FALL 2026 BAIS 7900

Lecture 3: Online Convex Optimization: OGD and FTRL

Lecturer: *Mengxiao Zhang* September 9th, 2026

1 Online Convex Optimization (OCO)

The expert problem from the previous lecture asks the learner to choose among a finite collection of actions. While this framework is useful in many applications, many operational and statistical decisions are instead continuous:

- an order quantity $q_t \in [0, Q]$ can take any value in an interval;
- a prediction rule is described by a coefficient vector $w_t \in \mathbb{R}^d$;
- a portfolio is a weight vector $w_t \in \Delta(N)$; and
- an allocation plan may lie in a high-dimensional feasible region.

OCO replaces the finite action set by a convex decision set $\Omega \subseteq \mathbb{R}^d$. Convexity is useful twice: it makes many offline optimization problems tractable, and it lets a gradient summarize how the current loss changes near the learner's decision.

The interaction on round $t = 1, \ldots, T$ is:

1. the learner chooses $w_t \in \Omega$ using only past information;
2. the environment selects a convex loss function $f_t : \Omega \to \mathbb{R}$; and
3. the learner suffers loss $f_t(w_t)$ and observes f_t (or only first-order feedback, namely a gradient or subgradient at w_t, as we will see later).

For a fixed comparator $u \in \Omega$, define $\mathrm{Reg}_T(u) \triangleq \sum_{t=1}^{T} \big(f_t(w_t) - f_t(u)\big)$. The regret is then defined as the difference from the best fixed decision in hindsight:

$$\mathrm{Reg}_T \triangleq \sum_{t=1}^{T} f_t(w_t) - \min_{u \in \Omega} \sum_{t=1}^{T} f_t(u). \tag{1}$$

1.1 Examples

Example 1.1 (Inventory)**.** A retailer chooses $q_t \in [0, Q]$ before demand d_t is known. With holding cost h and shortage cost b, the loss is

$$f_t(q) \triangleq h(q - d_t)_+ + b(d_t - q)_+.$$

This V-shaped function is convex. As discussed before, in many applications the retailer *does not* observe the complete loss function f_t after making her decision q_t, due to potential stock-out. However, the retailer can still observe a *(sub-)gradient* at her decision in this round.

Specifically, let $s_t \triangleq \min\{q_t, d_t\}$ denote the observed sales. A valid subgradient can be constructed as

$$g_t \triangleq \begin{cases} h, & s_t < q_t, \\ -b, & s_t = q_t. \end{cases}$$

Indeed, if $s_t < q_t$, then the retailer has leftover inventory and therefore knows that $d_t < q_t$, in which case the gradient is h. If $s_t = q_t$, then the retailer knows that $d_t \geq q_t$, even though the exact demand may be censored, and $-b$ is a valid subgradient.

At $q_t = d_t$, any value in $[-b, h]$ is a valid subgradient (so the choice $-b$ above remains valid). We will see later that receiving this information is already enough to achieve sublinear regret.

Example 1.2 (Online linear prediction)**.** At time t, a feature vector $x_t \in \mathbb{R}^d$ arrives and the learner predicts $\widehat{y}_t = \langle w_t, x_t \rangle$. The learner's loss is defined as the following square loss:

$$f_t(w) \triangleq \frac{1}{2}\big(\langle w, x_t \rangle - y_t\big)^2.$$

In this case, suppose that the learner observes the true label y_t after making her decision. Since both x_t and y_t are then known, the learner knows the complete loss function f_t. In particular, she can compute its gradient at w_t:

$$\nabla f_t(w_t) = \big(\langle w_t, x_t \rangle - y_t\big) x_t.$$

Example 1.3 (Portfolio choice)**.** Let $w_t \in \Delta(N)$ be portfolio weights and let $r_t \in \mathbb{R}^N_+$ contain the assets' gross returns. The negative log-growth loss

$$f_t(w) \triangleq -\ln \langle w, r_t \rangle$$

is convex wherever $\langle w, r_t \rangle > 0$. If the complete return vector r_t is revealed at the end of the round, then the learner also knows the complete loss function f_t after making her decision.

1.2 Reduction to Online Linear Optimization

It turns out that gradient information is already enough, thanks to convexity. Specifically, if f_t is differentiable, let $g_t \triangleq \nabla f_t(w_t)$. More generally, we may choose any subgradient $g_t \in \partial f_t(w_t)$. By the subgradient inequality for convex functions, for every $u \in \Omega$,

$$f_t(w_t) - f_t(u) \leq \langle g_t, w_t - u \rangle . \tag{2}$$

Summing over time shows that, for every fixed comparator u,

$$\mathrm{Reg}_T(u) \leq \sum_{t=1}^{T} \langle g_t, w_t - u \rangle = \sum_{t=1}^{T} (\widetilde{f}_t(w_t) - \widetilde{f}_t(u)) \leq \sum_{t=1}^{T} \widetilde{f}_t(w_t) - \min_{u^\star \in \Omega} \sum_{t=1}^{T} \widetilde{f}_t(u^\star). \tag{3}$$

The right-hand side is the regret for the linearized loss $\widetilde{f}_t(w) \triangleq \langle g_t, w \rangle$. Therefore, it is enough to design an algorithm that performs well on the sequence of observed gradients. For notational convenience, we write $L_t \triangleq \sum_{s=1}^{t} g_s$ and $L_0 \triangleq \mathbf{0}$ for the cumulative gradient through round t.

2 Online Gradient Descent

If one is familiar with statistical machine learning, gradient descent is probably the first method that comes to mind for minimizing a differentiable function. Suppose, for a moment, that the same convex loss function f is used on every round. Starting from w_t, the usual gradient descent step moves in the direction $-\nabla f(w_t)$, which locally decreases the loss. If the step leaves the feasible set Ω, we simply project it back.

In the online problem, the loss function can change from f_t to f_{t+1}. Nevertheless, after observing $g_t \in \partial f_t(w_t)$, we can still take one gradient step using the current loss and use the resulting point on the next round. This gives the Online Gradient Descent (OGD) update

$$w_{t+1} = \Pi_\Omega(w_t - \eta g_t), \qquad \Pi_\Omega(z) \triangleq \arg\min_{w \in \Omega} \|w - z\|_2 , \tag{4}$$

where $\eta > 0$ is the learning rate. We assume that Ω is nonempty, closed, and convex, so the Euclidean projection is well defined and unique. The complete pseudocode for online gradient descent is shown in Algorithm 1.

Algorithm 1: Online Gradient Descent (OGD)

Input: closed convex set Ω, learning rate $\eta > 0$, initial decision $w_1 \in \Omega$
for $t = 1, \dots, T$ **do**
  play w_t and observe $g_t \in \partial f_t(w_t)$;
  compute $w_{t+1} = \Pi_\Omega(w_t - \eta g_t)$;

There are two simple but useful observations. First, the update is legal: the learner uses g_t only after playing w_t, and hence only to construct w_{t+1}. Second, if $f_t = f$ for every t, then OGD is exactly the projected gradient descent algorithm used to minimize the fixed function f. The interesting point is that the same update continues to work even when the loss function changes on every round.

2.1 Regret guarantee for OGD

We now show that this very natural online version of gradient descent already achieves sublinear regret. The proof relies on the following basic property of Euclidean projection: for every $u \in \Omega$,

$$\|\Pi_\Omega(z) - u\|_2 \le \|z - u\|_2 . \tag{5}$$

In words, projecting onto Ω cannot increase the distance to a feasible comparator.

Theorem 2.1. *Suppose the feasible set and the observed gradients satisfy*

$$\|w - u\|_2 \le D \quad \textit{for all } w, u \in \Omega, \qquad \|g_t\|_2 \le G \quad \textit{for all } t.$$

Then projected OGD satisfies

$$\mathrm{Reg}_T(u) \le \frac{D^2}{2\eta} + \frac{\eta T G^2}{2} \qquad \textit{for every } u \in \Omega. \tag{6}$$

In particular, choosing $\eta = D/(G\sqrt{T})$ *gives* $\mathrm{Reg}_T(u) \le DG\sqrt{T}$.

Proof. Fix any $u \in \Omega$. Applying (5) to the OGD update and expanding the square gives

$$\|w_{t+1} - u\|_2^2 \le \|w_t - \eta g_t - u\|_2^2 = \|w_t - u\|_2^2 - 2\eta \langle g_t, w_t - u\rangle + \eta^2 \|g_t\|_2^2 .$$

Rearranging gives

$$\langle g_t, w_t - u\rangle \le \frac{\|w_t - u\|_2^2 - \|w_{t+1} - u\|_2^2}{2\eta} + \frac{\eta}{2}\|g_t\|_2^2 . \tag{7}$$

Summing (7) over t makes the squared-distance terms telescope. We therefore obtain

$$\sum_{t=1}^{T} \langle g_t, w_t - u\rangle \le \frac{\|w_1 - u\|_2^2 - \|w_{T+1} - u\|_2^2}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\|g_t\|_2^2 \le \frac{D^2}{2\eta} + \frac{\eta T G^2}{2}.$$

Finally, inequality (3) transfers the bound from the linearized losses to the original convex losses. □

The two terms in (6) have the usual learning-rate tradeoff. A small η makes the algorithm move slowly but enlarges the first term, whereas a large η reacts more strongly to each gradient and enlarges the second term. Balancing the two terms gives the $O(\sqrt{T})$ regret bound. In particular, $\mathrm{Reg}_T(u)/T = O(1/\sqrt{T})$, so the average regret vanishes.

2.2 Connection to statistical machine learning

The OGD update also gives a direct connection back to the statistical learning setting from the previous lecture. Suppose $Z_1, \dots, Z_T$ are independent samples from the same distribution and

$$f_t(w) \triangleq f(w; Z_t).$$

Define the population loss by $F(w) \triangleq \mathbb{E}[f(w; Z)]$. In this setting, OGD is exactly projected stochastic gradient descent: on round t, it takes a gradient step using the new sample Z_t.

Let

$$\overline{w}_T \triangleq \frac{1}{T}\sum_{t=1}^{T} w_t.$$

Because w_t depends only on the previous samples, it is independent of the new sample Z_t. The online-to-batch argument from the previous lecture, together with convexity of F, gives

$$\mathbb{E}[F(\overline{w}_T)] - F(u) \le \frac{\mathbb{E}[\mathrm{Reg}_T(u)]}{T} \le \frac{D^2}{2\eta T} + \frac{\eta G^2}{2}. \tag{8}$$

With $\eta = D/(G\sqrt{T})$, the expected excess population loss is at most $DG/\sqrt{T}$. Thus the same calculation can be read in two ways: adversarially, it is a regret guarantee for OGD; statistically, it is a generalization guarantee for stochastic gradient descent through online-to-batch conversion.

Examples of OGD Updates

- **Inventory.** For $q_t \in [0, Q]$, projected OGD becomes

  $$q_{t+1} = \Pi_{[0,Q]}(q_t - \eta g_t).$$

  If demand exceeds inventory, then $g_t = -b$, so the next order increases by ηb before clipping. If inventory is left over, then $g_t = h$, so the next order decreases by ηh. For example, if $Q = 100$, $q_t = 40$, $d_t = 55$, $b = 4$, and $\eta = 2$, then $q_{t+1} = 48$.

- **Online regression.** For the square loss introduced earlier, the update is

$$w_{t+1} = \Pi_\Omega \left(w_t - \eta \big(\langle w_t, x_t \rangle - y_t \big) x_t \right).$$

  An overprediction moves w_t against x_t, whereas an underprediction moves it toward x_t. If $\|x_t\|_2$, $|y_t|$, and $\|w\|_2$ are bounded, then the gradient is also bounded and Theorem 2.1 applies.

3 From Follow-the-Leader to Follow-the-Regularized-Leader

Let us take a step back and reconsider one (failed) algorithm that we have seen before: Follow-the-Leader (FTL). As discussed in the expert problem, FTL chooses the action that has incurred the smallest cumulative loss over the previous rounds. In online linear optimization, this idea leads to the update

$$w_t = \operatorname*{arg\,min}_{w \in \Omega} \langle L_{t-1}, w \rangle . \tag{9}$$

Here and below, arg min denotes one selected minimizer; when needed, ties are broken according to a fixed rule. In other words, FTL chooses the decision that would have been optimal if the game had ended at round $t-1$. This is a natural strategy, but it can fail badly.

From the previously discussed counterexample, we can see one important failure mode of FTL: it is very sensitive to newly observed losses. Although the cumulative loss changes by only one additional term on each round, the leader may jump from one end of the decision set to the other. FTL follows the historical leader too aggressively, even when that leader changes only because of a very small difference in cumulative loss.

In contrast, if we examine the behavior of OGD, we see that its decisions change in a much more controlled manner. Recall that OGD updates $w_{t+1} = \Pi_\Omega(w_t - \eta g_t)$. We have

$$\|w_{t+1} - w_t\|_2 = \|\Pi_\Omega(w_t - \eta g_t) - \Pi_\Omega(w_t)\|_2 \le \eta \|g_t\|_2 . \tag{10}$$

Therefore, if $\|g_t\|_2 \le G$, then $\|w_{t+1} - w_t\|_2 \le \eta G$. Unlike FTL, OGD does not completely change its decision in response to one new loss. Instead, it moves only slightly from the previous decision, with the amount of movement controlled by the learning rate.

This comparison suggests that stability is an important reason why OGD performs well and, at the same time, an important property missing from FTL. This leads to a natural question: can we retain the cumulative-loss perspective of FTL while preventing its decisions from changing too abruptly? Follow-the-Regularized-Leader answers this question by adding a regularizer to the cumulative loss.

Regularization and strong convexity. Regularization is already familiar from statistics and machine learning. Instead of minimizing only an empirical loss, one often adds a regularization term, such as an ℓ_2 penalty, to discourage extreme or overly complex solutions. For example, the regularized empirical-risk minimization problem takes the form

$$\min_{w \in \Omega} \left\{ \sum_{s=1}^{n} f_s(w) + \lambda \|w\|_2^2 \right\}.$$

In the online setting, regularization will serve a closely related purpose: it prevents the minimizer of the cumulative loss from changing too abruptly when a new loss is added.

The useful property of a regularizer is not only that it penalizes large decisions, but also that it has sufficient curvature. Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. A differentiable function $\psi : \Omega \to \mathbb{R}$ is called 1-*strongly convex with respect to* $\|\cdot\|$ if, for every $w, v \in \Omega$,

$$\psi(v) \geq \psi(w) + \langle \nabla\psi(w), v - w\rangle + \frac{1}{2}\|v - w\|^2 . \tag{11}$$

A standard example is $\psi(w) \triangleq \frac{1}{2}\|w\|_2^2$, which is 1-strongly convex with respect to the Euclidean norm. We will see shortly that strong convexity is exactly what allows us to control the movement between two consecutive decisions.

Follow-the-Regularized-Leader. Follow-the-Regularized-Leader (FTRL) adds a regularizer to the cumulative linearized loss. Given a learning rate $\eta > 0$ and a regularizer ψ, FTRL chooses

$$w_t = \arg\min_{w\in\Omega} \left\{ \langle L_{t-1}, w\rangle + \frac{1}{\eta}\psi(w) \right\}. \tag{12}$$

We assume that the minimizers exist. The complete procedure is summarized in Algorithm 2.

Algorithm 2: Follow-the-Regularized-Leader (FTRL)

Input: convex set Ω, learning rate $\eta > 0$, regularizer ψ
initialize $L_0 \triangleq \mathbf{0}$;
for $t = 1, \ldots, T$ **do**
  compute $w_t = \arg\min_{w\in\Omega} \left\{ \langle L_{t-1}, w\rangle + \frac{1}{\eta}\psi(w) \right\}$;
  play w_t and observe $g_t \in \partial f_t(w_t)$;
  update $L_t = L_{t-1} + g_t$;

At the beginning of round t, the vector L_{t-1} contains all the gradient information observed before this round. The term $\langle L_{t-1}, w\rangle$ therefore asks the learner to follow the decision that has performed well on the previous linearized losses. The additional term $\psi(w)/\eta$ discourages the minimizer from reacting too strongly to small changes in L_{t-1}. A smaller η places more weight on the regularizer and generally produces more conservative decisions. Notice also that, because $L_0 = \mathbf{0}$, $w_1 = \arg\min_{w\in\Omega} \psi(w)$. Thus, the first decision is the point preferred by the regularizer.

Be-the-Leader Lemma. To analyze FTRL, it is useful to introduce the decision that would be chosen after observing the loss on round t:

$$w_{t+1} = \arg\min_{w\in\Omega} \left\{ \langle L_t, w\rangle + \frac{1}{\eta}\psi(w) \right\}.$$

This is only an analytical device: the learner still plays w_t before observing g_t.

We first establish an inequality for these look-ahead decisions. Define $h_0(w) \triangleq \frac{1}{\eta}\psi(w)$ and $h_t(w) \triangleq \langle g_t, w\rangle$ for $t \geq 1$. Then the FTRL update can be written as $w_{t+1} = \arg\min_{w\in\Omega} \sum_{s=0}^{t} h_s(w)$. The following Be-the-Leader lemma says that if we are able to "look ahead", then we are able to achieve non-positive regret.

Lemma 3.1 (Be-the-Leader Lemma)**.** *For every* $u \in \Omega$, $\sum_{t=0}^{T} (h_t(w_{t+1}) - h_t(u)) \leq 0$.

Proof. We prove the result by induction on T. When $T = 0$, the claim follows because w_1 minimizes h_0 over Ω. Now suppose that the claim holds with $T = t - 1$. Since it holds for every comparator, we may apply it with $u = w_{t+1}$ and obtain

$$\sum_{\tau=0}^{t-1} h_\tau(w_{\tau+1}) \le \sum_{\tau=0}^{t-1} h_\tau(w_{t+1}).$$

Adding $h_t(w_{t+1})$ to both sides gives

$$\sum_{\tau=0}^{t} h_\tau(w_{\tau+1}) \le \sum_{\tau=0}^{t} h_\tau(w_{t+1}).$$

By the definition of w_{t+1}, we know that $\sum_{\tau=0}^{t} h_\tau(w_{t+1}) \le \sum_{\tau=0}^{t} h_\tau(u)$ for every $u \in \Omega$. Combining the last two inequalities proves the result. □

The Be-the-Leader lemma leads directly to the following regret decomposition.

Lemma 3.2. *For every $u \in \Omega$, FTRL satisfies*

$$\sum_{t=1}^{T} \langle g_t, w_t - u \rangle \le \frac{\psi(u) - \psi(w_1)}{\eta} + \sum_{t=1}^{T} \langle g_t, w_t - w_{t+1} \rangle . \tag{13}$$

Proof. Adding and subtracting w_{t+1} gives

$$\sum_{t=1}^{T} \langle g_t, w_t - u \rangle = \sum_{t=1}^{T} \langle g_t, w_{t+1} - u \rangle + \sum_{t=1}^{T} \langle g_t, w_t - w_{t+1} \rangle .$$

By Lemma 3.1, $\sum_{t=1}^{T} \langle g_t, w_{t+1} - u \rangle \le \frac{\psi(u) - \psi(w_1)}{\eta}$. Substituting this inequality into the preceding identity proves the claim. □

The first term on the right-hand side of (13) is the price of regularization. The second term measures the stability of the algorithm. If consecutive decisions are close, then this term is small.

Connection to the potential method. The preceding argument is closely related to the potential method used in the analysis of Hedge. To see the connection, define the optimal regularized cumulative loss

$$\Phi_t \triangleq \min_{w \in \Omega} \left\{ \langle L_t, w \rangle + \frac{1}{\eta} \psi(w) \right\} .$$

The change from Φ_{t-1} to Φ_t measures the effect of adding the new linear loss $\langle g_t, w \rangle$. Summing these changes over time produces a telescoping argument, just as summing changes of the log potential did in the analysis of Hedge. The particular potentials are different, but they play the same role. Lemma 3.1 packages this telescoping argument into a convenient inequality.

Strong convexity controls movement. It remains to control the stability term in (13). Let the dual norm of $\|\cdot\|$ be

$$\|z\|_* \triangleq \sup_{\|x\|\le 1} \langle z, x\rangle . \tag{14}$$

By the definition of the dual norm, we have $\langle x, y\rangle \le \|x\|_* \|y\|$. The following lemma shows that the per-round movement of FTRL is controlled by the learning rate multiplied by the norm of the gradient.

Lemma 3.3 (FTRL stability)**.** *Suppose that ψ is 1-strongly convex with respect to $\|\cdot\|$. Then*

$$\|w_{t+1} - w_t\| \le \eta \|g_t\|_* . \tag{15}$$

Consequently, $\langle g_t, w_t - w_{t+1}\rangle \le \eta \|g_t\|_*^2$.

Proof. Define

$$F_{t-1}(w) \triangleq \langle L_{t-1}, w\rangle + \frac{1}{\eta}\psi(w).$$

Because ψ is 1-strongly convex, F_{t-1} is $(1/\eta)$-strongly convex. Since w_t minimizes F_{t-1} over Ω, the first-order optimality condition gives

$$\langle \nabla F_{t-1}(w_t), w_{t+1} - w_t\rangle \ge 0.$$

Combining this inequality with strong convexity gives

$$F_{t-1}(w_{t+1}) \ge F_{t-1}(w_t) + \frac{1}{2\eta}\|w_{t+1} - w_t\|^2 . \tag{16}$$

Similarly, define

$$F_t(w) \triangleq \langle L_t, w\rangle + \frac{1}{\eta}\psi(w) = F_{t-1}(w) + \langle g_t, w\rangle .$$

Since w_{t+1} minimizes F_t over Ω, the same argument gives

$$F_t(w_t) \ge F_t(w_{t+1}) + \frac{1}{2\eta}\|w_{t+1} - w_t\|^2 . \tag{17}$$

Adding (16) and (17), and then canceling the common terms, gives

$$\langle g_t, w_t - w_{t+1}\rangle \ge \frac{1}{\eta}\|w_{t+1} - w_t\|^2 .$$

On the other hand, the dual-norm inequality gives $\langle g_t, w_t - w_{t+1}\rangle \le \|g_t\|_* \|w_t - w_{t+1}\|$. Combining the two inequalities yields

$$\frac{1}{\eta}\|w_{t+1} - w_t\|^2 \le \|g_t\|_* \|w_{t+1} - w_t\| .$$

If $w_{t+1} = w_t$, inequality (15) holds immediately. Otherwise, dividing both sides by $\|w_{t+1} - w_t\|$ gives

$$\|w_{t+1} - w_t\| \le \eta \|g_t\|_* .$$

Applying the dual-norm inequality once more gives $\langle g_t, w_t - w_{t+1}\rangle \le \|g_t\|_* \|w_t - w_{t+1}\| \le \eta \|g_t\|_*^2$. □

This lemma makes the role of strong convexity explicit. Because the regularized objective has sufficient curvature, adding one new linear loss cannot move its minimizer too far. In this sense, FTRL is a stabilized version of FTL.

The general FTRL guarantee. Define the range of the regularizer over Ω by

$$B_\psi \triangleq \max_{w\in\Omega} \psi(w) - \min_{w\in\Omega} \psi(w). \tag{18}$$

Theorem 3.4 (FTRL regret bound)**.** *Suppose that ψ is 1-strongly convex with respect to $\|\cdot\|$ and that $\|g_t\|_* \le G$ for every t. Then, for every $u \in \Omega$,*

$$\sum_{t=1}^{T} \langle g_t, w_t - u\rangle \le \frac{B_\psi}{\eta} + \eta G^2 T. \tag{19}$$

Consequently, choosing $\eta = \sqrt{\frac{B_\psi}{G^2 T}}$ gives $\sum_{t=1}^{T} \langle g_t, w_t - u\rangle \le 2G\sqrt{B_\psi T}$.

Proof. Combining Lemma 3.2 and Lemma 3.3 gives

$$\sum_{t=1}^{T} \langle g_t, w_t - u\rangle \le \frac{\psi(u) - \psi(w_1)}{\eta} + \eta \sum_{t=1}^{T} \|g_t\|_*^2.$$

Because w_1 minimizes ψ over Ω, we have $\psi(u) - \psi(w_1) \le B_\psi$. Moreover, $\|g_t\|_* \le G$ implies $\sum_{t=1}^{T} \|g_t\|_*^2 \le G^2 T$. Therefore, $\sum_{t=1}^{T} \langle g_t, w_t - u\rangle \le \frac{B_\psi}{\eta} + \eta G^2 T$. The stated choice of η balances the two terms. □

Finally, recall from (3) that $f_t(w_t) - f_t(u) \le \langle g_t, w_t - u\rangle$. Therefore, the same guarantee also bounds $\mathrm{Reg}_T(u)$. FTRL thus achieves $O(\sqrt{T})$ regret whenever the regularizer has bounded range and the observed subgradients have bounded dual norm.

4 Instances of FTRL

We now return to two familiar algorithms. An ℓ_2 regularizer gives an OGD-type method, while the negative-entropy regularizer gives exactly the Hedge algorithm from the previous lecture.

4.1 FTRL with ℓ_2 regularization and Online Gradient Descent

Choose the quadratic regularizer $\psi(w) \triangleq \frac{1}{2}\|w - w_1\|_2^2$. The FTRL update becomes

$$w_t = \arg\min_{w\in\Omega} \left\{ \langle L_{t-1}, w\rangle + \frac{1}{2\eta}\|w - w_1\|_2^2 \right\}, \qquad w_t = \Pi_\Omega\big(w_1 - \eta L_{t-1}\big). \tag{20}$$

The projection formula follows by completing the square. If we define the unconstrained point

$$u_t \triangleq w_1 - \eta L_{t-1},$$

then (20) can be written as $u_{t+1} = u_t - \eta g_t$, $w_{t+1} = \Pi_\Omega(u_{t+1})$. This method accumulates gradient steps in the unconstrained space and then projects, so it is often called *lazy OGD* (or dual averaging).

The projected OGD algorithm introduced earlier instead uses

$$u_{t+1} = w_t - \eta g_t, \qquad w_{t+1} = \Pi_\Omega(u_{t+1}).$$

Thus the two methods differ only in where the new gradient step starts. Lazy OGD starts from the accumulated unconstrained point u_t, whereas projected OGD starts from the current feasible decision w_t. If $\Omega = \mathbb{R}^d$, projection does nothing and the two updates coincide exactly. With constraints, they may generate different sequences, but both achieve $O(\sqrt{T})$ regret under the boundedness conditions used above.

The general FTRL theorem recovers this rate directly. The quadratic regularizer is 1-strongly convex with respect to $\|\cdot\|_2$, whose dual norm is again $\|\cdot\|_2$. If the diameter of Ω is at most D, then $B_\psi \le D^2/2$. Therefore, when $\|g_t\|_2 \le G$, Theorem 3.4 gives

$$\sum_{t=1}^{T} \langle g_t, w_t - u \rangle \le \frac{D^2}{2\eta} + \eta T G^2 = O(DG\sqrt{T})$$

after choosing the optimal learning rate. Combining this inequality with (3) gives the same order of regret for the original convex losses.

4.2 FTRL with (negative) entropy regularization = Hedge

For the expert problem, the decision set is $\Omega = \Delta(N)$ and the linearized loss vector is $g_t = \ell_t \in [0,1]^N$. If we use Euclidean geometry, then $\|\ell_t\|_2 \le \sqrt{N}$, which leads to an $O(\sqrt{TN})$ regret bound. This is much worse than the $O(\sqrt{T \ln N})$ guarantee of Hedge, so the quadratic regularizer is not using the structure of the simplex effectively.

Following the notation from the previous lecture, write the decision as $p_t \in \Delta(N)$ and choose the negative-entropy regularizer $\psi(p) \triangleq \sum_{i=1}^N p_i \ln p_i$, where $0 \ln 0 \triangleq 0$. The FTRL update is then

$$p_t = \arg\min_{p \in \Delta(N)} \left\{ \sum_{i=1}^{N} p_i L_{t-1,i} + \frac{1}{\eta} \sum_{i=1}^{N} p_i \ln p_i \right\}. \tag{21}$$

Because the cumulative losses are finite, the optimizer has strictly positive coordinates, so the following first-order calculation takes place in the relative interior of the simplex. Introducing a Lagrange multiplier for the constraint $\sum_i p_i = 1$ and solving the first-order conditions gives

$$p_{t,i} = \frac{\exp(-\eta L_{t-1,i})}{\sum_{j=1}^{N} \exp(-\eta L_{t-1,j})}. \tag{22}$$

This is exactly Hedge. In other words, Hedge is not a separate trick: it is FTRL with a regularizer that is well suited to distributions. The entropy term is minimized at the uniform distribution, so the algorithm balances small cumulative loss with a preference against becoming overly concentrated too quickly.

To apply Theorem 3.4, we use two standard facts. First, negative entropy is 1-strongly convex with respect to $\|\cdot\|_1$ on the simplex (with the strong-convexity inequality interpreted on the relative interior and extended to the boundary by continuity), and its range is $B_\psi = \ln N$. Second, the dual norm of $\|\cdot\|_1$ is $\|\cdot\|_\infty$, and $\|\ell_t\|_\infty \le 1$. Therefore, $\mathrm{Reg}_T \le \frac{\ln N}{\eta} + \eta T$. Choosing $\eta = \sqrt{\ln(N)/T}$ gives $\mathrm{Reg}_T \le 2\sqrt{T \ln N}$, recovering the same dependence on T and N that we proved for Hedge in the previous lecture.