# 6.883: Online Methods in Machine Learning

Alexander Rakhlin

## LECTURE 19 AND 20

## 1. ONLINE LINEAR OPTIMIZATION, CONTINUED

### 1.1 Recap: $B_1/B_\infty$ setting

Recall that we are considering the online linear optimization problem in $\mathbb{R}^N$

For $t = 1, \dots, n$
  Predict $\widehat{y}_t \in B_1$
  Observe costs $z_t \in B_\infty$

with regret defined as the difference

$$\sum_{t=1}^{n} \langle \widehat{y}_t, z_t \rangle - \min_{v \in B_1} \sum_{t=1}^{n} \langle v, z_t \rangle . \tag{1}$$

### 1.2 Recap: the relaxation approach

Let us recall the relaxation approach that we've been using in this course. To guarantee that (1) can be upper bounded for all sequences, we define a function **Rel** such that

$$-\min_{v \in B_1} \sum_{t=1}^{n} \langle v, z_t \rangle \le \mathbf{Rel}(z_{1:n}) \tag{2}$$

and such that

$$\inf_{\widehat{y}_t} \max_{z_t} \{ \langle \widehat{y}_t, z_t \rangle + \mathbf{Rel}(z_{1:t}) \} \le \mathbf{Rel}(z_{1:t-1}). \tag{3}$$

Any relaxation that satisfies (2) and (3) is called ***admissible***. A subtlety: a particular strategy $\widehat{y}'_t$ might not be the optimal solution in (3), yet it might still guarantee that

$$\max_{z_t} \left\{ \left\langle \widehat{y}'_t, z_t \right\rangle + \mathbf{Rel}(z_{1:t}) \right\} \le \mathbf{Rel}(z_{1:t-1}). \tag{4}$$

We will call such a choice an *admissible strategy for the given relaxation*. Indeed, in many cases, we might choose a relaxation but not be able to solve it exactly. Such is the case with random playout.

If **Rel** is an admissible relaxation, then we have that

$$\sum_{t=1}^{n} \langle \widehat{y}_t, z_t \rangle - \min_{v \in B_1} \sum_{t=1}^{n} \langle v, z_t \rangle \le \mathbf{Rel}(\varnothing), \tag{5}$$

and so the choice of **Rel** implies both an optimization problem to solve and the final regret bound.

### 1.3 A Rel for linear optimization

We derived the relaxation

$$\mathbf{Rel}(z_{1:t}) = \mathbb{E}_{u_{t+1:n}\sim D}\left\|\sum_{j=1}^{t} z_j + 6\sum_{s=t+1}^{n} u_s\right\|_\infty \tag{6}$$

where $D$ is uniform on $\{\pm 1\}^n$. This relaxation guarantees that

$$\forall z_1,\dots,z_n,\quad \sum_{t=1}^{n}\langle \widehat{y}_t, z_t\rangle \le \min_{v\in B_1}\sum_{t=1}^{n}\langle v, z_t\rangle + 6\sqrt{2n\log N}. \tag{7}$$

That is, for any sequence (which could even be adaptively chosen by Nature) the average cost is no more than the cost of the best fixed decision, plus an $O(\sqrt{(\log N)/n})$ term.

### 1.4 Algorithm

Let us take the relaxation (6) and develop an algorithm. Recall that we need to solve

$$\min_{\widehat{y}_t\in B_1}\max_{z_t\in B_\infty}\left\{\langle \widehat{y}_t, z_t\rangle + \mathbb{E}_{u_{t+1:n}}\left\|\sum_{j=1}^{t} z_j + 6\sum_{s=t+1}^{n} u_s\right\|_\infty\right\}. \tag{8}$$

Now this is in the form that can be used for random playout. Take the algorithm to be: on round $t$, draw $U = 6\sum_{s=t+1}^{n} u_s$ and solve

$$\widehat{y}_t = \operatorname*{argmin}_{\widehat{y}_t\in B_1}\max_{z_t\in B_\infty}\left\{\langle \widehat{y}_t, z_t\rangle + \|z_t + L_{t-1} + U\|_\infty\right\} \tag{9}$$

where $L_{t-1} = \sum_{j=1}^{t-1} z_j$ is the sum of past loss vectors. If the largest (in absolute value) coordinate of $L_{t-1} + U$ is separated by more than 4 from the second-largest value, the solution in (9) takes on a simpler-looking form

$$\widehat{y}_t = \operatorname*{argmin}_{v\in B_1}\ \langle v, L_t + U\rangle, \tag{10}$$

which is attained at a $\pm\boldsymbol{e}_j$ standard unit vector corresponding the the maximal coordinate (homework: prove this). The method (10) is called **Follow-the-Perturbed-Leader** (FTPL) [Han57, KV05].

The FTPL method has a nice interpretation when we consider experts. Let's think of the $B_1$ ball as the set of experts and their negations. Then FTPL says: compute cumulative costs of all experts (here, the experts are the coordinates), add an appropriately scaled random perturbation to each cumulative cost, and choose the best. (Perhaps, the interpretation is cleaner when we consider the probability simplex for $\widehat{y}_t$, as opposed to $B_1$)

While it can be shown that choosing the best response to the empirical performance (Fictitious Play) cannot ensure that (1) is sublinear for all sequences $z_{1:n}$, the FTPL method (Smoothed Fictitious Play) does the job. In fact, the method recovers the correct scaling of the regret bound (7), up to a multiplicative constant.

We just argued that (9) is equivalent to (10) when the gap between the top two coordinates (in absolute value) are separated by more than 4. The probability that this does not happen is small because we are perturbing the cumulative costs by $U$ which has independent coordinates, and each coordinate is an average of $n - t$ coin flips. We will skip

the analysis and just state that the probability that the gap is smaller than 4 is at most $O(N \times \exp\{-c(n-t)^2\})$ for some appropriate constant $c$. On the event of this probability, the per-round loss incurred by the algorithm is $\langle \widehat{y}_t, z_t \rangle \le 1$, and, hence, the extra expected loss when using (10) in place of (9) is $O(N \sum_{t=1}^{n} \exp\{-c(n-t)^2\})$, which is negligible.

We remark that the present proof requires a draw of $U$, which is a sum of $n-t$ vectors with Rademacher random variables. An almost identical proof works for vectors with Gaussians coordinates. The advantage of this choice is computational: we simply draw a vector with independent standard Gaussian components and re-scale it by $\sqrt{n-t}$.

Summary so far: we have derived the Follow-the-Perturbed-Leader algorithm that may be viewed as an alternative to Exponential Weights. As we will see later on the example of Online Shortest Path, FTPL is computationally attractive.

### 1.5 Admissibility

Unfortunately, we are not done yet. The bound $6\sqrt{2n \log N}$ was proved for the algorithm that solves (8) with relaxation (6). Yet, we solved the random-playout version of it. Is it an admissible strategy for this relaxation (as defined earlier)? We need to show that $\mathbf{Rel}(z_{1:t-1})$ is an upper bound on the value of (8) under the proposed strategy.

Let us take a closer look at the response of the max player in (9) when the gap between the top coordinates is at least 4. Suppose $\widehat{y}_t = \pm \boldsymbol{e}_j$ is the FTRL solution. What is the best response for $z_t$ and how much value can this response bring? The value of $z_t$ on coordinates other than $j$ is irrelevant, as the $\ell_\infty$ norm will be achieved on coordinate $j$, thanks to the gap. Moreover, both choices $z_t(j) = \pm 1$ give the same overall value to the objective (can you see why?), which is

$$\|L_{t-1} + U\|_\infty .$$

Now, recalling the "magic" random-playout lemma from the previous lecture (Lemma 1), the value of the overall strategy on the min max objective (8) is upper bounded by the expected value of the min max objective in (9), which is

$$\mathbb{E}_U \|L_{t-1} + U\|_\infty ,$$

which is precisely $\mathbf{Rel}(z_{1:t-1})$. To make the proof completely rigorous, we need to include the expected cost of a bad event of small probability.

### 1.6 Beyond $B_1/B_\infty$

There are very few steps in the above proof that required the particular form of the $\ell_1$ and $\ell_\infty$ norms. The online protocol can be stated for arbitrary pairs of dual norms (and even beyond, as we will see in the next example). The relaxation is modified appropriately and the random playout version of (8) is still possible. What may change is that (9) may no longer take on a simpler form (10), and the perturbation distribution $D$ may need to be taken differently.

## 2. APPLICATION: ONLINE SHORTEST PATH

Consider the problem of choosing a path from home to work, repeatedly for $n$ days. After arriving at work, the traffic information (delays on all the edges) are revealed, prompting us to adjust the strategy for the next day. Our goal is to have a small average time-to-work over the course of $n$ days. We hope that there exists at least one good path (best path in hindsight), and a modest goal is to incur average delay almost as small as that of this best path.

Let $G = (V, E)$ be an acyclic directed graph with a designated source $s$ and sink $t$. Let $m = |E|$ be the number of edges, and $M$ be the number of $s-t$ paths. Each path is associated with the vector $p \in \{0,1\}^m$, indicators of edges present in the path. Let $\mathcal{P} \subseteq \{0,1\}^m$ denote the set of all valid $s-t$ paths. Let $z_t \in [0,1]^m$ denote the delays on all the edges on day $t$. Our goal can be phrased as developing a (possibly randomized) strategy for choosing $\widehat{y}_t \in \mathcal{P}$ such that

$$\frac{1}{n}\sum_{t=1}^{n} \langle \widehat{y}_t, z_t \rangle - \min_{p \in \mathcal{P}} \frac{1}{n}\sum_{t=1}^{n} \langle p, z_t \rangle \tag{11}$$

is small.

How different is this problem from $B_1/B_\infty$ in the previous section? Well, the choice of the delays here is $[0,1]^m$ rather than $[-1,1]^m$, but that's not a big change. What is more crucial is that $B_1$ is replaced by $\mathcal{P}$ which is a (very structured!) subset of $\{0,1\}^m$. Its convex hull $\mathrm{conv}(\mathcal{P})$ is known as the *flow polytope*. The "size" of this set should come into play, but it's not clear at the moment how to measure the size (the answer is: Rademacher averages of the flow polytope).

One approach is to think of each path in $\mathcal{P}$ as an expert and lift the problem in the space $\mathbb{R}^{|\mathcal{P}|}$. However, the update of Exponential Weights, or FTPL in that space would require enumerating all $s-t$ paths, which might take prohibitively long (surprisingly, there is an efficient implementation of Exponential Weights, see Chapter 5 of [CBL06]). We will develop a method (in a different form due to [KV05]) that works in the original space $\mathbb{R}^m$. The method is:

> On round $t$, let $L_{t-1} \in \mathbb{R}^m$ denote the cumulative delays on all edges. Draw a random vector $U$ from a suitable distribution (defined later) and find the shortest path with respect to delays $L_{t-1} + U$.

One can see that this is a version of FTPL method. The computation is simply the shortest-path per round.

### 2.1 Towards a proof

Let's outline the key ingredients of the $B_1/B_\infty$ proof. First, we proved that for any $v \in \mathbb{R}^m$ and any distribution $p$ on $B_\infty$ (which may be chosen based on $v$),

$$\mathbb{E}_{z\sim p} \left\| v + (z - \mathbb{E}[z]) \right\|_\infty \le \mathbb{E}_{u\sim D} \left\| v + 6u \right\|_\infty \tag{12}$$

where $D$ is uniform. This condition is then used in the recursive proof. We then defined a random playout strategy, and showed that, given a gap between the top two coordinates, it is equivalent to FTPL.

All these steps can be extended to our $\mathrm{conv}(\mathcal{P})/[0,1]^m$ setting. Let us take a shortcut that ensures a version of (12) and also introduces an additional technique.

Take the (unnormalized) comparator term in (11) and define the last relaxation as an upper bound

$$-\min_{p\in\mathcal{P}}\sum_{t=1}^{n}\langle p, z_t\rangle = \max_{p\in\mathcal{P}}\left\langle p, \sum_{t=1}^{n} -z_t + \mathbb{E}\gamma\right\rangle \le \mathbb{E}_\gamma \max_{p\in\mathcal{P}}\left\langle p, \sum_{t=1}^{n} -z_t + \gamma\right\rangle \triangleq \mathbf{Rel}(z_{1:n}) \tag{13}$$

Here $\gamma$ is a $\mathbb{R}^m$-valued zero-mean random variable which we will specify later.

Now, there are two possible way to proceed. One is to prove a version of (12):

$$\mathbb{E}_{z\sim p}\mathbb{E}_\gamma \max_{p\in\mathcal{P}}\langle p, v+z+\gamma\rangle \le \mathbb{E}_{u\sim D}\mathbb{E}_\gamma \max_{p\in\mathcal{P}}\langle p, v+6u+\gamma\rangle\,, \tag{14}$$

for any $v$ and any zero-mean distribution $p$ supported on $[-1,1]^m$. However, since $\gamma$ already carries enough randomness, we will simply prove

$$\mathbb{E}_{z\sim p}\mathbb{E}_\gamma \max_{p\in\mathcal{P}}\langle p, v+z+\gamma\rangle \le \mathbb{E}_\gamma \max_{p\in\mathcal{P}}\langle p, v+\gamma\rangle\,, \tag{15}$$

If we choose $\gamma$ to be independent Gaussian with standard deviation $\sqrt{n}$ on each coordinate, we will ensure that with high probability the best path with respect to the delays $v+\gamma$ is $2K$-separated from the second-best path, where $K$ is the length of the longest $s-t$ path. Under this event, the vector $z$, no matter how chosen, cannot cause the best path to change. That is, on that event,

$$\operatorname*{argmax}_{p\in\mathcal{P}}\ \langle p, v+z+\gamma\rangle = \operatorname*{argmax}_{p\in\mathcal{P}}\ \langle p, v+\gamma\rangle\,.$$

The condition (15) then holds. It remains to ensure that the distribution of $\gamma$ has large enough standard deviation for this event to hold.

## 2.2 Proof

We now formally write out the recursion and solve it. Define $\mathbf{Rel}(z_{1:n})$ as in (13). We will see that the relaxations at other time steps are

$$\mathbf{Rel}(z_{1:t}) = \mathbb{E}_\gamma \max_{p\in\mathcal{P}}\left\langle p, -\sum_{s=1}^{t} z_t + \gamma\right\rangle \tag{16}$$

Let us consider the optimization problem at step $t$:

$$\min_{\widehat{y}_t\in\mathcal{P}}\max_{z_t\in[0,1]^m}\left\{\langle \widehat{y}_t, z_t\rangle + \mathbf{Rel}(z_{1:t})\right\}. \tag{17}$$

Using the minimax theorem, as before, the expression above is equal to

$$\max_{p_t\in\Delta([0,1]^m)}\left\{\min_{\widehat{y}_t\in\mathcal{P}}\langle \widehat{y}_t, \mathbb{E}[z_t]\rangle + \mathbb{E}_{z_t}\mathbf{Rel}(z_{1:t})\right\} \tag{18}$$

In the $B_1/B_\infty$ proof we used the fact that the min is simply the minus norm of $\mathbb{E}z_t$, and then proceeded with the triangle inequality. Here, we can no longer write the maximum as a norm, but the triangle inequality still holds in spirit:

$$\min_{\widehat{y}_t\in\mathcal{P}}\langle \widehat{y}_t, \mathbb{E}[z_t]\rangle + \mathbb{E}_{z_t}\mathbf{Rel}(z_{1:t}) \tag{19}$$

$$= \mathbb{E}_{z_t}\mathbb{E}_\gamma \max_{p\in\mathcal{P}}\left\{\langle p, -L_t+\gamma\rangle + \min_{\widehat{y}_t\in\mathcal{P}}\langle \widehat{y}_t, \mathbb{E}[z_t]\rangle\right\} \tag{20}$$

$$\le \mathbb{E}_{z_t}\mathbb{E}_\gamma \max_{p\in\mathcal{P}}\left\{\langle p, -L_{t-1}-z_t+\gamma\rangle + \langle p, \mathbb{E}[z_t]\rangle\right\} \tag{21}$$

where $L_t = \sum_{s=1}^{t} z_s$. We simply replaced the best choice $\widehat{y}_t$ with $p$ (this is essentially what triangle inequality does). We now use (15) to get rid of the worst-case choice $p_t$ over which we have no control. The upper bound is then becomes $\mathbf{Rel}(z_{1:t-1})$, thus proving the recursion.

Under the event that the gap between top paths is at least $2K$, the algorithm reduces to FTPL, just as before. In fact, this improves the result of [KV05] (and the one described in [CBL06]). The improvement is in terms of the regret upper bound.

To conclude a regret upper bound, we need to understand $\mathbf{Rel}(\varnothing)$. We have

$$\mathbf{Rel}(\varnothing) = \mathbb{E}_\gamma \max_{p \in \mathcal{P}} \langle p, \gamma \rangle \tag{22}$$

where $\gamma$ is a random vector that ensures the gap condition with high probability. If we treat all $s - t$ paths as independent, we upper bound (22) as

$$O(K\sqrt{\sigma n \log |\mathcal{P}|})$$

where the linear dependence on $K$ comes from the fact that the number of ones in the vector $p \in \mathcal{P}$ is at most $K$, and $\gamma(i) \sim N(0, n\sigma^2)$. Note that $\sigma$ is not a constant, but has mild dependence on problem parameters (homework: work it out!) More nuanced control of (22) is possible. The result can be seen as an improvement over [KV05, CBL06] which have an extra $m$ dependence.

## 3. APPLICATION: ONLINE RANKING

We briefly discuss another interesting problem, where the set of decisions is a combinatorial subset of the hypercube. As we have seen, the understanding of an achievable regret bound and computationally efficient algorithms rests on the geometry of the set of decisions. In the ranking problem, discussed further in Homework 3, the set of decisions needs to be represented in a non-trivial manner to admit a computationally efficient prediction method.

Let us describe the online ranking problem. There are $d$ teams that repeatedly play each other, for a total of $n$ games. Suppose on round $t$, the pair $(i_t, j_t)$ of two teams is announced, $i_t \neq j_t$. We are interested in predicting the outcome of each game. The prediction should be based both on the past outcomes of the games for the present pair $i_t$ and $j_t$, as well as on the outcomes of all other games (which may indirectly provide the information about team rankings). Nothing is assumed about the order in which teams play each other, and the outcomes are not assumed to be consistent with any ranking.

Notation: Let $\widehat{y}_t \in \{\pm 1\}$ denote our prediction (possibly randomized) of whether team $i_t$ will win against $j_t$. The outcome $y_t \in \{\pm 1\}$ is then observed.

Discussion: the experts bound and the independent-matches bound.

## References


[CBL06] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[Han57] J. Hannan. Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.

[KV05] Adam Tauman Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *J. Comput. Syst. Sci.*, 71(3):291–307, 2005.