Problem 1

In this problem, we consider an extension of the standard stochastic optimal control problems to a class of stochastic games known as zero-sum games. In this class of problems, we can have two players, one is trying to maximize a cost function and the other is trying to minimize it. Imagine a scenario where an inspector's job is to prevent dumping wastes at N known locations and he has to travel from one place to another in order to catch the "dumpers". The dumpers, in turn, are interested in destroying the environment and they track the inspector's most recent location. In this scenario, we call the decisions the inspector makes "control inputs" and the decisions that the dumpers make "adversary inputs". The adversary input does not know the actions of the control input, but knows the location (state) of the game. Modeling the uncertainty as an adversary can be more effective in many applications than the simpler noise model we have been considering.

Typical in such formulations is not to consider the strategy as the actual control input, but rather a probability distribution on the set of inputs. We certainly could have formulated our infinite horizon problems that way. As an example, if we considered the standard discounted cost problem on finite states, instead of taking the action to be $u \in U(x)$ (which is referred to as pure strategy), we could have mixed the actions by deciding a pdf $\pi(x)$ on the set of allowable inputs $U(x)$ (which is referred to as mixed strategy). This means that the control input is randomized at each given state according to $\pi(x)$.

Next we formulate the standard zero-sum game as a mimimax problem. Let $u \in U$ and $v \in V$, both finite sets of dimension r and s respectively. Consider a MDP with a finite number of states $S = 1, \ldots, N$, and with transition probabilities $P_{ij}(u, v)$. The input v is the adversary input and is selected as a stationary mixed policy according to a pdf $\gamma(x)$. Explicitly,

$$\gamma(x)(j) = \text{Prob}\{v(x) = v_j\} \quad j = 1, \ldots 2$$

If $\gamma(i)$ is represented as a row vector, then the matrix of dimension $N \times s$, $\Gamma = col(\gamma(i)$, represents a strategy. Similarly, the control input u, is selected as a stationary mixed policy according to a pdf $\pi(x)$. The strategy $N \times r$ matrix $\Pi = col(\pi(i))$. Define the cost function as:

$$J(i, \Pi, \Gamma) = \sum_{k=0}^{\infty} E\{\alpha^k g(x_k, u_k, v_k) | x_0 = i\}$$

where the expectation is taken over the future states and the different inputs, and $\alpha < 1$ is a discount factor.

The minimax problem can now be formulated as follows:

$$J^*(i) = \min_{\Pi} \max_{\Gamma} J(i, \Pi, \Gamma)$$

1. Consider the finite horizon version of this problem over a period of length M and final cost

equal to $\alpha^M J_0(x)$:

$$J_M(i) = \min_{\Pi} \max_{\Gamma} \sum_{k=0}^{M-1} E\{\alpha^k g(x_k, u_k, v_k) + \alpha^M J_0(x_M) | x_0 = i\}$$

Show that

$$\lim_{M \to \infty} J_M(i) = J^*(i)$$

2. Derive the corresponding Bellman equation for this problem. Use the previous part to show that value iteration converges to the optimal solution.

3. Assume that the transition probabilities do not depend on the actions of the adversary, v. Derive an LP formulation of this problem.

4. Show that if $\gamma(x) = \gamma$ independent of the state, then the optimal control strategy is pure (i.e. the distribution $\pi(x) = e_{j(x)}$ for some index $j(x)$ and e is the unit vector.