Algorithmic Game Theory, Complexity and Learning July 20, 2017

Lecture 1.4: Online Convex Optimization

Lecturer: Constantinos Daskalakis *Scribe: Constantinos Daskalakis*

In the previous lecture we studied a simple online learning setting where a learner chooses between two actions on each day, aiming to achieve an expected loss over a sequence of days that is at least as good as the loss of the best fixed action in hindsight. We saw that a simple tweak to the very simple Follow-the-Leader algorithm, namely the Follow-the-Regularized-Leader algorithm, achieved the latter property, which we called *no-regret.*

In this lecture we will see a vast generalization of the above setting, which is known as *online convex optimization* and we will also see how the Follow-the-Regularized-Leader algorithm and its analysis directly generalizes to this setting.

We will conclude by showing that the existence of no-regret algorithms for online convex optimization imply von-Neumann's minimax theorem for zero-sum games. Since the minimax theorem for zero-sum games is equivalent to strong LP duality, we will thus see how the existence of no-regret algorithms implies strong LP duality.

1 Online Convex Optimization: A General Framework

In the previous lecture we considered a learner who picks an action from a set $\{H, L\}$ by sampling a probability distribution $p_t \in [0,1]$. His expected loss was thus $f(p_t; \ell_t) = p_t \cdot \ell_t^H + (1-p_t) \cdot \ell_t^L$, where $\ell_t^H, \ell_t^L \in [0,1]$ were respectively the losses incurred by H and L as chosen by an adversary. *Online convex optimization* is the following generalization of this setting. On each day t:

1. the learner picks a vector p_t from a convex set S in $\mathbb{R}^d$
2. an adversary picks a convex function $f_t : S \to \mathbb{R}$, which is L-Lipscitz with respect to some norm $\|\cdot\|$ (i.e. $\|f_t(p) - f_t(p')\| \le L \cdot \|p - p'\|$) and differentiable[1]
3. the learner incurs a loss $f_t(p_t)$ and observes the function f_t (i.e. he can evaluate the function at any point after time t)[2]

Regret. The regret of an online convex optimization algorithm is the difference between the cumulative loss of the algorithm and the loss of the best fixed $p \in S$ in hindsight, in the worst-case over loss sequences:

$$\textsc{Regret}(T) = \sup_{\ell_1,\ldots,\ell_T} \left(\sum_{t=1}^{T} f_t(p_t) - \inf_{p \in S} \sum_{t=1}^{T} f_t(p) \right) \tag{1}$$

An algorithm is no-regret, if its regret grows sub-linearly with time, i.e., $\textsc{Regret}(T) = o(T)$. Equivalently, if its average regret $\textsc{Regret}(T)/T$ goes to zero as T goes to infinity.

Online convex optimization generalizes many problems that have been studied in online learning theory. We present here a few:

Example 1 (Expert Setting)**.** *In the expert setting, the learner wants to decide among K actions, often referred to as "experts" in the literature. Every day the algorithm picks a distribution $p_t \in \Delta_K$ over this K actions and draws an action i_t from this distribution. An adversary picks a loss vector $\ell_t = (\ell_t^1, \ldots, \ell_t^K)$, where ℓ_t^i is the loss of action i on day t. The algorithm receives an expected loss of*

[1] In fact we do not need to assume that the function is differentiable and everything we do in this note continues to hold if we replace gradients with sub-gradients.

[2] As you might uncover in the assignment, you only need to observe the gradient of the function f_t after day t, rather than being able to evaluate the function at every point.

$f_t(p_t) = \langle p_t, \ell_t \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product between two vectors. The goal of the algorithm is to compete with the loss of the best fixed action in hindsight, which equivalently can be written as $\sup_{p\in\Delta_K} \sum_{t=1}^{T} f_t(p)$.

It is easy to see that the expert setting is a special case of online convex optimization. The convex set $S \subseteq \mathbb{R}^K$ from which the learner is picking is the simplex and the loss function that the adversary picks every day is not an arbitrary convex function, but a linear function associated with a vector $\ell_t \in \mathbb{R}^K$. Without the simplex constraint, but for an arbitrary set S, the latter is known in the literature as online linear optimization. *As you might uncover in the assignment, online convex optimization can always be reduced to online linear optimization.*

Example 2 (Online Quadratic Optimization). *Every day the learner wants to pick a point $p_t \in S \subseteq \mathbb{R}^d$ and the adversary picks a point $z_t \in S$. The goal of the learner is to pick a point p_t such that it is close to z_t. More concretely, his loss is the squared euclidean norm, between p_t and z_t: $f_t(p_t) = \|p_t - z_t\|^2$. The goal of the learner is to pick a sequence of points such that they incur a loss which is comparable to the best fixed point in hindsight. This example is a special case of online convex optimization, with the extra property that the loss functions that the adverary picks are not just convex, but strongly convex. The latter can be leveraged to get regret bounds that are much better than what we will see in this lecture, i.e. regret that increases as $O(\log(T))$.*

Example 3 (Online Least Squares Linear Regression). *Every day the adversary picks a pair (x_t, y_t) of a point $x_t \in \mathbb{R}^d$ and an outcome $y_t \in R$. The learner wants to find a good linear function that approximates the relation between x_t and y_t in hindsight. Specifically, he wants to pick a point $p_t \in \mathbb{R}^d$, with $\|p_t\| \leq H$, and his loss at time t will be: $f_t(p_t) = (\langle p_t, x_t \rangle - y_t)^2$.*

The Follow-the-Regularized-Leader algorithm can be easily generalized to the online convex optimization setting as follows:

$$\text{Follow-the-Regularized-Leader (FTRL): } \tilde{p}_t = \underset{p\in S}{\operatorname{argmin}} \underbrace{\sum_{\tau=1}^{t-1} f_\tau(p)}_{F_{t-1}(p)} + \frac{1}{\eta} R(p) = \underset{p\in S}{\operatorname{argmin}} \tilde{F}_{t-1}(p) \tag{2}$$

In the above $R(p)$ is the regularizer function which we assume to be strongly convex. In the case where $S = \mathbb{R}$, in the last lecture, strong convexity simply meant that $R''(p) \geq 1$. In a multi-dimensional space, strong convexity is the generalization of that condition and is defined as follows:

Definition 1 (Strongly Convex Function). *A function $R : \mathbb{R}^d \to \mathbb{R}$ is $\frac{1}{\eta}$-strongly convex with respect to a norm $\|\cdot\|$, if for any p and p_0 in S:*

$$R(p) \geq R(p_0) + \langle \nabla R(p_0), p - p_0 \rangle + \frac{1}{2\eta}\|p - p_0\|^2 \tag{3}$$

where $\nabla R(\cdot)$ is the gradient of the function.

A pictorial way of viewing a strongly convex function is as follows: take the linearization of the function $R(p)$ around a point p_0 (i.e. consider the linear function $h(p) = R(p) + \langle \nabla R(p_0), p - p_0 \rangle$, depicted with a dashed blue in Figure 1). Then the value of the function at any other point $p \neq p_0$, should not only be above $h(p)$, but it should be above $h(p)$ by a large margin that grows with the squared distance of p from p_0, i.e. $\frac{1}{2\eta}\|p - p_0\|^2$.

Similar to what we did in the single-dimension, we can show more generally, that strong convexity leads to stability, i.e. that if two functions are strongly convex and their difference is a Lipschitz function, then their minima have to be close. We remind the reader the pictorial proof of this fact from the previous lecture in Figure 2.

Lemma 1 (Closeness of minima of strongly convex functions). *Consider two functions $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$, that are $\frac{1}{\eta}$-strongly convex with respect to some norm $\|\cdot\|$ and such that their difference $h(p) = g(p) - f(p)$ is an L-Lipschitz function with respect to the same norm. Then if $p_f = \operatorname{argmin}_{p\in S} f(p)$ and $p_g = \operatorname{argmin}_{p\in S} g(p)$, it must hold that: $\|p_f - p_g\| \leq \eta \cdot L$.*

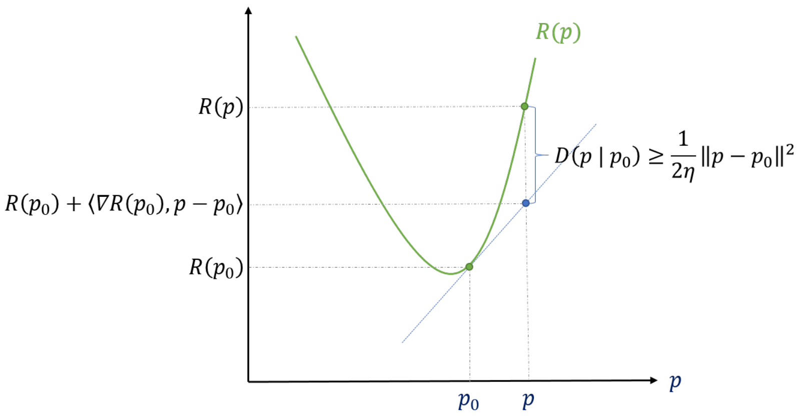
$R(p)$
$R(p)$
$D(p \mid p_0) \geq \frac{1}{2\eta}\|p - p_0\|^2$
$R(p_0) + \langle \nabla R(p_0), p - p_0 \rangle$
$R(p_0)$
p_0
p
p

Figure 1: Pictorial distribution of strong convexity of a function.

Proof: By the definition of strong convexity of $f(\cdot)$ for any $p \in S$:

$$\begin{aligned} A = f(p) - f(p_f) \geq\ & \langle \nabla f(p_0), p - p_f \rangle + \frac{1}{2\eta} \cdot \|p - p_f\|^2 && \text{(by } \tfrac{1}{\eta}\text{-strong convexity of } f) \\ \geq\ & \frac{1}{2\eta} \cdot \|p - p_f\|^2 && \text{(since } p_f \text{ is minimizer, } \langle \nabla f(p_0), p - p_f \rangle \geq 0) \end{aligned}$$

Similarly, we can show that $B = g(p) - g(p_g) \geq \frac{1}{2\eta} \cdot \|p - p_g\|^2$.

Finally, we can conclude by invoking Lipschitzness of the difference function:

$$\begin{aligned} L \cdot \|p_f - p_g\| \geq\ & h(p_f) - h(p_g) && \text{(by lipschitzness of } h) \\ =\ & \underbrace{g(p_f) - f(p_f)}_{C} - \underbrace{(g(p_g) - f(p_g))}_{D} && \text{(by definition of } h) \\ =\ & \underbrace{g(p_f) - g(p_g)}_{A} + \underbrace{f(p_g) - f(p_f)}_{B} && \text{(by re-arranging)} \\ \geq\ & \frac{1}{2\eta}\|p_f - p_g\|^2 + \frac{1}{2\eta}\|p_g - p_f\|^2 && \text{(by strict convexity of } f(\cdot) \text{ and } g(\cdot)) \\ =\ & \frac{1}{\eta}\|p_f - p_g\|^2 \end{aligned}$$

Dividing the above inequality by $\|p_f - p_g\|$ yields the Lemma. □

With this generalization of the stability lemma for strongly convex functions we are now read to prove the generalization of the regret bound for the FTRL algorithm from the previous lecutre.

Theorem 1. *The expected regret of FTRL with a* 1*-strongly convex regularizer* $R(p)$ *with respect to a norm* $\|\cdot\|$*, a parameter* η*, and* L*-Lipschitz loss functions with respect to the norm* $\|\cdot\|$*, is upper bounded by:*

$$\textsc{Regret}(T) \leq \frac{\max_{p \in S} R(p) - \min_{p \in S} R(p)}{\eta} + \eta \cdot L^2 \cdot T \tag{4}$$

Proof: We will show this again in two lemmas. Let's consider a slight modification of the FTRL algorithm which includes the next iteration's loss vector in the optimization:

$$\text{Be-the-Regularized-Leader (BTRL): } \tilde{p}_t^* = \operatorname*{argmin}_{p \in S} \tilde{F}_t(p) \tag{5}$$

Lemma 1.1 (BTRL vs FTRL)**.** *The regret of FTRL is at most the regret of BTRL plus* $L \cdot \sum_{t=1}^{T} |\tilde{p}_t - \tilde{p}_{t+1}|$*. The latter is upper bounded by* $\eta \cdot L^2 \cdot T$*.*

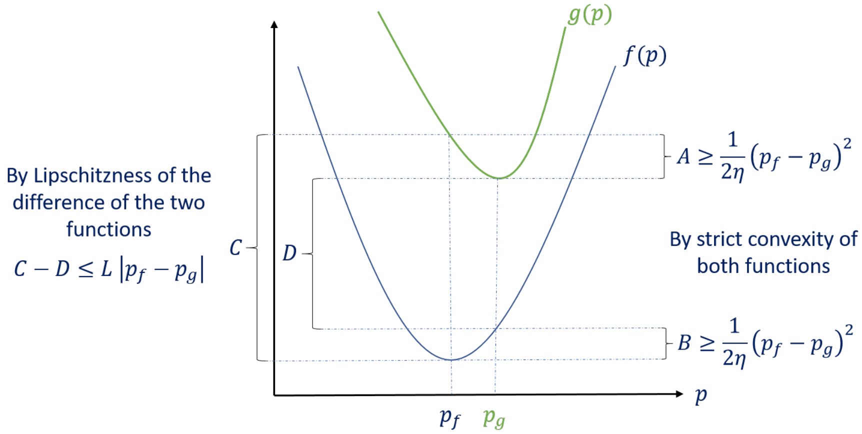
$g(p)$
$f(p)$
By Lipschitzness of the difference of the two functions
$C - D \leq L\,|p_f - p_g|$
C
D
$A \geq \frac{1}{2\eta}(p_f - p_g)^2$
By strict convexity of both functions
$B \geq \frac{1}{2\eta}(p_f - p_g)^2$
p
p_f
p_g

Figure 2: The proof of Lemma 1 follows immediately by noting that $C - D = A + B$ in the above figure, together with the fact that $C - D \leq L|p_f - p_g|$ by Lipschitzness of the difference of the two functions and $A + B \geq \frac{1}{\eta}(p_f - p_g)^2$ by the strict convexity of the two functions.

Proof: We will show that the loss of the FTL algorithm is at most the loss of the BTL algorithm plus the stability term.

$$\begin{aligned}
\sum_{t=1}^{T} f_t(\tilde{p}_t;\ell_t) = {} & \sum_{t=1}^{T} f_t(\tilde{p}_t^*) + \sum_{t=1}^{T} (f_t(\tilde{p}_t) - f_t(\tilde{p}_t^*)) \\
\leq {} & \sum_{t=1}^{T} f(p_t^*;\ell_t) + L \sum_{t=1}^{T} \|p_t - p_t^*\| && \text{(by } L\text{-Lipschitzness of } f_t\text{)} \\
\leq {} & \sum_{t=1}^{T} f(p_t^*;\ell_t) + \sum_{t=1}^{T} \|p_t - p_{t+1}\| && \text{(by the definition of FTRL and BTRL, } \tilde{p}_t^* = \tilde{p}_{t+1}\text{)}
\end{aligned}$$

The second part follows by observing that $\tilde{p}_t$ and $\tilde{p}_{t+1}$ are minimizers of $\tilde{F}_{t-1}(\cdot)$ and $\tilde{F}_t(\cdot)$ respectively. The latter are $\frac{1}{\eta}$-strongly convex functions. Moreover, $\tilde{F}_{t+1}(p) - \tilde{F}_t(p) = f_t(p)$, which is an L-Lipschitz function. Thus invoking Lemma 1, we get that $|\tilde{p}_t - \tilde{p}_{t+1}| \leq \eta \cdot L$. □

Lemma 1.2 (BTRL has small regret)**.** *The regret of BTRL is upper bounded by* $\frac{\max_{p\in S} R(p) - \min_{p\in S} R(p)}{\eta}$.

Proof: Let $f_0(p) = \frac{1}{\eta}R(p)$ and $\tilde{p}_0^* = \operatorname{argmin}_{p\in S} \frac{1}{\eta}R(p)$. Suppose by induction that:

$$\underbrace{\sum_{\tau=0}^{t} f_\tau(\tilde{p}_\tau^*)}_{\substack{\text{loss of BTL until time } t\\ \text{including fake loss at } \tau=0}} \quad \leq \quad \underbrace{\tilde{F}_t(\tilde{p}_t^*)}_{\substack{\text{cummulative loss of best fixed action until time } t\\ \text{including fake loss at } \tau=0}} \tag{6}$$

Then we conclude the induction step that:

$$\begin{aligned}
\sum_{\tau=0}^{t+1} f_\tau(\tilde{p}_\tau^*) = {} & f_{t+1}(\tilde{p}_{t+1}^*) + \sum_{\tau=0}^{t} f_\tau(\tilde{p}_\tau^*) \\
\leq {} & f_{t+1}(\tilde{p}_{t+1}^*) + \tilde{F}_t(\tilde{p}_t^*) && \text{(by induction hypothesis)} \\
\leq {} & f_{t+1}(\tilde{p}_{t+1}^*) + \tilde{F}_t(\tilde{p}_{t+1}^*) && \text{(by optimality of } \tilde{p}_t^*\text{)} \\
= {} & \tilde{F}_{t+1}(\tilde{p}_{t+1}^*) && \text{(by definition of } F_t(p)\text{)}
\end{aligned}$$

which concludes the induction step. Finally, the base case of $t = 0$ is satisfied by the definition of p_0^*.

Thus we can conclude that:

$$\sum_{t=0}^{T} f_t(\tilde{p}_t^*) \le \min_{p\in S} \sum_{t=0}^{T} f_t(p) \le \max_{p\in S} f_0(p) + \min_{p\in S} \sum_{t=1}^{T} f_t(p)$$

By re-arranging we get:

$$\sum_{t=1}^{T} f_t(\tilde{p}_t^*) - \min_{p\in S} \sum_{t=1}^{T} f_t(p) \le \max_{p\in S} \frac{1}{\eta} R(p) - \min_{p\in S} \frac{1}{\eta} R(p)$$

□

Combining the two lemmas yields the theorem. □

An easy corollary of the latter theorem is that:

Corollary 1. *Let $R^* = \max_{p\in S} R(p) - \min_{p\in S} R(p)$. Then expected regret of FTRL with $\eta = \sqrt{\frac{R^*}{T}}$ is at most:*

$$\textsc{Expected-Regret}(T) \le 2\sqrt{2R^* \cdot T} \tag{7}$$

1.1 Expert Setting and Negative Entropy Regularization

Let's now look at the expert setting with K actions and at the case when the regularizer is the negative entropy:

$$R(p) = \sum_{i=1}^{K} p^i \log(1 - p^i) \tag{8}$$

We first show that the entropy function is 1-strongly convex with respect to the ℓ_1 norm $\|x\|_1 = \sum_{i=1}^{K} |x^i|$.

Lemma 2 (Strong convexity of negative entropy)**.** *The negative entropy function is* 1*-strongly convex with respect to the ℓ_1 norm $\|x\|_1 = \sum_{i=1}^{K} |x^i|$.*

Proof: Strong convexity is equivalent to showing that $x^T \nabla^2 R(p) x \ge \|x\|_1^2$ for all $p \in \Delta_K$ and $x \in \mathbb{R}^d$, where $\nabla^2 R(\cdot)$ is the Hessian of function $R(\cdot)$, i.e. the matrix whose (i, j) entry is $\frac{\partial^2 R(p)}{\partial p^i \partial p^j}$.

For the case of the negative entropy function we have:

$$\frac{\partial^2 R(p)}{\partial p^i \partial p^j} = \begin{cases} \frac{1}{p_i} & \text{if } i = j \\ 0 & \text{o.w.} \end{cases} \tag{9}$$

Thus we have:

$$\begin{aligned} x^T \nabla^2 R(p) x &= \sum_{i=1}^{K} \frac{(x^i)^2}{p^i} \\ &= \left(\sum_{i=1}^{K} p^i\right) \cdot \left(\sum_{i=1}^{K} \frac{(x^i)^2}{p^i}\right) && \text{(since } p \text{ is a distribution)} \\ &\ge \sum_{i=1}^{K} \sqrt{p^i} \frac{|x^i|}{\sqrt{p^i}} && \text{(by Cauchy-Schwarz inequality)} \\ &\ge \sum_{i=1}^{K} |x^i| \end{aligned}$$

□

For this type of a regularizer we can find in closed form the optimal solution to the problem that FTRL:

$$\tilde{p}_t = \underset{p\in[0,1]}{\text{argmin}} \quad \langle p, \sum_{\tau=1}^{t-1} \ell_\tau \rangle + \frac{1}{\eta} \sum_{i=1}^{K} p^i \log(p^i)$$

By taking the first order condition of the Langrangian of the optimization problem, where we include the constraint that $\sum_{i=1}^K p^i = 1$

$$\mathcal{L}(p, \lambda) = \langle p, \sum_{\tau=1}^{t-1} \ell_\tau \rangle + \frac{1}{\eta} \sum_{i=1}^{K} p^i \log(p^i) + \lambda \left(1 - \sum_{i=1}^{K} p_i\right)$$

We can conclude that

$$\tilde{p_t^i} = \frac{\exp\{-\eta \cdot \sum_{\tau=1}^{t-1} \ell_\tau^i\}}{\sum_{j=1}^{K} \exp\{-\eta \cdot \sum_{\tau=1}^{t-1} \ell_\tau^j\}}$$

A nice way of looking at the latter algorithm, is that at each iteration we keep a weight w_t^i for each action $i \in [K]$, starting from $w_0^i = \frac{1}{K}$. Then at each iteration we update the weight of each algorithm as:

$$w_{t+1}^i = w_t^i \cdot \exp\{-\eta \cdot \ell_t^i\} \tag{10}$$

and at each iteration we play each action with probability proportional to its weight!

As we pointed out in the previous lecture, the above algorithm is a very well-known one, known under various names, such as *exponential weight updates*, *EXP*, *multiplicative weight updates*, *Hedge*, ..., and has found many applications in many areas of computer science even ones not directly related to online learning.

Our general regret bound for the Follow-the-Regularized-Leader algorithm implies the following corollary for the expert setting:

Corollary 2. *The Exponential Weight Updates algorithm with parameter* $\eta = \sqrt{\frac{\log(K)}{T}}$ *achieves expected regret:* $2\sqrt{\log(K) \cdot T}$ *in the expert setting with losses in* $[0, 1]$.

Proof: Since linear loss functions with losses in $[0, 1]$, are 1-Lipschitz with respect to the $\|\cdot\|_1$ norm and since the negative entropy is a 1-strongly convex regularizer with respect to the $\|\cdot\|_1$ norm, and also satisfies that $R(p) \in [-\log(K), 0]$, we get by Theorem 1 that the regret of Exponential Weight Updates is:

$$\text{Regret}(T) \leq \frac{\log(K)}{\eta} + \eta T \tag{11}$$

By the choice of η we get the corollary. □

Importantly, the regret of the Exponential Weights algorithm increases only logarithmically with the number of actions available!

2 Playing Zero-Sum Games with Online Learning Algorithms

In the first lecture we considered zero sum games between two players defined by a $m \times n$ loss matrix A (equivalently a payoff matrix), where m is the number of actions of the *row* player and n is the number of actions of the column player. If the row player plays i and the column player plays j, then the row player receives a loss of $A_{ij} \in [0, 1]$ and the column player a loss of $-A_{ij}$ (equivalently a reward of A_{ij}. Thus the row player is trying to minimize the entry of the matrix that is picked and the column player to maximize it.

If both players are randomizing, and if we denote with $x \in \Delta_m$ the probability distribution of the row player and with $y \in \Delta_n$ the probability distribution of the column player, then the expected loss of the row player is x^TAy and similarly this is the reward of the column player. Von-Neumann's minimax theorem states that:

$$\min_{x\in\Delta_m}\max_{y\in\Delta_n} x^TAy = \max_{y\in\Delta_n}\min_{x\in\Delta_m} x^TAy \tag{12}$$

Moreover, each of this optimization problems can be phrased as an LP and the pair $(\bar{x},\bar{y})$ of solutions to each of these problems constitutes a Nash equilibrium of the game.

We will show here that the proof of this theorem immediately follows by the existence of a no-regret algorithm for the expert setting. We will do this as follows: suppose that the zero-sum game is played repeatedly for T iterations. On each day t the row player picks a strategy x_t and the column player a strategy y_t based on the past. Then the row player incurs a loss $\langle x_t^T, Ay_t\rangle$ and the column player a loss $-\langle y_t, A^Tx_t\rangle$. Thus both players are essentially facing an online learning problem and in particular their problem is an expert setting problem.

Thus we will imagine both players deciding their strategy on each day based on a no-regret algorithm. We saw in the last section that such algorithms exist.

Theorem 2. *If there exists a no-regret algorithm for the expert setting, then*

$$\min_{x\in\Delta_m}\max_{y\in\Delta_n} x^TAy = \max_{y\in\Delta_n}\min_{x\in\Delta_m} x^TAy \tag{13}$$

Moreover, if we allow $(x_1,y_1),\ldots,(x_T,y_T)$ is history of play when we both player's use a no-regret algorithm with average regret $\epsilon(T)$, then the strategies $\bar{x}=\frac{1}{T}\sum_{t=1}^T x_t$ and $\bar{y}=\frac{1}{T}\sum_{t=1}^T y_t$ are an $2\epsilon(T)$-approximate Nash equilibrium, i.e.:

$$\bar{x}^TA\bar{y} \leq \min_x x^TA\bar{y} + 2\epsilon(T) \tag{14}$$

$$\bar{x}^TA\bar{y} \geq \max_y \bar{x}^TAy - 2\epsilon(T) \tag{15}$$

$$\tag{16}$$

Proof: First observe that by the $\epsilon(T)$-regret definition for each player we have:

$$V = \frac{1}{T}\sum_{t=1}^T x_t^TAy_t \leq \min_x \frac{1}{T}\sum_{t=1}^T x^TAy_t + \epsilon(T) = \min_x x^TA\bar{y} + \epsilon(T) \tag{17}$$

$$V = \frac{1}{T}\sum_{t=1}^T x_t^TAy_t \geq \max_y \frac{1}{T}\sum_{t=1}^T x_t^TAy - \epsilon(T) = \max_y \bar{x}^TAy - \epsilon(T) \tag{18}$$

Thus these two inequalities immediately yield:

$$\min_x x^TA\bar{y} \geq \max_y \bar{x}^TAy - 2\epsilon(T) \tag{19}$$

The latter essentially is the crucial property that we need. From here on it's a matter of combining this property with a bunch of trivial inequalities to conclude both parts of the theorem.

Part 1. For the minimax theorem we observe that:

$$\max_y\min_x x^TAy \geq \min_x x^TA\bar{y} \geq \max_y \bar{x}^TAy - 2\epsilon(T) \geq \min_x\max_y x^TAy - 2\epsilon(T) \tag{20}$$

Combining the above with the trivial inequality that $\max_y\min_x x^TAy \leq \min_x\max_y x^TAy$, and taking $T\to\infty$, which takes $\epsilon(T)\to 0$, yields the result.

Part 2. For the $2\epsilon(T)$-equilibrium we observe that:

$$\begin{aligned}
\bar{x}^T A \bar{y} \leq \quad & \max_y \bar{x}^T A y \leq \min_x x^T A \bar{y} + 2\epsilon(T) \\
\bar{x}^T A \bar{y} \geq \quad & \min_x x^T A \bar{y} \geq \max_y \bar{x}^T A y - 2\epsilon(T)
\end{aligned}$$

□

An intersting point to note, is that we did not really need both players to use a no-regret algorithm. The column player for instance, could be always best responding to the action of the row player. This is trivially a no-regret algorithm for the column player. This property has been utilized in the literature, when the actions of the column player are given implicitly and are exponential in the size of their representation, but when we have a polynomial time algorithm for computing the best action conditional on the strategy of the row player. Hence, in such situations we can still compute an approximate Nash equilibrium in polynomial time.

Moreover, as we will ask you in your assignment, the existence of no-regret algorithms for the more general setting of online convex optimization immediately imply the general version of von-Neumann's minimax theorem, which pertains to non-linear zero-sum games. In particular, consider a two player zero-sum game where player 1 picks a strategy $x \in S$ where S is convex subset of $\mathbb{R}^m$ and player 2 picks a strategy $y \in Q$ where Q is a convex subset of $\mathbb{R}^n$. When player 1 picks x and player 2 picks y, player 1 receives a loss of $c(x, y)$ and player 2 receives a loss of $-c(x, y)$. If the function $c(x, y)$ is convex in its first argument and concave in its second argument, then:

$$\min_{x \in S} \max_{y \in Q} c(x, y) = \max_{y \in Q} \min_{x \in S} c(x, y) \tag{21}$$

If we imagine this zero-sum game played repeatedly and each player using an online convex optimization algorithm to decide its strategy at each time step t, then we can prove the latter minimax theorem, as well as the fact that their average strategies $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$ and $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$, constitute an approximate Nash equilibrium of the zero-sum game.

3 Historical Remarks

The online learning framework dates back to the very early work of Hannan [Han57] on consistency and of Blackwell [B$^+$56] on approachability, which are both terms that are very closely related to achieving the no-regret condition. The exponential weights algorithm dates back to the early and very influential work of Littlestone and Warmuth [LW94] and Freund and Schapire [FS97] and Kivinen and Warmuth [KW97]. The analysis of Hedge, through the Follow-the-Regularized-Leader lens is due to Shalev-Shwartz and Singer [SSS07a, SSS07b]. A very comprehensive survey of Follow-the-Regularized-Leader and related algorithms can be found in [SS12].

The proof of the minimax theorem through online learning is due to [FS99], who analyzed the multiplicative weights algorithm in zero-sum games. No-regret learning in appropriately defined zero sum games has found tremendous applications in many areas of computer science (see e.g. some random samples [HHRW16], [PST95]) and a survey here [AHK12].

4 Acknowledgments

Based on lecture notes by Vasilis Syrgkanis for the Spring 2017 course "Algorithmic Game Theory and Data Science" by Daskalakis and Syrgkanis.

References

[AHK12] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(6):121–164, 2012.

[B^+56] David Blackwell et al. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[FS97] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, August 1997.

[FS99] Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1):79 – 103, 1999.

[Han57] James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.

[HHRW16] Justin Hsu, Zhiyi Huang, Aaron Roth, and Zhiwei Steven Wu. Jointly private convex programming. In *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 580–599, Philadelphia, PA, USA, 2016. Society for Industrial and Applied Mathematics.

[KW97] Jyrki Kivinen and Manfred K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Inf. Comput.*, 132(1):1–63, January 1997.

[LW94] N. Littlestone and M.K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212 – 261, 1994.

[PST95] Serge A. Plotkin, David B. Shmoys, and Éva Tardos. Fast approximation algorithms for fractional packing and covering problems. *Math. Oper. Res.*, 20(2):257–301, April 1995.

[SS12] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2012.

[SSS07a] Shai Shalev-Shwartz and Yoram Singer. Online learning: Theory, algorithms, and applications. 2007.

[SSS07b] Shai Shalev-Shwartz and Yoram Singer. A primal-dual perspective of online learning algorithms. *Mach. Learn.*, 69(2-3):115–142, December 2007.