MIT 6.S890 Topics in Multiagent Learning Fall 2024

Homework 1

Released on: Sept. 17, 2024 **Due on: Oct. 1, 2024, 11:59pm ET**

Instructions This homework contains six problems (each split into smaller tasks), for a total of 100 points. Please attach a readable printout of your code for Problem 6 at the end of your submission.

Collaboration policy You are welcome to discuss the problems with other students. However you should write up the solutions by yourself.

1 Another Look at the Nash Improvement Function (10 points)

In this problem, we will go through Nash's original proof to show that Nash equilibria are the fixed points of the improvement function.

Recall the Nash improvement function discussed in class: given player strategies $x_1 \in \Delta(A_1), \ldots, x_n \in \Delta(A_n)$, the Nash improvement function $\varphi : \Delta(A_1) \times \cdots \times \Delta(A_n) \to \Delta(A_1) \times \cdots \times \Delta(A_n)$ is given by:

$$\varphi_{i,a_i}(x_1, \ldots, x_n) := \frac{x_{i,a_i} + \ r_{i,a_i}(x_1, \ldots, x_n)\ ^+}{1 + \sum_{a_i' \in A_i} r_{i,a_i'}(x_1, \ldots, x_n)\ ^+}$$

for all player $i \in [n]$ and action $a_i \in A_i$, where $[r]^+ := \max\{0, r\}$ denotes the positive part of r and

$$r_{i,a_i}(x_1, \ldots, x_n) := u_i(a_i, x_{-i}) - u_i(x_1, \ldots, x_n)$$

is the expected regret that player i experiences with respect to action $a_i \in A_i$.

Problem 1.1 (2 points). Show that if $r_{i,a_i}(x_1, \ldots, x_n) > 0$ for some player $i \in [n]$ and action $a_i \in A_i$, then $\varphi_{i,a_i}(x_1, \ldots, x_n) > 0$.

Problem 1.2 (3 points). Show that the expect regret player i experiences satisfies

$$\sum_{a_i \in A_i} r_{i,a_i}(x_1, \ldots, x_n) \cdot x_{i,a_i} = 0.$$

Now, select any fixed point $(x_1^*, \ldots, x_n^*)$ of function φ, i.e., such that $\varphi_{i,a_i}(x_1^*, \ldots, x_n^*) = x_{i,a_i}^*$ for all players i and action $a_i \in A_i$.

Problem 1.3 (5 points). Show that the above results imply that, for any player $i \in [n]$ and action $a_i \in A_i$,

$$r_{i,a_i}(x_1^*, \ldots, x_n^*) \leq 0.$$

Conclude that $(x_1^*, \ldots, x_n^*)$ is a Nash equilibrium.

★ Hint: for the sake of contradiction, assume $r_{i,a_i}(x_1^*, \ldots, x_n^*) > 0$ for some player i and action a_i. Under this assumption, demonstrate that there exists an alternative action a_i' such that $\varphi_{i,a_i'}(x_1^*, \ldots, x_n^*) < x_{i,a_i'}^*$.

2 Symmetry and Nash Equilibria (15 points)

In this problem, you will show that if two players in a normal-form game are "symmetric", then there always exists a Nash equilibrium in which the players play the same strategy.

Problem 2.1 (15 points). Consider an n-player game in which the first two players have the following properties:

- they have the same action set $A_1 = A_2$;
- their utility functions u_1, u_2 satisfy the following property: for all strategy profiles $(x_1, x_2, x_3, \ldots, x_n)$,

$$u_1(x_1, x_2, x_3, \ldots, x_n) = u_2(x_2, x_1, x_3, \ldots, x_n).$$

Show that the game has a Nash equilibrium $(x_1^*, x_2^*, \ldots, x_n^*)$ in which $x_1^* = x_2^*$.

★ Hint: a strategy profile is a Nash equilibrium if and only if it is a fixed point of the Nash improvement function.

3 Properties of the Nash Equilibria (20 points)

In class, we mentioned that there exists a three-player game with rational payoffs where all Nash equilibria involve players using strategies with irrational probabilities. The following problem asks you to keep the course staff honest and compute the Nash equilibrium of that game.

Problem 3.1 (10 points). Consider the following three-player game where all players choose their actions simultaneously:

- Player 1 chooses between top and bottom.
- Player 2 chooses between left and right.
- Player 3 chooses between X and Y.

action X

	left	right
top	3, 0, 2	0, 2, 0
bottom	0, 1, 0	1, 0, 0

action Y

	left	right
top	1, 0, 0	0, 1, 0
bottom	0, 3, 0	2, 0, 3

Each cell represents the payoffs to Player 1, 2, and 3, respectively, for the corresponding combination of actions. Let x, y, z be the probability for each player to choose action top, left, and X, respectively. Show that

$$(x, y, z) = \frac{53}{46} - \frac{\sqrt{601}}{46}, \frac{\sqrt{601}}{24} - \frac{13}{24}, -\frac{23}{4} + \frac{\sqrt{601}}{4} \approx (0.619, 0.480, 0.379)$$

is the only Nash equilibrium of the game.

As mentioned, for two-player games the situation is different. This is because the problem can be reduced to a linear complementarity problem (LCP), which is guaranteed to have a rational solution. Recall that an LCP can be formulated in the form

$$\begin{aligned} \underset{z,w \in \mathbb{R}^m}{\text{find}} \quad & z \\ \text{s.t.} \quad & w = Mz + q \\ & z \geq 0, w \geq 0 \\ & z^\top w = 0, \end{aligned}$$

where M is a square matrix and q is a vector.

Problem 3.2 (10 points). Consider a generic two-player general-sum game. Show that finding Nash equilibrium in the game can be reduced to linear complementarity problem with a number of variables that is linear in the number of actions $|A_1| + |A_2|$ of the game.

★ Hint: the strategy profile (x_1, x_2) is a Nash equilibrium if and only if $r_{i,a_i}(x_1, x_2) \leq 0$ for all $i \in \{1, 2\}$ and action $a_i \in A_i$. Unfortunately, the obstacle is that the definition of r_i involves bilinear products of the form $x^\top U_1 y$ where x and y are the strategies. Use the complementarity condition $z^\top w = 0$ to express such a term, by possibly first introducing a dummy variable.

4 Properties of Coarse Correlated Equilibria (10 points)

Recall that a coarse correlated equilibrium is a probability distribution $\mu \in \Delta(A_1 \times \cdots \times A_n)$ over the joint action space such that for each player $i \in [n]$ and for all actions $a_i' \in A_i$, the following condition holds:

$$\mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu}[u_i(a_i', a_{-i})] \leq \mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu}[u_i(a_i, a_{-i})].$$

Since the condition is linear, we have the set of coarse correlated equilibria forms a convex polytope. We aim to explore the relationship between Nash equilibria and coarse correlated equilibria within this convex set.

Problem 4.1 (5 points). We say a game is *non-trivial* if the utility function satisfies $u_i(a_i, a_{-i}) \neq u_i(a_i', a_{-i})$ for some player i, action $a_i, a_i' \in A_i$ and $a_{-i} \in A_{-i}$. Show that in any non-trivial game, there is *no* Nash equilibrium lies in the interior of the convex polytope of coarse correlated equilibria.

Note: A similar relationship can be established for correlated equilibria.

In two-player zero-sum games, the two concepts have a closer connection: the marginal distributions of any coarse correlated equilibrium constitute a Nash equilibrium.

Problem 4.2 (5 points). Let $\mu \in \Delta(A_1 \times A_2)$ be a coarse correlated equilibrium of a two-player *zero-sum* game. Show that the *marginalization* of μ is a Nash equilibrium; formally, show that the pair of marginalized strategies (x_1, x_2) defined by

$$x_{1,a_1} \coloneqq \sum_{a_2' \in A_2} \mu_{a_1,a_2'}; \qquad x_{2,a_2} \coloneqq \sum_{a_1' \in A_1} \mu_{a_1',a_2}$$

for all $a_1 \in A_1, a_2 \in A_2$ is a Nash equilibrium of the game.

Recall that in class, we have shown that no-regret algorithms lead to approximate coarse correlated equilibria. This connection enables a different path to computing approximate Nash equilibria efficiently in two-player zero-sum games.

5 Dominated Strategies in (Coarse) Correlated Equilibria (20 points)

Consider a n-player general sum game where each player i has an action set A_i. The utility that player i receives is given by $u_i(a_1, \ldots, a_n)$ when each player j plays action $a_j \in A_j$.

As a reminder, an action $a_i \in A_i$ is dominated by another action a_i^* if, for every possible combination of actions $a_{-i} \in A_{-i}$ chosen by the other players, a_i^* always yields a higher payoff for player i than a_i, that is, $u_i(a_i^*, a_{-i}) > u_i(a_i, a_{-i})$.

Problem 5.1 (6 points). Show that a dominated action cannot be in the support of any correlated equilibrium in any n-player game. In other words, any correlated distribution of actions that is a correlated equilibrium must place zero probability mass on all action tuples that contain a_i.

Problem 5.2 (7 points). Show that the same does not hold for coarse correlated equilibria, by giving an example of a game where a dominated action is in the support of a coarse correlated equilibrium.

Problem 5.3 (7 points). We saw in class that no-external-regret algorithms converge to coarse correlated equilibria in general games. Consider the multiplicative weights update (MWU) algorithm, which assigns to every action a_i of player i at iteration t the probability mass

$$x_{i,a_i}^{(t)} \propto \exp\left(\eta \sum_{\tau<t} r_{i,a_i}(x_1^{(\tau)}, \dots, x_n^{(\tau)})\right)$$

where $\eta > 0$ is a learning rate parameter and

$$r_{i,a_i}(x_1^{(\tau)}, \dots, x_n^{(\tau)}) := \mathop{\mathbb{E}}_{(a_1^{(\tau)},\dots,a_n^{(\tau)})\sim(x_1^{(\tau)},\dots,x_n^{(\tau)})} u_i(a_i, a_{-i}^{(\tau)}) - u_i(a_1^{(\tau)}, \dots, a_n^{(\tau)})$$

is the expected regret of player i with respect to action $a_i \in A_i$ in episode τ. Show that as $t \to \infty$, the probability mass $x_{i,a_i}^{(t)}$ that the algorithm places on any *dominated* action a_i converges to 0.

6 Implementation of Learning Algorithms (25 points)

In this problem, you will implement the no-regret algorithms for normal-form games. The zip of the homework also contains a stub Python file to help you set up your implementation.

You will run your implementation of the two no-regret algorithms we have seen in class (regret matching and multiplicative weights update) on two games: rock-paper-superscissors (a simple variant of rock-paper-scissors, where beating paper with scissors gives a payoff of 2 instead of 1), and a small poker variant called Kuhn poker [Kuhn, 1950]. We focus on the normal-form representation of game, where the payoff matrix U_1 for player 1 is given in the zip of this homework; player 1 is on the rows, player 2 is on the columns. This means that the expected utility for player 1 given mixed strategies x and y for player 1 and 2 respectively is given by the formula

$$u_1(x, y) = x^\top U_1 y.$$

Since these games are 0-sum, the expected utility for player 2 is the opposite of player 1: $u_2(x, y) = -u_1(x, y)$. We denote by A_x and A_y the action set of player 1 and player 2.

As a reminder, the probability mass that player 1 puts on action a at each iteration t is given by

$$x_a^{(t)} \propto \left[\sum_{\tau<t} r_{1,a}(x^{(\tau)}, y^{(\tau)})\right]^+ \in \mathcal{X} := \Delta^{|A_x|}$$

in regret matching (RM), and

$$x_a^{(t)} \propto \exp\left(\eta \sum_{\tau<t} r_{1,a}(x^{(\tau)}, y^{(\tau)})\right)$$

in multiplicative weights update (MWU), where $[r]^+ := \max\{0, r\}$ denotes the positive part of r and

$$r_{1,a}(x, y) := u_1(a, y) - u_1(x, y) = e_a^\top U_1 y - x^\top U_1 y$$

is the expected regret of player 1 with respect to action $a \in A_x$. The updating rule for player 2 is symmetric.

6.1 Learning to best respond

A good smoke test when implementing regret minimization algorithms is to verify that they learn to best respond. In particular, you will verify that your implementations applied to player 1 learns a best response against player 2 when player 2 plays the *uniform* strategy $y^{(0)}$, that is, the strategy that picks all of the available actions with equal probability.

Let u be the sequence-form representation of the strategy for player 2 that selects each of the available actions with equal probability. When player 2 plays according to that strategy, the utility gradient for player 1 is given by $u_x := U_1 y$, where U_1 is the payoff matrix of the game for player 1 and y is the (mixed) strategy of player 2.

For each of the two games, take your no-regret algorithm for player 1, and let it output strategies $x^{(t)} \in \mathcal{X}$ while giving as feedback at each time t the same utility vector u_x. As $T \to \infty$, the average strategy

$$\bar{x}^{(T)} := \frac{1}{T} \sum_{t=1}^{T} x^{(t)} \in \mathcal{X} \tag{1}$$

will converge to a best response to the uniform strategy $y^{(0)}$, that is,

$$\lim_{T \to \infty} (\bar{x}^{(T)})^\top U_1 y^{(0)} = \max_{\hat{x} \in \mathcal{X}} \hat{x}^\top U_1 y^{(0)}.$$

If the above doesn't happen empirically, something is wrong with your implementation.

Problem 6.1 (10 points). In each of the two games, apply your regret matching (RM) and multiplicative weights update (MWU) implementation. You should use the learning rate $\eta = 0.01$ for MWU. At each time t, give as feedback to the algorithm the same utility vector $u_x := U_1 y^{(0)}$, where $y^{(0)} \in \mathcal{Y} := \Delta^{A_y}$ is the uniform random strategy for player 2, satisfying $y_a^{(0)} = 1/|A_y|$. Run the algorithm for 10000 iterations. After each iteration $T = 1, \ldots, 10000$, compute the value of $v^{(T)} := (\bar{x}^{(T)})^\top U_1 y^{(0)}$ where $\bar{x}^{(T)} \in \mathcal{X}$ is the average strategy output so far by RM, as defined in (1).

Plot $v^{(T)}$ as a function of T. Empirically, what is the limit you observe $v^{(T)}$ is converging to?

Your solution should include four plots (one for each game–algorithm combination). Don't forget to turn in your implementation.

★ Hint: in rock-paper-superscissor, $v^{(T)}$ should approach the value 1/3.

6.2 Learning a Nash equilibrium

Now that you are confident that your implementation of the algorithms is correct, you will use RM and MWU to converge to Nash equilibrium using the self-play idea recalled next.

The idea behind using regret minimization to converge to Nash equilibrium in a two-player zero-sum game is to use *self play*. We instantiate two regret minimization algorithms, $\mathcal{R}_x$ and $\mathcal{R}_y$, for the domains of the maximization and minimization problem, respectively. At each time t the two regret minimizers output strategies $x^{(t)}$ and $y^{(t)}$, respectively. Then, they receive as feedback the vectors $u_x^{(t)}, u_y^{(t)}$ defined as

$$u_x^{(t)} := U_1 y^{(t)}, \qquad u_y^{(t)} := -U_1^\top x^{(t)}, \tag{2}$$

where U_1 is player 1's payoff matrix.

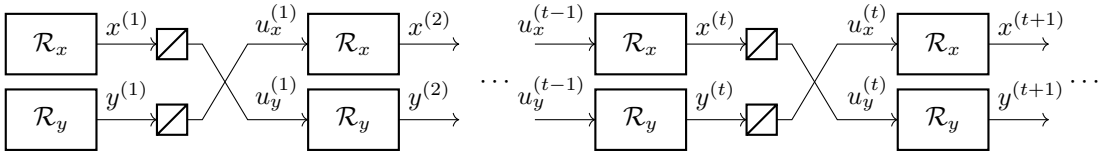

Figure 1: The flow of strategies and utilities in regret minimization for games. The symbol ⊠ denotes computation/construction of the utility vector.

We summarize the process pictorially in Figure 1.

A well known folk theorem establish that the pair of average strategies produced by the regret minimizers up to any time T converges to a Nash equilibrium, where convergence is measured via the *Nash equilibrium gap*

$$0 \leq \gamma(x,y) \coloneqq \max_{\hat{x}\in\mathcal{X}}\{\hat{x}^\top U_1 y\} - x^\top U_1 y \;+\; x^\top U_1 y - \min_{\hat{y}\in\mathcal{Y}}\{x^\top U_1 \hat{y}\} = \max_{\hat{x}\in\mathcal{X}}\{\hat{x}^\top U_1 y\} - \min_{\hat{y}\in\mathcal{Y}}\{x^\top U_1 \hat{y}\}.$$

A point $(x,y) \in \mathcal{X}\times\mathcal{Y}$ has zero Nash equilibrium gap if and only if it is a Nash equilibrium of the game.

Theorem 1. Consider the self-play setup summarized in Figure 1, where $\mathcal{R}_x$ and $\mathcal{R}_y$ are regret minimizers for the sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\mathrm{Reg}_x^{(T)}$ and $\mathrm{Reg}_y^{(T)}$ be the (sublinear) regret cumulated by $\mathcal{R}_x$ and $\mathcal{R}_y$, respectively, up to time T, and let $\bar{x}^{(T)}$ and $\bar{y}^{(T)}$ denote the average of the strategies produced up to time T, that is,

$$\bar{x}^{(T)} \coloneqq \frac{1}{T}\sum_{t=1}^{T} x^{(t)}, \qquad \bar{y}^{(T)} \coloneqq \frac{1}{T}\sum_{t=1}^{T} y^{(t)}. \tag{3}$$

Then, the Nash equilibrium gap $\gamma(\bar{x}^{(T)},\bar{y}^{(T)})$ of strategy profile $(\bar{x}^{(T)},\bar{y}^{(T)})$ satisfies

$$\gamma(\bar{x}^{(T)},\bar{y}^{(T)}) \leq \frac{\mathrm{Reg}_x^{(T)} + \mathrm{Reg}_y^{(T)}}{T} \to 0 \qquad \text{as } T\to\infty.$$

Problem 6.2 (10 points). Let your implementation of RM for player 1 and MWU (with learning rate $\eta = 0.01$) for player 2 play against each other. Plot the Nash equilibrium gap $\gamma(\bar{x}^{(T)},\bar{y}^{(T)})$ and the expected utility for player 1, $u_1(\bar{x}^{(T)},\bar{y}^{(T)})$, of the average strategies as a function of the number of iterations $T = 1,\ldots,10000$.

Your solution should include four plots (two for each game—one for the Nash equilibrium gap and one for the utility). Don't forget to turn in your implementation.

★ Hint: the Nash equilibrium gap should be going to zero. The expected utility of the average strategies in rock-paper-superscissor should approach the value 0.

6.3 Alternating RM$^+$

To achieve better performance in practice when learning Nash equilibria in two-player zero-sum games, people often make the following modifications to the setup of the previous subsection.

- Instead of regret matching, use the regret matching plus algorithm. As a reminder, the probability mass that player 1 puts on action a at each iteration t is given by

$$x_a^{(t)} \propto R_{1,a}^{(t)},$$

where the cumulative regret is maintained according to

$$R_{1,a}^{(t)} \leftarrow \left[R_{1,a}^{(t-1)} + r_{1,a}(x^{(t)}, y^{(t)}) \right]^{+}.$$

- Instead of using the classical self-play scheme described in Figure 1, people *alternate* the iterates and feedback as described in Figure 2, where the utility vector $u_x^{(t)}$ is as defined in (2), whereas

$$\tilde{u}_y^{(t)} := -U_1^\top x^{(t+1)}.$$

(Note that at the very beginning, $x^{(1)}$ does not participate in the computation of any utility vector).

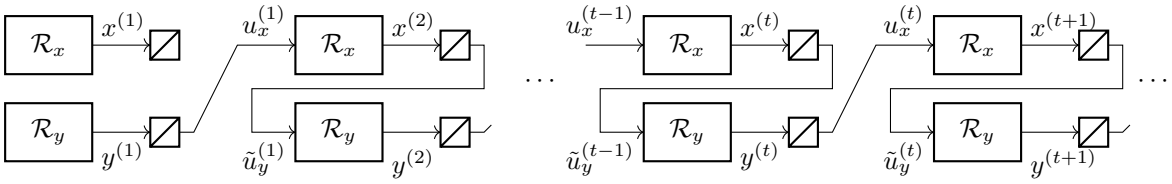

Figure 2: The alternation method for RM in games. The symbol ⊠ denotes computation/construction of the utility vector.

- Finally, the *linear average* of the strategies, defined as the weighted averages

$$\bar{\bar{x}}^{(T)} := \frac{2}{T(T+1)} \sum_{t=1}^{T} t \cdot x^{(t)}, \qquad \bar{\bar{y}}^{(T)} := \frac{2}{T(T+1)} \sum_{t=1}^{T} t \cdot y^{(t)}$$

is considered instead of the regular averages (3) when computing the Nash equilibrium gap.

Problem 6.3 (5 points). Modify your implementation of RM to match the RM$^+$ self-play setup described above. Run RM$^+$ for 10000 iterations, plotting the expected utility for player 1, $u_1(\bar{\bar{x}}^{(T)}, \bar{\bar{y}}^{(T)})$, and the Nash equilibrium gap $\gamma(\bar{\bar{x}}^{(T)}, \bar{\bar{y}}^{(T)})$ of the linear averages after each iteration T.

Your solution should include four plots (two for each game—one for the Nash equilibrium gap and one for the utility). Don't forget to turn in your implementation.

References

H. W. Kuhn. A simplified two-person poker. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games*, volume 1 of *Annals of Mathematics Studies, 24*, pages 97–103. Princeton University Press, Princeton, New Jersey, 1950.

6.S890 Topics in Multiagent Learning

Fall 2024