Lecture 10: FTRL and Adversarial MAB

Week 10

Lecturer: Tianyu Wang

1 Online Convex Optimization and Follow-The-Regularized-Leader

Online convex optimization setting. For $t = 1, 2, \cdots, T$:

- The environment picks a convex loss function f_t;
- The player picks $x_t \in \Delta$ (Δ is a convex domain) and suffers loss $f_t(x_t)$; The function f_t is revealed to the player only after the player has made her decision.

Motivation: Regression problems with streaming in data.

Algorithm 1 Follow-The-Regularized-Leader (FTRL)

Input: learning rate η. Convex domain: Δ. Regularizer: Φ.
Initialization: Pick $x_1 \in \arg\min_{x \in \Delta} \Phi(x)$.
for $t = 1, 2, \cdots, T-1$ **do**
    Play x_t and observe f_t.
    Compute $\nabla f_t(x_t)$.
    Pick $x_{t+1} \in \arg\min_{x\in\Delta} \left\{ \eta \sum_{s=1}^{t} \langle \nabla f_t(x_t), x \rangle + \Phi(x) \right\}$.
end for

Regret against $z \in \Delta$ is defined as

$$Reg_T(z) = \sum_{t=1}^{T} f_t(x_t) - f_t(z).$$

1.1 Bregman divergence

Definition 1.1 (Bregman divergence)**.** Let Φ be a strongly convex function. The Bregman divergence with respect to Φ is

$$B_\Phi(x|y) = \Phi(x) - \Phi(y) - \nabla\Phi(y)^\top (x - y).$$

Note that by convexity, we have

$$\Phi(x) \geq \Phi(y) + \nabla\Phi(y)^\top (x - y).$$

Thus the Bregman divergence is always non-negative.

By Taylor theorem, there exists some z such that $z = \alpha x + (1-\alpha)y$ and

$$B_\Phi(x|y) = \frac{1}{2}(x-y)^\top \nabla^2\Phi(z)(x-y).$$

1.2 Dual norm

Any positive definite matrix M defines a norm: $\|x\|_M := \sqrt{x^\top M x}$. For any norm $\|\|$, its dual norm is defined as

$$\|x\|^* = \sup_{y:\|y\|\leq 1} x^\top y.$$

Example. The Euclidean norm is the dual norm of itself.

Since Φ is strongly convex, $\nabla^2\Phi(z)$ is positive definite at any z. Thus $\nabla^2\Phi(z)$ defines a norm $\|\cdot\|_{\nabla^2\Phi(z)}$ and we can define its dual norm $\|\cdot\|^*_{\nabla^2\Phi(z)}$.

Proposition 1.2 (Holder's inequality). *For any $x, y \in \mathbb{R}^d$ and a norm $\|\cdot\|$ on $\mathbb{R}^d$, it holds that*

$$x^\top y \leq \|x\|\|y\|^*,$$

where $\|\cdot\|^*$ is the dual norm of $\|\cdot\|$.

Proof. When $x = 0$, the inequality is trivially true. When $x \neq 0$, let $u = \frac{x}{\|x\|}$ and we have

$$x^\top y \leq \|x\| u^\top y \leq \|x\| \sup_{u,\|u\|\leq 1} u^\top y \leq \|x\|\|y\|^*.$$

□

Proposition 1.3. *Let M be a positive definite (symmetric) matrix. Then it holds that*

$$\|x\|_{M^{-1}} = \|x\|^*_M.$$

1.3 Regret Analysis

The regret of Algorithm 1 can be bounded by the following Theorem.

Theorem 1.4. *Algorithm 1 attains for every $u \in \Delta$ the following bound on the regret:*

$$Reg_T(u) \leq 2\eta \sum_{t=1}^{T} \left(\|\nabla f_t(x_t)\|^*_{\nabla^2\Phi(z_t)}\right)^2 + \frac{\Phi(u) - \Phi(x_1)}{\eta},$$

where $z_t = \alpha_t x_t + (1-\alpha_t)x_{t+1}$ for some α_t.

Lemma 1.5. *For any $u \in \Delta$, it holds that*

$$\frac{\Phi(x_1)}{\eta} + \sum_{t=1}^{T-1} \nabla f_t(x_t)^\top x_{t+1} \leq \frac{\Phi(u)}{\eta} + \sum_{t=1}^{T-1} \nabla f_t(x_t)^\top u.$$

Proof. Note that $\Phi(x_1) \leq \Phi(u)$ by algorithm initialization, which implies that the statement is true when $T = 1$.

Suppose the statement is true for $T-1$. Then since

$$x_{T+1} \in \arg\min_{x \in \Delta} \left\{ \frac{\Phi(x)}{\eta} + \sum_{t=1}^{T} \nabla f_t(x_t)^\top x \right\},$$

we have

$$\begin{aligned}
\frac{\Phi(u)}{\eta} + \sum_{t=1}^{T} \nabla f_t(x_t)^\top u &\geq \frac{\Phi(x_{T+1})}{\eta} + \sum_{t=1}^{T} \nabla f_t(x_t)^\top x_{T+1} \\
&= \frac{\Phi(x_{T+1})}{\eta} + \sum_{t=1}^{T-1} \nabla f_t(x_t)^\top x_{T+1} + \nabla f_T(x_T)^\top x_{T+1} \\
&\geq \frac{\Phi(x_1)}{\eta} + \sum_{t=1}^{T-1} \nabla f_t(x_t)^\top x_{t+1} + \nabla f_T(x_T)^\top x_{T+1} \\
&= \frac{\Phi(x_1)}{\eta} + \sum_{t=1}^{T} \nabla f_t(x_t)^\top x_{t+1}
\end{aligned}$$

where second last line uses induction hypothesis.

□

Lemma 1.6. *The regret for Algorithm 1 satisfies*

$$Reg_T(u) \leq \sum_{t=1}^{T} \nabla f_t(x_t)^\top (x_t - x_{t+1}) + \frac{D_\Phi^2}{\eta}, \qquad \forall u \in \Delta,$$

where

$$D_\Phi = \sqrt{\sup_{x,y \in \Delta} \{\Phi(x) - \Phi(y)\}}$$

is the diameter of Δ with respect to the regularizer Φ.

Proof. By convexity of f_t, we have that

$$\sum_{t=1}^{T} f_t(x_t) - f_t(u) \leq \sum_{t=1}^{T} \nabla f_t(x_t)^\top (x_t - u)$$

By Lemma 1.5, we have that

$$\begin{aligned}
\sum_{t=1}^{T} \nabla f_t(x_t)^\top (x_t - u) &\le \sum_{t=1}^{T} \nabla f_t(x_t)^\top (x_t - x_{t+1}) + \frac{1}{\eta} (\Phi(u) - \Phi(x_1)) \\
&\le \sum_{t=1}^{T} \nabla f_t(x_t)^\top (x_t - x_{t+1}) + \frac{D_\Phi}{\eta}.
\end{aligned}$$

□

Proof of Theorem 1.4. Recall that $\Phi(x)$ is a strongly convex function and Δ is a convex set. Define:

$$\Psi_t(x) := \eta \sum_{s=1}^{t} \nabla f_s(x_s)^\top x + \Phi(x)$$

We have

$$\begin{aligned}
\Psi_t(x_t) &= \Psi_t(x_{t+1}) + \nabla \Psi_t(x_{t+1})^\top (x_t - x_{t+1}) + B_{\Psi_t}(x_t | x_{t+1}) \\
&\qquad \text{(by definition of Bregman divergence)} \\
&\ge \Psi_t(x_{t+1}) + B_{\Psi_t}(x_t | x_{t+1}) \\
&\qquad (\nabla \Psi_t(x_{t+1})^\top (x_t - x_{t+1}) \ge 0 \text{ since } x_{t+1} \text{ minimizes } \Psi_t \text{ over } \Delta) \\
&= \Psi_t(x_{t+1}) + B_\Phi(x_t | x_{t+1}). \quad \text{(the linear terms do not contribute to Bregman divergence)}
\end{aligned}$$

Then it holds that

$$\begin{aligned}
B_\Phi(x_t | x_{t+1}) &\le \Psi_t(x_t) - \Psi_t(x_{t+1}) \\
&= \Psi_{t-1}(x_t) - \Psi_{t-1}(x_{t+1}) + \eta \left(\nabla f_t(x_t)^\top (x_t - x_{t+1}) \right) \\
&\overset{(i)}{\le} \eta \left(\nabla f_t(x_t)^\top (x_t - x_{t+1}) \right). \qquad \text{(since } \Psi_{t-1}(x_t) \le \Psi_{t-1}(x_{t+1}))
\end{aligned}$$

Thus we gives

$$\begin{aligned}
\nabla f_t(x_t)^\top (x_t - x_{t+1}) &\le \| \nabla f_t(x_t) \|^*_{\nabla^2 \Phi(z_t)} \| x_t - x_{t+1} \|_{\nabla^2 \Phi(z_t)} \\
&= \| \nabla f_t(x_t) \|^*_{\nabla^2 \Phi(z_t)} \sqrt{2 B_\Phi(x_t | x_{t+1})} \\
&\le \| \nabla f_t(x_t) \|^*_{\nabla^2 \Phi(z_t)} \sqrt{2 \eta \nabla f_t(x_t)^\top (x_t - x_{t+1})}. \qquad \text{(by } (i))
\end{aligned}$$

This implies

$$\nabla f_t(x_t)^\top (x_t - x_{t+1}) \le 2\eta \left(\| \nabla f_t(x_t) \|^*_{\nabla^2 \Phi(z_t)} \right)^2.$$

Combine this with Lemma 1.6 finishes the proof. □

2 Adversarial Multi-Armed Bandits

Adversarial multi-armed bandit setting. There are K arms, and the player pulls the arms sequentially.

For $t = 1, 2, \cdots, T$:

- The environment picks a loss function l_t defined on the arms $[K]$, such that $l_{t,j} \in [0, 1]$ for all t, i;
- The player pulls arm $I_t \in [K]$ and suffers loss l_{t,I_t}; The loss of arm I_t is revealed to the player only after the player has made her decision. No information about other arms are revealed.

Algorithm 2 EXP3

1: **Input**: learning rate η.
2: **Initialization**: Let p_1 be the uniform distribution over $[K]$.
3: **for** $t = 1, 2, \cdots, T - 1$ **do**
4: Play $I_t \sim p_t$ and observe l_{t,I_t}.
5: Construct estimators for $l_{t,i}$:

$$\widehat{l}_{t,i} = \frac{l_{t,I_t} \mathbb{I}_{[I_t = i]}}{p_{t,i}}.$$

6: Pick the probability p_{t+1} such that $p_{t+1,i} = \frac{p_{t,i} \exp\left(-\eta \widehat{l}_{t,i}\right)}{\sum_{j=1}^{K} p_{t,j} \exp\left(-\eta \widehat{l}_{t,j}\right)}$.
7: **end for**

First of all, we have

$$\mathbb{E}\left[\widehat{l}_{t,i}\right] = l_{t,i}, \quad \forall t, i.$$

We will show that the EXP3 algorithm is a special case of FTRL, where Δ is the probability simplex and Φ is the negative entropy $\Phi(p) = \sum_{i=1}^{K} p_i \log p_i$.

Using the FTRL framework

- Define $\Delta := \{x \in \mathbb{R}^K : x_i \geq 0, \sum_{i=1}^{K} x_i = 1\}$, which is convex.
- Define $\Phi(x) = \sum_{i=1}^{T} x_i \log x_i$.

Let q^* be distribution with 1 on the hindsight optimal arm, and 0 on all other entries.
Note that, by Taylor's theorem,

$$\Phi(z_t) = B_\Phi(x_t | x_{t+1})$$

Let's pretend that the loss vectors chosen by the adversary are $\widehat{l_t}$. By Theorem 1.4, we have

$$\sum_{t=1}^{T} \widehat{l_t}^\top p_t - \sum_{t=1}^{T} \widehat{l_t}^\top q^* \leq 2\eta \sum_{t=1}^{T} \left(\|\widehat{l_t}\|^*_{\nabla^2\Phi(z_t)} \right)^2 + \frac{\Phi(q^*) - \Phi(p_1)}{\eta} \tag{1}$$

Note that $\nabla^2\Phi(z)$ is a diagonal matrix with $\frac{1}{z_i}$ on its diagonal and 0 on the off-diagonal entries.

$$\begin{aligned}
\left(\|\widehat{l_t}\|^*_{\nabla^2\Phi(z_t)} \right)^2 &= \left\| \widehat{l_t} \right\|^2_{(\nabla^2\Phi(z_t))^{-1}} \\
&= \widehat{l_t}^\top (\nabla^2\Phi(z_t))^{-1} \widehat{l_t} \\
&= \sum_{i=1}^{K} \frac{l_{t,i}^2 \mathbb{I}_{[I_t=i]}}{p_{t,i}^2} z_{t,i} \\
&= \sum_{i=1}^{K} \frac{l_{t,i}^2 \mathbb{I}_{[I_t=i]}}{p_{t,i}^2} \left(\alpha_t p_{t,i} + (1-\alpha_t) p_{t+1,i} \right),
\end{aligned} \tag{2}$$

for some $\alpha_t \in [0, 1]$.

For any i,

$$\begin{aligned}
\frac{p_{t+1,i}}{p_{t,i}} &= \frac{\exp(-\eta \widehat{l}_{t,i})}{\sum_{j=1}^{K} p_{t,j} \exp(-\eta \widehat{l}_{t,j})} \\
&= \frac{\exp(-\eta l_{t,i} \mathbb{I}_{[I_t=i]}/p_{t,i})}{\sum_{j=1}^{K} p_{t,j} \exp(-\eta l_{t,j} \mathbb{I}_{[I_t=j]}/p_{t,j})} \\
&\leq \begin{cases} \exp(-\eta l_{t,i}/p_{t,i}), & \text{if } I_t = i \\ \frac{1}{p_{t,k} \exp\left(-\eta l_{t,k}/p_{t,k}\right) + (1-p_{t,k})} & \text{if } I_t = k \text{ for some } k \neq i \end{cases} \\
&\leq \exp(\eta).
\end{aligned}$$

Plugging $\frac{p_{t+1,i}}{p_{t,i}} \leq \exp(\eta)$ into (2) gives

$$\begin{aligned}
\left(\|\widehat{l_t}\|^*_{\nabla^2(z_t)} \right)^2 &\leq \sum_{i=1}^{K} \frac{l_{t,i}^2 \mathbb{I}_{[I_t=i]}}{p_{t,i}} \left(\alpha_t + (1-\alpha_t) \exp(\eta) \right) \\
&\leq \exp(\eta) \sum_{i=1}^{K} \frac{l_{t,i}^2 \mathbb{I}_{[I_t=i]}}{p_{t,i}}.
\end{aligned} \tag{3}$$

Plugging the above result into (1), and taking expectation gives

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\widehat{l}_t^\top p_t - \sum_{t=1}^{T}\widehat{l}_t^\top q^*\right] &\le \mathbb{E}\left[2\eta\sum_{t=1}^{T}\left(\|\widehat{l}_t\|^*_{\nabla^2\Phi(z_t)}\right)^2 + \frac{\Phi(q^*)-\Phi(p_1)}{\eta}\right] \\
&\le \mathbb{E}\left[2\eta\sum_{t=1}^{T}\left(\|\widehat{l}_t\|^*_{\nabla^2\Phi(z_t)}\right)^2\right] + \frac{\Phi(p_1)}{\eta} && \text{(since } \Phi(u)\ge 0) \\
&= \mathbb{E}\left[2\eta\sum_{t=1}^{T}\left(\|\widehat{l}_t\|^*_{\nabla^2\Phi(z_t)}\right)^2\right] + \frac{\log K}{\eta} \\
&\le \mathbb{E}\left[2\eta\exp(\eta)\sum_{t=1}^{T}\sum_{i=1}^{K}\frac{l_{t,i}^2\mathbb{I}_{[I_t=i]}}{p_{t,i}}\right] + \frac{\log K}{\eta} && \text{(by Eq. 3)} \\
&= 2\eta\exp(\eta)\sum_{t=1}^{T}\sum_{i=1}^{K} l_{t,i}^2 + \frac{\log K}{\eta}. && \left(\text{since } \mathbb{E}\left[\tfrac{\mathbb{I}_{[I_t=i]}}{p_{t,i}}\right]=1\right) \\
&\le 2\eta\exp(\eta)TK + \frac{\log K}{\eta} && \text{(since } l_{t,i}\in[0,1].)
\end{aligned}
$$

By law of total expectation, we have $\mathbb{E}\left[\sum_{t=1}^{T}\widehat{l}_t^\top p_t - \sum_{t=1}^{T}\widehat{l}_t^\top q^*\right] = \sum_{t=1}^{T} l_t^\top p_t - \sum_{t=1}^{T} l_t^\top q^*$ which is the regret. For a derivation with more rigor, one needs to use the probability theory arguments discussed in previous lecture.

Acknowledgement

Reference: online convex optimization book by Elad Hazan.