Assignment 3

Lilian Kourti
CME 241

<u>Problem 1</u>

Bellman Equations for the Deterministic Policy $\pi_D : \mathcal{S} \rightarrow \mathcal{A}$:

$$
\begin{aligned}
V^{\pi_D}(s) &= Q^{\pi_D}(s, \pi_D(s)), \quad \forall s \in \mathcal{S} \\
Q^{\pi_D}(s,a) &= R(s,a) + \gamma \sum_{s' \in N} P(s,a,s') V^{\pi_D}(s'), \quad \forall s \in \mathcal{S},\ a \in \mathcal{A} \\
V^{\pi_D}(s) &= R(s, \pi_D(s)) + \gamma \sum_{s' \in N} P(s, \pi_D(s), s') V^{\pi_D}(s'), \quad \forall s \in \mathcal{S} \\
Q^{\pi_D}(s,a) &= R(s,a) + \gamma \sum_{s' \in N} P(s,a,s') Q^{\pi_D}(s', \pi_D(s')), \quad \forall s \in \mathcal{S},\ a \in \mathcal{A}
\end{aligned}
$$

Problem 2

We observe that the transition probabilities and the reward function are the same $\forall s \in \mathcal{S}$ and $a \in [0,1]$. The same holds true for the reward function. This means that the dynamics of the given infinite-state MDP are the same as the finite-state MDP with two states, let's say $s_1,\ s_2$ that has transition probabilities:

$$\mathbb{P}[s_2|s_1,a] = a, \mathbb{P}[s_1|s_1,a] = 1-a, \mathbb{P}[s_1|s_2,a] = a, \mathbb{P}[s_2|s_2,a] = 1-a,\ \ a \in [0,1]$$

and reward function:

$$R_T(s_1,a,s_2) = 1-a,\ \ R_T(s_1,a,s_1) = 1+a,\ \ R_T(s_2,a,s_1) = 1-a,\ \ R_T(s_2,a,s_2) = 1+a$$

Hence, in our initial MDP the Optimal Value Function V*(s) is the same $\forall s \in \mathcal{S}$ and it satisfies the Bellman Optimality Equations:

$$V^*(s) = \max_a\{R(s,a) + \gamma \sum_{s' \in S} P(s,a,s')V^*(s')\}$$

where

$$\begin{aligned} R(s,a) &= \mathbb{E}[R_{t+1}|S_t = s, A_t = a] \\ &= \sum_{s' \in S} P(s,a,s')R_T(s,a,s') \\ &= P(s,a,s+1)R_T(s,a,s+1) + P(s,a,s)R_T(s,a,s) \\ &= a(1-a) + (1-a)(1+a) \\ &= (1-a)(2a+1) \\ &= -2a^2 + a + 1 \end{aligned}$$

Therefore:

$$\begin{aligned} V^*(s) &= \max_{a \in [0,1]}\{-2a^2 + a + 1 + \gamma(aV^*(s) + (1-a)V^*(s)\} \\ \implies V^*(s) &= \max_{a \in [0,1]}\{-2a^2 + a + 1 + \gamma V^*(s)\} \\ \implies (1-\gamma)V^*(s) &= \max_{a \in [0,1]}\{-2a^2 + a + 1\} \\ \xRightarrow{\gamma=0.5} 0.5V^*(s) &= \max_{a \in [0,1]}\{-2a^2 + a + 1\} \\ \implies 0.5V^*(s) &= 9/8 \\ \implies V^*(s) &= 9/4 = 2.25, \quad \forall s \in \mathcal{S} \end{aligned}$$

The above maximization is achieved for $a = 1/4$, which means that the optimal policy $\forall s \in \mathcal{S}$ is $\pi^*(s) = 1/4$.

Problem 3

May be added later if time allows.

Problem 4

Our goal is to to minimize the infinite-horizon Expected Discounted-Sum of Costs when $\gamma = 0$ (myopic case).

$$
\begin{aligned}
V^*(s) &= \max_\pi V^\pi(s), \quad \forall s \in S \\
&= \max_a \{\mathbb{E}_{s' \sim N(s,\sigma^2)}[-e^{as'}|s]\} \\
&= \max_a \{-\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\sigma)^2}{2\sigma^2}} e^{ax}\, dx\} \\
&= \max_a \{-\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2-2xs+s^2-2\sigma^2 ax}{2\sigma^2}}\, dx\} \\
&= \max_a \{-\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2-2x(s+\sigma^2 a)+(s+\sigma^2 a)^2}{2\sigma^2}+as+\frac{\sigma^2 a^2}{2}}\, dx\} \\
&= \max_a \{-e^{as+\frac{\sigma^2 a^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-(s+\sigma^2 a))^2}{2\sigma^2}}\, dx\} \\
&= \max_a \{-e^{as+\frac{\sigma^2 a^2}{2}} \mathbb{E}_{x \sim N(s+\sigma^2 a, \sigma^2)}[1]\} \\
&= \max_a \{-e^{as+\frac{\sigma^2 a^2}{2}}\}
\end{aligned}
$$

The optimal policy satisfies $a^* = \arg\max_a \{-e^{as+\frac{\sigma^2 a^2}{2}}\}$. Hence, we solve for:

$$
\frac{\partial \{-e^{as+\frac{\sigma^2 a^2}{2}}\}}{\partial a} = 0 \quad \Longrightarrow \quad s + a^*\sigma^2 = 0
$$
$$
\Longrightarrow \quad a^* = -\frac{s}{\sigma^2}
$$

which achieves the maximum, since the function is concave (2nd derivative is negative). Therefore, the Optimal Action (Deterministic Policy) is:

$$
\pi^*(s) = -\frac{s}{\sigma^2}, \quad \forall s \in S
$$

and by substituting a^*, we get that the Optimal Value Function is:

$$
V^*(s) = -e^{-\frac{s^2}{\sigma^2}+\frac{s^2}{2\sigma^2}} = -e^{-\frac{s^2}{2\sigma^2}}, \quad \forall s \in S
$$

which results in an Optimal Cost equal to $e^{-\frac{s^2}{2\sigma^2}}, \quad \forall s \in S$.