Lecture 17

Instructor: Haipeng Luo

1 Adversarial Linear Bandit and FTRL

In the last lecture we discussed the Exp2 algorithm for adversarial linear bandit. The problem of Exp2 is that in general it does not admit an efficient algorithm (in some cases even finding the right exploration distribution is computationally expensive). This time we discuss a different and efficient approach from the seminal work [Abernethy et al., 2008] that borrows a deep and beautiful idea from convex optimization.

Recall the linear bandit problem under an adversarial environment: for each $t = 1, \ldots, T$,

1. learner picks action $w_t \in \Omega \subset \mathbb{R}^d$ while simultaneously environment picks $\ell_t \in \mathbb{R}^d$;
2. learner suffers and observes $w_t^\top \ell_t$ (assume $|w^\top \ell_t| \leq 1$ for any $w \in \Omega$).

Note that we switch to the notation w for an action and Ω for the set of actions to highlight its connection to the OCO setting and the fact that Ω is a compact convex set (instead of a discrete set as for Exp2). Once again we assume that the environment is oblivious and aim to minimize expected regret:

$$\mathbb{E}[\mathcal{R}_T] = \mathbb{E}\left[\sum_{t=1}^{T} w_t^\top \ell_t\right] - \min_{w \in \Omega} \sum_{t=1}^{T} w^\top \ell_t.$$

In the full information setting we have seen the OGD algorithm, an instance of FTRL, that works for general OCO problems. Here we will again consider using FTRL with some regularizer ψ and naturally feed it with some loss estimators $\widehat{\ell}_t$:

$$w_{t+1} = \operatorname*{argmin}_{w \in \Omega} \sum_{\tau=1}^{t} w^\top \widehat{\ell}_\tau + \frac{1}{\eta}\psi(w).$$

There are two main difficulties in this approach that are closely related to each other. The first one is about how to construct the loss estimators. If one looks at the way we construct estimators for Exp3 or Exp2, it is clear that randomization is the key. Therefore, a natural idea is to explore randomly around w_t (computed according to FTRL), instead of exactly playing w_t. One possibility is to simply explore a small ball centered at w_t. This is a reasonable strategy as we will see in the next lecture, but not an optimal one. The reason is that if w_t is close to the boundary of Ω in one direction, then the exploration ball needs to be very small, which limits the exploration in all other directions.

Another difficulty is in choosing the regularizer. In previous examples of FTRL, we have used some regularizer that is strongly convex in some norm $\|\cdot\|$ and have shown that the regret depends on the dual norm of the gradients of the loss functions, which in this case is $\|\widehat{\ell}_t\|_\star$. Since $\widehat{\ell}_t$ in general can have large coordinates (just think about the importance weighted estimators), it is important that the dual norm somehow magically cancels these large coordinates. Indeed, we have seen similar phenomenon in the analysis of Exp3 and Exp2.

It turns out that both difficulties can be simultaneously addressed using one special kind of regularizer, call *self-concordant barriers*, which is also the key concept behind the classic optimization

Algorithm 1: SCRiBLe

Input: learning rate $\eta > 0$ and a ν-self-concordant function ψ
for $t = 1, \dots, T$ **do**
  compute $w_t = \operatorname{argmin}_{w \in \Omega} \sum_{\tau=1}^{t-1} w^\top \widehat{\ell}_\tau + \frac{1}{\eta}\psi(w)$
  compute eigendecomposition $\nabla^2 \psi(w_t) = \sum_{i=1}^d \lambda_i v_i v_i^\top$
  sample $i_t \in [d]$ and $\sigma_t \in \{-1, +1\}$ uniformly at random
  play $\widetilde{w}_t = w_t + \frac{\sigma_t}{\sqrt{\lambda_{i_t}}} v_{i_t}$ and observe $\widetilde{w}_t^\top \ell_t$
  construct estimator $\widehat{\ell}_t = d(\widetilde{w}_t^\top \ell_t)\sigma_t \sqrt{\lambda_{i_t}} v_{i_t}$

algorithm *interior point method*. Instead of stating the definition of self-concordant barriers immediately, which might not be the most intuitive way to understand why it is helpful here, we will defer its definition to the last section and first state on-the-fly some useful properties of self-concordant barriers as we explain and analyze the algorithm.

The first property is about how the Hessian of a self-concordant barrier stretches the space. Specifically, for a point $w \in \text{int}(\Omega)$ ($\text{int}(\Omega)$ denotes the interior of Ω), we define a norm associated with the Hessian of ψ at w as $\|x\|_w = \|x\|_{\nabla^2\psi(w)} = \sqrt{x^\top \nabla^2 \psi(w) x}$ for any $x \in \mathbb{R}^d$. This is indeed a norm since a self-concordant barrier is strictly convex such that $\nabla^2\psi(w)$ is positive definite for any $w \in \text{int}(\Omega)$. The *Dikin ellipsoid* centered at w with radius r is then defined as the ellipsoid $\mathcal{E}_r(w) = \{x \in \mathbb{R}^d : \|x - w\|_w \le r\}$.

Property 1. *If ψ is a self-concordant barrier on Ω, then $\mathcal{E}_1(w) \subset \Omega$ for any $w \in$ int(Ω).*

In other words, the Hessian of a self-concordant barrier stretches the space in a way so that the unit Dikin ellipsoid is always contained in the action set. This implies that given w_t, we can safely explore within the Dikin ellipsoid $\mathcal{E}_1(w_t)$, and it has the hope of better making use the available space than simply using a ball.

Specifically, we will simply uniformly sample one of the end points of the principal axes of the ellipsoid and play this point. In other words, if $\sum_{i=1}^d \lambda_i v_i v_i^\top$ is the eigendecomposition of $\nabla^2\psi(w_t)$, we will then play $\widetilde{w}_t = w_t + \frac{\sigma_t}{\sqrt{\lambda_{i_t}}} v_{i_t}$ where $\sigma_t \in \{-1, +1\}$ is a uniformly random sign and $i_t \in [d]$ is also chosen uniformly at random. It is clear that in expectation we are playing the point w_t, that is $\mathbb{E}_t[\widetilde{w}_t] = w_t$ ($\mathbb{E}_t$ denotes the conditional expectation with respect to the random draw of i_t and σ_t).

With this sampling scheme, we can construct the loss estimator as $\widehat{\ell}_t = d(\widetilde{w}_t^\top \ell_t)\sigma_t \sqrt{\lambda_{i_t}} v_{i_t}$ so that it lies in the direction of the chosen principal axis. This is indeed an unbiased loss estimator since

$$\mathbb{E}_t\left[\widehat{\ell}_t\right] = \frac{1}{d}\sum_{i=1}^d \left(d\sqrt{\lambda_i} v_i \cdot \frac{1}{2}\left(\sum_{\sigma \in \{-1,+1\}} (w_t^\top \ell_t)\sigma + \frac{\sigma^2 v_i^\top \ell_t}{\sqrt{\lambda_i}} \right) \right) = \left(\sum_{i=1}^d v_i v_i^\top \right) \ell_t = \ell_t.$$

This completes all the details of the algorithm (see Algorithm 1), which is called SCRiBLe (Self-Concordant Regularization in Bandit Learning).

2 Regret Analysis

In this section we prove a regret bound for SCRiBLe. The first step is to simply invoke the BTL lemma: for any $u \in \Omega$,

$$\sum_{t=1}^T (w_t - u)^\top \widehat{\ell}_t \le \frac{\psi(u) - \psi(w_1)}{\eta} + \sum_{t=1}^T (w_t - w_{t+1})^\top \widehat{\ell}_t. \tag{1}$$

The rest of the proof will (slightly) deviate from the proof that we have seen since ψ is not necessarily strongly convex. To deal with last term, we apply Hölder's inequality to get $(w_t - w_{t+1})^\top \widehat{\ell}_t \le \|w_t - w_{t+1}\|_{w_t} \|\widehat{\ell}_t\|^\star_{w_t}$ where $\|x\|^\star_{w_t} = \sqrt{x^\top [\nabla^2\psi(w_t)]^{-1} x}$. The term $\|\widehat{\ell}_t\|^\star_{w_t}$ is exactly the dual

norm term mentioned previously. One can verify that the way we construct $\widehat{\ell}_t$ and the dual norm once again "work well" together, leading to important cancellation:

$$\|\widehat{\ell}_t\|_{w_t}^\star = d|\widetilde{w}_t^\top \ell_t|\sqrt{\lambda_{i_t}}\sqrt{v_{i_t}^\top [\nabla^2\psi(w_t)]^{-1} v_{i_t}} = d|\widetilde{w}_t^\top \ell_t|\sqrt{\lambda_{i_t}}\sqrt{\frac{1}{\lambda_{i_t}} v_{i_t}^\top v_{i_t}} \leq d.$$

Next we will show that the algorithm is stable in the sense that $\|w_t - w_{t+1}\|_{w_t} \leq 8\eta\|\widehat{\ell}_t\|_{w_t}^\star$ so that the second term of Eq. (1) is simply bounded by $8\eta T d^2$. To this end let $F_t(w) = \sum_{\tau=1}^{t-1} w^\top \widehat{\ell}_\tau + \frac{1}{\eta}\psi(w)$ so that $w_t = \operatorname{argmin}_w F_t(w)$. Then we have one one hand by optimality of w_t,

$$\begin{aligned} F_{t+1}(w_t) - F_{t+1}(w_{t+1}) &= (w_t - w_{t+1})^\top \widehat{\ell}_t + F_t(w_t) - F_t(w_{t+1}) \\ &\leq (w_t - w_{t+1})^\top \widehat{\ell}_t \leq \|w_t - w_{t+1}\|_{w_t} \|\widehat{\ell}_t\|_{w_t}^\star, \end{aligned} \tag{2}$$

and on the other hand by Taylor's theorem there exists an point ξ on the segment connecting w_t and w_{t+1} such that

$$\begin{aligned} F_{t+1}(w_t) - F_{t+1}(w_{t+1}) &= \nabla F_{t+1}(w_{t+1})^\top (w_t - w_{t+1}) + \tfrac{1}{2}(w_t - w_{t+1})^\top \nabla^2 F_{t+1}(\xi)(w_t - w_{t+1}) \\ &\geq \tfrac{1}{2}(w_t - w_{t+1})\nabla^2 F_{t+1}(\xi)(w_t - w_{t+1}) \\ &\qquad \text{(by first order optimality condition)} \\ &= \tfrac{1}{2\eta}\|w_t - w_{t+1}\|_\xi^2. \end{aligned} \tag{3}$$

If ψ was strongly convex we would have a direct lower bound for the last term. For self-concordant barriers, we need to use another property, which says that within the unit Dikin ellipsoid, the Hessian of every point is pretty close.

Property 2. *If ψ is a self-concordant barrier on Ω, then $\|h\|_{w'} \geq \|h\|_w (1 - \|w - w'\|_w)$ for any $w \in$ int(Ω), $w' \in \mathcal{E}_1(w)$ and $h \in \mathbb{R}^d$.*

Therefore, if we can first show a weaker stability result: $\|w_t - w_{t+1}\|_{w_t} \leq 1/2$ (which also implies $\|w_t - \xi\|_{w_t} \leq 1/2$), then we can use this property to lower bound $\|w_t - w_{t+1}\|_\xi$ by $\|w_t - w_{t+1}\|_{w_t}(1 - \|w_t - \xi\|_{w_t}) \geq \frac{1}{2}\|w_t - w_{t+1}\|_{w_t}$. Combining with Eq. (2) and Eq. (3) would then finish the proof for $\|w_t - w_{t+1}\|_{w_t} \leq 8\eta\|\widehat{\ell}_t\|_{w_t}^\star$.

We now show that $\|w_t - w_{t+1}\|_{w_t} \leq 1/2$ is indeed true. Since w_{t+1} is the minimizer of the convex function F_{t+1}, it suffices to show that $F_{t+1}(w') \geq F_{t+1}(w_t)$ for all w' on the boundary of $\mathcal{E}_{1/2}(w_t)$, that is, $\|h\|_{w_t} = 1/2$ for $h = w' - w_t$. Indeed, using Taylor's theorem again, we have for some ξ lying on the segment connecting w_t and w',

$$\begin{aligned} F_{t+1}(w') &= F_{t+1}(w_t) + \nabla F_{t+1}(w_t)^\top h + \frac{1}{2} h^\top \nabla^2 F_{t+1}(\xi) h \\ &= F_{t+1}(w_t) + \widehat{\ell}_t^\top h + \nabla F_t(w_t)^\top h + \frac{1}{2\eta}\|h\|_\xi^2 \\ &\geq F_{t+1}(w_t) + \widehat{\ell}_t^\top h + \frac{1}{2\eta}\|h\|_{w_t}^2 (1 - \|w_t - \xi\|_{w_t})^2 \\ &\qquad \text{(by first order optimality condition and Property 2)} \\ &\geq F_{t+1}(w_t) - |\widehat{\ell}_t^\top h| + \frac{1}{32\eta} \\ &\geq F_{t+1}(w_t) - \|\widehat{\ell}_t\|_{w_t}^\star \|h\|_{w_t} + \frac{1}{32\eta} \geq F_{t+1}(w_t) - \frac{d}{2} + \frac{1}{32\eta}, \end{aligned}$$

and therefore as long as $\eta \leq \frac{1}{16d}$ we have $F_{t+1}(w') \geq F_{t+1}(w_t)$ and thus $w_{t+1} \in \mathcal{E}_{1/2}(w_t)$.

Finally, it remains the bound the first term of Eq. (1). While the most natural choice of u is simply the best fixed point in hindsight $w_\star = \operatorname{argmin}_{w\in\Omega} \sum_{t=1}^T w^\top \ell_t$, $\psi(u)$ in this case will actually be infinity (since $w_\star$ is on the boundary and a barrier is unbounded on the boundary as we will discuss soon). Fortunately, if ψ is a ν-self-concordant barrier for some parameter $\nu > 0$ (defined in the next section), then the following property holds:

Property 3. *If ψ is a ν-self-concordant barrier on Ω, then for any $\epsilon > 0$ and $u \in \Omega$ such that $u + \epsilon(u - w_1) \in \Omega$ (where $w_1 = \operatorname{argmin}_{w \in \Omega} \psi(w)$), we have $\psi(u) - \psi(w_1) \leq \nu \ln\left(\frac{1}{\epsilon} + 1\right)$.*

Therefore we can set $u = \frac{1}{1+\epsilon}(w_\star - w_1) + w_1$ so that $u + \epsilon(u - w_1) = w_\star$ and $\psi(u) - \psi(w_1) \leq \nu \ln\left(\frac{1}{\epsilon} + 1\right)$. Note that u is not so far away from $w_\star$ and

$$\sum_{t=1}^{T} (u - w_\star)^\top \ell_t \leq \frac{\epsilon}{1+\epsilon} \sum_{t=1}^{T} (w_1 - w_\star)^\top \ell_t \leq 2T\epsilon.$$

Finally noting that $\mathbb{E}_t[\widetilde{w}_t^\top \ell_t] = w_t^\top \ell_t = \mathbb{E}_t[w_t^\top \widehat{\ell}_t]$ and combining everything we have

$$\begin{aligned}\mathbb{E}[\mathcal{R}_T] &= \mathbb{E}\left[\sum_{t=1}^{T} \widetilde{w}_t^\top \ell_t\right] - \sum_{t=1}^{T} w_\star^\top \ell_t = 2T\epsilon + \mathbb{E}\left[\sum_{t=1}^{T} (w_t - u)^\top \widehat{\ell}_t\right] \\ &\leq 2T\epsilon + \frac{\nu}{\eta} \ln\left(\frac{1}{\epsilon} + 1\right) + 8\eta T d^2.\end{aligned}$$

Picking $\epsilon = 1/T$ and the optimal η we have thus proved the following theorem:

Theorem 1. *With $\eta = \min\left\{\frac{1}{16d}, \sqrt{\frac{\nu \ln T}{T d^2}}\right\}$ SCRiBLe ensures $\mathbb{E}[\mathcal{R}_T] = \mathcal{O}(d\sqrt{\nu T \ln T} + d\nu \ln T)$.*

3 Definition and Examples of Self-concordant Barriers

For completeness we now give the formal definition of ν-self-concordant barriers. A function $\psi : \Omega \to \mathbb{R}$ is a barrier if $\psi(w) \to +\infty$ when w approaches the boundary of Ω. A function ψ is self-concordant if it is C^3 (that is, third-order differentiable) and strictly convex and satisfies the following Lipschitz Hessian condition

$$|\nabla^3 \psi(w)[h, h, h]| \leq 2 \|h\|_w^3 \quad \text{for all } w \in \text{int}(\Omega) \text{ and } h \in \mathbb{R}^d$$

where $\nabla^3 \psi(w)[h, h, h]$ is a shorthand for $\sum_{i,j,k \in [d]} \frac{\partial^3 \psi(w)}{\partial w_i \partial w_j \partial w_k} h_i h_j h_k$. Finally a function ψ is a ν-self-concordant barrier if it is a self-concordant barrier and also satisfies the Lipschitz condition

$$|\nabla \psi(w)^\top h| \leq \sqrt{\nu} \|h\|_w \quad \text{for all } w \in \text{int}(\Omega) \text{ and } h \in \mathbb{R}^d.$$

Based on these definitions, one can prove all the three properties we mentioned above.

A seminal result of [Nesterov and Nemirovskii, 1994] states that there *always* exists a ν-self-concordant barrier with $\nu = O(d)$ for a closed convex set $\Omega \subset \mathbb{R}^d$. Canonical examples include the following:

- $\psi(w) = -\sum_{i=1}^{d} \ln w_i$ is a d-self-concordant barrier for $\Omega = \mathbb{R}^d_+$ (verify yourself);
- $\psi(w) = -\sum_{j=1}^{m} \ln(a_j^\top w - b_j)$ is an m-self-concordant barrier for the polytope $\Omega = \{w \in \mathbb{R}^d : a_j^\top w \geq b_j \text{ for } j = 1, \ldots, m\}$;
- $\psi(w) = -\ln(1 - \|w\|_2^2)$ is a 1-self-concordant barrier for the unit ball $\Omega = \{w \in \mathbb{R}^d : \|w\|_2 \leq 1\}$.

The existence of an $O(d)$-self-concordant barrier implies that the regret of SCRiBLe can be as small as $\mathcal{O}(d^{3/2}\sqrt{T \ln T})$ for any Ω (note that the minimax regret for this problem is actually $\mathcal{O}(d\sqrt{T})$). To efficiently implement SCRiBLe, it is not hard to see that one only needs to be able to compute the Hessian of ψ efficiently (see [Abernethy et al., 2008] for details). The best example to showcase the power of SCRiBLe is the online-shortest-path problem where Ω is simply a polytope (a set of flows) so one can use the above concrete barrier and obtain a very efficient algorithm, while Exp2 is very difficult to implement efficiently (indeed in this case uniform exploration does not work provably).

References

Jacob D Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *21st Annual Conference on Learning Theory*, 2008.

Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.