The Johns Hopkins University
Department of Electrical and Computer Engineering

Foundations of Reinforcement Learning

Problem Set #5

Homework Rules:

- Whenever a proof is required, please provide a detailed justification of every step.
- Whenever Python/Matlab simulations are required, please attach a copy of the code. The code should be legible by an experienced reader.
- **Reference book**: Reinforcement Learning: An Introduction, 2nd edition, by Richard S. Sutton and Andrew G. Barto.
- Unless explicitly stated otherwise, exercises are expected to be solved by yourself using only Blackboard resources and your own lecture notes. Verbal and piazza discussions are acceptable, provided that do not involve any explicit writing of the solution.

Problems.

1. (a) Answer Sutton&Barto Exercise 4.3.
   Please start with

   $$q_\pi(s,a) = \mathbb{E}_\pi \left[G_t | S_t = s, A_t = a \right] ,$$

   and show all the derivation.

   (b) Equation (4.5) in the Sutton&Barto can be viewed as $v_{k+1} = \mathcal{T}_\pi(v_k)$, where $\mathcal{T}_\pi$ is an operator on value function v (See also the lecture note).
   Analogous to this, from (a) we have iteration $q_{k+1} = \mathcal{T}_\pi^q(q_k)$ to compute the state-action value function for policy π, where $\mathcal{T}_\pi^q$ is an operator on state-action value function q. Show $\mathcal{T}_\pi^q$ is γ-contracting, where $0 < \gamma < 1$ is the discounting factor.
   (Hint: any q can be represented by a nm-dimension vector, where n is the number of states and m is the number of action. Then $\mathcal{T}_\pi^q(q) = R_\pi^q + P_\pi^q q$ for some $R_\pi^q \in \mathbb{R}^{nm}$,$P_\pi^q \in \mathbb{R}^{nm \times nm}$.What are these R_π^q,P_π^q and what property of P_π^q leads to the contraction of $\mathcal{T}_\pi^q$?)

2. Answer Sutton&Barto Excercise 4.5.