Assignment 3 - Dynamic Programming

Jonathan Stumber
Christian Reiser
Magnus Ostertag

1. Proofs

1.1 a) Show that the Bellman optimality operator $\mathcal{T}$ is a γ-contraction.

Let $v, v' \in \mathcal{V}$ be two value functions in the value function space $\mathcal{V}$ with norm $||\cdot||_\infty$. We get:

$$
\begin{aligned}
||\mathcal{T}v - \mathcal{T}v'||_\infty &= \max_s \left|\left(\mathcal{T}v\right)(s) - \left(\mathcal{T}v'\right)(s)\right| \\
&= \max_s \left| \max_a \underbrace{\left[\sum_{s',r} p(s',r|s,a)\left[r + \gamma v(s')\right]\right]}_{:=g(a)} - \max_a \underbrace{\left[\sum_{s',r} p(s',r|s,a)\left[r + \gamma v'(s')\right]\right]}_{:=g'(a)} \right|
\end{aligned}
$$

Let's consider the inner part for a fix s.
WLOG we assume

$$\max_a g(a) \geq \max_a g'(a)\,.$$

Let

$$
\begin{aligned}
a_1 &:= \arg\max_a g(a) \\
a_2 &:= \arg\max_a g'(a)\,.
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow\ 0 &\leq |\max_a g(a) - \max_a g'(a)| \\
&= g(a_1) - g'(a_2) \\
&= g(a_1) - g'(a_1) + \underbrace{g'(a_1) - g'(a_2)}_{\leq 0} \\
&\leq g(a_1) - g'(a_1)
\end{aligned}
$$

The back term is ≤ 0 because $a_2 = \arg\max_a g'(a)$. So we can make the whole sum larger by leaving it out.

$$\begin{aligned}
&g(a_1) - g'(a_1) \\
&= \sum_{s',r} p(s', r|s, a_1) \left[(r + \gamma v(s')) - (r + \gamma v'(s'))\right] \\
&= \sum_{s',r} p(s', r|s, a_1)\gamma \underbrace{\left[v(s') - v'(s')\right]}_{\leq \max_s |v(s) - v'(s)|} \\
&\leq \gamma \max_s |v(s) - v'(s)| \underbrace{\sum_{s',r} p(s', r|s, a_1)}_{=1} \\
&= \gamma \max_s |v(s) - v'(s)|
\end{aligned}$$

This is independent from s.
Hence,

$$||\mathcal{T}v - \mathcal{T}v'||_\infty \leq \gamma ||v - v'||_\infty$$

and $\mathcal{T}$ is a γ-contraction.

$\square$

1.2 b) Assuming a general finite MDP (S, A, R, p, γ) where rewards are bounded: $r \in [r_{\mathbf{min}}, r_{\mathbf{max}}]$ for all $r \in \mathbb{R}$. Prove the following equations.

$$\frac{r_{\min}}{1-\gamma} \leq v(s) \leq \frac{r_{\max}}{1-\gamma} \tag{1}$$

$$|v(s) - v(s')| \leq \frac{r_{\max} - r_{\min}}{1-\gamma} \tag{2}$$

Using the definition of $v(s)$ with the geometric series we get:

$$
\begin{aligned}
v(s) &= \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i R_{t+i+1} \middle| S_t = s\right] \\
&\leq \mathbb{E}\left[\underbrace{\sum_{i=0}^{\infty} \gamma^i r_{\max}}_{=const.} \middle| S_t = s\right] \\
&= \sum_{i=0}^{\infty} \gamma^i r_{\max} \ \underbrace{\mathbb{E}\left[1 \middle| S_t = s\right]}_{=1} \\
&= r_{\max} \underbrace{\sum_{i=0}^{\infty} \gamma^i}_{=\frac{1}{1-\gamma}} \\
&= \frac{r_{\max}}{1-\gamma} \ .
\end{aligned}
$$

And with the same argument,

$$
v(s) \geq \frac{r_{\min}}{1-\gamma} \ .
$$

This proves (1).
WLOG, we assume $v(s) \geq v(s')$.
Hence,

$$
\begin{aligned}
|v(s) - v(s')| &= v(s) - v(s') \\
&\overset{(1)}{\leq} \frac{r_{\max}}{1-\gamma} - \frac{r_{\min}}{1-\gamma} \\
&= \frac{r_{\max} - r_{\min}}{1-\gamma} \ .
\end{aligned}
$$

□

2. Value Iteration

2.1 a) Implement the value iteration algorithm.

- How many steps does it need to converge? – 77
- Optimal value function

0.015	0.016	0.027	0.016
0.027	0.000	0.060	0.000
0.058	0.134	0.197	0.000
0.000	0.247	0.544	0.000

2.2 b) Optimal policy

1 3 2 3
0 0 0 0
3 1 0 0
0 2 1 0