Problem Set no. 1

Given: November 17, 2019

Exercise 1.1 (a) Describe a simple linear time algorithm for finding all the states from which a Markov chain stops with probability 1. (The running time should be linear in the number of non-zero transition probabilities.) (Hint: Start by finding all the states from which the Markov chain stops with a positive probability.)

(b) Extend the algorithm to find all states of a Markov Decision Process (MDP) from which the process ends with probability 1, no matter what the controller does. The running time should still be linear.

Exercise 1.2 Show that if an MDP on n states is *stopping*, i.e., stops from each initial state with probability 1, no matter what the controller does, then there exists $\varepsilon > 0$ such that from each state the probability that the process stops after at most n steps is at least ε. (Hint: rely on the algorithm developed in the previous exercise.)

Exercise 1.3 Show that if P is the probability transition matrix of a stopping Markov chain, then $(I - P)^{-1} = \sum_{r \geq 0} P^r$. (Note that $P^0 = I$.) (Prove that the infinite sum exists and is equal to the inverse.) In particular, $(I - P)^{-1}$ exists, all its entries are non-negative, and all entries on the diagonal are at least 1. (Hint: rely on the previous exercise to show that the sum is bounded by a geometric series.)

Exercise 1.4 (a) Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the *value iteration operator* of a discounted MDP, i.e., $T(\mathbf{y})_i = \min_{a \in A_i} c_a + \lambda \sum_{j \in S} p_{a,j} y_j$, for $\mathbf{y} = (y_i) \in \mathbb{R}^n$, $i \in S$, where $0 < \lambda < 1$ is the *discount factor*. Prove that T is a *contraction*, i.e., that $||T(\mathbf{x}) - T(\mathbf{y})||_\infty \leq \lambda ||\mathbf{x} - \mathbf{y}||_\infty$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

(b) Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the *value iteration operator* of a *stopping* MDP on n states, i.e., $T(\mathbf{y})_i = \min_{a \in A_i} c_a + \sum_{j \in S} p_{a,j} y_j$, for $\mathbf{y} \in \mathbb{R}^n$, $i \in S$. Prove that $T^{(n)}$, i.e., T iterated n times, is a *contraction*, i.e., there exists $0 < \lambda < 1$ such that $||T^{(n)}(\mathbf{x}) - T^{(n)}(\mathbf{y})||_\infty \leq \lambda ||\mathbf{x} - \mathbf{y}||_\infty$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

(c) In the previous item, prove that T is *not* necessarily a contraction.

Exercise 1.5 Let π be a positional strategy of a stopping MDP. Let π' be a policy obtained by choosing the best improving switch from each state, i.e., $\pi'_i = \arg\min_{a \in A_i} c_a + \sum_{j \in S} p_{a,j} y^\pi_j$, for every $i \in S$, where $\mathbf{y}^\pi = (y^\pi_i)$ are the values with respect to π. Prove that $y^{\pi'}_i \leq T(\mathbf{y}^\pi)_i$, for every $i \in S$. For which i's do we have $y^{\pi'}_i < T(\mathbf{y}^\pi)_i$?

Exercise 1.6 Let π be a positional strategy of a stopping MDP. Let $a \in A_i$ and $b \in A_j$, where $i \neq j$, be two improving switches with respect to π. Let $\pi' = \pi^{a,b}$ be the policy obtained from π by switching to a and b, i.e., $\pi'_i = a$, $\pi'_j = b$, and $\pi'_k = \pi_k$ for every $k \neq i, j$. Similarly, let π^a and π^b be the policies obtained by performing only one of the improving switches. Show that $\mathbf{y}^{\pi^{a,b}} \leq \mathbf{y}^\pi$, but that *not* necessarily $\mathbf{y}^{\pi^{a,b}} \leq \mathbf{y}^{\pi^a}$ and $\mathbf{y}^{\pi^{a,b}} \leq \mathbf{y}^{\pi^b}$. (Give an appropriate example.) Is it true that either $\mathbf{y}^{\pi^{a,b}} \leq \mathbf{y}^{\pi^a}$ or $\mathbf{y}^{\pi^{a,b}} \leq \mathbf{y}^{\pi^b}$?

Exercise 1.7 (a) Use the policy iteration algorithm to prove that any stopping MDP has a positional strategy that is optimal from each state without relying on Banach's fixed-point theorem.

(b) Do the same for stopping TBSGs.