CSE 250A: Assignment 7

Jiaxu Zhu A53094655

December 8, 2015

7.1 Policy improvement

(a) The state-value function is shown in Table 1.

s	$\pi(s)$	$V^{\pi}(s)$
0	0	-1.5
1	0	7.5

Table 1: The state-value function

(b) The greedy policy $\pi'(s)$ with respect to the state-value function $V^{\pi}(s)$ is shown in Table 2.

s	$\pi(s)$	$\pi'(s)$
0	0	1
1	0	0

Table 2: The greedy policy

7.2 Value and policy iteration

(a) The optimal state values are shown in Table. 3, according to the map.

(b) The optimal policies are shown in Table. 4, according to the map.

(c) The optimal policies are shown in Table. 5, according to the map, which agree with the results from part (b).

0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	72.98	73.80	74.63	0.00	-100.00	59.67	-100.00	0.00
71.39	72.17	0.00	75.47	74.40	64.89	68.95	59.67	0.00
0.00	0.00	77.20	76.34	0.00	-100.00	70.31	-100.00	0.00
0.00	0.00	78.07	0.00	0.00	0.00	80.33	0.00	0.00
0.00	79.83	78.94	0.00	0.00	-100.00	81.47	-100.00	0.00
0.00	80.72	0.00	84.41	85.36	86.31	92.20	93.67	100.00
0.00	81.63	82.55	83.47	0.00	90.52	91.62	92.64	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 3: Optimal State Value

*	*	*	*	*	*	*	*	*
*	EAST	EAST	SOUTH	*	*	SOUTH	*	*
EAST	NORTH	*	SOUTH	WEST	WEST	SOUTH	WEST	*
*	*	SOUTH	WEST	*	*	SOUTH	*	*
*	*	SOUTH	*	*	*	SOUTH	*	*
*	SOUTH	WEST	*	*	*	SOUTH	*	*
*	SOUTH	*	EAST	EAST	EAST	EAST	EAST	WEST
*	EAST	EAST	NORTH	*	EAST	EAST	NORTH	*
*	*	*	*	*	*	*	*	*

Table 4: Optimal State Value

*	*	*	*	*	*	*	*	*
*	EAST	EAST	SOUTH	*	*	SOUTH	*	*
EAST	NORTH	*	SOUTH	WEST	WEST	SOUTH	WEST	*
*	*	SOUTH	WEST	*	*	SOUTH	*	*
*	*	SOUTH	*	*	*	SOUTH	*	*
*	SOUTH	WEST	*	*	*	SOUTH	*	*
*	SOUTH	*	EAST	EAST	EAST	EAST	EAST	WEST
*	EAST	EAST	NORTH	*	EAST	EAST	NORTH	*
*	*	*	*	*	*	*	*	*

Table 5: Optimal State Value Using Policy Iteration

7.3 Effective horizon time

$$\begin{aligned}
\sum_{n\geq t}\gamma^n r_n &\leq \sum_{n\geq t}\gamma^n \\
&= \frac{\gamma^n(1-\gamma^\infty)}{1-\gamma} \\
&\leq \frac{\gamma^n}{1-\gamma} \\
&\leq \frac{e^{n(\gamma-1)}}{1-\gamma} \\
&= he^{-t/h}
\end{aligned}$$

7.4 Convergence of iterative policy evaluation

$$\begin{aligned}
\Delta_{k+1} &= \max_s |V_{k+1}(s) - V^\pi(s)| \\
&= \max_s |R(s) + \gamma\sum_{s'} P(s'|s,\pi(s))V_k(s') - R(s) - \gamma\sum_{s'} P(s'|s,\pi(s))V^\pi(s')| \\
&= \max_s |\gamma\sum_{s'} P(s'|s,\pi(s))(V_k(s') - V^\pi(s'))| \\
&= \max_s \gamma\sum_{s'} P(s'|s,\pi(s))|(V_k(s') - V^\pi(s'))| \\
&< \max_s \gamma\sum_{s'} P(s'|s,\pi(s))\Delta_k \\
&= \gamma\Delta_k \\
&< \gamma^k\Delta_1
\end{aligned}$$

It means that the error Δ_k decays exponentially fast in the number of iterations.

7.5 Value function for a random walk

(a)

$$\begin{aligned} V^{\pi}(s) &= R(s) + \gamma \sum_{s'=0}^{\infty} P(s'|s, \pi(s)) V^{\pi}(s') \\ &= R(s) + \gamma \sum_{s'=s}^{s+1} P(s'|s, \pi(s)) V^{\pi}(s') \end{aligned}$$

(b)

$$\begin{aligned} V^{\pi}(s) &= R(s) + \gamma \sum_{s'=s}^{s+1} P(s'|s, \pi(s)) V^{\pi}(s') \\ as + b &= s + \gamma[\frac{3}{4}(as + b) + \frac{1}{4}(as + a + b)] \\ ((1-\gamma)a - 1)s &= \frac{1}{4}\gamma a - (1-\gamma)b \end{aligned}$$

To make this equation satisfy for $s \in 0, 1, 2, ..., \infty$

$$\begin{aligned} a &= \frac{1}{1-\gamma} \\ b &= \frac{\gamma}{4(1-\gamma)^2} \end{aligned}$$

7.6 Value function for a random walk

(a)

$$\begin{aligned} V^{\pi}(s) &= R(s) + \gamma \sum_{s'=1}^{n} P(s'|s, \pi(s)) V^{\pi}(s') \\ &= R(s) + \gamma \sum_{s'=1}^{n} P(s'|s, 1) V^{\pi}(s') \\ &= R(s) + \gamma V^{\pi}(s) \\ V^{\pi}(s) &= \frac{R(s)}{1-\gamma} \end{aligned}$$

(b)

$$\begin{aligned} V^{\pi}(s) &= R(s) + \gamma \sum_{s'=1}^{n} P(s'|s, \pi(s)) V^{\pi}(s') \\ &= R(s) + \frac{\gamma}{n} \sum_{s'=1}^{n} V^{\pi}(s') \end{aligned}$$

(c)

$$\begin{aligned} v &= \frac{1}{n}\sum_s V^\pi(s) \\ &= \frac{1}{n}[\sum_{s\in S_0} V^\pi(s) + \sum_{s\in S_1} V^\pi(s)] \\ &= \frac{1}{n}\sum_{s\in S_0} R(s) + \frac{1}{n}\sum_{s\in S_0}\frac{\gamma}{n}\sum_{s'=1}^{n} V^\pi(s') + \frac{1}{n}\sum_{s\in S_1}\frac{R(s)}{1-\gamma}\} \\ &= -r + v\mu\gamma + \frac{r}{1-\gamma} \\ v &= \frac{\gamma r}{(1-\gamma)(1-\gamma\mu)} \end{aligned}$$

(d)

$$\begin{aligned} V^\pi(s) &= R(s) + \frac{\gamma}{n}\sum_{s'=1}^{n} V^\pi(s') \\ &= R(s) + \gamma v \\ &= R(s) + \frac{\gamma r}{(1-\gamma)(1-\gamma\mu)} \end{aligned}$$

(e)

$$\begin{aligned} \pi^\star(s) &= \arg\max_a Q^\star(s,a) \\ &= \arg\max_a [\sum_{s'=1}^{n} P(s'|s,a)V^\star(s')] \end{aligned}$$

For $\pi^\star(s) = 1$

$$\begin{aligned} \sum_{s'=1}^{n} P(s'|s,1)V^\star(s') &> \sum_{s'=1}^{n} P(s'|s,0)V^\star(s') \\ V(s) &> v^\star \end{aligned}$$

For $\pi^\star(s) = 0$

$$\begin{aligned} \sum_{s'=1}^{n} P(s'|s,0)V^\star(s') &\geq \sum_{s'=1}^{n} P(s'|s,1)V^\star(s') \\ V(s) &\leq v^\star \end{aligned}$$

Thus $\theta = v^\star$

7.7 Stochastic approximation

(a)

$$\begin{aligned} \sum_{k=1}^{\infty}\alpha_k &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + ... + \frac{1}{8} + ... \\ &> 1 + \frac{1}{2} + (\frac{1}{4} + \frac{1}{4}) + (\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}) + ... \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + ... \\ &= \infty \end{aligned}$$

$$\begin{aligned}
\sum_{k=1}^{\infty} \alpha_k^2 &= 1 + \frac{1}{2\times 2} + \frac{1}{3\times 3} + \frac{1}{4\times 4} + ... \\
&< 1 + \frac{1}{1\times 2} + + \frac{1}{2\times 3} + + \frac{1}{3\times 4} + ... \\
&= 1 + (1 - \frac{1}{2}) + (\frac{1}{2} - \frac{1}{3}) + (\frac{1}{3} - \frac{1}{4}) + ... \\
&< 2
\end{aligned}$$

(b) Suppose $\mu_k = (1/k)(x_1 + x_2 + ... + x_k)$ is true, then

$$\begin{aligned}
\mu_{k+1} &= \mu_k + \frac{1}{k+1}(x_{k+1} - \mu_k) \\
&= (1/k)(x_1 + x_2 + ... + x_k) + \frac{1}{k+1}[x_{k+1} - (1/k)(x_1 + x_2 + ... + x_k)] \\
&= (\frac{1}{k} - \frac{1}{k(k+1)})(x_1 + x_2 + ... + x_k) + \frac{1}{k+1}x_{k+1} \\
&= \frac{1}{k+1}(x_1 + x_2 + ... + x_k + x_{k+1})
\end{aligned}$$

And we have $\mu_1 = x_1$