Lecture 03:
Online Mirror Descent and Follow-the-Perturbed-Leader – No-Regret via Penalty and Perturbation

Lecturer: John Lazarsfeld

September 23, 2025

Abstract

Online Mirror Descent (OMD) and its analysis, equivalence between OMD and FTRL on linear losses, Follow-the-Perturbed-Leader (FTPL) and its analysis, relationship between FTPL and FTRL.

1 Online Mirror Descent

Recap from Lecture 02 and Overview. In the previous lecture, we introduced the Follow-the-Regularized-Leader family of online learning algorithms, and we proved a general regret bound that highlighted the importance of an algorithm's *stability* in this setting. The goal of this lecture is to introduce a second canonical family of algorithms called *Online Mirror Descent* (OMD), which achieves stability via *penalty function* (as opposed to regularization). We will motivate, introduce, and derive a similar general regret bound for this family, as well as discuss its connection with FTRL. Finally, we will introduce and analyze yet a third family of algorithms called Follow-the-Perturbed-Leader (FTPL), which achieves stability (in expectation) via random perturbations.

1.1 Motivating OMD: Stability via Penalty

Before introducing the general Online Mirror Descent family, we build some intuition using our previous analyses of Online Gradient Descent and the FTRL family. Suppose an instance of the general Online Convex Optimization setting with loss functions $\{f_t\}$. At time t, an online learning algorithm outputs an iterate $x_t \in \mathcal{X}$ and incurs a loss of $f_t(x_t)$. How should the algorithm choose its iterate x_{t+1}?

Exploitation via First-Order Approximation. As the learner is unaware of f_{t+1} when choosing x_{t+1}, the algorithm must balance the exploration/exploitation tradeoff (similarly as in OGD and FTRL). For this, under the hypothesis that f_{t+1} is similar to f_t, one natural strategy is to choose x_{t+1} to minimize the most recently observed loss function f_t. In particular, by convexity of f_t (and assuming differentiability), the learner can attempt to minimize the first order approximation of f_t around x_t, which is always a lower bound on f_t. Specifically, define $\tilde{f}_t(x) = f_t(x_t) + \langle \nabla f(x_t), x - x_t \rangle$. Then by the first-order convexity inequality, we have $f_t(x) \geq \tilde{f}_t(x)$ for all $x \in \mathcal{X}$.

Exploration via Penalty. On the other hand, as shown in the analysis of FTRL, in the online learning setting, the regret of an algorithm is related to the *stability* of its iterates. In particular, regret can scale with the *path-length* of the iterates $\sum_{t=1}^{T} \|x_t - x_{t+1}\|$, and thus an algorithm should additionally choose x_{t+1} not far from x_t. In FTRL, this property was enforced implicitly via the use of a (strongly-convex) regularizer. On the other hand, we could also imagine enforcing stability directly into an objective function via a penalty function like $\|x - x_{t+1}\|$.

Combining the Pieces. Combining these ideas (and in particular, using a squared Euclidean norm penalty term) leads to the following type of update rule at time $t+1$:

$$x_{t+1} := \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) + \frac{1}{2\eta} \|x - x_t\|_2^2 \right\} . \tag{1}$$

Here, similar to OGD and FTRL, the stepsize parameter $\eta > 0$ is used to balance the magnitudes exploration and exploitation terms.

Now substituting the definition of $\tilde{f}_t(x)$ into (1), observe that we can rewrite x_{t+1} as:

$$\begin{aligned} x_{t+1} &= \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ f_t(x_t) + \langle \nabla f_t(x_t), x - x_t \rangle + \frac{1}{2\eta} \|x_t - x\|_2^2 \right\} && (2) \\ &= \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ 2\eta \langle \nabla f_t(x_t), x - x_t \rangle + \|x_t - x\|_2^2 \right\} && (3) \\ &= \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \|\eta \nabla f_t(x_t)\|_2^2 + 2\eta \langle \nabla f_t(x_t), x - x_t \rangle + \|x_t - x\|_2^2 \right\} && (4) \\ &= \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \|x - (x_t - \eta \nabla f_t(x_t))\|_2^2 \right\} && (5) \\ &= \Pi_{\mathcal{X}} \Big(x_t - \eta \nabla f_t(x_t) \Big) . && (6) \end{aligned}$$

In other words, the strategy used in (1) exactly recovers the update rule for Online Gradient Descent from Lecture 01! In general, given the geometry of the decision space $\mathcal{X}$, we could also imagine using other types of norms as penalty functions. This is the precisely recipe for the Online Mirror Descent family. In particular, Online Mirror Descent replaces the ℓ_2^2 penalty $\|x - x_t\|_2^2$ with the *Bregman divergence* with respect to a function ψ. For this, we first recall some useful definitions from convex analysis:

Convex Analysis Refresher: Bregman Divergences. We recall definitions and properties of *Bregman divergences* (more details are given in (Orabona, 2019, Section 6.3)). Throughout, suppose $\mathcal{X} \subset \mathbb{R}$ is convex and compact, and $\psi : \mathcal{X} \to \mathbb{R}$ is strictly convex and differentiable.

Definition 1. *The Bregman Divergence with respect to ψ is the function $D_\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, where for $x, x' \in \mathcal{X}$:*

$$D_\psi(x, x') \;:=\; \psi(x) - \psi(x') - \langle \nabla \psi(x'), x - x' \rangle .$$

In words $D_\psi(x, x')$ is the difference between $\psi(x)$ and the first-order approximation of ψ at x'. The following properties additionally hold:

Lemma 2. *For $x, x' \in \mathcal{X}$: $D_\psi(x, x') \geq 0$. However, in general $D_\psi(x, x') \neq D_\psi(x', x)$.*

Lemma 3 (Relationship with Strong Convexity). *If ψ is α-strongly-convex with respect to $\|\cdot\|$ for some $\alpha > 0$, then for any $x, x' \in \mathcal{X}$:*

$$D_\psi(x, x') \geq \frac{\alpha}{2} \|x - x'\|^2 .$$

Lemma 4 (3-point Identity). *Let $x, y, z \in \mathcal{X}$. Then the following identity holds:*

$$\langle \nabla \psi(x) - \nabla \psi(y), z - x \rangle = D_\psi(z, y) - D_\psi(z, x) - D_\psi(x, y) .$$

We also have the following common examples of Bregman Divergences:

Example 5. Let $\psi(x) = \frac{1}{2}\|x\|_2^2$. Then $D_\psi(x, x') = \frac{1}{2}\|x - x'\|_2^2$.

Example 6. Let $\mathcal{X} = \Delta_n$, and let ψ be the negative entropy function $\psi(x) = \sum_{i=1}^n x_i \log x_i$. Then $D_\psi(x, x') = \sum_{i=1}^n x_i \log \frac{x_i}{x_i'}$, which is also known as the *KL-Divergence* $\mathrm{KL}(x, x')$.

1.2 OMD Algorithm and Regret Guarantee

We now formally state the Online Mirror Descent family of algorithms. In particular, similar to the FTRL family, each choice of ψ corresponds to a different instantiation of OMD.

Algorithm 1 Online Mirror Descent (OMD) for OCO Setting

Input: Initial $x_1 \in \mathcal{X}$; strictly convex $\psi : \mathcal{X} \to \mathbb{R}$. stepsize parameter $\eta > 0$.
for $t = 1, \dots, T$ **do**:
1. Play action $x_t \in \mathcal{X}$, and incur cost $f_t(x_t)$. Observe gradient feedback $\nabla f_t(x_t) \in \mathbb{R}^n$.
2. Update $x_{t+1} \in \mathcal{X}$ by

$$x_{t+1} := \left\{ \langle \nabla f_t(x_t), x \rangle + \frac{1}{\eta} D_\psi(x, x_t) \right\} . \tag{7}$$

end for

Remark 7. We make several remarks about the OMD algorithm:

- **Standard vs. Lazy OMD**: There is sometimes a distinction between standard OMD and a "lazy" variant. We focus here on the standard variant and defer more details on the lazy variant to Hazan et al. (2016, Section 5.3).
- **On gradient feedback model**: In the algorithm, we assume at each round that the learner observes feedback in the form of the gradients $\nabla f_t(x_t)$. Alternatively, we can assume the learner observes the full function f_t and is able to subsequently compute $\nabla f_t(x_t)$.

We now state the regret guarantee for OMD.

Theorem 8. *Let $\{x_t\}$ be the iterates of OMD (Algoritm 1) using $\psi : \mathcal{X} \to \mathbb{R}$ that is 1-strongly-convex with respect to $\|\cdot\|$ and stepsize $\eta > 0$ on any sequence of convex and differentiable losses $\{f_t\}$. Then for any $T \geq 1$, the following bounds hold:*

(1) Suppose for $B > 0$ that $x_1 \in \mathcal{X}$ satisfies $D_\psi(x, x_1) \leq B$ for all $x \in \mathcal{X}$. Then:

$$\mathrm{Reg}_{OMD}(T) \;\leq\; \frac{\eta}{2} \sum_{t=1}^{T} \|\nabla f_t(x_t)\|_\star^2 + \frac{B}{\eta} .$$

(2) Moreover, suppose for $L > 0$ that $\|\nabla f_t(x_t)\|_\star \leq L$ for all $t \geq 1$. Then setting $\eta := \sqrt{\frac{B}{2L^2T}}$:

$$\mathrm{Reg}_{OMD}(T) \;\leq\; L\sqrt{BT} . \tag{8}$$

Remark 9. We make several remarks about the regret bound for OMD in Theorem 8

- **Similarities with OGD and FTRL bounds**: At first glance, the bound in Part (1) appears more similar to the OGD regret bound in Lecture 01 than the FTRL regret bound from Lecture 02. In particular, observe that the bound in (1) does not directly depend on the stability or path length of the iterates $\sum_{t=1} \|x_t - x_{t+1}\|$. On the other hand, we assume that ψ is strongly-convex, and this property is used to establish the multiplicative factor of η in the first term (recall that strong-convexity was also needed to prove the η-stability of the iterates in FTRL). Thus intuitively, ψ can be viewed as the regularizer in the FTRL algorithm (this is made more precise in the next section).
- **On the constant** B: Notice that the constant B depends on the initialization $x_1 \in \mathcal{X}$. While this is a slight abuse of notation, the key point is that, depending on ψ, the Bregman divergence D_ψ can grow arbitrarily large, even if $\mathcal{X}$ is bounded.

1.3 Proof of OMD Regret Bound

Proof. We will prove Part (1) of the theorem. Part (2) is straightforward and identical to the proofs of OGD and FTRL. The proof of the Part (1) will be developed in several parts:

(i) **Setup using linearized losses**:
By the convexity of the loss functions, recall that it suffices to upper bound the regret $\mathrm{Reg}_{OMD}(T)$ with respect to $\{f_t\}$ by the regret using the *linearized losses* $\{\langle \nabla f_t(x_t), x\rangle\}$:

$$\mathrm{Reg}_{OMD}(T) \;=\; \min_{x\in\mathcal{X}}\left\{\sum_{t=1}^{T} f_t(x_t) - f_t(x)\right\} \;\leq\; \min_{x\in\mathcal{X}}\left\{\sum_{t=1}^{T}\langle \nabla f_t(x_t), x_t - x\rangle\right\}. \tag{9}$$

(ii) **Establish the "Mirror-Descent Lemma" at round** t:
Fix t. We will derive an upper bound on the quantity $\langle \nabla f_t(x_t), x_t - x^\star\rangle$, where we fix $x^\star \in \mathrm{argmin}_{x\in\mathcal{X}} \sum_{t=1}^{T}\langle \nabla f_t(x_t), x_t - x\rangle$. For this, observe first that we can write

$$\begin{aligned}\langle \nabla f_t(x_t), x_t - x^\star\rangle &= \langle \nabla f_t(x_t), x_{t+1} - x^\star\rangle + \langle \nabla f_t(x_t), x_t - x_{t+1}\rangle \qquad (10)\\ &\leq \langle \nabla f_t(x_t), x_{t+1} - x^\star\rangle + \|\nabla f_t(x_t)\|_\star \|x_t - x_{t+1}\|\,, \qquad (11)\end{aligned}$$

where the final line comes from the generalized Cauchy-Schwarz inequality. Our goal is now to bound the term $\langle \nabla f_t(x_t), x_{t+1} - x^\star\rangle$, and we will do this by using the optimality of x_{t+1} under the OMD update rule.

For this, we let $F_t : \mathcal{X} \to \mathbb{R}$ be the function

$$\begin{aligned}F_t(x) &= \langle \nabla f_t(x_t), x\rangle + \frac{1}{\eta} D_\psi(x, x_t)\\ &= \langle \nabla f_t(x_t), x\rangle + \frac{1}{\eta}\left(\psi(x) - \psi(x_t) - \langle \nabla\psi(x_t), x - x_t\rangle\right),\end{aligned}$$

where the second equality comes from the definition of $D_\psi(x, x_t)$. Now observe by definition of F_t that under the OMD update, $x_{t+1} = \mathrm{argmin}_{x\in\mathcal{X}} F_t(x)$. Thus, by the optimality of x_{t+1}, we have by first-order optimality conditions that $\langle \nabla F_t(x_{t+1}), x^\star - x_{t+1}\rangle \geq 0$, and using the definition of $\nabla F_t(x_{t+1})$, we have

$$\langle \nabla f_t(x_t) + \tfrac{1}{\eta}(\nabla\psi(x_{t+1}) - \nabla\psi(x_t)), x^\star - x_{t+1}\rangle \geq 0\,,$$

which by rearranging implies that

$$\begin{aligned}\langle \nabla f_t(x_t), x_{t+1} - x^\star\rangle &\leq \frac{1}{\eta}\langle \nabla\psi(x_{t+1}) - \nabla\psi(x_t), x^\star - x_{t+1}\rangle \qquad (12)\\ &= \frac{1}{\eta}\left(D_\psi(x^\star, x_t) - D_\psi(x^\star, x_{t+1}) - D_\psi(x_{t+1}, x_t)\right), \qquad (13)\end{aligned}$$

where in the final equality we apply the three-point identity of Lemma 4.

Now substituting (13) back into expression (9), we have

$$\langle \nabla f_t(x_t), x_t - x^\star\rangle \;\leq\; \frac{1}{\eta}\left(D_\psi(x^\star, x_t) - D_\psi(x^\star, x_{t+1})\right) - \frac{1}{\eta} D_\psi(x_{t+1}, x_t) + \|\nabla f_t(x_t)\|_\star \|x_t - x_{t+1}\|\,. \tag{14}$$

(iii) **Apply strong-convexity of** ψ **to simplify**:
Under the assumption that ψ is 1-strongly convex, we can further simplify the latter to terms of (14). For this, applying Lemma , we have for each t that

$$D_\psi(x_{t+1}, x_t) \geq \frac{1}{2}\|x_{t+1} - x_t\|^2\,.$$

It follows that we can write

$$-\frac{1}{\eta}D_\psi(x_{t+1}, x_t) + \|\nabla f_t(x_t)\|_\star \|x_t - x_{t+1}\| \leq \|\nabla f_t(x_t)\|_\star \|x_t - x_{t+1}\| - \frac{1}{2\eta}\|x_{t+1} - x_t\|^2 \tag{15}$$

$$\leq \frac{\eta}{2}\|\nabla f_t(x_t)\|_\star^2 \,. \tag{16}$$

where in the final inequality we apply the identity $az - \frac{b}{2}z^2 \leq \frac{a^2}{2b}$ for $a, b > 0$ and $z \in \mathbb{R}$.

Substituting into (14), we conclude for all t that

$$\langle \nabla f_t(x_t), x_t - x^\star \rangle \leq \frac{1}{\eta}\big(D_\psi(x^\star, x_t) - D_\psi(x^\star, x_{t+1})\big) + \frac{\eta}{2}\|\nabla f_t(x_t)\|_\star^2 \,. \tag{17}$$

(iv) **Sum and simplify**:
The final remaining step to bound $\mathrm{Reg}_{OMD}(T)$ is to sum the terms in (17) over all $t \in [T]$. For this, applying the inequality from (17), we find via telescoping that

$$\mathrm{Reg}_{OMD}(T) \leq \sum_{t=1}^{T} \langle \nabla f_t(x_t), x_t - x^\star \rangle \tag{18}$$

$$\leq \sum_{t=1}^{T} \frac{1}{\eta}\big(D_\psi(x^\star, x_t) - D_\psi(x^\star, x_{t+1})\big) + \frac{\eta}{2}\|\nabla f_t(x_t)\|_\star^2 \tag{19}$$

$$\leq \frac{D_\psi(x^\star, x_1) - D_\psi(x^\star, x_{T+1})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\|\nabla f_t(x_t)\|_\star^2 \tag{20}$$

$$\leq \frac{D_\psi(x^\star, x_1)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\|\nabla f_t(x_t)\|_\star^2 \,, \tag{21}$$

where in the final inequality we use the fact (Lemma 2) that $D_\psi(x^\star, x_{T+1}) \geq 0$.

Finally, recalling the assumption that $x_1 \in \mathcal{X}$ satisfies $D_\psi(x^\star, x_1) \leq D$ concludes the proof.

□

1.4 Geometric Perspective of OMD and Relationship with FTRL

In this section, we describe a more geometric perspective for OMD using the machinery of Fenchel conjugates and duality. We then discuss the relationship between OMD and FTRL.

Convex Analysis Refresher: Fenchel Conjugates. Here we review the notion of Fenchel conjugates. See Orabona (2019, Section 6.4.1) for more details.

Definition 10. *Let $\psi : \mathcal{X} \to \mathbb{R}$ be strictly convex and differentiable. Then the convex (Fenchel) conjugate of ϕ is the function $\phi^* : \mathbb{R}^n \to \mathbb{R}$ given by*

$$\phi^*(y) = \sup_{x \in \mathcal{X}} \langle x, y \rangle - \phi(x) \,.$$

One special class of regularizers ψ are those that satisfy the following *Legendre* property:

Definition 11. *Suppose $\psi : \mathcal{X} \to \mathbb{R}$ is strictly convex and differentiable. Then ψ is a Legendre function if $\|\nabla\psi(x)\| \to \infty$ as $x \to \partial\mathcal{X}$, and $\nabla\psi : \mathcal{X} \to$ range $\nabla\psi(\mathcal{X}) \subseteq \mathbb{R}^n$ is a bijection.*

In particular, if ψ satisfies the Legendre property of Definition 12, then the following holds:

Lemma 12. *Suppose $\psi : \mathcal{X} \to \mathbb{R}$ satisfies Definition . Then $\nabla\psi^* = (\nabla\phi)^{-1}$.*

Geometric view of Online Mirror Descent. Consider the iterates $\{x_t\}$ of running OMD with strictly convex $\psi : \mathcal{X} \to \mathbb{R}$ on loss functions $\{x_t\}$. For simplicity[1] , we will assume that ψ is Legendre and that all $x_t \in \operatorname{int} \mathcal{X}$. Then by definition of D_ψ, observe for $t \geq 1$ that we have

$$\begin{aligned} x_{t+1} &= \operatorname*{argmin}_{x \in \mathcal{X}} \langle \nabla f_t(x_t), x \rangle + \frac{1}{\eta}\big(\psi(x) - \psi(x_t) - \langle \nabla \psi(x_t), x - x_t \rangle\big) \\ &= \operatorname*{argmin}_{x \in \mathcal{X}} \langle \eta \nabla f_t(x_t) - \nabla \psi(x_t), x \rangle + \psi(x) . \end{aligned}$$

By the assumption that $x_t \in \operatorname{int} \mathcal{X}$, the first-order optimality conditions of x_{t+1} imply

$$\nabla \psi(x_{t+1}) = \nabla \psi(x_t) - \eta \nabla f_t(x_t) \quad \Longrightarrow \quad x_{t+1} = \nabla \psi^*\big(\nabla \psi(x_t) - \eta \nabla f_t(x_t)\big) . \tag{22}$$

Here, the implication comes from the fact that $\nabla \psi^* = (\nabla \psi)^{-1}$ when ψ is Legendre (Lemma 1.4). Thus expression (22) shows that the OMD update can be viewed as a certain *online gradient descent* in the dual space of gradients: first, $\nabla \psi$ maps the primal iterate x_t to this dual space, and after taking a gradient step $-\eta \nabla f_t(x_t)$, the inverse map $\nabla \psi^\star$ brings the dual iterate back to the primal space.

Online Mirror Descent and FTRL on Linear Losses. In the case of linear loss functions, this geometric view of OMD coincides with that of FTRL. For this, assume the experts setting with loss vectors $\{\ell_t\}$. Then observe that under FTRL with Legendre regularizer ψ, for all $t \geq 1$:

$$x_{t+1} := \operatorname*{argmin}_{x \in \mathcal{X}} \left\langle x, \sum_{k=1}^{T} \ell_k \right\rangle + \frac{1}{\eta} \psi(x) .$$

For simplicity we again assume that each $x_{t+1} \in \operatorname{int} \mathcal{X}$. Thus again by first-order optimality conditions, and letting $g_t = -\eta \sum_{k=1}^{t-1} \ell_k$, then

$$\nabla \psi(x_{t+1}) = \eta g_{t+1} \quad \Longrightarrow \quad x_{t+1} = \nabla \psi^*(g_t - \eta \ell_t) , \tag{23}$$

where the implication again follows by Lemma . For OMD in the experts setting (technically, with linear loss functions $f_t(x) = \langle x, \ell_t \rangle$), then under the assumption $x_1 := \min_{x \in \mathcal{X}} \psi(x)$, the update in expression (22) means

$$\nabla \psi(x_{t+1}) = -\eta \sum_{k=1}^{t} \ell_k = -\eta g_{t+1} \quad \Longrightarrow \quad x_{t+1} = \nabla \psi^*(g_t - \eta \ell_t) .$$

Thus in the linear experts setting, the iterates of OMD and FTRL coincide.

For more detailed discussion on the geometric interpretation and connection between OMD and FTRL, see (Orabona, 2019, Sections 6.4.2 and 7.3.1).

2 Follow-the-Perturbed-Leader

We have seen thus far the family of Follow-the-Regularized-Leader algorithms (Lecture 02) and the family of Online Mirror Descent algorithms (this lecture), and for both families we have derived general regret bounds that, under mild boundedness assumptions, can be instantiated to obtain $O(\sqrt{T})$ regret guarantees. The key tool used by both algorithms is to employ an explicit regularization or penalty function to deterministically enforce the stability of an algorithm.

In this section, we introduce a third family of algorithms called *Follow-the-Perturbed-Leader* (FTPL). FTPL is similar to FTRL in that it extends the leader-based approach of FTL, but now by adding *random perturbations* to the greedy objective function. For simplicity, we will focus in this section on the experts setting (e.g., linear losses over the simplex):

[1] See (Orabona, 2019, Sections 6.4.2 and 7.3.1) to generalize the argument beyond the Legendre and interior iterates assumption

Algorithm 2 Follow-the-Perturbed-Leader (FTPL) for Experts Setting

Input: Initial $x_1 \in \Delta_n$; perturbation distribution $\mathcal{D}$.
for $t = 1, \ldots, T$ **do**:
1. Play action $x_t \in \Delta_n$, and incur cost $\langle x_t, \ell_t \rangle$. Observe loss vector $\ell_t \in \mathbb{R}^n$.
2. Sample $\rho_t \sim \mathcal{D}$ (independently for each t). Select $x_{t+1} \in \Delta_n$ by

$$x_{t+1} \in \operatorname*{argmin}_{x \in \Delta_n} \left\langle x, \sum_{k=1}^{t} \ell_k + \rho_t \right\rangle, \tag{24}$$

end for

Remark 13. We make several remarks about the FTPL algorithm:

- **Choice of noise distribution:** As stated, we assume the noise distribution $\mathcal{X}$ is some general (but fixed) distribution. For simplicity, in these notes we will mainly consider the case where $\mathcal{D}$ is a uniform distribution over some subset of $\mathbb{R}^n$. However, studying the behavior of FTPL with other classes of distributions is an active area of research.
- **Tiebreaking assumptions:** Notice in (24) that we do not assume x_{t+1} is the unique element of the argmin set. In general, given that the optimization problem in the expression is linear, and due to the fact that we assume $\mathcal{X} = \Delta_n$, the argmin set will not necessarily be a singleton, and in general each x_{t+1} may be a vertex of Δ_n. We will assume there exists some arbitrary tiebreaking rule for selecting the exact iterate x_{t+1} in the case where the argmin set contains multiple elements.
- **Adversary model and expected regret:** Given the randomness of the algorithm, observe that each x_{t+1} is now a random variable whose distribution depends on $\mathcal{D}$. For this reason, we will primarily be concerned with *expected* regret bounds (e.g., bounds on the quantity $\mathbf{E}[\mathrm{Reg}_{FTPL}(T)]$, where the expectation is taken over the randomness of the sequence $\{\rho_t\}$). Moreover, we will assume in the online learning setup that the adversary has knowledge of the distribution $\mathcal{D}$, but not on the realization of the sample $\rho_t \sim \mathcal{D}$ at each round.

We will prove the following expected regret guarantees for FTPL:

Theorem 14. *Let $\{x_t\}$ be the iterates of FTPL (Algorithm 2) with perturbation distribution $\mathcal{D}$ with loss functions $\{f_t(x) := \langle x, \ell_t \rangle\}$. Then for any $T \geq 1$, the following bounds hold:*

(1) Over the randomness of $\{\rho_t\}$ and letting $\rho \sim \mathcal{D}$:

$$\mathbf{E}\left[\mathrm{Reg}_{FTPL}(T)\right] \leq \sum_{t=1}^{T} \mathbf{E}\left[\langle \ell_t, x_t - x_{t+1} \rangle\right] + 2 \cdot \mathbf{E}\left[\|\rho\|_\infty\right] .$$

(2) If $\mathcal{D} = \mathrm{Unif}([0, 1/\epsilon]^n)$ for $\epsilon > 0$, then over the randomness of $\{\rho_t\}$:

$$\mathbf{E}\left[\mathrm{Reg}_{FTPL}(T)\right] \leq \epsilon \sum_{t=1}^{T} \|\ell_t\|_\infty \|\ell_t\|_1 + \frac{2}{\epsilon} .$$

(3) If in addition $\|\ell_t\|_1 \leq G$ and $\|\ell_t\|_\infty \leq L$ for all $t \geq 1$ and $G, L > 0$, then setting $\epsilon := \sqrt{\frac{2}{TGL}}$:

$$\mathbf{E}\left[\mathrm{Reg}_{FTPL}(T)\right] \leq \sqrt{8GLT} .$$

2.1 FTPL Expected Regret Analysis

We now develop the proof of Theorem 14. The strategy is to re-use some of the tools and approaches for the FTRL regret bound from Lecture 02 (in particular, the regret guarantees for FTL and the coupling between FTL and FTRL).

Recap of FTL Regret for Experts Setting. We first recall the general regret bound for Follow-the-Leader (FTL) derived in Lecture 02. In particular, in the experts setting, we have the following bound (see Lemma 9 of Lecture 02).

Proposition 15. *Let $\{x_t\}$ be the iterates of FTL with loss functions $\{f_t(x) := \langle x, \ell_t \rangle\}$. Then*

$$\mathrm{Reg}_{FTL}(T) \leq \sum_{t=1}^{T} f_t(x_t) - f_t(x_{t+1}) = \sum_{t=1}^{T} \langle \ell_t, x_t - x_{t+1} \rangle \ .$$

Note that the statement of this proposition differs slightly from that of Lemma 9 in Lecture 02. The key difference is that Lemma 9 in Lecture 02 further simplifies the statement of the proposition above by applying the generalized Cauchy-Schwarz inequality at each term in the summation. For the proof of FTPL, we will find it more convenient to work with the more general statement above.

General Bound on Expected Regret of FTPL. Similar to FTRL, we can then derive the following expected regret bound for FTPL using Proposition 15:

Proposition 16. *Let $\{x_t\}$ be the iterates of FTPL with perturbation distribution $\mathcal{D}$ on loss functions $\{f_t(x) := \langle x, \ell_t \rangle\}$. Then over the randomness of $\{\rho_t\}$, and letting $\rho \sim \mathcal{D}$:*

$$\mathbf{E}\left[\mathrm{Reg}_{FTPL}(T)\right] \leq \sum_{t=1}^{T} \mathbf{E}\left[\langle \ell_t, x_t - x_{t+1} \rangle\right] + 2 \cdot \mathbf{E}\left[\|\rho\|_\infty\right] \ .$$

The proof of Proposition 16 follows similarly to that of Lemma 12 in Lecture 02 for FTRL. In particular, the strategy is to define the shifted set of loss functions $\{\hat{f}_t\}$ by

$$\begin{cases} \hat{f}_1(x) := \langle x, \rho \rangle \\ \hat{f}_{t+1}(x) := f_t(x) = \langle x, \ell_t \rangle \text{ for } t \geq 1 \ , \end{cases}$$

where each $\rho_t \sim \mathcal{D}$. Then, we can compare the iterates $\{\hat{x}_t\}$ produced by FTL on the sequence $\{\hat{f}_t\}$ to the iterates $\{x_t\}$ produced by FTPL on $\{f_t\}$, and we can apply (in expectation) the general regret bound for FTL from Proposition 15. The full details of the proof are left as an exercise.

Expected Regret of FTPL Under Uniform Perturbations. We now prove the main regret guarantee of FTPL using *uniform* noise (Part (2) of Theorem 14). We restate this guarantee in the following lemma (and note that Part (3) follows as a direct consequence):

Lemma 17. *Let $\{x_t\}$ be the iterates of FTPL with perturbation distribution $\mathcal{D} = \mathrm{Unif}([0, 1/\epsilon]^n)$ for a fixed $\epsilon > 0$ on loss functions $\{f_t(x) := \langle x, \ell_t \rangle\}$. Then over the randomness of the perturbations $\rho_t \sim \mathcal{D}$:*

$$\mathbf{E}\left[\mathrm{Reg}_{FTPL}(T)\right] \leq \epsilon \sum_{t=1}^{T} \|\ell_t\|_1 + \frac{2}{\epsilon} \ .$$

Proof. By Proposition 16, we have for any fixed distribution $\mathcal{D}$ (and for $\rho \sim \mathcal{D}$) that

$$\mathbf{E}\left[\mathrm{Reg}_{FTPL}(T)\right] \leq \sum_{t=1}^{T} \mathbf{E}\left[\langle \ell_t, x_t - x_{t+1} \rangle\right] + 2 \cdot \mathbf{E}\left[\|\rho\|_\infty\right] \ . \tag{25}$$

Under the setting $\mathcal{D} = \mathrm{Unif}([0, 1/\epsilon]^n)$, we will bound each term of (25) separately:

(i) **Bound on $\mathbf{E}[\|\rho\|_\infty]$:**

By definition of $\mathcal{D} := \text{Unif}([0, 1/\epsilon]^n)$, for $\rho \sim \mathcal{D}$, we have $|\rho(i)| \leq 1/\epsilon$ for all $i \in [n]$ deterministically. Thus

$$2\,\mathbf{E}[\|\rho\|_\infty] \leq \frac{2}{\epsilon} \tag{26}$$

(ii) **Bound on $\mathbf{E}[\langle \ell_t, x_t - x_{t+1} \rangle]$:**

Fix $t \geq 1$. We will prove that $\mathbf{E}[\langle \ell_t, x_t - x_{t+1} \rangle] \leq \|\ell_t\|_1 \cdot \epsilon$ under the randomness of $\rho_t \sim \text{Unif}([0, 1/\epsilon]^n)$. For this, we introduce the following pieces of notation: first, let $L_t = \sum_{k=1}^{t-1} \ell_k$. Then define

$$\begin{cases} F_t & = L_t + \rho_t \in \mathbb{R}^n \\ F_{t+1} & = L_t + \ell_t + \rho_{t+1} \in \mathbb{R}^n \\ \Theta_t & = \{L_t + [0, 1/\epsilon]^n\} \cap \{L_t + \ell_t + [0, 1/\epsilon]^n\} \subset \mathbb{R}^n \,. \end{cases} \tag{27}$$

Under the FTPL update rule, observe that we can then write $x_t \in \text{argmax}_{x \in \Delta_n} \langle F_t, x \rangle$ and $x_{t+1} \in \text{argmax}_{x \in \Delta_n} \langle F_{t+1}, x \rangle$. In particular, F_t and F_{t+1} are random variables (where the randomness comes from ρ_t and ρ_{t+1}, respectively), and thus the distributions of x_t and x_{t+1} are completely determined by the distributions of F_t and F_{t+1}, respectively.

Moreover, under the randomness of ρ_t and ρ_{t+1}, the following properties hold (which follow from the structure of $\mathcal{D} = \text{Unif}([0, 1/\epsilon]^n)$ and the independence of ρ_t and ρ_{t+1}:

- Property (a): $F_t \sim \text{Unif}(\{L_t + [0, 1/\epsilon]^n\})$ and $F_{t+1} \sim \text{Unif}(\{L_t + \ell_t + [0, 1/\epsilon]^n\})$.
- Property (b): $\Pr[F_t \in \Theta_t] = \Pr[F_{t+1} \in \Theta_t]$.
- Property (c): $\Pr[F_t \notin \Theta_t] \leq \sum_{i=1}^{d} \Pr[F_{t+1}(i) \notin \Theta_t(i)] \leq \sum_{i=1}^{d} \frac{|\ell_t(i)|}{1/\epsilon} = \epsilon \cdot \|\ell_t\|_1$.

Using the definition of conditional expectation, can then write:

$$\mathbf{E}[\langle x_t, \ell_t \rangle] = \mathbf{E}[\langle x_t, \ell_t \rangle | F_t \in \Theta_t] \cdot \Pr[F_t \in \Theta_t] + \mathbf{E}[\langle x_t, \ell_t \rangle | F_t \notin \Theta_t] \cdot \Pr[F_t \notin \Theta_t] \,.$$

and also:

$$\begin{aligned} \mathbf{E}[\langle x_{t+1}, \ell_t \rangle] &= \mathbf{E}[\langle x_{t+1}, \ell_t \rangle | F_{t+1} \in \Theta_t] \cdot \Pr[F_{t+1} \in \Theta_t] + \mathbf{E}[\langle x_{t+1}, \ell_t \rangle | F_{t+1} \notin \Theta_t] \cdot \Pr[F_{t+1} \notin \Theta_t] \\ &= \mathbf{E}[\langle x_t, \ell_t \rangle | F_t \in \Theta_t] \cdot \Pr[F_t \in \Theta_t] + \mathbf{E}[\langle x_{t+1}, \ell_t \rangle | F_{t+1} \notin \Theta_t] \cdot \Pr[F_{t+1} \notin \Theta_t] \,. \end{aligned}$$

Here, the final equality comes from the fact that $\Pr[F_t \in \Theta_t] = \Pr[F_{t+1} \in \Theta_t]$ from Property (b). Thus, the random variables x_t conditioned on $F_t \in \Theta_t$ and x_{t+1} conditioned on $F_{t+1} \in \Theta_t$ have the same distribution. Subtracting $\mathbf{E}[\langle x_t - x_{t+1}, \ell_t \rangle]$ and using properties (b) and (c) above, we find

$$\begin{aligned} \mathbf{E}[\langle x_t - x_{t+1}, \ell_t \rangle] &= \big(\mathbf{E}[\langle x_t, \ell_t \rangle | F_t \notin \Theta_t] - \mathbf{E}[\langle x_{t+1}, \ell_t \rangle | F_{t+1} \notin \Theta_t] \big) \cdot \Pr[F_t \notin \Theta_t] \\ &\leq \|\ell_t\|_\infty \cdot \Pr[F_t \notin \Theta_t] \\ &\leq \epsilon \cdot \|\ell_t\|_\infty \|\ell_t\|_1 \,. \end{aligned} \tag{28}$$

(iii) **Combining the pieces**:

Combining expressions (26) and (28) and summing over all $t \in [T]$, we conclude

$$\mathbf{E}[\text{Reg}_{FTPL}(T)] \leq \epsilon \sum_{t=1}^{T} \|\ell_t\|_\infty \|\ell_t\|_1 + \frac{2}{\epsilon} \,. \qquad \square$$

References

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.