Homework 1

ORF 418 - Fall 2025
Due: September 28 (Sunday), 2025 - midnight
version: September 15, 2025

1 (40 points). Consider the *inverted pendulum problem* in infinite horizon of Section 2.3 or of Lecture 5 (September 17). Suppose that

$$h = 0.01, \quad a = 0.1, \quad b = 5.1, \quad L = 2, \quad \rho = 0.81.$$

(You may modify and use the code from the lectures for the following computations.)

a (10 points). Compute the *A and B matrices.*

b (10 points). Suppose that $N = 1$ and

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

Write down the *Riccati equation.*

c (10 points). With the use of a *python package* compute the *solution* V of the Riccati equation and use it to compute the *gains matrix* F.

d (10 points). *Show* that the dynamics given by (A, B) is *controllable.*

---

2 (50 points). Consider the *armed bandit problem* discussed in Lecture 2 (September 8) with $K = 3$ *arms.* As in the lectures, let R_a is random reward if arm a is chosen and set $q(a) = \mathbb{E}[R_a]$. Suppose

$$q(1)=1, \quad q(2)=2, \quad q(3)=10.$$

Consider the ϵ-*greedy algorithm* with $\epsilon = 0.15$, and the *exploration* part of the algorithm *chooses each arm with equal probability.*

The purpose of this exercise is to show that this greedy algorithm will result in a nearly-optimal average reward as T goes to infinity. Let a_t be the action chosen at time t and let r_t be the resulting random reward. We recall several definitions from the lectures:

$$N_T(a) = \sum_{t=1}^{T} \mathbb{1}_{\{a_t=a\}},$$

$$\widehat{q}_T(a) = \frac{1}{N_T(a)} \sum_{t=1}^{T} r_t \, \mathbb{1}_{\{a_t=a\}} \qquad a = 1, 2, 3,$$

where $\mathbb{1}_{\{a_t=a\}}$ is either one or zero depending on if a_t is equal to a or not. In words, $N_T(a)$ is number of times the arm a is chosen up to time T and $\widehat{q}_T(a)$ is an unbiased estimator of $q(a)$ at time T.

a (10 points). *Show* that

$$\liminf_{T\to\infty} \frac{1}{T}\, N_T(a) \,>\, 0, \qquad a = 1, 2, 3.$$

b (10 points). With the help of part a, *argue* that

$$\lim_{T\to\infty} \widehat{q}_T(a) \,=\, q(a), \qquad a = 1, 2, 3.$$

c (15 points). Using the first two parts, for each $a = 1, 2, 3$, *compute*

$$p(a) := \lim_{T\to\infty} \frac{1}{T}\, N_T(a)\,.$$

d (15 points). *Compute*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} r_t\,.$$

f (0 points). *OPTIONAL (will not be graded):*
Run a simulation with $T = 1,000$ or more to test the above result (you may use the code available in Canvas). You may also experiment sending ϵ to zero (in a controlled manner) in the simulations to see how you can achieve the theoretical maximum of $q(3) = 10$.

---

3. (10 points) *Repeat all parts* of the second problem with a general ϵ.
