Problem Set 1

BY JACK YANSONG LI
University of Illinois Chicago
Email: yli340@uic.edu

1 Definition of game environments

Problem 1. *(15 pt)* Given a two-player normal-form game $(2, A_1 \times A_2, r_{\text{joint}})$ with player 2's policy $\pi \in \Delta(A_2)$ fixed. Construct a single-agent environment from the player 1's point of view.

Hint: The model of the single agent environment is also called *multi-armed bandit* defined as $(2, A, r)$, where the reduced action space A and the corresponding reward r should be defined by you.

Problem 2. *(15 pt)* Reformulate the team Markov game $(N, H, S, s_0, A_{\text{joint}}, \mathbb{P}, r, \gamma)$, where $|S| < \infty$, $|A_{\text{joint}}| < \infty$ into a indentical-interest normal-form game $(N, A'_{\text{joint}}, r')$. Show that the size of the joint action space $|A'_{\text{joint}}|$ grows exponentially with respect to $|S|$.

Problem 3. *(30 pt)* Extend the definition of stochastic game into partially observable setting (usually called *partially observable stochastic game*), based on the definition you just created

1. Define policies and cumulative reward.
2. Define regrets for player 1 given other player's joint policies.
3. Reformulate the *partially observable stochastic game* into a single agent environment in player 1's perspective given other player's joint policies. The single agent environment is called partially observable Markov decision process.

2 Regret analysis

Problem 4. *(15 pt)* Construct an example using normal-form game that shows the regret minimization of player 1 returns a better strategy than the cumulative reward maximization.

Hint: Construct a two-player identical interest game. Carefully design the reward so the reward varies dramatically with respect to player 2's action.

3 Markov decision process

Problem 5. *(25 pt)* Prove that the optimal policy of an infinite $(H = \infty)$ Markov decision process is stationary, i.e., $\mu^*: S \to \Delta(A)$ instead of $\mu^*: S \times \mathbb{N}_+ \to \Delta(A)$.