Homework 1

*Collaboration in the sense of discussion is allowed, however, the assignment is **individual** and the work you turn in should be entirely your own. See the collaboration and academic integrity statement here:* https://natanaso.github.io/ece276b. *Books may be consulted but not copied from. Please acknowledge **in writing** people you discuss the problems with and provide references for any books or papers you use.*

Submission

Upload your solutions on **Gradescope** by the deadline shown at the top right corner. You may use latex, scanned handwritten notes (write legibly!), or any other method to prepare a pdf file. Do not just write the final result. Present your work in detail, explaining your approach at every step.

Problems

In square brackets are the points assigned to each part.

1. Consider the Markov chain shown in Fig. 1.
   (a) [5 pts] What is the transition matrix P?
   (b) [5 pts] Specify a stationary distribution for this Markov chain.
   (c) [10 pts] What is the expected number of times that the chain is in state 1 given that the initial state is 1?
   (d) [10 pts] What is the expected number of steps to reach state 8 given that the initial state is 1?

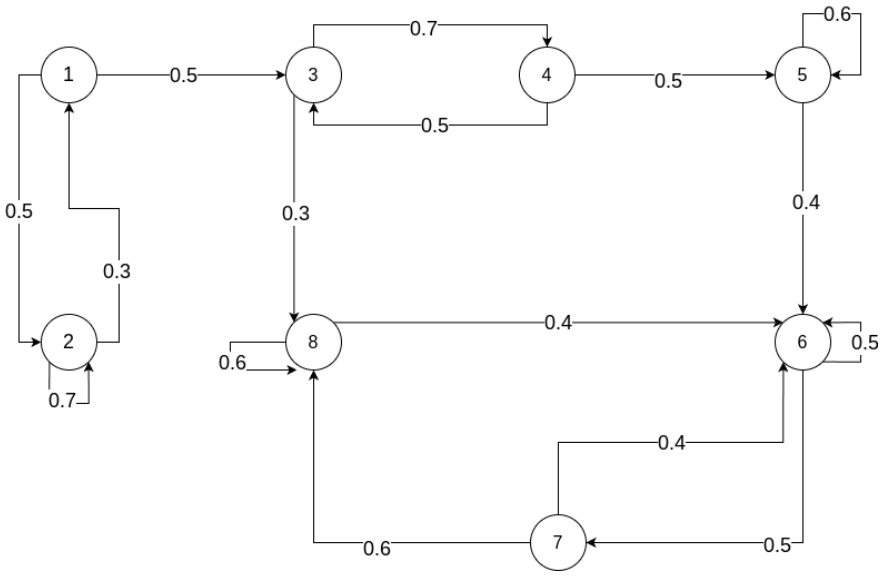

Figure 1: Markov chain with transition probabilities indicated on the edges.

2. Consider a Markov reward process $(\mathcal{X}, p_0, p_f, T, \ell, \mathfrak{q}, \gamma)$ with $\mathcal{X} = \{1, 2, 3\}$, $p_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$, $T = 4$, $\ell(x) = 2x$, $\mathfrak{q}(x) = -x$, $\gamma = 1$, and:

$$p_f(x_{t+1} = j | x_t = i) = P_{ij} \qquad \text{with} \qquad P = \begin{bmatrix} 0.1 & 0 & 0.9 \\ 0.7 & 0.3 & 0 \\ 0 & 0.4 & 0.6 \end{bmatrix}. \tag{1}$$

   (a) [10 pts] What is the probability that all states have been visited by $T = 4$?

   (b) [10 pts] Compute the value $V_0(1)$ of state 1.

3. Consider a cleaning robot, shown in Fig. 2, that needs to collect a used bucket (on the right) and also recharge its batteries (on the left). The cleaning robot is operating on a slippery floor and, as a result, its motion is not deterministic. The robot can choose to move either left or right. When moving in a certain direction, the robot succeeds with probability 0.8, remains in the same location with probability 0.15, and may even move in the opposite of the intended direction with probability 0.05. The robot cannot go outside of the grid in Fig. 2, and any probability likelihood for moving outside of the grid should be assigned to the nearest valid location. Upon reaching the bucket, the robot collects a reward of 5 but needs to recharge its battery subsequently. If the bucket location is visited again before the battery is recharged, the reward is 0. Recharging the battery costs 1 (i.e., a reward of -1) every time the robot reaches the recharging location. The reward in any of the other location is 0. Formulate the problem of minimizing the robot's cost over an infinite horizon as a Markov Decision Process (MDP) with horizon $T = \infty$ and discount factor $\gamma < 1$.

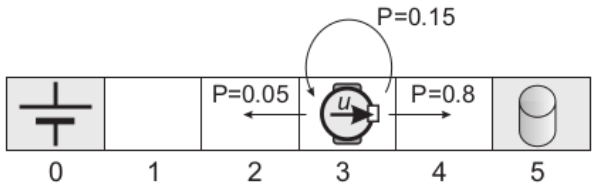

Figure 2: A cleaning robot on a slippery floor. The robot intends to move right but may instead stand still or move left, with different probabilities.

   (a) [5 pts] Define a state space $\mathcal{X}$ and a control space $\mathcal{U}$.

   (b) [5 pts] Define a motion model p_f describing the robot's transitions.

   (c) [5 pts] Define a stage cost function $\ell(\mathbf{x}, u)$.

4. Consider a system with motion model $f(x, u, w) = 2x - u + w$, where $x \in \mathbb{R}$ is the state, $u \in \mathbb{R}$ is the input and $w \in \mathbb{R}$ is the motion noise. Suppose that the motion noise is independent across time and of the state and is identically distributed with $\mathbb{E}[w_t] = 0$ and $\mathbb{E}[w_t^2] = 1$. Consider a finite-horizon optimal control problem with stage cost $\ell(x, u) = x^2$, terminal cost $\mathfrak{q}(x) = x^2$, discount factor $\gamma = 1$, and planning horizon $T = 2$. Let $\pi(x) := 2x - 1$ and $\mu(x) := 3x$ be two control policies.

   (a) [10 pts] Determine the value function $V_0^\pi(x)$ associated with policy π.

   (b) [10 pts] Determine the value function $V_0^\mu(x)$ associated with policy μ.

   (c) [10 pts] Determine the optimal value function $V_0^*(x)$.

   (d) [5 pts] What is the optimality gap between $V_0^*(x)$ and $V_0^\pi(x)$, $V_0^\mu(x)$ at the points x where $V_0^\pi(x) = V_0^\mu(x)$? In other words, compute $V_0^\pi(x) - V_0^*(x)$ for values of x where we have $V_0^\pi(x) = V_0^\mu(x)$.