Solutions to Assignment 6 of CPSC 368/516 (Spring'23)

March 4, 2023

1 Problem 1

Problem 1.1. *In this problem we apply the Multiplicative Weight Update (MWU) framework to approximately find equilibria in two player zero-sum games. Consider the following theorem about the MWU method. (Here, $\Delta_n := \{p : \sum_{i=1}^n p_i = 1 \text{ and } p \geq 0\}$.)*

Theorem 1.2 (MWU Theorem). *Let $g^t \in \mathbb{R}^n$ be vectors with $\|g^t\|_\infty \leq 1$ for every $t = 0, \ldots, T-1$, and let $0 < \delta \leq \frac{1}{2}$. Then, starting with $p^0 := \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$, the MWU algorithm produces a sequence of probability distributions $p^1, \ldots, p^{T-1} \in \Delta_n$ such that*

$$\sum_{t=0}^{T-1} \langle p^t, g^t \rangle - \inf_{p \in \Delta_n} \sum_{t=0}^{T-1} \langle p, g^t \rangle \leq \frac{\ln n}{\delta} + \delta T.$$

Let $A \in \mathbb{R}^{n \times m}$ be a matrix with $A(i,j) \in [0,1]$ for all $i \in [n]$ and $j \in [m]$. We consider a game between two players: the row player and a column player. The game consists of one round in which the row player picks one row $i \in \{1, 2, \ldots, n\}$ and the column player picks one column $j \in \{1, 2, \ldots, m\}$. The goal of the row player is to minimize the value $A(i,j)$ which they pay to the column player after such a round, the goal of the column player is the opposite (to maximize the value $A(i,j)$).

The min-max theorem asserts that:

$$\max_{q \in \Delta_m} \min_{i \in \{1,\ldots,n\}} \mathop{\mathbb{E}}_{J \leftarrow q} A(i,J) = \min_{p \in \Delta_n} \max_{j \in \{1,\ldots,m\}} \mathop{\mathbb{E}}_{I \leftarrow p} A(I,j). \tag{1}$$

Here $\mathbb{E}_{I \leftarrow p} A(I,j)$ is the expected loss of the row player when using the randomized strategy $p \in \Delta_n$ against a fixed strategy $j \in \{1, 2, \ldots, m\}$ of the column player. Similarly, define $\mathbb{E}_{J \leftarrow q} A(i,J)$. Formally,

$$\mathop{\mathbb{E}}_{I \leftarrow p} A(I,j) := \sum_{i=1}^n p_i A(i,j) \quad \text{and} \quad \mathop{\mathbb{E}}_{J \leftarrow q} A(i,J) := \sum_{j=1}^m q_j A(i,j)$$

Let opt be the common value of the two quantities in (1) corresponding to two optimal strategies $p^\star \in \Delta_n$ and $q^\star \in \Delta_m$ respectively. Our goal is to use the MWU framework to construct, for any $\varepsilon > 0$, a pair of strategies $p \in \Delta_n, q \in \Delta_m$ such that:

$$\max_j \mathop{\mathbb{E}}_{I \leftarrow p} A(I,j) \leq \mathrm{opt} + \varepsilon \qquad \text{and} \qquad \min_i \mathop{\mathbb{E}}_{J \leftarrow q} A(i,J) \geq \mathrm{opt} - \varepsilon.$$

1. *Prove the following "easier" direction of Equation (1):*

$$\max_{q \in \Delta_m} \min_{i \in \{1,\ldots,n\}} \mathop{\mathbb{E}}_{J \leftarrow q} A(i,J) \leq \min_{p \in \Delta_n} \max_{j \in \{1,\ldots,m\}} \mathop{\mathbb{E}}_{I \leftarrow p} A(I,j).$$

1. *Give an algorithm, which given $p \in \Delta_n$ constructs a $j \in \{1, 2, \ldots, m\}$ which maximizes $\mathbb{E}_{I \leftarrow p} A(I,j)$. What is the running time of your algorithm? Show that for such a choice of j we have $\mathbb{E}_{I \leftarrow p} A(I,j) \geq \mathrm{opt}$.*

We will follow the MWU scheme with $p^0, \ldots, p^{T-1} \in \Delta_n$ and the vector g^t at step t being $g^t := Aq^t$, where $q^t := e_j$ with j chosen as to maximize $\mathbb{E}_{I \leftarrow p^t} A(I,j)$. (Recall that e_j is the vector with 1 at coordinate j and 0 otherwise.)

1. *Prove that* $\|g^t\|_\infty \le 1$ *and* $\langle p^\star, g^t\rangle \le \mathrm{opt}$ *for every* $t = 0, 1, \ldots, T-1$.
2. *Use the MWU theorem mentioned above to show that for* T *large enough:*

$$\mathrm{opt} \le \frac{1}{T}\sum_{t=0}^{T-1} \langle p^t, g^t\rangle \le \mathrm{opt} + \varepsilon.$$

What is the smallest value of T *which suffices for this to hold? Conclude that for some* $1 \le t < T$ *it holds that* $\max_j \mathbb{E}_{I\leftarrow p^t} A(I,j) \le \mathrm{opt} + \varepsilon$.

3. *Let* $q \coloneqq \frac{1}{T}\sum_{t=0}^{T-1} q^t$. *Prove that for* T *as in part (d):* $\min_i \mathbb{E}_{J\leftarrow q} A(i,J) \ge \mathrm{opt} - \varepsilon$.
4. *What is the total running time of the whole procedure to find an* ε*-approximate pair of strategies* p *and* q *we set out to find at the beginning of this problem?*

1.1 Part 1

Fix any strategy $p \in \Delta_n$, we claim that

$$\begin{aligned} \max_{j\in[m]} \mathop{\mathbb{E}}_{I\leftarrow p} A(I,j) &= \max_{j\in[m]} p^\top A e_j && (2)\\ &= \max_{q\in\Delta_m} p^\top A q. && (3)\end{aligned}$$

Direction 1 ((2)≤(3)). Let $j^\star \coloneqq \mathrm{argmax}_{j\in[m]}\, p^\top A e_j$. We have

$$\begin{aligned} \max_{j\in[m]} p^\top A e_j &= p^\top A e_{j^\star}\\ &\le \max_{q\in\Delta_m} p^\top A q. && \text{(Using that } e_{j^\star} \in \Delta_m)\end{aligned}$$

Direction 2 ((2)≥(3)). Fix any $q \in \Delta_m$. We have

$$\begin{aligned} p^\top A q &= \langle p^\top A, q\rangle\\ &\le \left\|p^\top A\right\|_\infty \|q\|_1 && \text{(Holder's inequality)}\\ &= \left\|p^\top A\right\|_\infty && (q \text{ is in the } m\text{-dimensional simplex})\\ &= \max_{j\in[m]} \left|(p^\top A)_j\right|\\ &= \max_{j\in[m]} (p^\top A)_j && \text{(Using that } p, A \ge 0)\\ &= \max_{j\in[m]} p^\top A e_j\\ &= \max_{j\in[m]} \mathop{\mathbb{E}}_{I\leftarrow p} A(I,j).\end{aligned}$$

Combining the two directions we have, for any $p \in \Delta_n$,

$$\max_{j\in[m]} \mathop{\mathbb{E}}_{I\leftarrow p} A(I,j) = \max_{q\in\Delta_m} p^\top A q.$$

Since this holds for any $p \in \Delta_n$, we get

$$\min_{p\in\Delta_n} \max_{j\in[m]} \mathop{\mathbb{E}}_{I\leftarrow p} A(I,j) = \min_{p\in\Delta_n} \max_{q\in\Delta_m} p^\top A q.$$

Similarly, it can be proved that

$$\max_{q\in\Delta_m} \min_{i\in[n]} \mathop{\mathbb{E}}_{J\leftarrow q} A(i,J) = \max_{q\in\Delta_m} \min_{p\in\Delta_n} p^\top A q.$$

Using these, the required inequality reduces to

$$\max_{q\in\Delta_m}\min_{p\in\Delta_n} p^\top Aq \le \min_{p\in\Delta_n}\max_{q\in\Delta_m} p^\top Aq.$$

Let $p_1 \in \Delta_n$ and $q_1 \in \Delta_m$ be the strategies achieving $\max_{q\in\Delta_m}\min_{p\in\Delta_n} p^\top Aq$. Similarly, let $p_2 \in \Delta_n$ and $q_2 \in \Delta_m$ be the strategies achieving $\min_{p\in\Delta_n}\max_{q\in\Delta_m} p^\top Aq$. Now, we are ready to prove the required inequality

$$\begin{aligned}
\max_{q\in\Delta_m}\min_{p\in\Delta_n} p^\top Aq &= (p_1)^\top Aq_1 \\
&\le (p_2)^\top Aq_1 && \text{(Since } p_1 \text{ minimizes } (p)^\top Aq_1 \text{ for } p\in\Delta_n) \\
&\le (p_2)^\top Aq_2 && \text{(Since } q_2 \text{ maximizes } (p_2)^\top Aq \text{ for } q\in\Delta_m) \\
&= \min_{p\in\Delta_n}\max_{q\in\Delta_m} p^\top Aq.
\end{aligned}$$

1.2 Part 2

Consider the following algorithm.

Input: A strategy $p \in \Delta_n$ and a matrix $A \in \mathbb{R}^{n\times m}$

Output: An index $j^\star \in [m]$

---

1. **Compute** $p^\top A$ (in $O(nm)$ arithmetic operations)
2. **Set** $j^\star = \text{argmax}_{j\in[m]}(p^\top A)_i$ (in $O(m)$ arithmetic operations)
3. **Output** $j^\star$.

The algorithm computes $\mathbb{E}_{I\leftarrow p} A(I,j)$ for all $j \in [m]$ and outputs the $j^\star$ that maximizes this value. The algorithm runs in time $O(nm)$ (the running time of each step is given in the algorithm). Since $p \in \Delta_n$ and $j^\star \in [m]$, we have

$$\mathop{\mathbb{E}}_{I\leftarrow p} A(I,j^\star) = \max_{j\in[m]}\mathop{\mathbb{E}}_{I\leftarrow p} A(I,j^\star) \ge \min_{p\in\Delta_n}\max_{j\in\{1,\ldots,m\}} \mathbb{E}_{I\leftarrow p}A(I,j) = y^\star. \tag{4}$$

1.3 Part 3

Subpart (a). For every $t = 0,1,\ldots,T-1$, $g^t = Aq^t$, where $q^t = q_j$ for some $j \in [m]$ chosen at each iteration. Fix any $i \in [n]$, we will bound $|g_i^t|$:

$$\begin{aligned}
\left|g_i^t\right| &= \left|\sum_{j\in[m]} A_{ij}q_j^t\right| \\
&\le \left|\sum_{j\in[m]} q_j^t\right| && \text{(Using that } 0 \le A_{ij} \le 1 \text{ and } q^t \ge 0) \\
&\le 1. && \text{(Using that } q^t \in \Delta_m)
\end{aligned}$$

Since each coordinate of g^t has absolute value at most 1, it follows that $\|g^t\|_\infty \le 1$.

Subpart (b). Recall that $p^\star \in \Delta_n$ is the strategy $\text{argmin}_{p\in\Delta_n}\max_{j\in\{1,\ldots,m\}}\mathbb{E}_{I\leftarrow p}A(I,j)$. Thus, we have

$$\max_{j\in\{1,\ldots,m\}} \mathbb{E}_{I\leftarrow p^\star} A(I,j) = \min_{p\in\Delta_n}\max_{j\in\{1,\ldots,m\}} \mathbb{E}_{I\leftarrow p}A(I,j) = y^\star.$$

Further, for any $j^\star \in [m]$

$$\mathbb{E}_{I \leftarrow p^\star} A(I, j^\star) \leq \max_{j \in \{1,\ldots,m\}} \mathbb{E}_{I \leftarrow p^\star} A(I, j) = y^\star.$$

The claim follows since at each t, $g^t = e_j$ for some j, and hence,

$$\left\langle p^\star, g^t \right\rangle = (p^\star)^\top A e_j = \mathbb{E}_{I \leftarrow p^\star} A(I, j^\star) \leq y^\star. \tag{5}$$

1.4 Part 4

First observe that

$$\begin{aligned}
\min_{p \in \Delta_n} \frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p \right\rangle &\leq \frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p^\star \right\rangle && \text{(Since } p^\star \in \Delta_n\text{)} \\
&\leq \frac{1}{T} \sum_{t=0}^{T-1} y^\star && \text{(Using Equation (5))} \\
&= y^\star. && (6)
\end{aligned}$$

Now, since at each iteration $\|g^t\|_\infty \leq 1$, we can apply Theorem 1.2. Applying Theorem 1.2, we have that for $T = \Theta\left(\frac{\log n}{\varepsilon}\right)$.

$$\frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p^t \right\rangle - \min_{p \in \Delta_n} \frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p \right\rangle \leq \varepsilon.$$

Combining the above equation with Equation (6), we get

$$\frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p^t \right\rangle \leq \min_{p \in \Delta_n} \frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p \right\rangle + \varepsilon \leq y^\star + \varepsilon. \tag{7}$$

Since the minimum is at most the average, we have

$$\min_{t \in \{0,\ldots,T-1\}} \left\langle g^t, p^t \right\rangle \leq \frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p^t \right\rangle \leq y^\star + \varepsilon.$$

It remains to lower bound $\frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p^t \right\rangle$. Towards this, observe that at each iteration $t = 0, \ldots, T-1$.

$$\begin{aligned}
\left\langle p^t, g^t \right\rangle &= (p^t)^\top A q^t \\
&= \max_{j \in [m]} \mathop{\mathbb{E}}_{I \leftarrow p^t} A(I, j) \\
&\geq y^\star. && \text{(Using Equation (4))}
\end{aligned}$$

Thus,

$$\frac{1}{T} \sum_{t=0}^{T-1} \left\langle g^t, p^t \right\rangle \geq \frac{1}{T} \sum_{t=0}^{T-1} y^\star = y^\star. \tag{8}$$

1.5 Part 5

Let $g \coloneqq Aq$, i.e., $g = \frac{1}{T}\sum_{t=0}^{T-1} Aq^t$. Then we have

$$\begin{aligned}
\min_{i\in[n]} \langle g, e_i\rangle + \varepsilon = \min_{p\in\Delta_n} \langle g, p\rangle + \varepsilon & \\
= \min_{p\in\Delta_n} \frac{1}{T}\sum_{t=0}^{T-1} \left\langle g^t, p\right\rangle + \varepsilon & \\
\geq \frac{1}{T}\sum_{t=0}^{T-1} \left\langle g^t, p^t\right\rangle & \qquad \text{(Using Equation (7))} \\
\geq y^\star. & \qquad \text{(Using Equation (8))}
\end{aligned}$$

On rearranging we get $\min_{i\in[n]} \langle g, p\rangle \geq y^\star - \varepsilon$.

1.6 Part 6

The total iterations of the MWU algorithm is $T = \Theta\left(\varepsilon^{-1}\log n\right)$. In each iteration, it takes $O(nm)$ arithmetic operations to compute g^t. All other steps in an iteration require $O(n)$ arithmetic operations. Thus, the total number of arithmetic operations is $\Theta\left(\varepsilon^{-1} nm\log n\right)$.

2 Problem 2

Problem 2.1. *Consider a general linear feasibility problem that asks for a point x satisfying a system of inequalities*

$$\langle a_i, x\rangle \geq b_i$$

for $i = 1, 2, \ldots, m$, where $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$ and $b_1, b_2, \ldots, b_m \in \mathbb{R}$. The goal of this problem is to give an algorithm that, given an error parameter $\varepsilon > 0$, outputs a point x such that

$$\langle a_i, x\rangle \geq b_i - \varepsilon \tag{9}$$

for all i whenever there is a solution to the above system of inequalities. We also assume the existence of an oracle *that, given vector $p \in \Delta_m$, solves the following relaxed problem: does there exist an x such that*

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_i a_{ij} x_j \geq \sum_{i=1}^{m} p_i b_i. \tag{10}$$

Assume that when the oracle returns a feasible solution for a p, the solution x that it returns is not arbitrary but has the following property:

$$\max_i |\langle a_i, x\rangle - b_i| \leq 1.$$

Prove the following theorem:

Theorem 2.2. *There is an algorithm that, if there exists an x such that $\langle a_i, x\rangle \geq b_i$ for all i, outputs an $\bar{x}$ that satisfies (9). The algorithm makes at most $O\left(\frac{\ln m}{\varepsilon^2}\right)$ calls to the oracle for the problem mentioned in (10).*

We will use the MWU method (the version in Theorem 1.2) to design an algorithm that satisfies the claim in Theorem 2.2. Concretely, we claim that the following algorithm suffices.

- **Input:**
  - A number $\varepsilon > 0$,
  - a matrix $A \in \mathbb{R}^{m \times n}$,
  - an oracle $\mathcal{O}$ that, given $p \in \Delta_m$ outputs a point x that satisfies Equation (10), and
  - an the MWU oracle $\mathcal{M}$ which given vectors $g^t \in \mathbb{R}^m$ and $w^t \in \Delta_m$ with $\|g^t\|_\infty \leq 1$ and $0 < \delta \leq \frac{1}{2}$ outputs a point $p^t \in \mathbb{R}^m$ that satisfies Theorem 1.2
- **Output:** A point $\overline{x} \in \mathbb{R}^m$ that satisfies Equation (9)

---

1. Set $T \coloneqq \frac{4 \ln m}{\varepsilon^2}$, $\delta \coloneqq \sqrt{\frac{\ln m}{T}}$, and $p^0 \coloneqq \left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right) \in \Delta_m$
2. **For** $t \in \{0, 1, 2, \ldots, T-1\}$ **do:**
   (a) Query $x^t \coloneqq \mathcal{O}(p^t)$
   (b) Set $g^t \coloneqq Ax^t - b$
   (c) Query $p^{t+1} \coloneqq \mathcal{M}\left(g^t, p^t, \delta\right)$
3. **return** $\overline{x} \coloneqq \frac{1}{T}\sum_{t=0}^{T-1} x^t$

Let A be the $m \times n$ matrix whose i-th row is a_i. Theorem 1.2 implies the following inequality

$$\inf_{p \in \Delta_m} \sum_{t=0}^{T-1} \frac{\langle p, g^t \rangle}{T} \geq \sum_{t=0}^{T-1} \frac{\langle p^t, g^t \rangle}{T} - \frac{\ln m}{\delta T} - \delta.$$

Consequently, the following inequality holds

$$\min_{i \in [m]} \sum_{t=0}^{T-1} \frac{g_i^t}{T} \geq \sum_{t=0}^{T-1} \frac{\langle p^t, g^t \rangle}{T} - \frac{\ln m}{\delta T} - \delta. \tag{11}$$

To see how the above inequality implies the proof, observe that:

$$\begin{aligned}
\forall_{i \in [m]}, \quad (A\overline{x} - b)_i &= \frac{1}{T}\sum_{t=0}^{T-1} \left(Ax^t - b\right)_i && \text{(Using that } \overline{x} \coloneqq \tfrac{1}{T}\textstyle\sum_{t=0}^{T-1} x^t\text{)} \\
&= \frac{1}{T}\sum_{t=0}^{T-1} g_i^t \\
&\geq \sum_{t=0}^{T-1} \frac{\langle p^t, g^t \rangle}{T} - \frac{\ln m}{\delta T} - \delta && \text{(Using Equation (11))} \\
&= \sum_{t=0}^{T-1} \frac{\langle p^t, Ax^t - b \rangle}{T} - \frac{\ln m}{\delta T} - \delta \\
&\geq 0 - \frac{\ln m}{\delta T} - \delta \\
& && \text{(Using the construction of } x^t \text{ and } p_t\text{, and the fact that } \mathcal{O} \text{ satisfies Equation (10))} \\
&= \varepsilon. && \text{(Using that } \delta = \sqrt{\tfrac{\ln m}{T}} \text{ and } T = \tfrac{4 \ln m}{\varepsilon^2}\text{)}
\end{aligned}$$

The above inequality is equivalent to Equation (9).