Problem Set no. 1

Given: November 2, 2020
Due: November 15, 2020

Exercise 1.1 Let π be an arbitrary randomized and history-dependent policy for a total reward MDP Γ that satisfies the stopping condition. Argue directly, without relying on any of the results proved in class, that there is a *deterministic*, possibly history-dependent, policy π' such that $\textsc{val}^{\pi'}(\Gamma, i) \le \textsc{val}^{\pi}(\Gamma, i)$, for every $i \in S$.

Exercise 1.2 (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 1.3 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 1.4 Show that if P is the probability transition matrix of a stopping Markov chain, then $(I - P)^{-1} = \sum_{r \ge 0} P^r$. (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 1.5 (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 $||T(\mathbf{x}) - T(\mathbf{y})||_\infty \le \lambda ||\mathbf{x} - \mathbf{y}||_\infty$ 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 $||T^{(n)}(\mathbf{x}) - T^{(n)}(\mathbf{y})||_\infty \le \lambda ||\mathbf{x} - \mathbf{y}||_\infty$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

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

Exercise 1.6 (a) Write down the primal and dual LPs that correspond to an MDP with a *discount factor* $0 < \lambda < 1$. (b) Show directly, without relying on fact that the optimality equations for discounted MDPs have a solution, that the dual LP has at least one feasible solution. (c) Show directly that the dual LP for discounted MDPs is bounded. (d) Show that the dual LP for discounted MDPs has an optimal solution which is also a solution of the optimality equations for discounted MDPs.

(Bonus: Solve Exercise 1.6 for non-discounting MDPs that satisfy the stopping condition.)