Exercise 3

[Avik Banerjee (3374885), Soumyadeep Bhattacharjee (3375428)]

1 Proofs

a)

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

2 Value Iteration

a) The algorithm takes 42 iterations to converge. The optimal value function is:

```
[0.015 0.016 0.027 0.016
 0.027 0     0.059 0
 0.058 0.134 0.197 0
 0     0.246 0.544 0]
```

b) The optimal policy is:

```
[1 3 2 3
 0 0 0 0
 3 1 0 0
 0 2 1 0]
```