Assignment 1

Proving convergence of the value iteration algorithm for policy evaluation.

Consider a MDP specified as $(\mathcal{S}, \mathcal{A}, \mathbb{P}, \mathcal{R}, \gamma)$. Let $s \in \mathcal{S}$. Then the state value of s for a policy π is

$v^\pi(s) = \sum_a \pi(a|s) \sum_r p(r|s,a) + \gamma \sum_a \pi(a|s) \sum_{s'} p(s'|s,a) v^\pi(s')$.

To compute $v^\pi(s)$ we have the value iteration algorithm, that starts with an estimate $v_0(s), \forall s \in \mathcal{S}$ which is say initialized to 0. Then for every k we define

$v_{k+1}(s) = \sum_a \pi(a|s) \sum_r p(r|s,a) + \gamma \sum_a \pi(a|s) \sum_{s'} p(s'|s,a) v_k(s')$.

Prove that $\lim_{k \to \infty} v_k(s) = v^\pi(s), \forall s \in \mathcal{S}$.

Hints:

1. Try reading the proof in the textbook and see if you can write it out on your own
2. Read about "Contracting mapping" and "Contraction mapping theorem" from Wikipedia - can you show that the "mapping" which transforms the vector $\overline{v}_k(.)$ to $\overline{v}_{k+1}(.)$ is a contraction map? Can that be used to show convergence?