lambda_i(X_T^\top X_T) + \mu$, so

$$\|\theta - \hat\theta_T\|_{\min} = \frac{\sqrt{R}}{\sqrt{\mu + \lambda_{\max}\left(X_T^\top X_T\right)}}, \quad \|\theta - \hat\theta_T\|_{\max} = \frac{\sqrt{R}}{\sqrt{\mu + \lambda_{\min}\left(X_T^\top X_T\right)}}$$

21.2300. True or false. For each of the statements below, state whether it is true or false. If true, provide a concise explanation (two or three sentences at most) as to why it holds. If false, give a counterexample. Correct true/false answers with incorrect/missing explanations will receive half credit.

a) The continuous-time linear dynamical system $\dot x = x + Bu$ is never controllable if $u \in \mathbb{R}$, but it can be controllable if $u \in \mathbb{R}^2$.

b) A symmetric matrix $A \in \mathbb{R}^{n\times n}$ with eigenvectors v_1, v_2 corresponding to distinct eigenvalues λ_1, λ_2 exists such that the matrix $P = v_1 v_2^\top$ has exactly one non-zero eigenvalue.

c) If $A \in \mathbb{R}^{m\times n}$, then $\sigma_i(A^\top A + \mu I) \geq \mu$, where $\sigma_i(\cdot)$ is the ith singular value.

d) If $A \in \mathbb{R}^{n\times n}$ and $\kappa(A) = 1$, where $\kappa(\cdot)$ is the condition number of a matrix, then, $c \in \mathbb{R}$ exists such that $\|Ax\| = c\|x\|$ for any $x \in \mathbb{R}^n$.

e) Let $A^\dagger$ be the pseudo-inverse of $A \in \mathbb{R}^{m\times n}$. Then, $A^\dagger A x = x$ if $x \notin \mathrm{null}(A)$.

f) Let $A^\dagger$ be the pseudo-inverse of $A \in \mathbb{R}^{m\times n}$. Then, $AA^\dagger x = x$ if $x \in \mathrm{range}(A)$.

g) Suppose $h(t)$ is the impulse response of a continuous-time linear dynamical system with scalar inputs and outputs. If $h(t) = 0$ for $t > 5$, then the output at time t does not depend on any input from more than 5 seconds ago.

h) Suppose $A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \in \mathbb{R}^{m\times n}$ is given with the QR factorization $A = QR$. If the top left 3×4 block of R is given by $\begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, then $a_1 = 0$, $\{a_2, a_4\}$ is orthonormal, and $\mathrm{rank}(A) \leq n - 2$.

i) Suppose $A \in \mathbb{R}^{n\times n}$ and S is a single-point invariant set of the autonomous linear dynamical system $\dot{x} = Ax$. Then, $S = \{0\}$.

j) If $A \in \mathbb{R}^{4\times 4}$, then $\beta_0, \beta_1, \beta_2, \beta_3$ exist such that $e^A = \beta_0 I + \beta_1 A + \beta_2 A^2 + \beta_3 A^3$.

Please read the following problem setup and then answer the questions (k) through (o) based on it.

Consider the following discrete-time time dynamical system with nonlinear output:

$$\begin{aligned} x(t+1) &= Ax(t) + u(t) \\ y(t) &= x(t)^\top C x(t), \end{aligned}$$

where $A, C \in \mathbb{R}^{n\times n}$ are symmetric and $x(0) = 0$.

k) The problem of maximizing the output at time $t = 1$ with control inputs $u(t)$ such that $\|u(t)\| \leq 1$ is equivalent to maximizing $u(0)^\top C u(0)$ subject to the constraint $\|u(0)\| \leq 1$.

l) Even if C is not positive semi-definite, the maximum value of $y(1)$ is attained when $u(0) = \lambda_1 q_1$, where λ_1 is the largest eigenvalue of C with the corresponding unit-length eigenvector q_1.

m) Suppose $C > 0$. As a measure of the output robustness, we are interested in finding the maximum-norm input $u(0)$ that keeps $y(1) \leq 1$. This maximum-norm input is the longest semi-axis of the ellipsoid $\mathcal{E} = \{x : x^\top C x \leq 1\}$.

n) The maximum-norm input described in the previous part is given by $u(0) = \lambda_1^{-1/2} q_1$, where λ_1 is the largest eigenvalue of C with the corresponding unit-length eigenvector q_1.

o) Suppose $C > 0$ and the cost of applying the control input $u(t)$ is $u(t)^\top D u(t)$, where D is symmetric and positive semi-definite. The problem of finding the most costly control input that maintains $y(1) \leq 1$ is equivalent to maximizing $u(0)^\top C u(0)$ subject to the constraint $u(t)^\top D u(t) \leq 1$.

Solution.

a) **True.** The controllability matrix is $\mathcal{C} = \begin{bmatrix} B & AB \end{bmatrix} = \begin{bmatrix} B & B \end{bmatrix}$. So, for the system to be controllable, $\text{range}(\mathcal{C}) = \text{range}(B)$ has to be $\mathbb{R}^2$, which can happen if B has at least two columns, which means that u must be in $\mathbb{R}^m$ for some $m \geq 2$.

b) **False.** We know that the non-zero eigenvalues of AB and BA are the same, so if P has a non-zero eigenvalue, it is given by $\lambda = v_2^\top v_1$. However, v_1 and v_2 are eigenvectors of a real and symmetric matrix corresponding to distinct eigenvalues, which means that $v_1 \perp v_2$ and that $v_2^\top v_1 = 0$, so all the eigenvalues of P are zeros.

c) **True.** $\lambda_i(A^\top A + \mu I) = \mu + \lambda_i(A^\top A) \geq \mu$, where the inequality is the result of $A^\top A$ being positive semi-definite. Also, for PSD matrices we have that $\sigma_i = \lambda_i$, which yields $\sigma_i \geq \mu$.

d) **True.** Consider the SVD given by $A = U\Sigma V^\top$. If $\kappa = \sigma_{\max}/\sigma_{\min} = 1$, then $\Sigma = \sigma I$, which yields that $A = \sigma U V^\top$. So, A is a multiple of an orthogonal matrix and $\|Ax\| = \sigma\|x\|$.

e) **False.** Consider the SVD given by $A = U\Sigma V^\top$. Then, $A^\dagger A = VV^\top$, so $A^\dagger A x = x$ holds if and only if $x \in \text{range}(V) = \text{null}(A)^\perp$. Note that $x \in \text{null}(A)^\perp$ is not equivalent to $x \notin \text{null}(A)$, so the statement is incorrect.

f) **True.** Consider the SVD given by $A = U\Sigma V^\top$. Then, $AA^\dagger = UU^\top$, so $AA^\dagger x = x$ holds if and only if $x \in \text{range}(U) = \text{range}(A)$.

g) **True.** This immediately follows from the fact that the output is the convolution of the input and the impulse response, *i.e.*, $y(t) = \int_0^t h(t-\tau)u(\tau)\mathrm{d}\tau$.

h) **True.** Let q_1, q_2 be the first two columns of Q. Then, $a_1 = 0, a_2 = q_1, a_3 = q_1 + q_2, a_4 = q_2$. This confirms the first two statements. Also, since $\begin{bmatrix} a_1 & \cdots & a_4 \end{bmatrix}$ has no more than two independent columns, A can have at most $n-2$ independent columns, which confirms the third statement.

i) **False.** $S = \{x_0\}$ is a single-point invariant set if and only if x_0 is an equilibrium point, which means $Ax_0 = 0$ or $x_0 \in \text{null}(A)$.

j) **True.** This is true because all the higher-order terms in the expansion of e^A can be written as a linear combination of I, A, A^2, A^3 according to the Cayley-Hamilton theorem.

k) **True.** This is true since $y(1) = x(t)^\top C x(1) = \left(Ax(0) + u(0)\right)^\top C \left(Ax(0) + u(0)\right) = u(0)^\top C u(0)$

l) **False.** This is only true when $\lambda_1 > 0$. Otherwise, the optimal input will be $u(0) = 0$.

m) **True.** The set of feasible inputs is given by the ellipsoid $\mathcal{E}$, and the input with the maximum-norm is the longest semi-axis of the ellipsoid.

n) **False.** The maximum-norm input corresponds to the smallest eigenvalue, not the largest.

o) **False.** The problem is to maximize the cost, given by $u(t)^\top D u(t) \leq 1$, subject to the constraint of $y(1) = u(0)^\top C u(0) \leq 1$.