Andrew ElHabr
ISyE 6664
9/29/17

Homework 4

1. Using value iteration and setting $v^0 = (0,0,0)$ and $\epsilon = 0.001$, we find that the epsilon-optimal policy is $\pi^*_\epsilon : d(s_1) = a_{1,1}, d(s_2) = a_{2,2}, d(s_3) = a_{3,1}$ after 45 iterations. The algorithm was coded in R, and the output for v for selected iterations and the final output for d_ϵ are shown in the following:

V

	[,1]	[,2]	[,3]
[1,]	0.000000	4.00000	4.000000
[2,]	1.600000	7.20000	4.000000
[3,]	3.520000	7.20000	5.280000
⋮			
[18,]	8.191894	12.37810	10.517537
[19,]	8.227997	12.41403	10.553515
[20,]	8.256811	12.44281	10.582398
⋮			
[43,]	8.371412	12.55746	10.696994
[44,]	8.371549	12.55760	10.697130
[45,]	8.371657	12.55770	10.697239

d_epsilon

	[,1]
[1,]	1
[2,]	2
[3,]	1

2. Let u and v be such that $\mathcal{L}v(s) \geq \mathcal{L}u(s)$ for $s \in S$. Then

$$\sup_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)v(j)\} \geq \sup_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)u(j)\}.$$

Now, since we are not guaranteed that $a_s^* \in \arg\max_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)v(j)\}$ exists, let us consider an "epsilon-optimal action," a_s^ϵ, in which $a_s^\epsilon \in \arg\max_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)v(j) - \epsilon\}$ for some $\epsilon > 0$. The existence of a_s^ϵ is shown by the following:

$$\begin{aligned}
&\sup_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)v(j)\} \geq \sup_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)u(j)\} \\
\implies &\sup_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)v(j)\} - \epsilon \geq \sup_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)u(j)\} - \epsilon \\
\implies &\max_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)v(j) - \epsilon\} \geq \max_{a \in A_s} \{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)u(j) - \epsilon\}.
\end{aligned}$$

Note that

$$\mathcal{L}v(s) = \max_{a \in A_s}\{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)v(j) - \epsilon\} + \epsilon, \text{ and}$$
$$\mathcal{L}u(s) = \max_{a \in A_s}\{r(s,a) + \lambda \sum_{j \in S} p(j|s,a)u(j) - \epsilon\} + \epsilon.$$

Then,

$$\begin{aligned}
0 \leq \mathcal{L}v(s) - \mathcal{L}u(s) \leq & r(s,a_s^\epsilon) + \lambda \sum_{j \in S} p(j|s,a_s^\epsilon)v(j) + \epsilon - \Big(r(s,a_s^\epsilon) + \lambda \sum_{j \in S} p(j|s,a_s^\epsilon)u(j) + \epsilon\Big) \\
= & \lambda \sum_{j \in S} p(j|s,a_s^\epsilon)[v(j) - u(j)] \\
\leq & \lambda \sum_{j \in S} p(j|s,a_s^\epsilon)||v - u|| \\
= & \lambda||v - u||. \qquad\qquad \Big(\text{since } \sum_{j \in S} p(j|s,a_s^\epsilon) = 1\Big)
\end{aligned}$$

We can similarly show that $0 \leq \mathcal{L}u(s) - \mathcal{L}v(s) \leq \lambda||v-u||$. Thus, we have that $0 \leq |\mathcal{L}v(s) - \mathcal{L}u(s)| \leq \lambda||v-u||$. Finally, when taking the supremum of this last inequality over all $s \in S$, we have that $||\mathcal{L}v(s) - \mathcal{L}u(s)|| \leq \lambda||v-u||$, which proves that $\mathcal{L}$ is a contraction mapping by definition.

3. We just need to carry out one iteration of policy iteration in which you set your initial decision rule d_0 to d^∞ for which you already have the policy evaluation of $v_\lambda^* = (I - \lambda P_{d^\infty})^{-1} r_{d^\infty}$, which is usually the most difficult part of the policy iteration algorithm to compute. Then, you must complete the policy increment step, i.e. $d' \in \arg\max_d\{r_d + \lambda P_d v_\lambda^*\}$. If $d' = d^\infty$, then d^∞ is still optimal; otherwise, d^∞ is no longer optimal. (Note that this is essentially the same as determining if

$$r(s',a') + \lambda \sum_{j \in S} p(j|s',a')v_\lambda^*(j) \leq v_\lambda^*(s')$$

or not. If the above inequality is found to be true, then d^∞ is still optimal; otherwise, d^∞ is no longer optimal.)