Experts Problem and Online Convex Optimization

Chen-Yu Wei

References

- Lecture 1 and 2 of Haipeng's course

Learning from Experts (The Experts Problem)

Suppose there are N experts (actions)

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

Learner generates a distribution $p_t \in \Delta_N$

Adversary decides the loss of every expert $\ell_t = \left(\ell_t(1), \ldots, \ell_t(N)\right) \in \mathbb{R}^N$

Adversary reveals ℓ_t to the learner

Learner suffers loss $\langle p_t, \ell_t \rangle = \mathbb{E}_{i \sim p_t}\left[\ell_t(i)\right]$

A special case of Online Convex Optimization with

$\mathcal{X} = \Delta_N$ and $f_t(p) = \langle p, \ell_t \rangle$

$$\textbf{Regret} := \sum_{t=1}^{T} \langle p_t, \ell_t \rangle - \min_{i \in [N]} \sum_{t=1}^{T} \ell_t(i)$$

Follow the Leader (FTL)

Suppose there are N experts (actions)

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

Learner generates a distribution $p_t \in \Delta_N$

Adversary decides the loss of every expert $\ell_t = \left(\ell_t(1), \ldots, \ell_t(N)\right) \in \mathbb{R}^N$

Adversary reveals ℓ_t to the learner

Learner suffers loss $\langle p_t, \ell_t \rangle$

Define $\quad L_t(i) \coloneqq \sum_{s=1}^{t} \ell_s(i)$

"Follow the leader" strategy: $\quad i_t = \underset{i \in [N]}{\operatorname{argmin}} L_{t-1}(i) \quad$ and let $p_t = \mathrm{e}_{i_t}$

Follow the Leader (FTL)

Time	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	…
$\ell_t(1)$	0.5	0	1	0	1	0	…
$\ell_t(2)$	0	1	0	1	0	1	…

$$\textbf{Regret} = \sum_{t=1}^{T} \langle p_t, \ell_t \rangle - \min_{i \in [N]} \sum_{t=1}^{T} \ell_t(i)$$

$\geq T - 1$ $\quad \approx \frac{T}{2}$

$\approx (T-1) - \frac{T}{2} \approx \Theta\left(\frac{T}{2}\right) \neq o(T)$

Softmin

Replacing “min” by “softmin”:

$$p_t(i) = \frac{w_t(i)}{\sum_{j=1}^{N} w_t(j)} \quad \text{where} \quad w_t(i) = e^{-\eta L_{t-1}(i)} \quad \textbf{(Exponential weights)}$$

“Follow-the-leader” (hardmin) is a special case when $\eta = \infty$

Regret Guarantee for Exponential Weights

Theorem 1.

Assume that $\eta \ell_t(i) \geq -1$ for all t, i. Then Exponential Weight ensures

$$\text{Regret} \leq \frac{\log N}{\eta} + \eta \sum_{t=1}^{T} \sum_{i=1}^{N} p_t(i) \ell_t(i)^2$$

Assume $0 \leq \ell_t(i) \leq 1$

$$\text{Regret} \leq \frac{\log N}{\eta} + \eta T = 2\sqrt{T \log N} \leq o(T)$$

$$\eta = \sqrt{\frac{\log N}{T}}$$

Freund and Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, 1997.

Proof of Theorem 1

log = ln

Recall $L_t(i) = \sum_{s=1}^{t} \ell_s(i)$

and $p_t(i) = \frac{w_t(i)}{W_t}$ where $w_t(i) = \exp\left(-\eta L_{t-1}(i)\right)$, $W_t = \sum_{j=1}^{N} w_t(j)$

$$\ln \frac{W_{t+1}}{W_t} = \ln \frac{\sum_{i=1}^{N} w_{t+1}(i)}{W_t} = \ln \frac{\sum_{i=1}^{N} \exp\left(-\eta L_t(i)\right)}{W_t}$$

$w_t(i)$

$$= \ln \frac{\sum_{i=1}^{N} \exp\left(-\eta L_{t-1}(i)\right) \exp\left(-\eta \ell_t(i)\right)}{W_t} \qquad \left(L_t(i) = L_{t-1}(i) + \ell_t(i)\right)$$

$$= \ln \frac{\sum_{i=1}^{N} w_t(i) \exp\left(-\eta \ell_t(i)\right)}{W_t} = \ln \left(\sum_{i=1}^{N} p_t(i) \exp\left(-\eta \ell_t(i)\right) \right)$$

$$\ln \frac{W_{t+1}}{W_t} = \ln \left(\sum_{i=1}^{N} P_t(i) \exp\left(-\eta \ell_t(i)\right) \right)$$

$$\leq \ln \left(\sum_{i=1}^{N} P_t(i) \left(1 - \eta \ell_t(i) + \eta^2 \ell_t(i)^2\right) \right)$$

$$\leq \ln \left(\sum_{i=1}^{N} P_t(i) - \eta \sum_{i=1}^{N} P_t(i)\ell_t(i) + \eta^2 \sum_{i=1}^{N} P_t(i)\ell_t(i)^2 \right)$$

$$= \ln \left(1 - \eta \langle P_t, \ell_t \rangle + \eta^2 \sum_{i=1}^{N} P_t(i)\ell_t(i)^2 \right)$$

$$\leq -\eta \langle P_t, \ell_t \rangle + \eta^2 \sum_{i=1}^{N} P_t(i)\ell_t(i)^2$$

For real number $z \leq 1$

$$\exp(z) \leq 1 + z + z^2$$

$z = -\eta \ell_t(i)$

$$\ln(1+z) \leq z$$

$\sum_{t=1}^{T}$

Summing over $t = 1, 2, \cdots, T$

$$\ln \frac{W_{T+1}}{W_1} \leq -\eta \sum_{t=1}^{T} \langle P_t, \ell_t \rangle + \eta^2 \sum_{t=1}^{T} \sum_{i=1}^{N} P_t(i)\ell_t(i)^2$$

$$\ln \frac{W_{T+1}}{W_1} \leq -\eta \sum_{t=1}^{T} \langle p_t, \ell_t \rangle + \eta^2 \sum_{t=1}^{T} \sum_{i=1}^{N} p_t(i)\, \ell_t(i)^2$$

$i^* = \arg\min_i L_T(i)$

Recall $W_t = \sum_{i=1}^{N} \exp\left(-\eta L_{t-1}(i)\right)$

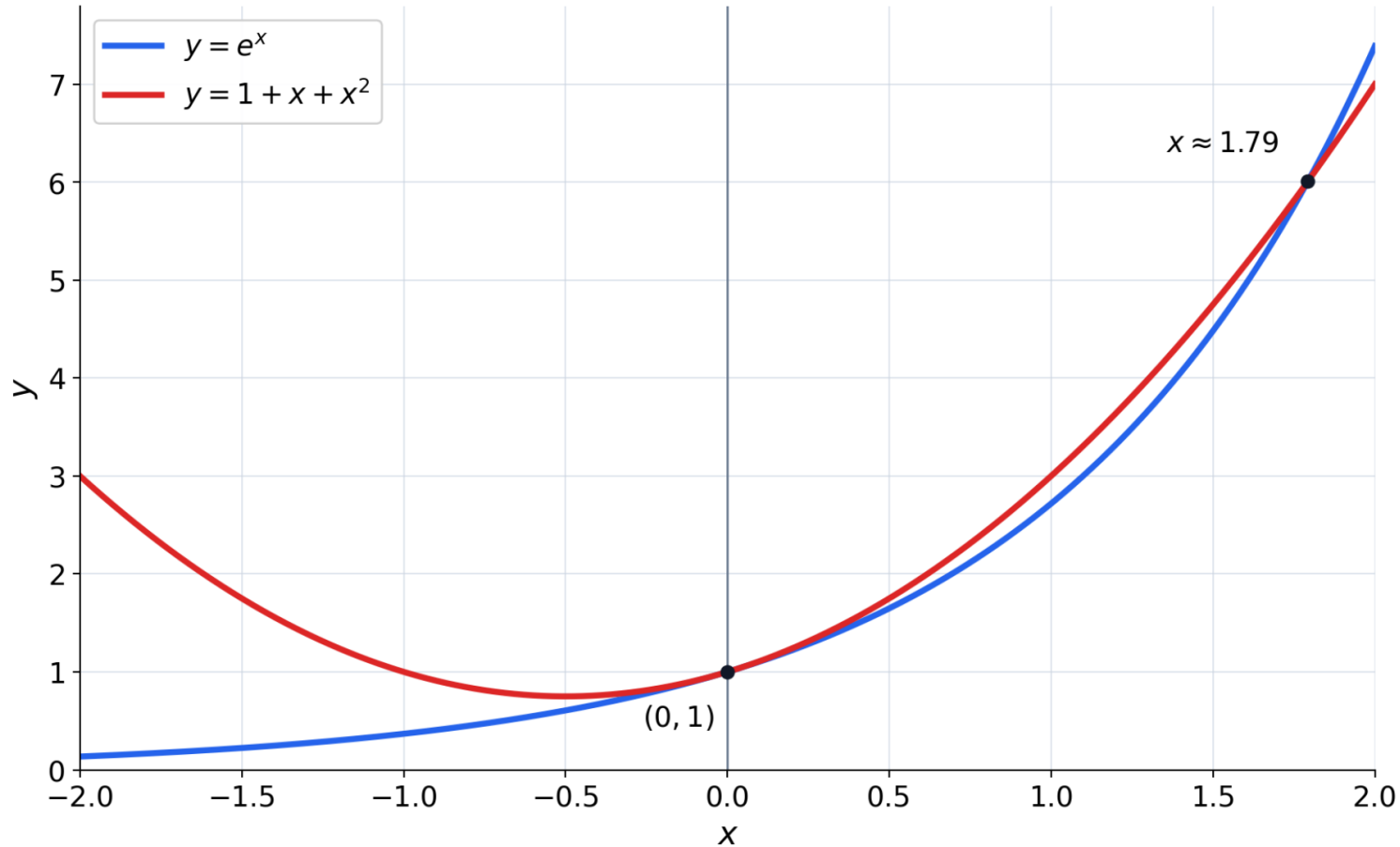
$$\begin{cases} W_1 = \sum_{i=1}^{N} \exp(0) = N \\ W_{T+1} = \sum_{i=1}^{N} \exp\left(-\eta L_T(i)\right) \geq \exp\left(-\eta L_T(i^*)\right) \end{cases}$$

$L_0(i) = 0$

$$\Rightarrow \ln \frac{W_{T+1}}{W_1} = \ln W_{T+1} - \ln W_1 \geq -\eta L_T(i^*) - \ln N$$

$$\Rightarrow -\eta L_T(i^*) - \ln N \leq -\eta \sum_{t=1}^{T} \langle p_t, \ell_t \rangle + \eta^2 \sum_{t=1}^{T} \sum_{i=1}^{N} p_t(i)\, \ell_t(i)^2$$

$$\Rightarrow \sum_{t=1}^{T} \langle p_t, \ell_t \rangle - L_T(i^*) \leq \frac{\ln N}{\eta} + \eta \sum_{t=1}^{T} \sum_{i=1}^{N} p_t(i)\, \ell_t(i)^2$$


$y = e^x$
$y = 1 + x + x^2$
$x \approx 1.79$
(0, 1)
7
6
5
4
3
2
1
0
y
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
x

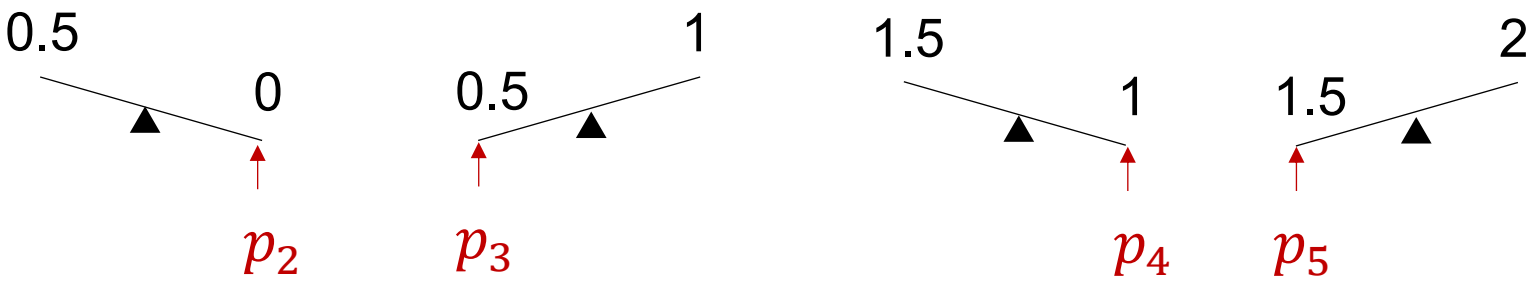
Why softmin works, while min does not?

(Discussion)

The Failure of Follow-The-Leader

Time	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	…
$\ell_t(1)$	0.5	0	1	0	1	0	…
$\ell_t(2)$	0	1	0	1	0	1	…

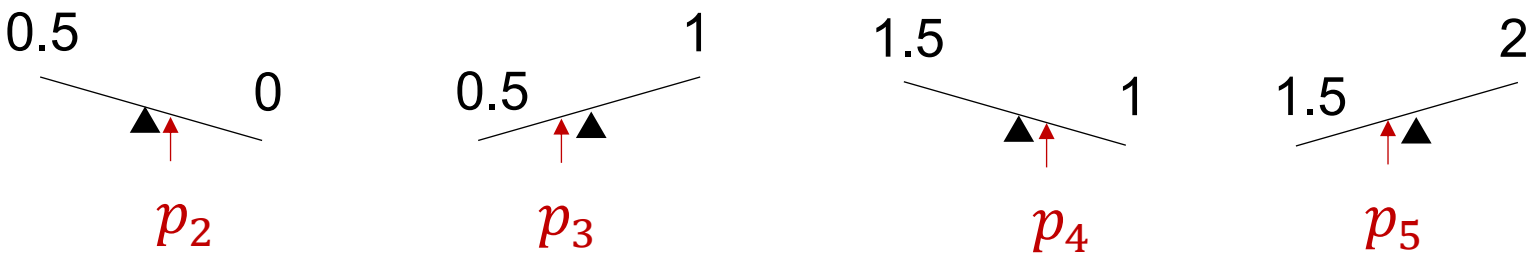
FTL: $i_t = \underset{i \in [N]}{\operatorname{argmin}} L_{t-1}(i)$



The Failure of Follow-The-Leader

Time	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	…
$\ell_t(1)$	0.5	0	1	0	1	0	…
$\ell_t(2)$	0	1	0	1	0	1	…

EW: $p_t(i) \propto e^{-\eta L_{t-1}(i)}$



Why Stability Helps?

(Discussion)

Be-The-Leader Lemma

Theorem 2. Follow the leader (FTL) strategy

$$x_t = \operatorname*{argmin}_{x \in \mathcal{X}} \sum_{s=1}^{t-1} f_s(x)$$

ensures

$$\text{Regret} \leq \sum_{t=1}^{T} (f_t(x_t) - f_t(x_{t+1}))$$

(This lemma doesn't need convexity of f_t or $\mathcal{X}$)

Proof of Theorem 2

Follow The Regularized Leader

(Generalizing the Exponential Weights algorithm)

An Equivalent Form of EW

$$\psi(p) \coloneqq \sum_{i=1}^{N} p(i) \ln p(i) \quad \text{(convex)}$$

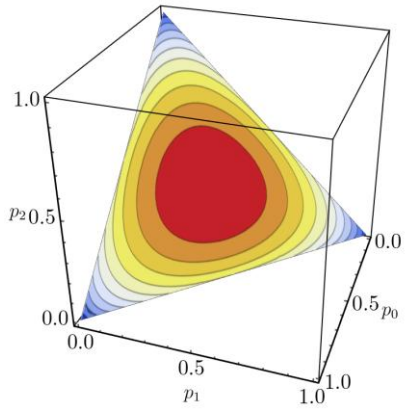
$$p_t(i) = \frac{e^{-\eta L_{t-1}(i)}}{\sum_{j=1}^{N} e^{-\eta L_{t-1}(j)}}$$

$\Leftrightarrow$

Verify yourself
(Lagrange multiplier)

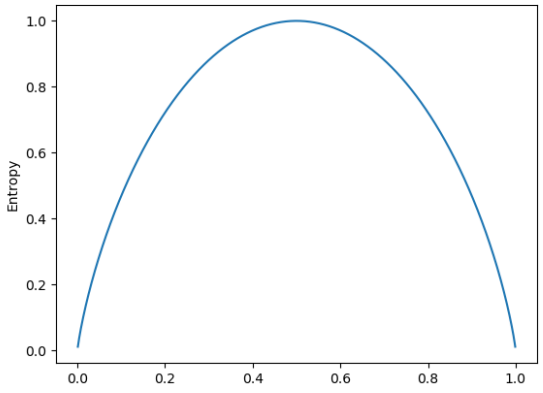
$$p_t = \underset{p \in \Delta_N}{\operatorname{argmin}} \left\{ \langle p, L_{t-1} \rangle + \frac{1}{\eta} \psi(p) \right\}$$

$$-\psi(p) = \sum_{i=1}^{N} p(i) \ln \frac{1}{p(i)}$$ is called the **(Shannon) entropy function** (concave)



(link)

N=3



(link)

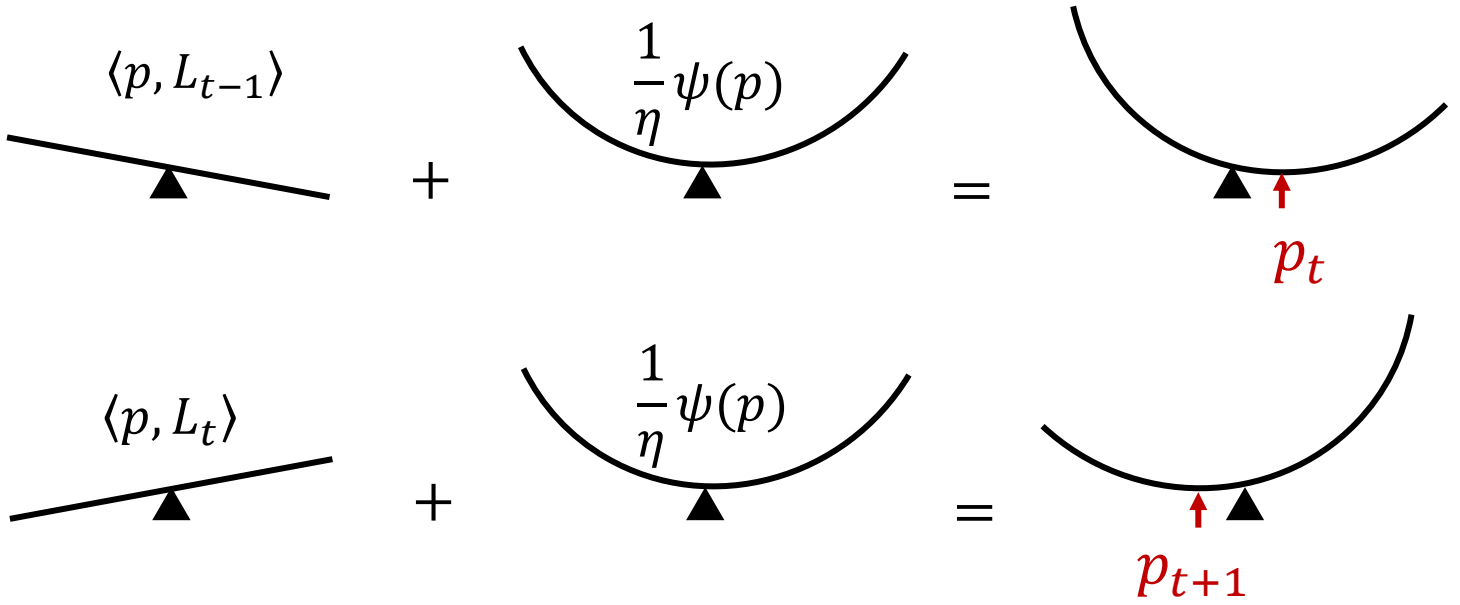
N=2

Follow the Regularized Leader (FTRL)

Regularizer

$$x_t = \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t-1} f_s(x) + \frac{1}{\eta}\psi(x) \right\}$$

If ψ is *curved*, it makes the update more stable:



(Strongly) Convex Functions

Bregman Divergence

FTRL for Online Convex Optimization (OCO)

Given: convex set $\mathcal{X} \subset \mathbb{R}^d$

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

Learner makes a decision $x_t \in \mathcal{X}$

Adversary decides a convex loss function $f_t: \mathcal{X} \to \mathbb{R}$

Adversary reveals f_t to the learner

Follow The Regularized Leader (FTRL)

$$x_t = \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t-1} f_s(x) + \frac{1}{\eta} \psi(x) \right\}$$

Regret Guarantee of FTRL in OLO

Theorem 3. Let ψ be μ-strongly-convex in norm $\|\cdot\|$.

Then FTRL in OCO ensures for any $x^\star \in \mathcal{X}$,

$$\sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x^\star) \leq \frac{\psi(x^\star) - \min_{u \in \mathcal{X}} \psi(u)}{\eta} + \frac{\eta}{2\mu} \sum_{t=1}^{T} \|\nabla f_t(x_t)\|_\star^2$$

$\|\cdot\|_\star$ is the dual norm of $\|\cdot\|$

Proof of Theorem 3 (Stability)

Proof of Theorem 3 (Stability)

Proof of Theorem 3

Special Case: Negative Shannon Entropy

$$\mathcal{X} = \Delta_N, \qquad \psi(x) = \sum_{i=1}^{N} x(i) \log x(i)$$

Special Case: Euclidean Norm

$\mathcal{X} = \text{general convex set}, \qquad \psi(x) = \frac{1}{2}\|x\|_2^2$

OCO with only gradient feedback

Given: convex set $\mathcal{X} \subset \mathbb{R}^d$

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

Learner makes a decision $x_t \in \mathcal{X}$

Adversary decides a convex loss function $f_t: \mathcal{X} \to \mathbb{R}$

Adversary reveals $\nabla f_t(x_t)$ to the learner

Online Mirror Descent

Online Mirror Descent

Implicit Online Mirror Descent

$$x_{t+1} = \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ f_t(x) + \frac{1}{\eta} D_\psi(x, x_t) \right\}$$

Online Mirror Descent

$$x_{t+1} = \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \langle x, \nabla f_t(x_t) \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\}$$

The regret guarantee and analysis are closely related to those of FTRL (will be studied in Homework 1)

Special Case: Negative Shannon Entropy

$$\mathcal{X} = \Delta_N, \qquad \psi(x) = \sum_{i=1}^{N} x(i) \log x(i)$$

Special Case: Euclidean Norm

$\mathcal{X} = \text{general convex set}, \qquad \psi(x) = \frac{1}{2}\|x\|_2^2$

Follow The Perturbed Leader

Online Linear Optimization with a Weaker Adversary

Given: convex set $\mathcal{X} \subset \mathbb{R}^d$

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

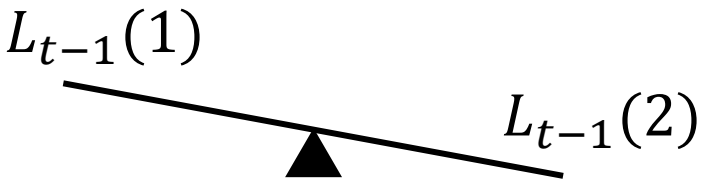
Adversary decides a **linear** loss function $f_t(x) = \langle x, \ell_t \rangle$
(not revealing it to the learner at this moment)

Learner makes a decision $x_t \in \mathcal{X}$

Adversary reveals ℓ_t to the learner

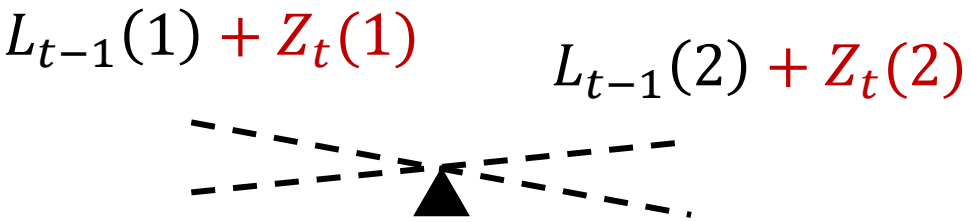
order swapped

Another Way to Avoid Exploitation by Adversary



Follow the Leader: $i_t = \underset{i \in [N]}{\operatorname{argmin}} L_{t-1}(i)$

As the learner will choose $i_t = 2$,
I'll make $\ell_t(2) \gg \ell_t(1)$ …



$Z_t(i) \sim \mathcal{N}(0, \sigma^2) \quad \forall i = 1,2$

Follow the Perturbed Leader:

$$i_t = \underset{i \in [N]}{\operatorname{argmin}} L_{t-1}(i) + Z_t(i)$$

Hard to predict the learner's choice…

FTPL is also ensuring “stability”

FTPL for OLO

Follow The Perturbed Leader (FTPL) for OLO

$$x_t = \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \sum_{s=1}^{t-1} \langle x, \ell_s \rangle + \langle x, z_t \rangle \right\}$$

where $z_t \in \mathbb{R}^d$ is drawn from some distribution $\mathcal{D}$.

$\mathcal{D}$ needs not be Gaussian, needs not be zero mean, and needs not be the same in every round.

When is FTPL more preferred than FTRL

For many common decision set $\mathcal{X}$, it is (computationally) much easier to solve

$$\operatorname*{argmin}_{x \in \mathcal{X}} \langle x, L \rangle$$

then

$$\operatorname*{argmin}_{x \in \mathcal{X}} \langle x, L \rangle + \psi(x)$$

for a given $L \in \mathbb{R}^d$.

When is FTPL more preferred than FTRL

Example: **Online Shortest Path**

Given: a graph, a source, and a sink

Let $\mathcal{Q} \subset \{0,1\}^{|E|}$ be the set of paths from the source to sink

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

Adversary decides the costs on all edges $\ell_t \in \mathbb{R}^{|E|}$

Learner chooses a path $q_t \in \mathcal{Q}$

Adversary reveals ℓ_t to the learner

$\mathcal{Q}$ above is not a convex set, but we can

- Run OCO in $\mathcal{X} = \text{convhull}(\mathcal{Q})$ (becoming a set of flows)
- Sample q_t such that $\mathbb{E}[q_t] = x_t$ in round t.