Problem 1

Consider a finite discounted MDP $M = (S, A, P, R, \gamma)$. In this problem, we study some properties of value iteration. The Bellman optimality equation for the optimal value function $V^* : S \to \mathbb{R}$, which we also write $V^* \in \mathbb{R}^S$, is

$$V^*(s) = \max_{a \in A} \left( \sum_{s' \in S} P(s'|s, a) (R(s, a, s') + \gamma V^*(s')) \right).$$

Define the Bellman optimality operator $F^* : \mathbb{R}^S \to \mathbb{R}^S$ as

$$F^*V(s) = \max_{a \in A} \left( \sum_{s' \in S} P(s'|s, a) (R(s, a, s') + \gamma V(s')) \right),$$

where $F^*V(s)$ is shorthand for $(F^*(V))(s)$. Note that $S$ is finite so value functions are vectors in $\mathbb{R}^{|S|}$. The operator $F^*$ maps vectors in $\mathbb{R}^{|S|}$ to vectors in $\mathbb{R}^{|S|}$.

Value iteration amounts to the repeated application of $F^*$ to an arbitrary initial value function $V_0 \in \mathbb{R}^{|S|}$.

a) Prove that $V^*$ is an unique fixed point of $F^*$, i.e., $F^*V^* = V^*$ and that if $F^*V = V$ and $F^*V' = V'$ for two value functions $V, V' \in \mathbb{R}^S$, then $V = V'$, i.e., $V(s) = V'(s)$ for all $s \in S$.

b) Prove that $(F^*)^kV_0$ converges to $V^*$ as $k \to \infty$ for any $V_0 \in \mathbb{R}^{|S|}$. Consider convergence in max-norm. The max-norm of a vector $u \in \mathbb{R}^d$ is defined as $||u||_{\infty} = \max_{i \in \{1, \ldots, d\}} |u_i|$.

c) Given the optimal value function $V^*$ write down the expression that recovers the optimal policy $\pi^*$ as a function of $V^*$ and the parameters of $M$.\footnote{ Actually, such a procedure works for any value function $V \in \mathbb{R}^S$. This procedure is called policy extraction.}