Optimal Private Streaming SCO in ℓ_p-geometry with Applications in High Dimensional Online Decision Making

Yuxuan Han [*1] Zhicong Liang [*1] Zhipeng Liang [*2] Yang Wang [1 2] Yuan Yao [1] Jiheng Zhang [1 2]

Abstract

Differentially private (DP) stochastic convex optimization (SCO) is ubiquitous in trustworthy machine learning algorithm design. This paper studies the DP-SCO problem with streaming data sampled from a distribution and arrives sequentially. We also consider the continual release model where parameters related to private information are updated and released upon each new data. Numerous algorithms have been developed to achieve optimal excess risks in different ℓ_p norm geometries, but none of the existing ones can be adapted to the streaming and continual release setting. We propose a private variant of the Frank-Wolfe algorithm with recursive gradients for variance reduction to update and reveal the parameters upon each data. Combined with the adaptive DP analysis, our algorithm achieves the first optimal excess risk in linear time in the case $1 < p \leq 2$ and the state-of-the-art excess risk meeting the non-private lower ones when $2 < p \leq \infty$. Our algorithm can also be extended to the case $p = 1$ to achieve nearly dimension-independent excess risk. While previous variance reduction results on recursive gradient have theoretical guarantee only in the i.i.d. setting, we establish such a guarantee in a non-stationary setting. To demonstrate the virtues of our method, we design the first DP algorithm for high-dimensional generalized linear bandits with logarithmic regret.

1. Introduction

Stochastic convex optimization (SCO) is a fundamental problem in machine learning, statistics, and operations research. The goal of SCO is to minimize a population loss function $F_P(\theta) = \mathbb{E}_{x\sim P}[f(\theta, x)]$ over a d-dimensional support set $\mathcal{C}$, with only access to the i.i.d. samples $\{x_t\}_{t=1}^n$ from some distribution P. The performance of an algorithm is measured in terms of the excess population risk of its solution θ, i.e., $F_P(\theta) - \min_{v\in\mathcal{C}} F_P(v)$. In practice, samples related to users' profiles might contain sensitive information; thus, it is important to solve stochastic convex optimization problems with differential privacy guarantees (DP-SCO) (Bassily et al., 2014; 2019; 2021a).

In this paper, we consider the DP-SCO with *streaming data*, where samples arrive sequentially and cannot be stored in memory for long. *Streaming data* has been studied in the context of online learning (Smale & Yao, 2006; Tarrès & Yao, 2014), online statistical inference (Vovk, 2001; 2009; Steinhardt et al., 2014; Fang et al., 2018), and online optimization (Cesa-Bianchi & Lugosi, 2006; Hazan, 2016; Hoi et al., 2021). In addition, parameter release is concerned due to privacy requirements. Our method can also accommodate *continual release* (Jain et al., 2021; Dwork et al., 2010; Chan et al., 2011), i.e., receives sensitive data as a stream of input and releases an output of it immediately after processing while satisfying DP requirements. A closely related setting considered as an extension in this paper is so-called online decision making (Slivkins, 2019; Lattimore & Szepesvári, 2020) where a decision needs to be made at each time, and the performance is measured in terms of accumulative regret, the gap between actual reward and the best possible reward, over the time. Recent work starts introducing streaming algorithms in (private) SCO into the solution of the online decision-making (Ding et al., 2021; Han et al., 2021) to enjoy high computational efficiency and flexibility to handle different reward structures. In particular, Han et al. (2021) proposes to solve private contextual bandits with stochastic gradient descent (SGD). However, the extension of other streaming algorithms, including the Frank-Wolfe and the stochastic mirror descent, remains elusive in this setting.

Compared with non-private SCO, private SCO depends on the dimension d inherently (Agarwal et al., 2012; Bassily et al., 2021b). Therefore, in DP-SCO, the optimal excess risk also has a crucial dependence on the space metric. Remarkable progress has been made in achieving optimal rate in ℓ_p norm with $1 \leq p \leq \infty$ as shown in Table 1. However,

[*] Equal contribution [1] Department of Mathematics [2] Department of Industrial Engineering and Decision Analytics. Correspondence to: Yuan Yao <yuany@ust.hk>, Jiheng Zhang <jiheng@ust.hk>.

Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022. Copyright 2022 by the author(s).

no existing rate-optimal DP-SCO algorithms can be adopted in the streaming and continual release setting. Previous works mainly rely on either Frank-Wolfe or mirror descent. Algorithms relying on Frank-Wolfe require a batch size of $\tilde{\Omega}(n)$ (Bassily et al., 2021b; Asi et al., 2021) for variance reduction, which is unacceptable in the streaming setting. Algorithms based on mirror descent require the same batch size and need a superlinear number of gradient query of $\tilde{\Omega}(n^{3/2})$ (Asi et al., 2021).

Our Contributions

We present a systematic study on the Frank-Wolfe algorithm with DP streaming data for various ℓ_p geometries ($1 \leq p \leq \infty$), as well as an application to private online decision making. Table 1 summarizes comparisons of our algorithm and the existing ones in terms of the excess population risk, the number of queried gradients, and required batch size. Detailed contributions are discussed below for different settings, followed by the application to a high dimensional generalized linear bandit problem with differential privacy. Note that in the streaming and continual release setting, the total time steps T equals the sample size n. So we will use n instead of T for the total number of iterations. Excess risk bounds denoted by t hold for every time step $t \in [n]$, while those denoted by n only hold after $\Omega(n)$ time steps.

Case of $1 < p \leq 2$**.** We generalize the recursive Frank-Wolfe algorithm proposed in (Xie et al., 2020) under non-private setting with ℓ_2 norm to private setting with general ℓ_p norm. The key observation is that the recursive variance reduction scheme can be written as a normalized incremental summation of gradients. According to this observation, we apply the tree-based mechanism as in Guha Thakurta & Smith (2013), and use an adaptive argument to show that noise with variance $\tilde{O}(\frac{1}{t^2\varepsilon^2})$ is enough to guarantee (ε, δ)-differential privacy. Such an analysis leads to a variance reduced gradient error bound of $\tilde{O}(\frac{1}{\sqrt{t}} + \frac{\sqrt{d}}{t\varepsilon})$ with high probability. The recursive gradient method we used here is closely related to Bassily et al. (2021b), while their algorithm uses $\frac{n}{2}$ samples for variance reduction, and their gradient error may be of the order $\tilde{O}(\frac{1}{\sqrt{n}} + \frac{\sqrt{d}}{\varepsilon n^{3/4}})$ in the worst-case. Our improvement on the variance reduction reduces their $\tilde{O}(\frac{1}{\sqrt{n}} + \frac{\sqrt{d}}{n^{3/4}\varepsilon})$ excess risk to $\tilde{O}(\frac{1}{\sqrt{t}} + \frac{\sqrt{d}}{t\varepsilon})$, which is optimal up to a logarithmic factor. Asi et al. (2021) achieves the optimal rates in terms of n and d at the cost of $O(n^{3/2})$ gradient queries while we achieve the same rate with only $O(n)$ gradient queries. Moreover, their rate will explode to $+\infty$ when p approaches 1, while our dependency on p is upper bounded by $\log d$.

Case of $2 < p \leq \infty$**.** The analysis above can be generalized to the case of $2 < p \leq \infty$. We achieve a rate of $\tilde{O}\big(\frac{d^{1/2-1/p}}{\sqrt{t}} + \frac{d^{1-1/p}}{t\varepsilon}\big)$, which matches the non-private lower bound $\Omega\big(\frac{d^{1/2-1/p}}{\sqrt{n}}\big)$ and is thus optimal when $d = \tilde{O}(n\varepsilon^2)$.

Case of $p = 1$**.** The challenge of this case is that the tree-based mechanism is no longer applicable to achieve a logarithmic dependence on d because the tree-based method will lead to an $O(\sqrt{d})$ factor. To overcome the difficulty, we combine the analysis of adaptive composition and Report Noisy Max mechanism (Dwork et al., 2014) to show that the noise with variance $O(\frac{\log d}{t\varepsilon})$ is enough to protect the privacy. Comparing with the rate-optimal algorithm with excess risk $O\big(\sqrt{\frac{\log d}{n}} + \big(\frac{\log d}{n\varepsilon}\big)^{2/3}\big)$ proposed in Asi et al. (2021), our algorithm can only achieve rate of $O(\sqrt{\frac{\log d}{t}} + \frac{\log d}{\sqrt{t\varepsilon}})$ in the streaming and continual release setting. We emphasize that such a gap is not due to the variance reduction analysis but the difficulty of the streaming setting. The analysis in Asi et al. (2021) relies on the privacy amplification via shuffling the dataset. However, storage of the whole dataset in the streaming setting is infeasible due to space constraints.

Strongly Convex. All the above results can be generalized to the case that population loss is strongly convex. Although it is appealing to use a folklore reduction from convex setting to strongly convex setting as in Asi et al. (2021); Feldman et al. (2020) to attain $O(\frac{1}{n})$ convergence rate, the reduction relies on the batch splitting. Specifically, a batch size in the order of $O(n/\log n)$ is required. However, the ground-truth time horizon n^* can hardly be known in advance in practice. Thus, one may need to overestimate the time horizon to ensure sufficient privacy protection. Once the estimated time horizon $n \gtrsim n^*/\log n^*$, the batch-based method will fail, and the last iteration only has the same guarantee as in the convex setting.

Private-Preserving Online Decision Making. A salient feature of our algorithm is that we provide $\tilde{O}(1/t)$ convergence guarantee for *each time step* while previous works (e.g., (Asi et al., 2021; Feldman et al., 2020)) can only hold after observing $\tilde{\Omega}(n)$ samples. Such a convergence result is not of purely intellectual interest. It is one of the foundations for extending our algorithm to the online decision-making setting. Despite the adaptivity of our algorithm to the streaming nature, it is highly non-trivial to extend the SCO guarantee to the online decision setting. The recursive gradient variance reduction method needs the stationary distribution assumption of coming data x_t. In contrast, the distribution of observation x_t will depend on the decision at step t in multi-arm bandit problems, thus varies over time t. By carefully analyzing the structure of bandit problems, we establish a new variance reduction guarantee that involves a total-variation term to describe the non-stationarity. Then we show that under suitable assumptions, such total-

Table 1. Bounds for excess population risk of (ε, δ)-DP-SCO. † denotes bounds in expectation while ‡ denotes bounds with high probability. And * denotes bounds without smoothness assumption. Here $\kappa = \min\{\frac{1}{p-1}, 2\log d\}$. Most of the DP-SCO algorithms require large batch size and thus fail to be adopted in streaming data, except for (Feldman et al., 2020). However their algorithm can only release the last variable for privacy protection and thus contradict to the requirements of continual release. We also compare our algorithms and some of the algorithms below empirically in Appendix C. And the results are summarized to Tabel 3 and 4.

Loss	ℓ_p	Theorem	Gradient Queries	Rate	Batch Size
Convex	$p = 2$	Thm. 3.2 (Bassily et al., 2019)	$O(\min\{n^{3/2}, \frac{n^{5/2}}{d}\})$	$O(\sqrt{\frac{1}{n}} + \frac{\sqrt{d}}{\epsilon n})^{\dagger}$	$O(\sqrt{\varepsilon n})$
		Thm. 3.5 (Feldman et al., 2020)	$O(\min\{n, \frac{n^2}{d}\})$	$O(\sqrt{\frac{1}{n}} + \frac{\sqrt{d}}{\epsilon n})^{\dagger}$	$O(\frac{\sqrt{d}}{\varepsilon})$
	$p = 1$	Thm. 7 (Asi et al., 2021)	$O(n)$	$\tilde{O}(\sqrt{\frac{\log d}{n}} + (\frac{\log d}{\varepsilon n})^{2/3})^{\dagger}$	$O(\frac{n}{\log^2 n})$
		Thm. 3.2 (Bassily et al., 2021b)	$O(n)$	$\tilde{O}(\frac{\log d}{\varepsilon\sqrt{n}})^{\dagger}$	$O(\frac{n}{2})$
		Theorem 3.12	$O(n)$	$\tilde{O}(\frac{\log d}{\varepsilon\sqrt{n}})^{\ddagger}$	1
	$1 < p < 2$	Thm. 5.4 (Bassily et al., 2021b)	$O(n)$	$\tilde{O}(\frac{\kappa}{\sqrt{n}} + \frac{\kappa\sqrt{d}}{\varepsilon n^{3/4}})^{\dagger}$	$O(\frac{n}{2})$
	$1 < p \leq 2$	Thm. 13 (Asi et al., 2021)	$O(n^{3/2})$	$\tilde{O}(\frac{1}{\sqrt{(p-1)n}} + \frac{\sqrt{d}}{(p-1)n\varepsilon})^{\dagger *}$	$O(\frac{n}{2})$
		Theorem 3.5	$O(n)$	$\tilde{O}(\sqrt{\frac{\kappa}{n}} + \frac{\sqrt{\kappa d}}{n\varepsilon})^{\ddagger}$	1
	$2 < p \leq \infty$	Prop. 6.1 (Bassily et al., 2021b)	$O(n^2)$	$\tilde{O}(\frac{d^{1/2-1/p}}{\sqrt{n}} + \frac{d^{1-1/p}}{\varepsilon n})^{\dagger *}$	$O(n)$
		Theorem 3.5	$O(n)$	$\tilde{O}(\frac{d^{1/2-1/p}}{\sqrt{n}} + \frac{d^{1-1/p}}{\varepsilon n})^{\ddagger}$	1
Strongly Convex	$p = 1$	Thm. 9 (Asi et al., 2021)	$O(n)$	$\tilde{O}(\frac{\log d}{n} + (\frac{\log d}{\varepsilon n})^{4/3})^{\dagger}$	$O(\frac{n}{2\log n})$
		Theorem 3.13	$O(n)$	$\tilde{O}(\frac{\log^2 d}{\varepsilon^2 n})^{\ddagger}$	1
	$1 < p \leq 2$	Theorem 3.8	$O(n)$	$\tilde{O}(\frac{\kappa}{n} + \frac{\kappa d}{\varepsilon^2 n^2})^{\ddagger}$	1
	$2 < p \leq \infty$	Theorem 3.8	$O(n)$	$\tilde{O}(\frac{d^{1-2/p}}{n} + \frac{d^{2-2/p}}{\varepsilon^2 n^2})^{\ddagger}$	1

variation term decays at a favorable rate to ensure the desired estimation error guarantee.

While our results generalize easily in the case of $1 < p \leq \infty$, we consider the high-dimensional (where $p = 1$) online decision-making problem (Bastani & Bayati, 2020), which has received lots of attention recently, to illustrate the generality of our method. While several remarkable progress has been made on the low-dimensional online decision-making problems with privacy guarantee recently (Chen et al., 2020; Shariff & Sheffet, 2018), no existing work provides sublinear regret bound in the high-dimensional setting even for linear rewards. Combining the new variance reduction guarantee mentioned above, we provide the first regret bound for DP high-dimensional generalized linear bandits.

2. Preliminaries

Notations. Let $(\mathbf{E}, \|\cdot\|)$ be a normed space of dimension d, and $\mathcal{C} \subseteq \mathbf{E}$ is a convex set of diameter D. Let $\langle\cdot\rangle$ be an arbitrary inner product over $\mathbf{E}$ (not necessarily inducing the norm $\|\cdot\|$). The dual norm over $\mathbf{E}$ is defined as $\|y\|_* := \max_{\|x\|\leq 1}\langle x, y\rangle$. With this definition, $(\mathbf{E}, \|\cdot\|_*)$ is also a d-dimensional normed space. We use $[K]$ to denote $\{1, 2, \cdots, K\}$ and for any $Z \in \mathbb{R}^d$ we denote $Z_{1:t} = \{Z_1, Z_2, \cdots, Z_t\}$. We denote $\mathbf{0}$ as an all-zero matrix whose size is adjusted according to the context. We adopt the standard asymptotic notations. For two non-negative sequences $\{a_n\}$ and $\{b_n\}$, we denote $\{a_n\} = O(\{b_n\})$ or $\{a_n\} \lesssim \{b_n\}$ iff $\limsup_{n\to\infty} a_n/b_n < \infty$, $a_n = \Omega(b_n)$ iff $b_n = O(a_n)$, and $a_n = \Theta(b_n)$ iff $a_n = O(b_n)$ and $b_n = O(a_n)$. We also use $\tilde{O}(\cdot)$, $\tilde{\Omega}(\cdot)$ and $\tilde{\Theta}(\cdot)$ to denote the respective meanings within multiplicative logarithmic factors in n and δ.

2.1. SCO with Streaming Data

Given a parameter set $\mathcal{C} \subset \mathbb{R}^d$, and an unknown distribution P over $\mathcal{X} \subset \mathbb{R}^d$ and a function $f : \mathcal{C} \times \mathcal{X} \to \mathbb{R}$, we consider the following optimization problem,

$$\min_{\theta\in\mathcal{C}} F_P(\theta) := \mathbb{E}_{x\sim P}[f(\theta, x)],$$

where F_P is assumed to be a convex function,

$$F_P(\theta) \geq F_P(\theta') + \langle\nabla F_P(\theta'), \theta - \theta'\rangle, \quad \forall\theta, \theta' \in \mathcal{C}.$$

We will abbreviate F_P as F when the context is clear for simplicity. In practice, the population loss $F(\cdot)$ is unknown and one can only access it via empirical approximation from a set of i.i.d. samples $\{x_i\}_{i=1}^n$. In the literature, the study of such SCO problems focuses on designing efficient algorithms to find a parameter θ over samples $\{x_i\}_{i=1}^n$ such that the excess population risk is acceptable.

In this work, we consider SCO under streaming and continual release setting. In each time t round, one sample $x_t \sim P$ arrives, and our algorithm needs to output a parameter θ_t with convergence guarantee regarding F. Here consider the following standard assumptions (e.g. Liang et al. (2019)).

Assumption 2.1 (Strongly-convex)**.** The population loss F is said to be μ-strongly convex if $\exists \mu \geq 0$, $F(\theta_1) \geq F(\theta_2) + \langle \nabla F(\theta_2), \theta_1 - \theta_2 \rangle + \dfrac{\mu}{2}\|\theta_1 - \theta_2\|^2$ for $\forall \theta_1, \theta_2 \in \mathcal{C}$.

Assumption 2.2 (Smoothness)**.** For any $\theta_1, \theta_2 \in \mathcal{C}$ and $x \in \mathcal{X}$, the loss function f is saied to be β-smooth if $\|\nabla f(\theta_1, x) - \nabla f(\theta_2, x)\|_* \leq \beta \|\theta_1 - \theta_2\|$.

Assumption 2.3. For any $\theta \in \mathcal{C}$ and $x \in \mathcal{X}$, the loss function f satisfies: $\|\nabla f(\theta, x) - \nabla F(\theta)\|_* \leq G$.

Assumption 2.4 (Lipschitz)**.** For any $\theta \in \mathcal{C}$ and $x \in \mathcal{X}$, the loss function f satisfies: $\|\nabla f(\theta, x)\|_* \leq L$.

2.2. Differential Privacy

Our work also extends to the privacy-preserving setting, where the sequence $(\theta_1, \ldots, \theta_n)$ satisfies the differential privacy constraint (see Definition 2.5) with respect to the data. Here we recall the definition of (ε, δ)-differential privacy.

Definition 2.5 (Differential Privacy (Dwork et al., 2014), (ε, δ)-DP)**.** A randomized algorithm $\mathcal{A}$ is said to be (ε, δ) differentially private if for any pair of datasets $\mathcal{D}$ and $\mathcal{D}'$ differing in one entry and any event $\mathcal{E}$ in the range of $\mathcal{A}$ it holds that $\mathbb{P}[\mathcal{A}(\mathcal{D}) \in \mathcal{E}] \leq e^{\varepsilon}\mathbb{P}[\mathcal{A}(\mathcal{D}') \in \mathcal{E}] + \delta$.

To design the DP-SCO algorithm under ℓ_p norm with $1 < p \leq 2$, we recall the generalized Gaussian mechanism proposed in (Bassily et al., 2021b) that leverages the regularity of the dual normed space.

Definition 2.6 (Regular Normed Space)**.** For a normed space $(\mathbf{E}, \|\cdot\|)$, we say that the norm $\|\cdot\|$ is κ-regular associated with $\|\cdot\|_+$, if there exists $1 \leq \kappa_+ \leq \kappa$ so that $\|\cdot\|_+$ is κ_+-smooth and $\|\cdot\|$ and $\|\cdot\|_+$ are equivalent with constant $\sqrt{\kappa/\kappa_+}$:

$$\|x\|^2 \leq \|x\|_+^2 \leq \frac{\kappa}{\kappa_+}\|x\|^2, \quad \forall x \in \mathbf{E}.$$

ℓ_q norm for $q \geq 1$ is a important class of regular norms, we specify the regularity constant κ_q and the associated smooth norm $\|\cdot\|_{q,+}$ later in Lemma 3.9 and Lemma 3.10.

Lemma 2.7 (Generalized Gaussian Distribution and Mechanism (Bassily et al., 2021b))**.** *Given a κ-regular norm $\|\cdot\|$ associated with smooth norm $\|\cdot\|_+$ in d-dimensional space, and the generalized Gaussian distribution $\mathcal{G}_{\|\cdot\|_+}(\mu, \sigma^2)$ with density:*

$$g(z; \sigma) = C(\sigma, d)\exp(-\|z - \mu\|_+^2/[2\sigma^2]),$$

where $C(\sigma, d) = \left(Area\{\|x\|_+ = 1\}\frac{(2\sigma^2)^{d/2}}{2}\Gamma(d/2)\right)^{-1}$, and Area is the $(d-1)$-dimension surface measure on $\mathbb{R}^d$, then for any function f with $\|\cdot\|$ sensitivity $s > 0$, we have that the mechanism output:

$$f + \mathcal{G}_{\|\cdot\|_+}(0, 2\kappa\log(1/\delta)s^2/\varepsilon^2)$$

is (ε, δ)-differentially private.

3. Differential Private SCO

3.1. ℓ_p-setup for $1 < p \leq \infty$

In this section, we provide a unified design analysis for optimization in ℓ_p geometry with $1 < p \leq 2$, which can be generalized to $2 < p \leq \infty$. As a consequence of the Hölder's inequality, the dual of ℓ_p norm is ℓ_q norm, where q satisfies $\frac{1}{p} + \frac{1}{q} = 1$, i.e., $q := \frac{p}{p-1}$.

Algorithm 1 DP-SCO with Streaming Data in ℓ_p-setup for $1 < p \leq \infty$.

```
Input: privacy parameters $(\varepsilon, \delta)$, $\{\rho_t\}_{t=1}^n = \{\eta_t\}_{t=1}^n = \frac{1}{1+t}$, $p$ considered in $\ell_p$, and its dual norm $\|\cdot\|_q$ associated with regular norm $\|\cdot\|_{q,+}$, initial point $\theta_0 = \theta_1 = 0 \in \mathcal{C}$
for $t = 1$ to $n$ do
    Compute and pass $g_t$ in Eq. (1) and $\sigma_+(q, \varepsilon, \delta)$ according to Theorem 3.1 into the tree-based mechanism (Algorithm 4).
    Get noisy summation $\tilde{G}_t = \text{noisy}(\sum_{i=1}^t g_i)$ from the tree-based mechanism (Algorithm 4).
    Set $d_t = \frac{1}{t+1} \cdot \tilde{G}_t$
    $v_t = \arg\min_{v \in \mathcal{C}} \langle d_t, v \rangle$.
    $\theta_{t+1} \leftarrow \theta_t + \eta_t(v_t - \theta_t)$.
end for
```

Our proposed algorithm is shown in Algorithm 1. At iteration t, we consider the following recursive gradient estimator d_t (Xie et al., 2020) as an unbiased estimator of the population gradient $\nabla F(\theta_t)$:

$$d_t = \nabla f(\theta_t, x_t) + (1 - \rho_t)(d_{t-1} - \nabla f(\theta_{t-1}, x_t)),$$

where $d_1 = \nabla f(\theta_1; x_1)$ and $\rho_t = \frac{1}{1+t}$.

A similar recursive gradient scheme is also used in Bassily et al. (2021b). However, they use additive noise to ensure the privacy of d_t at each iteration, which will accumulate linearly in t. To alleviate the influence of the noise induced by DP, they initialize d_1 with the first $\frac{n}{2}$ samples and begin to take mini-batch updates with batch size $\frac{\sqrt{n}}{2}$ for $\sqrt{n}$ iterations, which helps control the sensitivity of d_t and lower the number of noise accumulations. However, this strategy will lead to a gradient estimation error of $O(\frac{1}{n^{1/2}} + \frac{\sqrt{d}}{\varepsilon n^{3/4}})$. And it will also fail in the streaming setting where only one sample is available in initialization.

To improve the error rate and fit the streaming setting, our key observation is that the recursive gradient estimation d_t can be represented as the following summation of empirical gradients,

$$d_t = \frac{1}{t+1}\sum_{i=1}^{t} \underbrace{\big((i+1)\nabla f(\theta_i; x_i) - i\nabla f(\theta_{i-1}; x_i)\big)}_{g_i}. \tag{1}$$

Now we reduce the problem of privately releasing d_t in every step t to the problem of privately releasing the incremental summation of g_i in Eq. (1), which motivates us to apply the tree-based mechanism in Guha Thakurta & Smith (2013). In the tree-based mechanism, the leave nodes store the vectors g_i. Each internal node stores a private version of the summation of all the leaves in its sub-tree. In this case, any partial summation over g_i can be represented by at most $\lceil \log_2 n \rceil$ nodes. This critical property ensures that the noise induced by DP would not accumulate on d_t linearly in t. In this case, our algorithm fits in the streaming setting, where a relatively large number of iterations is required.

One difficulty of applying the tree-based mechanism is the sensitivity analysis. Suppose without loss of generality that for adjacent datasets $\mathcal{D} \sim \mathcal{D}'$, we have $x_1 \neq x_1'$. Such difference will affect the whole trajectory of the parameters: $\theta_i \neq \theta_i', \forall i \geq 2$. In other words, the sensitivity will be very large. Fortunately, we can show that such sensitivity can be dramatically reduced by the adaptive analysis similar to Guha Thakurta & Smith (2013). It turns out that noise with variance $\tilde{O}(\frac{1}{t^2\varepsilon^2})$ is enough to maintain (ε, δ)-differential privacy guarantee when reporting the t-th recursive gradient over the whole time horizon.

With the tree-based mechanism and the adaptive analysis mentioned above, we achieve a gradient error rate of $\tilde{O}(\frac{1}{\sqrt{n}} + \frac{\sqrt{d}}{\varepsilon t})$ (see Proposition 3.3). Furthermore, to report private incremental summation $\sum_{i=1}^{t} g_i$ for all $t \in [n]$, the amount of space required by the tree-based mechanism is $O(\log_2 n)$. Detailed description can be found in Algorithm 4 in Appendix.

In the following theorem, we characterize the privacy guarantee of Algorithm 1. The proof can be found in Appendix A.1.

Theorem 3.1 (Privacy Guarantee)**.** *Algorithm 1 is (ε, δ)-differentially private when σ_+^2 in is selected to be*

$$\sigma_+^2 = \frac{8(\lceil \log_2 n \rceil + 1)^2 \kappa_q \log((\lceil \log_2 n \rceil + 1)/\delta)(\beta D + L)^2}{\varepsilon^2}. \tag{2}$$

Existing results only concern the excess population risk in expectation (Bassily et al., 2021b), thus the moment information of generalized Gaussian mechanism is enough for their derivation. While in our high-probability analysis, the tail behaviour of generalized Gaussian mechanism is characterized.

Lemma 3.2 (Gamma Distribution)**.** *Assume that $Z \sim \mathcal{G}_{\|\cdot\|_+}(0, \sigma_+^2)$ in d-dimensional space, then $\|Z\|_+^2$ follows Gamma distribution $\Gamma(d/2, 2\sigma_+)$. Furthermore, $\|Z\|_+^2 - \mathbb{E}[\|Z\|_+^2]$ follows sub-Gamma$(2\sigma_+^4 d, 2\sigma_+^2)$, which implies that for any $\lambda > 0$, we have*

$$\mathbb{P}(\|Z\|_+^2 > \mathbb{E}[\|Z\|_+^2] + 2\sqrt{\sigma_+^4 d\lambda} + 2\sigma_+^2\lambda) \leq \exp(-\lambda).$$

As a result, we have the following high-probability variance reduction guarantee for the recursive gradient estimator. The proof of Lemma 3.2 can be found in Section A.2 while the proof of Proposition 3.3 is in Appendix A.3.

Proposition 3.3. *Under Assumption 2.2 and 2.3, with probability at least $1-\alpha$, for $t \in [n]$, Algorithm 1 satisfies:*

$$\|d_t - \nabla F(\theta_t)\|_q \lesssim \frac{(\sqrt{\kappa_q} + \sqrt{\log(1/\alpha)})(\beta D + G)}{\sqrt{t+1}} + \frac{\log n \cdot \sigma_+ \sqrt{d \log(\log n/\alpha)}}{t+1}.$$

Remark 3.4. Noticing that σ_+ is in scaling of $\tilde{O}(\frac{1}{\varepsilon})$, thus our gradient error for $1 < p \leq 2$ is in scaling of $\tilde{O}(\frac{1}{\sqrt{t}} + \frac{\sqrt{d}}{t\varepsilon})$, which improves over the $O(\frac{1}{\sqrt{n}} + \frac{\sqrt{d}}{\varepsilon n^{3/4}})$ in-expectation one in Bassily et al. (2021a) under the same condition.

Now we have the following convergence guarantee.

Theorem 3.5 (Convergence Guarantee for General Convexity)**.** *Consider Algorithm 1 with convex function F and assumptions 2.2 to 2.4, for $t \in [n]$, we have with probability at least $1-\alpha$,*

$$F(\theta_t) - F(\theta^*) \lesssim \frac{D(\beta D + G)\big(\sqrt{\kappa_q} + \sqrt{\log(n/\alpha)}\big)}{\sqrt{t}} + \frac{\log t\big(\beta D^2 + D\sigma_+\sqrt{d\log(\log n/\alpha)}\log n\big)}{t}.$$

Remark 3.6. Later, we will show that the result of Theorem 3.5 is nearly tight for $1 < p \leq 2$ and matches the best existing convergence rate for $2 < p \leq \infty$.

One known drawback of Frank-Wolfe is that its convergence rate is slow when the solution lies at the boundary, and it cannot be improved in general even the objection function is strongly convex (Lacoste-Julien & Jaggi, 2015; Garber & Hazan, 2015). In this case, additional assumption is necessary to improve the convergence rate of Frank-Wolfe in the strongly convex setting. In the following, we introduce a geometric assumption, which is typical for Frank-Wolfe in the strongly convex setting, even for the non-private case (Guélat & Marcotte, 1986; Lafond et al., 2015). Denoted by $\partial\mathcal{C}$ the boundary set of $\mathcal{C}$.

Assumption 3.7 ((Lafond et al., 2015))**.** There is a minimizer θ^* of F that lies in the interior of $\mathcal{C}$, i.e., $\gamma := \inf_{v\in\partial\mathcal{C}} \|v-\theta^*\| > 0$.

Theorem 3.8 (Convergence Guarantee for Strong Convexity)**.** *Consider Algorithm 1 with Assumptions 2.1, to 2.4 and 3.7, for $1 < p \le \infty$ and $t \in [n]$, we have with probability at least $1-\alpha$,*

$$\begin{aligned} & F(\theta_t) - F(\theta^*) \\ \lesssim & \frac{1}{\gamma^2\mu}\frac{D^2(\beta D + G)^2(\kappa_q + \log(n/\alpha))}{t} \\ & + \frac{1}{\gamma^2\mu}\frac{\left(\beta^2 D^4 + dD^2\sigma_+^2\log(\log n/\alpha)\log^2 n\right)\log n}{t^2}. \end{aligned}$$

Discussions about ℓ_p-setup for $1 < p \le 2$

Let recall the following lemma for $1 < p \le 2$.

Lemma 3.9 (Regularity for $q \ge 2$ (Bassily et al., 2021b))**.** *When $2 \le q \le \infty$, the ℓ_q norm is regular with*

$$\begin{aligned} &\kappa_q := \min\{q-1, e^2(\log d - 1)\}, \\ &\|\cdot\|_+ := \|\cdot\|_{\kappa_{q,+}}. \\ &\kappa_{q,+} := \min\{q-1, \log d - 1\} \end{aligned}$$

Now noticing that $q = \frac{p}{p-1} \in [2,\infty)$, we plug the κ_q in Lemma 3.9 into Theorem 3.5, Theorem 3.8 and Eq. (2) to get the convergence rate of Algorithm 1 when $1 < p \le 2$:

Convex:

$$F(\theta_t) - F(\theta^*) \lesssim \sqrt{\frac{\log(n/\alpha)}{t}} + \frac{\sqrt{d}\log(\log(n)/\delta)}{t\varepsilon} \quad (3)$$

Strongly Convex:

$$F(\theta_t) - F(\theta^*) \lesssim \frac{\log(n/\alpha)}{t} + \frac{d\log(\log(n)/\delta)}{t^2\varepsilon^2} \quad (4)$$

The bound in Eq. (3) is optimal, up to a logarithmic factor, comparing with the $\Omega(\frac{1}{\sqrt{t}} + \frac{\sqrt{d}}{t\varepsilon})$ lower bound shown in Bassily et al. (2021b) in the case of $1 < p \le 2$.

In strongly convex case, Eq. (4) is tight comparing with the $\Omega(\frac{1}{t} + \frac{d}{t^2\varepsilon^2})$ lower bound shown in Bassily et al. (2014) in the case of $p = 2$. And we conjecture that such bound is also tight for general $1 < p \le 2$. Deriving the corresponding lower bound is leaved for future exploration.

Table 2. SubOpt for Algorithm 1, NoisySFW and NoisySGD with $T = 2000$, $d = 10$ and $(1, 1/T)$-DP.

	Algorithm 1	NoisySFW	NoisySGD
$p = 1.5$	0.060 ± 0.032	0.89 ± 0.069	N.A.
$p = \infty$	0.058 ± 0.013	N.A.	0.038 ± 0.013
Complexity	$O(n)$	$O(n)$	$O(n^2)$

Discussions about ℓ_p-setup for $2 < p \le \infty$.

When $2 < p \le \infty$, we have $1 \le q < 2$ and the following lemma.

Lemma 3.10 (Regularity for $1 \le q < 2$)**.** *When $1 \le q < 2$, the ℓ_q norm is regular with*

$$\kappa_q = d^{1-2/p}, \quad \|\cdot\|_+ = d^{1/2-1/p}\|\cdot\|_2.$$

Despite noticing that regularity constant of ℓ_q norm has a worse dependence on d, we can still get a satisfactory convergence rate by plugging the constants in Lemma 3.10 to Theorem 3.5 and Theorem 3.8:

Convex:

$$F(\theta_t) - F(\theta^*) \lesssim d^{\frac{1}{2}-\frac{1}{p}}\sqrt{\frac{\log(n/\alpha)}{t}} + \frac{d^{1-\frac{1}{p}}\log(\log(n)/\delta)}{t\varepsilon}. \quad (5)$$

Strongly Convex:

$$F(\theta_t) - F(\theta^*) \lesssim d^{1-\frac{2}{p}}\frac{\log(n/\alpha)}{t} + \frac{d^{2-\frac{2}{p}}\log^2(\log(n)/\delta)}{t^2\varepsilon^2}. \quad (6)$$

Comparing with the optimal non-private lower bound $\Omega(\frac{d^{1/2-1/p}}{\sqrt{n}})$ (Agarwal et al., 2012) in convex setting when $2 < p \le \infty$, our result (5) nearly matches the optimal non-private rate and is optimal when $d = \tilde{O}(n\varepsilon^2)$.

The same private-SCO rate is also attained by Bassily et al. (2021b) using the the multi-pass noisy SGD in Bassily et al. (2020) for ℓ_2-setup. However, the multi-pass SGD has superlinear complexity while our algorithm incur linear complexity.

Numerical Comparisons

In addition to the theoretical results above, we conduct numerical experiments and compare Algorithm 1 with NoisySFW (Algorithm 3 in (Bassily et al., 2021b)), NoisySGD (Algorithm 2 in (Bassily et al., 2020)), and LocalMD (Algorithm 6 in (Asi et al., 2021)). Table 2 summarizes typical results with details in Appendix C, where SubOpt is an empirical estimation of suboptimality $(F(\theta_t) - F(\theta^*))/(F(\theta_0) - F(\theta^*))$. LocalMD (for $1 < p \le 2$) is left to Appendix C because its empirical performance is unacceptable. In summary, our algorithm has comparable or better empirical performance, wider application range, and significantly lower computation complexity. The code to reproduce our numerical results is shared in Github Repo.

3.2. ℓ_p-setup for $p = 1$

Algorithm 2 DP-SCO with Streaming Data in ℓ_p setup for $p = 1$.

```
Input: praivacy parameters $(\varepsilon, \delta)$, $\{\rho_t\}_{t=1}^n = \{\eta_t\}_{t=1}^n = \frac{1}{1+t}$, and initial point $\theta_0 = \theta_1 = 0 \in \mathcal{C}$.
for $t = 1$ to $n$ do
    if t=1 then
        $d_t = \nabla f(\theta_t, x_t)$.
    else
        $d_t = \nabla f(\theta_t, x_t) + (1-\rho_t)(d_{t-1} - \nabla f(\theta_{t-1}, x_t))$.
    end if
    $\forall v \in \mathcal{C}$, sample $\mathbf{n}_v^t \sim \text{Lap}\Big(\frac{4D(\beta D+L)\sqrt{\log n \cdot \log(1/\delta)}}{\varepsilon\sqrt{t}}\Big)$.
    $v_t = \arg\min_{v\in\mathcal{C}}(\langle d_t, v\rangle + \mathbf{n}_v^t)$.
    $\theta_{t+1} \leftarrow \theta_t + \eta_t(v_t - \theta_t)$.
end for
```

In this section, we consider the ℓ_p-setup for $p = 1$. In Algorithm 2, we combine the analysis of the adaptive composition, and the Report Noisy Max mechanism (Dwork et al., 2014) to ensure differential privacy, which reduces the $O(\sqrt{d})$ factor in the excess population risk incurred by the tree-based mechanism in Section 3.1. In the following, we characterize the privacy guarantee of Algorithm 2. The proof can be found in Appendix A.6.

Theorem 3.11 (Privacy Guarantee)**.** *Algorithm 2 is (ε, δ)-differentially private.*

Theorem 3.12 (Convergence Guarantee for General Convexity)**.** *Consider Algorithm 2 with convex function F, Assumption 2.2, 2.3, 2.4 and 3.7, for $t \in [n]$, we have with probability at least $1 - \alpha$,*

$$F(\theta_t) - F(\theta^*) \le \frac{3}{\sqrt{t+1}}(\beta D^2 + A),$$

where

$$A = 8D(\beta D + G)\sqrt{\log(8dn/\alpha)} + \dots + \frac{16D(\beta D + L)\log(4dn/\alpha)\sqrt{\log n \cdot \log(1/\delta)}}{\varepsilon}.$$

The gradient error in our algorithm (see Lemma A.7) is of the same rate $O(\frac{1}{n})$ as the one in Asi et al. (2021). Comparing with their excess population risk of $\tilde{O}(\sqrt{\frac{\log d}{n}} + (\frac{\log d}{n\varepsilon})^{2/3})$, our bound achieves the rate of $\tilde{O}(\sqrt{\frac{\log d}{t}} + \frac{\log d}{\sqrt{t}\varepsilon})$. However, the analysis in Asi et al. (2021) relies on the privacy amplification via shuffling the dataset, which is unacceptable in streaming setting. The proof of the above theorem can be found in Appendix A.7.

Theorem 3.13 (Convergence Guarantee for Strong Convexity)**.** *Consider Algorithm 2 with Assumption 2.1, 2.2, 2.3, 2.4 and 3.7, for $t \in [n]$, we have with probability at least $1 - \alpha$,*

$$F(\theta_t) - F(\theta^*) \le \frac{1}{t+1}\left(\frac{9(\beta D^2 + A)^2}{\gamma^2 \mu}\right),$$

where A is defined in Theorem 3.12.

The above theorem achieves a rate of $\tilde{O}(\frac{\log d}{t} + \frac{\log^2 d}{t\varepsilon})$ comparing with the rate of $\tilde{O}(\frac{\log d}{n} + (\frac{\log d}{n\varepsilon})^{4/3})$ in Asi et al. (2021), which relies on the privacy amplification via shuffling the dataset as we mentioned in the comment under Theorem 3.12. The proof of this theorem can be found in Appendix A.7.

4. Application in Generalized Linear Bandits

In this section, we consider the generalized contextual bandits with stochastic contexts, where a decision is made upon each new data (Li et al., 2017). Our proposed private Frank-Wolfe algorithm is potential to derive a satisfying estimator for smart decisions under a wide range of reward structures while providing sufficient privacy protection in this setting due to the streaming and continual release feasibility. However, we face some non-stationarity incurred by the decision process, which leads to a non-trivial difficulty when applying the recursive gradient for variance reduction. For the fluency of the presentation, we first formulate the contextual bandits model and further explain the difficulty in-depth.

At each time t, with individual-specific context X_t sampled from some distribution $\mathcal{P}$ on $\mathcal{X}$, the decision maker can take an action a_t from a finite set (arms) of size K to receive a reward randomly generated from the distribution depending on the context X_t and the chosen arm through its parameter θ_{a_t} via a generalized linear model: $r_t = \zeta(X_t^\top \theta_{a_t}^*) + \epsilon_t$, where $\zeta(\cdot)$ is an inverse link function. We assume the noise ϵ_t is sub-Gaussian (Wainwright, 2019) and conditional mean zero, i.e., $\mathcal{F}_t = \sigma(X_{1:t}, r_{1:t-1})$ and $\mathbb{E}[\epsilon_t|\mathcal{F}_t] = 0$. We use the standard notion of pseudo regret to measure the difference between expected rewards obtained by the action a_t and the best achievable expected reward in this round:

$$\text{Regret}(T) = \sum_{t=1}^{T} \zeta(X_t^\top \theta_{a_t^*}^*) - \zeta(X_t^\top \theta_{a_t}^*),$$

where $a_t^* = \text{argmax}_{i\in[K]} X_t^\top \theta_i^*$.

It is non-trivial to introduce the privacy guarantee in the design of the bandit algorithms. The standard notion of DP under continual observation would enforce to select almost the same action for different contexts and incur $\Omega(T)$ regret (Shariff & Sheffet, 2018). Here we utilize the more relaxed notion of *Joint Differential Privacy* under continuous observation (Shariff & Sheffet, 2018).

Definition 4.1 ((ε, δ)-Jointly Differential Privacy (JDP))**.** A randomized action policy $\mathcal{A} = (\mathcal{A}_t)_{t=1}^T$ is said to be (ε, δ)-jointly differentially private under continual observations if for any t and any pair of sequences $\mathcal{D}$ and $\mathcal{D}'$ differing in the t entry and any sequences of actions ranging from time $t+1$ to the end of sequence $\mathcal{E}_{>t}$, it holds for $\mathcal{A}_{>t}(\mathcal{D}) := (A_s(\mathcal{D}))_{s>t}$ that $\mathbb{P}[\mathcal{A}_{>t}(\mathcal{D}) \in \mathcal{E}_{>t}] \le e^{\varepsilon}\mathbb{P}[\mathcal{A}_{>t}(\mathcal{D}') \in \mathcal{E}_{>t}] + \delta$.

We present some standard assumptions in contextual bandits, and similar assumptions can be found in Bastani & Bayati (2020); Goldenshluger & Zeevi (2013); Bastani et al. (2020).

Assumption 4.2 (Optimal Arm Set)**.** We have a partition $[K] = K_{\text{sub}} \cup K_{\text{opt}}$, so that for every arm $i \in K_{\text{sub}}$,

$$P(i = \text{argmax}_{j\in[K]} X^\top \theta_j^*) = 0.$$

Moreover, we suppose there exists a $h_{sub} > 0$ such that

1. $\max_{i\in[K]} X^\top\theta_i^* - h_{sub} > X^\top\theta_j^* \quad \forall j \in K_{sub}, X \in \mathcal{X}$.
2. For $U_i := \{X | X^\top\theta_i^* - h_{sub} > \max_{j\neq i} X^\top\theta_j^*\}$ we have $P(X \in U_i) > u$ for some $u > 0$.

Assumption 4.3 (Eigenvalue)**.** We assume that $\mathbb{E}[XX^\top | X \in U_i] \succeq \lambda I_d, \forall i \in [K]$, for some $\lambda > 0$.

Assumption 4.4 (Margin Condition)**.** There exists a constant ℓ so that for the sets

$$\Gamma_i := \{\theta : \|\theta - \theta_i^*\|_1 \le \ell\}, \forall i \in U,$$

and for $\theta_i \in \Gamma_i, \forall i \in U$, we have,

$$P(X^\top\theta_{i_t}^* - \text{argmax}_{j\in K_{opt}, j\neq i_t^*} X^\top\theta_j^* \le h) \le \nu h,$$

where $i_t^* := \text{argmax}_{i\in K_{opt}} X_t^\top \theta_i^*$ for some $\nu > 0$.

Intuitively, Assumption 4.2 implies that the positive probability for the strictly optimality to holds. Assumptoin 4.4 prohibit small errors in the parameter estimation to incur wrong decision. Assumption 4.3 is necessary for the estimation error of the underlying parameter to decrease in a desirable rate ans similar assumptions have been adopted in (Han et al., 2021). These assumptions all holds for a very wide class of continuous and discrete covariate distributions (Bastani & Bayati, 2020; Han et al., 2021). Next we impose the standard regularity assumption on the reverse link function (Li et al., 2017; Ren et al., 2020; Chen et al., 2020) which includes widely-used linear model and logistic regression.

Assumption 4.5. There exist μ and β such that $0 < \mu \le \zeta'(z) \le \beta$ for any $|z| \le C$, where C is some given constant.

Based on the above assumptions, we design differentially private high-dimensional GLM bandits (Algorithm 3). Our algorithm follows the similar procedure of Bastani & Bayati (2020) to use two sets of estimators: the forced-sampling estimators $\{\theta_{t_0,j}\}_{j\in[K]}$ constructed using i.i.d. samples to select a pre-selected set of arms; and the all-sample estimators $\{\theta_{t,j}\}_{t>t_0, j\in[K]}$ to greedily choose the "best" arm among the pre-selected set. Another ingredient of our algorithm is the so-called synthetic update, i.e., adding the noisy all-zero contexts and zero rewards to the collected samples for the unselected arm. This ingredient is similar to Han et al. (2021) while they focus on local differential privacy. For our synthetic update, we have the following privacy guarantee and the proof is deferred to Appendix B.

Algorithm 3 Differential Private High Dimensional Bandit

```
Input: time horizon $T$; warm up period length $t_0$; privacy parameter $(\varepsilon, \delta)$, initial parameters $\theta_{0,i}, i \in [K]$
Initialize $\mathcal{I}_i = \emptyset$ for $i \in [K]$
for $i = 1$ to $K$ do
    for $t = 1$ to $t_0$ do
        Observe the context $X_{it_0+t}$.
        Pull arm $i$ and receive $r_{it_0+t}$.
        Add $(X_{it_0+t}, r_{it_0+t})$ to $\mathcal{I}_i$
        Update $\theta_{t,i}$ via running the $t$-th step of Algorithm 2 over $\mathcal{I}_i$ .
    end for
end for
for $t \ge Kt_0 + 1$ do
    Observe the context $X_t$.
    Compute the set of pre-selected arms:
        $\hat{K}_t = \{i \in [K] : \zeta(X_t^\top\theta_{t_0,i}) > \max_{j\in[K]} \zeta(X_t^\top\theta_{t_0,j}) - \frac{h_{sub}}{2}\}$
    Compute the greedy action
        $a_t = \text{argmax}_{a\in\hat{K}_t} \zeta(X_t^\top\theta_{t,i})$
    Select $a_t$-th arm and receive $r_t$.
    Add $(X_t, r_t)$ to $\mathcal{I}_{a_t}$. Add $(\mathbf{0}, \zeta(0))$ to $\mathcal{I}_i$ for $i \neq a_t$.
    Update $\theta_{t,i}$ via running the $t$-th step of Algorithm 2 over $\mathcal{I}_i$ for all $i \in [K]$.
end for
```

Theorem 4.6 (Privacy Guarantee)**.** *Algorithm 3 is (ε, δ)-JDP.*

Although it is natural to run Algorithm 2 for estimators for any arm $i \in [K]$, we are in fact facing various loss functions, say $F_t(\theta_t) := \mathbb{E}[\nabla f_t(\theta_{t,i}; x_{t,a_t}, y_t)|\mathcal{F}_{t-1}]$, for each time t. While all of the loss functions share the same minimizers θ_i^*, $d_{t-1} - \nabla f(\theta_{t-1,i}, x_t)$ in Algorithm 2 is not mean zero and thus the recursive gradient is not an unbiased estimator for the population gradient. As in the SCO setting, to show that the norm of the gradient estimation error $\Delta_t = d_t - \nabla F_t(\theta_{t,i})$ converges to zero sufficiently fast, we reformulate Δ_t as the sum of a sequence $\{\zeta_{t,\tau}\}_{\tau=1}^t$. Our SCO results enjoy the i.i.d. nature of the data and thus $\{\zeta_{t,\tau}\}_{\tau=1}^t$ is a martingale difference sequence which can be controlled by an Azuma-Hoeffding-type concentration inequality. In the bandits setting, after the forced-sampling period, the sample distribution for each arm evolves by time, and thus the sequence is no longer conditional mean zero. To overcome the difficulty, we develop a new lemma on bridging the gradient error to the total variance difference of distributions between each time step.

Lemma 4.7. *For each arm $i \in K_{opt}$, suppose that the greedy action begins to be picked at t_0, then for any $t > t_0$ we have with probability at least $1-\alpha$,*

$$\|\Delta_t\|_\infty \lesssim \sqrt{\log((d+T)/\alpha)}\Big(\frac{(MD+\beta)}{\sqrt{t}} + \dots$$
$$+ \frac{\beta DM}{t}\Big(C_{sc}(\frac{\alpha}{d+t_0})\sqrt{t_0} + \nu \sum_{\tau=t_0+1}^{t} \|\theta_{\tau-1,i} - \theta_i^*\|_1\Big)\Big)$$

where $C_{sc}(\alpha) = O(\log(dT/\alpha))$ is specified in the complete version in Lemma B.2.

Such lemma provides a guideline on tuning the warm-up stage length of the algorithm. In particular, it implies that polylog(T) length of warm-up is sufficient to get a $O(\frac{1}{\sqrt{t}})$-decayed gradient estimation error for each arm $i \in K_{\text{opt}}$ if the previous estimators converge to the underlying one at sufficiently fast rates. Such a low gradient estimation error is sufficient for the fast parameter convergence in the consequent time steps.

As far as we know, this is the first attempt to directly apply variance reduction in a non-stationary environment, which is sharply contrast to the previous solutions. In reinforcement learning (RL), as pointed out by (Papini et al., 2018), variance reduction can potentially improve much the sample efficiency since the collection of the samples requires the agent to interact with the environment, which could be costly. However, the sampling trajectories is generated by an RL algorithm. Thus the direct usage of the variance reduction also suffer from the changing distribution of the collected sample once their RL algorithm improves based on previous experience. This also applies to the bandits setting which shares the similar spirit in the data collection process. In overcome this, previous work (Sutton et al., 2016; Papini et al., 2018; Xu et al., 2020), mainly employ importance sampling to correct the distribution shift and construct an unbiased estimator for the policy gradient with respect to the snapshot policy. However, importance sampling is prone to high variance, e.g., (Thomas et al., 2015). In contrast, we carefully exploit the structure of our bandits problem and shows that the bias of our gradient estimator is implicitly self-corrected in a satisfying rate, which recovers the convergence rate in the i.i.d. setting in a painless manner in the algorithm design.

We prove the convergence rate of the estimation error by induction in Appendix B.1, and here we present the corresponding theorem.

Theorem 4.8 (Estimation Error). *For the full-sample estimator $\theta_{t,i}$, when $t \geq t_0$, for every arm $i \in K_{opt}$, we have with probability at least $1-\alpha$,*

$$\lambda\mu u\|\theta_{t,i} - \theta_i^*\|_1^2 \leq F_t(\theta_{t,i}) - F_t(\theta_i^*) \leq \frac{C_{in}(\alpha)}{t},$$

for some constant $C_{in}(\alpha) = O(\frac{\log^2(dT/\alpha)\log(T)}{\varepsilon^2})$ specified in Appendix B.1.

Now we are ready to present our regret bound by converting the estimation error to regret, whose formal proof is given in Appendix B.2.

Theorem 4.9 (Regret bound). *With probability at least $1-\alpha$, Algorithm 3 achieves the following regret bound*

$$\textit{Regret}(T) \leq t_0 + M^2\beta^2 C_{in}(\alpha/(4|K_{opt}|))\log(T) + \dots$$
$$+ 2M\beta\sqrt{C_{in}(\alpha/(4|K_{opt}|))\log(T)\log(4/\alpha)}$$
$$= O\left(\frac{\log^2(dT/\alpha)\log^2 T}{\varepsilon^2}\right).$$

Remark 4.10. This regret has a sublinear growth rate, and it is the first regret bound for DP high-dimensional generalized linear bandits. In particular, the upper bound above has only a poly-logarithmic growth concerning dimension d, as desired in high dimensional scenarios. Compared with the regret bound $O(\log^2(dT))$ without DP in Bastani & Bayati (2020), our upper bound contains an extra $O(\log^2 T)$ factor, which is due to our simplified proof to shed light on the main idea. We leave the refinement as future directions.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China / Research Grants Council Joint Research Scheme Grant HKUST635/20, Guangdong-Hong Kong-Macao Joint Laboratory for Data-Driven Fluid Mechanics and Engineering Applications, Hong Kong Research Grant Council (HKRGC) Grant 16308321, 16214121, 16208120, 16308518, ITF UIM/390, as well as awards from Smale Institute of Mathematics of Computation, Tencent AI Lab, Si Family Foundation, and Microsoft Research-Asia. Zhicong Liang is supported by Hong Kong PhD Fellowship Scheme by HKRGC.

References

Agarwal, A., Bartlett, P. L., Ravikumar, P., and Wainwright, M. J. Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization. *IEEE Transactions on Information Theory*, 58(5):3235–3249, 2012.

Asi, H., Feldman, V., Koren, T., and Talwar, K. Private stochastic convex optimization: Optimal rates in L1 geometry. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 393–403. PMLR, 18–24 Jul 2021.

Bassily, R., Smith, A., and Thakurta, A. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pp. 464–473. IEEE, 2014.

Bassily, R., Feldman, V., Talwar, K., and Guha Thakurta, A. Private stochastic convex optimization with optimal rates. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Bassily, R., Feldman, V., Guzmán, C., and Talwar, K. Stability of stochastic gradient descent on nonsmooth convex losses. In *Advances in Neural Information Processing Systems*, volume 33, pp. 4381–4391, 2020.

Bassily, R., Guzmán, C., and Menart, M. Differentially private stochastic optimization: New results in convex and non-convex settings. *Advances in Neural Information Processing Systems*, 34, 2021a.

Bassily, R., Guzman, C., and Nandi, A. Non-euclidean differentially private stochastic convex optimization. In Belkin, M. and Kpotufe, S. (eds.), *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 474–499. PMLR, 15–19 Aug 2021b.

Bastani, H. and Bayati, M. Online decision making with high-dimensional covariates. *Operations Research*, 68 (1):276–294, 2020.

Bastani, H., Bayati, M., and Khosravi, K. Mostly exploration-free algorithms for contextual bandits. *Management Science*, 67, 07 2020. doi: 10.1287/mnsc.2020.3605.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, learning, and games*. Cambridge university press, 2006.

Chan, T.-H. H., Shi, E., and Song, D. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):1–24, 2011.

Chen, X., Simchi-Levi, D., and Wang, Y. Privacy-preserving dynamic personalized pricing with demand learning. *Available at SSRN 3700474*, 2020.

Ding, Q., Hsieh, C.-J., and Sharpnack, J. An efficient algorithm for generalized linear bandit: Online stochastic gradient descent and thompson sampling. In *International Conference on Artificial Intelligence and Statistics*, pp. 1585–1593. PMLR, 2021.

Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pp. 715–724, 2010.

Dwork, C., Roth, A., et al. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

Fang, Y., Xu, J., and Yang, L. Online bootstrap confidence intervals for the stochastic gradient descent estimator. *The Journal of Machine Learning Research*, 19(1):3053–3073, 2018.

Feldman, V., Koren, T., and Talwar, K. Private stochastic convex optimization: optimal rates in linear time. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 439–449, 2020.

Garber, D. and Hazan, E. Faster rates for the frank-wolfe method over strongly-convex sets. In *International Conference on Machine Learning*, pp. 541–549. PMLR, 2015.

Goldenshluger, A. and Zeevi, A. A linear response bandit problem. *Stochastic Systems [electronic only]*, 3, 01 2013. doi: 10.1214/11-SSY032.

Guélat, J. and Marcotte, P. Some comments on wolfe's 'away step'. *Mathematical Programming*, 35(1):110–119, 1986.

Guha Thakurta, A. and Smith, A. (nearly) optimal algorithms for private online learning in full-information and bandit settings. *Advances in Neural Information Processing Systems*, 26:2733–2741, 2013.

Han, Y., Liang, Z., Wang, Y., and Zhang, J. Generalized linear bandits with local differential privacy. In *Advances in Neural Information Processing Systems*, volume 34, pp. 26511–26522, 2021.

Hazan, E. Introduction to online convex optimization. *Found. Trends Optim.*, 2(3–4):157–325, 2016. ISSN 2167-3888.

Hoi, S. C., Sahoo, D., Lu, J., and Zhao, P. Online learning: A comprehensive survey. *Neurocomputing*, 459:249–289, 2021.

Jain, P., Raskhodnikova, S., Sivakumar, S., and Smith, A. The price of differential privacy under continual observation. *arXiv preprint arXiv:2112.00828*, 2021.

Lacoste-Julien, S. and Jaggi, M. On the global linear convergence of frank-wolfe optimization variants. *Advances in neural information processing systems*, 28, 2015.

Lafond, J., Wai, H.-T., and Moulines, E. On the online frank-wolfe algorithms for convex and non-convex optimizations. *arXiv preprint arXiv:1510.01171*, 2015.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Li, L., Lu, Y., and Zhou, D. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pp. 2071–2080. PMLR, 2017.

Liang, Z., Wang, B., Gu, Q., Osher, S., and Yao, Y. Differentially private federated learning with laplacian smoothing. In *NeurIPS Workshop on Federated Learning for Data Privacy and Confidentiality, Vancouver, Canada, Dec. 8-14, 2019*, 2019.

Papini, M., Binaghi, D., Canonaco, G., Pirotta, M., and Restelli, M. Stochastic variance-reduced policy gradient. In *International conference on machine learning*, pp. 4026–4035. PMLR, 2018.

Pinelis, I. Optimum bounds for the distributions of martingales in banach spaces. *The Annals of Probability*, pp. 1679–1706, 1994.

Ren, Z., Zhou, Z., and Kalagnanam, J. R. Batched learning in generalized linear contextual bandits with general decision sets. *IEEE Control Systems Letters*, 2020.

Shariff, R. and Sheffet, O. Differentially private contextual linear bandits. *arXiv preprint arXiv:1810.00068*, 2018.

Slivkins, A. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019.

Smale, S. and Yao, Y. Online learning algorithms. *Foundation of Computational Mathematics*, 6(2):145–170, 2006.

Steinhardt, J., Wager, S., and Liang, P. The statistics of streaming sparse regression. *arXiv preprint arXiv:1412.4182*, 2014.

Sutton, R. S., Mahmood, A. R., and White, M. An emphatic approach to the problem of off-policy temporal-difference learning. *The Journal of Machine Learning Research*, 17 (1):2603–2631, 2016.

Tarrès, P. and Yao, Y. Online learning as stochastic approximations of regularization paths: Optimality and almost-sure convergence. *IEEE Transactions on Information Theory*, 60(9):5716–5735, 2014.

Thomas, P., Theocharous, G., and Ghavamzadeh, M. High-confidence off-policy evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

Vovk, V. Competitive on-line statistics. *International Statistical Review*, 69:213–248, 2001.

Vovk, V. On-line predictive linear regression. *The Annals of Statistics*, 37(3):1566–1590, 2009.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Xie, J., Shen, Z., Zhang, C., Wang, B., and Qian, H. Efficient projection-free online methods with stochastic recursive gradient. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 6446–6453, 2020.

Xu, P., Gao, F., and Gu, Q. Sample efficient policy gradient methods with recursive variance reduction. In *International Conference on Learning Representations*, 2020.

Appendix For Private Streaming SCO in ℓ_p geometry with Applications in High Dimensional Online Decision Making

A. Proof of Section 3

Algorithm 4 Private Tree based aggregation protocol (Guha Thakurta & Smith, 2013)

```
1: **Input:** $\langle z_1, z_2, .., z_n \in \mathbb{R}^d\rangle$ (in an online sequence), noise level $\sigma_+(q, \varepsilon, \delta)$
2: **Initialization**: Define a binary tree $A$ of size $2^{\lceil \log_2 n\rceil+1} - 1$ with leaves $z_1, z_2, ..., z_n$.
3: **Online Phase:** At each iteration $t \in [n]$, execute Steps 4 till 23
4: Accept $z_t$ from the data stream.
5: Let *path* $= \{z_t \to \cdots \to \text{root}\}$ be the path from $z_t$ to the root.
6: **Tree update:** Step 7 till 11
7: $\Lambda \leftarrow$ First node in *path* that is left child in the tree. Let *path*$_\Lambda = \{z_t \to \cdots \to \Lambda\}$.
8: **for** $\alpha$ **in** path **do**
9:    $\alpha \leftarrow \alpha + z_t$
10:   **If** $\alpha \in \text{path}_\Lambda$, then **then** $\alpha \leftarrow \alpha + \mathbf{n}$ where $\mathbf{n} \sim \mathcal{G}_{\|\cdot\|_q,+}(0, \sigma_+^2)$.
11: **end for**
12: **Output Private Partial Sum:** Step 13 till 23
13: Initial Vector $v \in \mathbb{R}^d$ to zero. Let $b \leftarrow \lceil \log_2 n\rceil + 1$-bit binary representation of $t$.
14: **for all** $i$ **in** $[\lceil \log_2 n\rceil + 1]$ **do**
15:   **if** bit $b_i = 1$ **then**
16:     **if** $i$-th node in *path* (denoted by *path*$(i)$) is a left child in $A$ **then**
17:       $v \leftarrow v + \textit{path}(i)$,
18:     **else**
19:       $v \leftarrow v + \text{left sibling}\big(\textit{path}(i)\big)$.
20:     **end if**
21:   **end if**
22: **end for**
23: **return** The noisy partial sum $v$.
```

A.1. Proof of Theorem 3.1

Proof of Theorem 3.1. We expend d_t as follow

$$\begin{aligned}
d_t &= \nabla f(\theta_t, x_t) + (1-\rho_t)(d_{t-1} - \nabla f(\theta_{t-1}, x_t)) \\
&= \sum_{i=1}^{t} \Bigg(\prod_{k=i+1}^{t} (1-\rho_k)\nabla f(\theta_i, x_i) - \prod_{k=i}^{t}(1-\rho_k)\nabla f(\theta_{i-1}, x_i)\Bigg) \\
&= \frac{1}{t+1}\sum_{i=1}^{t}\Big((i+1)\nabla f(\theta_i, x_i) - i\nabla f(\theta_{i-1}, x_i)\Big),
\end{aligned} \tag{7}$$

where the last inequality is due to the fact that $\rho_t = \frac{1}{t+1}$. If we consider the tree based mechanism in Algorithm 4, each sample x_i is involved in at most $\lceil \log_2 n\rceil + 1$ nodes in the tree. And all partial summations can also be determined by at most $\lceil \log_2 n\rceil$ nodes. The privacy analysis of the partial sum now reduces to the privacy analysis of the tree.

Suppose adjacent datasets $\mathcal{D}$ and $\mathcal{D}'$ differ by sample x_i and x_i', then for any sets $B = (B_1, B_2, ..., B_{2^{\lceil \log_2 n\rceil+1}-1})$ corresponding to the post-order traversal of the binary tree, it suffices to prove that

$$\mathbb{P}(A_1(\mathcal{D}) \in B_1, ..., A_{2^{\lceil \log_2 n\rceil+1}-1}(\mathcal{D}) \in B_{2^{\lceil \log_2 n\rceil+1}-1}) \leq e^{\varepsilon}\mathbb{P}(A_1(\mathcal{D}') \in B_1, ..., A_{2^{\lceil \log_2 n\rceil+1}-1}(\mathcal{D}') \in B_{2^{\lceil \log_2 n\rceil+1}-1}) + \delta.$$

For node A_m including x_i, suppose that it stores the summation $\sum_{j=k}^{l}\big((j+1)\nabla f(\theta_j,x_j)-j\nabla f(\theta_{j-1},x_j)\big)$, we have then conditioned on $A_1(\mathcal{D})=A_1(\mathcal{D}'),...,A_{m-1}(\mathcal{D})=A_{m-1}(\mathcal{D}')$, $\theta_j(\mathcal{D})=\theta_j(\mathcal{D}')=\theta_j,\forall j\le l$. Thus the difference between x_i and x_i' will cause the difference between

$$(i+1)\nabla f(\theta_i,x_i)-i\nabla f(\theta_{i-1},x_i)\quad\text{and}\quad(i+1)\nabla f(\theta_i,x_i')-i\nabla f(\theta_{i-1},x_i').$$

which has ℓ_q sensitivity $2(\beta D+L)$ because

$$\begin{aligned}&\|\big((i+1)\nabla f(\theta_i,x_i)-i\nabla f(\theta_{i-1},x_i)\big)-\big((i+1)\nabla f(\theta_i,x_i')-i\nabla f(\theta_{i-1},x_i')\big)\|_q\\ \le\ &2i\beta\|\theta_i-\theta_{i-1}\|_p+\|\nabla f(\theta_i,x_i)-\nabla f(\theta_i,x_i')\|_q\\ \le\ &2(\beta D+L).\end{aligned}$$

According to the above sensitivity, and the using the fact that $\|\cdot\|_{q,+}$ is $\kappa_{q,+}$-smooth, we can now apply the generalized Gaussian in Lemma 2.7. We add noise $\mathcal{G}_{\|\cdot\|_+}(0,8(\lceil\log_2 n\rceil+1)^2\kappa_q\log((\lceil\log_2 n\rceil+1)/\delta)(\beta D+L)^2/\varepsilon^2)$ independently to each node to ensure that each node is $(\varepsilon/(\lceil\log_2 n\rceil+1),\delta/(\lceil\log_2 n\rceil+1))$-differentially private.

We recall that each sample x_i is involved in at most $\lceil\log_2 n\rceil+1$ nodes in the tree. We denote the path from x_i to the root of the tree as $path_i$, where $|path_i|\le\lceil\log_2 n\rceil+1$. And here we use p to denote the density of $(A_1(\mathcal{D}),...,A_{2^{\lceil\log_2 n\rceil+1}-1}(\mathcal{D}))$ and p' for its counterpart regarding dataset $\mathcal{D}'$. Then for any $B=(B_1,B_2,...,B_{2^{\lceil\log_2 n\rceil+1}-1})$, we have

$$\begin{aligned}&\mathbb{P}(A_1(\mathcal{D})\in B_1,...,A_{2^{\lceil\log_2 n\rceil+1}-1}(\mathcal{D})=B_{2^{\lceil\log_2 n\rceil+1}-1})\\ =&\int_{B_1\times,...,\times B_{2^{\lceil\log_2 n\rceil+1}-1}}p(a_1,...,a_{2^{\lceil\log_2 n\rceil+1}-1})da_1...da_{2^{\lceil\log_2 n\rceil+1}-1}\\ =&\int\prod_{m\in path_i}p(a_m|a_1,...,a_{m-1})\cdot\prod_{m\notin path_i}p(a_m|a_1,...,a_{m-1})da_1...da_{2^{\lceil\log_2 n\rceil+1}-1}.\end{aligned}$$

Notice that for any $m\notin path_i$, $p(a_m|a_1,...,a_{m-1})=p'(a_m|a_1,...,a_{m-1})$. For $m\in path_i$,

$$\begin{aligned}\int_{B_m}p(a_m|a_1,...,a_{m-1})da_m&=\int_{B_m}p(a_m|a_1,...,a_{m-1})-1\wedge e^{\varepsilon/(\lceil\log_2 n\rceil+1)}p'(a_m|a_1,...,a_{m-1})da_m+\dots\\&\quad+\int_{B_m}1\wedge e^{\varepsilon/(\lceil\log_2 n\rceil+1)}p'(a_m|a_1,...,a_{m-1})da_m\\&\le\delta/(\lceil\log_2 n\rceil+1)+\int_{B_m}1\wedge e^{\varepsilon/(\lceil\log_2 n\rceil+1)}p'(a_m|a_1,...,a_{m-1})da_m.\end{aligned}$$

Applying the above inequality to any node in $path_i$, we have

$$\begin{aligned}&\int\prod_{m\in path_i}p(a_m|a_1,...,a_{m-1})\cdot\prod_{m\notin path_i}p(a_m|a_1,...,a_{m-1})da_1...da_{2^{\lceil\log_2 n\rceil+1}-1}\\ \le\ &e^{\varepsilon}\int\prod_{m\in path_i}p'(a_m|a_1,...,a_{m-1})\cdot\prod_{m\notin path_i}p'(a_m|a_1,...,a_{m-1})da_1...da_{2^{\lceil\log_2 n\rceil+1}-1}+\delta\\ =\ &e^{\varepsilon}\mathbb{P}(A_1(\mathcal{D}')\in B_1,...,A_{2^{\lceil\log_2 n\rceil+1}-1}(\mathcal{D}')=B_{2^{\lceil\log_2 n\rceil+1}-1})+\delta,\end{aligned}$$

which concludes the proof.

□

A.2. Proof of Lemma 3.2

Proof of Lemma 3.2. Since each Z_j are i.i.d. $\mathcal{G}_{\|\cdot\|_+}(0,\sigma_+^2)$, we have

$$\begin{aligned}\mathbb{P}(\|Z_j\|_+^2>\lambda)&=C(\sigma_+,d)\text{Area}\{\|x\|_+=1\}\int_{r^2>\lambda}r^{d-1}\exp(-\frac{r^2}{2\sigma_+^2})dr\\&=\frac{1}{2}C(\sigma_+,d)\text{Area}\{\|x\|_+=1\}\int_{r>\lambda}r^{d/2-1}\exp(-\frac{r}{2\sigma_+^2})dr.\end{aligned}$$

By

$$C(\sigma_+, d)\text{Area}\{\|x\|_+ = 1\} = \frac{1}{(2\sigma_+^2)^{d/2} \cdot \Gamma(d/2)/2},$$

we know that the tail of $\|Z_j\|_+^2$ is exactly the tail of $\Gamma(d/2, 2\sigma_+^2)$ at λ, which means $\|Z_j\|_+^2$ follows $\Gamma(d/2, 2\sigma_+^2)$. Thus $\|Z_j\|_+^2 - \mathbb{E}[\|Z_j\|_+^2]$ is subGamma$(2\sigma_+^4 d, 2\sigma_+^2)$, then the standard tail bound of sub-Gamma distribution implies

$$P(\|Z_j\|_+^2 > \mathbb{E}[\|Z_j\|_+^2] + 2\sqrt{\sigma_+^4 d\lambda} + 2\sigma_+^2\lambda) \leq \exp(-\lambda) \tag{8}$$

□

A.3. Proof of Proposition 3.3

Proposition A.1 (Azuma-Hoeffding inequality in regular space)**.** *Given the κ-smooth norm $\|\cdot\|$ and a vector-valued martingale difference sequence $\mathbf{d}_t$ with respect to $\{\mathcal{F}_t\}_t$, we have if*

$$\mathbb{E}[\exp(\|\mathbf{d}_t\|^2/\sigma_t^2)|\mathcal{F}_{t-1}] \leq \exp(1), \quad \forall t, \tag{9}$$

then

$$\mathbb{P}\Big(\Big\|\sum_{i=1}^{t}\mathbf{d}_i\Big\| \geq (\sqrt{2e\kappa} + \sqrt{2}\lambda)\Big(\sum_{i=1}^{t}\sigma_i^2\Big)^{1/2}\Big) \leq 2\exp(-\lambda^2/64).$$

We provide the a detailed version of Proposition 3.3 in the following proposition.

Proposition A.2. *We denote $\Delta_t = d_t - \nabla F(\theta_t)$. Assume Assumption 2.2 and 2.3, for $t \in [n]$, we have that with probability at least $1-\alpha$, Algorithm 1 will satisfies*

$$\|\Delta_t\|_q \leq (\sqrt{2e\kappa_q} + 8\sqrt{4\log(2/\alpha)})\frac{2(\beta D + G)}{\sqrt{t+1}} + \lceil\log_2 n\rceil\frac{\sigma_+}{t+1}\big(d + 2\sqrt{d\log(2\lceil\log_2 n\rceil/\alpha)} + 2d\log(2\lceil\log_2 n\rceil/\alpha)\big)^{1/2}.$$

Proof. We first reformulate $\Delta_t = d_t - \nabla F(\theta_t)$ as the sum of a martingale difference sequence. We denote M_t the set of node indices used when reporting d_t and Z the noise in the tree based mechanism in Algorithm 4 . For $t \geq 1$, we have

$$\begin{aligned}
\Delta_t &= \frac{1}{1+t}\sum_{j\in M_t} Z_j + \nabla f(\theta_t, x_t) + (1-\rho_t)(d_{t-1} - \nabla f(\theta_{t-1}, x_t)) - \nabla F(\theta_t) \\
&= \frac{1}{1+t}\sum_{j\in M_t} Z_j + (1-\rho_t)\Delta_{t-1} + \rho_t(\nabla f(\theta_t, x_t) - \nabla F(\theta_t)) + \dots \\
&\quad + (1-\rho_t)\big(\nabla f(\theta_t, x_t) - \nabla f(\theta_{t-1}, x_t) - (\nabla F(\theta_t) - \nabla F(\theta_{t-1}))\big) \\
&= \frac{1}{1+t}\sum_{j\in M_t} Z_j + \prod_{k=2}^{t}(1-\rho_k)\epsilon_1 + \sum_{\tau=2}^{t}\Big(\rho_\tau\prod_{k=\tau+1}^{t}(1-\rho_k)\big(\nabla f(\theta_\tau, x_\tau) - \nabla F(\theta_\tau)\big) + \dots \\
&\quad + \prod_{k=\tau}^{t}(1-\rho_k)\big(\nabla f(\theta_\tau, x_\tau) - \nabla f(\theta_{\tau-1}, x_\tau) - (\nabla F(\theta_\tau) - \nabla F(\theta_{\tau-1}))\big)\Big) \\
&\triangleq \frac{1}{t+1}\sum_{j\in M_t} Z_j + \zeta_{t,1} + \sum_{\tau=2}^{t}\zeta_{t,\tau}
\end{aligned} \tag{10}$$

Recall that $\Delta_1 = \nabla f(\theta_1, x_1) - \nabla F(\theta_1)$. And we observe that $\mathbb{E}[\zeta_{t,\tau}|\mathcal{F}_{\tau-1}] = 0$ where $\mathcal{F}_\tau$ is the σ-field generated by $\{x_1, x_2, ..., x_{\tau-1}\}$. Therefore, $\{\zeta_{t,\tau}\}_{\tau=1}^{t}$ is a martingale difference sequence. In what follows, we derive upper bounds of $\|\zeta_{t,\tau}\|_q$. We start by observing that for any $\tau = 1, 2, ..., t$,

$$\prod_{k=\tau}^{t}(1-\rho_k) = \prod_{k=\tau}^{t}(1-\frac{1}{k+1}) = \prod_{k=\tau}^{t}\frac{k}{k+1} = \frac{\tau}{t+1}. \tag{11}$$

We can bound $\|\zeta_{t,1}\|_q$:

$$\|\zeta_{t,1}\|_q \leq \frac{1}{t+1}\|\nabla f(\theta_1, x_1) - \nabla F(\theta_1)\|_q \leq \frac{G}{t+1} \triangleq c_{t,1},$$

where the second inequality follows from Assumption 2.3. For $\tau > 1$,

$$\begin{aligned}\|\zeta_{t,\tau}\|_q \leq \prod_{k=\tau}^{t}(1-\rho_k)\big(\|\nabla f(\theta_\tau, x_\tau) - \nabla f(\theta_{\tau-1}, x_\tau)\|_q + \|\nabla F(\theta_\tau) - \nabla F(\theta_{\tau-1})\|_q\big) + \dots \\ + \rho_\tau \prod_{k=\tau+1}^{t}(1-\rho_k)\|\nabla f(\theta_\tau, x_\tau) - \nabla F(\theta_\tau)\|_q \\ \leq 2\beta\|\theta_\tau - \theta_{\tau-1}\|_p \prod_{k=\tau}^{t}(1-\rho_k) + G\rho_\tau \prod_{k=\tau+1}^{t}(1-\rho_k) \qquad (12) \\ = 2\beta\eta_{\tau-1}\|v_{\tau-1} - \theta_{\tau-1}\|_p \prod_{k=\tau}^{t}(1-\rho_k) + G\rho_\tau \prod_{k=\tau+1}^{t}(1-\rho_k) \\ \leq \frac{2(\beta D + G)}{t+1} \triangleq c_{t,\tau},\end{aligned}$$

where the second inequality follows from Assumption 2.2 and 2.3, and the last inequality is due to $\eta_\tau = \rho_\tau$ and the definition of D. Now according to Proposition A.1, we have

$$\mathbb{P}\left(\Big\|\Delta_t - \frac{1}{1+t}\sum_{j\in M_t} Z_j\Big\|_q \geq (\sqrt{2e\kappa_q} + \sqrt{2}\lambda)\Big(\sum_{\tau=1}^{t} c_{t,\tau}^2\Big)^{1/2}\right) \leq 2\exp(-\lambda^2/64), \qquad (13)$$

We can bound $\sum_{\tau=1}^{t} c_{t,\tau}^2$ as

$$\sum_{\tau=1}^{t} c_{t,\tau}^2 = c_{t,1}^2 + \sum_{\tau=2}^{t} c_{t,\tau}^2 = \left(\frac{G}{t+1}\right)^2 + \sum_{\tau=2}^{t}\left(\frac{2\beta D + G}{t+1}\right)^2 \leq \sum_{\tau=1}^{t}\left(\frac{2\beta D + 2G}{t+1}\right)^2 \leq \frac{4(\beta D + G)^2}{t+1}.$$

Plugging the above bound into Eq. (13) and setting

$$\lambda = 8\sqrt{\log(2/\alpha_1)},$$

we have with probability at least $1 - \alpha_1$,

$$\Big\|\Delta_t - \frac{1}{t+1}\sum_{j\in M_t} Z_j\Big\|_q \leq (\sqrt{2e\kappa_q} + 8\sqrt{2\log(2/\alpha_1)})\frac{2(\beta D + G)}{\sqrt{t+1}}.$$

According to Lemma 3.2, we know that $\|Z_j\|_{q,+}^2$ follows Gamma distribution $\Gamma(d/2, \sigma_+)$. Selecting $\lambda = \log(\lceil\log_2 n\rceil/\alpha_2)$, and by $\mathbb{E}[\|Z_j\|_{q,+}^2] = \sigma_+^2 d$, we get with probability at least $1 - \alpha_2/\lceil\log_2 n\rceil$,

$$\|Z_j\|_{q,+}^2 \leq \sigma_+^2 d + 2\sigma_+^2\sqrt{d\log(\lceil\log_2 n\rceil/\alpha_2)} + 2\sigma_+^2 d\log(\lceil\log_2 n\rceil/\alpha_2).$$

Thus with probability at least $1 - \alpha_2$, we have

$$\max_{j\in M_t}\|Z_j\|_{q,+}^2 \leq \sigma_+^2 d + 2\sigma_+^2\sqrt{d\log(\lceil\log_2 n\rceil/\alpha_2)} + 2\sigma_+^2 d\log(\lceil\log_2 n\rceil/\alpha_2), \qquad (14)$$

here we use the fact that $|M_t| \leq \lceil\log_2 n\rceil$. Thus with probability at least $1 - \alpha_2$,

$$\begin{aligned}\Big\|\sum_{j\in M_t} Z_j\Big\|_{q,+} &\leq \lceil\log_2 n\rceil \max_{j\in M_t}\|Z_j\|_{q,+} \\ &\leq \lceil\log_2 n\rceil\sigma_+\big(d + 2\sqrt{d\log(\lceil\log_2 n\rceil/\alpha_2)} + 2d\log(\lceil\log_2 n\rceil/\alpha_2)\big)^{1/2}.\end{aligned}$$

According to the norm equivalent property in Definition 2.6, we have

$$\Big\| \sum_{j\in M_t} Z_j \Big\|_q \le \Big\| \sum_{j\in M_t} Z_j \Big\|_{q,+}.$$

As a result, by setting $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, we have with probability at least $1-\alpha$,

$$\|\Delta_t\|_q \le (\sqrt{2e\kappa_q} + 8\sqrt{4\log(2/\alpha)})\frac{2(\beta D + G)}{\sqrt{t+1}} + \frac{\lceil \log_2 n \rceil \sigma_+ \big(d + 2\sqrt{d\log(2\lceil \log_2 n\rceil/\alpha)} + 2d\log(2\lceil \log_2 n\rceil/\alpha)\big)^{1/2}}{t+1}.$$

□

A.4. Proof of Theorem 3.5

We provide a detailed version of Theorem 3.5 in the following Theorem.

Theorem A.3. *Consider Algorithm 1 with convex function F, Assumption 2.2, 2.3 and 2.4, for $t \in [n]$, we have with probability at least $1-\alpha$,*

$$\begin{aligned} F(\theta_t) - F(\theta^*) \le & \frac{2(\sqrt{2e\kappa_q} + 8\sqrt{4\log(2n/\alpha)})D(\beta D + G)}{\sqrt{t}} + \frac{(\log t + 1)\beta D^2}{2t} + \dots \\ & + \frac{\log t}{t}\cdot D\lceil \log_2 n\rceil \sigma_+ \big(d + 2\sqrt{d\log(2n\lceil \log_2 n\rceil/\alpha)} + 2d\log(2n\lceil \log_2 n\rceil/\alpha)\big)^{1/2}. \end{aligned} \tag{15}$$

Proof. We start from β-smoothness:

$$\begin{aligned} F(\theta_{t+1}) &\le F(\theta_t) + \langle \nabla F(\theta_t), \theta_{t+1} - \theta_t\rangle + \frac{\beta}{2}\|\theta_{t+1} - \theta_t\|_p^2 \\ &\le F(\theta_t) + \eta_t\langle \nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2 \beta D^2}{2}. \end{aligned}$$

We subtract $F(\theta^*)$ from both sides, and denote $h_t = F(\theta_t) - F(\theta^*)$. We have

$$\begin{aligned} h_{t+1} &\le h_t + \eta_t\langle \nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\ &= h_t + \eta_t\langle \nabla F(\theta_t) - d_t, v_t - \theta_t\rangle + \eta_t\langle d_t, v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\ &\le h_t + \eta_t\langle \nabla F(\theta_t) - d_t, v_t - \theta_t\rangle + \eta_t\langle d_t, \theta^* - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\ &= h_t + \eta_t\langle d_t - \nabla F(\theta_t), \theta^* - v_t\rangle + \eta_t\langle \nabla F(\theta_t), \theta^* - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\ &\le h_t + \eta_t D\|d_t - \nabla F(\theta_t)\|_q + \eta_t\langle \nabla F(\theta_t), \theta^* - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\ &\le (1-\eta_t)h_t + \eta_t D\|d_t - \nabla F(\theta_t)\|_q + \frac{\eta_t^2\beta D^2}{2}. \end{aligned}$$

where the second inequality is due to definition of v_t. According to Proposition A.2, with probability at least $1 - t\alpha'$, we have

$$\begin{aligned}
h_{t+1} &\leq (1-\eta_t)h_t + \frac{\beta D^2}{2(t+1)^2} + \dots \\
&\quad + \frac{1}{t+1} D\Big(\frac{2(\sqrt{2e\kappa_q} + 8\sqrt{4\log(2/\alpha')})(\beta D + G)}{\sqrt{t+1}} + \dots \\
&\qquad + \frac{\lceil \log_2 n \rceil \sigma_+ \big(d + 2\sqrt{d\log(2\lceil \log_2 n\rceil/\alpha')} + 2d\log(2\lceil \log_2 n\rceil/\alpha')\big)^{1/2}}{t+1} \Big) \\
&\leq (1-\eta_t)h_t + \frac{1}{(t+1)^{3/2}} \underbrace{2(\sqrt{2e\kappa_q} + 8\sqrt{4\log(2/\alpha')})D(\beta D + G)}_{C_1} + \dots \\
&\quad + \frac{1}{(t+1)^2}\Big(\underbrace{D\lceil \log_2 n\rceil \sigma_+ \big(d + 2\sqrt{d\log(2\lceil \log_2 n\rceil/\alpha')} + 2d\log(2\lceil \log_2 n\rceil/\alpha')\big)^{1/2} + \beta D^2/2}_{C_2} \Big).
\end{aligned}$$

Then we have

$$\begin{aligned}
h_{t+1} &= (1-\eta_t)h_t + \frac{C_1}{(t+1)^{3/2}} + \frac{C_2}{(t+1)^2} \\
&= h_1 \prod_{\tau=1}^{t}(1-\eta_\tau) + \sum_{k=1}^{t}\left(\frac{C_1}{(k+1)^{3/2}} + \frac{C_2}{(k+1)^2} \right) \prod_{\tau=k+1}^{t}(1-\eta_\tau) \\
&= \frac{1}{t+1}h_1 + \sum_{k=1}^{t}\left(\frac{C_1}{(k+1)^{3/2}} + \frac{C_2}{(k+1)^2} \right) \prod_{\tau=k+1}^{t}(1-\eta_\tau) \\
&= \frac{1}{t+1}h_1 + \frac{1}{t+1}\sum_{k=1}^{t}\left(\frac{C_1}{(k+1)^{1/2}} + \frac{C_2}{(k+1)} \right) \\
&\leq \frac{1}{t+1}h_1 + \frac{C_1}{\sqrt{t+1}} + \frac{C_2 \log t}{t+1}.
\end{aligned}$$

Now setting $\alpha' = \frac{\alpha}{n}$, and recalling that $h_1 \leq \frac{\beta D^2}{2}$ according to β-smoothness lead to the desired result. □

A.5. Proof of Theorem 3.8

We firstly introduce the following lemma.

Lemma A.4 (Lemma 6 in (Lafond et al., 2015)). *Assume Assumption 3.7, and the population loss function F satisfies Assumption 2.1 and 2.2, then*

$$\left(\max_{\theta\in\mathcal{C}} \langle \nabla F(\theta_t), \theta_t - \theta\rangle \right)^2 \geq 2\mu\gamma^2 h_t \quad \textit{and} \quad \beta D^2 \geq \gamma^2\mu.$$

where $h_t = F(\theta_t) - F(\theta^*)$.

We provide a detailed version of Theorem 3.8 in the following Theorem.

Theorem A.5. *Consider Algorithm 1 with Assumption 2.1, 2.2, 2.3, 2.4 and 3.7, for $t \in [n]$, we have with probability at least $1-\alpha$,*

$$\begin{aligned}
F(\theta_t) - F(\theta^*) &\leq \frac{18}{\gamma^2\mu} \frac{4D^2(\sqrt{2e\kappa_q} + 8\sqrt{4\log(2n/\alpha)})^2(\beta D + G)^2}{t+1} + \dots \\
&\quad + \frac{18}{\gamma^2\mu} \frac{\Big(D\lceil \log_2 n\rceil \sigma_+ \big(d + 2\sqrt{d\log(2n\lceil \log_2 n\rceil/\alpha)} + 2d\log(2T\lceil \log_2 n\rceil/\alpha)\big)^{1/2} + \beta D^2/2 \Big)^2 \log n}{(t+1)^2}.
\end{aligned}$$

Proof. We denote $h_t = F(\theta_t) - F(\theta^*)$, and $\tilde{\theta}_t := \arg\max_{\theta\in\mathcal{C}}(\langle \nabla F(\theta_t), \theta_t - \theta\rangle)^2$ in Lemma A.4. We start from

β-smoothness:

$$\begin{aligned}
h_{t+1} &\leq h_t + \eta_t\langle\nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\
&= h_t + \eta_t\langle d_t, v_t - \theta_t\rangle - \eta_t\langle d_t - \nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\
&\leq h_t + \eta_t\langle d_t, \tilde{\theta}_t - \theta_t\rangle - \eta_t\langle d_t - \nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\
&= h_t + \eta_t\langle \nabla F(\theta_t), \tilde{\theta}_t - \theta_t\rangle + \eta_t\langle d_t - \nabla F(\theta_t), \tilde{\theta}_t - v_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\
&\leq h_t + \eta_t\|d_t - \nabla F(\theta_t)\|_q D - \eta_t\gamma\sqrt{2\mu h_t} + \frac{\eta_t^2\beta D^2}{2}
\end{aligned} \tag{16}$$

where the first inequality is due to the definition of v and the last inequality comes from Lemma A.4. According to Proposition A.2, with probability at least $1 - t\alpha'$, we have

$$\begin{aligned}
h_{t+1} \leq \sqrt{h_t}(\sqrt{h_t} - \eta_t\gamma\sqrt{2\mu}) &+ \frac{1}{(t+1)^{3/2}}\underbrace{2D(\sqrt{2e\kappa_q} + 8\sqrt{4\log(2/\alpha')})(\beta D + G)}_{C_1} + \ldots \\
&+ \frac{1}{(t+1)^2}\Big(\underbrace{D\lceil\log_2 n\rceil\sigma_+\big(d + 2\sqrt{d\log(2\lceil\log_2 n\rceil/\alpha')} + 2d\log(2\lceil\log_2 n\rceil/\alpha')\big)^{1/2} + \beta D^2/2}_{C_2}\Big) \\
= \sqrt{h_t}(\sqrt{h_t} - \eta_t\gamma\sqrt{2\mu}) &+ \frac{1}{(t+1)^{3/2}}C_1 + \frac{1}{(t+1)^2}C_2,
\end{aligned} \tag{17}$$

Now the claim holds by induction. We assume that

$$h_t \leq \frac{1}{t+1}\cdot\frac{18C_1^2}{\gamma^2\mu} + \frac{1}{(t+1)^2}\cdot\frac{18C_2^2\log^2 n}{\gamma^2\mu} \triangleq \frac{1}{t+1}A + \frac{1}{(t+1)^2}B.$$

For $t = 1$, according to Eq. (17), we have

$$h_2 \leq h_1 + \frac{C_1}{2\sqrt{2}} + \frac{C_2}{4} \leq \frac{9C_1^2}{\gamma^2\mu} + \frac{9C_2^2}{2\gamma^2\mu},$$

where the second inequality comes from Lemma A.4 that $\beta D^2 \geq \gamma^2\mu$.

For $t \geq 1$. There are two cases.

Case 1. $\sqrt{h_t} - \eta_t\gamma\sqrt{2\mu} \leq 0$:

Since $\eta = \frac{1}{t+1}$, Eq. (17) yields,

$$h_{t+1} \leq \frac{1}{(t+1)^{3/2}}C_1 + \frac{1}{(t+1)^2}C_2 \leq \frac{C_1^2}{\gamma^2\mu(t+1)^{3/2}} + \frac{C_1^2}{\gamma^2\mu(t+1)^2} \leq \frac{1}{t+1}\cdot\frac{18C_1^2}{\gamma^2\mu} + \frac{1}{(t+1)^2}\cdot\frac{18C_2^2\log^2 n}{\gamma^2\mu}.$$

where the second inequality comes from Lemma A.4 that $\beta D^2 \geq \gamma^2\mu$.

Case 2. $\sqrt{h_t} - \eta_t\gamma\sqrt{2\mu} > 0$:

According to Eq. (17) and the assumption that $h_t \le \frac{A}{t+1} + \frac{B}{(t+1)^2}$, we have

$$\begin{aligned}
&h_{t+1} - \frac{A}{t+2} - \frac{B}{(t+2)^2} \\
&\le A\Big(\frac{1}{t+1} - \frac{1}{t+2}\Big) + B\Big(\frac{1}{(t+1)^2} - \frac{1}{(t+2)^2}\Big) + \dots \\
&\quad + \frac{C_1}{(t+1)^{3/2}} + \frac{C_2}{(t+1)^2} - \frac{\gamma}{t+1}\sqrt{2\mu\Big(\frac{A}{t+1} + \frac{B}{(t+1)^2}\Big)} \\
&= \frac{A}{(t+1)^2} + \frac{2B}{(t+1)^3} + \frac{C_1}{(t+1)^{3/2}} + \frac{C_2}{(t+1)^2} - \frac{\gamma}{2(t+1)^{3/2}}\sqrt{2\mu A} - \frac{\gamma}{2(t+1)^2}\sqrt{2\mu B} \qquad (18)\\
&\le \frac{A}{(t+1)^2} + \frac{2B}{(t+1)^3} + \frac{C_1}{(t+1)^{3/2}} + \frac{C_2}{(t+1)^2} - \frac{3C_1}{(t+1)^{3/2}} - \frac{3C_2 \log n}{(t+1)^2} \\
&= \frac{A}{(t+1)^2} + \frac{2B}{(t+1)^3} - \frac{2C_1}{(t+1)^{3/2}} - \frac{2C_2 \log n}{(t+1)^2} \\
&\le \frac{2}{(t+1)^{3/2}}\Big(\frac{A}{(t+1)^{1/2}} + \frac{B}{(t+1)^{3/2}} - C_1 - \frac{C_2 \log n}{(t+1)^{1/2}}\Big)
\end{aligned}$$

Define

$$t_0 := \inf\Big\{t \ge 1 : \frac{A}{(t+1)^{1/2}} + \frac{B}{(t+1)^{3/2}} - C_1 - \frac{C_2 \log n}{(t+1)^{1/2}} \le 0\Big\}.$$

According to the definition of A and B, t_0 exists. For those $t \ge t_0$, the RHS of Eq. (18) is negative, then the proof is done. For those $t \ge t_0$, we have

$$C_1 + \frac{C_2 \log n}{(t+1)^{1/2}} \le \frac{A}{(t+1)^{1/2}} + \frac{B}{(t+1)^{3/2}},$$

which is equivalent to

$$\frac{C_1}{(t+1)^{1/2}} + \frac{C_2 \log n}{t+1} \le \frac{A}{t+1} + \frac{B}{(t+1)^2}.$$

To finish the proof, it suffices to prove that

$$h_t \le \frac{C_1}{(t+1)^{1/2}} + \frac{C_2 \log n}{t+1},$$

which is demonstrated in Theorem A.3. Now we conclude the proof by setting $\alpha' = \alpha/n$. □

A.6. Proof of Theorem 3.11

Proof. Consider two adjacent datasets $\mathcal{D}$ and $\mathcal{D}'$, and their corresponding d_t and d_t'. We denote the sensitivity of $\langle d_t, v\rangle$ as s_t, namely $s_t := \max_{v\in\mathcal{C}} \max_{\mathcal{D}\simeq\mathcal{D}'} |\langle d_t - d_t', v\rangle|$. Then

$$s_t \le \max_{\mathcal{D}\simeq\mathcal{D}'} D\|d_t - d_t'\|_\infty.$$

Now we upper bound the sensitivity of $\|d_t - d_t'\|_\infty$. According to Eq. (7), we know that

$$d_t = \frac{1}{t+1}\sum_{i=1}^{t}\Big((i+1)\nabla f(\theta_i, x_i) - i\nabla f(\theta_{i-1}, x_i)\Big).$$

If adjacent datasets $\mathcal{D}$ and $\mathcal{D}'$ differ in data point x_i and x_i', then

$$\begin{aligned}
\|d_t - d_t'\|_\infty &= \frac{1}{t+1}\Big\|\Big((i+1)\nabla f(\theta_i, x_i) - i\nabla f(\theta_{i-1}, x_i)\Big) - \Big((i+1)\nabla f(\theta_i, x_i') - i\nabla f(\theta_{i-1}, x_i')\Big)\Big\|_\infty \\
&= \frac{1}{t+1}\Big\|i\Big(\nabla f(\theta_i, x_i) - \nabla f(\theta_{i-1}, x_i)\Big) - i\Big(\nabla f(\theta_i, x_i') - \nabla f(\theta_{i-1}, x_i')\Big) + \Big(\nabla f(\theta_i, x_i) - \nabla f(\theta_{i-1}, x_i')\Big)\Big\|_\infty \\
&\le \frac{2}{t+1}(i\beta\|\theta_i - \theta_{i-1}\|_1 + L) \le \frac{2}{t+1}(\beta\|v_t - \theta_{i-1}\|_1 + L) \\
&\le \frac{2}{t+1}(\beta D + L).
\end{aligned}$$

where the first inequality is due to β-smoothness and L-Lipschitz of F. Now we have

$$s_t \leq \frac{2D(\beta D + L)}{t+1}.$$

We denote the selected v_t in each iteration as random variable A_t. For any $v_1, v_2, ..., v_n \in \mathcal{C}$, we have

$$\log \frac{\mathbb{P}(A_1 = v_1, A_2 = v_2, ..., A_n = v_T | \mathcal{D})}{\mathbb{P}(A'_1 = v_1, A'_2 = v_2, ..., A'_n = v_n | \mathcal{D}')} = \sum_{t=1}^{n} \log \frac{\mathbb{P}(A_t = v_t | A_{t-1} = v_{t-1}, ..., A_1 = v_1, \mathcal{D})}{\mathbb{P}(A'_t = v_t | A'_{t-1} = v_{t-1}, ..., A'_1 = v_1, \mathcal{D}')} := \sum_{t=1}^{n} c_t(v_t, ..., v_1).$$

For each c_t, since we condition on $A_1 = v_1, A_2 = v_2, ..., A_{t-1} = v_{t-1}$, the randomness of A_t totally comes from the noise $\mathbf{n}_v^t \sim \text{Lap}\Big(\frac{2s_t\sqrt{t \cdot \log n \cdot \log(1/\delta)}}{\varepsilon}\Big)$. According to the Report Noisy Max Mechanism in Claim 3.9 in (Dwork et al., 2014), we have

$$|c_t| \leq \frac{\varepsilon}{2\sqrt{t \cdot \log n \cdot \log(1/\delta)}} := \varepsilon_t.$$

Then according to Lemma 3.18 in (Dwork et al., 2014), we have

$$\mathbb{E}[c_t | v_1, v_2, ..., v_{t-1}] \leq \varepsilon_t(e^{\varepsilon_t} - 1).$$

Now, according to Azuma-Hoeffding's inequality, we have

$$\mathbb{P}\Bigg(\sum_{t=1}^{n} c_t \geq \sum_{t=1}^{n} \varepsilon_t(e^{\varepsilon_t} - 1) + \sqrt{2\log(1/\delta)}\sqrt{\sum_{t=1}^{n} \varepsilon_t^2}\Bigg) \leq \delta.$$

So we can get (ε', δ)-DP, where

$$\varepsilon' = \sum_{t=1}^{n} \varepsilon_t^2 + \sqrt{2\log(1/\delta)}\sqrt{\sum_{t=1}^{n} \varepsilon_t^2} \leq \varepsilon,$$

which concludes the proof. □

A.7. Proof of Theorem 3.12

Firstly, we would like to introduce a proposition and a lemma.

Proposition A.6. *(Theorem 3.5 in (Pinelis, 1994)) Let $\zeta_1, \zeta_2, ..., \zeta_t \in \mathbb{R}^d$ be a vector-valued martingale difference sequence w.r.t. a filtration $\{\mathcal{F}_t\}$, i.e. for each $\tau \in 1, 2, ..., t$, we have $\mathbb{E}[\zeta_\tau | \mathcal{F}_{\tau-1}] = 0$. Suppose that $\|\zeta_\tau\|_2 \leq c_\tau$ almost surely. Then, $\forall t \geq 1$,*

$$\mathbb{P}\Bigg(\Bigg\|\sum_{\tau=1}^{t} \zeta_\tau\Bigg\|_2 \geq \lambda\Bigg) \leq 4\exp\Bigg(-\frac{\lambda^2}{4\sum_{\tau=1}^{t} c_\tau^2}\Bigg).$$

Lemma A.7. *Assume Assumption 2.2 and 2.3, for $t \in [n]$, we have that with probability at least $1 - \alpha_1$, Algorithm 2 will statisfies*

$$\|\Delta_t\|_\infty := \|d_t - \nabla F(\theta_t)\|_\infty \leq \frac{4(\beta D + G)\sqrt{\log(4d/\alpha_1)}}{\sqrt{t+1}}. \tag{19}$$

Proof of Lemma A.7. This proof is similar to the proof of Lemma 1 in (Xie et al., 2020), except that we consider the $\|\cdot\|_1$ norm and its dual norm $\|\cdot\|_\infty$, and apply the Proposition A.6 in a different way. Reformulating $\Delta_t = d_t - \nabla F(\theta_t)$ as the sum of a martingale difference sequence. For $t \geq 1$, we have

$$\begin{aligned}
\Delta_t &= \nabla f(\theta_t, x_t) + (1 - \rho_t)(d_{t-1} - \nabla f(\theta_{t-1}, x_t)) - \nabla F(\theta_t) \\
&= (1 - \rho_t)\epsilon_{t-1} + \rho_t(\nabla f(\theta_t, x_t) - \nabla F(\theta_t)) + \ldots \\
&\quad + (1 - \rho_t)\big(\nabla f(\theta_t, x_t) - \nabla f(\theta_{t-1}, x_t) - (\nabla F(\theta_t) - \nabla F(\theta_{t-1}))\big) \\
&= \prod_{k=2}^{t}(1 - \rho_k)\epsilon_1 + \sum_{\tau=2}^{t}\Bigg(\rho_\tau \prod_{k=\tau+1}^{t}(1 - \rho_k)\big(\nabla f(\theta_\tau, x_\tau) - \nabla F(\theta_\tau)\big) + \ldots \\
&\quad + \prod_{k=\tau}^{t}(1 - \rho_k)\big(\nabla f(\theta_\tau, x_\tau) - \nabla f(\theta_{\tau-1}, x_\tau) - (\nabla F(\theta_\tau) - \nabla F(\theta_{\tau-1}))\big)\Bigg) \triangleq \zeta_{t,1} + \sum_{\tau=2}^{t} \zeta_{t,\tau}
\end{aligned} \tag{20}$$

Recall that $\Delta_1 = \nabla f(\theta_1, x_1) - \nabla F(\theta_1)$. And we observe that $\mathbb{E}[\zeta_{t,\tau}|\mathcal{F}_{\tau-1}] = 0$ where $\mathcal{F}_\tau$ is the σ-field generated by $\{x_1, x_2, ..., x_{\tau-1}\}$. Therefore, $\{\zeta_{t,\tau}\}_{\tau=1}^t$ is a martingale difference sequence. In what follows, we derive upper bounds of $\|\zeta_{t,\tau}\|_\infty$. We start by observing that for any $\tau = 1, 2, ..., t$,

$$\prod_{k=\tau}^{t}(1-\rho_k) = \prod_{k=\tau}^{t}(1-\frac{1}{k+1}) = \prod_{k=\tau}^{t}\frac{k}{k+1} = \frac{\tau}{t+1} \tag{21}$$

We can bound $\|\zeta_{t,1}\|_\infty$ as follows:

$$\|\zeta_{t,1}\|_\infty \le \frac{1}{t+1}\|\nabla f(\theta_1, x_1) - \nabla F(\theta_1)\|_\infty \le \frac{G}{t+1} := c_{t,1},$$

where the first inequality is due to Assumption 2.3. For $\tau > 1$,

$$\begin{aligned}\|\zeta_{t,\tau}\|_\infty \le & \prod_{k=\tau}^{t}(1-\rho_k)\big(\|\nabla f(\theta_\tau, x_\tau) - \nabla f(\theta_{\tau-1}, x_\tau)\|_\infty + \|\nabla F(\theta_\tau) - \nabla F(\theta_{\tau-1})\|_\infty\big) + \dots \\ & + \rho_\tau \prod_{k=\tau+1}^{t}(1-\rho_k)\|\nabla f(\theta_\tau, x_\tau) - \nabla F(\theta_\tau)\|_\infty \\ \le & 2\beta\|\theta_\tau - \theta_{\tau-1}\|_1 \prod_{k=\tau}^{t}(1-\rho_k) + G\rho_\tau \prod_{k=\tau+1}^{t}(1-\rho_k) \\ = & 2\beta\eta_{\tau-1}\|v_{\tau-1} - \theta_{\tau-1}\|_1 \prod_{k=\tau}^{t}(1-\rho_k) + G\rho_\tau \prod_{k=\tau+1}^{t}(1-\rho_k) \\ \le & \frac{2\beta D + G}{t+1} := c_{t,\tau}.\end{aligned}$$

where the second inequality follows from Assumption 2.2 and 2.3, and the last inequality is due to $\eta_\tau = \rho_\tau = \frac{1}{\tau+1}$ and the definition of D. Now we denote the i-th element of Δ_t as $\Delta_{t,i}$ for $i \in 1, 2, ..., d$. According to Proposition A.6, we have

$$\mathbb{P}\big(|\Delta_{t,i}| \ge \lambda\big) \le 4\exp\left(-\frac{\lambda^2}{4\sum_{\tau=1}^t c_{t,\tau}^2}\right). \tag{22}$$

We can bound $\sum_{\tau=1}^t c_{t,\tau}^2$ as

$$\sum_{\tau=1}^t c_{t,\tau}^2 = c_{t,1}^2 + \sum_{\tau=2}^t c_{t,\tau}^2 = \left(\frac{G}{t+1}\right)^2 + \sum_{\tau=2}^t\left(\frac{2\beta D + G}{t+1}\right)^2 \le \sum_{\tau=1}^t\left(\frac{2\beta D + 2G}{t+1}\right)^2 \le \frac{4(\beta D + G)^2}{t+1}.$$

Plugging in the above bound and and setting $\lambda = \frac{4(\beta D+G)\sqrt{\log(4d/\alpha_1)}}{\sqrt{t+1}}$, for some $\alpha_1 \in (0, 1)$, we have with probability $1 - \alpha_1/d$,

$$|\Delta_{t,i}| \le \frac{4(\beta D + G)\sqrt{\log(4d/\alpha_1)}}{\sqrt{t+1}}$$

Then

$$\mathbb{P}(\|\Delta_t\|_\infty \le \lambda) = 1 - \mathbb{P}(\|\Delta_t\|_\infty > \lambda) \ge 1 - \sum_{i=1}^d \mathbb{P}(|\Delta_{t,i}| \ge \lambda) = 1 - \alpha_1,$$

where the first inequality comes from the union bound. In other word, with probability at least $1 - \alpha_1$, we have

$$\|\Delta_t\|_\infty \le \frac{4(\beta D + G)\sqrt{\log(4d/\alpha_1)}}{\sqrt{t+1}}.$$

□

Now we are ready to prove Theorem 3.12.

Proof of Theorem 3.12. We denote $h_t = F(\theta_t) - F(\theta^*)$, and $\tilde{v}_t \coloneqq \arg\min_{v\in\mathcal{C}}(d_t, v)$. We start from β-smoothness:

$$\begin{aligned}
h_{t+1} &\leq h_t + \eta_t\langle\nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\
&= h_t + \eta_t\langle d_t, v_t - \theta_t\rangle - \eta_t\langle d_t - \nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\
&= h_t + \eta_t\langle d_t, \tilde{v}_t - \theta_t\rangle - \eta_t\langle d_t - \nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle \\
&\leq h_t + \eta_t\langle d_t, \theta^* - \theta_t\rangle - \eta_t\langle d_t - \nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle \\
&= h_t + \eta_t\langle \nabla F(\theta_t), \theta^* - \theta_t\rangle + \eta_t\langle d_t - \nabla F(\theta_t), \theta^* - v_t\rangle + \frac{\eta_t^2\beta D^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle \\
&\leq (1-\eta_t)h_t + \eta_t D\|d_t - \nabla F(\theta_t)\|_\infty + \frac{\eta_t^2\beta D^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle.
\end{aligned} \tag{23}$$

To upper bound $\eta_t\langle d_t, v_t - \tilde{v}_t\rangle$, notice that

$$\langle d_t, v_t - \tilde{v}_t\rangle = \min_{v\in\mathcal{C}}\left(\langle v, d_t\rangle + \mathbf{n}_v^t\right) - \min_{v\in\mathcal{C}}\langle v, d_t\rangle \leq 2\max_{v=1,\ldots,2d}|\mathbf{n}_v^t| \tag{24}$$

with $\mathbf{n}_v^t \overset{i.i.d.}{\sim}$ Laplace$(0, \frac{4D(\beta D+L)\sqrt{\log n\cdot\log(1/\delta)}}{\sqrt{t}\varepsilon})$, we have by integrating the tail density

$$\mathbb{P}(\max_v|\mathbf{n}_v^t| > \lambda) \leq \sum_{v=1}^{2d} P(|\mathbf{n}_v^t| > \lambda) \leq 2d\exp\Big(-\frac{\sqrt{t}\varepsilon\lambda_t}{4D(\beta D+L)\sqrt{\log n\cdot\log(1/\delta)}}\Big).$$

selecting $\lambda_t = \frac{4D(\beta D+L)\sqrt{\log n\cdot\log(1/\delta)}}{\sqrt{t}\varepsilon}\cdot\log(2d/\alpha_2)$ we get then with probability at least $1-\alpha_2$,

$$\max_v|\mathbf{n}_v^t| \leq \frac{4D(\beta D+L)\sqrt{\log n\cdot\log(1/\delta)}}{\sqrt{t}\varepsilon}\cdot\log(2d/\alpha_2). \tag{25}$$

According to Eq. (23), (24), (25) and Lemma A.7, at iteration t, we have with probability at least $1-t(\alpha_1+\alpha_2)$,

$$\begin{aligned}
h_{t+1} &\leq (1-\eta_t)h_t + \frac{\eta_t^2\beta D^2}{2} + \ldots \\
&\quad + \frac{\eta_t}{\sqrt{t+1}}\underbrace{\left(8D(\beta D+G)\sqrt{\log(4d/\alpha_1)} + \frac{16D(\beta D+L)\log(2d/\alpha_2)\sqrt{\log n\cdot\log(1/\delta)}}{\varepsilon}\right)}_{A} \\
&= (1-\eta_t)h_t + \frac{\beta D^2}{2(t+1)^2} + \frac{A}{(t+1)^{3/2}}.
\end{aligned} \tag{26}$$

Now we prove $h_t \leq \frac{3}{\sqrt{t+1}}(\beta D^2 + A)$ by induction. For $t=1$, we have

$$h_2 \leq \frac{1}{2}\big(F(\theta_1) - F(\theta^*)\big) + \frac{\beta D^2}{8} + \frac{A}{3^{3/2}} \leq \frac{3}{\sqrt{2}}(\beta D^2 + A),$$

where the last inequality is due to $F(\theta_1) - F(\theta^*) \leq \frac{\beta D^2}{2}$ by the smoothness of F. Now we suppose $h_t \leq \frac{3}{\sqrt{t+1}}(\beta D^2 + A)$ for $t \geq 1$. For $t+1$, according to Eq. (26), we have

$$\begin{aligned}
h_{t+1} - \frac{3}{\sqrt{t+2}}(\beta D^2 + A) &\leq 3(\beta D^2 + A)\left(\frac{1}{\sqrt{t+1}} - \frac{1}{\sqrt{t+2}}\right) - \frac{2(\beta D^2 + A)}{(t+1)^{3/2}} \\
&\leq \frac{3(\beta D^2 + A)}{2(t+1)^{3/2}} - \frac{2(\beta D^2 + A)}{(t+1)^{3/2}} \leq 0,
\end{aligned}$$

where the second inequality is due to $\frac{1}{(t+1)^{1/2}} - \frac{1}{(t+2)^{1/2}} \leq \frac{1}{2(t+1)^{3/2}}$. And now we conclude the proof by setting $\alpha_1 = \alpha_2 = \frac{\alpha}{2n}$.

□

A.8. Proof of Theorem 3.13

Proof of Theorem 3.13. We define $\tilde{v}_t := \arg\min_{v\in\mathcal{C}}(d_t, v)$ and $\tilde{\theta}_t := \arg\max_{\theta\in\mathcal{C}}(\langle\nabla F(\theta_t), \theta_t - \theta\rangle)^2$ in Lemma A.4. And we denote that $h_t = F(\theta_t) - F(\theta^*)$. According to β-smoothness, we have

$$\begin{aligned} h_{t+1} &\leq F(\theta_t) + \eta_t\langle\nabla F(\theta_t), v_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} \\ &= h_t + \eta_t\langle\nabla F(\theta_t) - d_t, v_t - \theta_t\rangle + \eta_t\langle d_t, \tilde{v}_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle \\ &\leq h_t + \eta_t\|\nabla F(\theta_t) - d_t\|_\infty D + \eta_t\langle d_t, \tilde{\theta}_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle \\ &\leq h_t + 2\eta_t\|\nabla F(\theta_t) - d_t\|_\infty D + \eta_t\langle\nabla F(\theta_t), \tilde{\theta}_t - \theta_t\rangle + \frac{\eta_t^2\beta D^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle \\ &\leq h_t + 2\eta_t\|\nabla F(\theta_t) - d_t\|_\infty D - \eta_t\gamma\sqrt{2\mu h_t} + \frac{\eta_t^2\beta D^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle, \end{aligned} \tag{27}$$

where the last inequality follows from Lemma A.4. According to Eq. (24), (25) and (27), Lemma A.7, at iteration t, we have with probability at least $1 - t(\alpha_1 + \alpha_2)$,

$$\begin{aligned} h_{t+1} \leq &\sqrt{h_t}(\sqrt{h_t} - \eta_t\gamma\sqrt{2\mu}) + \frac{\eta_t^2\beta D^2}{2} + \dots \\ &+ \frac{\eta_t}{\sqrt{t+1}}\underbrace{\left(8D(\beta D + G)\sqrt{\log(4d/\alpha_1)} + \frac{16D(\beta D + L)\log(2d/\alpha_2)\sqrt{\log n \cdot \log(1/\delta)}}{\varepsilon}\right)}_{A} \end{aligned} \tag{28}$$

Now the claim holds by induction. For simplicity, we denote

$$B := \frac{9(\beta D^2 + A)^2}{\gamma^2\mu}.$$

Firstly, for $t = 1$, according to Eq. (28) we have

$$h_2 \leq F(\theta_1) - F(\theta^*) + \frac{\beta D^2}{8} + \frac{A}{3^{3/2}} \leq \frac{B}{2}.$$

where the last inequality is due to Lemma A.4 and the fact that $F(\theta_1) - F(\theta^*) \leq \frac{\beta D^2}{2}$. Suppose that $h_t \leq \frac{B}{t+1}$ for some $t \geq 1$. There are two cases.

Case 1. $\sqrt{h_t} - \eta_t\gamma\sqrt{2\mu} \leq 0$:

Then since $\eta_t = \frac{1}{t+1}$, Eq. (28) yields

$$h_{t+1} \leq \frac{\beta D^2}{2(t+1)^2} + \frac{A}{(t+1)^{3/2}} \leq \frac{\beta D^2 + A}{t+1} \leq \frac{2}{t+2}\frac{(\beta D^2 + A)^2}{\gamma^2\mu} \leq \frac{B}{t+2}.$$

where the third inequality is due to Lemma A.4 and the last inequality is from the definition of B.

Case 2. $\sqrt{h_t} - \eta_t\gamma\sqrt{2\mu} > 0$:

According to Eq. (28) and the assumption that $h_t \leq \frac{B}{t+1}$, we have

$$\begin{aligned} h_{t+1} - \frac{B}{t+2} &\leq B\left(\frac{1}{t+1} - \frac{1}{t+2}\right) + \frac{\beta D^2}{2(t+1)^2} + \frac{A}{(t+1)^{3/2}} - \frac{\gamma\sqrt{2\mu B}}{(t+1)^{3/2}} \\ &\leq \frac{1}{(t+1)^{3/2}}\left(\frac{B}{(t+1)^{1/2}} + \beta D^2 + A - \gamma\sqrt{2\mu B}\right) \\ &\leq \frac{1}{(t+1)^{3/2}}\left(\frac{B}{(t+1)^{1/2}} - 3(\beta D^2 + A)\right), \end{aligned} \tag{29}$$

where the last inequality comes from the definition of B. Define

$$t_0 \coloneqq \inf\{t \geq 1 : \frac{B}{(t+1)^{1/2}} - 3(\beta D^2 + A) \leq 0\}.$$

According to the definition of B, t_0 exists. For any $t \geq t_0$, the RHS of Eq. (29) is negative, and the proof is done. For those $t < t_0$, we have

$$3(\beta D^2 + A) \leq \frac{B}{(t+1)^{1/2}},$$

which is equivalent to

$$\frac{3(\beta D^2 + A)}{(t+1)^{1/2}} \leq \frac{B}{t+1}.$$

To conclude the proof, it suffices to show that the following inequality holds,

$$h_t \leq \frac{3(\beta D^2 + A)}{(t+1)^{1/2}}. \tag{30}$$

which is demonstrated in Theorem 3.12. Finally, we conclude the proof be setting $\alpha_1 = \alpha_2 = \frac{\alpha}{2n}$. □

B. Proof of Section 4

In this section we establish the privacy protection for our Algorithm 3 and the convergence result for the forced-sample estimators and full-sample estimators. We prove the convergence of estimators for any given arm i and use θ_t to represent $\theta_{t,i}$ and θ^* to represent θ_i^* for notation simplicity.

Proof of Theorem 4.6

Proof. By post-processing property, we only need to guarantee that the sequence $(\theta_1, \ldots, \theta_T)$ is (ε, δ) differentially private. In fact, we have for each sequence $\{\nu_{i+1}, \ldots, \nu_T\} \subset \mathcal{C}$. Suppose condition on $a_i(\mathcal{D}) = j_i$ and $a_i(\mathcal{D}') = j_i'$, we have then

$$\begin{aligned} &\frac{P(\theta_{i+1} = \nu_{i+1}, \ldots, \theta_T = \nu_T | \mathcal{D})}{P(\theta_{i+1} = \nu_{i+1}, \ldots, \theta_T = \nu_T | \mathcal{D}')} = \frac{P(\theta_{i+1} = \nu_{i+1} | \mathcal{D})}{P(\theta_{i+1} = \nu_{i+1} | \mathcal{D}')} \\ &= \frac{P(\theta_{i+1,1} = \nu_{i+1,1} \ldots, \theta_{i+1,K} = \nu_{i+1,K} | \mathcal{D})}{P(\theta_{i+1,1} = \nu_{i+1,1} \ldots, \theta_{i+1,K} = \nu_{i+1,K} | \mathcal{D}')} \\ &= \frac{P(\theta_{i+1,j_i} = \nu_{i+1,j_i}, \theta_{i+1,j_i'} = \nu_{i+1,j_i'} | \mathcal{D})}{P(\theta_{i+1,j_i'} = \nu_{i+1,j_i'}, \theta_{i+1,j_i} = \nu_{i+1,j_i} | \mathcal{D}')} \end{aligned}$$

Now by the synthetic update method, we have the above ratio is smaller or equal than ε with probability at least $1 - \delta$, which implies the (ε, δ)-differential privacy guarantee of $(\theta_1, \ldots, \theta_T)$. □

Lemma B.1. *Consider the arm i with $i \in K_{opt}$. Suppose the action a_τ (and $a_{\tau-1}$) depend only on θ_τ(and $\theta_{\tau-1}$), then we have for $P_i(\cdot|\theta)$ the distribution of $X_{t,i} := X_t \mathbf{1}\{a_t = i\}$ condition on θ (in particular, such distribution is independent of t),*

$$\|\mathbb{E}[\nabla F_{\tau-1}(\theta_{\tau-1}) - \nabla F_\tau(\theta_{\tau-1}) | \mathcal{F}_{\tau-1}]\|_\infty \leq 2\beta\|\theta_{\tau-1} - \theta^*\|_1 M \cdot \mathbb{E}_{\theta_\tau}\left[\|P_i(\cdot|\theta_{\tau-1}) - P_i(\cdot|\theta_\tau)\|_{TV} | \mathcal{F}_{\tau-1}\right]$$

Moreover, when both $a_{\tau-1}$ and a_τ are greedy actions, we have

$$\|P_i(\cdot|\theta_{\tau-1}) - P_i(\cdot|\theta_\tau)\|_{TV} \leq 2\beta\eta_{\tau-1}D.$$

Proof. Denote $P_i(\cdot|\theta)$ as the distribution of $X_{t,i}$ under greedy action condition on θ, and $\mathbb{E}^{\tau-1}[\cdot]$ the expectation condition on $\mathcal{F}_{\tau-1}$, then notice that $\nabla f(\theta; \mathbf{0}, r) = \mathbf{0}$ for every θ, r, we have

$$\begin{aligned}
&\|\mathbb{E}[D_\tau|\mathcal{F}_{\tau-1}]\|_\infty \\
=&\big\|\mathbb{E}^{\tau-1}_{X_{\tau-1,i}}[\mathbb{E}_r[\nabla f(\theta_{\tau-1,i}; X, r)|X]] - \mathbb{E}^{\tau-1}_{X_{\tau,i}}[\mathbb{E}_r[\nabla f(\theta_{\tau-1,i}; X, r)|X]]\big\|_\infty \\
=&\|\int_{\mathcal{X}} \big(\zeta(x^\top\theta_{\tau-1,i}) - \zeta(x^\top\theta_i^*)\big)xdP_i(x|\theta_{\tau-1}) - \int_\Theta\int_{\mathcal{X}} \big((\zeta(x^\top\theta_{\tau-1,i}) - \zeta(x^\top\theta_i^*)\big)xdP_i(x|\theta_\tau)dP(\theta_\tau)\|_\infty \\
\le& \int_\Theta\left[\int_{\mathcal{X}} |(\zeta(x^\top\theta_{\tau-1,i}) - \zeta(x^\top\theta_i^*)| \cdot \|x\|_\infty \cdot |p_i(x|\theta_{\tau,i}) - p_i(x|\theta_{\tau-1,,i})|d\nu\right]dP(\theta_\tau) \\
\le& 2\beta\|\theta_{\tau-1,i} - \theta_i^*\|M \cdot \mathbb{E}^{\tau-1}_{\theta_\tau}\big[\|P_i(\cdot|\theta_{\tau-1}) - P_i(\cdot|\theta_\tau)\|_{TV}\big]
\end{aligned}$$

Thus the first part is proved. On the other hand, notice that

$$\|\theta_\tau - \theta_{\tau-1}\|_1 = \eta_{\tau-1}\|v_{\tau-1} - \theta_{\tau-1}\|_1 \le \eta_{\tau-1}D$$

we get then

$$\begin{aligned}
\|P_i(\cdot|\theta_{\tau-1}) - P_i(\cdot|\theta_\tau)\|_{TV} &= \frac{1}{2}\int_{\mathcal{X}} |p_i(x|\theta_{\tau-1}) - p_i(x|\theta_\tau)|d\nu \\
&\le \frac{1}{2}(\int_S + \int_{S^c}) \cdot \big|p_i(x|\theta_{\tau-1}) - p_i(x|\theta_\tau)\big|d\nu
\end{aligned}$$

with $S := \big\{x \in \mathcal{X} : \mathbf{1}\{a(x|\theta_{\tau-1}) = i\} = \mathbf{1}\{a(x|\theta_\tau) = i\}\big\}$ and $a(x|\theta) \in [K]$ is the greedy action given context x and estimator θ, in particular we have $\mathbf{0} \in S$. Clearly we have the distribution of $X\mathbf{1}\{a(X|\theta_\tau) = i\}$ equals to the distribution of $X\mathbf{1}\{a(X|\theta_{\tau-1}) = i\}$ on S, thus

$$\frac{1}{2}(\int_S + \int_{S^c}) \cdot \big|p_i(x|\theta_{\tau-1}) - p_i(x|\theta_\tau)\big|d\nu = \frac{1}{2}\int_{S^c} \big|p_i(x|\theta_{\tau-1}) - p_i(x|\theta_\tau)\big|d\nu.$$

On the other hand, if we denote $p(z)$ the distribution of X, then by $\mathbf{0} \notin S^c$,

$$\begin{aligned}
\int_{S^c} \big|p_i(x|\theta_{\tau-1}) - p_i(x|\theta_\tau)\big|d\nu &\le 2\int_{\mathcal{X}}\int_{S^c} p_i(x|\theta_{\tau-1}, z)p(z)d\nu dz \\
&= 2\int_{S^c} p(z)\int_{S^c} p_i(x|\theta_{\tau-1}, z)d\nu dz \\
&\le 2\int_{S^c} p(z)dz \\
&= 2P(X \in S^c).
\end{aligned}$$

And by assumption 4.4

$$\begin{aligned}
P(X \in S^c) =& P\big(\mathbf{1}\{a(X|\theta_\tau) = i\} \neq_d \mathbf{1}\{a(X|\theta_{\tau-1}) = i\}\big) \\
=& P\big(a(X|\theta_\tau) = i, a(X|\theta_{\tau-1}) \neq i\big) + P\big(a(X|\theta_\tau) \neq i, a(X|\theta_{\tau-1}) = i\big) \\
\le& P(\max_{j\neq i} X^\top(\theta_{\tau,i} - \theta_{\tau,j}) > 0, \max_{j\neq i} X^\top(\theta_{\tau-1,i} - \theta_{\tau-1,j}) \le 0) \\
&+ P(\max_{j\neq i} X^\top(\theta_{\tau,i} - \theta_{\tau,j}) > 0, \max_{j\neq i} X^\top(\theta_{\tau-1,i} - \theta_{\tau-1,j}) \le 0) \\
\le& 2P(\max_{j\neq i} X^\top(\theta_{\tau-1,i} - \theta_{\tau-1,j}) < \eta_{\tau-1}D) \\
\le& 2\nu\eta_{\tau-1}D.
\end{aligned}$$

We get

$$\|P_i(\cdot|\theta_{\tau-1}) - P_i(\cdot|\theta_\tau)\|_{TV} \le 2\nu\eta_{\tau-1}D.$$

□

Lemma B.2. *For each arm $i \in K_{opt}$, suppose that the greedy action begins to be picked at time t_0, then for any $t > t_0$ we have with probability at least $1-\alpha$,*

$$\begin{aligned}\Delta_t \lesssim \frac{\beta}{t}\big(MD + C_{sc}(\alpha/2(d+t_0))(M + \sqrt{\log((d+T)/\alpha)})\sqrt{t_0} + DM\nu \sum_{\tau=t_0+1}^{t} \|\theta_{\tau-1,i} - \theta_i^*\|_1\big) \\ + \frac{(MD+\beta)\sqrt{\log((d+T)/\alpha)}}{\sqrt{t}},\end{aligned}$$

where

$$\Delta_t = \|d_t - \nabla F_t(\theta_t)\|_\infty,$$

$$A(\alpha) = 8D(\beta D + G)\sqrt{\log(8dT/\alpha)} + \frac{16D(\beta D + L)\log(4dT/\alpha)\sqrt{\log T \cdot \log(1/\delta)}}{\varepsilon},$$

and

$$C_{sc}(\alpha) = \sqrt{\frac{9(\beta D^2 + A(\alpha))^2}{u\mu\lambda}}.$$

Proof. For each arm i, denote $X_{t,i} := X_t \mathbf{1}\{a_t = i\}, r_{t,i} := r_t \mathbf{1}\{a_t = i\}$. Moreover we introduce $f_t(\theta) := f(\theta; X_{t,a_t}, y_t)$, $F_t(\theta) := \mathbb{E}[\nabla f_t(\theta)|\mathcal{F}_{t-1}]$ and $\Delta_t := d_t - \nabla F_t(\theta_t)$. Then

$$\begin{aligned}\Delta_t &= d_t - \nabla F_t(\theta_t) \\ &= \nabla f_t(\theta_t) + (1-\rho_t)(d_{t-1} - \nabla f_t(\theta_{t-1})) - \nabla F_t(\theta_t) \\ &= (1-\rho_t)\epsilon_{t-1} + \rho_t(\nabla f_t(\theta_t) - \nabla F_t(\theta_t)) \\ &\quad + (1-\rho_t)\big(\nabla f_t(\theta_t) - \nabla f_t(\theta_{t-1}) - (\nabla F_t(\theta_t) - \nabla F_{t-1}(\theta_{t-1}))\big) \\ &\le \prod_{k=2}^{t}(1-\rho_k)\epsilon_1 + \sum_{\tau=2}^{t}\prod_{k=\tau}^{t}(1-\rho_k)\underbrace{\big(\nabla f_\tau(\theta_\tau) - \nabla f_\tau(\theta_{\tau-1}) - (\nabla F_\tau(\theta_\tau) - \nabla F_{\tau-1}(\theta_{\tau-1}))\big)}_{D_\tau} \\ &\quad + \sum_{\tau=2}^{t}\rho_\tau \prod_{k=\tau+1}^{t}(1-\rho_k)\big(\nabla f_\tau(\theta_\tau) - \nabla F_\tau(\theta_\tau)\big) \qquad (31) \\ &= \prod_{k=2}^{t}(1-\rho_k)\epsilon_1 + \underbrace{\sum_{\tau=2}^{t}\prod_{k=\tau}^{t}(1-\rho_k)\big(D_\tau - \mathbb{E}[D_\tau|\mathcal{F}_{\tau-1}]\big)}_{\text{I}} \\ &\quad + \underbrace{\sum_{\tau=2}^{t}\rho_\tau \prod_{k=\tau+1}^{t}(1-\rho_k)\big(\nabla f_\tau(\theta_\tau) - \nabla F_\tau(\theta_\tau)\big)}_{\text{II}} + \underbrace{\sum_{\tau=2}^{t}\mathbb{E}[\prod_{k=\tau}^{t}(1-\rho_k)D_\tau|\mathcal{F}_{\tau-1}]}_{\text{III}}.\end{aligned}$$

First we bound III. Now by the theorem 3.13, we have with probability at least $1-(t_0-1)\alpha_1$,

$$\|\theta_{s,i} - \theta_i^*\|_1 \le \frac{C_{sc}((t_0-1)\alpha_1)}{\sqrt{s}}, \quad \forall s \le t_0 - 1.$$

Thus by the Lemma B.1 and the fact that

$$\mathbb{E}_{\theta_{t_0}}\big[\|P_i(\cdot|\theta_{t_0-1}) - P_i(\cdot|\theta_{t_0})\|_{TV}|\mathcal{F}_{t_0-1}\big] \le 1,$$

we have with probability at least $1-(t_0-1)\alpha_1$,

$$\begin{aligned}\sum_{\tau=2}^{t}\prod_{k=\tau}^{t}(1-\rho_k)\mathbb{E}[D_\tau|\mathcal{F}_{\tau-1}] &= \frac{t_0+1}{t+1}2\beta\|\theta_{t_0-1} - \theta^*\|_1 M + \sum_{\tau=t_0+1}^{t}\frac{\tau+1}{t+1}2\beta\|\theta_{\tau-1} - \theta^*\|_1 M \cdot 2\nu\eta_{\tau-1}D \\ &\le \frac{\beta M}{t+1}\Big(2C_{sc}((t_0-1)\alpha_1)\cdot\sqrt{t_0+1} + \sum_{\tau=t_0+1}^{t}(\tau+1)2\|\theta_{\tau-1} - \theta^*\|_1 \cdot 2\nu\eta_{\tau-1}D\Big)\end{aligned}$$

Now to bound Δ_t, it sufficient to bound I+II, i.e.,

$$\sum_{\tau=2}^{t}\prod_{k=\tau}^{t}(1-\rho_k)\big(D_\tau-\mathbb{E}[D_\tau|\mathcal{F}_{\tau-1}]\big)+\sum_{\tau=2}^{t}\rho_\tau\prod_{k=\tau+1}^{t}(1-\rho_k)\big(\nabla f_\tau(\theta_\tau)-\nabla F_\tau(\theta_\tau)\big).$$

For the second summation, we have by the same argument as in proof of Lemma A.7, with probability at least $1-\alpha_2$

$$\|\sum_{\tau=2}^{t}\rho_\tau\prod_{k=\tau+1}^{t}(1-\rho_k)\big(\nabla f_\tau(\theta_\tau)-\nabla F_\tau(\theta_\tau)\big)\|_\infty\le\frac{4MD\sqrt{\log(4d/\alpha_2)}}{\sqrt{t+1}}$$

For the first summation, notice that

$$\begin{aligned}\|D_\tau\|_\infty&\le 2MD\eta_\tau+\|\nabla F_\tau(\theta_{\tau-1})-\nabla F_{\tau-1}(\theta_{\tau-1})\|_\infty\\&=2MD\eta_\tau+\mathbb{E}\big[\|\mathbb{E}[\nabla F_\tau(\theta_{\tau-1})|\mathcal{F}_{\tau-1}]-\nabla F_{\tau-1}(\theta_{\tau-1})\|_\infty\big]\\&\le 2MD\eta_\tau+2\beta\|\theta_{\tau-1,i}-\theta_i^*\|_1 M\cdot\mathbb{E}[\|P_i(\cdot|\theta_{\tau-1})-P_i(\cdot|\theta_\tau)\|_{TV}]\end{aligned}$$

Now by

$$\mathbb{E}[\|P_i(\cdot|\theta_{\tau-1})-P_i(\cdot|\theta_\tau)\|_{TV}]\le\left\{\begin{array}{cc}0&\tau<t_0\\1&\tau=t_0,\\D\nu\eta_\tau&\tau>t_0.\end{array}\right.$$

And $\|\theta_{\tau-1,i}-\theta_i^*\|_1\le D$ we have by setting $M_\tau=\tau\big(D_\tau-\mathbb{E}[D_\tau|\mathcal{F}_{i-1}]\big)$, then

$$\|M_\tau\|_\infty\le\left\{\begin{array}{cc}2MD&\tau<t_0\\2M(D+\beta C_{sc}(\alpha_1)\sqrt{t_0})&\tau=t_0,\ \text{with probability at least }1-\alpha_1\\2MD(1+\beta)&\tau>t_0.\end{array}\right.$$

And thus apply Azuma-Hoeffding's inequality to each components $M_{\tau,i}$ with M_{t_0} replaced by $M'_{t_0}:=M_{t_0}\mathbf{1}\{M_{t_0}\le 2M(D+\beta C_{sc}(\alpha_1)\sqrt{t_0})\}$, we have with probability at least $1-d\alpha_1$,

$$\begin{aligned}|M'_{t_0,i}+\sum_{\tau\ne t_0}M_{\tau,i}|&\le\left(2M\sqrt{t_0D^2+(D+\beta C_{sc}(\alpha_1)\sqrt{t_0})^2+(t-t_0)D^2(1+\beta)^2}+\mathbb{E}[M'_{t_0,i}|\mathcal{F}_{t_0-1}]\right)\cdot\sqrt{\log(1/\alpha_1)}\\&\le 4M\big[D(1+\beta)\sqrt{t}+(D+\beta C_{sc}(\alpha_1)\sqrt{t_0})\big]\cdot\sqrt{\log(1/\alpha_1)}\end{aligned}$$

normalizing the summation by t and notice that $M'_{t_0}\ne M_{t_0}$ with probability at most α_1, we have then with probability at least $1-d\alpha_1$

$$\sum_{\tau=2}^{t}\prod_{k=\tau}^{t}(1-\rho_k)\big(D_\tau-\mathbb{E}[D_\tau|\mathcal{F}_{\tau-1}]\big)\le\big(\frac{2MD(1+\beta)}{\sqrt{t}}+\frac{D+\beta C_{sc}(\alpha_1)\sqrt{t_0}}{t}\big)\cdot\sqrt{\log(1/\alpha_1)},\forall i\in[d]$$

Now combining all bounds and set $\alpha_1=\alpha(d+t_0)/2,\alpha_2=\alpha/2$, we get with probability at least $1-\alpha$,

$$\begin{aligned}\|\Delta_t\|&\le\frac{M\beta D}{t+1}+\big(\frac{2MD(1+\beta)}{\sqrt{t}}+\frac{D+C_{sc}(\alpha(d+t_0)/2)\beta\sqrt{t_0}}{t}\big)\cdot\sqrt{\log(2(d+t_0)/\alpha)}+\frac{4MD\sqrt{\log(8d/\alpha)}}{\sqrt{t+1}}\\&\quad+\frac{\beta M}{t+1}\left(3C_{sc}((t_0-1)\alpha/(2(d+t_0)))\cdot\sqrt{t_0+1}+\sum_{\tau=t_0+1}^{t}(\tau+1)2\|\theta_{\tau-1,i}-\theta_i^*\|_1\cdot 2\nu\eta_{\tau-1}D\right)\\&\lesssim\frac{\beta}{t}\big(MD+C_{sc}((d+t_0)\alpha/2)(M+\sqrt{\log((d+T)/\alpha)})\sqrt{t_0}+DM\nu\sum_{\tau=t_0+1}^{t}\|\theta_{\tau-1}-\theta^*\|_1\big)\\&\quad+\frac{(MD+\beta)\sqrt{\log((d+T)/\alpha)}}{\sqrt{t}}\end{aligned}$$

as claimed. □

Lemma B.3. *As long as t_0 is selected so that with probability at least $1-\alpha$,*

$$\frac{C_{sc}(\alpha)}{\sqrt{t_0}} \leq \min\{\frac{h_{sub}}{4\beta M}, \ell\}.$$

Then we have with probability at least $1-\alpha$, the following claim holds for all $t_0 \leq t \leq T$ and $i \in K_{opt}$,

$$\begin{aligned}
&\hat{U}_t \cap U^c = \emptyset, \\
&u\lambda \leq \mathbb{E}[X_{t,i}X_{t,i}^\top] \\
&a_t^* \in \hat{K}_t.
\end{aligned}$$

In particular, that implies the GLM loss

$$F_t(\theta) := \mathbb{E}[f(\theta; X_{t,i}, r_{t,i})|\mathcal{F}_{t-1}]$$

is $\mu\lambda u$-strongly convex in ℓ_1-geometry.

Proof. Firstly, notice that as long as $\sup_{i\in[K]}\|\theta_{t_0,i} - \theta_i^*\|_1 < \dfrac{h_{sub}}{4\beta M}$, we have for each t, denote

$$i_t := \text{argmax}_{i\in[K]}\zeta(X_t^\top \theta_{t_0,i}), i_t^* := \text{argmax}_{i\in[K]}\zeta(X_t^\top \theta_i^*),$$

then

$$\begin{aligned}
P(\hat{K}_t \cap K_{sub} \neq \emptyset) &= P(\exists j \in K_{sub} \text{ s.t. } \zeta(X_t^\top \theta_{t_0,j}) > \zeta(X_t^\top \theta_{t_0,i_t}) - h_{sub}/2) \\
&\leq P(\exists j \in K_{sub}c \text{ s.t. } \zeta(X_t^\top \theta_{t_0,j}) > \zeta(X_t^\top \theta_{t_0,i_t^*}) - h_{sub}/2) \\
&\leq P(\exists j \in K_{sub} \text{ s.t. } \zeta(X_t^\top \theta_j^*) + \frac{h_{sub}}{4} > \zeta(X_t^\top \theta_{i_t^*}^*) - \frac{3h_{sub}}{4}) \\
&= 0.
\end{aligned}$$

Thus the first claim holds.
To prove the second claim, notice that for every $t \geq K * t_0$, we have condition on the $\sup_{i\in[K]}\|\theta_{t_0,i} - \theta_i^*\|_1 < \dfrac{h_{sub}}{4\beta M}$, for every $i \in K_{opt}$,

$$\begin{aligned}
P(a_t = i) &\geq P(\hat{K}_t = \{i\}) \\
&\geq P(\hat{K}_t = \{i\}, X_t \in U_i) \\
&= P(\zeta(X_t^\top \theta_{Kt_0,i}) > \max_{j\neq i}\zeta(X_t^\top \theta_{Kt_0,j}) + h_{sub}/2, \zeta(X_t^\top \theta_i^*) > \max_{j\neq i}\zeta(X_t^\top \theta_j^*) + h_{sub}) \\
&\geq P(\sup_{i\in[K]}\|\theta_{s_0,i} - \theta_i^*\|_1 < \frac{h_{sub}}{4\beta M}, \zeta(X_t^\top \theta_i^*) > \max_{j\neq i}\zeta(X_t^\top \theta_j^*) + h_{sub}) \\
&= P(X_t \in U_j) \geq u.
\end{aligned}$$

Thus we have

$$\lambda u \leq P(X_t \in U_i) \cdot \mathbb{E}[X_t X_t^\top | X_t \in U_i] \leq \mathbb{E}[X_{t,i}X_{t,i}^\top].$$

To prove the third claim, note that for any $i_t = \text{argmax}_{i\in[K]} X_t^\top \theta_{t_0,i}$

$$\begin{aligned}
\zeta(X_t^\top \theta_{t_0,i_t}) - \zeta(X_t^\top \theta_{t_0,i_t^*}) &= \zeta(X_t^\top \theta_{t_0,i_t}) - \zeta(X_t^\top \theta_{t_0,i_t}^*) \\
&\quad + \zeta(X_t^\top \theta_{t_0,i_t}^*) - \zeta(X_t^\top \theta_{t_0,i_t^*}^*) \\
&\quad + \zeta(X_t^\top \theta_{t_0,i_t^*}^*) - \zeta(X_t^\top \theta_{t_0,i_t^*})) \\
&\leq \frac{h_{sub}}{2} + \zeta(X_t^\top \theta_{t_0,i_t}^*) - \zeta(X_t^\top \theta_{t_0,i_t^*}^*) \\
&\leq \frac{h_{sub}}{2}.
\end{aligned}$$

Thus $i_t^* \in \hat{K}_i$. □

B.1. Proof of Theorem 4.8

Proof. Suppose at time t, we have with probability at least $1-\alpha$ that

$$h_\tau \leq \frac{C_{in}(\alpha)}{\tau}, (\text{thus } \|\theta_\tau - \theta^*\|_1 \leq \sqrt{\frac{C_{in}(\alpha)}{u\lambda\mu\tau}}), \quad \forall \tau \leq t-1,$$

then condition on such event, from Lemma B.2 and the same argument in (25), we have for $h_t := F_t(\theta_t) - F_t(\theta^*)$, with probability at least $1-\alpha_1-\alpha_2$,

$$\begin{aligned}
h_t &\leq h_{t-1} + 2\eta_t\|\nabla F_t(\theta_t) - d_t\|_\infty D - \eta_t\gamma\sqrt{2\mu h_{t-1}} + \frac{\eta_t^2 LD^2}{2} + \eta_t\langle d_t, v_t - \tilde{v}_t\rangle \\
&\leq h_{t-1} - \eta_t\gamma\sqrt{2\mu h_{t-1}} + 2\eta_t\frac{D\beta}{t}\big(MD + C_{sc}(\alpha_1/2(d+t_0))(M + \sqrt{\log((d+T)/\alpha_1)})\sqrt{t_0}\big) \\
&\quad + \frac{2\eta_t DM\nu\sqrt{t}\beta C_{in}^{1/2}}{\sqrt{\mu}t} + \frac{2\eta_t D(MD+\beta)\sqrt{\log((d+T)/\alpha_1)}}{\sqrt{t}} + \frac{\eta_t^2 LD^2}{2} \\
&\quad + \eta_t\frac{4D(\beta D + L)\sqrt{\log T \cdot \log(1/\delta)}}{\sqrt{t}\varepsilon} \cdot 4\log(2d/\alpha_2) \\
&:= h_{t-1} - \eta_t\gamma\sqrt{2\mu h_{t-1}} + \frac{G_1 C_{in}^{1/2}}{t^{3/2}} + \frac{G_2}{t^{3/2}\varepsilon} + \frac{G_3}{t^{3/2}} + \frac{G_4}{t^2},
\end{aligned}$$

where

$$\begin{aligned}
G_1 &:= \frac{2\beta D^2 M\nu}{\sqrt{\mu}}, \\
G_2 &:= \frac{4D(\beta D + L)\sqrt{\log T \cdot \log(1/\delta)}}{\varepsilon} \cdot 4\log(2d/\alpha_2), \\
G_3 &:= D(MD+\beta)\sqrt{\log((d+T)/\alpha_1)}, \\
G_4 &:= D\beta\big(MD + C_{sc}(\alpha_1/2(d+t_0))(M + \sqrt{\log((d+T)/\alpha_1)})\sqrt{t_0}\big) + \frac{LD^2}{2}.
\end{aligned}$$

For notation simplicity we abbreviate C_{in} for $C_{in}(\alpha)$ below.

Case1: $h_{t-1} - \eta_t\gamma\sqrt{2\mu h_{t-1}} \leq 0$: i.e.

$$h_t \leq \frac{G_1 C_{in}^{1/2}}{t^{3/2}} + \frac{G_2}{t^{3/2}\varepsilon} + \frac{G_3}{t^{3/2}} + \frac{G_4}{t^2}.$$

To ensure the induction, we need

$$\begin{aligned}
&\frac{G_1 C_{in}^{1/2}}{t^{3/2}} + \frac{G_2}{t^{3/2}\varepsilon} + \frac{G_3}{t^{3/2}} + \frac{G_4}{t^2} \leq \frac{C_{in}}{t} \\
\iff &\frac{G_1 C_{in}^{1/2}}{t^{1/2}} + \frac{G_2}{t^{1/2}\varepsilon} + \frac{G_3}{t^{1/2}} + \frac{G_4}{t} \leq C_{in}
\end{aligned}$$

As long as $t \geq \max\{G_1^2, \frac{G_2^2}{\log(dT/\alpha)}, G_3^2, \frac{G_4}{\log(dT/\alpha)}\} \Rightarrow t \geq \dfrac{\log(dT/\alpha)\log(T)}{\varepsilon^2}$, we can choose $C_{in} = \max\{4, \frac{9\log^2(dT/\alpha)}{\varepsilon^2}\}$ to satisfy the above inequality.

Case2: $h_{t-1} - \eta_t\gamma\sqrt{2\mu h_{t-1}} > 0$ **:**

$$\begin{aligned}
h_t &= h_{t-1} - \eta_t\gamma\sqrt{2\mu h_{t-1}} + \frac{G_1 C_{in}^{1/2}}{t^{3/2}} + \frac{G_2}{t^{3/2}\varepsilon} + \frac{G_3}{t^{3/2}} + \frac{G_4}{t^2} \\
&\leq \frac{C_{in}}{t} - \eta_t\gamma\frac{\sqrt{2\mu}C_{in}^{1/2}}{\sqrt{t}} + \frac{G_1 C_{in}^{1/2}}{t^{3/2}} + \frac{G_2}{t^{3/2}\varepsilon} + \frac{G_3}{t^{3/2}} + \frac{G_4}{t^2},
\end{aligned}$$

i.e.

$$h_{t+1} - \frac{C_{in}}{t+1} \le \Big(\frac{C_{in}}{t} - \frac{C_{in}}{t+1}\Big) - \eta_t \gamma \frac{\sqrt{2\mu} C_{in}^{1/2}}{\sqrt{t}} + \frac{G_1 C_{in}^{1/2}}{t^{3/2}} + \frac{G_2}{t^{3/2}\varepsilon} + \frac{G_3}{t^{3/2}} + \frac{G_4}{t^2},$$

i.e. we need to choose C_{in} so that RHS is not greater than zero,

$$\text{i.e. } (\gamma\sqrt{2\mu} - G_1) C_{in}^{1/2} - \frac{C_{in}}{t^{1/2}} \ge \frac{G_2}{t^{1/2}\varepsilon} + \frac{G_3}{t^{1/2}} + \frac{G_4}{t}.$$

Choosing $\gamma \ge \frac{3G_1}{2\sqrt{2\mu}}$, and as long as

$$C_{in} = \tilde{C} \frac{\log^2(dT/\alpha)}{\varepsilon^2} \text{ for some } \tilde{C} \text{ independent of } d, T \text{ and } \alpha,$$

$$t \ge \max\{\frac{G_2^2}{\log(dT/\alpha)}, G_3^2, \frac{G_4}{\log(dT/\alpha)}, \frac{C_{in}}{9\log(dT/\alpha)}\} \Rightarrow t \gtrsim \frac{\log(dT/\alpha)\log(T)}{\varepsilon^2},$$

the claim holds.

Finally we need to ensure the induction holds when $t = t_0$. Note that when $t \le t_0$, the convergence result is given by Theorem 3.13. Thus to ensure the induction holds we also need $C_{in} \ge \lambda\beta\mu(C_{sc}(\alpha))^2$. In conclusion, we choose $\alpha_1 = \alpha_2 = \frac{\alpha}{2T}$, $C_{in} = \max\{4, \frac{9\log(dT)}{\varepsilon^2}, \frac{36\log^2(dT)}{(\varepsilon G_1)^2}, \lambda\beta\mu(C_{sc}(\alpha))^2\} = O(\frac{\log^2(dT/\alpha)\log(T)}{\varepsilon^2})$, and $t_0 = \max\{G_1^2, \frac{G_2^2}{\log(dT/\alpha)}, G_3^2, \frac{G_4}{\log(dT/\alpha)}, \frac{C_{in}}{9\log(dT/\alpha)}\} = O(\frac{\log(dT/\alpha)\log(T)}{\varepsilon^2})$ to ensure the induction holds.

□

B.2. Proof of Theorem 4.9

Proof. We define event

$$E_{t,1} \coloneqq \{\hat{U}_t \cap U^c = \emptyset, a_t^* \in \hat{U}_t\},$$

$$E_{t,2} \coloneqq \{\sup_{i \in K_{opt}} \|\theta_{t,i} - \theta_i^*\|_1 \le \frac{\sqrt{C_{in}(\alpha/(4|K_{opt}|)}}{2\sqrt{t}}\},$$

$$E_t \coloneqq \{E_{t,1}, E_{t,2}, \Delta_t \le \frac{M\beta\sqrt{C_{in}(\alpha/(4|K_{opt}|))}}{\sqrt{t}}\}.$$

and we use C_{in} for $C_{in}(\alpha/(4|K_{opt}|))$ for notation simplicity in the following. Using the similar argument as in Theorem 4.8, we can verify that condition on the event $\sup_{i \in K_{opt}} \|\theta_{t,i} - \theta_i^*\|_1 \le \frac{\sqrt{C_{in}}}{2\sqrt{t}}$ and $\Delta_t \le \frac{M\beta\sqrt{C_{in}}}{\sqrt{t}}$, we must have $a_t = a_t^*$. Thus with probability at least $1 - \frac{3\alpha}{4}$, the event $E_t \cup E_{t,1}^c \cup E_{t,2}^c$ holds, thus,

$$\begin{aligned}\text{Regret}(T) &= \Big(\sum_{t \le t_0} + \sum_{t > t_0}\Big)\big(\zeta(X_t^\top \theta_t^*) - \zeta(X_t^\top \theta_{a_t}^*)\big) \\ &\le \sum_{t \le t_0} (\zeta(X_t^\top \theta_t^*) - \zeta(X_t^\top \theta_{a_t}^*)) + \sum_{t > t_0} (\zeta(X_t^\top \theta_t^*) - \zeta(X_t^\top \theta_{a_t}^*))\mathbf{1}\{E_t \cup E_{t,1}^c \cup E_{t,2}^c\}.\end{aligned}$$

Note that the choice of $t_0 = O(\frac{\log(dT/\alpha)\log(T)}{\varepsilon^2})$ as stated in Theorem 4.8 and the range of the reward can be bounded. Now it remains to bound the second term. Let $A_t = \frac{M\beta\sqrt{C_{in}}}{\sqrt{t}}\mathbf{1}\{\Delta_t \le \frac{M\beta\sqrt{C_{in}}}{\sqrt{t}}\}$ and the second term is upper bounded by $\sum_{t_0 < t \le T} A_t$. We have $\sum_{t_0 < t \le T} (\frac{M\beta\sqrt{C_{in}}}{\sqrt{t}})^2 \le M^2\beta^2 C_{in}\log(T)$. Note that $P(A_t = \frac{M\beta\sqrt{C_{in}}}{\sqrt{t}}) \le \nu\frac{M\beta\sqrt{C_{in}}}{\sqrt{t}}$ by Assumption 4.4 and thus $\mathbb{E}[\sum_{t_0 < t \le T} A_t] \le \nu M^2\beta^2 C_{in}\log(T)$. We apply Hoeffding's inequality and can conclude that with probability at least $1 - \alpha/4$

$$\sum_{t \ge t_0} A_t \le \nu M^2\beta^2 C_{in}\log(T) + 2M\beta\sqrt{C_{in}\log(T)\log(4/\alpha)}.$$

Putting all the terms together, and we arrive the desired conclusion. □

Table 3. SubOpt for NoisySFW and Algorithm 1 with $p = 1.5$.

	$T = 1000, d = 5$	$T = 1000, d = 10$
NoisySFW	0.52± 0.057	0.95± 0.095
Algo. 1	0.017± 0.010	0.20± 0.048
	$T = 2000, d = 5$	$T = 2000, d = 10$
NoisySFW	0.44± 0.029	0.89± 0.069
Algo. 1	0.0024± 0.001	0.060± 0.032

Table 4. SubOpt for NoisySGD and Algorithm 1 with $p = \infty$.

	$T = 1000, d = 5$	$T = 1000, d = 10$
NoisySGD	0.026 ± 0.018	0.053 ± 0.015
Algo. 1	0.036 ± 0.022	0.092 ± 0.021
	$T = 2000, d = 5$	$T = 2000, d = 10$
NoisySGD	0.013 ± 0.0037	0.038 ± 0.013
Algo. 1	0.015 ± 0.0059	0.058 ± 0.013

C. Experiments

In this section, we present experiment results to demonstrate the efficacy and efficiency of our algorithm. We consider the linear regression setting $y = X^\top \theta + \epsilon$, where the design matrix $X \in \mathbb{R}^{d\times n}$, true parameter $\theta \in \mathbb{R}^d$, output $y \in \mathbb{R}^n$, and $\epsilon \sim N(0, \nu^2)$ is a noise vector. We define the loss function as $\mathcal{L}(\hat{\theta}, X) = \frac{1}{n}\sum_{i=1}^n (y_i - \langle x_i, \hat{\theta}\rangle)^2$ for any given estimation $\hat{\theta}$, where y_i is the i-th entry of y and x_i is the i-th column of X. Therefore, the excess risk will be $F(\hat{\theta}) = \mathbb{E}[\mathcal{L}(\hat{\theta})]$. Here we will use the loss function over a separate testing set as an empirical estimation of the excess risk, which we denote as $\mathcal{L}(\hat{\theta}, X_{\text{test}})$. And we further introduce suboptimality as SubOpt $= \frac{\mathcal{L}(\hat{\theta}, X_{\text{test}}) - \mathcal{L}(\theta, X_{\text{test}})}{\mathcal{L}(\theta_0, X_{\text{test}}) - \mathcal{L}(\theta, X_{\text{test}})}$, to demonstrate utility intuitively. Here θ_0 is zero vector, serving as the initialization of all algorithms.

We choose $p = 1.5$ and $p = \infty$ as our geometries. And $(1, 1/T)$-DP is guaranteed. We compare our Algorithm 1 with NoisySFW (Algorithm 3 in Bassily et al. (2021b)), LocalMD (Algorithm 6 in Asi et al. (2021)) when $p = 1.5$, and with NoisySGD (Algorithm 2 in Bassily et al. (2020)) when $p = \infty$. We generate T samples i.i.d. from a normal distribution with mean zero and standard deviation 0.05, and then normalize them by their q-norm to ensure each sample maintain unit q-norm. We also generate the true underlying parameter θ by setting all its entries to 1 and then normalized it by its p-norm. The size of the testing set is 10000.

We show the SubOpt in Table 3 and Table 4. All results are based on 10 independent runs. To achieve the best performance for each algorithm, we will scale their default learning rate by a grid of scaling factors, and report the best SubOpt. One thing we need to mention is that LocalMD does not converge regardless of the learning rate scaling. We suspect that this is due to the large constants before their Bregman divergence, and the standard deviation of their Gaussian noise. As we can see, Algorithm 1 outperforms NoisySFW in $p = 1.5$. What's more, it achieves comparable result to NoisySGD in $p = \infty$. And we recall that the number of gradient query for Algorithm 1 is linear, while that of NoisySGD is superlinear.