Exercise 3

Kai Schneider

May 10, 2021

Task 1 Proofs

a.)

given $(\mathscr{T}\upsilon)(s) = \max_a \sum_{s',r} p(s',r|s,a)[r+\gamma\upsilon(s')]$ show that $\mathscr{T}$ is a γ-contraction:

To prove: $||\mathscr{T}\upsilon - \mathscr{T}\upsilon'|| \leqslant \gamma||\upsilon - \upsilon'||$

$$
\begin{aligned}
||\mathscr{T}\upsilon - \mathscr{T}\upsilon'|| &= ||\max_a \sum_{s',r} p(s',r|s,a)[r+\gamma\upsilon] - \max_a \sum_{s',r} p(s',r|s,a)[r+\gamma\upsilon']|| \\
&= ||\max_a \left[\sum_{s',r} p(s',r|s,a)[r+\gamma\upsilon] - \sum_{s',r} p(s',r|s,a)[r+\gamma\upsilon']\right]|| \\
&= ||\max_a \sum_{s',r} p(s',r|s,a)\left[[r+\gamma\upsilon] - [r+\gamma\upsilon']\right]|| \\
&= ||\max_a \gamma \sum_{s',r} p(s',r|s,a)[\upsilon - \upsilon']|| \\
&= \gamma||\underbrace{\max_a \sum_{s',r} p(s',r|s,a)[\upsilon - \upsilon']||}_{0\leq\star\leq 1} \\
&\leq \gamma||\upsilon - \upsilon'||
\end{aligned}
$$

□

b.)

finite $MDP(S,A,R,p,\gamma)$, $r \in [r_{min}, r_{max}] \;\; \forall r \in R$

I:

prove $\frac{r_{min}}{1-\gamma} \leq \upsilon(s) \leq \frac{r_{max}}{1-\gamma}$

$$\text{with:} \quad \upsilon = \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i R_{t+i+1} | S_t = s\right]$$

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i r_{min} | S_t = s\right] \leq \upsilon \leq \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i r_{max} | S_t = s\right]$$

$$\mathbb{E}\left[r_{min} \sum_{i=0}^{\infty} \gamma^i | S_t = s\right] \leq \upsilon \leq \mathbb{E}\left[r_{max} \sum_{i=0}^{\infty} \gamma^i | S_t = s\right]$$

$$\mathbb{E}\left[\frac{r_{min}}{1-\gamma} | S_t = s\right] \leq \upsilon \leq \mathbb{E}\left[\frac{r_{max}}{1-\gamma} | S_t = s\right] \quad \text{for } 0 \leq \gamma \leq 1$$

$$\Leftrightarrow \frac{r_{min}}{1-\gamma} \leq \upsilon \leq \frac{r_{max}}{1-\gamma}$$

□

II:

prove $|\upsilon(s) - \upsilon(s')| \leq \frac{r_{max} - r_{min}}{1-\gamma}$

$$\text{again with:} \quad \upsilon = \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i R_{t+i+1} | S_t = s\right]$$

$$|\upsilon(s) - \upsilon(s')| \leq \left| \mathbb{E}\left[\frac{r_{max}}{1-\gamma} | S_t = s\right] - \mathbb{E}\left[\frac{r_{min}}{1-\gamma} | S_t = s'\right] \right| \quad \text{see (I)}$$

$$|\upsilon(s) - \upsilon(s')| \leq \left| \frac{r_{max}}{1-\gamma} - \frac{r_{min}}{1-\gamma} \right| = \frac{r_{max} - r_{min}}{1-\gamma}$$

□

Task 2 Programming

a.)

The algorithm converges in 43 steps with the folling optimal value function:

0.01543432	0.01559069	0.02744009	0.01568004
0.02685371	0.0	0.05978021	0.0
0.0584134	0.13378315	0.1967357	0.0
0.0	0.2465377	0.54419553	0.0

b.)

The optimal policy is:
(2, 3, 2, 3, 0, 0, 0, 0, 3, 1, 0, 0, 0, 2, 1, 0)