17-740 Spring 2025
Algorithmic Foundations of Interactive Learning

Lecture 7: Computing Equilibria II

Lecturer: Steven Wu
Scribe: Eryn Ma, Naveen Raman, Jingwu Tang

7.1 Approximate Equilibria

7.1.1 Recap: Minimax Theorem

In this lecture, we aim to extend our prior discussion of the minimax theorem to consider approximate equilibria in situations where $\mathcal{X}$ and $\mathcal{Y}$ are convex sets. To begin, first recall the minimax theorem:

$$\max_{x\in\Delta(R)} \min_{y\in\Delta(C)} x^T My = \min_{y\in\Delta(C)} \max_{x\in\Delta(R)} x^T My = \mathrm{Val}(M) \tag{7.1}$$

We can naturally extend the two-player games to situations where the actions x and y are chosen from convex and compact (closed + bounded) sets $\mathcal{X}$ and $\mathcal{Y}$. Now, the utility or payoff can be represented as $U(x, y)$ rather than $x^T My$. We consider the setting in which the utility function U is concave in x and convex in y. Under this scenario, the minimax theorem contines to hold:

$$\max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} U(x, y) = \min_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} U(x, y) \tag{7.2}$$

7.1.2 ϵ-Nash Equilibria

We now aim to extend to this to scenarios that are *approximate* Nash Equilibria, or more formally, ϵ-Nash Equilibria. We first define ϵ-Nash Equilibria as follows:

ϵ-Nash Equilibrium - A pair of actions, $\hat{x}$, $\hat{y}$ is an ϵ Nash Equilibrium if

$$U(\hat{x}, \hat{y}) - \min_y U(\hat{x}, y) \le \epsilon \tag{7.3}$$

$$\max_x U(\hat{x}, y) - U(\hat{x}, \hat{y}) \le \epsilon \tag{7.4}$$

A natural way to find an ϵ-Nash Equilibrium is for two players to play a no-regret strategy. We will formally prove that statement:

Theorem 1 *Suppose two players play No Regret algorithms. That is, let* $\text{Reg}_X = \max_{x \in \mathcal{X}} \sum_{t=1}^T U(x, y^t) - U(x^t, y^t)$*, and similarly let* $\text{Reg}_Y = \min_{y \in \mathcal{Y}} \sum_{t=1}^T U(x^t, y^t) - U(x^t, y)$*. Then the strategy* $\bar{x} = \frac{1}{T} \sum_{t=1}^T x^t$ *and* $\bar{y} = \frac{1}{T} \sum_{t=1}^T y^t$ *is a* $\frac{\text{Reg}_X + \text{Reg}_Y}{T}$ *Nash Equilibrium.*

Proof:

$$\frac{\text{Reg}_X + \text{Reg}_Y}{T} \tag{7.5}$$

$$= \frac{1}{T}(\min_{y \in \mathcal{Y}} \sum_{t=1}^T U(x^t, y^t) - U(x^t, y) + \max_{x \in \mathcal{X}} \sum_{t=1}^T U(x, y^t) - U(x^t, y^t)) \tag{7.6}$$

$$= \frac{1}{T} \sum_{t=1}^T \max_{x \in \mathcal{X}} U(x, y^t) - \frac{1}{T} \sum_{t=1}^T \min_y U(x^t, y) \tag{7.7}$$

$$= \frac{1}{T} \sum_{t=1}^T \max_{x \in \mathcal{X}} U(x, y^t) - U(\bar{x}, \bar{y}) + U(\bar{x}, \bar{y}) - \frac{1}{T} \sum_{t=1}^T \min_y U(x^t, y) \tag{7.8}$$

$$\geq \max_{x \in \mathcal{X}} U(x, \bar{y}) - U(\bar{x}, \bar{y}) + U(\bar{x}, \bar{y}) - \min_y U(\bar{x}, y) \tag{7.9}$$

The last inequality is from Jensen's inequality, and this proves that we achieve a Nash Equilibrium bounded by the sum of the regrets. Therefore, playing No Regret algorithms can lead to a small ϵ Nash Equilibrium. ∎

7.2 Algorithms

Note that Theorem 1 holds regardless of the specific algorithms the players follow. We now introduce two types of dynamic strategies that ensure the average play of the players converges to an ϵ-NE.

7.2.1 No-Regret vs. No-Regret (NRNR) Dynamics

The first type of dynamics is no-regret vs. no-regret (NRNR), where both players follow no-regret algorithms. A typical example is when both players employ Follow-the-Regularized-Leader (FTRL).

Max player: $x_t = \arg\min_{x \in \mathcal{X}} - \sum_{\tau=1}^{t-1} U(x, y^\tau) + R(x)$

Min player: $y_t = \arg\min_{y \in \mathcal{Y}} \sum_{\tau=1}^{t-1} U(x^\tau, y) + R(y)$

From the regret bound of FTRL, we know that both $\mathrm{Reg}_X, \mathrm{Reg}_Y = O(\sqrt{T})$. Consequently, the average play of both players converges to an approximate Nash equilibrium.

7.2.2 No-Regret vs. Best-Response (NRBR) Dynamics

The second type of dynamics is no-regret vs. best-response (NRBR), where one player follows a no-regret algorithm while the other plays the best response to the no-regret player's strategy. We consider a case where the x-player plays FTRL while the y-player plays the best response.

Max player: $x_t = \arg\min_{x \in \mathcal{X}} - \sum_{\tau=1}^{t-1} U(x, y^\tau) + R(x)$

Min player: $y^t = \arg\min_{y \in \mathcal{Y}} U(x^t, y)$

From the regret bound of FTRL, we have $\mathrm{Reg}_X = O(\sqrt{T})$. Since y_t always plays the best response, we have $\mathrm{Reg}_Y \leq 0$. Therefore, the average play of the players will converge to an approximate Nash equilibrium.

7.2.3 Revisit Max Entropy

We consider the following maximum entropy problem:

$$\begin{aligned} \min_{p \geq 0} \quad & \sum_i p(i) \ln \frac{1}{p(i)} := \mathcal{H}(p) \quad \text{(OPT)} \\ \text{s.t.} \quad & \sum_i p(i) f(i) = c \\ & \sum_i p(i) = 1 \end{aligned} \tag{7.10}$$

To solve the constraint optimization problem, we introduce the Lagrangian dual variables γ and λ, we write the Lagrangian:

$$\min_{\gamma, \lambda} \max_{p \geq 0} \mathcal{H}(p) - \gamma \left(\sum_i p(i) - 1 \right) - \lambda \left(\sum_i p(i) f(i) - c \right) \tag{7.11}$$

Here, γ and λ penalize deviations from the constraints, ensuring that:

$$\sum_i p(i) - 1 = 0, \quad \sum_i p(i) f(i) = c. \tag{7.12}$$

Now, we can view it as a minimax game, and computing the minimax equilibrium with the NRBR dynamics introduced above.

$$U(p,(\gamma,\lambda)) := \mathcal{H}(p) - \gamma\left(\sum_i p(i) - 1\right) - \lambda\left(\sum_i p(i)f(i) - c\right) \tag{7.13}$$

- The dual variables γ and λ are updated using no-regret algorithms, specifically online gradient descent.

- The primal variable p follows a best-response dynamics given by:

$$p^t(i) \propto \exp(-\lambda^t f(i)). \tag{7.14}$$

From Theorem 1, we know that $(\overline{p}, \overline{\lambda}, \overline{\gamma})$ converge to an approximate NE.

Now suppose we have an 0-NE for the minimax game. This ensures that $\min\limits_{\gamma,\lambda} U(\overline{p},(\gamma,\lambda)) =$ OPT, which implies that both constraints are exactly satisfied.

7.2.4 Extensive-form games with incomplete information

We begin by introducing extensive-form games (EFGs) with incomplete information.

Figure 7.1 provides an example of an EFG. An EFG is represented as a graph, where blue nodes indicate that it is player 1's turn, while red nodes indicate that it is player 2's turn. In such games, players may have incomplete information, meaning they might not know their exact position in the game tree, even if they have perfect recall—that is, they remember the sequence of actions taken by all players. This uncertainty arises due to the presence of a fictitious nature player, which determines certain elements of the game. For instance, in Figure 7.1, nature decides whether the card is a King or an Ace. As a result, player 2 cannot distinguish between certain nodes when making their move. Consequently, they must play the same strategy at these nodes, which together form what is known as an *information set*. For a more detailed introduction to EFGs, we refer the reader to [1].

There are two common representations for strategies in extensive-form games.

The first representation is sequence-form strategies, which model each player's strategies as sequences walking down the tree. Using sequence-form strategies, the strategy space for both players becomes convex and compact, making game a bilinear problem. This representation shows that the minimax theorem holds in extensive-form games. A detailed introduction to sequence-form strategies can be found in [1].

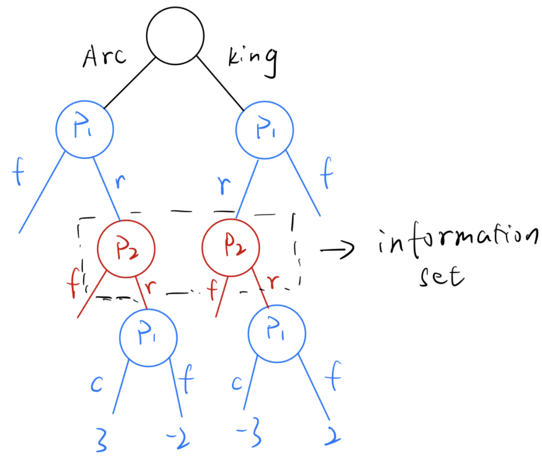

Figure 7.1: Example of EFG with imcomplete information

The second representation are behavior strategies, where a player's strategy is defined as a probability distribution over actions at each decision point. Under this representation, we are able to design no regret algorithms for EFG, such as Counterfactual Regret Minimization (CFR). CFR maintains an independent instance of a no-regret algorithm at each decision node, enabling efficient learning in extensive-form games. For a detailed discussion of CFR, we refer the reader to [2].

7.3 *

References

[1] 15-888 lecture 2: Representation of strategies in tree-form decision spaces, 2021.

[2] 15-888 lecture 5: Regret circuits and the counterfactual regret minimization (cfr) paradigm, 2021.