# Incentives and Machine Learning (CSC 482A/581A)
# Lectures 16–18

Nishant Mehta

## 1 Finite games and Nash equilibria

Recall that in an $n$-player normal form game, each player $i \in [n]$ has a set $A_i$ of strategies and a cost function $c_i : A_1 \times \cdots \times A_n \to \mathbb{R}$. If each player plays a mixed strategy $p_i \in \Delta(A_i)$, then we call the joint probability distribution $p = p_1 \times \cdots p_n$ a mixed strategy profile. Player $i$'s cost given strategy profile $p$ is defined as $c_i(p) = \mathsf{E}_{a \sim p}[c_i(a)]$. Assuming the game is finite so each set $A_i$ is discrete, and using our usual notation $A = A_1 \times \ldots A_n$, this expectation can be rewritten as

$$\mathsf{E}_{a \sim p}[c_i(a)] = \sum_{a \in A} p(a) c_i(a) = \sum_{a=(a_1,\ldots,a_n) \in A} p_1(a_1) \cdots p_n(a_n) \cdot c_i(a_1, \ldots, a_n).$$

In the first week of the course, we saw that there are finite games that do not have any pure strategy Nash equilibria. On the other hand, Von Neumann's minimax theorem shows that every finite game does have a mixed strategy Nash equilibrium. Given the existence of a mixed strategy Nash equilibrium (hereafter, abbreviated to Nash equilibrium), one wonders whether the players of a game, by repeatedly interacting with their opponents, could figure out how to converge to a Nash equilibrium.

Unfortunately, it would appear that the answer for general finite games is no: the problem of approximating a Nash equilibrium is PPAD-complete. In this course, we won't worry too much about the meaning of the complexity class PPAD. The rough message here is that for general, finite games, we do not know of polynomial-time algorithms for approximating Nash equilibria. If indeed there is no polynomial time algorithm for approximating Nash equilibria, then there is no reason to believe that $n$ agents, each with polynomial computation, will be able to interact sequentially in such a way as to converge to a Nash equilibrium.

This all might seem like very bad news. Indeed, what is the point of a stability concept like a Nash equilibrium if agents (or even a coordinator) cannot actually figure out how to reach this equilibrium? For example, would we care about such downstream questions like what the social cost of an equilibrium is if the agents will never find that equilibrium? This bad news is just one of the reasons that, a bit later in the course, we'll switch to looking at relaxations of mixed strategy Nash equilibria such as correlated equilibria and coarse correlated equilibria. These latter concepts are also of interest in their own right.

In this lecture, we will focus on the special subclass of games that are two-player zero-sum games. For these games, it *is* possible to tractably approximate Nash equilibria; moreover, this approximation can happen simply by the two agents repeatedly playing the game and adapting their strategies. Specifically, we'll look at a method called No-Regret Dynamics. In this method, each player plays a no-regret algorithm[1] for a specific online learning game derived from the two-player

[1] Recall that an online learning algorithm is no-regret if its average regret decays to zero, or, equivalently, its regret is sublinear.

zero-sum game. We then form a mixed strategy profile by taking, for each player, the empirical frequency distribution of the player's strategies. Remarkably, we'll show that this strategy profile is an approximate Nash equilibrium.

## 2 Two-player zero-sum games

In a *zero-sum game*, the losses of the players sum to zero. Therefore, in a two-player zero-sum game, the loss of the second player is simply the negation of the loss of the first player. Since there are only two players, let us adopt simpler notation. We don't need to use $n$ to refer to the number of players anymore, so we let $m = |A_1|$ and $n = |A_2|$; i.e., Player 1 has $m$ strategies and Player 2 has $n$ strategies. Also, since there are only two players, we won't use $p$ to refer to a joint mixed strategy profile. Instead, we will use $p$ (instead of $p_1$) for Player 1's mixed strategy and $q$ (instead of $p_2$) for Player 2's mixed strategy. Finally, since there are only two players, for any strategy profile $(p, q)$, we can express the cost of each player via a game matrix $M \in [0,1]^{m \times n}$. Under pure strategy profile $(i, j) \in A_1 \times A_2$, Player 1's cost is $M_{ij}$ and Player 2's cost is $-M_{ij}$. More generally, for mixed strategy profile $(p, q)$, Player 1's expected cost is $p^\top M q$ and Player 2's expected cost is $-p^\top M q$. Therefore, from the perspective of Player 1's cost, Player 1 can be viewed as the "min" player and Player 2 can be viewed as the "max" player.

Now, suppose that $(\bar{p}, \bar{q})$ is a Nash equilibrium. Since both players are best-responding, we have both

$$\bar{p}^\top M \bar{q} = \min_{p \in \Delta_m} p^\top M \bar{q} \leq p^\top M \bar{q} \quad \text{for all } p \in \Delta_m$$

and

$$\bar{p}^\top M \bar{q} = \max_{q \in \Delta_n} \bar{p}^\top M q \geq \bar{p}^\top M q \quad \text{for all } q \in \Delta_n$$

Consequently, we have the sequence of inequalities

$$\begin{aligned}
\min_{p \in \Delta_m} \max_{q \in \Delta_n} p^\top M q &\leq \max_{q \in \Delta_n} \bar{p}^\top M q \\
&= \bar{p}^\top M \bar{q} \\
&= \min_{p \in \Delta_m} p^\top M \bar{q} \leq \max_{q \in \Delta_n} \min_{p \in \Delta_m} p^\top M q.
\end{aligned}$$

Also, it is easy to show the opposite inequality (try to show this as an exercise):

$$\max_{q \in \Delta_n} \min_{p \in \Delta_m} p^\top M q \leq \min_{p \in \Delta_m} \max_{q \in \Delta_n} p^\top M q.$$

Hence, the existence of a Nash equilibrium $(\bar{p}, \bar{q})$ implies the fundamental *equality*:

$$\min_{p \in \Delta_m} \max_{q \in \Delta_n} p^\top M q = \max_{q \in \Delta_n} \min_{p \in \Delta_n} p^\top M q. \tag{1}$$

When this equality is satisfied (as it is when we have a Nash equilibrium), we call the quantity on either side of the equality the *value of the game*.

Results showing equalities of the form (1) are known as minimax theorems. Thus far, we haven't actually shown that (1) holds for two-player zero-sum games. We have only shown (1) holds if there exists a Nash equilibrium.

A fundamental result of game theory is that for two-player zero-sum games, (1) holds.

**Theorem 1** (Von Neumann's Minimax Theorem[2])**.** *For a two-player zero-sum game, we have*

$$\min_{p\in\Delta_m}\max_{q\in\Delta_n} p^\top Mq = \max_{q\in\Delta_n}\min_{p\in\Delta_m} p^\top Mq.$$

Various proofs of the theorem exist that make use of fixed-point theorems. Our proof of the theorem will make use of no-regret learning algorithms. The proof also is constructive: it will reveal a Nash equilibrium, and as we will see, it will also yield an algorithm for efficiently finding an approximate Nash equilibrium.

# 3 From no-regret learning to approximate Nash equilibria

We begin this section by proving Theorem 1. After that, we show how the proof algorithmically gives us an approximate Nash equilibrium. We will call the algorithmic procedure developed in the proof No-Regret Dynamics.

## 3.1 The proof

*Proof (of Theorem 1).* The idea of the proof is to have the players play a repeated game over $T$ rounds. From the perspective of each player, this game will be an instance of Decision-Theoretic Online Learning, and each player will run a no-regret algorithm. Letting $p_1, \ldots, p_T$ be the sequence of strategies of Player 1 and $q_1, \ldots, q_T$ be the sequence of strategies of Player 2, the loss of Player 1 in round $t$ will be $\ell_t^{(\mathrm{I})} := Mq_t$ while the loss of Player 2 in round $t$ will be $\ell_t^{(\mathrm{II})} := -M^\top p_t$.

Let us assume that Player 1's no-regret algorithm has average regret at most $\varepsilon_{T,1} = o(1)$, so

$$\frac{1}{T}\sum_{t=1}^{T} p_t^\top Mq_t \leq \min_{p\in\Delta_m} \frac{1}{T}\sum_{t=1}^{T} pMq_t + \varepsilon_{T,1}. \tag{2}$$

Similarly, let us assume Player 2's (possibly different) no-regret algorithm has average regret at most $\varepsilon_{T,2} = o(1)$, so

$$\max_{q\in\Delta_n} \frac{1}{T}\sum_{t=1}^{T} pMq_t \leq \frac{1}{T}\sum_{t=1}^{T} p_t^\top Mq_t + \varepsilon_{T,2}. \tag{3}$$

Let $\bar{p} = \frac{1}{T}\sum_{t=1}^{T} p_t$ and $\bar{q} = \frac{1}{T}\sum_{t=1}^{T} q_t$ be the simple averages of Player 1's and Player 2's plays respectively. Observe that

$$\begin{aligned}
\min_{p\in\Delta_m}\max_{q\in\Delta_n} p^\top Mq &\leq \max_{q\in\Delta_n} \bar{p}^\top Mq \\
&= \max_{q\in\Delta_n} \frac{1}{T}\sum_{t=1}^{T} p_t^\top Mq \\
&\leq \frac{1}{T}\sum_{t=1}^{T} p_t^\top Mq_t + \varepsilon_{T,2} \qquad\qquad (4)\\
&\leq \min_{p\in\Delta_m} \frac{1}{T}\sum_{t=1}^{T} p^\top Mq_t + \varepsilon_{T,1} + \varepsilon_{T,2} \\
&= \min_{p\in\Delta_m} p^\top M\bar{q} + \varepsilon_{T,1} + \varepsilon_{T,2} \qquad \leq \max_{q\in\Delta_n}\min_{p\in\Delta_m} p^\top Mq + \varepsilon_{T,1} + \varepsilon_{T,2},
\end{aligned}$$

[2] Here, we only present von Neumann's minimax theorem in a simplified form. The full version holds in the more general situation where $p^\top Mq$ is replaced by $f(x, y)$ for a function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ that is convex in $x$ (for fixed $y$), concave in $y$ (for fixed $x$), and for $\mathcal{X}$ and $\mathcal{Y}$ compact convex subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively.

where the first inequality applies (3) and the second inequality applies (2).

Taking $T \to \infty$ implies that $\min_{p\in\Delta_m} \max_{q\in\Delta_n} p^\top Mq \le \max_{q\in\Delta_n} \min_{p\in\Delta_m} p^\top Mq$, as desired. □

### 3.2 Using the proof to find an approximate Nash equilibrium

The above proof, as short as it is, actually does more than prove Theorem 1. It also shows a way to find an $\varepsilon$-approximate (mixed strategy) Nash equilibrium (and, taking $T \to \infty$, it recovers an actual Nash equilibrium).

Recall that a mixed strategy profile $(\bar{p}, \bar{q})$ is an $\varepsilon$-approximate Nash equilibrium if both Player 1's strategy is an $\varepsilon$-best response, i.e.,

$$\bar{p}^\top M\bar{q} \le \min_{p\in\Delta_m} p^\top M\bar{q} + \varepsilon \tag{5}$$

and Player 2's strategy is an $\varepsilon$-best response, i.e.,

$$\bar{p}^\top M\bar{q} \ge \max_{q\in\Delta_n} \bar{p}^\top Mq - \varepsilon. \tag{6}$$

Both these inequalities can easily be shown to hold using (4). For (5), observe that

$$\begin{aligned}\bar{p}^T M\bar{q} &\le \max_{q\in\Delta_n} \bar{p}^T Mq \\ &\le \min_{p\in\Delta_m} p^T M\bar{q} + \varepsilon_{T,1} + \varepsilon_{T,2},\end{aligned}$$

where the first line is trivial and the second line is from (4). Likewise, for (6),

$$\begin{aligned}\bar{p}^T M\bar{q} &\ge \min_{p\in\Delta_m} p^T M\bar{q} \\ &\ge \max_{q\in\Delta_n} \bar{p}^T Mq - \varepsilon_{T,1} - \varepsilon_{T,2}.\end{aligned}$$

where again the first line is trivial and the second line is from (4).

Hence, $(\bar{p}, \bar{q})$ is an $(\varepsilon_{T,1} + \varepsilon_{T,2})$-approximate Nash equilibrium.

### 3.3 A concrete instantiation of the proof

The proof of Theorem 1 is somewhat abstract in that we did not specify a particular pair of no-regret learning algorithms to use, and as such, we did not specify the average regret bound sequences $(\varepsilon_{T,1})_{T\ge1}$ and $(\varepsilon_{T,2})_{T\ge1}$. Let us now make things concrete. For Player 1, recall from the proof of Theorem 1 that the loss vector in round $t$ is $\ell_t^{(\mathrm{I})} = Mq_t$ with $M \in [0,1]^{m\times n}$ and $q_t \in \Delta_n$. Therefore, we have $\ell_t^{(\mathrm{I})} \in [0,1]^m$, and we may use Hedge which, from Section 3 of the previous lecture, enjoys an average regret bound of $\varepsilon_{T,1} = 2\sqrt{\frac{\log m}{T}}$ since the number of actions is now $m$. Similarly, for Player 2, we have $\ell_t^{(\mathrm{II})} = -M^T p_t$ with $p_t \in \Delta_m$, so we have $\ell_t^{(\mathrm{II})} \in [-1,0]^n$. It is not hard to see that the proof of the regret bound for Hedge still goes through when the loss range is $[-1,0]$ (I suggest you verify this), and since the number of actions is now $n$, using Hedge once gives an average regret bound of $\varepsilon_{T,2} = 2\sqrt{\frac{\log n}{T}}$.

Therefore, running No-Regret Dynamics (the method from the proof) for $T$ rounds, we have that the resulting pair $(\bar{p}, \bar{q})$ is an $\frac{2(\sqrt{\log m}+\sqrt{\log n})}{T}$-approximate Nash equilibrium. A natural question is if this is the fastest rate of convergence possible. To pose this question in a sensible way, one has to be careful about the scope of protocols considered. When running Hedge, note that each player:

- does not know how many actions the other player has;
- does not know the game matrix $M$ but instead knows only her loss vector induced from the opponent's mixed strategy;
- uses only a constant amount of private storage, where the constant is proportional to her number of actions; basically, Hedge can accomplish its computation by storing a constant number of (possibly cumulative) loss vectors.

In this sense, the distributed protocol — distributed among the two players — is strongly uncoupled.

It turns out that for protocols of the above type, it is possible to achieve a much faster convergence rate of order $O\left(\frac{\log m + \log n}{T}\right)$. In the interest of time, we won't look into how to achieve faster rates of convergence here, but if interested, just ask, and I'll direct you to lecture notes and papers.