Felix Bühler - 2973140
Jan Leusmann - 2893121
Jamie Ullerich - 3141241

Exercise 3

Task 1 - Proofs

a) Show that: $\|\mathcal{T}^\pi v - \mathcal{T}^\pi v'\|_\infty \leq \gamma \|v - v'\|_\infty$

per definition:

$$\begin{aligned}
\|\mathcal{T}^\pi v - \mathcal{T}^\pi v'\|_\infty &= \max_s |(\mathcal{T}^\pi v)(s) - (\mathcal{T}^\pi v')(s)| \rightarrow \text{insert equation given on exercise sheet} \\
&= \max_s \left| \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)[r + \gamma v(s')] - \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)[r + \gamma v'(s')] \right| \\
&= \max_s \left| \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)[r + \gamma v(s') - r - \gamma v'(s')] \right| \\
&= \gamma \max_s \left| \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)[v(s') - v'(s')] \right| \\
&\leq \gamma \max_s \left| \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a) \max_s |[v(s') - v'(s')]| \right| \\
&= \gamma \max_s |v(s) - v(s')| \\
&= \gamma \|v - v'\|_\infty
\end{aligned}$$

b) Equation 1: $\frac{r_{min}}{1-\gamma} \leq v(s) \leq \frac{r_{max}}{1-\gamma}$
Since $v(s)$ is defined as: $v(s) = E[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ...|S_t = s]$, we see that this is a geometric series, which converges for $\gamma < 1$. Therefore, since the reward is bounded, the lower bound converges to $\frac{r_{min}}{1-\gamma}$. Analogue for the upper bound.

The second equation follows from this statement.

Task 2 - Value Iteration

a) It takes 46 steps to converge.

Optimal value function:

0.015	0.016	0.027	0.016
0.027	0.	0.06	0.
0.058	0.134	0.197	0.
0.	0.247	0.544	0.

b) The optimal policy is: [1 3 2 3 0 0 0 0 3 1 0 0 0 2 1 0]