Advanced Machine Learning

Bandit Convex Optimization

MEHRYAR MOHRI MOHRI@

COURANT INSTITUTE & GOOGLE RESEARCH

Set-Up

- Convex set C.
- For $t = 1$ to T do
  - predict $\mathbf{x}_t \in C$.
  - receive convex loss function $f_t : C \to \mathbb{R}$.
  - incur loss $f_t(\mathbf{x}_t)$.
- Bandit setting: only loss revealed, no gradient information.
- Regret of algorithm $\mathcal{A}$:

$$R_T(\mathcal{A}) = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \inf_{\mathbf{x} \in C} \sum_{t=1}^{T} f_t(\mathbf{x}).$$

Single-Point Gradient Estimate

(Flaxman et al., 2005)

- Definitions:
  - $\mathbb{B} = \{\mathbf{x} \in \mathbb{R}^N \colon \|\mathbf{x}\| \leq 1\}$.
  - $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^N \colon \|\mathbf{x}\| = 1\}$.
  - $\widehat{f}(\mathbf{x}) = \operatorname*{E}_{\mathbf{v} \in \mathbb{B}}[f(\mathbf{x} + \delta \mathbf{v})]$: smoothed version of $f(\mathbf{x})$.
- Lemma: fix $\delta > 0$. Then, the following equality holds:

$$\operatorname*{E}_{\mathbf{u} \in \mathbb{S}}[f(\mathbf{x} + \delta \mathbf{u})\mathbf{u}] = \frac{\delta}{N} \nabla \widehat{f}(\mathbf{x}).$$

Proof

- By Stokes' theorem,

$$\nabla \int_{\delta\mathbb{B}} f(\mathbf{x}+\mathbf{v})d\mathbf{v} = \int_{\delta\mathbb{S}} f(\mathbf{x}+\mathbf{u})\frac{\mathbf{u}}{\|\mathbf{u}\|}d\mathbf{u}.$$

- Thus,

$$\begin{aligned}
\nabla \widehat{f}(\mathbf{x}) = \nabla \left[\frac{\int_{\delta\mathbb{B}} f(\mathbf{x}+\mathbf{v})d\mathbf{v}}{\mathrm{vol}_N(\delta\mathbb{B})}\right] &= \frac{\int_{\delta\mathbb{S}} f(\mathbf{x}+\mathbf{v})d\mathbf{v}}{\mathrm{vol}_N(\delta\mathbb{B})} \\
&= \frac{\int_{\delta\mathbb{S}} f(\mathbf{x}+\mathbf{v})d\mathbf{v}}{\mathrm{vol}_{N-1}(\delta\mathbb{S})}\frac{\mathrm{vol}_{N-1}(\delta\mathbb{S})}{\mathrm{vol}_N(\delta\mathbb{B})} \\
&= \mathop{\mathrm{E}}_{\mathbf{u}\in\mathbb{S}}[f(\mathbf{x}+\delta\mathbf{u})\mathbf{u}]\,\frac{N}{\delta}.
\end{aligned}$$

Algorithm

(Flaxman et al., 2005)

- Assume that C centered in the origin and let $C_\delta = \frac{1}{1-\delta} C$.

$\text{FKM}(T)$

$\mathbf{y}_1 \leftarrow \mathbf{0}$
for $t \leftarrow 1$ **to** T **do**
    $\mathbf{u}_t \leftarrow \textsc{Sample}(\mathbb{S})$
    $\mathbf{x}_t \leftarrow \mathbf{y}_t + \delta \mathbf{u}_t$
    $\textsc{loss} \leftarrow \textsc{Receive}(f_t(\mathbf{x}_t))$
    $\mathbf{g}_t \leftarrow \frac{N}{\delta} f_t(\mathbf{x}_t) \mathbf{u}_t$
    $y_{t+1} \leftarrow \Pi_{C_\delta}(\mathbf{y}_t - \eta \mathbf{g}_t)$

Analysis

- Assumptions:
  - $\text{diam}(C) \leq D$.
  - f_t bounded by M and G-Lipschitz.

- Theorem: the regret of the FKM algorithm is bounded by

$$\frac{D^2}{2\eta} + \frac{\eta M^2 N^2 T}{2\delta^2} + \delta(D+2)GT.$$

  - choosing $\eta = \frac{\delta D}{MN\sqrt{T}}$ and $\delta = \sqrt{\frac{DMN}{(D+2)G}\frac{1}{T^{\frac{1}{4}}}}$ yields the upper bound

$$2\sqrt{D(D+2)GMN}\,T^{\frac{3}{4}} = O(\sqrt{N}T^{\frac{3}{4}}).$$

Proof

- Let x^*_δ be the projection of x^* on C_δ, then $\|x^* - x^*_\delta\| \leq \delta D$. Thus, since f_ts are G-Lipschitz,

$$\sum_{t=1}^{T} \big(\mathrm{E}[f_t(\mathbf{x}_t)] - f_t(\mathbf{x}^*) \big)$$

$$= \sum_{t=1}^{T} \big(\mathrm{E}[f_t(\mathbf{x}_t)] - \mathrm{E}[\widehat{f}_t(\mathbf{x}_t)] + \mathrm{E}[\widehat{f}_t(\mathbf{x}_t)] - \widehat{f}_t(\mathbf{x}^*_\delta) + \widehat{f}_t(\mathbf{x}^*_\delta) - f_t(\mathbf{x}^*_\delta) + f_t(\mathbf{x}^*_\delta) - f_t(\mathbf{x}^*) \big)$$

$$\leq \sum_{t=1}^{T} \big(\mathrm{E}[\widehat{f}_t(\mathbf{x}_t)] - \widehat{f}_t(\mathbf{x}^*_\delta) \big) + 2\delta GT + \delta DGT$$

$$\leq \sum_{t=1}^{T} \big(\mathrm{E}[\widehat{f}_t(\mathbf{x}_t)] - \widehat{f}_t(\mathbf{x}^*_\delta) \big) + \delta(D+2)GT.$$

Proof

- Lemma: fix a sequence of convex and differentiable functions $u_1, \ldots, u_T \colon C \to \mathbb{R}$ and $\eta > 0$. Let $\mathbf{z}_0, \ldots, \mathbf{z}_T \in C$ be defined by $\mathbf{z}_0 = 0$ and $\mathbf{z}_{t+1} = \Pi_C(\mathbf{z}_t - \eta \mathbf{g}_t)$, where $\mathbf{g}_t$s are random variables such that
  - $\mathrm{E}[\mathbf{g}_t | \mathbf{z}_t] = \nabla u_t(\mathbf{z}_t)$ and $\|\mathbf{g}_t\| \leq G$; then,

$$\mathrm{E}\left[\sum_{t=1}^{T} u_t(\mathbf{z}_t)\right] - \min_{\mathbf{z} \in C} \sum_{t=1}^{T} u_t(\mathbf{z}) \leq \mathrm{E}[R_T(\mathrm{PSGD}, \mathbf{g}_1, \ldots, \mathbf{g}_T)].$$

- Proof: define h_t by $h_t(\mathbf{z}) = u_t(\mathbf{z}) + [\mathbf{g}_t - \nabla u_t(\mathbf{z}_t)] \cdot \mathbf{z}$. Then, $\nabla h_t(\mathbf{z}_t) = \mathbf{g}_t$, $\mathrm{E}[h_t(\mathbf{z}_t)] = \mathrm{E}[u_t(\mathbf{z}_t)]$ since $\mathrm{E}[\mathbf{g}_t | \mathbf{z}_t] = \nabla u_t(\mathbf{z}_t)$ and for any fixed $\mathbf{z}$, $\mathrm{E}[h_t(\mathbf{z})] = \mathrm{E}[u_t(\mathbf{z})]$. Thus, running deterministic PSGD on h_ts is equivalent to expected PSGD on the fixed functions u_ts.

Proof

- Regret bound for online projected gradient descent:

$$
\begin{aligned}
&\sum_{t=1}^{T} \left(\mathrm{E}[\widehat{f}_t(\mathbf{x}_t)] - \widehat{f}_t(\mathbf{x}^*_\delta) \right) \\
&\leq \sum_{t=1}^{T} \mathrm{E}\left[\mathbf{g}_t \cdot (\mathbf{x}_t - \mathbf{x}^*_\delta) \right] \\
&= \sum_{t=1}^{T} \frac{1}{2\eta} \mathrm{E}\left[\|\mathbf{x}_t - \mathbf{x}^*_\delta\|^2] + \eta^2 \|\mathbf{g}_t\|^2 - \|\mathbf{x}_t - \eta \mathbf{g}_t - \mathbf{x}^*_\delta\|^2 \right] \\
&\leq \sum_{t=1}^{T} \frac{1}{2\eta} \mathrm{E}\left[\|\mathbf{x}_t - \mathbf{x}^*_\delta\|^2] + \eta^2 M^2 \frac{N^2}{\delta^2} - \|\mathbf{x}_{t+1} - \mathbf{x}^*_\delta\|^2 \right] \qquad \text{(prop. of proj.)} \\
&\leq \frac{1}{2\eta} \mathrm{E}\left[\|\mathbf{x}_1 - \mathbf{x}^*_\delta\|^2] + \eta^2 M^2 \frac{N^2}{\delta^2} - \|\mathbf{x}_{T+1} - \mathbf{x}^*_\delta\|^2 \right] \\
&\leq \frac{1}{2\eta} \left[\|\mathbf{x}_1 - \mathbf{x}^*_\delta\|^2] + \eta^2 M^2 \frac{N^2}{\delta^2} T \right] \leq \frac{1}{2\eta} \left[D^2 + \eta^2 M^2 \frac{N^2}{\delta^2} T \right].
\end{aligned}
$$

References

- J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In COLT, pp. 263–274, 2008.
- Agarwal, A., O. Dekel, and L. Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In COLT, pp. 28–40, 2010.
- Bubeck, S. and R. Eldan. Multi-scale exploration of convex functions and bandit convex optimization. CoRR abs/1507.06580, 2015.
- V. Dani, T. Hayes, and S. Kakade. The price of bandit information for online optimization. In NIPS, 2008.
- Ofer Dekel, Ronen Eldan, Tomer Koren. Bandit Smooth Convex Optimization: Improving the Bias-Variance Tradeoff. NIPS 2015: 2926-2934.

References

- A. Flaxman, A. T. Kalai, and H. B. McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In SODA, pp. 385–394, 2005.
- Hazan, E. and K. Y. Levy (2014). Bandit convex optimiza- tion: Towards tight bounds. In NIPS, pp. 784–792.
- Saha, A. and A. Tewari (2011). Improved regret guaran- tees for online smooth convex optimization with bandit feedback. In AISTATS, pp. 636–642.
- Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In ICML, pages 928–936, 2009.