CS292A Convex Optimization: OCO with Bandits Feedback **Spring 2019**

Lecture 15: June 4

Lecturer: Yu-Xiang Wang *Scribes: Omid Askarisichani, Rachel Redberg*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

15.1 Recap (MAB)

K actions at every round, T rounds
A sequence of (bounded) rewards $r_1, r_2, \ldots, r_T \in [0,1]^K$
Choose losses beforehand: $\ell_1, \ell_2, \ldots, \ell_T \in [0,1]^K$

Regret: $\mathbf{E} \sum_{t=1}^{T} \ell_t(a_t) - \min \sum_{t=1}^{T} \ell_t(u) \ \rightarrow$
Replace a_t by x_t, the probability of taking action a: $\sum_{t=1}^{T} \ell_t(x_t) - \sum_{t=1}^{T} \ell_t(u)$

Apply on regret algorithm in the full info setting.

$X_t \in \mathcal{A}(\ell_1, \ell_2, \ldots, \ell_{t-1}) \ \rightarrow$
$X_t \in \tilde{\mathcal{A}}(\ell_1(a_1), \ell_2(a_2), \ldots, \ell_{t-1}(a_{t-1}))$
Reduction approach: regret bound calculated only from A, not $\tilde{A}$. Get a stochastic estimate of $\ell_1, \ldots, \ell_T$ given only the observations.

Idea: stochastic approximation of ℓ_t
This is an unbiased estimate provided $x_t > 0$:

$$\hat{\ell}_t(i) = \begin{cases} \ell_t(i)/x_t(i) & \text{if } i = a_t \\ 0 & \text{otherwise} \end{cases}$$

1. ϵ-greedy: OGD to $\{\hat{\ell}_1, \hat{\ell}_2, \ldots, \hat{\ell}_{t-1}\} \ \Rightarrow \ T^{\frac{3}{4}}\sqrt{K}$
2. EXP3 (exponential-weighted algorithm):
   A = Hedge (FTRL with entropy regularization) to $\{\hat{\ell}_1, \hat{\ell}_2, \ldots, \hat{\ell}_{t-1}\} \ \Rightarrow \ \sqrt{TK\log(\cdot)}$

(Adversarial) Contextual Bandits
Player declares strategy $\mathcal{A}$.
Adversary chooses $x_1, x_2, x_3, \ldots, x_T \in \mathbb{R}^d$. ($x_t$ are context vectors)
Adversary chooses $\ell_1, \ell_2, \ldots, \ell_T \in \mathbb{R}^d$. (Losses are bounded in l_∞ norm by 1: $||\ell_i||_\infty \leq 1$.)
Player is given $\mathcal{H}, h \in \mathcal{H}, h(x) \to a$
$h : X \to \mathcal{A}$
$x_i \to a_i$
(Every element in the hypothesis class is an "expert".)
Contrast to stochastic contextual bandits: sequence is drawn i.i.d.

EXP4: Exponential Weighting algorithm for Explore-Exp-loit with Experts

estimate $\hat{l}_t \in \mathbb{R}^K$
$T[:, :, L_t] \in \mathbb{R}^{|\mathcal{H}| \times K} \Rightarrow \sqrt{TK \log(|\mathcal{H}|)}$
This can be used for deep learning because there is no assumption that the hypothesis class $\mathcal{H}$ is convex (the only assumption is that the loss function is convex). But the runtime is $\Theta(|\mathcal{H}|)$ (linear in $|\mathcal{H}|$), so this is not efficient for large $|\mathcal{H}|$.

Can apply polynomial computation: See the paper "Taming the Monster", by Agarwal, Langford [agarwal2014].

15.2 OCO with Bandits Feedback

Expert Advice, MAB $\longrightarrow$ OCO (how?)

Algorithm 1

procedure EXPERTADVICES
  Setup: K feedback convex.
  Player A.
  Adversary chooses $f_1, \ldots, f_T : K \to \mathbb{R}$.
  for $t = 1, 2, \ldots, T$ **do**
    Player plays $X_t \sim A\left(f(X_1), f(X_2), \ldots, f(X_{t-1})\right)$.
    Player observes and suffers loss $f_t(X_t)$.

The regret for the algorithm is

$$\begin{aligned}\text{Regret} &= E\sum_{t=1}^{T} f_t(X_t) - \sum_{t=1}^{T} f_t(u), \quad \forall \text{ fixed } u. \\ &\leq E\sum_{t=1}^{T} < \nabla f_t(X_t), X_t > - \sum_{t=1}^{T} < \nabla f_t(X_t), v >, \quad \forall \text{ fixed } v.\end{aligned}$$

Algorithm 2 Reduction to Bandit Convex Optimization

procedure REDUCTIONBANDITCO
  Input: Convex set K, first order(full info) OCO A.
  $X_1 = A(\emptyset)$.
  for $t = 1, 2, \ldots, T$ **do**
    Sample $y_t \sim D_t$, such that $E[y_t] = X_t$.
    Play y_t, observe $f_t(y_t)$, generate g_t, such that $E[g_t] = \nabla f_t(X_t)$.
    $X_{t+1} = A(g_1, g_2, \ldots, g_t)$.

Lemma 15.1 *Let $u \in K$ fixed, $\forall f_1, \ldots, f_t : K \to \mathbb{R}$, and they are differentiable.*
Assume $Regret_T(A) \leq B_A(\nabla_1 f(X_1), \nabla_2 f(X_2), \ldots, \nabla_t f(X_t))$ in full info.
If in addition, $E[g_t | X_1, f_1, X_2, f_2, \ldots, X_t, f_t] = \nabla f_t(X_t)$, then

$$Regret_{Alg1,T} \leq E[B_A(g_1, \ldots, g_T].$$

Example (SGD):

$$\begin{aligned} E[g_t] &= \nabla f(X_t) \\ E[||g_t - Eg_t||_2^2] &\leq \delta^2 \\ E[||g_t||_2^2] &\leq G^2 + \delta^2 \end{aligned}$$

Proof: Let $h_t(X) = f_t(X) + (g_t - \nabla f_t(X_t)^T X$, we know the following

1. $\nabla h_t(X_t) = \nabla f_t(X_t) + g_t - \nabla f_t(X_t) = g_t$.
2. $$\begin{aligned} E[h_t(X)] &= E[f_t(X)] + E\left[(g_t - \nabla f_t(X_t))^T X\right] \\ &= E[f_t(X)] + E\left[E[(g_t - \nabla f_t(X_t))^T X | X_1, f_1, X_2, f_2, \ldots, X_t, f_t]\right] \\ &= E[f_t(X)] + E[0^T X] \\ &= E[f_t(X)]. \end{aligned}$$
3. $E[h_t(X_t)] = E[f_t(X_t)]$.

By regret bound of $A, \forall$ fixed $u \in K$,

$$\sum_{t=1}^{T} h_t(X_t) - \sum_{t=1}^{T} h_t(u) \leq B_A(g_1, \ldots, g_T).$$

Take expectation: we apply item 3 on first term, item 2 on the second term, and eventually have

$$E\sum_{t=1}^{T} f_t(X_t) - E\sum_{t=1}^{T} f_t(u) \leq E[B_A(g_1, \ldots, g_T)].$$

■

How do we estimate the gradient without a gradient?

$f : \mathbb{R} \to \mathbb{R}$
$f'(x) = \lim_{\delta \to 0} \frac{f(x+\delta)-f(x-\delta)}{2\delta}$

Let

$$g(x) = \begin{cases} \frac{f(x+\delta)}{\delta} & \text{with prob. } \frac{1}{2} \\ -\frac{f(x-\delta)}{\delta} & \text{with prob. } \frac{1}{2} \end{cases}$$

$$\mathbf{E}[g(x)] = \frac{1}{2}\frac{f(x+\delta)}{\delta} - \frac{1}{2}\frac{f(x-\delta)}{\delta} = \frac{f(x+\delta)-f(x-\delta)}{2\delta} = f'(x).$$

$f : \mathbb{R}^n \to \mathbb{R}$

$$B_\delta = \{x \mid ||x||_2 \le \delta\}, \quad S_\delta = \{x \mid ||x||_2 = \delta\}$$

$\hat{f}_\delta(X) = E_{v \sim \text{uniform}(B_1)}[f(X+\delta v)].$

Example (linear):
$f(X) =< l, X > \longrightarrow \hat{f}_\delta(X) = f(X).$

$g(X) = f(X+\delta u).u, \quad u \sim S_1 \in \mathbb{R}^n.$

Lemma 15.2 $E_{u \sim S_1}[f(X+\delta u).u] = \frac{\delta}{n}\nabla \hat{f}_\delta(X).$

Proof: We use Stokes Theorem:

$$\int_\Sigma \frac{d\mu}{dX} = \oint_{\delta\Sigma} dX.$$

Which in here we can write

$$\nabla \int_{B_\delta} f(X+v)dv = \int_{S_\delta} f(X+u)\frac{u}{||u||}du.$$

Thus, we have

$$\begin{aligned} \hat{f}_\delta(X) =& \frac{\int_{B_\delta} f(X+v)dv}{\text{Vol}(B_\delta)} E_{v \sim S}[f(x+\delta u).u] \\ =& \frac{\int_{S_\delta} f(X+u)\frac{u}{||u||}du}{\text{Vol}(S_\delta)} \\ =& \frac{\text{Vol}(B_\delta)}{\text{Vol}(S_\delta)} \\ =& \frac{\delta}{n}. \end{aligned}$$

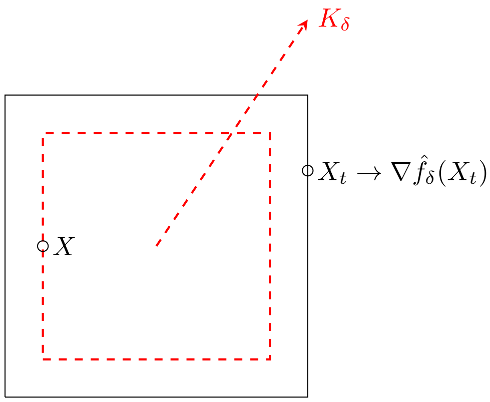
■

Algorithm 3

---

procedure FKM
  Input K, $0 \in K$, $B_1 \subset K$, $||f_t(X)|| \leq 1$, for $\forall X \in K$.
  for $t = 1, 2, \ldots, T$ **do**
    Draw $u_t \sim S$, $y_t = X_t + \delta u$.
    Play y_t, $f_t(y_t)$, let $g_t = \frac{n}{\delta} f_t(y_t).u_t$.
    Update $X_{t+1} = \pi_{K_\delta}[X_t - \eta g_t]$.

Apply full info A to K_δ such that $g_t = f(X + \delta u).u$ and $u \sim S_\delta$ uniformly.



To prove this we need to:

- Cost $K \to K_delta$
- Cost $f_t \to f_{t_delta}$
- Bound $E[||g_t||^2]$ with n, δ, f_t
- Choose δ and η carefully

References

[1] Agarwal, Alekh, Daniel Hsu, Satyen Kale, John Langford, Lihong Li, and Robert Schapire. (2014) Taming the monster: A fast and simple algorithm for contextual bandits. *In International Conference on Machine Learning*:1638–1646.