# Lecture 09

[skip MWA is ERM]

Other ERM algs:

e.g. Sampled ficticious play. Li. Tewari, 2017 (is Hannan consistent

FTPL, etc.

Last time:

$$(\text{ERM}) \quad \limsup_{T\to\infty} \frac{1}{T} \underbrace{\max_{a_i \in A_i} \left( \sum_{t=0}^{T-1} u_i(a_i, a_{-i}^t) - u_i(a_i^t, a_{-i}^t) \right)}_{ER_i} \leq 0. \quad \text{a.s.}$$

randomization by all players

- If all players use ERM strategies, then a.s.

$$\limsup_{T\to\infty} \sum_{a\in A} q^T(a)\left(u_i(a_i', a_{-i}) - u(a)\right) \leq 0 \quad (\text{limiting behaviors})$$

$\Rightarrow$ Coarse Correlated Equilibrium:

$$CCE := \left\{ q \in \Delta(A) : \sum_{a\in A} q(a) u_i(a_i', a) \leq \sum_{a\in A} q(a) u_i(a), \ \forall\, i, a_i' \right\}$$

- Recall

$$CE := \left\{ q \in \Delta(A) : \sum_{a_{-i}\in A_{-i}} q(a_{-i}|a_i) u_i(a_i', a) \leq \sum_{a_{-i}\in A_{-i}} q(a_{-i}|a_i) u_i(a), \ \forall\, i, a_i, a_i' \right\}$$

$$\Longleftrightarrow \quad CE := \left\{ q \in \Delta(A) : \sum_{a_{-i}\in A_{-i}} q(a_i, a_{-i}) u_i(a_i', a) \leq \sum_{a_{-i}\in A_{-i}} q(a_i, a_{-i}) u_i(a), \ \forall\, i, a_i, a_i' \right\}$$

- $CE \subseteq CCE$

  ? $\qquad$ ERM

Consider the following Internal Regret Minimization:

$$\limsup_{T\to\infty} \frac{1}{T} \max_{a_i, a_i' \in A_i} \sum_{t=0}^{T-1} \mathbb{I}(a_i^t = a_i)\left(u_i(a_i', a_{-i}^t) - u_i(a_i, a_{-i}^t)\right) \quad \text{a.s.}$$

(replace every $a_i^t = a_i$ by $a_i'$)

- If all players use IRM strategies, then

$$\limsup_{T\to\infty} \sum_{a_i \in A_{-i}} q(a)\left(u_i(a_i', a_{-i}) - u_i(a_i, a_{-i})\right) \leq 0$$

$\Rightarrow$ Theorem 9.1

If all players use IRM strategies, then the joint empiracal distribution approaches the set of correlated equilibria (a.s.).

- What strategies are IRM?
- Consider this generalized / unified regret definition:

$$R_{\mathcal{F}} := \max_{F \in \mathcal{F}} \left\{ U^T_{\pi,\mathcal{F}} - U_\pi \right\}$$

Modification rule

Given a mixed strategy $p^t$ from $\pi$, and a modification rule $F^t$:

$$f^t = F^t(p^t)$$

where $f_i^t = \sum_{j: F^t(j)=i} p_j^t$ $\qquad A \to A$

- $\mathcal{F} := \{F_t : t \geq 1\}$
- $U^T_{\pi,\mathcal{F}}$ is the profit induced by $(f^t : t \geq 1)$.

1. Recall external regret:

$$ER := \underbrace{U^T_{\theta,\max}}_{\max_{\sigma\in\theta} U^T_\sigma} - U^T_\pi$$

$\sigma \in \theta \to$ comparison class of algs

equivalent to $\mathcal{F}^{ex} := \{F_i : i \in A\}$ where $F_i(j) = i \quad \forall\, i,j \in A$

2. Internal regret =

$$IR := \max_{i,j \in A} \left\{ \sum_{t=1}^T p_i^t (u_j^t - u_i^t) \right\}$$

mixed strategies by $\pi$ ; fixing other players' strategies

Q? equivalent to $\mathcal{F}^{in} := \left\{ F_{ij} : \begin{cases} F_{ij}(i) = j \\ F_{ij}(i') = i' \quad \forall\, i' \neq i \end{cases},\ \begin{matrix} i,j \in A \\ i \neq j \end{matrix} \right\}$

$|\mathcal{F}^{in}| = |A|(|A|-1)$

- In general, how many distinct $F : A \to A$ are there?

$|A|^{|A|}$

- Each $F$ swaps the current online action $i$ w/ $F(i)$

Swap regret : $SR := \max_{F \in \mathcal{F}^{sw}} \left\{ U^T_{\pi \cdot F} - U_\pi \right\}$

$$= \sum_{i=1}^{|A|} \max_{j \in A} \left\{ \sum_{t=1}^{T} p_i^t ( u_j^t - u_i^t ) \right\}$$

- We see examples of ERM policies, such as MWA, sampled FP, leading to CCE.

What are IRM policies for CE?

- A general reduction from ER to IR (in fact, SR).

[ Blum & Mansour, COLT 2005 ]

Create $N = |A|$ copies of the external-regret procedures $\pi_1, \ldots, \pi_N$, each of them satisfying

not exactly ERM, can generalize

$\forall$ sequence of $T$ payoffs (losses) $(u^t : t = 1 \ldots T)$, $\forall j \in \{1, \ldots, N\}$,

$$U^T_{\pi_i} = \sum_{t=1}^{T} u^t_{\pi_i} \leq \sum_{t=1}^{T} u_j^t + R = U_j^T + R$$

We combine the $N$ procedures to one master procedure as follows. At each time $t$, each procedure $\pi_i$ outputs a distribution $q_i^t = ( q_{ij}^t : j = 1, \ldots, N )$.

fraction it assigns to action $j$

$\{ q_{ij}^t, i, j \in A \}$

Compute $p_j^t = \sum_i p_i^t q_{ij}^t \iff p^t = Q^t p^t$

a stationary distribution of the Markov Process defined by $Q^t$.

[ exercise: show this $p^t$ always exists ]

Choosing $p^t$ can be regarded in two equivalent ways:

(1). Using $p^t$ to select action $j$ w/ probability $p_j^t$ (actual payoff $p^t \cdot u^t$)

(2). Select procedure $\pi_i$ w/ probability $p_i^t$ and then use $\pi_i$ to select the action ( $Q^t p^t$ )

After receiving a payoff ( full information model ) $u^t$, we return to each $\pi_i$ the vector $p_i^t u^t$. So, procedure $\pi_i$ experiences

$$(p_i^t u^t) \cdot q_i^t = p_i^t (q_i^t \cdot u^t)$$

inner product

Since $\pi_i$ is an $R$ external regret procedure, $\forall$ action $j$,

$$\sum_{t=1}^{T} p_i^t (q_i^t \cdot u^t) \leq \sum_{t=1}^{T} p_i^t u_j^t + R \quad \text{——} \ (*)$$

Summing the payoffs of the $N$ procedures, at a given time $t$,

$$\sum_i p_i^t (q_i^t \cdot u^t) = \underbrace{(Q^t p^t)}_{= p^t} \cdot u^t = \underbrace{p^t \cdot u^t}_{\text{actual payoff}}$$

Hence, $(*)$ gives ( summing over $i = 1, \dots, N$ ) $\qquad u_\pi^T = \sum_{t=1}^{T} p^t \cdot u^t$

$$u_\pi^T \leq \sum_{i=1}^{N} \sum_{t=1}^{T} p_i^t u_{F(i)}^t + NR = u_{\pi, F}^T + NR$$

for any function $F: \underbrace{\{1, \dots, N\}}_{A} \to \underbrace{\{1, \dots, N\}}_{A}$.

In summary.

**Theorem 9.2**

Given an $R$ external regret procedure, the constructed master online procedure $\pi$ has the following guarantee. $\forall\ F: A \to A$,

$$U_\pi \leq U_{\pi \cdot F} + NR$$

i.e., the swap regret of $\pi$ is at most $NR$.

Hence.

**Corollary 9.1**

combining MWA. for example

There exists an online algorithm $\pi$ s.t. for every function $F: A \to A$, we have that

$$U_\pi \leq U_{\pi, F} + O(N\sqrt{T \log N})$$

i.e., the swap regret of $\pi$ is at most $O(N\sqrt{T\log N})$.

- Read Section 4.5, 4.6 in the book, which can be found in our reading materials for more detailed discussions on the partial information setting.

- More on online learning algs.

- Recall: MWA assumptions:

(1) The set of allowed actions for the player is the probability simplex

$$K_N = \{ p \in \mathbb{R}_+^N, \ \sum_{i=1}^{N} p_i = 1 \}$$

(2) Loss / Payoff functions are linear, $f^t(p) = m^t \cdot p$ where $m^t \in \mathbb{R}^N$ is normalized s.t. $|m_i^t| \leq 1 \ \forall \ i \in \{1, \dots, N\}$. This ensures $f^t(p) \in [-1, 1], \forall p$.

Given this, MWA is a special instance of

Follow the Regularized leader (FTRL)

$$p^{t+1} = \arg\min_{p \in K_N} \left\{ \eta \sum_{j \leq t} m^j \cdot p + \underbrace{R(p)}_{= -H(p)} \right\}$$

- Now, let's turn to online convex optimization (OCO):

Goal: solve $\min_{x \in K} \sum_{t=1}^{T} f_t(x)$ online

- $K$ is bounded, convex, closed
- $f_t : K \to \mathbb{R}$ is convex.

Similar to our previous discussion for MWA, the FTL / ficticious play scheme below fails in the worst case:

$$x_{t+1} = \arg\min_{x \in K} \sum_{\tau=1}^{t} f_\tau(x)$$

- Consider $K = [-1, 1]$, $f_1(x) = \frac{1}{2}x$, $f_\tau(x)$ for $\tau = 2, \dots, T$ alternate between $-x$ and $x$. Thus,

$$\sum_{\tau=1}^{t} f_\tau(x) = \begin{cases} \frac{1}{2}x & t \text{ is odd} \\ -\frac{1}{2}x & t \text{ is even} \end{cases}$$

$\Rightarrow$ FTL strategy keeps shifting beween $x_t = -1$ and $x_t = 1$, making the wrong choices.

- Consider regularization functions $R: K \to \mathbb{R}$ that are strongly convex, smooth, and twice differentiable.

Hence, by strong convexity, the Hessian $\nabla^2 R(x) \succ 0$ is positive definite.

Define the diameter of $K$ as

$$D_R := \left( \max_{x,y \in K} \{ R(x) - R(y) \} \right)$$

- Dual norm of $\|\cdot\|$ : $\quad \|y\|^* := \sup_{\|x\| \leq 1} \{ x^T y \}$.

Dual norm of the matrix norm $\|x\|_A = \sqrt{x^T A x}$ : $\quad \|x\|_A^* = \|x\|_{A^{-1}}$

For notational simplicity, we write

Generalized Cauchy Swartz inequality : $x^T y \leq \|x\|_A \cdot \|y\|_A^*$

any norm

$$\|x\|_y := \|x\|_{\nabla^2 R(y)}$$

$$\|x\|_y^* := \|x\|_{\nabla^{-2} R(y)}$$

- Difference beween the value of the regularization function at $x$ and the value of the 1st order Taylor approximation :

Bregman divergence: $\quad B_R(x \| y) := R(x) - R(y) - \nabla R(y)^T (x-y)$.

For twice differentiable functions, Taylor expansion and the mean-value theorem implies

$$B_R(x \| y) = \frac{1}{2} \|x-y\|_z^2$$ for some $z$, and $\exists \alpha \in [0,1]$ s.t. $z = \alpha x + (1-\alpha) y$.

Thus, the Bregman divergence defines a local norm, which has a dual norm $\|\cdot\|_{x,y}^* := \|\cdot\|_z^*$

Finally, we write $\|\cdot\|_t = \|\cdot\|_{x_t, x_{t+1}}$ so that $B_R(x_t \| x_{t+1}) = \frac{1}{2} \|x_t - x_{t+1}\|_t^2$

To wrap up, let's summarize and revisit this FTRL meta-algorithm:

Algorithm FTRL

Input: $\eta > 0$, regularization function $R$, and $K$

let $x_1 = \arg\min_{x \in K} \{ R(x) \}$

for $t = 1, \dots, T$ do

play $x_t$ and receive cost $f_t(x_t)$ (can relax from the full information setting to the bandit gradient setting, $\nabla_t := \nabla f_t(x_t)$)

update

$$x_{t+1} = \arg\min_{x \in K} \left\{ \eta \sum_{s=1}^{t} f_s(x) + R(x) \right\}$$

relax to $\nabla_s^T \cdot x$

since $\forall t$

$(\square) \leftarrow f_t(x_t) - f_t(x^*) \leq \nabla_t^T (x_t - x^*)$

by convexity

end for

- Theoretical Guarantee:

Theorem 9.3

The FTRL algorithm attains $\forall u \in K$ the following regret bound:

$$ER(\text{FTRL}) \leq 2\eta \sum_{t=1}^{T} (\|\nabla_t\|_t^*)^2 + \frac{R(u) - R(x_1)}{\eta}.$$

- Note that if $\|\nabla_t\|_t^* \leq G_R$ $\forall t$, then optimizing over $\eta$ gives a bound of $2 D_R G_R \sqrt{2T}$.

Proof: Lemma: FTRL guarantees $ER(\text{FTRL}) \leq \sum_{t=1}^{T} \nabla_t^T (x_t - x_{t+1}) + \frac{1}{\eta} D_R^2$.

$\downarrow$

proof: Define $g_0(x) := \frac{1}{\eta} R(x)$,

$g_t(x) := \nabla_t^T \cdot x$

By $(\square)$, it suffices to bound $\sum_{t=1}^{T} [g_t(x) - g_t(u)]$. We first show that $\forall u \in K$,

$$\sum_{t=0}^{T} g_t(u) \geq \sum_{t=0}^{T} g_t(x_{t+1})$$

To see it, we use induction on $T$:

- Induction base: by definition, $x_1 = \underset{x \in K}{\arg\min}\, R(x)$, thus $g_0(u) \geq g_0(x_1)\ \forall u$.
- Induction step:

Assume for $T$, we have $\sum_{t=0}^{T} g_t(u) \geq \sum_{t=0}^{T} g_t(x_{t+1})$

For $T+1$, since $x_{T+2} = \underset{x \in K}{\arg\min} \left\{ \sum_{t=0}^{T+1} g_t(x) \right\}$,

$$
\begin{aligned}
\sum_{t=0}^{T+1} g_t(u) &\geq \sum_{t=0}^{T+1} g_t(x_{T+2}) \\
&= \sum_{t=0}^{T} g_t(x_{T+2}) + g_{T+1}(x_{T+2}) \\
&\geq \underbrace{\sum_{t=0}^{T} g_t(x_{t+1})}_{\text{induction hypothesis } u = x_{T+2}} + g_{T+1}(x_{T+2}) \\
&= \sum_{t=0}^{T+1} g_t(x_{t+1})
\end{aligned}
$$

As a conclusion,

$$
\begin{aligned}
\sum_{t=1}^{T} [g_t(x_t) - g_t(u)] &\leq \sum_{t=1}^{T} [g_t(x_t) - g_t(x_{t+1})] + [g_0(u) - g_0(x_1)] \\
&= \sum_{t=1}^{T} [g_t(x_t) - g_t(x_{t+1})] + \frac{1}{\eta}[R(u) - R(x_1)] \\
&\leq \sum_{t=1}^{T} [g_t(x_t) - g_t(x_{t+1})] + \frac{1}{\eta} D_R^2
\end{aligned}
$$

#

Now, since $R(x)$ is a convex function and $K$ is a convex set.

Denote $\Phi_t(x) := \eta \sum_{s=1}^{t} \nabla_s^T x + R(x)$.

The Taylor expansion implies

$$
\begin{aligned}
\Phi_t(x) &= \Phi_t(x_{t+1}) + (x_t - x_{t+1})^T \nabla \Phi_t(x_{t+1}) + B_{\Phi_t}(x_t \| x_{t+1}) \\
&\geq \Phi_t(x_{t+1}) + B_{\Phi_t}(x_t \| x_{t+1}) \quad \text{since } x_{t+1} \text{ minimizes } \Phi_t \text{ over } K \\
&= \Phi_t(x_{t+1}) + B_R(x_t \| x_{t+1}) \quad \text{since the term } \nabla_s^T \cdot x \text{ is linear, it won't affect the Bregman divergence.}
\end{aligned}
$$

Rearrange the terms, we get

$$B_R(x_t \| x_{t+1}) \leq \Phi_t(x_t) - \Phi_t(x_{t+1})$$

$$= \left(\Phi_{t-1}(x_t) - \Phi_{t-1}(x_{t+1})\right) + \eta \nabla_t^T (x_t - x_{t+1})$$

$$\leq \eta \nabla_t^T (x_t - x_{t+1}) \quad (x_t \text{ minimizes } \Phi_{t-1} \text{ again})$$

Using our notation, the generalized Cauchy - Swartz inequality.

$$\nabla_t^T (x_t - x_{t+1}) \leq \|\nabla_t\|_t^* \cdot \|x_t - x_{t+1}\|_t$$

$$= \|\nabla_t\|_t^* \cdot \left(2 B_R(x_t \| x_{t+1})\right)^{\frac{1}{2}}$$

$$\leq \|\nabla_t\|_t^* \cdot \left(2 \eta \nabla_t^T (x_t - x_{t+1})\right)^{\frac{1}{2}}$$

$$\Rightarrow \nabla_t^T (x_t - x_{t+1}) \leq 2\eta \left(\|\nabla_t\|_t^*\right)^2.$$

Substituting this into the lemma complete the proof of Theorem 9.3.

#

- connection to online mirror descent (OMD)

OMD is a general class of 1st order methods extending GD.

It has two versions, agile and lazy.

**Algorithm OMD**

Input: $\eta > 0$, regularization function $R(x)$.

Let $y_1$ be s.t. $\nabla R(y_1) = 0$

Let $x_1 = \underset{x \in K}{\arg\min}\ B_R(x \| y_1)$

for $t = 1, \dots, T$, do

play $x_t$

observe the loss $f_t$, let $\nabla_t = \nabla f_t(x_t)$

update $y_t$ according to

$\nabla R(y_{t+1}) = \nabla R(y_t) - \eta \nabla_t$ Lazy

$\nabla R(y_{t+1}) = \nabla R(x_t) - \eta \nabla_t$ Agile

Project according to $B_R$:

$x_{t+1} = \underset{x \in K}{\arg\min}\ B_R(x \| y_{t+1})$

end for

- Both two version have regret bound guarantees similar to FTRL
- For instance,

Theorem 9.4 ( equivalence between FTRL and lazy OMD)

Let $f_1, \dots, f_T$ be linear cost functions. The lazy OMD and FTRL are identical

i.e.,

$$x_t = \underset{x \in K}{\arg\min}\; B_R(x \,\|\, y_t) = \underbrace{\underset{x \in K}{\arg\min} \left( \eta \sum_{s=1}^{t-1} \nabla_s^T \cdot x + R(x) \right)}_{(0)}$$

proof: The optimal solution

$x^*$ of the unconstrained optimization (0) satisfies

$$\nabla R(x_t^*) = -\eta \sum_{s=1}^{t-1} \nabla_s$$

By the lazy OMD update rule:

$$\nabla R(y_t) = -\eta \sum_{s=1}^{t-1} \nabla_s$$

$$\Rightarrow \quad \nabla R(x_t^*) = \nabla R(y_t)$$

Since $R$ is strictly convex, $x_t^* = y_t$

Hence, 
$$\begin{aligned} B_R(x \,\|\, y_t) &= R(x) - R(y_t) - (\nabla R(y_t))^T (x - y_t) \\ &= R(x) - R(y_t) + \eta \sum_{s=1}^{t-1} \nabla_s^T \cdot (x - y_t) \end{aligned}$$

independent of $x$

$\Rightarrow$ It's equivalent to

minimize $R(x) + \eta \sum_{s=1}^{t-1} \nabla_s^T \cdot x$ over $K$ #

- Now, we apply the general regret bound in Theorem 9.3 to concrete examples of $R(x)$.

Case I: $R(x) = x \log x$

$$\Rightarrow \nabla R(x) = 1 + \log x$$

$$K = \{ x \in \mathbb{R}_+^n : \sum x_i = 1 \} \Rightarrow x_{t+1}(i) = \frac{x_t(i) \cdot e^{-\eta \nabla_t(i)}}{\sum_{j=1}^{n} x_t(j) \cdot e^{-\eta \nabla_t(j)}}$$

If costs are in $[-1, 1]$,

$\|\nabla_t\|_t^* \leq \|\nabla_t\|_\infty \leq 1 =: G_R$

The diameter satisfies $D_R^2 \leq \log n$

$\Rightarrow \quad ER(MWA) \leq 2 D_R G_R \sqrt{2T} \leq 2\sqrt{2T\log n}$.

Case II $\quad R(x) = \frac{1}{2}\|x - x_0\|_2^2$.

arbitrary $x_0 \in K$

$\nabla R(x) = x - x_0$

$\Rightarrow \quad x_t = \mathrm{Proj}_K(y_t), \quad y_t = y_{t-1} - \eta \nabla_{t-1} \quad$ (lazy)

$x_t = \mathrm{Proj}_K(y_t), \quad y_t = x_{t-1} - \eta \nabla_{t-1} \quad$ (agile)

exactly online GD.

$$\Rightarrow \quad ER(OGD) \leq \frac{1}{\eta} D_R^2 + 2\eta \sum_{t=1}^{T} (\|\nabla_t\|_t^*)^2$$

$$\leq \frac{1}{2\eta} D^2 + 2\eta \sum_{t=1}^{T} \|\nabla_t\|^2$$

HW ($\|\cdot\|_t$ reduces to $\|\cdot\|_2$)

$$\leq 2GD\sqrt{T}$$

$\max_{x,y \in K} \|x - y\|$ Euclidean diameter

$\|\nabla f_t(x)\| \leq G, \ \forall t, \ x \in K$

- What about nonconvex costs?

Read the book by Elad Hazan for more detailed discussions.