# Introduction to Online Learning (CSC 482A/581A) - Lecture 7

Nishant Mehta

## 1 Game theory basics

Thus far in the course, we have seen no-regret algorithms for many problem settings. We now shift our focus to game theory. Remarkably, we will see that no-regret algorithms can be used to find equilibria in games. Our focus here will be restricted to Nash equilibria for two-player zero-sum games, both of which we will introduce below. However, the ideas developed here, with some additional work, have a much further reach. Our coverage will be kept basic and will serve as a proof of concept of how no-regret learning algorithms can be used to approximate equilibria in games.

### 1.1 N-player normal form games

An $N$-player finite game expressed in normal form consists of $N$ players, where

- each player $k \in [N]$ has $n_k$ actions;
- if all the players play $(i_1, \dots, i_N) \in [n_1] \times \cdots \times [n_N]$, meaning that each player $j \in [N]$ plays action $i_j$, then the loss of player $k$ is $\ell^{(k)}(i_1, \dots, i_N) \in [0, 1]$.

We also allow players to randomize, by adopting a *mixed strategy*. If player $k$ plays mixed strategy $\pi^{(k)} \in \Delta_{n_k}$, then it randomizes by drawing its action according to distribution $\pi^{(k)}$. Let

$$\pi = \pi^{(1)} \times \cdots \times \pi^{(N)}$$

be a *mixed strategy profile*; it is the joint distribution over mixed strategies. Note that $\pi$ is a product distribution, so the draws of the players are independent. The expected loss of player $k$ under mixed strategy profile $\pi$ is

$$\mathsf{E}_{(I_1,\dots,I_N)\sim\pi}\left[\ell^{(k)}(I_1, \dots, I_N)\right] = \sum_{i_1=1}^{n_1} \cdots \sum_{i_N=1}^{n_N} \pi_{i_1}^{(1)} \cdots \pi_{i_N}^{(N)} \ell^{(k)}(i_1, i_2, \dots, i_N).$$

### 1.2 Nash equilibrium

A fundamental concept in game theory is that of a *Nash equilibrium*. Informally, we say $\pi$ is a Nash equilibrium if no player benefits from a unilateral deviation; that is, for each player, the best thing for her to do is stick with her mixed strategy from $\pi$ (assuming all other players stick with their respective mixed strategies from $\pi$). We now give a formal definition.

**Definition 1.** We say $\pi$ is a Nash equilibrium if, for all players $k \in [N]$, for all $\pi'^{(k)} \in \Delta_{n_k}$, for all strategy profiles $\pi'$ satisfying $\pi' = \pi^{(1)} \times \cdots \times \pi^{(k-1)} \times \pi'^{(k)} \times \pi^{(k+1)} \times \pi^{(N)}$, we have

$$\mathsf{E}_{(I_1,\dots,I_N)\sim\pi}[\ell(I_1, \dots, I_N)] \leq \mathsf{E}_{(I_1,\dots,I_N)\sim\pi'}[\ell(I_1, \dots, I_N)].$$

### 1.3 Two-player zero-sum games

We will restrict our study of $N$-player games to the case of two-player zero-sum games. In a *zero-sum*, 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. Let $m = n_1$ and $n = n_2$, so that Player 1 has $m$ actions and Player 2 has $n$ actions. We use $w$ (instead of $\pi^{(1)}$) for Player 1's mixed strategy and $x$ (instead of $\pi^{(2)}$) for Player 2's mixed strategy. Finally, since there are only two players, for any strategy profile $w \times x$, we can express the loss of each player via a game matrix $A \in [0,1]^{m \times n}$ for $A_{ij} = \ell(i,j)$. In a zero-sum game, Player 1's expected loss is $w^\top A x$ and Player 2's loss is $-w^\top A x$.[1] Therefore, Player 1 can be viewed as the "min" player (in terms of its loss) and Player 2 can be viewed as the max player (in terms of Player 1's loss, which is the negation of Player 2's loss).

Now, suppose that $(\bar{w}, \bar{x})$ is a Nash equilibrium. Then we have both:

$$\bar{w}^\top A \bar{x} \leq w^\top A \bar{x} \quad \text{for all } w \in \Delta_m \qquad\qquad \bar{w}^\top A \bar{x} \geq \bar{w}^\top A x \quad \text{for all } x \in \Delta_n.$$

and hence both

$$\bar{w}^\top A \bar{x} = \min_{w \in \Delta_m} w^\top A \bar{x} \qquad\qquad \bar{w}^\top A \bar{x} = \max_{x \in \Delta_n} \bar{w}^\top A x.$$

Consequently, we have the sequence of inequalities

$$\min_{w \in \Delta_m} \max_{x \in \Delta_n} w^\top A x \leq \max_{x \in \Delta_n} \bar{w}^\top A x = \bar{w}^\top A \bar{x} = \min_{w \in \Delta_m} w^\top A \bar{x} \leq \max_{x \in \Delta_n} \min_{w \in \Delta_m} w^\top A x.$$

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

$$\max_{x \in \Delta_n} \min_{w \in \Delta_m} w^\top A x \leq \min_{w \in \Delta_m} \max_{x \in \Delta_n} w^\top A x.$$

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

$$\min_{w \in \Delta_m} \max_{x \in \Delta_n} w^\top A x = \max_{x \in \Delta_n} \min_{w \in \Delta_m} w^\top A x. \tag{1}$$

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. Note that thus far, we have not at all showed that (1) holds for two-player zero-sum games. We have only showed it 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*

$$\min_{w \in \Delta_m} \max_{x \in \Delta_n} w^\top A x = \max_{x \in \Delta_n} \min_{w \in \Delta_m} w^\top A x.$$

Various proofs exist that make use of fixed-point theorems. Our proof of this fact 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 (approximate in a sense that we define later).

[1] The careful reader may note that Player 2's loss is therefore in the range $[-1, 0]$ rather than $[0, 1]$. This should not be cause for any real concern, and in the end, the loss range is still 1, and that is all that will really matter for our purposes.

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

## 2 From no-regret learning to approximate Nash equilibria

We begin this section by proving Theorem 1. After that, we introduce the notion of approximate Nash equilibria and show how the proof algorithmically gives us an approximate Nash equilibrium.

### 2.1 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 online linear optimization, and each player will run a no-regret algorithm. Letting $w_1, \dots, w_T$ be the sequence of plays of Player 1 and $x_1, \dots, x_T$ be the sequence of players of Player 2, the loss of Player 1 in round $t$ will be $\ell_t^{(\mathrm{I})} := Ax_t$ while the loss of Player 2 in round $t$ will be $\ell_t^{(\mathrm{II})} := -A^\top w_t$.

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

$$\frac{1}{T}\sum_{t=1}^{T} w_t^\top A x_t \le \min_{w \in \Delta_m} \frac{1}{T}\sum_{t=1}^{T} w A x_t + \varepsilon_{T,1}. \tag{2}$$

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

$$\max_{x \in \Delta_n} \frac{1}{T}\sum_{t=1}^{T} w A x_t \le \frac{1}{T}\sum_{t=1}^{T} w_t^\top A x_t + \varepsilon_{T,2}. \tag{3}$$

Letting $\bar{w} = \frac{1}{T}\sum_{t=1}^{T} w_t$ and $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$ be the simple averages of Player 1's and Player 2's plays respectively, we have

$$\begin{aligned}
\min_{w \in \Delta_m} \max_{x \in \Delta_n} w^\top A x &\le \max_{x \in \Delta_n} \bar{w}^\top A x \\
&= \max_{x \in \Delta_n} \sum_{t=1}^{T} w_t^\top A x \\
&\le \sum_{t=1}^{T} w_t^\top A x_t + \varepsilon_{T,2} \qquad (4) \\
&\le \min_{w \in \Delta_m} \sum_{t=1}^{T} w^\top A x_t + \varepsilon_{T,1} + \varepsilon_{T,2} \\
&= \min_{w \in \Delta_m} w^\top A \bar{x} + \varepsilon_{T,1} + \varepsilon_{T,2} \qquad \le \max_{x \in \Delta_n} \min_{w \in \Delta_m} w^\top A x + \varepsilon_{T,1} + \varepsilon_{T,2},
\end{aligned}$$

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

Taking $T \to \infty$ implies that $\min_w \max_x w^\top A x \le \max_x \min_w w^\top A x$, as desired. □

### 2.2 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 approximate notion of a Nash equilibrium (and, taking $T \to \infty$, it recovers an actual Nash equilibrium). To understand this, we first need to introduce this approximate notion.

**Definition 2** (Approximate Nash equilibrium)**.** We say a pair of mixed strategies $(\bar{w}, \bar{x})$ is an $\varepsilon$-approximate Nash equilibrium if both of the following are true:

$$\bar{w}^\top A\bar{x} \leq w^\top A\bar{x} + \varepsilon \quad \text{for all } w \in \Delta_m$$
$$\bar{w}^\top A\bar{x} \geq \bar{w}^\top Ax - \varepsilon \quad \text{for all } x \in \Delta_n.$$

The above two conditions can be equivalently rewritten as:

$$\bar{w}^\top A\bar{x} \leq \min_{w \in \Delta_m} w^\top A\bar{x} + \varepsilon$$
$$\bar{w}^\top A\bar{x} \geq \max_{x \in \Delta_n} \bar{w}^\top Ax - \varepsilon.$$

In words, if $(\bar{w}, \bar{x})$ is an $\varepsilon$-approximate Nash equilibrium, then each player can benefit by at most $\varepsilon$ from a unilateral deviation from the strategy profile $\bar{w} \times \bar{x}$.

Note that a bound of the form

$$\max_{x \in \Delta_n} \bar{w}^\top Ax \leq \min_{w \in \Delta_m} w^\top A\bar{x} + \varepsilon \tag{5}$$

implies that $(\bar{w}, \bar{x})$ is an $\varepsilon$-approximate Nash equilibrium. Indeed, for the first inequality,

$$\bar{w}^\top A\bar{x} \leq \max_{x \in \Delta_n} \bar{w}^\top Ax \leq \min_{w \in \Delta_m} w^\top A\bar{x} + \varepsilon,$$

while for the second inequality,

$$\bar{w}^\top A\bar{x} \geq \min_{w \in \Delta_m} w^\top A\bar{x} \geq \max_{x \in \Delta_n} \bar{w}^\top Ax - \varepsilon.$$

Consequently, the sequence of inequalities (4) implies that $(\bar{w}, \bar{x})$ is an $(\varepsilon_{T,1} + \varepsilon_{T,2})$-approximate Nash equilibrium, since we have

$$\max_{x \in \Delta_n} \bar{w}^\top Ax \leq \min_{w \in \Delta_m} w^\top A\bar{x} + \varepsilon_{T,1} + \varepsilon_{T,2}.$$

Moreover, taking $T \to \infty$, the pair $(\bar{w}, \bar{x})$ is a Nash equilibrium.

### 2.3 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 $(\varepsilon_{T,1})_{T \geq 1}$ and $(\varepsilon_{T,2})_{T \geq 1}$. Let us now make things concrete. For Player 1, recall from the proof of Theorem 1 that the loss vector in round $t$ is $\ell_t^{(\mathrm{I})} = Ax_t$ with $A \in [0,1]^{m \times n}$ and $x_t \in \Delta_n$. Therefore, we have $\ell_t^{(\mathrm{I})} \in [0,1]^m$, and we may use Decreasing Hedge which, from Section 5 of Lecture 5, enjoys an average regret bound of $\varepsilon_{T,1} = \sqrt{\frac{2 \log m}{T}}$ since the number of actions is now $m$. Similarly, for Player 2, we have $\ell_t^{(\mathrm{II})} = -A^T w_t$ with $w_t \in \Delta_m$, so we have $\ell_t^{(\mathrm{II})} \in [-1,0]^n$. As only the infinity norm of loss vector matters for the regret bound, and since the number of actions is now $n$, using Decreasing Hedge once again gives an average regret bound of $\varepsilon_{T,2} = \sqrt{\frac{2 \log n}{T}}$.

Therefore, "running the proof" for $T$ rounds, we have that the resulting pair $(\bar{w}, \bar{x})$ is an $\frac{\sqrt{2 \log m} + \sqrt{2 \log n}}{T}$-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 $A$ 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 $O\left(\frac{\log m+\log n}{T}\right)$. To get this faster rate, we now introduce Optimistic Follow-the-Regularized-Leader (Optimistic FTRL).

# 3 Optimistic FTRL

The idea of Optimistic FTRL is to run FTRL with one extra loss, namely, a prediction of the loss function for the current round. For simplicity, we will restrict attention to the case of linear losses, so that the gradient of the loss is simply a loss vector $\ell_t \in \mathbb{R}^d$ and the predicted loss (equivalently, predicted gradient) in round $t$ is a vector $M_t \in \mathbb{R}^d$. Also for simplicity, we will consider the case of a fixed (so, not time-varying) regularizer $\frac{1}{\eta}\psi$ for some fixed, positive learning rate $\eta$. In this setup, in round $t$, Optimistic FTRL plays

$$w_t = \operatorname*{arg\,min}_{w \in V} \left\{ \left\langle w, \sum_{s=1}^{t-1} \ell_s + M_t \right\rangle + \frac{1}{\eta}\psi(w) \right\}. \tag{6}$$

Just like in our analysis of FTRL, our analysis of Optimistic FTRL centers around Be-the-Regularized-Leader (BTRL). In round $t$, BTRL uses not the predicted loss but rather the actual loss $\ell_t$; hence, it plays

$$w_t^* = \operatorname*{arg\,min}_{w \in V} \left\{ \left\langle w, \sum_{s=1}^{t} \ell_s \right\rangle + \frac{1}{\eta}\psi(w) \right\}. \tag{7}$$

## 3.1 Regret bound for Optimistic FTRL

Optimistic FTRL admits the following regret bound.

**Theorem 2.** *Let $V \subseteq \mathbb{R}^d$ be a closed convex set. Let $\ell_1, \dots, \ell_T \in \mathbb{R}^d$ be a sequence of loss vectors and let regularizer $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$ be $\lambda$-strongly convex with respect to a norm $\|\cdot\|$. Without loss of generality, assume (by shifting, if necessary) that $\inf_{u \in V} \psi(u) = 0$. For any sequence of loss predictions $(M_t)_{t \in [T]}$ and any comparator $u \in V$, Optimistic FTRL has regret*

$$\sum_{t=1}^{T} \langle \ell_t, w_t - u \rangle \le \frac{\psi(u)}{\eta} + \sum_{t=1}^{T} \|\ell_t - M_t\|_* \|w_t - w_t^*\| - \frac{\lambda}{2\eta} \sum_{t=1}^{T} \left( \|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right).$$

*Proof.* We begin with a decomposition of the instantaneous regret:

$$\langle \ell_t, w_t - u \rangle = \langle \ell_t - M_t, w_t - w_t^* \rangle + \langle M_t, w_t - w_t^* \rangle + \langle \ell_t, w_t^* - u \rangle$$

We control the first term via the following inequality (which holds by the definition of the dual norm):

$$\langle \ell_t - M_t, w_t - w_t^* \rangle \le \|\ell_t - M_t\|_* \|w_t - w_t^*\|;$$

note that the sum of the RHS already appears in the regret bound.

The main work is controlling the second two terms in the regret decomposition. For arbitrary $u \in V$, we will establish the inequality

$$\sum_{t=1}^{T} \langle M_t, w_t - w_t^* \rangle + \sum_{t=1}^{T} \langle \ell_t, w_t^* - u \rangle \le \frac{\psi(u)}{\eta} - \frac{\lambda}{2\eta} \sum_{t=1}^{T} \left( \|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right) \tag{8}$$

by induction on $T$. The base case of $T = 0$ is trivial from the nonnegativity of the regularizer. Let $Q_t = \frac{\lambda}{2\eta} \sum_{s=1}^{t} \left( \|w_s - w_s^*\|^2 + \|w_s - w_{s-1}^*\|^2 \right)$ denote the quadratic terms up till round $t$. Assume the inequality holds up to round $t-1$, so that

$$\sum_{s=1}^{t-1} \langle M_s, w_s - w_s^* \rangle + \sum_{s=1}^{t-1} \langle \ell_s, w_s^* - u \rangle \le \frac{\psi(u)}{\eta} - Q_{t-1}.$$

We begin with the rewrite

$$\sum_{s=1}^{t-1} \langle M_s, w_s - w_s^* \rangle + \sum_{s=1}^{t-1} \langle \ell_s, w_s^* \rangle \le \left\langle \sum_{s=1}^{t-1} \ell_s, u \right\rangle + \frac{\psi(u)}{\eta} - Q_{t-1}.$$

Next, take $u = w_{t-1}^*$ and add $\langle M_t, w_t - w_t^* \rangle + \langle \ell_t, w_t^* \rangle$ to both sides, giving

$$\begin{aligned} &\sum_{s=1}^{t} \langle M_s, w_s - w_s^* \rangle + \sum_{s=1}^{t} \langle \ell_s, w_s^* \rangle \\ &\le \left\langle \sum_{s=1}^{t-1} \ell_s, w_{t-1}^* \right\rangle + \frac{\psi(w_{t-1}^*)}{\eta} - Q_{t-1} + \langle M_t, w_t - w_t^* \rangle + \langle \ell_t, w_t^* \rangle. \end{aligned}$$

From the optimality of BTRL for the objective of (7) and the $\lambda$-strong convexity of $\psi$, the above is at most

$$\begin{aligned} &\left\langle \sum_{s=1}^{t-1} \ell_s, w_t \right\rangle + \frac{\psi(w_t)}{\eta} - Q_{t-1} + \langle M_t, w_t - w_t^* \rangle + \langle \ell_t, w_t^* \rangle - \frac{\lambda}{2\eta} \|w_t - w_{t-1}^*\|^2 \\ &= \left\langle \sum_{s=1}^{t-1} \ell_s + M_t, w_t \right\rangle + \frac{\psi(w_t)}{\eta} - Q_{t-1} - \langle M_t, w_t^* \rangle + \langle \ell_t, w_t^* \rangle - \frac{\lambda}{2\eta} \|w_t - w_{t-1}^*\|^2. \end{aligned}$$

Similarly, from the optimality of Optimistic FTRL for the objective of (6) and the $\lambda$-strong convexity of $\psi$, the above is at most

$$\begin{aligned} &\left\langle \sum_{s=1}^{t-1} \ell_s + M_t, w_t^* \right\rangle + \frac{\psi(w_t^*)}{\eta} - Q_{t-1} - \langle M_t, w_t^* \rangle + \langle \ell_t, w_t^* \rangle - \frac{\lambda}{2\eta} \|w_t - w_{t-1}^*\|^2 - \frac{\lambda}{2\eta} \|w_t - w_t^*\|^2 \\ &= \left\langle \sum_{s=1}^{t} \ell_s, w_t^* \right\rangle + \frac{\psi(w_t^*)}{\eta} - Q_t \\ &= \left\langle \sum_{s=1}^{t} \ell_s, u \right\rangle + \frac{\psi(u)}{\eta} - Q_t, \end{aligned}$$

where $u \in V$ is arbitrary. □

### 3.2 Faster convergence to Nash equilibria using Optimistic FTRL

Now, let us see how the theorem implies fast convergence to an approximate Nash equilibrium. We take the setup from Section 2.3. In addition, let $M_t^{(\mathrm{I})} = \ell_{t-1}^{(\mathrm{I})} = Ax_{t-1}$ be Player I's predicted loss vector in round $t$; similarly, let $M_t^{(\mathrm{II})} = \ell_{t-1}^{(\mathrm{II})} = -A^T w_{t-1}$ be Player II's predicted loss vector in round $t$. Finally, take the regularizer $\psi$ to be the negative Shannon entropy, so that:

- for Player I, $\psi(w) = \sum_{j=1}^m w_j \log w_j$ for $w \in \Delta_m$;
- for Player II, $\psi(x) = \sum_{j=1}^n x_j \log x_j$ for $x \in \Delta_n$.

Recall that $\psi$ is 1-strongly convex with respect to $\|\cdot\|_1$, so that we may take $\lambda = 1$ in Theorem 2.

With these settings, we have the following corollary to Theorem 2

**Corollary 1.** *With the settings above, taking $\eta = \sqrt{1/8}$, the sum of the regrets of Players I and II is at most*

$$\sqrt{8}(\log m + \log n).$$

*Proof.* We will first show that Player I's regret is bounded as

$$\frac{\log m}{\eta} + \eta \sum_{t=1}^T \|x_t - x_{t-1}\|_*^2 - \frac{1}{8\eta} \sum_{t=1}^T \|w_{t-1} - w_t\|^2 \tag{9}$$

and, similarly and symmetrically, Player II's regret is bounded as

$$\frac{\log n}{\eta} + \eta \sum_{t=1}^T \|w_t - w_{t-1}\|_*^2 - \frac{1}{8\eta} \sum_{t=1}^T \|x_{t-1} - x_t\|^2.$$

By selecting $\eta = \frac{1}{8\eta}$, there is a nice symmetry between the two regret bounds, and their sum gives the result.

It suffices to establish (9) as Player II's regret bound can be bounded in a completely analogous way. From Theorem 2, Player I's regret is bounded by

$$\frac{\log m}{\eta} + \sum_{t=1}^T \|\ell_t^{(\mathrm{I})} - \ell_{t-1}^{(\mathrm{I})}\|_* \|w_t - w_t^*\| - \frac{1}{2\eta} \sum_{t=1}^T \left( \|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right)$$

From the Fenchel-Young inequality (Lemma 2 of Lecture 6), the above is at most

$$\begin{aligned}
&\frac{\log m}{\eta} + \sum_{t=1}^T \left( \inf_{\rho > 0} \left\{ \frac{\rho}{2} \|\ell_t^{(\mathrm{I})} - \ell_{t-1}^{(\mathrm{I})}\|_*^2 + \frac{1}{2\rho} \|w_t - w_t^*\|^2 \right\} \right) - \frac{1}{2\eta} \sum_{t=1}^T \left( \|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right) \\
&\overset{(\rho = 2\eta)}{\leq} \frac{\log m}{\eta} + \sum_{t=1}^T \left( \eta \|\ell_t^{(\mathrm{I})} - \ell_{t-1}^{(\mathrm{I})}\|_*^2 + \frac{1}{4\eta} \|w_t - w_t^*\|^2 \right) - \frac{1}{2\eta} \sum_{t=1}^T \left( \|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right) \\
&= \frac{\log m}{\eta} + \sum_{t=1}^T \left( \eta \|\ell_t^{(\mathrm{I})} - \ell_{t-1}^{(\mathrm{I})}\|_*^2 - \frac{1}{4\eta} \|w_t - w_t^*\|^2 \right) - \frac{1}{2\eta} \sum_{t=1}^T \|w_t - w_{t-1}^*\|^2 \\
&\leq \frac{\log m}{\eta} + \eta \sum_{t=1}^T \|\ell_t^{(\mathrm{I})} - \ell_{t-1}^{(\mathrm{I})}\|_*^2 - \frac{1}{4\eta} \sum_{t=1}^T \left( \|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right) \\
&\leq \frac{\log m}{\eta} + \eta \sum_{t=1}^T \|\ell_t^{(\mathrm{I})} - \ell_{t-1}^{(\mathrm{I})}\|_*^2 - \frac{1}{4\eta} \sum_{t=1}^T \left( \|w_{t-1} - w_{t-1}^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right) \\
&\leq \frac{\log m}{\eta} + \eta \sum_{t=1}^T \|\ell_t^{(\mathrm{I})} - \ell_{t-1}^{(\mathrm{I})}\|_*^2 - \frac{1}{8\eta} \sum_{t=1}^T \|w_{t-1} - w_t\|^2,
\end{aligned} \tag{10}$$

where the last inequality uses the fact for any vectors $a, b, c \in \mathbb{R}^d$ and any norm $\|\cdot\|$,

$$\|a - c\|^2 \le 2\|a - b\|^2 + 2\|b - c\|^2.$$

Now, noting that

$$\|\ell_t^{(\mathrm{I})} - \ell_{t-1}^{(\mathrm{I})}\|_* = \|A(x_t - x_{t-1})\|_* \le \|x_t - x_{t-1}\|_1, \tag{11}$$

try to show this as an exercise, but this footnote[3] contains a proof. The regret bound (9) follows by combining (11) with (10). □

# 4 Improved regret under gradual variations

The idea of Optimistic FTRL (actually, they introduced Optimistic Mirror Descent) is due to Chiang et al. (2012). Their motivation was not at all related to faster convergence to Nash equilibria but instead to obtain lower regret in situations where the loss functions change slowly. Since we have restricted attention to online linear optimization (OLO), we will explore such a result for OLO. In order to have an adaptive algorithm (adaptive to the unknown amount of variation in the sequence of loss vectors, we first present an extension of Theorem 2 for time-varying (but non-decreasing in the sense of a non-increasing learning rate) regularizers.

**Theorem 3.** *Let $V \subseteq \mathbb{R}^d$ be a closed convex set. Let $\ell_1, \dots, \ell_T \in \mathbb{R}^d$ be a sequence of loss vectors and let regularizer $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$ be $\lambda$-strongly convex with respect to a norm $\|\cdot\|$. Without loss of generality, assume (by shifting, if necessary) that $\inf_{u \in V} \psi(u) = 0$. Let the sequence of regularizers be $\psi_t = \frac{1}{\eta_t}\psi$. For any sequence of loss predictions $(M_t)_{t \in [T]}$ and any comparator $u \in V$, Optimistic FTRL has regret*

$$\sum_{t=1}^T \langle \ell_t, w_t - u\rangle \le \frac{\psi(u)}{\eta_T} + \sum_{t=1}^T \|\ell_t - M_t\|_* \|w_t - w_t^*\| - \frac{\lambda}{2\eta_t} \sum_{t=1}^T \left(\|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2\right).$$

We leave the proof of this result as an exercise. For some guidance on how to prove this result, start with the proof of Theorem 2, but rather than proving (8) by induction, instead prove

$$\sum_{t=1}^T \langle M_t, w_t - w_t^*\rangle + \sum_{t=1}^T \langle \ell_t, w_t^* - u\rangle \le \frac{\psi(u)}{\eta_T} - \sum_{t=1}^T \frac{\lambda}{2\eta_t} \left(\|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2\right)$$

by induction. Hence, use the modified cumulative quadratic term

$$Q_t = \sum_{s=1}^t \frac{\lambda}{2\eta_t} \left(\|w_s - w_s^*\|^2 + \|w_s - w_{s-1}^*\|^2\right).$$

The proof should still go through since the regularizer is nonnegative and the learning rate is non-increasing (which allows to change $\frac{\psi(u)}{\eta_{t-1}}$ to $\frac{\psi(u)}{\eta_t}$ in the induction step).

Let us see how to use Theorem 3 to get an improved regret bound under gradual variations. We begin by taking the choice $M_t = \ell_{t-1}$ in Theorem 3 and then upper bounding the resulting regret

[3] Note that $\max_{j \in [n]} \left[\sum_{i=1}^m A_i (x_{t,i} - x_{t-1,i})\right]_j \le \max_{j \in [n]} \left[\sum_{i=1}^m \mathbf{1} |x_{t,i} - x_{t-1,i}|\right]_j = \|x_t - x_{t,i}\|_1$.

bound of the theorem as

$$\begin{aligned}
&\frac{\psi(u)}{\eta_T} + \sum_{t=1}^{T} \|\ell_t - \ell_{t-1}\|_* \|w_t - w_t^*\| - \frac{\lambda}{2\eta_t} \sum_{t=1}^{T} \left( \|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right) \\
&\leq \frac{\psi(u)}{\eta_T} + \sum_{t=1}^{T} \left( \frac{\eta_t}{2\lambda} \|\ell_t - \ell_{t-1}\|_*^2 + \frac{\lambda}{2\eta_t} \|w_t - w_t^*\|^2 \right) - \frac{\lambda}{2\eta_t} \sum_{t=1}^{T} \left( \|w_t - w_t^*\|^2 + \|w_t - w_{t-1}^*\|^2 \right) \\
&= \frac{\psi(u)}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2\lambda} \|\ell_t - \ell_{t-1}\|_*^2 - \frac{\lambda}{2\eta_t} \sum_{t=1}^{T} \|w_t - w_{t-1}^*\|^2 \\
&\leq \frac{\psi(u)}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2\lambda} \|\ell_t - \ell_{t-1}\|_*^2.
\end{aligned}$$

Let us denote the squared loss difference terms as $v_t = \|\ell_t - \ell_{t-1}\|_*^2$, so that the above may be rewritten as

$$\frac{\psi(u)}{\eta_T} + \frac{1}{2\lambda} \sum_{t=1}^{T} \eta_t v_t.$$

It will be convenient to also define the cumulative term $V_t = \sum_{s=1}^{t} v_s$. Now, it would be a nice world if we could select the learning rate $\eta_t$ with knowledge of $v_1, v_2, \ldots, v_t$. If we could do this, then setting $\eta_t = \sqrt{\frac{\lambda}{V_t}}$ would yield the regret bound

$$\psi(u)\sqrt{\frac{V_T}{\lambda}} + \sqrt{\frac{V_T}{\lambda}} = (\psi(u) + 1)\sqrt{\frac{V_T}{\lambda}}.$$

However, unfortunately, we use learning rate $\eta_t$ at the *start* of round $t$, at which point we only have knowledge of $v_1, v_2, \ldots, v_{t-1}$. One choice one might consider is $\eta_t = \sqrt{\frac{\lambda}{V_{t-1}}}$, but this blows up the regret bound since $V_0 = 0$. To fix this issue, suppose we know an upper bound on dual norms of the losses of the form $\max_{t \in [T]} \|\ell_t\|_* \leq G$ for a positive constant $G$. Then, from the triangle inequality, we have can bound the loss variation between any two rounds as $\max_{t \in [T]} v_t \leq 4G^2$. If we set the learning rate as $\eta_t = \sqrt{\frac{\lambda}{V_{t-1} + 4G^2}}$, from $V_t \leq V_{t-1} + 4G^2$ it follows that

$$\eta_t v_t \leq \sqrt{\frac{\lambda}{V_t}},$$

and hence we get the regret bound

$$\psi(u)\sqrt{\frac{V_T + 4G^2}{\lambda}} + \sqrt{\frac{V_T}{\lambda}}.$$

If one has available a known upper bound $R$ on $\psi(u)$ — for instance, the negative Shannon entropy (shifted to be nonnegative) has $\psi(u) \leq \log d$ for all $u \in V = \Delta_d$ — then a better choice of learning rate is $\eta_t = \sqrt{\frac{R\lambda}{V_{t-1}}}$, which gives the regret bound

$$\sqrt{\frac{R(V_T + 4G^2)}{\lambda}} + \sqrt{\frac{RV_T}{\lambda}}.$$

In particular, for using "Optimistic Decreasing Hedge" (Optimistic FTRL with the negative Shannon entropy regularizer) in Decision-Theoretic Online Learning gives a regret at most

$$\sqrt{V_T + 4}\left(\sqrt{\log d} + 1\right).$$

## Bibliographical notes

1. The first few sections heavily draw from Section 7.1 of Cesa-Bianchi and Lugosi (2006).

2. The notion of an $\varepsilon$-approximate Nash equilibrium is standard. See Daskalakis et al. (2006), for example.

3. As far as I am aware, Freund and Schapire (1999) were the first to show convergence to approximate Nash equilibria using no-regret learning algorithms. The ideas of this paper are similar to those in Freund and Schapire (1997), which introduced AdaBoost and the setting of Decision-Theoretic Online Learning.

4. Until the breakthrough work of Daskalakis et al. (2011), the best known convergence rate for Nash equilibria of two-player zero-sum games using no-regret learning algorithms was $O(1/\sqrt{T})$. The same work well-describes the notion of strongly uncoupled distributed protocols. The no-regret algorithm introduced by Daskalakis et al. (2011), however, is not that simple. Rakhlin and Sridharan (2013) were the first to give a simple no-regret algorithm achieving a $O(\frac{\log T}{T})$ convergence rate. Both the aforementioned works developed no-regret algorithms that in the nice case (when both players use the prescribed algorithm) achieve the fast convergence rate of $O(\frac{\log T}{T})$, while still ensuring $O(1/\sqrt{T})$ regret if the other player deviates by using a different algorithm. The latter fallback guarantee is a form of robustness and requires some light additional complexity to the algorithm of Rakhlin and Sridharan (2013); stripping away this complexity allows their algorithm to achieve a convergence rate of $O(1/T)$.

5. The predecessor of Optimistic FTRL is Optimistic Miror Descent, introduced by Chiang et al. (2012) for the special case of $M_t = \nabla \ell_t$ and generalized (and named Optimistic Mirror Descent) by Rakhlin and Sridharan (2013). Optimistic FTRL was introduced by Syrgkanis et al. (2015), and its presentation and the analysis given here is essentially from the same paper (which also provides a clean presentation and analysis of Optimistic Mirror Descent). The main difference is that we adapted the presentation to the case of two-player zero sum games and used losses instead of utilities.

6. The first bound for gradual variations is due to Chiang et al. (2012), which is a journal paper that combines two works, each from a different set of authors. I will complete this point soon. There is quite a bit of history here.

## References


Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Conference on Learning Theory*, pages 6–1. JMLR Workshop and Conference Proceedings, 2012.

Constantinos Daskalakis, Aranyak Mehta, and Christos Papadimitriou. A note on approximate nash equilibria. In *Internet and Network Economics: Second International Workshop, WINE 2006, Patras, Greece, December 15-17, 2006. Proceedings 2*, pages 297–306. Springer, 2006.

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 235–254. SIAM, 2011.

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

Yoav Freund and Robert E Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26, 2013.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. *Advances in Neural Information Processing Systems*, 28, 2015.