Dynamic Programming

Exercise 1

(a) Proof: see slide 17 (out of 32) of lecture 3

(b) For every s and π:

$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ...$
$minG_t = \sum_{k=0}^{\infty} \gamma^k r_{min} < (G_t|s)$
$maxG_t = \sum_{k=0}^{\infty} \gamma^k r_{max} > (G_t|s)$

$$
\begin{aligned}
&\implies \mathbb{E}[minG_t] < v(s) < \mathbb{E}[maxG_t] \\
&\implies \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k r_{min}] < v(s) < \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k r_{max}] \\
&\implies \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k r_{min}] < v(s) < \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k r_{max}] \\
&\implies \sum_{k=0}^{\infty} \gamma^k r_{min} < v(s) < \sum_{k=0}^{\infty} \gamma^k r_{max} \\
&\implies r_{min} \sum_{k=0}^{\infty} \gamma^k < v(s) < r_{max} \sum_{k=0}^{\infty} \gamma^k
\end{aligned}
$$

$$
\begin{aligned}
\sum_{k=0}^{\infty} \gamma^k &= \lim_{k\to\infty} 1 + \gamma + \gamma^2 + ... + \gamma^k \\
&= \lim_{k\to\infty} \frac{(1-\gamma)(1+\gamma+\gamma^2+...+\gamma^k)}{1-\gamma} \\
&= \lim_{k\to\infty} \frac{1+\gamma-\gamma+\gamma^2-\gamma^2...-\gamma^{k+1}}{1-\gamma} \\
&= \lim_{k\to\infty} \frac{1-\gamma^{k+1}}{1-\gamma} \\
&= \frac{1}{1-\gamma}
\end{aligned}
$$

$$\implies \frac{r_{min}}{1-\gamma} < v(s) < \frac{r_{max}}{1-\gamma}$$

For every $v(s)$ and $v(s')$:

$$v(s) \in [\frac{r_{min}}{1-\gamma}; \frac{r_{max}}{1-\gamma}]$$
$$v(s') \in [\frac{r_{min}}{1-\gamma}; \frac{r_{max}}{1-\gamma}]$$
$$\implies |v(s) - v(s')| < |\frac{r_{min}}{1-\gamma} - \frac{r_{max}}{1-\gamma}| = \frac{r_{max} - r_{min}}{1-\gamma}$$

Exercise 2

(a)

```
Iterations:  43

Optimal value function:
[0.01543432 0.01559069 0.02744009 0.01568004 0.02685371 0.
0.05978021 0.         0.0584134  0.13378315 0.1967357  0.
0.         0.2465377  0.54419553 0.        ]
```

(b)

```
Computed policy: [2 3 2 3 0 0 0 0 3 1 0 0 0 2 1 0]
```