CSE 522: Algorithms and Uncertainty Spring 2017

Lecture 12 — May 11, 2017

Lecturer: Nikhil R. Devanur

1 Bandit Algorithms

We will consider a general reduction from an algorithm in the full information setting to a bandit one. Let ALG_1 be a given algorirthm for Online linear optimization. We will design a generic algorithm ALG_2 using ALG_1 as a blackbox. Let $\mathcal{H}_t$ denote all random variables that are realized up to and including round t. We now define ALG_2, where the generic steps are steps 2 and 4.

For $t = 1..T$ do

1. Let $\mathbf{w}_t$ be the prediction of ALG_1 in round t.
2. Pick a randomized action $\mathbf{a}_t$.
3. Observe $\ell_t(\mathbf{a}_t) = \ell_t \cdot \mathbf{a}_t$.
4. Pick a random variable $\tilde{\ell}_t$ such that $\mathbb{E}[\tilde{\ell}_t|\mathcal{H}_{t-1}] = \ell_t$.
5. Simulate ALG_1 with $\tilde{\ell}_t$.

Theorem 1. *Let* $\text{REGRET}(\text{ALG})(\ell_1, \ell_2, \ldots, \ell_t)$ *be the regret of an algorithm* ALG *on a given sequence of loss vectors.*

$$\mathbb{E}[\text{REGRET}(\text{ALG}_2)(\ell_1, \ell_2, \ldots, \ell_t)] \leq \mathbb{E}[\text{REGRET}(\text{ALG}_1)(\tilde{\ell_1}, \tilde{\ell_2}, \ldots, \tilde{\ell_T}) + \sum_{t=1}^{T} \ell_t \cdot (\mathbf{a}_t - \mathbf{w}_t)].$$

Proof.

$$\begin{aligned}
\mathbb{E}[\sum_{t=1}^{T} \ell_t \cdot \mathbf{a}_t] &= \mathbb{E}[\sum_{t=1}^{T} \ell_t \cdot \mathbf{w}_t + \ell_t \cdot (\mathbf{a}_t - \mathbf{w}_t)] \\
&= \mathbb{E}[\sum_{t=1}^{T} \tilde{\ell}_t \cdot \mathbf{w}_t + \ell_t \cdot (\mathbf{a}_t - \mathbf{w}_t)] \qquad (1) \\
&= \mathbb{E}[\text{REGRET}(\text{ALG}_1)(\tilde{\ell_1}, \tilde{\ell_2}, \ldots, \tilde{\ell_T}) + \min_{\mathbf{u} \in S}\{\sum_{t=1}^{T} \tilde{\ell}_t \cdot \mathbf{u}\} + \sum_{t=1}^{T} \ell_t(\mathbf{a}_t - \mathbf{w}_t)] \\
&\leq \mathbb{E}[\text{REGRET}(\text{ALG}_1)(\tilde{\ell_1}, \tilde{\ell_2}, \ldots, \tilde{\ell_T}) + \sum_{t=1}^{T} \ell_t(\mathbf{a}_t - \mathbf{w}_t)] + \min_{\mathbf{u} \in S}\{\mathbb{E}[\sum_{t=1}^{T} \tilde{\ell}_t \cdot \mathbf{u}]\} \qquad (2) \\
&= \mathbb{E}[\text{REGRET}(\text{ALG}_1)(\tilde{\ell_1}, \tilde{\ell_2}, \ldots, \tilde{\ell_T}) + \sum_{t=1}^{T} \ell_t(\mathbf{a}_t - \mathbf{w}_t)] + \min_{\mathbf{u} \in S}\{\sum_{t=1}^{T} \ell_t \cdot \mathbf{u}\}
\end{aligned}$$

In the above sequence of inequalities, the unnumbered lines are straightforward. Inequality (2) holds because expectation of min is smaller than min of expectation. Equality (1) is a little tricky; here's the explanation. Note that $\mathbf{w}_t$ is completely determined by $\mathcal{H}_{t-1}$. Therefore, conditioned on $\mathcal{H}_{t-1}$, it is deterministic. This gives us that

$$\mathbb{E}[\tilde{\ell_t} \cdot \mathbf{w}_t | \mathcal{H}_{t-1}] = \ell_t \cdot \mathbf{w}_t,$$

where the expectation is over the randomness in round t. Now we take the expectation over $\mathcal{H}_{t-1}$ on both sides to get that

$$\mathbb{E}[\tilde{\ell_t} \cdot \mathbf{w}_t] = \mathbb{E}_{\mathcal{H}_{t-1}}[\mathbb{E}[\tilde{\ell_t} \cdot \mathbf{w}_t | \mathcal{H}_{t-1}]] = \mathbb{E}[\ell_t \cdot \mathbf{w}_t].$$

From the final inequality in this sequence, we have that

$$\begin{aligned}\mathbb{E}[\textsc{regret}(\textsc{alg}_2)(\ell_1, \ell_2, \ldots, \ell_t)] &= \mathbb{E}[\sum_{t=1}^{T} \ell_t \cdot \mathbf{a}_t] - \min_{\mathbf{u} \in S}\{\sum_{t=1}^{T} \ell_t \cdot \mathbf{u}\} \\ &\leq \mathbb{E}[\textsc{regret}(\textsc{alg}_1)(\tilde{\ell_1}, \tilde{\ell_2}, \ldots, \tilde{\ell_T}) + \ell_t(\mathbf{a}_t - \mathbf{w}_t)].\end{aligned}$$

□

1.1 Algorithm for the multi-armed bandit (MAB) problem

We will now apply the framework to derive an algorithm for the multi-armed bandit problem. We will use the multiplicative weight update algorithm, or its equivalent, FTRL with the entropic regularizer, as our $\textsc{alg}_1$. In order to define $\textsc{alg}_2$, we need to specify what to do in Steps 2 and 4. Step 2 is actually quite natural: recall that S is the unit simplex, therefore $\mathbf{w}_t$ is a probability distribution. Step 2 simply picks one of the corners of the simplex from this probability distribution. Step 4 is the tricky one. Notice that for any fixed a, we have that

$$\mathbb{E}_{\mathbf{a}_t \sim \mathbf{w}_t}[\mathbf{1}(\mathbf{a} = \mathbf{a}_t)\ell_t(\mathbf{a}_t)] = \mathbf{w}_t(\mathbf{a})\ell_t(\mathbf{a}).$$

$$\Leftrightarrow \mathbb{E}_{\mathbf{a}_t \sim \mathbf{w}_t}[\mathbf{1}(\mathbf{a} = \mathbf{a}_t)\ell_t(\mathbf{a}_t)/\mathbf{w}_t(\mathbf{a}_t)] = \ell_t(\mathbf{a}).$$

Recall that we represent each action $\mathbf{a}$ with its indicator vector, i.e., it has a 1 for the coordinate corresponding to $\mathbf{a}$ and 0 everywhere else. By abuse of notation, we will use $\mathbf{a}$ to denote this vector too. Then let

$$\tilde{\ell_t} := \mathbf{a}_t \ell_t(\mathbf{a}_t)/\mathbf{w}_t(\mathbf{a}_t) \text{ and notice that } \mathbb{E}_{\mathbf{a}_t \sim \mathbf{w}_t}[\tilde{\ell_t}] = \ell_t.$$

This defines Step 4 of the algorithm.

Let's try to apply Theorem 1 with the usual regret bound for the experts problem. One quantity we need to bound is the maximum loss. However, note that the coordinates of $\tilde{\ell_t}$ are unbounded, since $\mathbf{w}_t(\mathbf{a}_t)$ might be arbitrarily small.

We therefore define the following alternate algorithm, which fixes the problem of small $\mathbf{w}_t(\mathbf{a}_t)$ by making sure it is always above a certain threshold. Let $\mathbf{1}$ define the all ones vector, and let $\gamma \in [0, 1]$ be a parameter. Define a probability distribution $\mathbf{w}'_t$ as follows.

$$\mathbf{w}'_t := (1 - \gamma)\mathbf{w}_t + \gamma \mathbf{1}/n.$$

In step 2, the algorithm samples an arm $\mathbf{a}_t$ from the probability distribution $\mathbf{w}'_t$. Step 4 is similar to what we did before:

$$\tilde{\ell}_t := \mathbf{a}_t \ell_t(\mathbf{a}_t)/\mathbf{w}'_t(\mathbf{a}_t) \text{ and notice that } \mathbb{E}_{\mathbf{a}_t \sim \mathbf{w}'_t}[\tilde{\ell}_t] = \ell_t.$$

Since $\mathbf{w}'_t(\mathbf{a}) \geq \gamma/n$ for all $\mathbf{a}$, we now have that

$$\|\tilde{\ell}_t\|_\infty \leq \frac{n}{\gamma}.$$

However, in fixing this, we have introduced an additional error since now the following term is no longer zero.

$$\mathbb{E}_{\mathbf{a}_t \sim \mathbf{w}'_t}[\mathbf{a}_t - \mathbf{w}_t] = \mathbf{w}'_t - \mathbf{w}_t = \gamma(\mathbf{1}/n - \mathbf{w}_t).$$

From this we can bound the extra error term as

$$\mathbb{E}[\ell_t \cdot (\mathbf{a}_t - \mathbf{w}_t)] = \ell_t \cdot (\mathbf{1}/n - \mathbf{w}_t)\gamma \leq \gamma. \tag{3}$$

Now recall that the regret of the FTRL algorithm when the domain S is the unit simplex, and R is the entropic regularizer, is bounded above by

$$2G\sqrt{T \log n},$$

where G is an upper bound on the ℓ_∞ norms of the loss vectors. As argued above $G = \frac{n}{\gamma}$. Using this and (3) in Theorem 1, we get that a regret bound of

$$2\frac{n}{\gamma}\sqrt{T \log n} + \gamma T.$$

This is minimized at

$$\gamma = \sqrt{2n}\sqrt[4]{\frac{\log n}{T}},$$

giving a regret bound of

$$2\sqrt{2n}\sqrt[4]{\log n}T^{3/4}.$$

This bound is suboptimal. We will next show a more refined bound for the experts problem that will let us analyze the original algorithm. For this we go back to FTRL.

1.2 Back to FTRL

Recall the FTRL algorithm:

Definition 1. *The **Follow the Regularized Leader (FTRL)** strategy is to set*

$$\mathbf{w}_{t+1} = \arg\min_{\mathbf{w} \in S} \left[\eta \ell_{1..t} \cdot \mathbf{w} + R(\mathbf{w})\right],$$

*where $R(\cdot)$ is a **strongly convex** regularizer (see definition below).*

Strong convexity

Definition 2. *Let $f : S \to \mathbb{R}$ be a convex function, where S is a convex set. Then $f(\cdot)$ is β-**strongly convex** w.r.t. norm $||\cdot||$ if for all $\mathbf{u}, \mathbf{v} \in S$*

$$f(\mathbf{v}) \geq f(\mathbf{u}) + \nabla f(\mathbf{u}) \cdot (\mathbf{v} - \mathbf{u}) + \frac{\beta}{2}||\mathbf{u} - \mathbf{v}||^2.$$

In other words, the function grows faster than linearly in every direction.

Definition 3. *Suppose that R is a convex function, Then the **Bregman divergence** $B_f(\mathbf{x}||\mathbf{y})$ is defined as follows*

$$B_f(\mathbf{x}||\mathbf{y}) = f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})$$

The Bregman divergence from $\mathbf{x}$ to $\mathbf{y}$ is the difference between $f(\mathbf{x})$ and its linear approximation via the first-order Taylor expansion of f around $\mathbf{y}$. Since the function is convex, this is always nonnegative. It is not symmetric, but behaves in many ways like a distance function.

Definition 4. *Let $f : S \to \mathbb{R}$ be a convex function, where S is a convex set. Then $f(\cdot)$ is α-**smooth** w.r.t. norm $||\cdot||$ if for all $\mathbf{u}, \mathbf{v} \in S$*

$$f(\mathbf{v}) \leq f(\mathbf{u}) + \nabla f(\mathbf{u}) \cdot (\mathbf{v} - \mathbf{u}) + \frac{\alpha}{2}||\mathbf{u} - \mathbf{v}||^2.$$

Remark 1. *We can rewrite the definition of β-strongly convex as follows:*

$$B_f(\mathbf{u}||\mathbf{v}) \geq \frac{\beta}{2}||\mathbf{u} - \mathbf{v}||^2. \tag{4}$$

Similarly we can rewrite the definition of α-smooth as follows:

$$B_f(\mathbf{u}||\mathbf{v}) \leq \frac{\alpha}{2}||\mathbf{u} - \mathbf{v}||^2. \tag{5}$$

Definition 5 (Fenchel conjugate)**.** *For function f with domain S, the Fenchel conjugate of f is defined as*

$$f^*(\mathbf{y}) = \max_{\mathbf{w} \in S} \{\mathbf{y} \cdot \mathbf{w} - f(\mathbf{w})\}.$$

For simplicity, assume that the domain of R is S. Then, using the definition of the Fenchel conjugate, one can see the connection between FTRL and R^* as follows.

$$R^*(-\eta \ell_{1..t}) = \max_{\mathbf{w} \in S} (-\eta \ell_{1..t} \cdot \mathbf{w} - R(\mathbf{w})) = \min_{\mathbf{w} \in S} (\eta \ell_{1..t} \cdot \mathbf{w} + R(\mathbf{w})).$$

Recall that for convex functions, arg max is the gradient. Therefore we have that

$$\mathbf{w}_{t+1} = \nabla R^*(-\eta \ell_{1..t}).$$

Lemma 1 (Strong/Smooth Duality)**.** *Let R be a closed, convex function. R is β-strongly convex w.r.t $\|\cdot\|$ if and only if R^* is $1/\beta$-strongly smooth w.r.t. $\|\cdot\|_*$.*

We now show the following regret bound for FTRL.

Lemma 2.

$$\forall\ u \in S, \eta \sum_{t=1}^{T} (\mathbf{w}_t - u) \cdot \ell_t \leq R(\mathbf{u}) - R(\mathbf{w}_1) + \sum_{t=1}^{T} B_{R^*}(-\eta \ell_{1..t} \| - \eta \ell_{1..t-1}).$$

Proof. For simplicity of notation, we give the proof assuming $\eta = 1$. The steps are identical for general η.

From the definition of Fenchel conjugate we have that

$$-u \cdot \ell_{1..T} \leq R(u) + R^*(-\ell_{1..T}).$$

We write $R^*(-\ell_{1..T})$ as a telescopic sum:

$$R^*(-\ell_{1..T}) = \sum_{t=1}^{T} R^*(-\ell_{1..t}) - R^*(-\ell_{1..t-1}) + R^*(0).$$

The final term is simplified as

$$R^*(0) = \max_{\mathbf{w} \in S} \{0 \cdot \mathbf{w} - R(\mathbf{w})\} = -\min_{\mathbf{w} \in S} \{R(\mathbf{w})\} = -R(\mathbf{w}_1).$$

Adding these equalities and inequalities, we get that

$$\sum_{t=1}^{T} (\mathbf{w}_t - u) \cdot \ell_t \leq R(\mathbf{u}) - R(\mathbf{w}_1) + \sum_{t=1}^{T} \mathbf{w}_t \cdot \ell_t + R^*(-\ell_{1..t}) - R^*(-\ell_{1..t-1}).$$

Recall that $\mathbf{w}_t = \nabla R^*(-\ell_{t-1})$ note that $\ell_{1..t} = \ell_{1..t-1} + \ell_t$, and plug in the definition of Bregman divergence to get

$$\begin{aligned} B_{R^*}(-\ell_{1..t} \| - \ell_{1..t-1}) &= R^*(-\ell_{1..t}) - R^*(-\ell_{1..t-1}) - \nabla R^*(-\ell_{1..t-1}) \cdot (-\ell_t) \\ &= R^*(-\ell_{1..t}) - R^*(-\ell_{1..t-1}) + \mathbf{w}_t \cdot \ell_t. \end{aligned}$$

This completes the proof. □

Using this, we get the following corollary.

Corollary 1. *The regret of FTRL with entropic regularizer, with parameter η, is bounded by*

$$\frac{\log n}{\eta} + \eta \sum_{t=1}^{T} \sum_{\mathbf{a} \in \mathbf{A}} \mathbf{w}_t(\mathbf{a}) \ell_t(\mathbf{a})^2.$$

1.3 Near optimal algorithm for MAB

We now show how to use this more refined regret bound for FTRL to analyze the first algorithm proposed in this lecture. Note that from Theorem 1 and Corollary 1, we essentially need to bound

$$\begin{aligned} \mathbb{E}_{\mathbf{a}_t \sim \mathbf{w}_t}[\sum_{\mathbf{a} \in \mathbf{A}} \mathbf{w}_t(\mathbf{a}) \tilde{\ell}_t(\mathbf{a})^2] &= \sum_{\mathbf{a} \in \mathbf{A}} \mathbf{w}_t(\mathbf{a})^2 \frac{\ell_t(\mathbf{a})^2}{\mathbf{w}_t(\mathbf{a})^2} \\ &= \sum_{\mathbf{a} \in \mathbf{A}} \ell_t(\mathbf{a})^2 \\ &\leq n. \end{aligned}$$

Now picking $\eta = \sqrt{\log n/nT}$, we get the following lemma.

Lemma 3. *The MAB algorithm has regret*

$$O(\sqrt{nT \log n}).$$