Stochastic Continuous Submodular Maximization: Boosting via Non-oblivious Function

Qixin Zhang[1,2] Zengde Deng[2] Zaiyi Chen[2] Haoyuan Hu[2]
Yu Yang[1]

[1]School of Data Science
City University of Hong Kong

[2]Cainiao Network

International Conference on Machine Learning, July 2022

Table of Contents

Motivation and Challenge

Frank-Wolfe Algorithm:

- Projection-free
- the optimal (1-1/e)-approximation
- Hard to extend: varience reduction(Mokhtari et al.(2018)), meta-action(Streeter and Golovin.(2008), Chen et al.(2018b))

Gradient Ascent/Descent Algorithm:

- Fast convergence
- 1/2-approximation
- Easy to extend in complicated settings, e.g., stochastic, online, and delay feedback(Quanrud and Khashabi, 2015).

In order to design efficient algorithm in the complex optimization scenarios, we naturally want to know

Problems

How to boost the Gradient Ascent from 1/2-approximation to (1-1/e)-approximation?

Non-oblivious Function

For any monotone continuous γ-weakly DR-submodular function $f : \mathcal{X} \rightarrow \mathbb{R}_{+}$, we consider this auxiliary function F whose gradient at point $\mathbf{x}$ allocates different weights to the gradient $\nabla f(z * \mathbf{x})$, i.e.,

$$\nabla F(\mathbf{x}) = \int_0^1 w(z) \nabla f(z * \mathbf{x}) \mathrm{d}z, \tag{1}$$

where $w(z)$ is the weight function in $[0, 1]$.

Lemma

For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, *we have*

$$\langle \mathbf{y} - \mathbf{x}, \nabla F(\mathbf{x}) \rangle \geq \left(\gamma \int_0^1 w(z) dz \right) \left(f(\mathbf{y}) - \theta(w) f(\mathbf{x}) \right),$$

where $\theta(w) = \max_{f,\mathbf{x}} \theta(w, f, \mathbf{x})$,
$\theta(w, f, \mathbf{x}) = \frac{w(1) + \int_0^1 (\gamma w(z) - w'(z)) \frac{f(z*\mathbf{x})}{f(\mathbf{x})} \mathrm{d}z}{\gamma \int_0^1 w(z) \mathrm{d}z}$ *and* $f(\mathbf{x}) > 0$.

The gradient ascent with small step size usually converges to the stationary point. According to the Lemma, the stationary point of auxiliary function F provides a $1/\theta(w)$-approximation guarantee.

Factor-revealing Optimization Problem

In order to find the optimal $w(z)$, we consider

$$
\begin{aligned}
\min_{w} \theta(w) = \min_{w} \max_{f,\boldsymbol{x}} & \frac{w(1) + \int_0^1 (\gamma w(z) - w'(z)) \frac{f(z*\boldsymbol{x})}{f(\boldsymbol{x})} \mathrm{d}z}{\gamma \int_0^1 w(z) \mathrm{d}z} \\
\text{s.t. } & w(z) \geq 0, \\
& w(z) \in C^1[0,1], \\
& f(\boldsymbol{x}) > 0, \\
& \nabla f(\boldsymbol{x}_1) \geq \gamma \nabla f(\boldsymbol{y}_1) \geq \boldsymbol{0}, \forall \boldsymbol{x}_1 \leq \boldsymbol{y}_1 \in \mathcal{X}.
\end{aligned}
\tag{2}
$$

The optimal weight function

At first glance, the factor-revealing problem looks challenging to solve. Fortunately, we could directly find the optimal solution, which is provided in the following theorem.

Theorem

For factor-revealing problem, $\hat{w}(z) = e^{\gamma(z-1)} \in \arg\min_w \theta(w)$ *and* $\min_w \theta(w) = \frac{1}{1-e^{-\gamma}}$.

Therefore, we take $w(z) = e^{\gamma(z-1)}$, the stationary point of F provides a (1-$e^{-\gamma}$)-approximation.

Unbiased Gradient Estimate of Non-oblivious function

We introduce a new random variable $\mathbf{Z}$ where $\Pr(\mathbf{Z} \leq z) = \int_0^z \frac{\gamma e^{\gamma(u-1)}}{1-e^{-\gamma}} I(u \in [0,1]) \mathrm{d}u$ where I is the indicator function. When the number z is sampled from r.v. $\mathbf{Z}$, we consider $\widetilde{\nabla} F(\boldsymbol{x}) = \frac{1-e^{-\gamma}}{\gamma} \widetilde{\nabla} f(z * \boldsymbol{x})$ as an estimator of $\nabla F(\boldsymbol{x})$.

Boosting Gradient Ascent

Algorithm 1 Boosting Gradient Ascent

1: Set $\triangle_t = 1$ when $t < T$ and $\triangle_T = 1 + \ln(\tau)$ where $\tau = \max(\frac{1}{\gamma}, \frac{r^2(\mathcal{C})L}{c})$.
2: Set $\triangle = \sum_{t=1}^{T} \triangle_t$
3: **Initialize** any $\boldsymbol{x}_1 \in \mathcal{X}$.
4: **for** $t \in [T]$ **do**
5: Compute $\widetilde{\nabla} F(\boldsymbol{x}_t)$ according to the Unbiased Estimate
6: Set $\boldsymbol{y}_{t+1} = \boldsymbol{x}_t + \eta_t \widetilde{\nabla} F(\boldsymbol{x}_t)$
7: $\boldsymbol{x}_{t+1} = \arg\min_{\boldsymbol{z} \in \mathcal{C}} \left\| \boldsymbol{z} - \boldsymbol{y}_{t+1} \right\|$
8: **end for**
9: Choose a number $l \in [T]$ with the distribution $\mathrm{P}(l = t) = \frac{\triangle_t}{\triangle}$

Boosting Gradient Ascent

Theorem

Assume $\mathcal{C} \in \mathcal{X}$ is a bounded convex set and f is L-smooth, and the gradient oracle $\widetilde{\nabla} f(\boldsymbol{x})$ is unbiased with $\mathbb{E}(\|\widetilde{\nabla} f(\boldsymbol{x}) - \nabla f(\boldsymbol{x})\|^2 | \boldsymbol{x}) \leq \sigma^2$. Let $\eta_t = \frac{1}{\frac{\sigma\gamma\sqrt{t}}{\mathrm{diam}(\mathcal{C})} + L_\gamma}$ and $c = O(1)$ in Algorithm 1, then we have

$$\mathbb{E}(f(\boldsymbol{x}_l)) \geq \left(1 - e^{-\gamma} - O\left(\frac{1}{T}\right)\right) OPT - O\left(\frac{1}{\sqrt{T}}\right),$$

where $OPT = \max_{\boldsymbol{x} \in \mathcal{C}} f(\boldsymbol{x})$.

Online Boosting Gradient Ascent with Delay

Algorithm 2 Online Boosting Delayed Gradient Ascent

1: **Initialize:** any $\boldsymbol{x}_1 \in \mathcal{C}$.
2: **for** $t \in [T]$ **do**
3: Play $\boldsymbol{x}_t$, then observe reward $f_t(\boldsymbol{x}_t)$
4: Query $\widetilde{\nabla} F_t(\boldsymbol{x}_t) = \frac{1-e^{-\gamma}}{\gamma}\widetilde{\nabla} f_t(z_t * \boldsymbol{x}_t)$
5: Receive feedback $\widetilde{\nabla} F_s(\boldsymbol{x}_s)$, where $s \in \mathcal{F}_t$
6: $\boldsymbol{y}_{t+1} = \boldsymbol{x}_t + \eta \sum_{s\in\mathcal{F}_t} \widetilde{\nabla} F_s(\boldsymbol{x}_s)$
7: $\boldsymbol{x}_{t+1} = \arg\min_{\boldsymbol{z}\in\mathcal{C}} \left\| \boldsymbol{z} - \boldsymbol{y}_{t+1} \right\|$
8: **end for**

Online Boosting Gradient Ascent with Delay

Theorem

Assume $\mathcal{C} \subseteq \mathcal{X}$ is a bounded convex set and each f_t is monotone, differentiable, and weakly DR-submodular with γ. Meanwhile, the gradient oracle is unbiased $\mathbb{E}(\widetilde{\nabla} f_t(\boldsymbol{x})|\boldsymbol{x}) = \nabla f_t(\boldsymbol{x})$ and $\max_{t\in[T]}(\|\widetilde{\nabla} F_t(\boldsymbol{x}_t)\|) = \frac{1-e^{-\gamma}}{\gamma}\max_{t\in[T]}(\|\widetilde{\nabla} f_t(\boldsymbol{x}_t)\|)$. Let $\eta = \frac{diam(\mathcal{C})}{\max_{t\in[T]}(\|\widetilde{\nabla} F_t(\boldsymbol{x}_t)\|)\sqrt{D}}$ in Algorithm 2, then we have

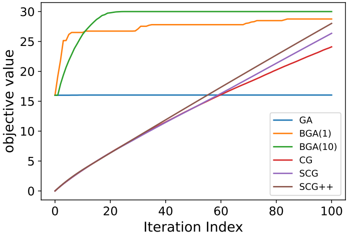
$$(1 - e^{-\gamma})\max_{\boldsymbol{x}\in\mathcal{C}}\sum_{t=1}^{T} f_t(\boldsymbol{x}) - \mathbb{E}(\sum_{t=1}^{T} f_t(\boldsymbol{x}_t)) = O(\sqrt{D}),$$

where $D = \sum_{i=1}^{T} d_t$ and $d_t \in \mathbb{Z}_+$ is a positive delay for the information about f_t.

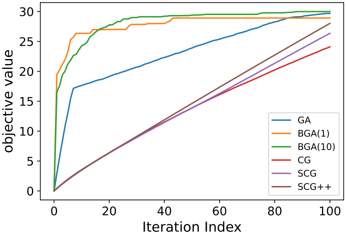
Special Case

Hassani(2017) introduced a special continuous DR-submodular function f_k coming from the multilinear extension of a set cover function. Also, they also verified that $\boldsymbol{x}_{loc} = (\underbrace{1, 1, \ldots, 1}_{k}, 0, \ldots, 0)$ is a local maximum with $(1/2 + 1/(2k))$-approximation to the global maximum. Then, we compare our boosting gradient ascent with the previous algorithms in this special case with Gaussian noise.

Special Case

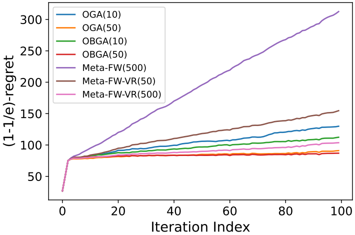


(a) Special Case (Local Max)

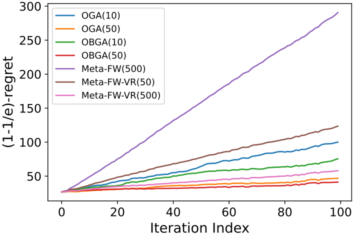


(b) Special Case (Origin)

Simulated Online Submodular QP



(c) Online QP with Delays



(d) Online QP without Delays

Connections with Filmus&Ward(2014)

We consider a continuous DR-submodular function $\bar{F} : [0,1]^n \to \mathbf{R}_+$ from the multilinear relaxation of a submodular set function $\bar{f} : 2^\Omega \to \mathbf{R}_+$,i.e., $\bar{F}(\boldsymbol{x}) = \sum_{S \in 2^\Omega} \bar{f}(S) \prod_{i \in S} x_i \prod_{j \in \Omega \setminus S} (1 - x_j)$. If taking the same boosting policy for $\bar{F}$, we could obtain a non-oblivious function $\bar{G}(\boldsymbol{x}) = \int_0^1 \frac{e^{z-1}}{z} \bar{F}(z * \boldsymbol{x}) dz$(Theorem 2). Also, Filmus&Ward(2014) define a non-oblivious set function $\bar{g}(A) = \sum_{B \subset A} m_{|A|-1,|B|-1} \bar{f}(B)$ for $\bar{f}$, where $m_{a,b} = \int_0^1 \frac{e^p}{e-1} p^b (1-p)^{a-b} dp$. By ignoring a constant factor, we could view $\bar{G}$ as the multilinear extension of $\bar{g}$.

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful comments. Zhang and Yang's research is supported in part by the Hong Kong Research Grants Council (ECS 21214720), City University of Hong Kong (Project 9610465) and Alibaba Group through Alibaba Innovative Research (AIR) Program.

Thank You!!