MIT 6.S890 Topics in Multiagent Learning Fall 2024

Homework 1 - Reference Solution

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

Problem 1

1.1

According to the definition, we have $x_{i,a_i} \geq 0$. Together with $r_{i,a_i}(x_1, \dots, x_n) > 0$, we can immediately conclude that

$$\varphi_{i,a_i}(x_1, \dots, x_n) := \frac{x_{i,a_i} + \ \ r_{i,a_i}(x_1, \dots, x_n) \ ^{+}}{1 + \ \ {}_{a_i' \in A_i} \ \ r_{i,a_i'}(x_1, \dots, x_n) \ ^{+}} > 0.$$

1.2

The equation can be verified according to

$$\begin{aligned}
{}_{a_i \in A_i} \ \ r_{i,a_i}(x_1, \dots, x_n) \cdot x_{i,a_i} &= {}_{a_i \in A_i} \ \ u_i(a_i, x_{-i}) - u_i(x_1, \dots, x_n) \ \ \cdot x_{i,a_i} \\
&= {}_{a_i \in A_i} \ \ u_i(a_i, x_{-i}) \cdot x_{i,a_i} - u_i(x_1, \dots, x_n) \ \ {}_{a_i \in A_i} \ \ x_{i,a_i} \\
&= u_i(x_1, \dots, x_n) - u_i(x_1, \dots, x_n) \\
&= 0,
\end{aligned}$$

where third equality holds since ${}_{a_i \in A_i} \ u_i(a_i, x_{-i}) \cdot x_{i,a_i} = u_i(x_1, \dots, x_n)$ from the linearity of utility function, and ${}_{a_i \in A_i} \ x_{i,a_i} = 1$ from the definition of strategy.

1.3

Towards a contradiction, assume there is some player $i \in [n]$ with action $a_i \in A_i$ such that

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

According to problem 1.1, we have $\varphi_{i,a_i}(x_1^*, \dots, x_n^*) > 0$. This further implies that

$$r_{i,a_i}(x_1^*, \dots, x_n^*) \cdot x_{i,a_i}^* = \varphi_{i,a_i}(x_1^*, \dots, x_n^*) \cdot x_{i,a_i}^* > 0.$$

Furthermore, combining this with the result in problem 1.2, which shows that the summation over A_i is zero, we conclude that there exists another action $\hat{a}_i \in A_i$ such that

$$r_{i,\hat{a}_i}(x_1^*, \dots, x_n^*) \cdot x_{i,\hat{a}_i} < 0.$$

Since $x_{i,\hat{a}_i} \geq 0$, it follows that $r_{i,\hat{a}_i}(x_1^*, \dots, x_n^*) < 0$. Finally, we have

$$\begin{aligned}
\varphi_{i,\hat{a}_i}(x_1^*, \dots, x_n^*) &= \frac{x^*_{i,\hat{a}_i} + \ \ r_{i,\hat{a}_i}(x_1^*, \dots, x_n^*)^{\ +}}{1 + \ \ {}_{a_i' \in A_i} \ r_{i,a_i'}(x_1^*, \dots, x_n^*)^{\ +}} \\
&\leq \frac{x^*_{i,\hat{a}_i}}{1 + r_{i,a_i}(x_1^*, \dots, x_n^*)} \\
&< x^*_{i,\hat{a}_i},
\end{aligned}$$

contradicting the fact that x^* is a fixed point. Therefore, we have $r_{i,a_i}(x_1^*, \dots, x_n^*) \leq 0$, implying that x^* is a Nash equilibrium.

Problem 2

2.1

Consider a restricted strategy space $S = \{x_0 \in \Delta(A_1) = \Delta(A_2), x_3 \in \Delta(A_3), \dots, x_n \in \Delta(A_n) : (x_0, x_0, x_3, \dots, x_n)\}$. We will show the desired Nash equilibrium exists by applying the fixed-point theorem to the restricted strategy space. Consider the Nash improvement function $\varphi : S \to \Delta(A_1) \times \cdots \times \Delta(A_n)$ given by

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

For each action $a \in A_1 = A_2$ and strategy profile $(x_0, x_0, x_3, \dots, x_n) \in S$, we have that

$$\begin{aligned}
\varphi_{1,a}(x_0, x_0, x_3, \dots, x_n) &= \frac{x_{0,a} + \ \ r_{1,a}(x_0, x_0, x_3, \dots, x_n)^{\ +}}{1 + \ \ {}_{a_i' \in A_i} \ r_{i,a_i'}(x_0, x_0, x_3, \dots, x_n)^{\ +}} \\
&= \frac{x_{0,a} + \ \ r_{2,a}(x_0, x_0, x_3, \dots, x_n)^{\ +}}{1 + \ \ {}_{a_i' \in A_i} \ r_{i,a_i'}(x_0, x_0, x_3, \dots, x_n)^{\ +}} \\
&= \varphi_{2,a}(x_0, x_0, x_3, \dots, x_n),
\end{aligned}$$

where the second equality is given by $u_1(x_1, x_2, \dots, x_n) = u_2(x_2, x_1, \dots, x_n)$. This shows that the codomain of φ is also S. According to Nash's fixed-point theorem, we have that a desired fixed point $(x_0^*, x_0^*, \dots, x_n^*)$ is a Nash equilibrium exists.

Problem 3

3.1

Denote by $(x, y, z) \in [0, 1]^3$ the probabilities that the players choose actions top, left, and X, respectively, in the Nash equilibrium. The utility for each player taking a specific action is given by:

$$\begin{aligned}
u_1(\text{top}) &= y(1 + 2z) \\
u_1(\text{bottom}) &= (1 - y)(2 - z) \\
u_2(\text{left}) &= (1 - x)(3 - 2z) \\
u_2(\text{right}) &= x(1 + z) \\
u_3(\text{X}) &= 2xy \\
u_3(\text{Y}) &= 3(1 - x)(1 - y)
\end{aligned}$$

Firstly, we show that there is no Nash equilibrium where Player 1 or Player 2 takes a pure strategy.

- If $x = 0$, the values of Player 2's actions, left and right, are $3 - 2z \geq 1$ and 0, respectively. Thus, taking action left will always dominate taking action right. This indicates $x = 0 \Rightarrow y = 1$.
- Conversely, when $x = 1$, the values of Player 2's actions are 0 and $1 + z \geq 1$, respectively. So Player 2 will always take action right. This indicates $x = 1 \Rightarrow y = 0$.
- Similarly, if $y = 0$, the values of Player 1's actions, top and bottom, are 0 and $2 - z \geq 1$, respectively. Player 1 will always take action bottom, indicating that $y = 0 \Rightarrow x = 0$.
- When $y = 1$, the values of Player 1's actions are $1 + 2z \geq 1$ and 0, respectively. Player 1 will always take action top, indicating that $y = 1 \Rightarrow x = 1$.

However, there is no suitable assignment for both x and y. This rules out any pure strategy profile for Player 1 and Player 2.

The above reasoning shows that $x, y \in (0, 1)$. According to the definition of Nash equilibrium, both players are indifferent between the actions in their support. This implies that

$$u_1(\text{top}) = u_1(\text{bottom}), \quad u_2(\text{left}) = u_2(\text{right})$$

which indicates that

$$\frac{y}{1-y} = \frac{2-z}{1+2z}, \quad \frac{1-x}{x} = \frac{1+z}{3-2z}. \tag{1}$$

Next, we show that there is no Nash equilibrium where Player 3 takes a pure strategy.

- If $z = 0$, the value of action Y is no less than that of action X. Furthermore, we have $x = \frac{3}{4}$ and $y = \frac{2}{3}$ from (1). However, this leads to a contradiction since $u_3(X) = 1 > \frac{1}{4} = u_3(Y)$.
- If $z = 1$, the value of action X is no less than that of action Y. Furthermore, we have $x = \frac{1}{3}$ and $y = \frac{1}{4}$ from (1). However, this again leads to a contradiction since $u_3(X) = \frac{1}{6} < \frac{3}{8} = u_3(Y)$.

As a result, we have $z \in (0, 1)$. This suggests that

$$u_3(X) = u_3(Y) \Rightarrow \frac{2}{3} \cdot \frac{y}{1-y} = \frac{1-x}{x}. \tag{2}$$

Substituting (2) into (1) gives

$$\frac{2}{3} \cdot \frac{2-z}{1+2z} = \frac{1+z}{3-2z}.$$

Solving the quadratic equation gives $z = -\frac{23}{4} \pm \frac{\sqrt{601}}{4}$. Since we have to ensure $z > 0$, the only valid solution remaining is $z = -\frac{23}{4} + \frac{\sqrt{601}}{4}$. Substituting this into (1) yields the only Nash equilibrium as

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

Remarks: This problem provides some intuition on why Nash equilibrium is generally hard to compute. The equilibrium condition only holds for actions in the support of the Nash equilibrium. Thus, we need to determine the support of the solution, which is difficult in general. Once the support is identified, one can compute the Nash equilibrium by simply solving the polynomial equations.

3.2

Denote by U_1 and U_2 the reward vectors of Player 1 and Player 2, respectively, i.e., $u_1(x_1, x_2) = x_1^\top U_1 x_2$ and $u_2(x_1, x_2) = x_1^\top U_2 x_2$. Consider the following program:

$$\begin{aligned}
\min_{\substack{r_1, x_1 \in \mathbb{R}^{|A_1|} \\ r_1, x_1 \in \mathbb{R}^{|A_2|} \\ v_1, v_2 \in \mathbb{R}}} \quad & 0 \\
\text{s.t.} \quad & x_1^\top r_1 = 0, \; x_2^\top r_2 = 0 \\
& r_1 = U_1 x_2 - v_1 \mathbf{1} \\
& r_2 = U_2^\top x_1 - v_2 \mathbf{1} \\
& x_1 \geq 0, \; x_2 \geq 0, \; x_1^\top \mathbf{1} = 1, \; x_2^\top \mathbf{1} = 1 \\
& r_1 \geq 0, \; r_2 \geq 0
\end{aligned}$$

where $\mathbf{1}$ stands for the all-ones vector. It is straightforward to verify that the program can be converted to a linear complementarity problem.

Now, we show that any solution to the problem, (x_1, x_2), is a Nash equilibrium of the game. Correspondingly, r_1 and r_2 encode the regret of taking each action. We will demonstrate that the strategy profile x_1 is a best response of Player 1 to x_2. According to the constraints $x_1 \geq 0$ and $x_1^\top \mathbf{1} = 1$, x_1 is a probability distribution. Thus, there must be some action $a_1 \in A_1$ such that $x_{1,a_1} > 0$. Given $x_1^\top r_1 = 0$, this indicates that there is some action $a_1 \in A_1$ such that $r_{1,a_1} = 0$. From the constraint $r_1 \geq 0$, we have

$$v_1 = \max_{a_1 \in A_1} e_{1,a_1}^\top U_1 x_2 = \max_{a_1 \in A_1} u_1(a_1, x_2),$$

which is the best response to the game. This shows that

$$r_{1,a_1} := (U_1 x_2 - v_1 \mathbf{1})_{a_1} = u_1(x_1, x_2) - \max_{a_1 \in A_1} u_1(a_1, x_2),$$

indicating that r_1 indeed encodes the regret of taking each action. Now, for each action $a_1 \in A_1$, either $x_{1,a_1} = 0$, meaning the action is never taken, or $r_{1,a_1} = 0$, meaning there is no regret in taking the action. This indicates that any action in the support of x_1 is a best response to x_2, which concludes that x_1 is a best response to x_2. Similarly, we can show that x_2 is a best response to x_1. Therefore, (x_1, x_2) is a Nash equilibrium of the game.

Problem 4

4.1

Consider a non-trivial game with $u_i(a_i, a_{-i}) \neq u_i(\hat{a}_i, a_{-i})$ for Player i. Denote by $\mu = (a_i, a_{-i})$ and $\hat{\mu} = (\hat{a}_i, a_{-i})$ two spike distributions. We have

$$\mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu}[u_i(a_i, a_{-i})] = u_i(a_i, a_{-i}) \neq u_i(\hat{a}_i, a_{-i}) = \mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\hat{\mu}}[u_i(a_i, a_{-i})]. \tag{3}$$

Moreover, for any action $a_i' \in A_i$, we have

$$\mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu}[u_i(a_i', a_{-i})] = u_i(a_i', a_{-i}) = \mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\hat{\mu}}[u_i(a_i', a_{-i})]. \tag{4}$$

Combining (3) and (4), we have that there exist some policy $\bar{\mu} \in \{\mu, \hat{\mu}\}$ such that

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

This indicates that the corresponding constraints are all non-trivial.

Towards a contradiction, assume there exists a Nash equilibrium $\mu^* = (x_1, \dots, x_n)$ that lies in the interior of the convex polytope of coarse correlated equilibria. In this case, for any action $a_i' \in A_i$, we have

$$\mathbb{E}_{(a_1,\dots,a_n)\sim\mu^*}[u_i(a_i', a_{-i})] < \mathbb{E}_{(a_1,\dots,a_n)\sim\mu^*}[u_i(a_i, a_{-i})], \tag{5}$$

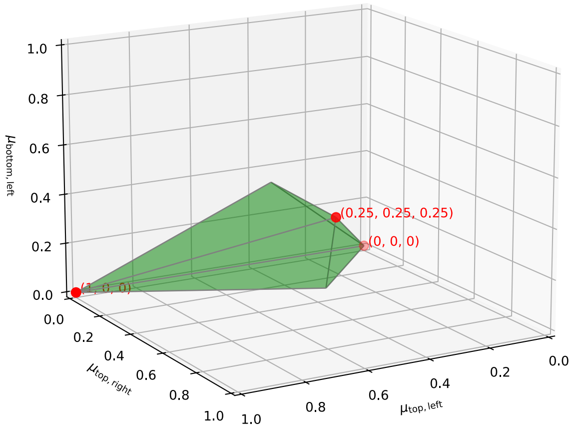
as they are all non-trivial constraints. Now, consider the best response $a_i^* \in A_i$ to x_{-i}. We have

$$\mathbb{E}_{(a_1,\dots,a_n)\sim\mu^*}[u_i(a_i', a_{-i})] = u_i(a_i', x_{-i}) \geq u_i(x_i, x_{-i}) = \mathbb{E}_{(a_1,\dots,a_n)\sim\mu^*}[u_i(a_i, a_{-i})].$$

where the equality holds since μ^* is a product distribution. However, this contradicts (5). This indicates that some non-trivial constraint must hold with equality, implying that μ^* cannot lie in the interior of the convex polytope of coarse correlated equilibria.

Remarks: One may wonder what geometric features are characterized by the problem. We provide a concrete instance for better understanding. Consider a two-player game defined by the payoff matrix on the left:

	left	right
top	1, 1	0, 0
bottom	0, 0	1, 1



The 3D plots on the right visualize the coarse correlated equilibrium and Nash equilibria of the game. The axes represent the probabilities of the joint actions: top and left, top and right, and bottom and left, respectively. The probability of the joint action bottom and right can be determined accordingly. The set of CCEs is represented by the green polytope, and the red points indicate the NEs. As one can see, all NEs lie on the boundary of the CCE polytope.

4.2

According to the definition of coarse correlated equilibrium, we have that for any action $a_1' \in A_1$ and $a_2' \in A_2$,

$$\mathbb{E}_{(a_1,a_2)\sim\mu}[u_1(a_1', a_2)] \leq \mathbb{E}_{(a_1,a_2)\sim\mu}[u_1(a_1, a_2)] \tag{6}$$

$$\mathbb{E}_{(a_1,a_2)\sim\mu}[u_2(a_1, a_2')] \leq \mathbb{E}_{(a_1,a_2)\sim\mu}[u_2(a_1, a_2)] \tag{7}$$

From the definition of two-player zero-sum game, we have that $u_1(a_1, a_2) = -u_2(a_1, a_2)$. Plugging this relation into (6), combining with (7), we have that for any action $a_1' \in A_1$ and $a_2' \in A_2$,

$$\mathbb{E}_{(a_1,a_2)\sim\mu}[u_2(a_1', a_2)] \geq \mathbb{E}_{(a_1,a_2)\sim\mu}[u_2(a_1, a_2)] \geq \mathbb{E}_{(a_1,a_2)\sim\mu}[u_2(a_1, a_2')]. \tag{8}$$

Consider the marginal distribution

$$x_{1,a_1} \coloneqq \sum_{a'_2 \in A_2} \mu_{a_1,a'_2}; \qquad x_{2,a_2} \coloneqq \sum_{a'_1 \in A_1} \mu_{a'_1,a_2}.$$

We have that for any action $a'_2 \in A_2$,

$$\begin{aligned} u_2(x_1, x_2) &= \sum_{a'_1 \in A_1} x_{1,a'_1} \cdot u_2(a'_1, x_2) \\ &= \sum_{a'_1 \in A_1} x_{1,a'_1} \cdot \mathop{\mathbb{E}}_{(a_1,a_2)\sim\mu} [u_2(a'_1, a_2)] \\ &\geq \sum_{a'_1 \in A_1} x_{1,a'_1} \cdot \mathop{\mathbb{E}}_{(a_1,a_2)\sim\mu} [u_2(a_1, a'_2)] \\ &= \mathop{\mathbb{E}}_{(a_1,a_2)\sim\mu} [u_2(a_1, a'_2)] = u_2(x_1, a'_2) \end{aligned}$$

where the first, second, and the last equations are given by the linearity of expectation, and the inequality follows from (8). This indicates that x_2 is a best response to strategy profile x_1 for Player 2. Similarly, we have that x_1 is a best response to x_2. As a result, we conclude that (x_1, x_2) is a Nash equilibrium.

Problem 5

5.1

Towards a contradiction, assume μ is a correlated equilibrium where Player i plays a dominated action $a'_i \in A_i$ with non-zero probability. That is, $\mu_{a'_i,a_{-i}} > 0$ for some $a_{-i} \in A_{-i}$. Denote by a^*_i the action that dominates a'_i. Consider the mapping $\phi_i : A_i \mapsto A_i$ where $\phi(a'_i) = a^*_i$ and $\phi(a_i) = a_i$ for any other $a_i \in A_i \setminus \{a'_i\}$.

Next, consider the improvement from deviating according to ϕ under the policy μ. We have:

$$\mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu} [u_i(\phi_i(a_i), a_{-i})] - \mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu} [u_i(a_i, a_{-i})] = \sum_{a_1 \in A_1} \cdots \sum_{a_n \in A_n} \mu_{a_1,\ldots,a_n} \left(u_i(\phi_i(a_i), a_{-i}) - u_i(a_i, a_{-i}) \right).$$

For any joint action $a_i \in A_i \setminus \{a'_i\}$ and $a_{-i} \in A_{-i}$, we have $u_i(\phi_i(a_i), a_{-i}) = u_i(a_i, a_{-i})$ by the definition of ϕ. Furthermore, for any $a_{-i} \in A_{-i}$, we have $u_i(\phi_i(a'_i), a_{-i}) = u_i(a^*_i, a_{-i}) > u_i(a'_i, a_{-i})$. Thus, we get:

$$\mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu} [u_i(\phi_i(a_i), a_{-i})] - \mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu} [u_i(a_i, a_{-i})] \geq \mu_{a'_i,a_{-i}} \left(u_i(\phi_i(a'_i), a_{-i}) - u_i(a'_i, a_{-i}) \right) > 0,$$

where the last inequality follows from the assumption $\mu_{a'_i,a_{-i}} > 0$.

This contradicts the assumption that μ is a correlated equilibrium. Therefore, a dominated action cannot be in the support of a correlated equilibrium.

5.2

Consider the following two-player game:

	left	right
top	6, 0	0, 0
middle	0, 0	6, 0
bottom	2, 0	8, 0

Table 1: Caption

It is clear that action bottom is dominated by action middle. Consider the strategy profile μ with support $\mu_{\text{top,left}} = \mu_{\text{middle,right}} = 0.5$. We verify that μ is a coarse correlated equilibrium. Since both actions are indifferent for Player 2, we only need to verify that Player 1 cannot deviate to improve their payoff. We first compute the expected payoff for Player 1:

$$\mathop{\mathbb{E}}_{(a_1,a_2)\sim\mu}[u_1(a_1,a_2)] = \mu_{\text{top,left}} \cdot u_1(\text{top},\text{left}) + \mu_{\text{middle,right}} \cdot u_1(\text{middle},\text{right}) = 6$$

For each action $a_1' \in A_1$, we have:

$$\begin{aligned}\mathop{\mathbb{E}}_{(a_1,a_2)\sim\mu}[u_1(\text{top},a_2)] &= \mu_{\text{top,left}} \cdot u_1(\text{top},\text{left}) + \mu_{\text{middle,right}} \cdot u_1(\text{top},\text{right}) = 3\\ \mathop{\mathbb{E}}_{(a_1,a_2)\sim\mu}[u_1(\text{middle},a_2)] &= \mu_{\text{top,left}} \cdot u_1(\text{middle},\text{left}) + \mu_{\text{middle,right}} \cdot u_1(\text{middle},\text{right}) = 3\\ \mathop{\mathbb{E}}_{(a_1,a_2)\sim\mu}[u_1(\text{bottom},a_2)] &= \mu_{\text{top,left}} \cdot u_1(\text{bottom},\text{left}) + \mu_{\text{middle,right}} \cdot u_1(\text{bottom},\text{right}) = 5\end{aligned}$$

In general, we have $\mathbb{E}_{(a_1,a_2)\sim\mu}[u_1(a_1',a_2)] \le \mathbb{E}_{(a_1,a_2)\sim\mu}[u_1(a_1,a_2)]$ for any action $a_1' \in A_1$, indicating that μ is a coarse correlated equilibrium where the dominated action is played with non-zero probability.

5.3

Consider the ratio between the mass MWU assigns to a dominated action a_i and the action a_i^*, which dominates a_i. We have that

$$\begin{aligned}\frac{x_{i,a_i}^{(t)}}{x_{i,a_i^*}^{(t)}} &= \frac{\exp\ \eta\ \ {}_{\tau<t}\, r_{i,a_i}(x_1^{(\tau)},\dots,x_n^{(\tau)})}{\exp\ \eta\ \ {}_{\tau<t}\, r_{i,a_i^*}(x_1^{(\tau)},\dots,x_n^{(\tau)})}\\ &= \exp\ \eta \underset{\tau<t}{\quad}\ r_{i,a_i}(x_1^{(\tau)},\dots,x_n^{(\tau)}) - r_{i,a_i^*}(x_1^{(\tau)},\dots,x_n^{(\tau)})\\ &= \exp\ \eta \underset{\tau<t}{\quad}\ u_i(a_i,x_{-i}^{(\tau)}) - u_i(a_i^*,x_{-i}^{(\tau)}) \quad ,\end{aligned}$$

where the last equality follows from the definition of regret $r_{i,a}$. Let

$$\epsilon := \inf_{x_{-i}}\ u_i(a_i^*,x_{-i}) - u_i(a_i,x_{-i})\ > 0$$

denote the minimal sub-optimality gap of the dominated action a_i. We have that

$$\lim_{t\to\infty} \frac{x_{i,a_i}^{(t)}}{x_{i,a_i^*}^{(t)}} \le \lim_{t\to\infty} \exp(-\eta t\epsilon) = 0.$$

Since x_i is a probability distribution, it satisfies that $0 \le x_{i,a_i^*}^{(t)} \le 1$. Thus, we conclude that $\lim_{t\to\infty} x_{i,a_i}^{(t)} = 0$, indicating that the probability that the algorithm plays the dominated action a_i converges to 0.

Problem 6

We present the performance results of the reference implementation below. Details of the implementation are omitted.

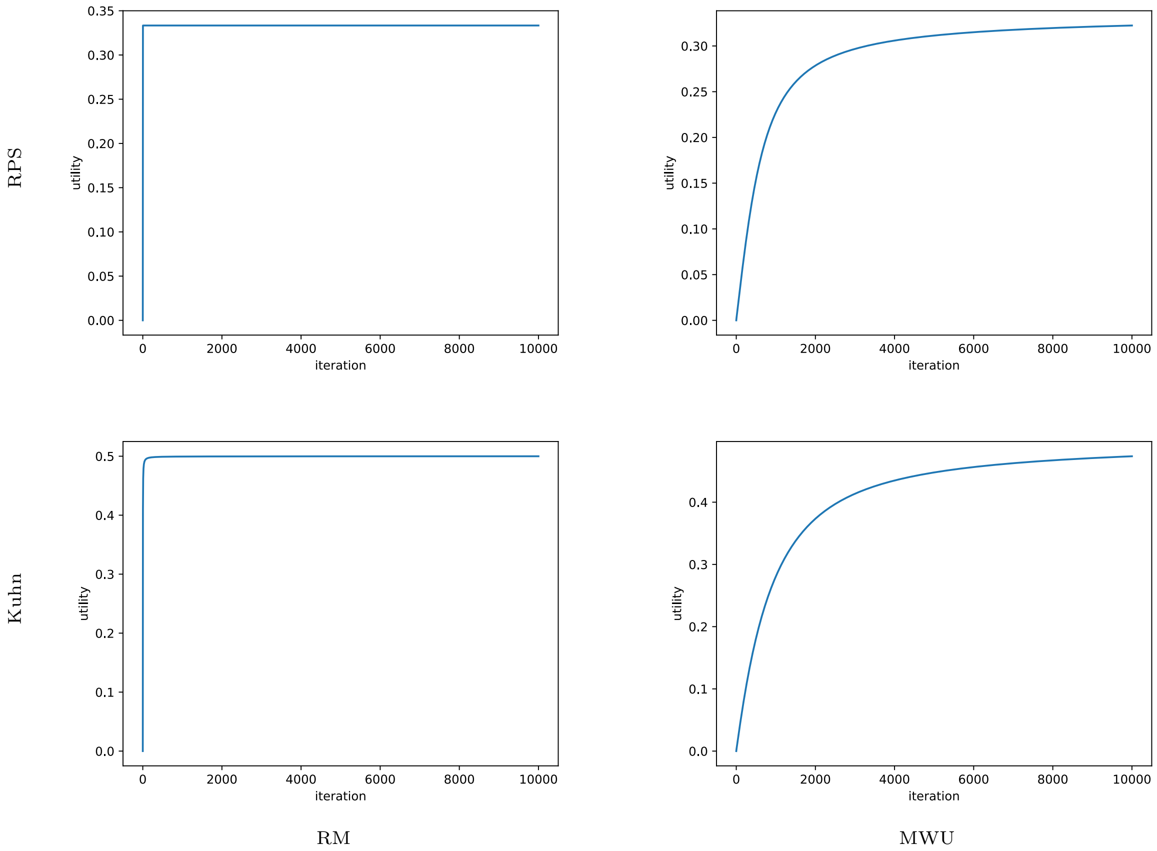


Table 2: Problem 6.1

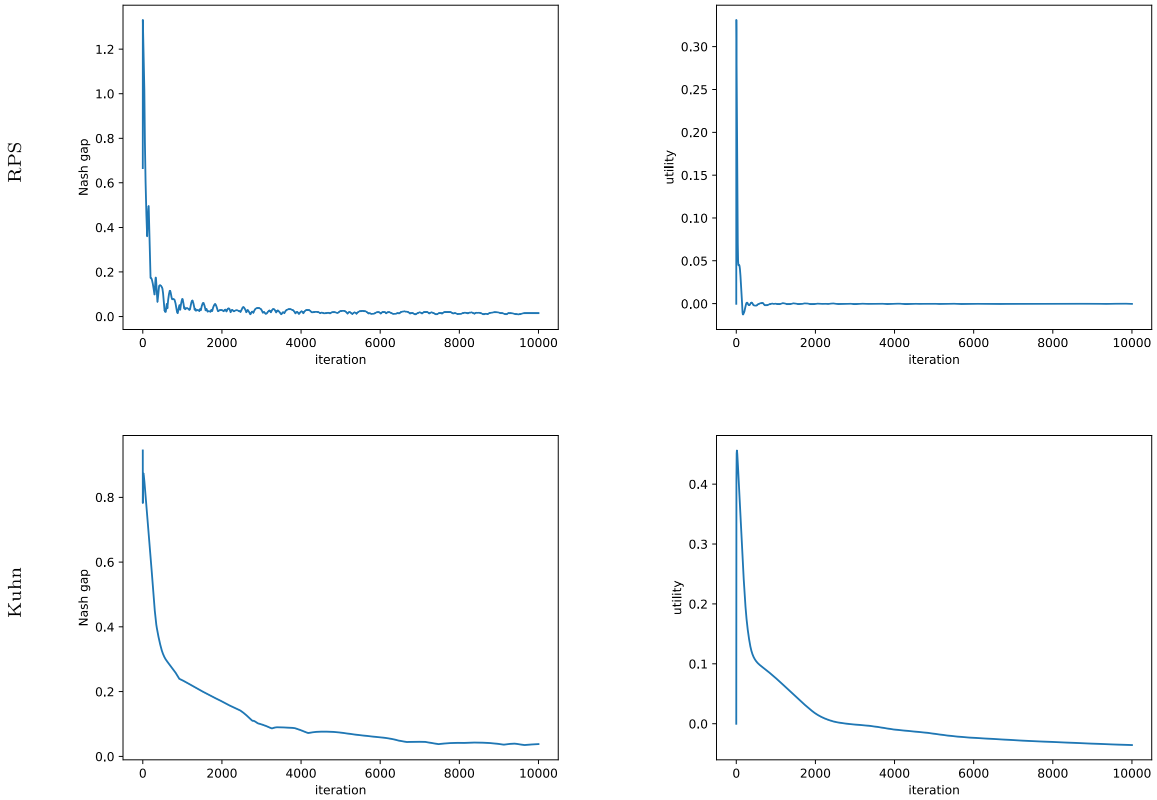


Table 3: Problem 6.2

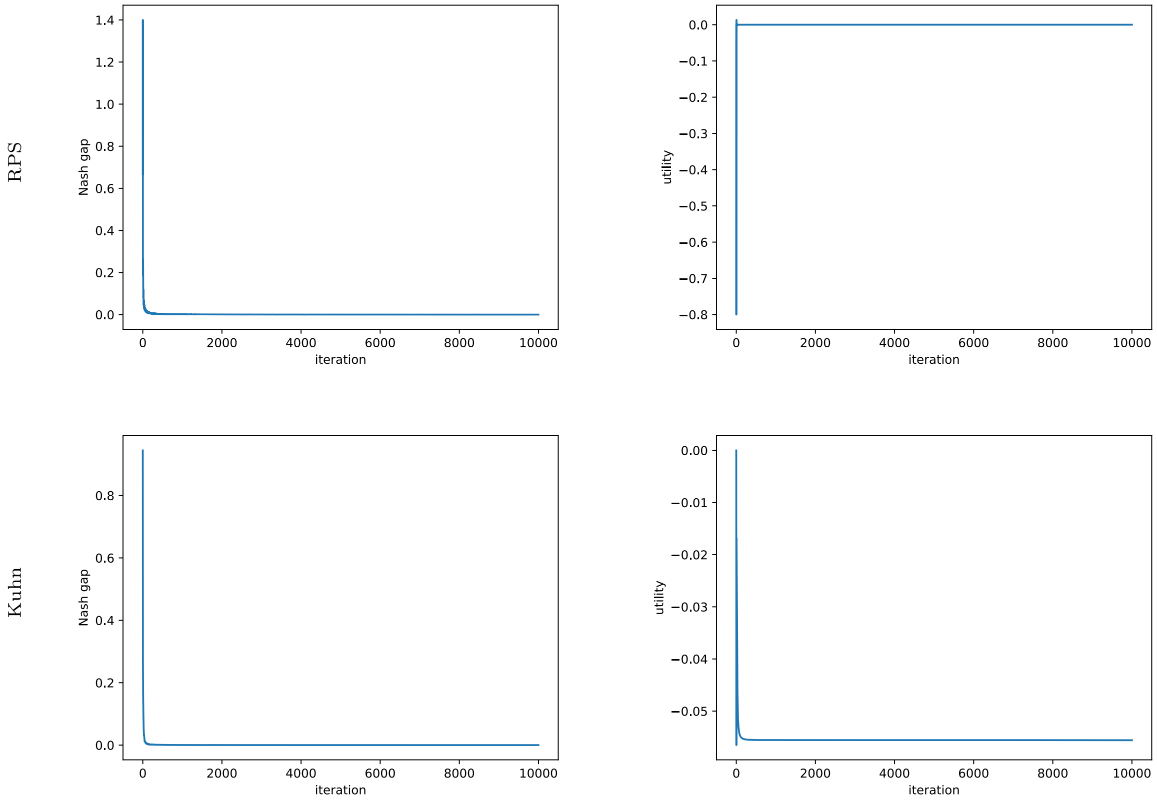


Table 4: Problem 6.3

6.S890 Topics in Multiagent Learning
Fall 2024