CS229M/STATS214: Machine Learning Theory

Lecturer: Tengyu Ma
Scribe: Kevin Guo

Lecture # 16
Mar 10th, 2021

16.1 Review and overview

In the last lecture, we introduced the general framework of *online learning* and *online convex optimization (OCO)*. As a reminder, the OCO problem is as follows: for each time $t = 1, \dots, T$, the following occur:

1. The player chooses an action $w_t \in \Omega$, where Ω is an abstract convex set.
2. The adversary chooses a convex function $f_t : \Omega \mapsto \mathbb{R}$.
3. The player suffers the loss $f_t(w_t)$ and observes the entire loss function $f_t(\cdot)$.

The player's goal is to choose actions $w_1, \dots, w_T$ in order to minimize the *regret* relative to the best action in hindsight:

$$\text{Regret} := \sum_{t=1}^{T} f_t(w_t) - \inf_{w \in \Omega} \sum_{t=1}^{T} f_t(w). \tag{16.1}$$

Meanwhile, we assume that the adversary simply plays a prespecified sequence of functions $f_1, \dots, f_T$, i.e. it does not adapt to the player's choices. This is known as an "oblivious" adversary.

In the last lecture, we introduced the "Follow the Leader" (FTL) algorithm, where the player chooses the action that would have performed best on the previous rounds of the game. We showed a counterexample (the expert problem with two experts) where the FTL algorithm performed poorly, that is, it achieved $O(T)$ regret instead of the $O(\sqrt{T})$ regret we usually expect. In this lecture, we will discuss remedies for the FTL algorithm and apply them to simple examples.

16.2 "Follow the leader" and "Be the leader" strategies

One strategy the player might adopt is *"Follow the Leader" (FTL)*. At time t, the FTL strategy simply chooses the action that would have performed best on the first $t-1$ rounds of the game:

$$w_t = \operatorname*{argmin}_{w \in \Omega} \sum_{i=1}^{t-1} f_i(w). \tag{16.2}$$

This strategy is a natural one as it implements the philosophy "the future will be like the past". Unfortunately, in OCO, we do not assume that the future is like the past. As a result, the FTL strategy does not have optimal regret. One can show that in many examples, its regret scales as $O(T)$, which is not very good. For example, if all the f_t's are bounded, then even the trivial strategy that picks a single w up front and sets $w_t = w$ for all t will achieve $O(T)$ regret.

A better strategy is called *"Be the Leader" (BTL)*. At time t, the BTL strategy chooses the action that would have performed best on $f_1, \cdots, f_{t-1}$ *and* f_t. In other words, the BTL action at time t is w_{t+1}, as defined in (16.2). Note that this is an “illegal” choice for the action because w_{t+1} depends on f_t: in online convex optimization, the action at time t is required to be chosen *before* seeing the function f_t. Nevertheless, we can still gain some useful insights by analyzing this procedure. In particular, the following lemma shows that the BTL strategy is worth emulating because it achieves very good regret.

Lemma 16.1. *The BTL strategy has non-positive regret. That, is, if w_t is defined as in* (16.2), *then*

$$\text{\textit{BTL regret}} = \sum_{t=1}^{T} f_t(w_{t+1}) - \min_{w\in\Omega} \sum_{t=1}^{T} f_t(w) \leq 0, \tag{16.3}$$

for any T and any sequence of functions $f_1, \cdots, f_T$.

Proof. We prove the lemma by induction on T. (16.3) holds trivially for $T = 1$. Suppose that (16.3) holds for all $t \leq T-1$ and any $f_1, \cdots, f_{T-1}$. Now we wish to extend (16.3) to time $t = T$. Let f_T be any function. Since $w_{T+1} = \operatorname{argmin}_w \sum_{t=1}^{T} f_t(w)$, we can write:

$$\begin{aligned}
\sum_{t=1}^{T} f_t(w_{t+1}) - \min_{w\in\Omega} \sum_{t=1}^{T} f_t(w) &= \sum_{t=1}^{T} f_t(w_{t+1}) - \sum_{t=1}^{T} f_t(w_{T+1}) && (16.4)\\
&= \sum_{t=1}^{T-1} f_t(w_{t+1}) - \sum_{t=1}^{T-1} f_t(w_{T+1}) \qquad \text{(final summands cancel)} && (16.5)\\
&\leq \sum_{t=1}^{T-1} f_t(w_{t+1}) - \min_{w\in\Omega} \sum_{t=1}^{T-1} f_t(w) && (16.6)\\
&\leq 0. \qquad \text{(induction hypothesis)} && (16.7)
\end{aligned}$$

□

A useful consequence of this lemma is a regret bound for the FTL strategy.

Lemma 16.2. (FTL regret bound) *Again, let w_t be as in (16.2). The FTL strategy has the regret guarantee*

$$\text{\textit{FTL regret}} = \sum_{t=1}^{T} f_t(w_t) - \min_{w\in\Omega} \sum_{t=1}^{T} f_t(w) \leq \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})]. \tag{16.8}$$

Proof.

$$\begin{aligned}
\text{FTL regret} &= \sum_{t=1}^{T} f_t(w_t) - \min_{w\in\Omega} \sum_{t=1}^{T} f_t(w) && (16.9)\\
&= \sum_{t=1}^{T} f_t(w_{t+1}) - \min_{w\in\Omega} \sum_{t=1}^{T} f_t(w) + \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})] && (16.10)\\
&\leq 0 + \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})], && (16.11)
\end{aligned}$$

where the last inequality is due to (16.3).

□

Lemma 16.2 tells us that if terms $f_t(w_t) - f_t(w_{t+1})$ are small (e.g. w_t does not change much from round to round), then the FTL strategy can have small regret. It suggests that the player should adopt a *stable* policy, i.e. one where the terms $f_t(w_t) - f_t(w_{t+1})$ are small. It turns out that following this intuition will lead to a strategy that improves the regret all the way to $O(\sqrt{T})$ in certain cases.

16.3 "Follow the regularized leader" strategy

Now, we discuss a OCO strategy aims to improve the stability of FTL by controlling the differences $f_t(w_t) - f_t(w_{t+1})$. To describe the method, we will first need a preliminary definition.

Definition 16.3. We say that a differentiable function $\phi : \Omega \mapsto \mathbb{R}$ is *α-strongly-convex* with respect to the norm $||\cdot||$ on Ω if we have

$$\phi(x) \geq \phi(y) + \langle \nabla f(y), x - y \rangle + \frac{\alpha}{2}\|x - y\|^2 \tag{16.12}$$

for any $x, y \in \Omega$.

Remark 16.4. If ϕ is convex, then we know that $f(x)$ has a linear lower bound $\phi(y) + \langle \nabla f(y), x - y \rangle$. Being α-strong-convex means that $f(x)$ has a quadratic lower bound, the RHS of (16.12). This quadratic lower bound is very useful in proving theorems in optimization.

Remark 16.5. If $\nabla^2 f(y) \succeq \alpha I$ for all y, then f is α-strongly-convex. This follows directly from writing the second-order Taylor expansion of f around y.

Given a 1-strongly-convex function $\phi(\cdot)$, which we call a *regularizer*, we can implement the *"Follow the Regularized Leader" (FTRL)* strategy. At time t, this strategy chooses the action

$$w_t = \operatorname*{argmin}_{w \in \Omega} \left[\sum_{i=1}^{t-1} f_i(w) + \frac{1}{\eta}\phi(w) \right], \tag{16.13}$$

where $\eta > 0$ is a tuning parameter that we will tune later.

16.3.1 Regularization and stability

To understand why we might use the FTRL policy, we first establish that it achieves the intended goal of controlling the differences $f_t(w_t) - f_t(w_{t+1})$. Actually, we will show a more general result that adding a regularizer induces stability for any convex objective.

Lemma 16.6. (Regularizers induce stability) *Let F and f be functions taking Ω into $\mathbb{R}$, and assume that F is α-strongly-convex with respect to the norm $\|\cdot\|$ and that f is convex. Let $w = \operatorname{argmin}_{z\in\Omega} F(z)$ and $w' = \operatorname{argmin}_{z\in\Omega}[f(z) + F(z)]$. Then*

$$0 \leq f(w) - f(w') \leq \frac{1}{\alpha}\|\nabla f(w)\|_*^2, \tag{16.14}$$

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

Proof. By strong convexity,

$$F(w') - F(w) \geq \langle \nabla F(w), w' - w \rangle + \frac{\alpha}{2}\|w - w'\|^2 \tag{16.15}$$
$$\geq \frac{\alpha}{2}\|w - w'\|^2, \tag{16.16}$$

where in the second step we used the fact that the KKT optimality conditions for w imply $\langle \nabla F(w), w' - w \rangle \geq 0$. (Informally, if $\Omega = \mathbb{R}^d$, then $\nabla F(w) = 0$ as w minimizes F. If Ω is a convex subset of $\mathbb{R}^d$, then the gradient $\nabla F(w)$ must be perpendicular to the tangent to Ω at w; otherwise, we could move in the direction of the negative gradient and project back to the set Ω to lower the value of F.) Since $F + f$ is also α-strongly convex, exactly the same argument implies:

$$[F(w) + f(w)] - [F(w') + f(w')] \geq \frac{\alpha}{2}\|w - w'\|^2. \tag{16.17}$$

Adding these two inequalities gives

$$f(w) - f(w') \geq \alpha\|w - w'\|^2. \tag{16.18}$$

Since this lower bound is clearly positive, this shows $0 \leq f(w) - f(w')$.

Next, we prove the upper bound on $f(w) - f(w')$. Rearranging the inequality (16.18), we obtain

$$\|w - w'\| \leq \sqrt{\frac{1}{\alpha}[f(w) - f(w')]}. \tag{16.19}$$

Since f is convex, we have $f(w') \geq f(w) + \langle \nabla f(w), w' - w \rangle$. Rearranging this gives

$$\begin{aligned} f(w) - f(w') &\leq \langle \nabla f(w), w - w' \rangle \\ &\leq \|\nabla f(w)\|_* \cdot \|w - w'\| && \text{(by Cauchy-Schwarz)} \\ &\leq \|\nabla f(w)\|_* \sqrt{\frac{1}{\alpha}[f(w) - f(w')]}. && \text{(by (16.19))} \end{aligned}$$

Since $f(w) - f(w') \geq 0$, we can square both sides of this inequality to conclude that

$$[f(w) - f(w')]^2 \leq ||\nabla f(w)||_*^2 \frac{1}{\alpha}[f(w) - f(w')]. \tag{16.20}$$

Dividing both sides of this expression by $f(w) - f(w')$ gives the desired upper bound. □

Remark 16.7. Consider the special case where $\nabla f(w) = 0$. In this situation, w is the minimizer of both F and f, and hence is the minimizer of $F + f$. This implies that $w = w'$, and the inequalities in (16.14) become equalities.

16.3.2 Regret of FTRL

We are now ready to prove a regret bound for the FTRL procedure, based on the idea that strongly convex regularizers induce stability.

Theorem 16.8. (Regret of FTRL) *Let ϕ be a 1-strongly-convex regularizer with respect to the norm $\|\cdot\|$ on Ω. Then the FTRL algorithm (16.13) satisfies the regret guarantee*

$$\textit{FTRL regret} = \sum_{t=1}^{T} f_t(w_t) - \operatorname*{argmin}_{w\in\Omega} \sum_{t=1}^{T} f_t(w) \le \frac{D}{\eta} + \eta \sum_{t=1}^{T} \|\nabla f_t(w_t)\|_*^2, \tag{16.21}$$

where $D = \max_{w\in\Omega} \phi(w) - \min_{w\in\Omega} \phi(w)$.

Remark 16.9. Suppose that for all t and w, we have the uniform bound $||\nabla f_t(w)||_* \le G$. Then Theorem 16.8 implies that the regret is upper bounded by $D/\eta + \eta GT$. Optimizing this upper bound over η by taking $\eta = \sqrt{\dfrac{D}{TG^2}}$ gives the guarantee

$$\text{FTRL regret} \le 2\sqrt{DG} \times \sqrt{T}. \tag{16.22}$$

In other words, optimally-tuned FTRL can achieve $O(\sqrt{T})$ regret in many cases.

Proof. For convenience, define $f_0(w) = \phi(w)/\eta$. Then the FTRL policy can be written as

$$w_t = \operatorname*{argmin}_{w\in\Omega} \sum_{i=0}^{t-1} f_i(w), \tag{16.23}$$

i.e. FTRL is just FTL with an additional "round" of play at time zero. Thus, by Lemma 16.2 with time starting from $t = 0$, we have

$$\sum_{t=0}^{T} f_t(w_t) - \operatorname*{argmin}_{w\in\Omega} \sum_{t=0}^{T} f_t(w) \le \sum_{t=0}^{T} [f_t(w_t) - f_t(w_{t+1})]. \tag{16.24}$$

For any $t \ge 1$, applying Lemma 16.6 with $F(w) = \sum_{i=0}^{t-1} f_i(w)$ (which is $1/\eta$-strongly-convex) and $f(w) = f_t(w)$ gives the bound $f_t(w_t) - f_t(w_{t+1}) \le \eta||\nabla f_t(w_t)||_*^2$. Plugging this into the preceding display gives the upper bound:

$$\sum_{t=0}^{T} f_t(w_t) - \operatorname*{argmin}_{w\in\Omega} \sum_{t=0}^{T} f_t(w) \le f_0(w_0) - f_0(w_1) + \eta \sum_{t=1}^{T} \|\nabla f_t(w_t)\|_*^2. \tag{16.25}$$

Next, we need to relate the LHS of the above display (which starts at time $t = 0$) to the actual regret of FTRL (which starts at time $t = 1$). To do this, define $w^* = \operatorname{argmin}_{w\in\Omega} \sum_{t=1}^{T} f_t(w)$. Then,

$$\begin{aligned}
\sum_{t=0}^{T} f_t(w_t) - \operatorname*{argmin}_{w\in\Omega} \sum_{t=0}^{T} f_t(w) &\ge \sum_{t=0}^{T} f_t(w_t) - \sum_{t=0}^{T} f_t(w^*) &(16.26)\\
&= f_0(w_0) - f_0(w^*) + \underbrace{\left(\sum_{t=1}^{T} f_t(w_t) - \operatorname*{argmin}_{w\in\Omega} \sum_{t=1}^{T} f_t(w) \right)}_{\text{Regret of FTRL}}. &(16.27)
\end{aligned}$$

Combining this inequality with (16.25) gives

$$\text{Regret of FTRL} \leq f_0(w_0) - f_0(w_1) + f_0(w^*) - f_0(w_0) + \eta \sum_{t=1}^{T} \|\nabla f_t(w_t)\|_*^2 \tag{16.28}$$

$$= \frac{\phi(w^*) - \phi(w_1)}{\eta} + \eta \sum_{t=1}^{T} \|\nabla f_t(w_t)\|_*^2 \tag{16.29}$$

$$\leq \frac{D}{\eta} + \eta \sum_{t=1}^{T} \|\nabla f_t(w_t)\|_*^2. \tag{16.30}$$

This concludes the proof of the theorem. □

16.3.3 Applying FTRL to online linear regression

We apply the FTRL algorithm to a concrete machine learning problem. Let $\Omega = \{\omega : \|w\|_2 \leq 1\}$, and let $f_t(\omega) = \frac{1}{2}(y_t - \omega^\top x_t)^2$ for some observation pair (x_t, y_t) satisfying $\|x_t\|_2 \leq 1$ and $|y_t| \leq 1$. This corresponds to a problem where we are trying to make accurate predictions using a linear model, but we do not assume any structure on the observation sequence (x_t, y_t) beyond boundedness.

Consider using FTRL in this problem with a ridge regularizer, $\phi(\omega) = \frac{1}{2}\|w\|_2^2$. One can check that ϕ is 1-strongly-convex with respect to the ℓ_2-norm, and also that $D = \max_{\omega \in \Omega} \phi(\omega) - \min_{\omega \in \Omega} \phi(\omega) = \frac{1}{2}$. Moreover, for all t and w we have

$$\nabla f_t(w) = -(y_t - w^\top x_t) x_t, \tag{16.31}$$

$$\|\nabla f_t(w)\|_2 \leq |y_t - w^\top x_t| \cdot \|x_t\|_2 \tag{16.32}$$

$$\leq 2 \cdot 1 = 2. \tag{16.33}$$

Therefore, by choosing $\eta = \sqrt{1/(8T)}$ and applying the FTRL regret theorem (Theorem 16.8), we can obtain the regret guarantee

$$\sum_{t=1}^{T} (y_t - w_t^\top x_t)^2 - \min_{\|w\|_2 \leq 1} \sum_{t=1}^{T} (y_t - w^\top x_t)^2 \leq 4\sqrt{T}. \tag{16.34}$$

16.3.4 Applying FTRL to the expert problem

For the expert problem, recall that the action space is $\Delta(N)$ and $f_t = \langle \ell_t, p \rangle$, where $\ell_t \in [0,1]^N$. As a first attempt at applying FTRL to this problem, we set $\phi(p) = \frac{1}{2}\|p\|_2^2$. With this choice,

$$D = \max_{p \in \Delta(N)} \phi(p) - \min_{p \in \Delta(N)} \phi(p) \tag{16.35}$$

$$\leq \max_{p \in \Delta(N)} \frac{1}{2}\|p\|_2^2 \tag{16.36}$$

$$\leq \max_{p \in \Delta(N)} \frac{1}{2}\|p\|_1^2 \tag{16.37}$$

$$= \frac{1}{2}. \tag{16.38}$$

Also,

$$\|\nabla f_t\|_2 = \|\ell_t\|_2 \leq \sqrt{N}. \tag{16.39}$$

Thus, the regret bound is $O(G\sqrt{DT}) = O(\sqrt{NT})$. This is optimal dependency on T, but not good dependency on N. In the next lecture, we will see that we can obtain $\log N$ dependency by using a different regularizer.