# Assignment 2

Kartik Bharadwaj CS20S020

April 24, 2021

**Declaration**: I pledge that I have not copied or given any unauthorized assistance on this assignment.

## 1 Problem 1

We are given:

$$ce \leq f(x') - x' \leq de \tag{1}$$

Also, contraction mapping definition says:

$$||f(x^*) - f(x')||_\infty \leq \alpha ||x^* - x'||_\infty \tag{2}$$

$$\|(x^* - x') - (f(x') - x')\|_\infty \leq \alpha \|x^* - x'\|_\infty \qquad (\because f(x^*) = x^*)$$

$$\big| \|x^* - x'\|_\infty - \|f(x') - x'\|_\infty \big| \leq \alpha \|x^* - x'\|_\infty \qquad \text{(Using triangle inequality)}$$

$$\implies -\alpha \|x^* - x'\|_\infty \leq \|x^* - x'\|_\infty - \|f(x') - x'\|_\infty \leq \alpha \|x^* - x'\|_\infty$$

$$\implies -\|f(x') - x'\|_\infty \leq (\alpha - 1) \|x^* - x'\|_\infty \tag{3}$$

$$\implies \|x^* - x'\|_\infty \leq \frac{1}{1-\alpha} \|f(x') - x'\|_\infty \tag{4}$$

Removing norm from (4) on both sides,

$$\implies x^* - x' \leq \frac{1}{1-\alpha} [f(x') - x'] \tag{5}$$

Using (1) in (5),

$$\implies x^* - x' \leq \frac{de}{1-\alpha}$$

$$\implies x^* \leq x' + \frac{|d|}{1-\alpha} e \tag{6}$$

Using (2) in (4),

$$\frac{\|x^* - f(x')\|_\infty}{\alpha} \leq \frac{1}{1-\alpha} \|f(x') - x'\|_\infty \tag{7}$$

Removing norm from (7) on both sides,

$$\implies x^* - f(x') \leq \frac{\alpha}{1-\alpha} [f(x') - x'] \tag{8}$$

From (1) and (8), we have:

$$\implies x^* \leq f(x') + \frac{\alpha |d|}{1-\alpha} e \tag{9}$$

Removing norm from (3) on both sides,

$$\frac{[f(x') - x']}{1-\alpha} \leq x^* - x' \tag{10}$$

Using (1) in (10),

$$\implies \frac{ce}{1-\alpha} \leq x^* - x'$$

$$\implies x^* \geq x' - \frac{|c|}{1-\alpha} e \tag{11}$$

Rewriting (2) as,

$$\frac{-||f(x^*) - f(x')||_\infty}{\alpha} \geq -||x^* - x'||_\infty \tag{12}$$

Using (3) in (12),

$$\implies \frac{-||f(x^*) - f(x')||_\infty}{\alpha} \geq -\frac{||f(x') - x'||_\infty}{1-\alpha}$$
$$\implies ||f(x^*) - f(x')||_\infty \geq \frac{\alpha}{1-\alpha}||f(x') - x'||_\infty \tag{13}$$

Removing norm from (13) and using (1) in it,

$$\implies x^* - f(x') \geq \frac{\alpha c}{1-\alpha}e$$
$$\implies x^* \geq f(x') - \frac{\alpha|c|}{1-\alpha}e \tag{14}$$

Finally, combining (1), (6), (9), (11), & (14), we have:

$$x' - \frac{|c|}{1-\alpha}e \leq f(x') - \frac{\alpha|c|}{1-\alpha}e \leq x^* \leq f(x') + \frac{\alpha\ |d|}{1-\alpha}e \leq x' + \frac{|d|}{1-\alpha}e$$

# 2 Problem 2

## 2.1 Part (a)

$$\tilde{p_{ij}} = \frac{p_{ij} - m_j}{1 - \sum_{k=1}^n m_k}$$
$$\implies \sum_{j=1}^n \tilde{p_{ij}} = \frac{\sum_{j=1}^n p_{ij} - \sum_{j=1}^n m_j}{1 - \sum_{k=1}^n m_k}$$

Since $\sum_{j=1}^n p_{ij} = 1$,

$$\sum_{j=1}^n \tilde{p_{ij}} = \frac{1 - \sum_{j=1}^n m_j}{1 - \sum_{k=1}^n m_k} = 1$$

Therefore, $\tilde{p_{ij}}$ are transition probabilities.

## 2.2 Part (b)

Using Bellman's Equation:

$$\tilde{J}(i) = \min_{a\in A}\Big[g(i,a) + \tilde{\alpha}\sum_{j=1}^n \tilde{p_{ij}}(a)\tilde{J}(j)\Big]$$

Substituting the values of $\tilde{\alpha}$ and $\tilde{p_{ij}}(a)$,

$$\tilde{J}(i) = \min_{a\in A}\Big[g(i,a) + \alpha(1 - \sum_{k=1}^n m_k)\sum_{j=1}^n \frac{p_{ij}(a) - m_j}{1 - \sum_{k=1}^n m_k}\tilde{J}(j)\Big]$$
$$\tilde{J}(i) = \min_{a\in A}\Big[g(i,a) + \alpha\sum_{j=1}^n (p_{ij}(a) - m_j)\tilde{J}(j)\Big]$$
$$\tilde{J}(i) = \min_{a\in A}\Big[g(i,a) + \alpha\sum_{j=1}^n p_{ij}(a)\tilde{J}(j) - \alpha\sum_{k=1}^n m_k\tilde{J}(k)\Big]$$

Minimizing over actions,

$$\tilde{J}(i) + \frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}e = \min_{a\in A}\Big[g(i,a) + \alpha\sum_{j=1}^n p_{ij}(a)\tilde{J}(j) - \alpha\sum_{k=1}^n m_k\tilde{J}(k) + \frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}\Big]$$
$$\tilde{J}(i) + \frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}e = \min_{a\in A}\Big[g(i,a) + \alpha\sum_{j=1}^n p_{ij}(a)\tilde{J}(j) - \alpha\sum_{k=1}^n m_k\tilde{J}(k)\big(1 - \frac{1}{1-\alpha}\big)\Big]$$
$$\tilde{J}(i) + \frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}e = \min_{a\in A}\Big[g(i,a) + \alpha\sum_{j=1}^n p_{ij}(a)\tilde{J}(j) + \alpha\frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}\Big]$$

We know that $\sum_{j=1}^n p_{ij} = 1$,

$$\tilde{J}(i) + \frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}e = \min_{a\in A}\Big[g(i,a) + \alpha\sum_{j=1}^n p_{ij}(a)\tilde{J}(j) + \alpha\sum_{j=1}^n p_{ij}(a)\frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}\Big]$$
$$\tilde{J}(i) + \frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}e = \min_{a\in A}\Bigg[g(i,a) + \alpha\sum_{j=1}^n p_{ij}(a)\Big(\tilde{J}(j) + \frac{\alpha\sum_{k=1}^n m_k\tilde{J}(k)}{1-\alpha}\Big)\Bigg]$$

Comparing the above equation with the Bellman equation for the original question, we have:

$$J^*(i) = \tilde{J}(i) + \frac{\alpha \sum_{k=1}^{n} m_k \tilde{J}(k)}{1-\alpha} e$$

# 3 Problem 3

## 3.1 Part (a)

Let $\mathcal{X} = \{1, 2, \ldots, T\}$. Let's assume we are in state $x_k \in \mathcal{X}$, where $x_k \neq T$. Then,

$$\tilde{P}(x_{k+1}|x_k, a, \text{heads}) = \begin{cases} 0 & \text{if } x_{k+1} = T \\ P(x_k, a, x_{k+1}) & \text{if } x_{k+1} \in \mathcal{X} \end{cases}$$

$$\tilde{P}(x_{k+1}|x_k, a, \text{tails}) = \begin{cases} 1 & \text{if } x_{k+1} = T \\ 0 & \text{if } x_{k+1} \in \mathcal{X} \end{cases}$$

Using $p(\text{heads}) = 1 - \beta$,

$$\tilde{P}(x_k, a, x_{k+1}) = \tilde{P}(x_{k+1}|x_k, a, \text{heads}) \cdot p(\text{heads}) + \tilde{P}(x_{k+1}|x_k, a, \text{tails}) \cdot p(\text{tails})$$

$$\tilde{P}(x_k, a, x_{k+1}) = \begin{cases} 1 - \beta \cdot P(x_k, a, x_{k+1}) & \text{if } x_{k+1} \in \mathcal{X} \\ \beta & \text{if } x_{k+1} = T \end{cases}$$

Similarly, if $x_k = T$,

$$\tilde{P}(x_k, a, x_{k+1}) = \begin{cases} 1 & \text{if } x_{k+1} = T \\ 0 & \text{if } x_{k+1} \in \mathcal{X} \end{cases}$$

## 3.2 Part (b)

Discount factor of the MDP variant: $1 - \beta$. The MDP variant will continue to be of discounted type, if the discount factor $\alpha$ of the original MDP is 1. Discounting the game's rewards by a factor of $\alpha$ is same as playing without discounting ($\alpha = 1$) but where the probability that the game ends is $\beta$.

# 4 Problem 4

Given an initial state $x_0 \in \mathcal{X}$, in a finite horizon MDP:

$$J_\pi(x_0) = \mathbb{E}_{x_1,\ldots,x_N}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), x_{k+1})\right]$$

where, $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$.

Hence, the optimization objective is:

$$J_{\pi^*}(x_0) = \min_{\pi in \Pi} J_\pi(x_0)$$

where, $\pi^* = \arg\min_{\pi \in \Pi} J_\pi(x_0)$.

Let $\mathcal{X}$ and $\mathcal{A}$ be finite. Then, the PI algorithm converges to the optimal policy after at most $|A|^{|S|}$ iterations, where $|S|$ is the number of states and $|A|$ is the number of actions. Not that we don't need to have policies as proper and stationary for finite MDP since the policies will depend on horizon $\mathcal{H}$.

Deriving the policy evaluation step:

We know that $J_\pi(x) = J_0(x)$. Then, for $k = \{N-1, N-2, \ldots, 0\} \& \forall x \in \mathcal{X}$

$$\begin{aligned} J_N &= g_N \\ J_k &= \mathcal{T}_{\mu_k} J_{k+1} \\ \implies J_\pi &= \mathcal{T}_{\mu_0} \mathcal{T}_{\mu_1} \ldots \mathcal{T}_{\mu_{N-1}} J_N \\ \implies J_\pi &= \mathcal{T}^\pi J_N \end{aligned}$$

In finite MDP, since we have finite number of policies, the termination condition is met for a specific $k$. Now, let's assume we get policy $\pi'$ in the policy improvement step. Then, the policy $\pi'$ is optimal if:

$$J_{\pi'} = \mathcal{T}^{\pi'} J_{\pi'} = \mathcal{T} J_{\pi'}$$

Hence, $\pi'$ is optimal and $J_{\pi'} = J^*$.

Therefore, for finite MDP, the PI algorithm is:

**Algorithm 1** Policy Iteration

**repeat**
  **Initialize:** $J_N(x_N) = g_N(x_N), \pi(x) \forall x \in \mathcal{X}$;
  *# Policy Evaluation*
  **for** k $\in \{N-1, N-2, \ldots, 0\}$ **do**
    **for** each $x_k \in \mathcal{X}$ **do**
      $J_k(x_k) = \sum_{x_{k+1} \in \mathcal{X}} P(x_k, \mu_k(x_k), x_{k+1}) \left[ g(x_k, \mu_k(x_k), x_{k+1}) + J_{k+1}(x_{k+1}) \right]$
    **end for**
  **end for**
  *# Policy Improvement*
  $done = 1$;
  **for** k $\in \{N-1, N-2, \ldots, 0\}$ **do**
    **for** each $x_k \in \mathcal{X}$ **do**
      $b = \mu_k(x_k)$
      $\mu_k(x_k) = \arg\min_{a \in A(x_k)} \sum_{x_{k+1} \in \mathcal{X}} P(x_k, \mu_k(x_k), x_{k+1}) \left[ g(x_k, \mu_k(x_k), x_{k+1}) + J_{k+1}(x_{k+1}) \right]$
      **if** $b \neq \mu_k(x_k)$ **then**
        $done = 0$;
      **end if**
    **end for**
  **end for**
**until** done=1

# 5 Problem 5

## 5.1 Part (a)

1. State space $x_k = \{T_1, T_2\}$; $T_1$ = Type-I, $T_2$ = Type-II
2. Action space $\mu_k(x_k) = \{I, NI\}$; $I$ = Incentivize, $NI$ = Not Incentivize
3. Transition probabilities
   (a) $P(x_{k+1} = T_2 | x_k = T_1, a_k = I) = p_{12}^{I} = 0.75$
   (b) $P(x_{k+1} = T_1 | x_k = T_1, a_k = NI) = p_{11}^{NI} = 0.75$
   (c) $P(x_{k+1} = T_2 | x_k = T_2, a_k = I) = p_{22}^{I} = 0.8$
   (d) $P(x_{k+1} = T_2 | x_k = T_2, a_k = NI) = p_{22}^{NI} = 0.4$
4. Single stage reward/profit: Depends only on current state and action
   (a) $g(T_1, I) = 2500$
   (b) $g(T_1, NI) = 2500$
   (c) $g(T_2, I) = 15000$
   (d) $g(T_2, NI) = 10000$

# References


- **Problem 1**: Contraction mapping, Lecture Notes
- **Problem 2**: DPOC Vol I and II, Lecture Notes
- **Problem 3**: Lecture Notes, Discussed with Richa Verma (CS20D020)
- **Problem 4**: Policy Iteration proof, Lecture Notes;
- **Problem 5**: Lecture Notes