ECE5984: Reinforcement Learning

Assignment #1

Nima Mohammadi

nimamo@vt.edu

Problem 1.

Consider a finite discrete-time Markov chain (DTMC) $\{s_n\}$ taking values in $\{1, 2\}$ with transition probability matrix

$$P = \begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix}$$

where $P_{ij} = \mathbb{P}\left(s_{n+1} = j \mid s_n = i\right)$. Let $\{Y_n\}$ be a different random process defined as

$$Y_n = \begin{cases} s_n, & \text{with probability } 0.7 \\ s_n - 1 & \text{with probability } 0.3 \end{cases}$$

1. Find the stationary distribution of P, i.e., find π such that $\pi P = \pi$.

The stationary distribution π is the row vector such that $\pi = \pi P$. Therefore, we can find our stationary distribution by solving the following linear system:

$$\begin{aligned} 0.3\pi_1 + 0.6\pi_2 &= \pi_1 \\ 0.7\pi_1 + 0.4\pi_2 &= \pi_2 \\ \pi_1 + \pi_2 &= 1 \end{aligned}$$

which corresponds to solving the linear system below:

$$\begin{bmatrix} -0.7 & 0.6 \\ 0.7 & -0.6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

which gives the solution:

$$\pi_1 = 6/13, \quad \pi_2 = 7/13$$

2. Find $\lim_{n\to\infty} P(s_n = 1 \mid Y_n = 1)$. [Hint: Use Bayes rule formula].

By Bayes rule we have

$$P(s_n = 1 \mid Y_n = 1) = \frac{P(Y_n = 1 \mid s_n = 1)\, P(s_n = 1)}{P(s_n = 1)}$$

where $P(s_n = 1) = \pi_1$ and $P(s_n = 2) = \pi_2$.

Then, we have

$$\lim_{n\to\infty} P(s_n = 1 \mid Y_n = 1) = \frac{P(Y_n = 1 \mid s_n = 1)\lim_{n\to\infty} P(s_n = 1)}{\lim_{n\to\infty} P(Y_n = 1)} = \frac{.7\frac{6}{13}}{.7\frac{6}{13} + .3\frac{7}{13}}$$
$$= \frac{4.2}{6.3} = 0.666$$

-1

Calculate separately using Total probability Theory. (See Solution)

Problem 2. We have the following facts

1. Let $\mathcal{S}$ be a bounded set of real numbers, i.e., $\exists D < \infty$ such that $|x| \le D$ for all $x \in \mathcal{S}$. Then there exists $\bar{D} < \infty$ such that
   - $x \le \bar{D}$ for all $x \in \mathcal{S}$
   - Given any $\epsilon > 0$ there exists $y \in \mathcal{S}$ s.t. $y \ge \bar{D} - \epsilon$

   In other words, $\bar{D}$ is the least upper bound or supremum of $\mathcal{S}$. Similar, there exists greatest lower bound or infimum of $\mathcal{S}$.

2. Consider an infinite sequence of real numbers $\{x_n\}_{n=1}^{\infty}$. Then there is a monotone subsequence of $\{x_n\}_{n=1}^{\infty}$, i.e., there exists $\{x_{n_1}, x_{n_2}, \ldots\}, n_1 \le n_2 \le \ldots$, that is either non-decreasing or non-increasing.

Questions:

1. Let $\{x_n\}_{n=1}^{\infty}$ be a non-decreasing upper bounded sequence of real numbers. Show that $\lim_{n\to\infty} x_n$ exists and finite.

For the non-decreasing upper-bounded sequence we have a least upper bound (LUB) by the fact mentioned in the problem, which we denote by u. Then for $\epsilon > 0$, we know that $(u-\epsilon)$ can not be an upper bound of the sequence, since $(u-\epsilon) < u$ being an LUB would contradict u being an LUB. Therefore, there exists some N, for which $u - \epsilon < x_n$. Since u is the LUB, then $x_n \le u;\ \forall n$.

As it is a non-decreasing sequence, then for $n \ge N$ we have $x_n \ge x_N$, and consequently $(u-\epsilon) < x_N \le x_n \le u < (u+\epsilon)$. Then,

$$\begin{aligned}
&\Leftrightarrow \text{if } n \ge N \text{ then } (u-\epsilon) < x_n < (u+\epsilon) \\
&\Leftrightarrow \text{if } n \ge N \text{ then } -\epsilon < x_n < +\epsilon \\
&\Leftrightarrow \text{if } n \ge N \text{ then } |x_n - u| < \epsilon
\end{aligned}$$

Therefore by the definition of limit we have $\lim_{n\to\infty}\{x_n\} = u$. Without loss of generality we can prove this for the lower monotone lower-bounded case as well.

2. Let $\{x_n\}_{n=1}^{\infty}$ be a bounded sequence, i.e., $\exists M < \infty$ such that $|x_n| \le M$. Show that $\{x_n\}_{n=1}^{\infty}$ has a convergent subsequent.

There are multiple approaches for proving this theorem, including repeated bisection. This can also be proven following the steps below. We say that m is a peak of the sequence $\{x_n\}_{n=1}^{\infty}$, if $n > m$ implies $x_n < x_m$, that is all subsequent terms of the peak point are of smaller value. Now, we would have either of these two cases:

1) $\{x_n\}_{n=1}^{\infty}$ has **infinite** number of peaks $k_0 < k_1 < k_2 < \cdots < k_n < \cdots$, then $\{x_{k_n}\}$ is strictly decreasing and monotone.

2) $\{x_n\}_{n=1}^{\infty}$ has **finite** number of peaks. Then for some N we have x_N as the **last** peak. Then $k_0 := N+1$ is **not** a peak, and there exists $k_1 > k_0$ such that $a_{k_1} > a_{k_0}$. Similarly, by repetition $\exists k_{n+1} > k_n$ such that $a_{k_{n+1}} > a_{k_n}$. Then $\{x_{kn}\}_{n=1}^{\infty}$ would be increasing and monotone.

In both cases we have monotone subsequences and as stated bounded, then via the first part we have proven that bounded sequences have a convergent subsequence. This property basically states that no matter how 'random' a sequence x_n may be, as long as it is bounded, then some part of it must converge.

3. A sequence of real numbers $\{x_n\}_{n=1}^{\infty}$ is Cauchy if given any $\epsilon > 0$ there exists N_ϵ(N depends on ϵ) s.t. $|x_n - x_m| < \epsilon$ for all $n, m > N_\epsilon$. Show that every Cauchy sequence is bounded.

By the triangle inequality we have $|x_n| - |x_m| \le |x_n - x_m|$. Then we could set $m := N_\epsilon + 1$ and have $|x_n| - |x_{N_\epsilon+1}| < \epsilon$ for $\forall n > N_\epsilon$. With some rearrangement we have $|x_n| < \epsilon + |x_{N_\epsilon+1}|$. Then $|x_n| \le \max\{|x_0|, |x_1|, \cdots, |x_N|, |x_{N+1}|, \epsilon + |x_{N+1}|\}$. Therefore, x_n is bounded within $\pm\max\{|x_0|, |x_1|, \cdots, |x_{N+1}|, \epsilon + |x_{N+1}|\}$.

4. Show that if a Cauchy sequence of real numbers $\{x_n\}_{n=1}^{\infty}$ has a convergent sub sequence, then the sequence $\{x_n\}_{n=1}^{\infty}$ must converge.

Assume the subsequence x_{n_k} converges to L. Then for $\varepsilon > 0$,

$$\begin{aligned} \exists N_1 \text{ such that } \quad & r \geqslant N_1 \implies \left|a_{n_r} - L\right| < \varepsilon/2 \\ \exists N_2 \text{ such that } \quad & m, n \geqslant N_2 \implies |a_m - a_n| < \varepsilon/2 \end{aligned}$$

Put $s := \min\{r \mid n_r \geqslant N_2\}$ and put $N = n_s$.Then

$$\begin{aligned} m, n \geqslant N \quad & \implies |a_m - a_n| \\ & \leqslant \left|a_m - a_{n_s}\right| + \left|a_{n_s} - l\right| \\ & < \quad \varepsilon/2 \quad + \quad \varepsilon/2 \quad = \varepsilon \end{aligned}$$

5. Show that every Cauchy sequence of real numbers is convergent.

As already shown in **part 3**, every Cauchy sequence is bounded. Also from **part 2**, we know that the bounded sequence would have a convergent subsequence. At last, per **part 4**, we know that if its has a convergent subsequnce, then the Cauchy sequence must converge.

Problem 3.
Recall the definition of an MDP from the second lecture. Let $\mathcal{S} = \{s_1, \ldots, s_n\}$ be an MC with transition probability P. X is called a controlled MC if $\mathbf{P}$ can be controlled, i.e., $\mathbf{P} = \left[P_{ij}(a)\right]$ where a is a control action. At time k, the state is $s_k \in \mathcal{X}$, we take an action $a_k = \mu_k(s_k)$, and it incurs a (bounded) cost $r(s_k, a_k)$, where w.l.o.g we assume $c \geq 0$. Here μ_k is a mapping from state to action. The goal is to choose $\{a_k\}$ to maximize

$$V_\pi(i) = \lim_{N\to\infty} \mathbb{E}\left[\sum_{k=0}^{N} \gamma^k r(s_k, a_k) \mid s_0 = i\right]$$

where $\gamma \in (0,1)$ is called the discount factor and $\pi = \left[\mu_0, \mu_1, \ldots\right]$ is the policy. When μ_k does not depend on time k, i.e., $\mu_k = \mu$, we call the policy is stationary and with some abuse of notation denote it as μ.
Policy evaluation Let consider a subproblem, where we want to estimate the vector value function V_μ for a given stationary policy μ. We know from class that V_μ satisfies the so-called Bellman equation

$$V_\mu(i) = \mathbb{E}[r(i, \mu(i))] + \gamma \sum_j P_{ij}(\mu(i)) V_\mu(j)$$

or in vector form

$$V_\mu = \mathbb{E}[r] + \gamma \mathbf{P}_\mu V_\mu$$

where $r = [r(i, \mu(i))]$ is a vector. In class, we have a theorem to show the existence and uniqueness of the solution of this Bellman equation. We mentioned that there are two ways to do it: using *algebra* or the classic *fixed point* theorems.
Questions of the algebra proof:

1. Consider matrix norm induced by the vector norm defined in class, i.e.,

$$\|\mathbf{P}\|_p = \max_{\|y\|_p=1} \|\mathbf{P}y\|_p$$

Let λ_i be the eigenvalues of $\mathbf{P}$. Show that

$$\max_i |\lambda_i| \leq \|\mathbf{P}\|_p, \quad \forall p \geq 1$$

Hint: Using the definition of the eigenvalues of a matrix.

The left hand-side of the inequality, that is the largest absolute value of the eigenvalues, refers to the *spectral radius* of the matrix which is often denote by $\rho(.)$,

$$\rho(A) = \max_i |\lambda_i|$$

We can see that for any induced matrix norm, if $(\tilde{\lambda}, \tilde{\mathbf{x}})$ is the eigenvalue-eigenvector pair which maximizes $|\lambda|$ for P, with $\tilde{\mathbf{x}}$ normalized to satisfy $\|\tilde{\mathbf{x}}\| = 1$, then

$$\rho(P) = |\tilde{\lambda}| = \|\tilde{\lambda}\tilde{\mathbf{x}}\| = \|P\tilde{\mathbf{x}}\| \leq \|P\|$$

So, any induced norm is always bounded below by the spectral radius.

Questions of the fixed point theorem proof:

1. Let T be a continuous mapping from $\mathcal{S} \to \mathcal{S}$ where $\mathcal{S}$ is a closed set. Suppose that T satisfies a contraction property, i.e., $\exists \gamma \in (0,1)$ such that

$$\|T(x) - T(y)\| \leq \gamma \|x - y\|$$

$\|\cdot\|$ can be any norm. Show that
(a) There **exists** a **unique** x^* s.t. $T(x^*) = x^*$
(b) The fixed point iteration starting with x_0

$$x_{k+1} = T(x_k)$$

converges to x^*, i.e., $\lim_{k\to\infty} x_k = x^*$.

We define $T^k(x)$ as the kth composition of T with itself.
Consider any $x_0 \in \mathcal{S}$ that is **not** a fixed point. We define the sequence $\{x_k\}$ by $x_k = T(x_{k-1})$ for all $k \in \mathbb{N}$. Then we will show that $\{x_k\}$ is Cauchy and therefore it is convergent. For all $k \geq 1$, using the contraction property, we have:

$$\begin{aligned}
\|x_{k+1} - x_k\| &= \|T(x_k) - T(x_{k-1})\| \\
&\leq \gamma \|x_k - x_{k-1}\| \\
&\vdots \\
&\leq \gamma^k \|x_1 - x_0\|
\end{aligned}$$

Now for any $m, n \in \mathbb{N}$ such that $m > n+1$, via the triangle inequality, we have

$$\begin{aligned}
\|x_m - x_n\| &\leq \|x_m - x_{m-1}\| + \|x_{m-1} - x_n\| \\
&\leq \|x_m - x_{m-1}\| + \|x_{m-1} - x_{m-2}\| + \|x_{m-2} - x_n\| \\
&\vdots \\
&\leq \|x_m - x_{m-1}\| + \ldots + \|x_{n+1} - x_n\|
\end{aligned}$$

Then with the help of the first inequality we obtain

$$\begin{aligned}
\|x_m - x_n\| &\leq (\gamma^{m-1} + \gamma^m + \ldots + \gamma^n \| \|x_1 - x_0\| \\
&= \gamma^n (1 + \gamma^2 + \ldots + \gamma^{m-n-1} \| \|x_1 - x_0\| \\
&< \gamma^1 (1 + \gamma^2 + \ldots) \|x_1 - x_0\| \\
&= \frac{\gamma^n}{1-\gamma} \|x_1 - x_0\|
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \|x_m - x_n\| = 0 \quad \forall m > n+1$$

So, $\{x_k\}$ is in fact Cauchy. Since $\mathcal{S}$ is closed, there exists $x^* \in \mathcal{S}$ such that $\lim_{k\to\infty} x_k = x^*$
Hence for any $\epsilon > 0$, there exists $m \in \mathbb{N}$ such that for all $n \geq m$,

$$\|x_n - x^*\| < \epsilon/2$$

Then, via the triangle inequality and the contraction property, we have

$$\begin{aligned}
\|T(x^*) - x^*\| &\leq \|T(x^*) - x_{n+1}\| + \|x_{n+1} - x^*\| \\
&= \|T(x^*) - T(x_n)\| + \|x_{n+1} - x^*\| \\
&\leq \gamma\|x^* - x_n\| + \|x_{n+1} - x^*\| \\
&< \epsilon
\end{aligned}$$

Hence $\|T(x^*) - x^*\| = 0$ and x^* is a unique fixed point of T, with

$$\lim_{n\to\infty} T^n(x) = x^*$$

for all $x \in \mathcal{S}$.
If there was another $\hat{x} \neq x^*$ such that $T(\hat{x}) = \hat{x}$, we would have a contradiction since

$$\begin{aligned}
0 < \|x^* - \hat{x}\| &= \|T(x^*) - T(\hat{x})\| \\
&\leq \gamma\|x^* - \hat{x}\| < \|x^* - \hat{x}\|
\end{aligned}$$

2. Next let T be the right-hand side of the Bellman equation, i.e. for all i

$$\left(TV_\mu\right)(i) = \mathbb{E}[r(i,\mu(i))] + \gamma \sum_{j\in\mathcal{S}} P_{ij}(\mu(i))V_\mu(j)$$

(a) Given any V, V' such that $V(i) \leq V'(i)$ for all i. Show that the following Monotonicity property holds

$$(TV)(i) \leq \left(TV'\right)(i)$$

$$\begin{aligned}
(TV)(i) &= \mathbb{E}[r(i,\mu(i))] + \gamma \sum_{j\in\mathcal{S}} P_{ij}(\mu(i))V(j) \\
&\leq \mathbb{E}[r(i,\mu(i))] + \gamma \sum_{j\in\mathcal{S}} P_{ij}(\mu(i))V'(j) = \left(TV'\right)(i)
\end{aligned}$$

(b) Let q be a scalar. Show that

$$(T(V+q))(i) = (TV)(i) + \gamma q$$

$$\begin{aligned}
\left(T(V+q)\right)(i) &= \mathbb{E}[r(i,\mu(i))] + \gamma \sum_{j\in\mathcal{S}} P_{ij}(\mu(i))(V+q)(j) \\
&= \mathbb{E}[r(i,\mu(i))] + \gamma \sum_{j\in\mathcal{S}} P_{ij}(\mu(i))V(j) + \gamma q \sum_{j\in\mathcal{S}} P_{ij}(\mu(i)) \\
&= (TV)(i) + \gamma q
\end{aligned}$$

(c) Using the two properties above show that T is contractive under maximum-norm, i.e., for all V, V'

$$\left\|TV - TV'\right\|_\infty \leq \gamma\left\|V - V'\right\|_\infty$$

Hint: Note that here the contraction only holds for the maximum norm. Then the first step is to consider $d = \max_i \left|V(i) - V'(i)\right|$. Recall that we consider finite-time MC, i.e., the set of states i is finite. Thus, d is well-defined.

To prove this notice that

$$\left|\max_a f(a) - \max_a g(a)\right| \le \max_a |f(a) - g(a)|$$

Then,

$$\begin{aligned}
\left\| TV(i) - TV'(i) \right\|_\infty &= |\max_i \left[r(i,\mu(i)) + \gamma \sum_{j\in\mathcal{S}} P_{ij}(\mu(i)) V(j) \right] \\
&\qquad - \max_{i'} \left[r(i,\mu(i)) + \gamma \sum_{j\in\mathcal{S}} P_{ij}(\mu(i')) V'(j) \right] | \\
&\le \max_i \left| \left[r(i,\mu(i)) + \sum_{j\in\mathcal{S}} P_{ij}(\mu(i)) V(j) \right] - \left[r(i,\mu(i)) + \sum_{j\in\mathcal{S}} P_{ij}(\mu(i)) V'(j) \right] \right| \\
&\le \gamma \max_i \sum_{j\in\mathcal{S}} P_{ij}(\mu(i)) \| V(j) - V'(j) \| \\
&\le \gamma \left\| V - V' \right\|_\infty \max_i \sum_{j\in\mathcal{S}} P_{ij}(\mu(i)) = \gamma \left\| V - V' \right\|_\infty
\end{aligned}$$