Homework #1
(Introduction and Structural Properties of MDP)

INF8250AE – Introduction to Reinforcement Learning (Fall 2025)

- **Deadline:** Tuesday, October 7, 2025 at **16:59**.
- **Submission:** You need to submit one PDF file including all your answers. You can produce the file however you like (e.g. LaTeX, Microsoft Word, very neat handwriting, etc) as long as it is readable. Points will be deducted if we have a hard time reading your solutions.
- **Late Submission:** 10% of the marks will be deducted for each day late, up to a maximum of 3 days. After that, no submissions will be accepted.
- **Collaboration:** Homework assignments can be done in a group of at most two students. You must clearly specify the role of each team member in solving the assignment.

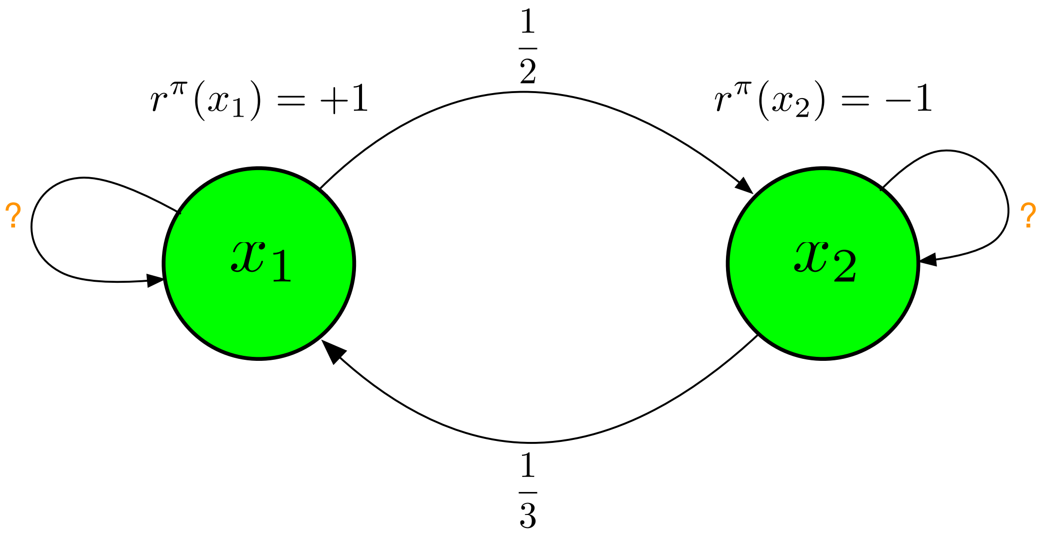

Figure 1: MDP of Exercise 5

Exercise 1. **[10pt]** *Write down the Bellman equation for V^π for a deterministic dynamical system $x_{t+1} = f(x_t, a_t)$ (see Example 1.2 in the textbook) and a deterministic policy $\pi : \mathcal{X} \to \mathcal{A}$.*

Exercise 2. **[10pt]** *Prove that the Bellman operator $T^\pi : \mathcal{B}(\mathcal{X} \times \mathcal{A}) \to \mathcal{B}(\mathcal{X} \times \mathcal{A})$ applied on Q satisfies the monotonicity property.*

Exercise 3. **[30pt]** *Describe a real-world application that can be formulated as an MDP. We encourage you to choose a problem close to your application domain (if your research is related to an application), but feel free to pick any other problem too. Describe what the state space, action space, transition model, and reward are. Do you formulate it as a finite horizon task, or an episodic one, or a continuing one?*

You do not need to be precise in the description of the transition model and reward (no formula is needed). Qualitative description is enough.

Exercise 4. **[10pt]** *Prove that for any $Q \in \mathcal{B}(\mathcal{X} \times \mathcal{A})$ and any $\pi \in \Pi$, we have*

$$\|Q - Q^\pi\|_\infty \leq \frac{\|Q - T^\pi Q\|_\infty}{1 - \gamma}.$$

Exercise 5. **[30pt]** *Consider a 2-state MDP with a fixed policy that induces the transition probabilities and the reward shown in Figure 1. Assume that $\gamma = 0.9$.*

1. **[3pt]** *Write down the missing components of $\mathcal{P}^\pi$:*

$$\mathcal{P}^\pi = \begin{bmatrix} \frac{1}{2} & ? \\ ? & ? \end{bmatrix}.$$

2. **[2pt]** *What is the immediate reward the agent receives?*

3. **[5pt]** *What is the expected reward an agent starting from state x_1 receives after moving* 1 *step in the environment? (clarification: not the immediate reward at x_1, but the one at the next step.) What about* 2 *steps? And* 10*? Do the same calculations (steps: 1, 2, 10) for an agent starting from state x_2.*

   Hint: You can compute this by calculating $(\mathcal{P}^\pi)^k r^\pi$. You need to write one or two lines of code for that calculation; don't do it manually. You don't need to include the code, but if you and your answer is incorrect, we may provide feedback on why the code is wrong.

4. **[5pt]** *The value function function V^π for continuing task is*

$$V^\pi(x) = \mathbb{E}\left[\sum_{t\geq 1} \gamma^{t-1} r^\pi(X_t) \mid X_1 = x\right].$$

   We can approximate this value by truncating at the finite horizon T:

$$V_T^\pi(x) = \mathbb{E}\left[\sum_{t\geq 1}^{T} \gamma^{t-1} r^\pi(X_t) \mid X_1 = x\right].$$

   The value of V_T^π approximates V^π, and it becomes more accurate as T increases. We can compute $V_T^\pi(x)$ by

$$V_T^\pi = \sum_{k\geq 0}^{T-1} \gamma^k (\mathcal{P}^\pi)^k r^\pi.$$

   Now we ask you to calculate V_T^π for $T \in \{2, 10, 20, 50, 100\}$.

5. **[5pt]** *Write down the Bellman equation $V^\pi = T^\pi V^\pi$ explicitly, that is, the exact values of $\mathcal{P}^\pi$ and r^π should appear.*

6. **[8pt]** *Compute the value function V^π (for continuing task) by solving the Bellman Equation manually (so no Value Iteration). This is a bit tedious, but doable with a pen and paper.*

7. **[2pt]** *Compare your result with V_T^π that you got in part 4 of this question. What T is enough so that $\|V^\pi - V_T^\pi\|_\infty$ is smaller than* 0.1*?*

Exercise 6. [10pt] *Suppose that the "multiplicative" Bellman operator is defined as*

$$(T^\pi_{mult} V)(x) \triangleq r^\pi(x) \int \mathcal{P}^\pi(\mathrm{d}x'|x) V(x'),$$

for any $x \in \mathcal{X}$. Answer the following questions:

1. **[5pt]** *Is the operator* T^{π}_{mult} *monotonic?*
   If yes, prove it. If not, what assumptions do we need to make in order to guarantee its monotonicity? Hint: Note that even though $2 \times 4 \leq 2 \times 5$, it is not true that $(-2) \times 4 \leq (-2) \times 5$.

2. **[5pt]** *Is it a contraction operator? Specify the contraction factor.*
   If yes, prove it. If not, what conditions do you need for that to hold?