AAE 590SC: Stochastic Control

Problem Set 1

Travis Hastreiter

4 September, 2025

Contents

Code Listing

Problem 1

1.1.1)

Consider fully observable linear state-space model $x_{t+1} = Ax_t + Bu_t$ where (A, B) is a stabilizable pair. Provide an algorithm to compute K s.t. the static state feedback controller $u_t = Kx_t$ stabilizes the closed-loop system.
Solution: This problem can be solved using a discrete time infinite horizon LQR controller. The control is defined as $u^* = Kx = -(R + B^T SB)^{-1}B^T SAx$ so $K = -(R + B^T SB)^{-1}B^T SA$. Where S is computed by solving the Discrete Algebraic Riccati Equation (DARE) for its one positive semidefinite matrix solution.

$$S = Q + A^T SA - (A^T SB)(R + B^T SB)^{-1}(B^T SA)$$

Solving the DARE is a very common procedure for control software and it has many efficient implementations.

1.1.2)

Consider the partially observable linear state-space model

$$x_{t+1} = Ax_t + Bu_t,\ y_t = Cx_t$$

Solution Extending the solution for part 1.1.1), adding in a discrete "infinite horizon" kalman filter allows for a discrete infinite horizon LQG controller to be developed to stabilize the system. Following the procedure for part 1.1.1), the LQR feedback gain K_r can be computed through solving the DARE. The Kalman gain matrix can be solved in an analogous way because the Kalman filter and LQR are dual problems of each other. By transposing the A matrix, swapping the B matrix with C^T, changing R to be the measurement noise matrix V_n, changing Q to be the dynamics noise matrix V_d, and changing S to P for clarity, solving DARE gives the matrix necessary for computing the Kalman gain $K_f = -(V_n + CPC^T)^{-1}CPA^T$.

$$P = V_d + APA^T - (APC^T)(V_n + CPC^T)^{-1}(CPA^T)$$

The state estimate gets updated as

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k - K_f(y_k - \hat{y}_k)$$

1.1.3)

Consider state space model from part 1.1.2) where a static output feedback policy $u = Ky_t$ must make the closed-loop system is stable. What do we know about such problems. Solution Reference 6 says when the number of inputs and outputs are small, there exists constructive static output feedback pole placement (SOFPP) methods. Also, determining if SOFPP problem can be solved for a given system and set of poles is NP-hard. Reference 7 talks aabout the following approaches to solving this problem

- Iterative LMI heuristic approaches
- LMI with rank constraints
- Decoupled Lyapunov matrices

Problem 2

(1.2) Three-Body Problem

Next, we consider the spacecraft orbital motion in a three-body system.

(a) Circular Restricted Three-Body Problem (CR3BP)

Consider the Earth-Moon system using the circular restricted three-body problem (CR3BP). The parameter values for this problem are provided in Table 1.

Table 1: Assumed dynamical parameter values (Problem (1.2))

Parameter	Symbol	Value	Unit
Earth-Moon distance	$d_{\text{Earth-Moon}}$	3.8475×10^5	km
Earth-Moon barycenter GM	$G(M_{\text{Earth}} + M_{\text{Moon}})$	4.0350×10^5	km^3/s^2
Mass ratio	μ	1.2151×10^{-2}	-

Simulate the spacecraft's orbital motion in the CR3BP system with the initial conditions and propagation times given in Table 2. Plot each trajectory in the synodic frame in the dimensional system, marking the positions of the Earth and Moon.

Table 2: Initial conditions for CR3BP simulations (non-dimensional)

IC #	x_0	y_0	z_0	$\dot{x}_0$	$\dot{y}_0$	$\dot{z}_0$	Propagation time
IC-1	1.2	0	0	0	-1.06110124	0	6.20628
IC-2	0.85	0	0.17546505	0	0.2628980369	0	2.5543991
IC-3	0.05	-0.05	0	4.0	2.6	0	15.0

(b) Third-Body Perturbation in the ECI Frame

The orbital motion under the influence of Earth and Moon gravity can also be modeled as a perturbed two-body problem, where the Moon exerts a third-body perturbation. To do this, re-define the ECI frame such that:

- $\hat{\boldsymbol{n}}_1$ is aligned with the Earth-Moon line at the epoch.
- $\hat{\boldsymbol{n}}_3$ is aligned with the normal vector of the Earth-Moon orbital plane.

The Moon's third-body perturbation in the ECI frame is given by:

$$\boldsymbol{a}_{\text{moon}} = -\mu_{\text{moon}} \left(\frac{\boldsymbol{r} - \boldsymbol{r}_{\text{moon}}}{\|\boldsymbol{r} - \boldsymbol{r}_{\text{moon}}\|_2^3} + \frac{\boldsymbol{r}_{\text{moon}}}{\|\boldsymbol{r}_{\text{moon}}\|_2^3} \right),$$

where:

- $\mu_{\text{moon}} = 4.9028 \times 10^3 \text{ km}^3/\text{s}^2$,
- $\boldsymbol{r}_{\text{moon}}$ is the Moon's position relative to the ECI frame origin.

For simplicity, assume that the Moon's orbit is circular (similar to the CR3BP model), but about Earth instead of the barycenter.

Convert the initial conditions in Table 4 to position and velocity vectors in the ECI frame and propagate the system using the perturbed two-body equations under the Moon's third-body perturbation. Show the results in both:

- The ECI frame.
- The synodic frame.

(c) Consistency of Results

Compare the results obtained in parts (a) and (b). Discuss the consistency between the two approaches. If discrepancies are observed, explain the possible causes.

(d) Extra Credit: Additional Perturbation

For extra credit, choose another perturbing force of your choice and derive its explicit mathematical expression. Include it in the three-body simulation (either in the CR3BP or ECI frame, at your discretion). Run the simulation with the same initial conditions from Table 4 and compare the results to the previous cases.

The chosen perturbation does not need to be covered in class.

Table 3: Feedback Control Policies Comparison

Controller	State/Output	Static/Dynamic	Determinisitic/Random
LQR	State	Static	Deterministic
LQG	State	Dynamic	Deterministic
P-Control	Output	Static	Determinisitic
PI-Control	Output	Dynamic	Deterministic
Q-Learning	Output	Dynamic	Randomized
Finite State Controller	Output	Static	Randomized

Finite state controller is randomized if the transitions from each node are determined by a probability distribution conditioned by the observation. However, if the state transitions are deterministic then the whole finite state controller is deterministic.

An important observation is that state feedback is a special case of output feedback where your outputs give you full state information.

Problem 3

(1.3)

Find an example of a random process $\{C_t\}, t = 0, 1, 2, \ldots$

$$\text{s.t. } \mathbb{E}(\sum_{\text{t = 0}}^{\infty} C_t) \neq \sum_{\text{t = 0}}^{\infty} \mathbb{E}(C_t)$$

Theorem: (Linearity of Expectation) Let $R_0, R_1, \dots$ be random variables s.t.

$$\sum_{\text{i = 0}}^{\infty} \mathbb{E}(|R_i|) \text{ converges}$$

then

$$\mathbb{E}(\sum_{\text{i = 0}}^{\infty} R_i) = \sum_{\text{i = 0}}^{\infty} \mathbb{E}(R_i)$$

In light of this theorem, an example where $\mathbb{E}(\sum_{\text{t = 0}}^{\infty} C_t) \neq \sum_{\text{t = 0}}^{\infty} \mathbb{E}(C_t)$ requires $\sum_{\text{i = 0}}^{\infty} \mathbb{E}(|R_i|)$ to diverge.

This question illustrates the Gambler's paradox which will be described next. Gambler's Paradox: Consider of a fair Roulette wheel. The expectation of the sum of your expected wins for each bet is zero ($\mathbb{E}(R_i) = 0$) Therefore,

$$\sum_{\text{i = 0}}^{\infty} \mathbb{E}(R_i) = 0$$

However, looking at the problem a different way, the probability of winning at least once as the number of spins approaches infinity is obviously one because it is a fair roulette wheel. Therefore, if a strategy can be devised that gaurentees that if the wheel lands on green and the player wins that the player can end the game and walk away with a net gain in money, then $\mathbb{E}(\sum_{\text{i = 0}}^{\infty} R_i) > 0$.

A well known strategy to achieve this is called "bet doubling." This strategy involves betting a starting amount, say 10, and every time the player loses they double their bet. If the player wins they stop playing and go home. The nth bet is $10 * 2^{n-1}$ whil the probability you make the bet (you haven't won yet) is 2^{-n} so $C_n = 10 * 2^{n-1}$ Therefore,

$$\mathbb{E}(|C_n|) = 10 * 2^{n-1} * 2^{-n} = 20$$
$$\implies \sum_{\text{i = 0}}^{\infty} \mathbb{E}(|C_i|) = 20 + 20 + \dots$$

This means that $\sum_{\text{i = 0}}^{\infty} \mathbb{E}(|C_i|)$ diverges so the theorem does not hold so

$$\mathbb{E}(\sum_{\text{t = 0}}^{\infty} C_t) \neq \sum_{\text{t = 0}}^{\infty} \mathbb{E}(C_t)$$

In solving this I referenced https://eng.libretexts.org/Bookshelves/Computer$_Science/Programming_and_Computation_Fun$

Problem 4

(1.4) Risk-Sensitive Control

Consider polynomial expansion of $\mathcal{L}_\theta$ in θ and derive first two leading terms. Explain why varying $\theta \in \mathbb{R}$ controls the degree of risk sensitivity in the optimal control policy.

The risk sensitive control policy is

$$\min_{\mu_0:T-1} \mathcal{L}_\theta,\ \theta \in \mathbb{R}$$
$$\text{where } \mathcal{L}_\theta = \begin{cases} \frac{1}{\theta}\log(\mathbb{E}(\exp(\theta C_{0:T}))) & \text{if } \theta \neq 0 \\ \mathbb{E}(C_{0:T}) & \text{if } \theta = 0 \end{cases}$$

Assume $\theta \neq 0$ then $\mathcal{L}_\theta = \frac{1}{\theta}\log(\mathbb{E}(\exp(\theta C_{0:T})))$. The Maclaurin series in θ of $\exp(a\theta)$ is

$$\exp(a\theta) = 1 + a\theta + \frac{a^2\theta^2}{2} + \dots$$

The Maclaurin series in x of $\log(1+x)$ is

$$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots$$

Substituting the Macluarin series for $\exp(a\theta)$ into $\mathcal{L}_\theta$ gives

$$\mathcal{L}_\theta = \frac{1}{\theta}\log\left(\mathbb{E}(1 + \theta C_{0:T} + \frac{(\theta C_{0:T})^2}{2} + \dots)\right)$$

Using linearity of expectation,

$$\mathcal{L}_\theta = \frac{1}{\theta}\log\left(1 + (\theta\mathbb{E}C_{0:T} + \frac{\theta^2}{2}\mathbb{E}(C_{0:T})^2 + \dots)\right)$$

Defining $x = (\theta\mathbb{E}C_{0:T} + \frac{\theta^2}{2}\mathbb{E}(C_{0:T})^2 + \dots)$, the Maclaurin seris for $\log(1+x)$ can be used

$$\mathcal{L}_\theta = \frac{1}{\theta}((\theta\mathbb{E}C_{0:T} + \frac{\theta^2}{2}\mathbb{E}(C_{0:T}^2) + \dots) - \frac{1}{2}(\theta(\mathbb{E}C_{0:T})^2 + \text{H.O.T}) + \text{H.O.T}) + \text{H.O.T}$$

Therefore, using the first two leading terms $\mathcal{L}_\theta$ can be approximated by

$$\mathcal{L}_\theta \approx \mathbb{E}C_{0:T} + \frac{1}{2}(\mathbb{E}(C_{0:T}^2) - (\mathbb{E}C_{0:T})^2)\theta$$

Noticing that $\text{Var}(C_{0:T}) = \mathbb{E}(C_{0:T}^2) - (\mathbb{E}C_{0:T})^2$

$$\mathcal{L}_\theta \approx \mathbb{E}C_{0:T} + \frac{1}{2}\text{Var}(C_{0:T})\theta$$

It is now obvious that

$$\begin{cases} \theta > 0 & \text{means that Var}(C_{0:T}) \text{ is penalized (risk averse)} \\ \theta < 0 & \text{means that Var}(C_{0:T}) \text{ is rewarded (risk seeking)} \end{cases}$$

In solving this I referenced https://laurentlessard.com/teaching/me7247/lectures/lecture