Lecture 3

Instructor: Haipeng Luo

1 Instances of FTRL

In the last lecture we study FTRL, a general online learning algorithm, and prove the following regret bound

$$\mathcal{R}_T \le \frac{D}{\eta} + \eta \sum_{t=1}^{T} \|\nabla f_t(w_t)\|_\star^2 ,$$

where $D = \max_{w\in\Omega} \psi(w) - \min_{w\in\Omega} \psi(w)$. Here we present two concrete instances of FTRL.

1.1 Recovering Hedge

Recall that in the expert problem $\Omega = \Delta(N)$ and $f_t(p) = \langle p, \ell_t \rangle$. If we pick the (negative) entropy function as the regularizer, that is, $\psi(p) = \sum_{i=1}^N p(i) \ln p(i)$, one can verify that the solution of

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

is exactly the Hedge algorithm, that is, $p_t(i) \propto \exp(-\eta \sum_{\tau=1}^{t-1} \ell_t(i))$. In other words, Hedge is just one special case of FTRL.

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 in this case the definition of strong convexity

$$\psi(p) - \psi(q) \le \langle \nabla\psi(p), p - q \rangle - \tfrac{1}{2} \|p - q\|_1^2$$

is equivalent to

$$\frac{1}{2} \|p - q\|_1^2 \le \sum_{i=1}^{N} q(i) \ln \frac{q(i)}{p(i)} \stackrel{\text{def}}{=} \mathrm{KL}(q, p).$$

The latter turns out to be exactly the Pinsker's inequality, which states that the Kullback-Leibler divergence of two distributions is lower bounded by half of their L_1 distance square.

Also notice that the dual norm of the L_1 norm is the L_∞ norm and by boundedness of losses we have $\|\nabla f_t(p_t)\|_\infty = \|\ell_t\|_\infty \le 1$. Moreover, the (negative) entropy function has maximal value 0 (when the distribution concentrates on one coordinate) and minimum value $-\ln N$ (when the distribution is uniform), and thus $D = \ln N$. Therefore, applying the FTRL regret bound we again arrive at

$$\mathcal{R}_T \le \frac{\ln N}{\eta} + T\eta,$$

the same bound we proved last time using a different potential-based argument.

1.2 Online Gradient Descent

In the next example we consider an arbitrary OCO problem and pick $\psi(w) = \frac{1}{2} \|w\|_2^2$. The FTRL algorithm becomes

$$w_t = \operatorname*{argmin}_{w\in\Omega} \sum_{\tau=1}^{t-1} f_\tau(w) + \frac{1}{2\eta} \|w\|_2^2 .$$

One can (approximately) solve this convex optimization problem using standard methods. However, it turns out that it is without loss of generality to assume that f_t is a linear function. To see this, note that by convexity the regret can be bounded as

$$\mathcal{R}_T = \max_{w\in\Omega} \sum_{t=1}^{T} \left(f_t(w_t) - f_t(w)\right) \leq \max_{w\in\Omega} \sum_{t=1}^{T} \left\langle \nabla f_t(w_t), w_t - w \right\rangle .$$

Therefore, we can imagine that the loss function is actually a linear function $f'_t(w) = \langle f_t(w_t), w\rangle$, and a regret bound for this linear problem is clearly also a regret bound for the original problem. With this reduction, we rewrite the above FTRL as

$$w_t = \operatorname*{argmin}_{w\in\Omega} \left\langle w, \sum_{\tau=1}^{t-1} \nabla f_\tau(w_\tau) \right\rangle + \frac{1}{2\eta} \|w\|_2^2 . = \operatorname*{argmin}_{w\in\Omega} \left\| w + \eta \sum_{\tau=1}^{t-1} \nabla f_\tau(w_\tau) \right\|_2^2 ,$$

which means w_t is the L_2 projection of $u_t = -\eta \sum_{\tau=1}^{t-1} \nabla f_\tau(w_\tau)$ onto Ω. This algorithm is called *Online Gradient Descent* (OGD) [Zinkevich, 2003]. To see the connection to the regular gradient descent, note that OGD can be equivalently written as

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

while regular gradient descent would instead do

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

In fact, there is little real difference between these two algorithms and one can prove the same guarantee for both of them. Below we apply the general FTRL guarantee to prove a regret bound.

Indeed, one can easily verify that $\psi(w) = \frac{1}{2}\|w\|_2^2$ is strongly convex with respect to the L_2 norm. Note that the dual norm of the L_2 norm is itself. So if we let G be an upper bound on all the gradients, that is, $\|\nabla f_t(w_t)\|_2 \leq G$, then the regret of OGD is bounded by

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

where the last step is by picking the optimal η.

Examples Consider the online regression problem where $\Omega = \{w \in \mathbb{R}^d : \|w\|_2 \leq 1\}$ 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 $x_t \in \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$ 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 $\max_{w\in\Omega} \|w\|_2 = 1$, 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. Note that for the simplex, $\max_{p\in\Delta(N)} \|p\|_2 \leq \max_{p\in\Delta(N)} \|p\|_1 = 1$, but $\|\ell_t\|_2 \leq \sqrt{N}$. Thus OGD's regret is $\mathcal{O}(\sqrt{TN})$ in this case, which has an exponentially worse dependence on N compared to Hedge.

2 Follow the Perturbed Leader and Combinatorial Problems

As we have seen, FTRL uses regularization to stabilize the algorithm. Here, we introduce another very different approach, *perturbation*. To motivate this approach, we consider the following online combinatorial problems.

Let $S = \{v_1, \dots, v_M\}$ be a set of combinatorial actions such that $v_j \in \{0,1\}^N$ and $\|v_j\|_1 \leq m$ for some integer $m \leq N$ and all $j \in [M]$. The decision space for the learner is the convex hull of S, that is, $\Omega = \left\{ \sum_{j=1}^{M} p(j) v_j : p \in \Delta(M) \right\}$. Thus, each point in Ω specifies a distribution over these combinatorial actions or in other words a randomized strategy. We consider linear loss functions so that $f_t(w) = \langle w, \ell_t\rangle$ for some $\ell_t \in [0,1]^N$. Finally, for simplicity we restrict our attention to oblivious environments so that $\ell_1, \dots, \ell_T$ are decided before the game starts.

The expert problem is clearly a special case where S consists of all the standard basis vectors in $\mathbb{R}^N$ and $m = 1$. Another example is when $S = \{v \in \{0,1\}^N : \|v\|_1 = m\}$ so that $\Omega = \{w \in [0,1]^N : \|w\|_1 = m\}$ (recall the multiple-product recommendation example in Lecture 1).

Yet another important example is the online shortest path problem. 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, then the loss (e.g. delay) for each edge is revealed and the player suffers the total loss of all the edges on the picked path. This can be formulated as a special case of the above combinatorial problem by setting S 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). Note that m is the length of the longest path in S.

One can again use the FTRL approach to tackle this problem, but it is not often clear how to solve the optimization problem in FTRL. Instead, we consider a different approach called *Follow the Perturbed Leader* (FTPL) [Kalai and Vempala, 2005], which only requires a linear optimization step over Ω. Specifically, let ℓ_0 be a uniformly random draw from $[0, 1/\eta]^N$ for some $\eta > 0$. Then at round t FTPL plays

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

In other words, w_t is the leader according the cumulative losses plus some perturbation ℓ_0. Note that this leader can always be some point in S due to linearity. Moreover, for many problems this linear optimization admits efficient algorithm. For example, for the expert problem this is trivially solved by picking the best coordinate. For the online shortest path problem, this can be solved by a shortest path algorithm such as Dijkstra's algorithm.

It remains to prove the regret bound of FTPL. To this end, we first show that perturbation provides stability in expectation.

Lemma 1 (Stability of FTPL)**.** *FTPL with parameter η ensures that*

$$\mathbb{E}[f_t(w_t) - f_t(w_{t+1})] \le mN\eta$$

where the expectation is with respect to the random draw of ℓ_0.

Proof. To make the dependence explicit, define $h_t(\ell_0) = \left\langle \operatorname{argmin}_{w\in\Omega} \sum_{\tau=0}^{t-1} \ell_\tau, \ell_t \right\rangle$. We then have

$$\begin{aligned}
\mathbb{E}[f_t(w_t) - f_t(w_{t+1})] &= \mathbb{E}[h_t(\ell_0) - h_t(\ell_0 + \ell_t)] \\
&= \eta^N \int_{\ell_0 \in [0,1/\eta]^N} h_t(\ell_0) - h_t(\ell_0 + \ell_t) d\ell_0 \\
&= \eta^N \left(\int_{\ell_0 \in [0,1/\eta]^N} h_t(\ell_0) d\ell_0 - \int_{\ell_0 \in \ell_t + [0,1/\eta]^N} h_t(\ell_0) d\ell_0 \right) \\
&\qquad \text{(change of variable)} \\
&\le \eta^N \int_{\ell_0 \in [0,1/\eta]^N \setminus \ell_t + [0,1/\eta]^N} h_t(\ell_0) d\ell_0 \\
&\le m\eta^N \int_{\ell_0 \in [0,1/\eta]^N \setminus \ell_t + [0,1/\eta]^N} d\ell_0 \qquad (h_t(\ell_0) \le m) \\
&= m \Pr\left(\exists i : \ell_0(i) \le \ell_t(i)\right) \\
&\le m \sum_{i=1}^N \Pr\left(\ell_0(i) \le 1\right) \qquad \text{(union bound and } \ell_t(i) \le 1\text{)} \\
&= mN\eta.
\end{aligned}$$

□

With the stability lemma, we can prove the following bound using similar argument as in FTRL.

Theorem 1. *FTPL with parameter η ensures that*

$$\mathbb{E}[\mathcal{R}_T] \leq \frac{m}{2\eta} + mNT\eta$$

where the expectation is with respect to the random draw of ℓ_0. With η optimally set to $\sqrt{1/(2NT)}$ we thus have $\mathbb{E}[\mathcal{R}_T] = \mathcal{O}(m\sqrt{TN})$.

Proof. Note that by the BTL lemma, with $w^\star = \operatorname{argmin}_{w\in\Omega} \sum_{t=1}^T f_t(w)$, we again have

$$\begin{aligned}
\mathcal{R}_T &= \sum_{t=1}^T f_t(w_t) - \sum_{t=1}^T f_t(w^\star) \\
&\leq \sum_{t=1}^T f_t(w_t) - \sum_{t=0}^T f_t(w_{t+1}) + f_0(w^\star) \\
&= f_0(w^\star) - f_0(w_1) + \sum_{t=1}^T (f_t(w_t) - f_t(w_{t+1})).
\end{aligned}$$

Applying the stability lemma and realizing $\mathbb{E}[f_0(w^\star) - f_0(w_1)] \leq \mathbb{E}[\langle \ell_0, w^\star\rangle] = \langle \mathbb{E}[\ell_0], w^\star\rangle \leq \frac{m}{2\eta}$ finish the proof. □

Note that the bound is suboptimal in general. For example, in the expert problem it has polynomial dependence on N instead of logarithmic dependence. However, with a more sophisticated noise distribution (rather than uniform), the regret bounds can often be improved to be optimal. In fact, it is well-known that in the expert problem, FTPL with Gumbel noise is *equivalent* to Hedge!

A final remark is that to deal with non-oblivious environments, it turns out that one only needs to draw a fresh sample of ℓ_0 at the beginning of each round. The intuition is that this will prevent a non-oblivious environment, whose strategy can depend on the player's actions, from figuring out ℓ_0 gradually. A formal proof can be found in [Hutter and Poland, 2005].

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.