CSCI 659 Lecture 2

Spring 2026

Instructor: Haipeng Luo

1 A General Algorithmic Framework

In the last lecture, we discussed a special case of OCO, that is, the expert problem, and the classical Hedge algorithm for this problem. For a general OCO problem, how should we design a no-regret algorithm? (Recall the OCO setting: each round the learner decides $w_t \in \Omega$, while the environment decides a convex loss function $f_t : \Omega \to \mathbb{R}$).

Before any further discussion, we first make the following observation: it is sufficient to solve the case where each f_t is a *linear* function. To see this, notice that by convexity, we have $f_t(w_t) - f_t(w) \le \langle \nabla f_t(w_t), w_t - w \rangle$ (draw a picture to convince yourself) and thus

$$\max_{w \in \Omega} \sum_{t=1}^{T} (f_t(w_t) - f_t(w)) \le \max_{w \in \Omega} \sum_{t=1}^{T} \langle \nabla f_t(w_t), w_t - w \rangle .$$

Therefore, we only need to solve a different OCO instance where the loss function $f'_t(w) = \langle \ell_t, w \rangle$ is linear with $\ell_t = \nabla f_t(w_t)$. There are a couple of things worth noticing in this argument:

- First, this shows that knowing the gradient of f_t at the played action w_t is already enough, and we do not really need full information about f_t. More importantly, many algorithms (such as those to be discussed in this lecture) only need to store the cumulative gradients, leading to T-independent space complexity.
- Second, in this reduction, the linear loss function generally becomes adaptively chosen (that is, dependent on w_t), even if the original adversary is oblivious (unless f_t itself is a linear function). Therefore, we do generally need an algorithm that is able to deal with an adaptive adversary.
- Third, while this reduction is sufficient to derive some regret bounds, it might not be the best way to do so, especially if f_t has some curvature making the linear approximation too loose.

In light of this reduction, for the rest of this lecture, we assume that the loss functions are linear and parameterized by $\ell_1, \ldots, \ell_T$.

1.1 The Importance of Stability

Once again, the natural first attempt is the FTL approach: $w_t = \operatorname{argmin}_{w \in \Omega} \left\langle w, \sum_{s=1}^{t-1} \ell_s \right\rangle$, which we already know suffers linear regret in the worst case. If we take a closer look at the worst case instance discussed in Lecture 1, we see that FTL exhibits highly unstable behavior — it alternates betweens two very different decisions. Motivated by this, we explore the idea of stabilizing the algorithm. This might not sound very intuitive at first glance: the loss functions can be changing in some arbitrary way, so why is it a good thing to have a stable learner? One answer is that, recall that the goal of the learner is only to compete with a *fixed* action, so there is really no point in "chasing" the loss functions all the time. Instead, one should stick around and not move too far away from the current best fixed action.

To make this intuition more formal, in this lecture we consider stabilizing the FTL approach by adding an auxiliary loss function $\Psi : \Omega \to \mathbb{R}$ that is differentiable and convex:

$$w_t \in \operatorname*{argmin}_{w \in \Omega} \left\langle w, \sum_{s=1}^{t-1} \ell_s \right\rangle + \Psi(w), \tag{1}$$

and we will discuss two different types of Ψ, leading to two different classical approaches. Before that, we first derive an intermediate regret bound for this general form, which will further illustrate the importance of stability and the role of Ψ.

First, recall that the Bregman divergence $D_\Psi : \Omega \times \Omega \to \mathbb{R}_+$ with respect to Ψ is defined as

$$D_\Psi(w, u) = \Psi(w) - \Psi(u) - \langle \nabla \Psi(u), w - u \rangle ,$$

which is simply the gap between Ψ and its first order approximation at u. Note that this is always nonnegative due to the convexity of Ψ and zero when $w = u$, but in general asymmetric between w and u (thus not a metric). Taylor theorem also tells us that there exists ξ between w and u such that $D_\Psi(w, u) = \frac{1}{2}\|w - u\|^2_{\nabla^2 \Psi(\xi)}$,[1] meaning that the Bregman divergence is roughly measuring some squared quadratic norm of $w - u$.

Question 1. *What is the Bregman divergence with respect to a linear function? What about a quadratic function?*

We will be using the following simple fact.

Lemma 1. *Suppose that $F : \Omega \to \mathbb{R}$ is convex and differentiable, and $w^\star \in \Omega$ minimizes F. Then $F(w^\star) \le F(w) - D_F(w, w^\star)$ holds for any $w \in \Omega$.*

Proof. By definition, this is equivalent to $\langle \nabla F(w^\star), w - w^\star \rangle \ge 0$, which is exactly the first-order optimality condition for $w^\star$ being a minimizer of Φ. □

What this lemma says is that $w^\star$ being a minimizer of F actually tells us a bit more than the trivial fact $F(w^\star) \le F(w)$. With this, we prove the following useful lemma.

Lemma 2 (Minimum Value Difference)**.** *Let $w^\star \in \Omega$ be a minimizer of the function $\langle w, L \rangle + \Psi(w)$ for some arbitrary L and Φ be its minimum value; similarly, let $\bar{w}^\star$ and $\bar{\Phi}$ be a minimizer and the minimum value respectively of the function $\langle w, \bar{L} \rangle + \Psi(w)$ for some arbitrary $\bar{L}$. Then we have*

$$\bar{\Phi} - \Phi \le \left\langle w^\star, \bar{L} - L \right\rangle - D_\Psi(w^\star, \bar{w}^\star).$$

Proof. Simply apply Lemma 1 with $F(w) = \langle w, \bar{L} \rangle + \Psi(w)$ and note that $D_F = D_\Psi$:

$$\begin{aligned} \bar{\Phi} - \Phi &= F(\bar{w}^\star) - \Phi && \text{(by definition)} \\ &\le F(w^\star) - \Phi - D_\Psi(w^\star, \bar{w}^\star) && \text{(Lemma 1)} \\ &= \left\langle w^\star, \bar{L} - L \right\rangle - D_\Psi(w^\star, \bar{w}^\star). && \text{(by definition)} \end{aligned}$$

□

Now we are ready to show the following intermediate regret bound.

Lemma 3. *Algorithm described by Eq. (1) ensures for any $u \in \Omega$:*

$$\begin{aligned} \sum_{t=1}^T \langle w_t - u, \ell_t \rangle &\le \underbrace{\Psi(u) - \min_{w \in \Omega} \Psi(w)}_{\textit{penalty term}} + \underbrace{\sum_{t=1}^T \langle w_t - w_{t+1}, \ell_t \rangle}_{\textit{stability term}} - \sum_{t=1}^T D_\Psi(w_{t+1}, w_t) \\ &\le B_\Psi + \sum_{t=1}^T \langle w_t - w_{t+1}, \ell_t \rangle - \sum_{t=1}^T D_\Psi(w_{t+1}, w_t), \end{aligned}$$

where $B_\Psi = \max_{w \in \Omega} \Psi(w) - \min_{w \in \Omega} \Psi(w)$ is the range of Ψ.

[1] For a PSD matrix M, the notation $\|w\|_M$ denotes $\sqrt{w^\top M w}$, which is equivalent to $\|M^{\frac{1}{2}} w\|_2$.

In this bound, the first term is the penalty or error term, caused by the fact that Eq. (1) is not directly minimizing the cumulative loss, but instead that plus the auxiliary term Ψ. The second term, called stability term, exactly demonstrates why having a stable algorithm is important: the closer w_t and w_{t+1} are (in terms of their loss for the loss function at time t precisely), the smaller the regret. Finally, for this lecture, we can simply ignore the third nonpositive term, but in the future we will see that this is critical in some cases. In fact, after ignoring this term and rearranging, Lemma 3 is simply saying (try to verify yourself)

$$\Psi(w_1) + \sum_{t=1}^{T} \langle w_{t+1}, \ell_t \rangle \le \Psi(u) + \sum_{t=1}^{T} \langle u, \ell_t \rangle$$

for any $u \in \Omega$, that is, the one-step lookahead strategy w_{t+1}, has negative regret (when the auxiliary loss is considered as well). This is often known as the Be-the-Leader lemma (where the leader refers to the one-step lookahead strategy).

Proof of Lemma 3. Let $\Phi_t = \min_{w\in\Omega} \left\langle w, \sum_{s\le t} \ell_s \right\rangle + \Psi(w) = \left\langle w_{t+1}, \sum_{s\le t} \ell_s \right\rangle + \Psi(w_{t+1})$. Then Lemma 2 tells us

$$\Phi_{t-1} - \Phi_t \le - \langle w_{t+1}, \ell_t \rangle - D_\Psi(w_{t+1}, w_t).$$

Adding $\langle w_t, \ell_t \rangle$ to both sides, summing over t, telescoping, and rearranging leads to

$$\sum_{t=1}^{T} \langle w_t, \ell_t \rangle - \Phi_T \le -\Phi_0 + \sum_{t=1}^{T} \langle w_t - w_{t+1}, \ell_t \rangle - \sum_{t=1}^{T} D_\Psi(w_{t+1}, w_t).$$

It thus remains to point out that $\Phi_0 = \Psi(w_1) = \min_{w\in\Omega} \Psi(w)$ and $\Phi_T \le \left\langle u, \sum_{s\le T} \ell_s \right\rangle + \Psi(u)$ for any u by definition. □

Therefore, our goal will be to find Ψ that stabilizes the algorithm while having a relatively small penalty. In the rest of this lecture, we discuss two general approaches to doing so.

2 Follow the Regularized Leader

The first approach is by taking Ψ to be some *strongly convex* function, often called a *regularizer*, to stabilize the algorithm. This approach is called Follow-the-Regularized-Leader, or FTRL for short. Note that regularization is also a widely-used technique in statistical learning (in fact, being able to stabilize the learning algorithm is one explanation of why it helps reduce generalization error).

Formally, let $\Psi = \frac{1}{\eta}\psi$ for some learning rate $\eta > 0$ and a strongly convex function $\psi : \Omega \to \mathbb{R}$. Strong convexity means: for any $w, u \in \Omega$, the following holds:[2]

$$\psi(u) - \psi(w) \le \langle \nabla\psi(u), u - w \rangle - \tfrac{1}{2} \|u - w\|^2 \tag{2}$$

for some norm $\|\cdot\|$. The following observations might help you understand this concept better:

- Compared to the convexity property $\psi(u) - \psi(w) \le \langle \nabla\psi(u), u - w \rangle$, there is an extra nonpositive term in the right-hand side of Eq. (2) (thus indeed "stronger").
- Rearranging Eq. (2) gives $\frac{1}{2} \|w - u\|^2 + \langle \nabla\psi(u), w - u \rangle + \psi(u) \le \psi(w)$, which, for quadratic norm $\|\cdot\|$, says that at any point w, the function $\psi(w)$ is completely above a certain quadratic (the left-hand side) with the same function value and gradient at u, thus illustrating a certain curvature of ψ.
- By definition, Eq. (2) is the same as $\frac{1}{2} \|w - u\|^2 \le D_\psi(w, u) = \frac{1}{2}\|w - u\|^2_{\nabla^2\Psi(\xi)}$ for some ξ between u and w, so strong convexity provides a nice lower bound for the Bregman divergence.

[2] More precisely, this is the definition of ψ being 1-strongly convex. Due to the scaling parameter of $1/\eta$, it does not matter whether ψ is 1-strongly convex or α-strongly convex for some other parameter α.

Due to the curvature, the minimizer of a strongly convex function is unique (verify this yourself), and more importantly, does not move significantly when the function is slightly changed, that is, the minimizer is stable. Formally, we prove the following (one should contrast this with Lemma 2).

Lemma 4 (Minimizer Difference)**.** *Let* $w^\star = \operatorname{argmin}_{w\in\Omega} \langle w, L\rangle + \frac{1}{\eta}\psi(w)$ *and* $\bar{w}^\star = \operatorname{argmin}_{w\in\Omega} \langle w, \bar{L}\rangle + \frac{1}{\eta}\psi(w)$ *for some arbitrary* L *and* $\bar{L}$, $\eta > 0$, *and strongly convex* ψ *(with respect to some norm* $\|\cdot\|$*). We have*

$$\|w^\star - \bar{w}^\star\|^2 \le \eta \left\langle w^\star - \bar{w}^\star, \bar{L} - L\right\rangle, \tag{3}$$

and consequently

$$\|w^\star - \bar{w}^\star\| \le \eta\|L - \bar{L}\|_\star, \tag{4}$$

where $\|\cdot\|_\star$ *is the dual norm.*[3]

Proof. Similar to Lemma 2, define $\Phi = \langle w^\star, L\rangle + \frac{1}{\eta}\psi(w^\star)$ and $\bar{\Phi} = \left\langle \bar{w}^\star, \bar{L}\right\rangle + \frac{1}{\eta}\psi(\bar{w}^\star)$. Then we have

$$\begin{aligned}\bar{\Phi} - \Phi &\le \left\langle w^\star, \bar{L} - L\right\rangle - \tfrac{1}{\eta}D_\psi(w^\star, \bar{w}^\star) && \text{(Lemma 2)}\\ &\le \left\langle w^\star, \bar{L} - L\right\rangle - \tfrac{1}{2\eta}\|w^\star - \bar{w}^\star\|^2. && \text{(definition of strong convexity)}\end{aligned}$$

On the other hand, noticing the symmetry between $\bar{\Phi}$ and Φ, we in fact also have

$$\begin{aligned}\Phi - \bar{\Phi} &\le \left\langle \bar{w}^\star, L - \bar{L}\right\rangle - \tfrac{1}{\eta}D_\psi(\bar{w}^\star, w^\star) && \text{(Lemma 2)}\\ &\le \left\langle \bar{w}^\star, L - \bar{L}\right\rangle - \tfrac{1}{2\eta}\|w^\star - \bar{w}^\star\|^2. && \text{(definition of strong convexity)}\end{aligned}$$

Summing up the two inequalities and rearranging proves Eq. (3). To prove Eq. (4), it suffices to further apply Hölder's inequality:

$$\left\langle w^\star - \bar{w}^\star, L - \bar{L}\right\rangle \le \|w^\star - \bar{w}^\star\|\|L - \bar{L}\|_\star$$

and divide both sides by $\|w^\star - \bar{w}^\star\|$. □

Note that the stability level is naturally controlled by the parameter η. Combining this stability lemma and the intermediate regret bound Lemma 3, we obtain the following regret bound for FTRL, where one see that the tradeoff between the penalty term and stability term is exactly governed by η.

Theorem 1. *The FTRL strategy:* $w_t = \operatorname{argmin}_{w\in\Omega} \left\langle w, \sum_{s<t}\ell_s\right\rangle + \frac{1}{\eta}\psi(w)$ *ensures for any loss sequence* $\ell_1, \ldots, \ell_T$*:*

$$\mathcal{R}_T = \max_{w\in\Omega}\sum_{t=1}^T \langle w_t - w, \ell_t\rangle \le \frac{B_\psi}{\eta} + \eta\sum_{t=1}^T \|\ell_t\|_\star^2 .$$

Therefore, if we further have $\max_t \|\ell_t\|_\star \le G$ *for some* $G > 0$, *then setting* $\eta = \sqrt{\frac{B_\psi}{TG^2}}$ *leads to regret bound* $\mathcal{R}_T = \mathcal{O}(G\sqrt{TB_\psi})$.

Proof. Note that by the definition of FTRL, Lemma 4 exactly tells us $\|w_t - w_{t+1}\| \le \eta\|\ell_t\|_\star$. Therefore, the regret can be bounded as

$$\begin{aligned}\mathcal{R}_T &\le \frac{B_\psi}{\eta} + \sum_{t=1}^T \langle w_t - w_{t+1}, \ell_t\rangle && \text{(Lemma 3)}\\ &\le \frac{B_\psi}{\eta} + \sum_{t=1}^T \|w_t - w_{t+1}\|\|\ell_t\|_\star && \text{(Hölder's inequality)}\\ &\le \frac{B_\psi}{\eta} + \eta\sum_{t=1}^T \|\ell_t\|_\star^2, && \text{(Lemma 4)}\end{aligned}$$

which finishes the proof. □

We conclude by pointing out that having a uniform bound G on $\|\ell_t\|_\star$, which we recall is just $\|\nabla f_t(w_t)\|_\star$, is simply saying that the loss function f_t is G-Lipschitz with respect to the norm $\|\cdot\|_\star$.

[3] Given a norm $\|\cdot\|$ (called primal norm), its dual norm is defined as $\|u\|_\star = \max_{\|w\|\le 1}\langle u, w\rangle$. The most important examples of primal-dual norm pair for this course are $(\|\cdot\|_2, \|\cdot\|_2)$ and $(\|\cdot\|_1, \|\cdot\|_\infty)$.

3 Instances of FTRL

We now discuss concrete instantiations of FTRL for different problems.

3.1 Online Gradient Descent

For an arbitrary action space Ω, we can pick $\psi(w) = \frac{1}{2}\left\|w\right\|_2^2$, a very standard regularizer in machine learning. The induced FTRL is thus

$$w_t = \operatorname*{argmin}_{w\in\Omega} \left\langle w, \sum_{s<t} \ell_s \right\rangle + \frac{1}{2\eta}\left\|w\right\|_2^2 = \operatorname*{argmin}_{w\in\Omega} \left\| w + \eta \sum_{s<t} \ell_s \right\|_2^2,$$

which means w_t is the L_2 projection of $u_t = -\eta \sum_{s<t} \ell_s$ onto Ω. This algorithm is called (lazy) *Online Gradient Descent* (OGD) [Zinkevich, 2003], which can be equivalently written as (after plugging in the original meaning of ℓ_t, that is, $\nabla f_t(w_t)$)

$$u_{t+1} = u_t - \eta \nabla f_t(w_t); \quad w_{t+1} = \operatorname*{argmin}_{w\in\Omega} \left\|w - u_{t+1}\right\|_2 .$$

A closely related variant (which enjoys similar regret bounds and is also referred to as OGD) is the following strategy that actively projects:

$$u_{t+1} = w_t - \eta \nabla f_t(w_t); \quad w_{t+1} = \operatorname*{argmin}_{w\in\Omega} \left\|w - u_{t+1}\right\|_2 .$$

If f_t stays the same over time, then the strategy above is apparently the standard gradient descent for optimizing this function. OGD shows that even if the function is changing over time, making a gradient descent step at each round is still a good strategy.

To apply the general regret bound from Theorem 1, we point out that $\psi(w) = \frac{1}{2}\left\|w\right\|_2^2$ is strongly convex with respect to the L_2 norm (verify it yourself). The dual norm of the L_2 norm is itself, so if G is such that $\max_t \left\|\nabla f_t(w_t)\right\|_2 \le G$, then the regret of OGD is bounded by

$$\mathcal{R}_T \le \frac{\max_{w\in\Omega}\left\|w\right\|_2^2}{2\eta} + \eta T G^2 = \mathcal{O}\left(\max_{w\in\Omega}\left\|w\right\|_2 G\sqrt{T}\right),$$

where the last step is by picking the optimal η. It can be shown that this bound is minimax optimal.

Examples Consider the online regression problem where $\Omega = \{w \in \mathbb{R}^d : \left\|w\right\|_2 \le 1\}$ is a set of linear predictors with bounded norm, and $f_t(w) = \frac{1}{2}(\langle w, x_t\rangle - y_t)^2$ is the square loss for an example $x_t \in \{x \in \mathbb{R}^d : \left\|x\right\|_2 \le 1\}$ and its label $y_t \in [-1, 1]$. Then because $\nabla f_t(w) = (\langle w, x_t\rangle - y_t)x_t$, we have $G = 2$ and $\max_{w\in\Omega}\left\|w\right\|_2 = 1$, and therefore OGD has regret $\mathcal{O}(\sqrt{T})$, *independent* of the dimension of the problem d.

Next consider using OGD for the expert problem (where $\Omega = \Delta(N)$ and $\ell_t \in [0, 1]^N$). In this case we have $\max_{p\in\Delta(N)}\left\|p\right\|_2 \le \max_{p\in\Delta(N)}\left\|p\right\|_1 = 1$, and $\left\|\ell_t\right\|_2 \le \sqrt{N}$. Thus OGD's regret is $\mathcal{O}(\sqrt{TN})$ in this case, which has *exponentially worse* dependence on N compared to Hedge.

3.2 Recovering Hedge for the Expert Problem

OGD fails to achieve the optimal regret bound for the expert problem because the squared L_2 norm regularizer does not fully exploit the particular structure of the problem. Instead, for the simplex, a classical regularizer is the (negative) Shannon entropy function (we switch the notation from w to p for convention): $\psi(p) = \sum_{i=1}^N p(i)\ln p(i)$. Then one can verify that the induced FTRL strategy

$$p_t = \operatorname*{argmin}_{p\in\Delta(N)} \left\langle p, \sum_{s<t} \ell_s \right\rangle + \frac{1}{\eta}\sum_{i=1}^N p(i)\ln p(i)$$

is exactly the Hedge algorithm, that is, $p_t(i) \propto \exp(-\eta \sum_{s<t} \ell_s(i))$ (verify this yourself). In other words, Hedge is just one special case of FTRL. This also gives us more intuition about Hedge: the negative entropy is minimized when p is the uniform distribution, so Hedge is trying to minimize the

cumulative loss while not being too far away from the uniform distribution (in case some currently bad experts become good in the future).

To apply the FTRL regret bound, we use the fact that the entropy function is strongly convex with respect to the L_1 norm. To see this, note that it suffices to prove $\|p-q\|_1^2 \le \|p-q\|_{\nabla^2\psi(\xi)}^2$ for any $\xi \in \Delta(N)$ between two distributions p and q, which is indeed true via a direct application of the Cauchy-Schwarz inequality:

$$\begin{aligned}\|p-q\|_1^2 &= \left(\sum_{i=1}^N |p(i)-q(i)|\right)^2 = \left(\sum_{i=1}^N \sqrt{\xi(i)}\cdot\frac{|p(i)-q(i)|}{\sqrt{\xi(i)}}\right)^2 \\ &\le \left(\sum_{i=1}^N \xi(i)\right)\left(\sum_{i=1}^N \frac{(p(i)-q(i))^2}{\xi(i)}\right) = \|p-q\|_{\nabla^2\psi(\xi)}^2.\end{aligned} \tag{5}$$

Therefore, to apply the general regret bound in Theorem 1, it suffices to do the following two simple calculations: first, the range of the entropy function is $B_\psi = \ln N$; second, the dual norm of the L_1 norm is the L_∞ norm, and by the definition of the problem (that ℓ_t is a vector in $[0,1]^N$), we have $\|\ell_t\|_\infty \le 1$ (note that this step is the key improvement compared to using OGD in this setting). Combining these facts, Theorem 1 implies the following regret bound for Hedge: $\mathcal{R}_T \le \frac{\ln N}{\eta} + T\eta$, the same bound we proved last time using a different potential-based argument.

Question 2. *In this case, can you recognize the connection between the Φ_t defined in the proof of Lemma 3, and the potential function (using the same notation Φ_t) we used in Lecture 1?*

3.3 Combinatorial Problems

Next, we consider a generalization of the expert problem that has a certain combinatorial structure. Let $A = \{a_1, \ldots, a_K\}$ be a set of *combinatorial actions* such that $a_j \in \{0,1\}^N$ and $\max_j \|a_j\|_1 \le m$ for some integer $m \le N$. Roughly speaking, the learner needs to select one of these combinatorial actions at each time. For randomized strategy, this means that the decision space for the learner is the convex hull of A, that is, $\Omega = \left\{\sum_{j=1}^K p(j)a_j : p \in \Delta(K)\right\} \subseteq [0,1]^N$, so that each point in Ω specifies a distribution over these combinatorial actions. We consider linear loss functions $f_t(w) = \langle w, \ell_t\rangle$ for some $\ell_t \in [0,1]^N$. Note that if $w = \sum_{j=1}^K p(j)a_j$, then $\langle w, \ell_t\rangle = \sum_{j=1}^K p(j)\langle a_j, \ell_t\rangle$ is clearly the expected loss if one selects a random combinatorial action according to p.

The expert problem is clearly a special case where A consists of all standard basis vectors in $\mathbb{R}^N$ and $m = 1$. Another example is the so-called m-set problem, where each combinatorial action corresponds to a set of m experts (that is, instead of picking one expert each time, we now need to pick m of them). Formally, in this case we have $A = \{a \in \{0,1\}^N : \|a\|_1 = m\}$ and $\Omega = \{w \in [0,1]^N : \|w\|_1 = m\}$ (recall the multiple-product recommendation example in Lecture 1).

Yet another important example is the online shortest path problem (useful for online routing for example). In this problem, a direct acyclic graph with N edges, a source vertex, and a destination vertex is given. Each round the player first randomly picks a path to send a package, then the loss (e.g. delay) for each edge is revealed, and the player suffers the total loss of all the edges on the selected path. This can be formulated as a special case of the above combinatorial problem by setting A to be the set of all paths starting from the source and ending at the destination (that is, a path is represented by a vector in $\{0,1\}^N$ so that each coordinate indicates whether the corresponding edge is on the path or not). Ω is now the set of all unit flows for this graph. Also note that m is the length of the longest path in A.

How do we solve such problems? The first natural approach is to reduce it to another standard expert problem. Earlier we point out that the latter is a special case of the former, but in fact the reverse is also true: we can simply treat each of the K combinatorial actions a_j as an expert with loss $\langle a_j, \ell_t\rangle$ at time t. Noticing that now the loss of each expert can be at most m, we see that applying Hedge to this problem leads to a regret bound of $\mathcal{O}(m\sqrt{T\ln K}) = \mathcal{O}\left(m\sqrt{mT\ln N}\right)$.

From this discussion we see again the importance of having only logarithmic dependence on the total number of experts: K is potentially exponential in m, so we can only afford $\ln K$ in the regret. However, this does not address the computational issue since running Hedge (naively) for this problem does still require $\mathcal{O}(K)$ time complexity.

To address this, we consider another approach: run FTRL directly over Ω with a suitable regularizer. The regularizer we use is the generalized entropy $\psi(w) = m\sum_{i=1}^N w(i)\ln w(i)$, which looks identical to the standard entropy except for being scaled up by m and extended from the simplex to the space of Ω. The induced FTRL strategy:

$$w_t = \operatorname*{argmin}_{w\in\Omega} \left\langle w, \sum_{s<t} \ell_s \right\rangle + \frac{m}{\eta}\sum_{i=1}^N w(i)\ln w(i)$$

is an N-dimensional convex problem with a polytope constraint set Ω. Therefore, as long as Ω is not too complicated, that is, described by poly(N, m) number of linear constraints (which is indeed the case for m-set and online shortest path), then one can apply any standard convex optimization method (such as the Interior Point Method) to find w_t efficiently.

After addressing the computational issue, we now consider the regret of this method. Via an argument similar to Eq. (5), one can show that the generalized entropy is also strongly convex with respect to the L_1 norm (verify it yourself). In addition, the range of ψ is now $B_\psi \le m^2 \ln\frac{N}{m}$, and we still have $\|\ell_t\|_\infty \le 1$. Applying Theorem 1 thus shows a regret bound of $\frac{m^2\ln\frac{N}{m}}{\eta} + \eta T$, which is $\mathcal{O}\big(m\sqrt{T\ln\frac{N}{m}}\big)$ with the optimal η and is $\sqrt{m}$ times better than what Hedge achieves. To conclude, by directly applying FTRL with a suitable regularizer over Ω, we not only resolve the computational issue, but also achieve an even better regret bound (optimal in fact).

4 Follow the Perturbed Leader

Even though the FTRL strategy for the combinatorial problems amounts to solving a convex problem that often admits polynomial time complexity, the actual running time might still be quite large. Is there a more efficient approach? In particular, is it possible to solve this problem based on only an offline linear optimization oracle that solves problems of the form $\operatorname{argmin}_{w\in\Omega}\langle w, L\rangle$? Notice that for m-set, this simply corresponds to finding the m smallest coordinates, while for online shortest path, this corresponds to finding the shortest path of a given graph, which are all "simple" problems.

Motivated by these questions, researchers developed another type of no-regret algorithms called Follow-the-Perturbed-Leader (FTPL) [Kalai and Vempala, 2005]. The idea is to introduce stability via *random perturbation/noise*. Specifically, we let $\Psi(w)$ be a (random) linear function $\langle w, \ell_0\rangle$ where ℓ_0 is drawn from some distribution, and the strategy at time t is simply

$$w_t \in \operatorname*{argmin}_{w\in\Omega} \left\langle w, \sum_{s=0}^{t-1} \ell_s \right\rangle,$$

which can clearly be solved by one call to the aforementioned linear optimization oracle, making the algorithm highly efficient.

Why is FTPL stable? The key is really in the randomness of ℓ_0, since we already know that linear functions do not admit stable minimizers. However, if the distribution of ℓ_0 is dispersed enough, then in expectation the minimizer should not change significantly between two rounds. To illustrates this idea, we focus on the combinatorial setup and one particular distribution of the noise, and prove the following stability lemma (even though the same idea is applicable more generally). For simplicity, we assume w_t is always selected as one of the combinatorial actions $a \in A$ (since the minimum value of a linear function over a polytope can always be achieved by one of its vertices), and the tie is broken in some deterministic way. Also, we restrict our attention to an oblivious adversary who decides the loss sequence ahead of time, so that the only randomness in the following discussion comes from ℓ_0. (To deal with adaptive adversary, it turns out that it suffices to resample ℓ_0 at every time to avoid leaking the randomness to the adversary; see [Hutter and Poland, 2005, Lemma 12].)

Lemma 5. *For the combinatorial problem described in Section 3.3, consider running the FTPL strategy* $w_t \in \operatorname{argmin}_{w\in\Omega}\left\langle w, \sum_{s=0}^{t-1}\ell_s\right\rangle$ *where each coordinate of* ℓ_0 *is an i.i.d. sample of the Laplace distribution with density function* $h(x) = \frac{\eta}{2}e^{-\eta|x|}$ *for some parameter* $\eta > 0$. *Then we have* $\mathbb{E}[\langle w_t - w_{t+1}, \ell_t\rangle] \le \eta N m$.

Proof. Slightly abusing the notation, let $h(\ell_0) = \Pi_{i=1}^N h(\ell_0(i)) = \frac{\eta}{2} e^{-\eta\|\ell_0\|_1}$ be the density function of the noise vector. For any combinatorial action $a_j \in A$, define $p_t(j)$ to be the probability of the event $w_t = a_j$ (with respect to the randomness of ℓ_0), which can be written as

$$\begin{aligned} p_t(j) &= \int_{\ell_0 \in \mathbb{R}^N} \mathbf{1}\left[a_j = \operatorname*{argmin}_{w\in\Omega} \left\langle w, \sum_{s=0}^{t-1} \ell_s \right\rangle\right] h(\ell_0) d\ell_0 \\ &= \int_{\ell_0 \in \mathbb{R}^N} \mathbf{1}\left[a_j = \operatorname*{argmin}_{w\in\Omega} \left\langle w, \sum_{s=0}^{t} \ell_s \right\rangle\right] h(\ell_0 + \ell_t) d\ell_0. \\ &\qquad\qquad\qquad\qquad\qquad\qquad (\text{change of variable: from } \ell_0 \text{ to } \ell_0 + \ell_t) \end{aligned}$$

Since $h(\ell_0 + \ell_t) = \frac{\eta}{2} e^{-\eta\|\ell_0+\ell_t\|_1} \le \frac{\eta}{2} e^{-\eta\|\ell_0\|_1 + \eta\|\ell_t\|_1} = h(\ell_0)e^{\eta\|\ell_t\|_1}$, we continue with

$$p_t(j) \le e^{\eta\|\ell_t\|_1} \int_{\ell_0 \in \mathbb{R}^N} \mathbf{1}\left[a_j = \operatorname*{argmin}_{w\in\Omega} \left\langle w, \sum_{s=0}^{t} \ell_s \right\rangle\right] h(\ell_0) d\ell_0 = e^{\eta\|\ell_t\|_1} p_{t+1}(j).$$

Finally, this means

$$\begin{aligned} \mathbb{E}[\langle w_t - w_{t+1}, \ell_t\rangle] &= \sum_{j=1}^K (p_t(j) - p_{t+1}(j)) \langle a_j, \ell_t\rangle \le \sum_{j=1}^K (1 - e^{-\eta\|\ell_t\|_1}) p_t(j) \langle a_j, \ell_t\rangle \\ &\le \sum_{j=1}^K \eta\|\ell_t\|_1 \cdot p_t(j) \langle a_j, \ell_t\rangle \le \eta N m \sum_{j=1}^K p_t(j) = \eta N m, \end{aligned}$$

where the second inequality uses the fact $1 - e^{-z} \le z$ for all z. □

With this stability lemma, we prove the following regret bound.

Theorem 2. *The FTPL strategy described in Lemma 5 ensures $\mathbb{E}[\mathcal{R}_T] \le \frac{2m}{\eta}(1 + \ln N) + \eta T N m$, which is $\mathcal{O}(m\sqrt{TN\ln N})$ with the optimally tuned η.*

Proof. It suffices to apply Lemma 3 and figure out the expected value of $\max_w \langle w, \ell_0\rangle - \min_w \langle w, \ell_0\rangle$, which by symmetry of the Laplace distribution, is $2\mathbb{E}[\max_{w\in\Omega} \langle w, \ell_0\rangle]$, and bounded by $2m\mathbb{E}[\|\ell_0\|_\infty]$. For any value of $b > 0$, we further bound $\mathbb{E}[\|\ell_0\|_\infty]$ as

$$\begin{aligned} \mathbb{E}[\|\ell_0\|_\infty] &= \int_0^\infty \Pr\left[\|\ell_0\|_\infty \ge x\right] dx \le b + \int_b^\infty \Pr\left[\|\ell_0\|_\infty \ge x\right] dx \\ &\le b + \sum_{i=1}^N \int_b^\infty \Pr\left[|\ell_0(i)| \ge x\right] dx = b + N \int_b^\infty e^{-\eta x} dx = b + \frac{N}{\eta} e^{-\eta b}, \end{aligned}$$

where the first equality uses the standard fact: for any nonnegative random variable X with density function g, $\mathbb{E}[X] = \int_0^\infty xg(x)dx = \int_0^\infty \int_0^x g(x)dydx = \int_0^\infty \int_y^\infty g(x)dxdy = \int_0^\infty \Pr[X \ge y]dy$; and the second inequality is by a union bound. Picking $b = \frac{1}{\eta}\ln N$ leads to the minimum upper bound $\frac{1}{\eta}(1 + \ln N)$. Combining this fact, Lemma 5, and Lemma 3 thus finishes the proof. □

We conclude with two remarks. First, the role of η is exactly the same as the learning rate in FTRL, that is, to balance between the penalty term and the stability term. Second, the regret bound we prove above has the undesirable $\sqrt{N}$ dependence. However, this is merely an artifact of the loose analysis of the stability term. In HW1, we will improve it and replace the $\sqrt{N}$ dependence with $\sqrt{m}$, which then basically matches the regret bound of running Hedge inefficiently over all combinatorial actions, while enjoying the favorable time complexity of one linear optimization per round.

References

Marcus Hutter and Jan Poland. Adaptive online prediction by following the perturbed leader. *Journal of Machine Learning Research*, 6(Apr):639–660, 2005.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, 2003.