Bandit Convex Optimization

Scott Yang

March 7, 2017

Table of Contents

Learning scenario

- Compact convex action set $\mathcal{K} \subset \mathbb{R}^d$.
- For $t = 1$ to T:
  - Predict $x_t \in \mathcal{K}$.
  - Receive convex loss function $f_t : \mathcal{K} \to \mathbb{R}$.
  - Incur loss $f_t(x_t)$.
- *Bandit setting*: only loss revealed, no other information.
- Regret of algorithm $\mathcal{A}$:

$$\mathrm{Reg}_T(\mathcal{A}) = \sum_{t=1}^{T} f_t(x_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x).$$

Related settings

- Online convex optimization: ∇f_t (and maybe $\nabla^2 f_t, \nabla^3 f_t, \ldots$) known at each round.
- Multi-armed bandit: $\mathcal{K} = \{1, 2, \ldots, K\}$ discrete.
- Zero-th order optimization: $f_t = f$.
- Stochastic bandit convex optimization: $f_t(x) = f(x) + \epsilon_t$, $\epsilon_t \sim \mathcal{D}$ noisy estimate.
- Multi-point bandit convex optimization: Query f_t at points $(x_{t,i})_{i=1}^m$, $m \geq 2$.

Table of Contents

"Gradient descent without a gradient"

- Online gradient descent algorithm:

$$x_{t+1} \leftarrow x_t - \eta \nabla f_t(x_t).$$

- BCO setting: $\nabla f_t(x_t)$ is not known!
- BCO idea:
  - Find $\widehat{g}_t$ such that $\widehat{g}_t \approx \nabla f_t(x_t)$.
  - Update

$$x_{t+1} \leftarrow x_t - \eta \widehat{g}_t.$$

- Question: how do we pick $\widehat{g}_t$?

Single-point gradient estimates (one dimension)

- By the fundamental theorem of calculus:

$$f'(x) \approx \frac{1}{2\delta} \int_{-\delta}^{\delta} f'(x+y)dy = \frac{1}{2\delta} [f(x+\delta) - f(x-\delta)]$$

$$= \mathop{\mathbb{E}}_{z \sim \mathcal{D}} \left[\frac{1}{\delta} f(x+z) \frac{z}{|z|} \right]$$

where $\mathcal{D}(z) = \delta$ w.p. $\frac{1}{2}$ and $= -\delta$ w.p. $\frac{1}{2}$.

- With enough regularity (e.g. f Lipschitz),

$$\frac{d}{dx} \frac{1}{2\delta} \int_{-\delta}^{\delta} f(x+y)dy = \frac{1}{2\delta} \int_{-\delta}^{\delta} f'(x+y)dy.$$

Single-point gradient estimates (higher dimensions)

- $\mathbb{B}_1 = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$.
- $\mathbb{S}_1 = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.
- $\int_A dy = |A|$.
- By Stokes' theorem,

$$\begin{aligned}\nabla f(x) &\approx \frac{1}{|\delta \mathbb{B}_1|} \int_{\delta \mathbb{B}_1} \nabla f(x+y) dy = \frac{1}{|\delta \mathbb{B}_1|} \int_{\delta \mathbb{S}_1} f(x+z) \frac{z}{|z|} dz \\ &= \frac{1}{|\delta \mathbb{B}_1|} \int_{\mathbb{S}_1} f(x+\delta z) z dz = \frac{|\mathbb{S}_1|}{|\mathbb{S}_1||\delta \mathbb{B}_1|} \int_{\mathbb{S}_1} f(x+\delta z) z dz \\ &= \frac{|\mathbb{S}_1|}{|\delta \mathbb{B}_1|} \mathbb{E}_{z \sim U(\mathbb{S}_1)} [f(x+\delta z)] = \frac{d}{\delta} \mathbb{E}_{z \sim U(\mathbb{S}_1)} [f(x+\delta z)] .\end{aligned}$$

- With enough regularity on f,

$$\nabla \frac{1}{|\delta \mathbb{B}_1|} \int_{\delta \mathbb{B}_1} f(x+y) dy = \frac{1}{|\delta \mathbb{B}_1|} \int_{\delta \mathbb{B}_1} \nabla f(x+y) dy.$$

Projection method [Flaxman et al, 2005]

- Let $\widehat{f}(x) = \frac{1}{|\delta\mathbb{B}_1|}\int_{\delta\mathbb{B}_1} f(x+y)dy$.
- Estimate $\nabla\widehat{f}(x)$ by sampling on $\delta\mathbb{S}_1(x)$
- Project gradient descent update to keep samples inside $\mathcal{K}$: $K_\delta = \frac{1}{1-\delta}K$.

BANDITPGD(T, η, δ):

- $x_1 \leftarrow 0$.
- For $t = 1, 2, \ldots, T$:
  - $u_t \leftarrow$ SAMPLE($\mathbb{S}_1$)
  - $y_t \leftarrow x_t + \delta u_t$
  - PLAY(y_t)
  - $f_t(y_t) \leftarrow$ RECEIVELOSS(y_t)
  - $\widehat{g}_t \leftarrow \frac{d}{\delta} f_t(y_t) u_t$
  - $x_{t+1} \leftarrow \Pi_{K_\delta}(x_t - \eta\widehat{g}_t)$.

Analysis of BANDITPGD

Theorem (Flaxman et al, 2005)

Assume $\text{diam}(\mathcal{K}) \leq D$, $|f_t| \leq C$, *and* $\|\nabla f_t\| \leq L$. *Then after* T *rounds, the (expected) regret of the* BANDITPGD *algorithm is bounded by:*

$$\frac{D^2}{2\eta} + \frac{\eta C^2 d^2 T}{2\delta^2} + \delta(D+2)LT.$$

In particular, by setting $\eta = \frac{\delta D}{Cd\sqrt{T}}$ *and* $\delta = \sqrt{\frac{DCd}{(D+2)LT^{1/2}}}$, *the regret is upper bounded by:* $\mathcal{O}(d^{1/2}T^{3/4})$.

Proof of BANDITPGD regret

- For any $x \in \mathcal{K}$, let $x_\delta = \Pi_{K_\delta}(x)$.
- $\widehat{f}_t(z) \geq f_t(z)$.
- Then

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}[f_t(y_t) - f_t(x^*)] \\
&= \sum_{t=1}^{T} \mathbb{E}\Bigg[f_t(y_t) - f_t(x_t) + f_t(x_t) - \widehat{f}_t(x_t) + \widehat{f}_t(x_t) - \widehat{f}_t(x_\delta^*) \\
&\quad + \widehat{f}_t(x_\delta^*) - f_t(x_\delta^*) + f_t(x_\delta^*) - f_t(x^*)\Bigg] \\
&\leq \sum_{t=1}^{T} \mathbb{E}\left[\widehat{f}_t(x_t) - \widehat{f}_t(x_\delta^*)\right] + [2\delta LT + \delta DLT] \\
&\leq \sum_{t=1}^{T} \mathbb{E}\left[\widehat{f}_t(x_t) - \widehat{f}_t(x_\delta^*)\right] + \delta(D+2)LT.
\end{aligned}$$

Proof of BanditPGD regret

- $\mathbb{E}\left[\|\widehat{g}_t\|_2\right] \leq \frac{C^2 d^2}{\delta^2}$.
- Thus,

$$\sum_{t=1}^{T} \mathbb{E}\left[\widehat{f}_t(x_t) - \widehat{f}_t(x_\delta^*)\right] \leq \sum_{t=1}^{T} \mathbb{E}\left[\nabla \widehat{f}_t(x_t) \cdot (x_t - x_\delta^*)\right]$$

$$= \sum_{t=1}^{T} \mathbb{E}\left[\widehat{g}_t \cdot (x_t - x_\delta^*)\right]$$

$$= \sum_{t=1}^{T} \frac{1}{2\eta} \mathbb{E}\left[\eta^2 \|\widehat{g}_t\|^2 + \|x_t - x_\delta^*\|^2 - \|x_t - \eta \widehat{g}_t - x_\delta^*\|^2\right]$$

$$\leq \sum_{t=1}^{T} \frac{1}{2\eta} \mathbb{E}\left[\eta^2 \frac{C^2 d^2}{\delta^2} + \|x_t - x_\delta^*\|^2 - \|x_{t+1} - x_\delta^*\|^2\right]$$

$$\leq \frac{1}{2\eta} \mathbb{E}\left[\|x_1 - x_\delta^*\|^2 + \eta^2 \frac{C^2 d^2}{\delta^2}\right] \leq \frac{1}{2\eta}\left[D^2 + \eta^2 \frac{C^2 d^2}{\delta^2}\right].$$

Table of Contents

Revisiting projection

- Goal of projection: Keep $x_t \in \mathcal{K}_\delta$ so that $y_t \in \mathcal{K}$.
- Total "cost" of projection: δDLT.
- Deficiency: completely separate from gradient descent update.
- Question: is there a better way to ensure that $y_t \in \mathcal{K}$?

Gradient Descent to Follow-the-Regularized-Leader

- Let $\widehat{g}_{1:t} = \sum_{s=1}^{t} \widehat{g}_s$.
- Proximal form of gradient descent:

$$x_{t+1} \leftarrow x_t - \eta \widehat{g}_t$$

$$x_{t+1} \leftarrow \operatorname*{argmin}_{x \in \mathbb{R}^d} \eta \widehat{g}_{1:t} \cdot x + \|x\|^2$$

- Follow-the-Regularized-Leader: for $\mathcal{R} : \mathbb{R}^d \to \mathbb{R}$,

$$x_{t+1} \leftarrow \operatorname*{argmin}_{x \in \mathbb{R}^d} \widehat{g}_{1:t} \cdot x + \mathcal{R}(x),$$

- BCO "wishlist" for $\mathcal{R}$:
  - Want to ensure that x_{t+1} stays inside $\mathcal{K}$.
  - Want enough "room" so that $y_{t+1} \in \mathcal{K}$ as well.

Self-concordant barriers

Definition (Self-concordant barrier (SCB))

Let $\nu \geq 0$. A C^3 function $\mathcal{R} : \text{int}(\mathcal{K}) \to \mathbb{R}$ is a ν-self-concordant barrier for $\mathcal{K}$ if for any sequence $(z_s)_{s=1}^{\infty} \subset \text{int}(\mathcal{K})$, with $z_s \to \partial\mathcal{K}$, we have $\mathcal{R}(z_s) \to \infty$, and for all $x \in \mathcal{K}$ and $y \in \mathbb{R}^n$, the following inequalities hold:

$$|\nabla^3 \mathcal{R}(x)[y, y, y]| \leq 2\|y\|_x^3, \quad |\nabla \mathcal{R}(x) \cdot y| \leq \nu^{1/2}\|y\|_x,$$

where $\|z\|_x^2 = \|z\|_{\nabla^2 \mathcal{R}(x)}^2 = z^\top \nabla^2 \mathcal{R}(x) z$.

Examples of barriers

- $\mathcal{K} = \mathbb{B}_1$:
$$\mathcal{R}(x) = -\log(1 - \|x\|^2)$$
is 1-self-concordant.
- $\mathcal{K} = \{x : a_i^\top x \leq b_i\}_{i=1}^m$:
$$\mathcal{R}(x) = \sum_{i=1}^{m} -\log(b_i - a_i^\top x)$$
is m-self-concordant.
- Existence of "universal barrier" [Nesterov & Nemirovski, 1994]: every closed convex domain $\mathcal{K}$ admits a $\mathcal{O}(d)$-self-concordant barrier.

Properties of self-concordant-barriers

- *Translation invariance*:
  for any constant $c \in \mathbb{R}$, $\mathcal{R} + z$ is also a SCB (so wlog, we assume $\min_{z \in \mathcal{K}} \mathcal{R}(z) = 0$.)
- *Dikin ellipsoid contained in interior*:
  let $\mathcal{E}(x) = \{y \in \mathbb{R}^n : \|y\|_x \leq 1\}$. Then for any $x \in \text{int}(\mathcal{K})$, $\mathcal{E}(x) \subset \text{int}(\mathcal{K})$.
- *Logarithmic growth away from boundary*:
  for any $\epsilon \in (0, 1]$, let $y = \text{argmin}_{z \in \mathcal{K}} \mathcal{R}(z)$ and $\mathcal{K}_{y,\epsilon} = \{y + (1 - \epsilon)(x - y) : x \in \mathcal{K}\}$. Then for all $x \in \mathcal{K}_{y,\epsilon}$,
  $$\mathcal{R}(x) \leq \nu \log(1/\epsilon).$$
- *Proximity to minimizer*: If $\|\nabla \mathcal{R}(x)\|_{x,*} \leq \frac{1}{2}$, then
  $$\|x - \text{argmin}\,\mathcal{R}\|_x \leq 2\|\nabla \mathcal{R}(x)\|_{x,*}.$$

Adjusting to the local geometry

- Let $A \succ 0$ SPD matrix.
- Sampling around A instead of Euclidean ball:

$$u \sim \textsc{Sample}(\mathbb{S}_1), \quad x \leftarrow y + \delta A u$$

- Smoothing over A instead of Euclidean ball:

$$\widehat{f}(x) = \mathbb{E}_{u \sim U(\mathbb{S}_1)}[f(x + \delta A u)].$$

- One-point gradient estimate based on A:

$$\widehat{g} = \frac{d}{\delta} f(x + \delta A u) A^{-1} u, \quad \mathbb{E}_{u \sim U(\mathbb{S}_1)} \left[\widehat{g}\right] = \nabla \widehat{f}(x).$$

- Local norm bound:

$$\|\widehat{g}\|_{A^2}^2 \leq \frac{d^2}{\delta^2} C^2.$$

BANDITFTRL [Abernethy et al, 2008; Saha and Tewari 2011]

BANDITFTRL($\mathcal{R}$, δ, η, T, x_1)

- For $t \leftarrow 1$ to T:
  - $u_t \leftarrow$ SAMPLE($U(\mathbb{S}_1)$).
  - $y_t \leftarrow x_t + \delta(\nabla^2\mathcal{R}(x_t))^{-1/2} u_t$.
  - PLAY(y_t).
  - $f_t(y_t) \leftarrow$ RECEIVELOSS(y_t).
  - $\widehat{g}_t \leftarrow \frac{d}{\delta} f_t(y_t) \nabla^2\mathcal{R}(x_t) u_t$.
  - $x_{t+1} \leftarrow \text{argmin}_{x \in \mathbb{R}^d} \eta \widehat{g}_{1:t}^{\top} x + \mathcal{R}(x)$.

Analysis of BanditFTRL

Theorem (Abernethy et al, 2008; Saha and Tewari, 2011)

Assume $\mathrm{diam}(\mathcal{K}) \leq D$*. Let* $\mathcal{R}$ *be a self-concordant-barrier for* $\mathcal{K}$*,* $|f_t| \leq C$*, and* $\|\nabla f_t\| \leq L$*. Then the regret of* BanditFTRL *is upper bounded as follows:*

- *If* $(f_t)_{t=1}^T$ *are linear functions, then* $\mathrm{Reg}_T(\text{BanditFTRL}) = \tilde{\mathcal{O}}(T^{1/2})$.
- *If* $(f_t)_{t=1}^T$ *have Lipschitz gradients, then* $\mathrm{Reg}_T(\text{BanditFTRL}) = \tilde{\mathcal{O}}(T^{2/3})$.

Proof of BanditFTRL regret: linear case

Approximation error of smoothed losses:

- $x^* = \text{argmin}_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x)$
- $x_\epsilon^* \in \text{argmin}_{y \in \mathcal{K}, \text{dist}(y, \partial\mathcal{K}) > \epsilon} \|y - x^*\|$
- Because f_t are linear,

$$\text{Reg}_T(\text{BanditFTRL}) = \mathbb{E}\left[\sum_{t=1}^{T} f_t(y_t) - f_t(x^*)\right]$$

$$= \mathbb{E}\Bigg[\sum_{t=1}^{T} f_t(y_t) - \widehat{f}_t(y_t) + \widehat{f}_t(y_t) - \widehat{f}_t(x_t) + \widehat{f}_t(x_t) - \widehat{f}_t(x_\epsilon^*) + \widehat{f}_t(x_\epsilon^*)$$

$$- f_t(x_\epsilon^*) + f_t(x_\epsilon^*) - f_t(x^*)\Bigg]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} \widehat{f}_t(x_t) - \widehat{f}_t(x_\epsilon^*)\right] + \epsilon LT = \mathbb{E}\left[\sum_{t=1}^{T} \widehat{g}_t^\top (x_t - x_\epsilon^*)\right] + \epsilon LT.$$

Proof of BanditFTRL regret: linear case

Claim: for any $z \in \mathcal{K}$,

$$\sum_{t=1}^{T} \widehat{g}_t^\top (x_{t+1} - z) \leq \frac{1}{\eta}\mathcal{R}(z).$$

- $T = 1$ case is true by definition of x_2.
- Assuming statement is true for $T - 1$:

$$\sum_{t=1}^{T} \widehat{g}_t^\top x_{t+1} = \sum_{t=1}^{T-1} \widehat{g}_t^\top x_{t+1} + \widehat{g}_T^\top x_{T+1} \leq \frac{1}{\eta}\mathcal{R}(x_T) + \sum_{t=1}^{T-1} \widehat{g}_t^\top x_T + \widehat{g}_T^\top x_{T+1}$$
$$\leq \frac{1}{\eta}\mathcal{R}(x_{T+1}) + \sum_{t=1}^{T-1} \widehat{g}_t^\top x_{T+1} + \widehat{g}_T^\top x_{T+1} \leq \frac{1}{\eta}\mathcal{R}(z) + \sum_{t=1}^{T} \widehat{g}_t^\top z.$$

Proof of BanditFTRL regret: linear case

$$\mathbb{E}\left[\sum_{t=1}^{T}\widehat{f}_t(x_t)-\widehat{f}_t(x_\epsilon^*)\right] \le \sum_{t=1}^{T}\mathbb{E}\left[\widehat{f}_t(x_t)-\widehat{f}_t(x_{t+1})+\widehat{f}_t(x_{t+1})-\widehat{f}_t(x_\epsilon^*)\right]$$

$$\le \sum_{t=1}^{T}\mathbb{E}\left[\|\widehat{g}_t\|_{x_t,*}\|x_t-x_{t+1}\|_{x_t}\right]+\frac{1}{\eta}\mathcal{R}(x_\epsilon^*)$$

Proximity to minimizer for SCB:

- Recall:
  - $x_{t+1} = \operatorname{argmin}_{x\in\mathcal{K}} \eta\widehat{g}_{1:t}^\top x + \mathcal{R}(x)$
  - $F_t(x) = \eta\widehat{g}_{1:t}^\top x + \mathcal{R}(x)$ is a SCB.
- Proximity bound: $\|x_t - x_{t+1}\|_{x_t} \le \|\nabla F_t(x_t)\|_{x_t,*} = \eta\|\widehat{g}_t\|_{x_t,*}$

Proof of BANDITFTRL regret: linear case

$$\text{Reg}_T(\text{BANDITFTRL}) \leq \epsilon LT + \sum_{t=1}^{T} \mathbb{E}\left[\eta \|\widehat{g}_t\|_{x_t,*}^2\right] + \frac{1}{\eta}\mathcal{R}(x_\epsilon^*)$$

- By the local norm bound: $\mathbb{E}\left[\eta \|\widehat{g}_t\|_{x_t,*}^2\right] \leq \frac{C^2 d^2}{\delta^2}$.
- By the logarithmic growth of the SCB: $\frac{1}{\eta}\mathcal{R}(x_\epsilon^*) \leq \nu \log\left(\frac{1}{\epsilon}\right)$.

$$\Rightarrow \text{Reg}_T(\text{BANDITFTRL}) \leq \epsilon LT + T\eta\frac{C^2 d^2}{\delta^2} + \frac{\nu}{\eta}\log\left(\frac{1}{\epsilon}\right).$$

Table of Contents

Issues with BANDITFTRL in the non-linear case

- Approximation error of $f \sim \widehat{f}$:
  - $\sim \delta T$ for $\mathcal{C}^{0,1}$ functions
  - $\sim \delta^2 T$ for $\mathcal{C}^{1,1}$ functions
- Variance of gradient estimates: $\mathbb{E}\left[\eta \|\widehat{g}_t\|^2_{x_t,*}\right] \leq \frac{C^2 d^2}{\delta^2}$
- Regret for non-linear loss functions:
  - $\mathcal{O}(T^{3/4})$ for $\mathcal{C}^{0,1}$ functions
  - $\mathcal{O}(T^{2/3})$ for $\mathcal{C}^{1,1}$ functions
- Question: can we reduce the variance of the gradient estimates to improve the regret?

Variance reduction

- Observation [Dekel et al, 2015]: If $\bar{g}_t = \frac{1}{k+1} \sum_{i=0}^{k} \widehat{g}_{t-i}$, then

$$\|\bar{g}_t\|^2_{x_t,*} = \mathcal{O}\left(\frac{C^2 d^2}{\delta^2 (k+1)}\right).$$

- Note: averaged gradient $\bar{g}_t$ is no longer an unbiased estimate of $\nabla \widehat{f}_t$.
- Idea: If f_t is sufficiently regular, then the bias will still be manageable.

Improving variance reduction via "optimism"

- Optimistic FTRL [Rakhlin and Sridharan, 2013]:

$$x_{t+1} \leftarrow \underset{x \in \mathcal{K}}{\operatorname{argmin}} (g_{1:t} + \tilde{g}_{t+1})^\top x + \mathcal{R}(x)$$

$$\sum_{t=1}^{T} f_t(x_t) - f_t(x^*) \leq \eta \sum_{t=1}^{T} \|g_t - \tilde{g}_t\|_{x_t,*} + \frac{1}{\eta}\mathcal{R}(x^*)$$

- By re-centering the averaged gradient at each step, we can further reduce the variance:

$$\tilde{g}_t = \frac{1}{k+1} \sum_{i=1}^{k} \widehat{g}_{t-i}.$$

- Variance of re-centered averaged gradients:

$$\|\bar{g}_t - \tilde{g}_t\|_{x_t,*}^2 = \frac{1}{(k+1)^2} \|\widehat{g}_t\|_{x_t,*}^2 = \mathcal{O}\left(\frac{C^2 d^2}{\delta^2 (k+1)^2}\right).$$

BanditFTRL-VR

BanditFTRL-VR($\mathcal{R}$, δ, η, k, T, x_1)

- For $t \leftarrow 1 \rightarrow T$:
  - $u_t \leftarrow$ Sample$(U(\mathbb{S}_1))$
  - $y_t \leftarrow x_t + \delta(\nabla^2 \mathcal{R}(x_t))^{-\frac{1}{2}} u_t$
  - Play(y_t)
  - $f_t(y_t) \leftarrow$ ReceiveLoss(y_t)
  - $\widehat{g}_t \leftarrow \frac{d}{\delta} f_t(y_t)(\nabla^2 \mathcal{R}(x_t))^{-\frac{1}{2}} u_t$
  - $\bar{g}_t \leftarrow \frac{1}{k+1} \sum_{i=0}^{k} \widehat{g}_{t-i}$
  - $\tilde{g}_{t+1} \leftarrow \frac{1}{k+1} \sum_{i=1}^{k} \widehat{g}_{t+1-i}$
  - $x_{t+1} \leftarrow \text{argmin}_{x \in \mathbb{R}^d} \eta(\bar{g}_{1:t} + \tilde{g}_{t+1})^\top x + \mathcal{R}(x)$

Analysis of BANDITFTRL-VR

Theorem (Mohri & Y., 2016)

Assume $\operatorname{diam}(\mathcal{K}) \leq D$*. Let* $\mathcal{R}$ *be a self-concordant-barrier for* $\mathcal{K}$*,* $|f_t| \leq C$*, and* $\|\nabla f_t\| \leq L$*. Then the regret of* BANDITFTRL *is upper bounded as follows:*

- *If* $(f_t)_{t=1}^T$ *are Lipschitz, then* $\mathrm{Reg}_T(\text{BANDITFTRL-VR}) = \tilde{\mathcal{O}}(T^{\frac{11}{16}})$.
- *If* $(f_t)_{t=1}^T$ *have Lipschitz gradients, then* $\mathrm{Reg}_T(\text{BANDITFTRL-VR}) = \tilde{\mathcal{O}}(T^{\frac{8}{13}})$.

Proof of BANDITFTRL-VR regret: Lipschitz case

Approximation: real to smoothed losses

- Relate global optimum x^* to projected optimum x^*_ϵ.
- Use Lipschitz property of losses to relate y_t to x_t and f_t to $\widehat{f}_t$.

$$
\begin{aligned}
\mathrm{Reg}_T(\text{BANDITFTRL-VR}) &= \mathbb{E}\left[\sum_{t=1}^{T} f_t(y_t) - f_t(x^*)\right] \\
&\leq \epsilon LT + 2L\delta DT + \sum_{t=1}^{T} \mathbb{E}\left[\widehat{f}_t(x_t) - \widehat{f}(x^*_\epsilon)\right].
\end{aligned}
$$

Proof of BANDITFTRL-VR regret: Lipschitz case

Approximation: smoothed to averaged losses

$$\sum_{t=1}^{T} \mathbb{E}\left[\widehat{f}_t(x_t) - \widehat{f}(x_\epsilon^*)\right] = \sum_{t=1}^{T} \mathbb{E}\Bigg[\frac{1}{k+1}\sum_{i=0}^{k}\left(\widehat{f}_t(x_t) - \widehat{f}_{t-i}(x_{t-i})\right)$$

$$+ \frac{1}{k+1}\sum_{i=0}^{k}\left(\widehat{f}_{t-i}(x_{t-i}) - \bar{f}_t(x_\epsilon^*)\right) + \frac{1}{k+1}\sum_{i=0}^{k}\left(\bar{f}_t(x_\epsilon^*) - \widehat{f}_t(x_\epsilon^*)\right)\Bigg]$$

$$\leq \frac{Ck}{2} + LT \sup_{\substack{t\in[1,T] \\ i\in[0,k\wedge t]}} \mathbb{E}\left[\|x_{t-i} - x_t\|_2\right] + \sum_{t=1}^{T} \mathbb{E}\left[\bar{g}_t^\top (x_t - x_\epsilon^*)\right].$$

Proof of BANDITFTRL-VR regret: Lipschitz case

FTRL analysis on averaged gradients with re-centering:

$$\sum_{t=1}^{T} \mathbb{E}\left[\bar{g}_t^\top (x_t - x_\epsilon^*)\right] \leq \frac{2C^2 d^2 \eta T}{\delta^2 (k+1)^2} + \frac{1}{\eta}\mathcal{R}(x_\epsilon^*).$$

Cumulative analysis:

$$\begin{aligned}\mathrm{Reg}_T(\text{BANDITFTRL-VR}) \leq \epsilon L T + 2L\delta D T + \frac{Ck}{2} + \frac{2C^2 d^2 \eta T}{\delta^2 (k+1)^2} + \frac{1}{\eta}\mathcal{R}(x_\epsilon^*) \\ + LT \sup_{\substack{t \in [1,T] \\ i \in [0, k \wedge t]}} \mathbb{E}\left[\|x_{t-i} - x_t\|_2\right].\end{aligned}$$

Proof of BANDITFTRL-VR regret: Lipschitz case

Stability estimate for the actions

- Want to bound: $\sup_{\substack{t\in[1,T] \\ i\in[0,k\wedge t]}} \mathbb{E}\left[\|x_{t-i} - x_t\|_2\right]$.
- Fact: Let D be the diameter of $\mathcal{K}$. For any $x \in \mathcal{K}$ and $z \in \mathbb{R}^d$,

$$D^{-1}\|z\|_{x,*} \le \|z\|_2 \le D\|z\|_x.$$

- By triangle inequality and equivalence of norms,

$$\mathbb{E}\left[\|x_{t-i} - x_t\|_2\right] \le \sum_{s=t-i}^{t-1} \mathbb{E}\left[\|x_s - x_{s+1}\|_2\right]$$

$$\le D \sum_{s=t-i}^{t-1} \mathbb{E}\left[\|x_s - x_{s+1}\|_{x_s}\right] \le D \sum_{s=t-i}^{t-1} 2\eta \mathbb{E}\left[\|\bar{g}_s + \tilde{g}_{s+1} - \tilde{g}_s\|_{x_s,*}\right].$$

Proof of BANDITFTRL-VR regret: Lipschitz case

- $\bar{g}_s + \tilde{g}_{s+1} - \tilde{g}_s = \frac{1}{k+1} \sum_{i=0}^{k} \widehat{g}_{s-i} + \frac{1}{k+1} \widehat{g}_s$
- Thus,

$$
\begin{aligned}
&\mathbb{E}\left[\|\bar{g}_s + \tilde{g}_{s+1} - \tilde{g}_s\|_{x_s,*}^2\right] \\
&\leq \frac{3}{k^2} \left\| \sum_{i=0}^{k-1} \mathbb{E}_{s-i}\left[\widehat{g}_{s-i}\right] \right\|_{x_s,*}^2 + \frac{3}{k^2} \mathbb{E}\left[\left\| \sum_{i=0}^{k-1} \widehat{g}_{s-i} - \mathbb{E}_{s-i}[\widehat{g}_{s-i}] \right\|_{x_s,*}^2 \right] + \frac{3}{k^2} L \\
&\leq \frac{3}{k^2} L + 2D^2 L^2 + \frac{3}{k^2} \mathbb{E}\left[\left\| \sum_{i=0}^{k-1} \widehat{g}_{s-i} - \mathbb{E}_{s-i}[\widehat{g}_{s-i}] \right\|_{x_s,*}^2 \right].
\end{aligned}
$$

Proof of BanditFTRL-VR regret: Lipschitz case

- Fact: $\forall 0 \le i \le k$ such that $t - i \ge 1$,

$$\frac{1}{2}\|z\|_{x_{t-i},*} \le \|z\|_{x_t,*} \le 2\|z\|_{x_{t-i},*}.$$

- Because the terms in the sum make up martingale difference,

$$\mathbb{E}\left[\left\|\sum_{i=0}^{k-1} \widehat{g}_{s-i} - \mathbb{E}_{s-i}[\widehat{g}_{s-i}]\right\|_{x_s,*}^2\right] \le 4\mathbb{E}\left[\left\|\sum_{i=0}^{k-1} \widehat{g}_{s-i} - \mathbb{E}_{s-i}[\widehat{g}_{s-i}]\right\|_{x_{s-k},*}^2\right]$$

$$\le 4\sum_{i=0}^{k-1} \mathbb{E}\left[\left\|\widehat{g}_{s-i} - \mathbb{E}_{s-i}[\widehat{g}_{s-i}]\right\|_{x_{s-k},*}^2\right]$$

$$\le 16\sum_{i=0}^{k-1} \mathbb{E}\left[\left\|\widehat{g}_{s-i} - \mathbb{E}_{s-i}[\widehat{g}_{s-i}]\right\|_{x_{s-i},*}^2\right]$$

$$\le 16\sum_{i=0}^{k-1} \mathbb{E}\left[\left\|\widehat{g}_{s-i}\right\|_{x_{s-i},*}^2\right] \le 16\sum_{i=0}^{k-1} \frac{C^2 d^2}{\delta^2} = 16k\frac{C^2 d^2}{\delta^2}.$$

Proof of BanditFTRL-VR regret: Lipschitz case

- By combining the components of the stability estimate,

$$\mathbb{E}\left[\|x_{t-i} - x_t\|_2\right] \le 2\eta D \sum_{s=t-i}^{t-1} \sqrt{\frac{3}{k^2} L + 2D^2 L^2 + \frac{3}{k^2} 16k \frac{C^2 d^2}{\delta^2}}.$$

- By the previous calculations,

$$\mathrm{Reg}_T(\text{BanditFTRL-VR}) \le \epsilon LT + 2L\delta DT + \frac{Ck}{2} + \frac{2C^2 d^2 \eta T}{\delta^2 (k+1)^2} + \frac{1}{\eta}\log(1/\epsilon) + LTD2\eta k \sqrt{\frac{3}{k^2} L + 2D^2 L^2 + \frac{48}{k^2} \frac{C^2 d^2}{\delta^2}}.$$

- Now set $\eta = T^{-11/16} d^{-3/8}$, $\delta = T^{-5/16} d^{3/8}$, $k = T^{1/8} d^{1/4}$.

Discussion of BANDITFTRL-VR regret: Lipschitz gradient case

- Approximation of real to smoothed losses incurs a $\delta^2 D^2 T$ penalty instead of δDT.
- Rest of analysis also leads to some changes in constants.
- General regret bound:

$$\mathrm{Reg}_T(\textsc{BanditFTRL-VR}) \leq \epsilon LT + H\delta^2 D^2 T + Ck + \frac{2C^2 d^2 \eta T}{\delta^2 (k+1)^2} + \frac{1}{\eta}\log(1/\epsilon) + (TL + DHT)2\eta kD\sqrt{\frac{3}{k^2}L + 2D^2L^2 + \frac{48}{k^2}\frac{C^2d^2}{\delta^2}}.$$

- Now set $\eta = T^{-8/13} d^{-5/6}$, $\delta = T^{-5/26} d^{1/3}$, $k = T^{1/13} d^{5/3}$.

Table of Contents

Other BCO methods

Strongly convex loss functions:

- Augment $\mathcal{R}$ in BANDITFTRL with additional regularization.
- $\mathcal{C}^{0,1}$ [Agarawal et al, 2010]: $\mathcal{O}(T^{2/3})$ regret
- $\mathcal{C}^{1,1}$ [Hazan & Levy, 2014]: $\mathcal{O}(T^{1/2})$ regret

Other types of algorithms:

- Ellipsoid method-based algorithm [Hazan and Li, 2016]: $\mathcal{O}(2^{d^4} \log(T)^{2d} T^{1/2})$.
- Kernel-based algorithm [Bubeck et al, 2017]: $\mathcal{O}(d^{9.5} T^{1/2})$

Table of Contents

Conclusion

- BCO is a flexible framework for modeling learning problems with sequential data and very limited feedback.
- BCO generalizes many existing models of online learning and optimization.
- State-of-the-art algorithms leverage techniques from online convex optimization and interior-point methods.
- "Efficient" algorithms obtaining optimal guarantees in $\mathcal{C}^{0,1}$, $\mathcal{C}^{1,1}$ cases are still open.