Local Differential Privacy for Bayesian Optimization

Xingyu Zhou,[1*] **Jian Tan** [2]
[1] ECE, The Ohio State University
[2] Alibaba Group, Sunnyvale
zhou.2055@osu.edu, j.tan@alibaba-inc.com

Abstract

Motivated by the increasing concern about privacy in nowadays data-intensive online learning systems, we consider a black-box optimization in the nonparametric Gaussian process setting with local differential privacy (LDP) guarantee. Specifically, the rewards from each user are further corrupted to protect privacy and the learner only has access to the corrupted rewards to minimize the regret. We first derive the regret lower bounds for any LDP mechanism and any learning algorithm. Then, we present three almost optimal algorithms based on the GP-UCB framework and Laplace DP mechanism. In this process, we also propose a new Bayesian optimization (BO) method (called MoMA-GP-UCB) based on median-of-means techniques and kernel approximations, which complements previous BO algorithms for heavy-tailed payoffs with a reduced complexity. Further, empirical comparisons of different algorithms on both synthetic and real-world datasets highlight the superior performance of MoMA-GP-UCB in both private and non-private scenarios.

Introduction

We consider the problem of maximizing an unknown function f over a set $\mathcal{D}$ via sequentially querying it and received only bandit feedback, i.e., when we query at x, we observe a possibly noisy evaluation of $f(x)$. This model has been a main focus in machine learning research, e.g., the classic multi-armed bandit (MAB) setting (Lai and Robbins 1985), linear bandit setting (Abbasi-Yadkori, Pál, and Szepesvári 2011) and the general Bayesian optimization (Shahriari et al. 2015), with each one generalizing the previous one. It also finds broad applications in many real-world systems, including medical experiments, online shopping websites and recommender systems (Li et al. 2010). These systems adaptively make a decision and receive rewards (feedback) from the user to simultaneously learn insightful facts and maximize the profit.

Recently, privacy has become a key issue in the above mentioned online learning systems. Users have become increasingly concerned about directly sharing their online information or activities to these systems, since these activities may reveal their private information. For example, a customer of an online shopping website is not willing to tell the website that he or she has purchased medicines for mental issues. Another example is the medical experiments in which the patient could reject to share the actual effects of the treatment due to privacy concerns. This stimulates the need to have a mechanism that further corrupts the feedbacks from each user to protect privacy, which exactly fits the *locally differential private* (LDP) model (Kasiviswanathan et al. 2011; Duchi, Jordan, and Wainwright 2013).

In contrast to the standard differential privacy model (Dwork, Roth et al. 2014), in which the learner collects the true data while releasing a private output to protect privacy, in the LDP model, the learner only has access to corrupted input data from the users. Hence, LDP often provides a much stronger privacy protection for the user and is more appealing in real applications, especially for the systems mentioned above (Cormode et al. 2018). To the best of our knowledge, in the setting of online learning with bandit feedback, LDP model has only been studied theoretically very recently. For example, in (Gajane, Urvoy, and Kaufmann 2018; Ren et al. 2020), the authors investigate MAB with LDP guarantee. (Zheng et al. 2020) studies LDP in the linear (contextual) bandit setting. However, LDP in the most general scenario, i.e., Bayesian optimization (BO), remains an important open problem.

Motivated by this, in this paper, we investigate the locally differentially private BO, in which the rewards are further corrupted to protect privacy. Specifically, we consider a Gaussian process (GP) model for BO (also called Gaussian process bandit setting), which directly generalizes both MAB and linear bandit setting. The main contributions of this paper can be summarized as follows.

Contributions. (i) We first derive the regret lower bounds for any LDP mechanism and any learning algorithm. (ii) Then, we present three almost optimal algorithms based on the GP-UCB framework and Laplace DP mechanism. (iii) Our two new methods developed for handling LDP also contribute to BO under heavy-tailed payoffs in general. In particular, one is a new truncation method that can be applied to any sub-Weibull rewards. The other one, called MoMA-GP-UCB, is based on median-of-means techniques and kernel approximations, which complements previous BO algorithms for general heavy-tailed payoffs (Chowdhury and

[*]This work was done during the internship of the first author at Alibaba Group, Sunnyvale, USA.

Gopalan 2019) with a reduced complexity. (iv) We further conduct empirical comparisons of different algorithms over both synthetic and real-world datasets, which demonstrate the superior performance of our new MoMA-GP-UCB algorithm in both private and non-private settings.

Related Work

In the traditional non-private case, a line of BO methods based on Gaussian process (GP) and upper confidence bound (UCB) have been analyzed in both sub-Gaussian (Srinivas et al. 2009; Chowdhury and Gopalan 2017) and heavy-tailed scenarios (Chowdhury and Gopalan 2019). Kernel approximation is recently proposed to reduce complexity of GP-based BO algorithms (Mutny and Krause 2018; Calandriello et al. 2019). In the private BO case, (Kusner et al. 2015) studies how to privately release the BO outputs to protect privacy (e.g., hyper-parameters of machine learning model), and hence it belongs to the traditional DP perspective rather than LDP.

LDP model has been previously considered in MAB setting (Gajane, Urvoy, and Kaufmann 2018; Basu, Dimitrakakis, and Tossou 2019; Ren et al. 2020). Recently, it is generalized to linear contextual bandits in which both the rewards and the contexts are corrupted for privacy (Zheng et al. 2020). There are also some other types of DP considered in MAB and linear bandit setting (not comparable to our work). Due to space limitations, we refer readers to (Ren et al. 2020; Zheng et al. 2020) and the references therein.

Problem Statement and Preliminaries

We consider a sequential decision-making problem over a set $\mathcal{D}$. A learning policy is adopted to select an action $x_t \in \mathcal{D}$ at each discrete time slot $t = 1, 2, \ldots$ with the corresponding reward observation $y_t = f(x_t) + \eta_t$, i.e., y_t could be a noisy version of $f(x_t)$. Then, the reward y_t will be further corrupted to protect privacy, and only the private response $\bar{y}_t$ is revealed to the learning agent. The action x_t is chosen based on the arms played and the private rewards obtained until $t-1$, denoted by the history $\mathcal{H}_{t-1} = \{(x_s, \bar{y}_s) : s \in [t-1]^1\}$. The objective is to simultaneously preserve LDP and minimize the cumulative regret defined as

$$R_T = \sum_{t=1}^{T} f(x^*) - f(x_t), \tag{1}$$

where $x^* = \arg\max_{x \in \mathcal{D}} f(x)$ (assuming the maximum is attained).

Definition 1 ((ϵ, δ)-LDP)**.** *A randomized mechanism $M : \mathcal{D} \to \mathcal{Z}$ is said to protect (ϵ, δ)-LDP if for any $x, x' \in \mathcal{D}$, and any measurable subset $E \in \mathcal{Z}$, there is*

$$\mathbb{P}\{M(x) \in E\} \leq e^{\epsilon}\mathbb{P}\{M(x') \in E\} + \delta,$$

for $\epsilon \geq 0$ and $\delta \geq 0$. Moreover, if $\delta = 0$, we say it protects ϵ-LDP.

[1]For any positive integer, we define [m]:={1,2, …, m}

Note that, if not explicitly stated, LDP in this paper means ϵ-LDP (stronger than (ϵ, δ)-LDP).

Noise Assumptions. We assume that the noise η_t has zero mean conditioned on the history and is bounded by R almost surely. We also address the case of unbounded noise at the end of the paper.

Regularity Assumptions. Attaining a sub-linear regret is in general infeasible for an arbitrary reward function f over a very large space without any assumptions on the structure of f. In this paper, we assume that $\mathcal{D}$ is compact and f has a bounded norm in the RKHS of functions $\mathcal{D} \to \mathbb{R}$, corresponding a kernel function $k : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$. This RKHS denoted by $\mathcal{H}_k(\mathcal{D})$ is completely determined by its kernel function with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ that satisfies the reproducing property: $f(x) = \langle f, k(x, \cdot) \rangle_{\mathcal{H}}$ for all $f \in H_k(\mathcal{D})$. The norm for the RKHS is given by $\|f\|_{\mathcal{H}} = \sqrt{\langle f, f \rangle_{\mathcal{H}}}$, which measures the smoothness of f. We assume $\|f\|_{\mathcal{H}} \leq B$ and $B < \infty$ is a known constant. Moreover, we assume a bounded variance by restricting $k(x, x) \leq 1$. Note that two commonly used kernels *Squared Exponential* and *Matérn* satisfy the bounded variance assumption, defined as:

$$k_{\text{SE}}(x, x') = \exp\left(-s^2/2l^2\right)$$
$$k_{\text{Matérn}}(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{s\sqrt{2\nu}}{l}\right)^{\nu} B_\nu \left(\frac{s\sqrt{2\nu}}{l}\right),$$

where $l > 0$ and $\nu > 0$ are hyper-parameters, $s = \|x - x'\|_2$ specifies the similarity between two points, and $B_\nu(\cdot)$ is the modified Bessel function.

Surrogate GP Model[2]. A Gaussian process, denoted by $\mathcal{GP}(\mu(\cdot), k(\cdot, \cdot))$, is a collection of (possibly infinitely many) random variables $f(x), x \in \mathcal{D}$, such that every finite subset of random variables $\{f(x_i), i \in [m]\}$ is jointly Gaussian with mean $\mathbb{E}[f(x_i)] = \mu(x_i)$ and covariance $\mathbb{E}[(f(x_i) - \mu(x_i))(f(x_j) - \mu(x_j))] = k(x_i, x_j)$, where $i, j \in [m]$ and $m \in \mathbb{N}$. By conditioning GPs on available observations, one can obtain a non-parametric surrogate Bayesian model over the space of functions. In particular, we use $\mathcal{GP}(0, k(\cdot, \cdot))$ as an initial prior on the unknown black-box function f, and a Gaussian likelihood with the noise variables η_t drawn independently across t from $\mathcal{N}(0, \lambda)$. Conditioned on a set of past observations $\mathcal{H}_t = \{(x_s, y_s), s \in [t]\}$, by the properties of GPs (Rasmussen 2003), the posterior distribution over f is $\mathcal{GP}(\mu_t(\cdot), k_t(\cdot, \cdot))$, where

$$\mu_t(x) = k_t(x)^T (K_t + \lambda I)^{-1} y_{1:t} \tag{2}$$
$$k_t(x, x') = k(x, x') - k_t(x)^T (K_t + \lambda I)^{-1} k_t(x')$$
$$\sigma_t^2(x) = k_t(x, x), \tag{3}$$

and $k_t(x) = [k(x_1, x), \ldots, k(x_t, x)]^T$ and $K_t = [k(u, v)]_{u,v \in \mathcal{H}_t}$. Therefore, for every $x \in \mathcal{D}$, the posterior distribution of $f(x)$, given $\mathcal{H}_t$ is $\mathcal{N}(\mu_t(x), \sigma_t^2(x))$. The following term often plays a key role in the regret bounds of

[2]The surrogate GP model described above (i.e., a GP prior and a Gaussian likelihood) is only used for the algorithm design.

GP based algorithms.

$$\gamma_t := \gamma_t(k, \mathcal{D}) = \max_{A \subset \mathcal{D}: |A| = t} \frac{1}{2} \ln |I_t + \lambda^{-1} K_A|,$$

where $K_A = [k(x, x')]_{x,x' \in A}$. Roughly speaking, γ_t is the maximum mutual information that can be obtained about the GP prior from t samples corrupted by a Gaussian channel $\mathcal{N}(0, \lambda)$. It is a function of the kernel k and domain $\mathcal{D}$. For instance, if $\mathcal{D}$ is compact and convex, then we have $\gamma_t = O((\ln t)^{d+1})$ for k_{SE}, $O(t^{\frac{d(d+1)}{2\nu + d(d+1)}} \ln t)$ for $k_{\text{Matérn}}$, and $O(d \ln t)$ for a linear kernel (Srinivas et al. 2009).

Lower Bounds

In this section, we derive the lower bounds for both k_{SE} and $k_{\text{Matérn}}$ under any LDP mechanism and any learning algorithm, as presented in the following theorem.

Theorem 1. *Let $\mathcal{D} = [0, 1]^d$ for some $d \in \mathbb{N}$. Fix a kernel $k \in \{k_{SE}, k_{Matérn}\}$, $B > 0$, $\epsilon > 0$, $T \in \mathbb{Z}$, $\delta \in (0, 1)$, $\alpha \in (0, 1]$ and $v > 0$. Given any learning algorithm, any ϵ-LDP mechanism, there exists a function $f \in \mathcal{H}_k(\mathcal{D})$ with $\|f\|_{\mathcal{H}} \leq B$, and a reward distribution satisfying $\mathbb{E}\left[|y_t|^{1+\alpha} | \mathcal{F}_{t-1}\right] \leq v$ for all $t \in [T]$, such that the following hold, respectively*

- $\mathbb{E}[R_T] = \Omega\left(v^{\frac{1}{1+\alpha}} T^{\frac{1}{1+\alpha}} \zeta^{\frac{-2\alpha}{1+\alpha}} \left(\ln \frac{B^{(1+\alpha)/\alpha} T \zeta^2}{v^{1/\alpha}}\right)^{\frac{d\alpha}{2+2\alpha}}\right)$, *where $\zeta = e^\epsilon - 1$, if $k = k_{SE}$*
- $\mathbb{E}[R_T] = \Omega\left(v^{\frac{\nu}{\nu(1+\alpha)+d\alpha}} T^{\frac{\nu+d\alpha}{\nu(1+\alpha)+d\alpha}} \tilde{\zeta} B^{\frac{d\alpha}{\nu(1+\alpha)+d\alpha}}\right)$, *where $\tilde{\zeta} = \zeta^{\frac{-2\alpha}{1+\alpha} + \frac{2d\alpha^2}{(1+\alpha)(\nu(1+\alpha)+d\alpha)}}$ and $\zeta = e^\epsilon - 1$, if $k = k_{Matérn}$.*

Remark 1. *For a small ϵ and $\alpha = 1$, and hence $\zeta \approx \epsilon$, the regret lower bounds in Theorem 1 have an additional factor of $1/\epsilon$ in front of the lower bounds for non-private case in (Chowdhury and Gopalan 2019)*[3].

Proof Sketch of Theorem 1. The proof follows the standard techniques in (Scarlett, Bogunovic, and Cevher 2017; Chowdhury and Gopalan 2019), which provide lower bounds for *non-private* BO under $i.i.d$ Gaussian noise (or heavy-tailed payoffs). The key challenge is handle the additional requirement of ϵ-LDP. To this end, we aim to relate the Kullback-Leibler (KL) divergence between two distributions P_1 and P_2 to the KL divergence between two new distributions M_1 and M_2, which are the distributions transformed from P_1 and P_2 according to a given ϵ-LDP mechanism. Inspired by (Basu, Dimitrakakis, and Tossou 2019), we resort to Theorem 1 of (Duchi, Jordan, and Wainwright 2013) and Pinskers inequality. More specifically, by Theorem 1 of (Duchi, Jordan, and Wainwright 2013), we have

$$D_{kl}(M_1||M_2) + D_{kl}(M_2||M_1) \leq 4(e^\epsilon - 1)^2 ||P_1 - P_2||^2_{TV}.$$

Then, by Pinskers inequality, we have

$$||P_1 - P_2||^2_{TV} \leq 2 D_{kl}(P_1||P_2).$$

Thus, roughly speaking, there is an additional term $(e^\epsilon - 1)^2$. The full proof is in Appendix. □

[3] for $k = k_{\text{Matérn}}$, it holds for a large ν.

Algorithms and Upper Bounds

In this section, we will present three algorithms that are able to achieve nearly optimal regret while guaranteeing ϵ-LDP. All the three algorithms rely on adding additional Laplace noise on the reward (i.e., Laplace mechanism in DP) to provide privacy guarantee. Note that, due to the additional Laplace noise, the rewards received by the learner are now no longer sub-Gaussian, and hence standard algorithms will not work. As a result, the three algorithms mainly differ in the way of handling the issue of non-sub-Gaussian rewards.

Laplace Mechanism

A commonly used mechanism in the areas of DP is the Laplace mechanism, which adds independent Laplace noise to the data point. For any $\mathcal{L} > 0$, the PDF of the Laplace($\mathcal{L}$) (i.e., mean is zero) is given by

$$\text{Laplace}(\mathcal{L}) : l(x \mid \mathcal{L}) = (2\mathcal{L})^{-1} \exp(-|x|/\mathcal{L}).$$

Thus, it is with mean 0 and variance $2\mathcal{L}^2$. The Laplace mechanism used in this paper is stated in Curator 1 and its theoretical guarantee is given by Lemma 1.

Curator 1 Convert-to-Laplace (CTL(ϵ))

On receiving a reward observation y_t:

return $\bar{y}_t := y_t + L$, where $L \sim \text{Laplace}(\mathcal{L})$ and $\mathcal{L} = \frac{2(B+R)}{\epsilon}$.

Lemma 1. *CTL(ϵ) guarantees ϵ-LDP.*

Proof. See Appendix. □

Adaptively Truncated Approximate (ATA) Algorithm and Regret

One direct way of handling non-sub-Gaussian rewards in BO is to utilize the recently developed technique for heavy-tailed payoffs (Chowdhury and Gopalan 2019). In particular, the authors show that when combining a good feature approximation (e.g., Nyström approximation) and a feature adaptive truncation of rewards (e.g., TOFU in (Shao et al. 2018)), one can obtain a regret bound roughly $\tilde{O}(\gamma_T T^{\frac{1}{1+\alpha}})$, when the $(1+\alpha)$-th moment of the reward is finite and $\alpha \in (0, 1]$. Hence, when $\alpha = 1$, it recovers the regret bounds under sub-Gaussian rewards (Chowdhury and Gopalan 2017).

Thus, it is natural to adapt ATA-GP-UCB introduced in (Chowdhury and Gopalan 2019) to handle the non-sub-Gaussian payoffs caused by the Laplace noise in the LDP setting, which leads to the LDP-ATA-GP-UCB, as described in Algorithm 1.

Further, by adapting the regret analysis of ATA-GP-UCB in (Chowdhury and Gopalan 2019), we have the following theorem for the regret upper bound of LDP-ATA-GP-UCB.

Theorem 2. *Let $f \in \mathcal{H}_k(\mathcal{D})$ with $\|f\|_{\mathcal{H}} \leq B$ for all $x \in \mathcal{D}$ and noise η_t is bounded by R. Fix $\epsilon > 0$, $\varepsilon \in (0, 1)$ and set $\rho = \frac{1+\varepsilon}{1-\varepsilon}$, and $v = B^2 + R^2 + 8(B+R)^2/\epsilon^2$. Then, for any $\delta \in (0, 1]$, LDP-ATA-GP-UCB with parameters $q = 6\rho \ln(4T/\delta)/\varepsilon^2$, $b_t = \sqrt{v / \ln(4m_t T/\delta)}$ and $\beta_{t+1} = B(1+$*

Algorithm 1 LDP-ATA-GP-UCB

Input: Parameters λ, B, R, $\epsilon > 0$, $\{b_t\}_{t\geq 1}$, $\{\beta_t\}_{t\geq 1}$, and q.
Set: $\tilde{\mu}_0(x) = 0$ and $\tilde{\sigma}_0(x) = k(x,x)$ for all $x \in \mathcal{D}$.
for $t = 1, 2, 3, \ldots, T$ **do**
  Play $x_t = \arg\max_{x\in\mathcal{D}} \tilde{\mu}_{t-1}(x) + \beta_t(x)\tilde{\sigma}_{t-1}(x)$
  Receive private response $\bar{y}_t$ from CTL(ϵ)
  Set m_t as the dimension of $\tilde{\varphi}_t$
  Set $\tilde{\Phi}_t^T = [\tilde{\varphi}_t(x_1), \ldots, \tilde{\varphi}_t(x_t)]$ and $\tilde{V}_t = \tilde{\Phi}_t^T\tilde{\Phi}_t + \lambda I_{m_t}$
  Find the rows $u_1^T \ldots, u_{m_t}^T$ of $\tilde{V}_t^{-1/2}\tilde{\Phi}_t^T$
  Set $\hat{r}_i = \sum_{\tau=1}^t u_{i,\tau}\bar{y}_\tau \mathbb{1}_{|u_{i,\tau}\bar{y}_\tau| \leq b_\tau}$ for $i \in [m_t]$
  Set $\tilde{\theta}_t = \tilde{V}_t^{-1/2}[\hat{r}_1, \ldots, \hat{r}_{m_t}]^T$
  Set $\tilde{\mu}_t(x) = \tilde{\varphi}_t(x)^T\tilde{\theta}_t$
  Set $\tilde{\sigma}_t^2(x) = k(x,x) - \tilde{\varphi}_t(x)^T\tilde{\varphi}_t(x) + \lambda\tilde{\varphi}_t(x)^T\tilde{V}_t^{-1}\tilde{\varphi}_t(x)$
end for

$\frac{1}{\sqrt{1-\varepsilon}}) + 4\sqrt{\ln(4m_tT/\delta)vm_t/\lambda}$, *with probability at least* $1-\delta$, *has regret bound*

$$R_T = O\left(\hat{\rho}B\sqrt{T\gamma_T} + \frac{\rho}{\varepsilon^2}\varphi_T\right),$$

in which $\hat{\rho} := \rho\left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)$, *and* $\varphi_T := \gamma_T\sqrt{T}\sqrt{v\ln(T/\delta)\ln(\frac{\gamma_T T\ln(T/\delta)}{\delta})}$.

Remark 2. *Note that by substituting the value of v into the regret bound, we obtain that $R_T = \tilde{O}(\gamma_T\sqrt{T}/\epsilon)$. That is, it has a factor of $1/\epsilon$ compared to the non-private case, which matches the same scaling of ϵ in the lower bounds as shown in Theorem 1. Moreover, LDP-ATA-GP-UCB enjoys the same scaling with respect to both γ_T and T as in the state-of-the-art non-private sub-Gaussian case.*

Although the LDP-ATA-GP-UCB algorithm achieves almost optimal regret bound, it might be a 'overkill' for the LDP setting. In particular, the original setting for the ATA-GP-UCB algorithm in (Chowdhury and Gopalan 2019) only assumes at most a finite variance. However, in our LDP setting, the corrupted reward $\bar{y}$ has all the moments being bounded and enjoys an exponential-type tail. In other words, although the additional Laplace noise causes the corrupted reward to be no longer sub-Gaussian, it still enjoys better properties compared to the general conditions for the ATA-GP-UCB algorithm to work. Therefore, it seems that there is some hope that we can design simple algorithm to achieve the same regret bound. Another issue of ATA-GP-UCB is its computational complexity. As pointed out by (Shao et al. 2018) in the linear bandit setting (ATA-GP-UCB reduces to TOFU), for each round, it needs to truncate all the historical payoffs, which leads to a high complexity.

Based on the discussions above, in the following, we will propose two novel algorithms that are also able to achieve almost optimal regret while substantially reducing the implementation and computational complexity of LDP-ATA-GP-UCB.

Algorithm 2 LDP-TGP-UCB

Input: Parameters B, R, $\epsilon > 0$, λ, δ.
Set: $K = B^2 + R^2 + 2\mathcal{L}^2$
for $t = 1, 2, 3, \ldots, T$ **do**
  Set $b_{t-1} = B + R + \mathcal{L}\ln(t-1)$
  Set $\beta_t = B + \frac{2\sqrt{2}}{\sqrt{\lambda}}b_{t-1}\sqrt{\gamma_{t-1} + \ln(1/\delta)} + \frac{1}{\sqrt{\lambda}}\sqrt{K(\ln(t-1)+1)}$
  Play $x_t = \arg\max_{x\in\mathcal{D}} \hat{\mu}_{t-1}(x) + \beta_t\sigma_{t-1}(x)$
  Receive private response $\bar{y}_t$ from CTL(ϵ).
  Set $\hat{y}_t = \bar{y}_t\mathbb{1}_{|\bar{y}_t|\leq b_t}$ and $\hat{Y}_t = [\hat{y}_1, \ldots, \hat{y}_t]^T$
  Set $\hat{\mu}_t(x) = k_t(x)^T(K_t + \lambda I)^{-1}\hat{Y}_t$
  Set $\sigma_t^2(x) = k(x,x) - k_t(x)^T(K_t + \lambda I)^{-1}k_t(x)$
end for

Raw Reward Truncation Algorithm and Regret

In this section, instead of using the sophisticated truncation in the feature space as in LDP-ATA-GP-UCB, we turn to adopt the simple truncation on the raw rewards. In the general heavy-tail reward setting (with at most a finite variance), (Chowdhury and Gopalan 2019) proposed TGP-UCB algorithm which truncates the reward to zero if it is larger than a truncated point b_t for round t. Specifically, for a finite variance case, the truncated point b_t is $\theta(t^{1/4})$ in TGP-UCB, which finally leads to an regret bound of $\tilde{O}(T^{3/4})$. Hence, it has an additional factor $O(T^{1/4})$ when compared to the regret bound for the sub-Gaussian case. This means that we cannot directly adopt TGP-UCB to achieve the same regret bound as in LDP-ATA-GP-UCB of the last section.

However, as pointed before, the corrupted reward $\bar{y}$ has a nice exponential tail property. This suggests that a truncated point of order $O(\ln t)$ is enough, which will only in turn incurs an additional $O(\ln T)$ factor in the regret. Based on this idea, we propose the LDP-TGP-UCB algorithm, as described in Algorithm 2.

Moreover, by refining the regret analysis of TGP-UCB in (Chowdhury and Gopalan 2019), we can obtain the following theorem for the regret bound of LDP-TGP-UCB.

Theorem 3. *Fix $\epsilon > 0$. Let $f \in \mathcal{H}_k(\mathcal{D})$ with $\|f\|_{\mathcal{H}} \leq B$ for all $x \in \mathcal{D}$ and the noise η_t is bounded by R for all t. Then, for any $\delta \in (0,1]$, LDP-TGP-UCB achieves, with probability at least $1-\delta$, the regret bound*

$$R_T = O\left(\vartheta\sqrt{\ln T\gamma_T T} + \vartheta\sqrt{T}\ln T\sqrt{\gamma_T(\gamma_T + \ln(1/\delta))}\right),$$

where $\vartheta = (B+R)/\epsilon$.

Remark 3. *As in LDP-ATA-GP-UCB, LDP-TGP-UCB is also able to achieve regret bound $\tilde{O}(\gamma_T\sqrt{T}/\epsilon)$. The advantage of LDP-TGP-UCB is its simple implementation in the sense that each reward is only trimmed once.*

Proof Sketch of Theorem 3. As in most GP-UCB like algorithms (Chowdhury and Gopalan 2017, 2019), the key step boils down to establishing a (high-probability) confidence interval bound, i.e., in our setting,

$$|f(x) - \hat{\mu}_t(x)| \leq \beta_{t+1}\sigma_t(x).$$

To this end, with some linear algebra calculations, we have

$$|f(x) - \hat{\mu}_t(x)| \leq \left(B + \lambda^{-1/2} \|\sum_{\tau=1}^{t} \hat{\eta}_\tau \varphi(x_\tau)\|_{V_t^{-1}} \right) \sigma_t(x),$$

where $\hat{\eta}_t = \hat{y}_t - f(x_t)$, $\varphi(x) := k(x, \cdot)$, which maps $x \in \mathbb{R}^d$ to RKHS H associated with kernel function k and $V_t = \Phi_t^T \Phi_t + \lambda I_{\mathcal{H}}$, $\Phi_t := [\varphi(x_1)^T, \ldots, \varphi(x_t)^T]^T$.

The key term is $\|\sum_{\tau=1}^{t} \hat{\eta}_\tau \varphi(x_\tau)\|_{V_t^{-1}}$, which can be handled by the self-normalized inequality if $\hat{\eta}_\tau$ is sub-Gaussian. However, in our setting, it is not. To overcome this issue, we will divide it into two parts. In particular, similar to (Chowdhury and Gopalan 2019), we define $\xi_t = \hat{\eta}_t - \mathbb{E}[\hat{\eta}_t \mid \mathcal{F}_{t-1}]$. Now, the key term can be written as

$$\begin{aligned} &\|\sum_{\tau=1}^{t} \hat{\eta}_\tau \varphi(x_\tau)\|_{V_t^{-1}} \\ &= \underbrace{\|\sum_{\tau=1}^{t} \xi_\tau \phi(x_\tau)\|_{V_t^{-1}}}_{\mathcal{T}_1} + \underbrace{\|\sum_{\tau=1}^{t} \mathbb{E}[\hat{\eta}_\tau | \mathcal{F}_{\tau-1}] \phi(x_\tau)\|_{V_t^{-1}}}_{\mathcal{T}_2}. \quad (4) \end{aligned}$$

For $\mathcal{T}_1$, note that $\xi_t = \hat{y}_t - \mathbb{E}[\hat{y}_t \mid \mathcal{F}_{t-1}]$, which is bounded by $2b_t$, and hence sub-Gaussian. Thus, by the self-normalized inequality for the RKHS-valued process in (Durand, Maillard, and Pineau 2018; Chowdhury and Gopalan 2019), we can bound $\mathcal{T}_1$ as follows

$$\mathcal{T}_1 \leq 2b_t \sqrt{2(\gamma_t + \ln(1/\delta))}.$$

For $\mathcal{T}_2$, with some linear algebra, we can first bound it as $\sqrt{\sum_{\tau=1}^{t} \mathbb{E}[\hat{\eta}_\tau | \mathcal{F}_{\tau-1}]^2}$. Further, note that $\mathbb{E}[\hat{\eta}_\tau | \mathcal{F}_{\tau-1}] = -\mathbb{E}\left[\bar{y}_\tau \mathbb{1}_{|\tilde{y}_\tau| > b_\tau} | \mathcal{F}_{\tau-1}\right]$. Hence, by Cauchy-Schwartz inequality with $b_\tau = B + R + \mathcal{L} \ln \tau$, we have

$$\mathbb{E}[\hat{\eta}_\tau | \mathcal{F}_{\tau-1}]^2 \leq \mathbb{E}\left[\bar{y}_\tau^2 | \mathcal{F}_{\tau-1}\right] \mathbb{P}(|L| > \mathcal{L} \ln \tau) \leq K \frac{1}{\tau},$$

where $K := B^2 + R^2 + 2\mathcal{L}^2$. The last inequality holds since $|L| \sim \text{Exp}(1/\mathcal{L})$. Therefore, by the property of Harmonic sum, we have

$$\mathcal{T}_2 \leq \sqrt{K(\ln t + 1)}.$$

Hence, the (high-probability) confidence interval bound is obtained by setting

$$\beta_{t+1} = B + \frac{2\sqrt{2}}{\sqrt{\lambda}} b_t \sqrt{\gamma_t + \ln(1/\delta)} + \frac{1}{\sqrt{\lambda}} \sqrt{K(\ln t + 1)}.$$

The full proof is relegated to Appendix. □

It is worth pointing out the truncation trick used in LDP-TGP-UCB also sheds light on the regret bounds for non-private BO under payoffs that are beyond sub-Gaussian, e.g., sub-Weibull which includes sub-Gaussian and sub-exponential as special cases (Vladimirova et al. 2019). More specifically, according to (Vladimirova et al. 2019), a random variable X is said to be a sub-Weibull with tail parameter θ, i.e., $X \sim \text{subW}(\theta)$, if for some constants a and b such that

$$\mathbb{P}(|X| \geq x) \leq a \exp(-b x^{1/\theta}), \quad \text{for all } x > 0. \quad (5)$$

It can be seen that sub-Gaussian and sub-exponential distributions are special cases of sub-Weibull with $\theta = 1/2$ and $\theta = 1$, respectively. Thus, instead of choosing the truncation point $b_t = O(\ln t)$ as in LDP-TGP-UCB, one turn to choose $b_t = O((\ln t)^\theta)$, which in turn only incurs a log factor in the regret bound. As a result, with this simple truncation, the non-private BO under sub-Weibull noise is still $\tilde{O}(\gamma_T \sqrt{T})$.

Median of Means Approximate (MoMA) Algorithm and Regret

In this section, we will introduce a new BO method that is able to achieve almost optimal regret bounds under general heavy-tailed payoffs. Hence, the LDP setting is just a special case. This new method is mainly inspired by the MENU algorithm in (Shao et al. 2018), which is introduced to handle the heavy-tailed payoffs in the linear bandit setting. Moreover, it has been shown to have a lower complexity than TOFU (which is the core of ATA-GP-UCB). The key idea in MENU is based on median of means techniques (Bubeck, Cesa-Bianchi, and Lugosi 2013). The main challenge in generalizing MENU to the BO setting is to handle the possibly infinite feature dimension associated with the kernel function. To this end, we will again use kernel approximation techniques (i.e., Nyström approximation) developed in (Calandriello et al. 2019; Chowdhury and Gopalan 2019). However, instead of updating the approximations after every iteration as in ATA-GP-UCB, in our new method the approximation is only updated after each 'epoch', which is composed of multiple iterations. This further reduces its complexity.

This new algorithm is called MoMA-GP-UCB, which is presented in Algorithm 3. With the aid of Fig. 1 (adapted from (Shao et al. 2018)), we can easily see that the total number of T iterations are divided into N epochs, each consisted of k iterations. The algorithm will loop over each $n = 1, \ldots, N$ epoch. Within each epoch n, a point x_n is selected in a GP-UCB fashion, and the selected point x_n will be played k times with corresponding rewards. Then, the kernel approximation terms are updated, i.e., $\tilde{\varphi}_n$,$\tilde{\Phi}_n$ and $\tilde{V}_n$. Following this update, it will calculate k least-square-estimates (LSE), each is based on the rewards along each row $j \in [k]$ (e.g., using the data in the pink row to generate the pink LSE, and similarly green data for the green LSE). Next, it applies median-of-means techniques to find the best LSE $\tilde{\theta}_{n,k^*}$ for epoch n. Finally, the posterior mean and variance are updated.

Now, we have the following theorem for the regret bound of MoMA-GP-UCB under general heavy-tailed payoffs.

Theorem 4. *Let $f \in \mathcal{H}_k(\mathcal{D})$ with $\|f\|_{\mathcal{H}} \leq B$ for all $x \in \mathcal{D}$. Assume that $\mathbb{E}\left[|\eta_t|^{1+\alpha} \mid \mathcal{F}_{t-1}\right] \leq c$. Fix $\varepsilon \in (0, 1)$ and set $\rho = \frac{1+\varepsilon}{1-\varepsilon}$. Then, for any $\delta \in (0, 1]$, MoMA-GP-UCB with parameters $q = 6\rho \ln(4T/\delta)/\varepsilon^2$, $k = \lceil 24 \ln\left(\frac{4eT}{\delta}\right) \rceil$, and $\beta_{n+1} = B(1 + \frac{1}{\sqrt{1-\varepsilon}}) + 3\left((9m_n c)^{\frac{1}{1+\alpha}} n^{\frac{1-\alpha}{2(1+\alpha)}}\right)$, with*

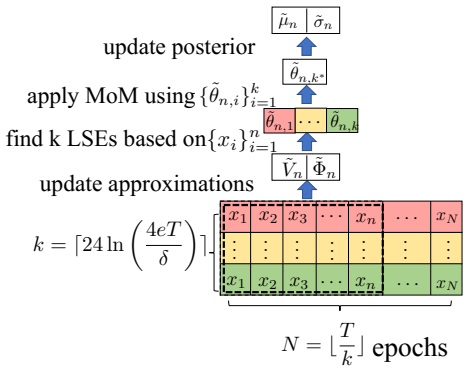


Figure 1: Illustration of MoMA-GP-UCB

probability at least $1-\delta$*, has regret bound*

$$R_T = O\left(\hat{B}\sqrt{\gamma_T T \ln\frac{T}{\delta}} + Z \ln\frac{T}{\delta} c^{\frac{1}{1+\alpha}} T^{\frac{1}{1+\alpha}} \gamma_T^{\frac{3+\alpha}{2(1+\alpha)}}\right),$$

where $\hat{B} = \rho B(1+\frac{1}{\sqrt{1-\varepsilon}})$ *and* $Z = \left(\frac{\rho^{3+\alpha}}{\varepsilon^2}\right)^{\frac{1}{1+\alpha}}$.

Remark 4. *Note that when* $\alpha = 1$*, MoMA-GP-UCB recovers the same regret bound* $\tilde{O}(\gamma_T\sqrt{T})$ *as in the sub-Gaussian case. Moreover, for the special linear kernel case, substituting* $\gamma_T = O(d \ln T)$*, the bound in Theorem 4 recovers the regret bound in (Shao et al. 2018) up to logarithmic factor.*

Proof Sketch of Theorem 4. The proof is mainly inspired by (Shao et al. 2018; Chowdhury and Gopalan 2019). The key step is again a (high-probability) confidence interval bound, i.e.,

$$|f(x) - \tilde{\mu}_n(x)| \le \beta_{n+1}\tilde{\sigma}_n(x). \tag{6}$$

Assume that the kernel approximation is ε-accurate, the LHS of Eq. (6) can be bounded by

$$|f(x) - \tilde{\mu}_n(x)| \le B(1+\frac{1}{\sqrt{1-\varepsilon}})\tilde{\sigma}_n(x) + \lambda^{-1/2}\|\tilde{V}_n^{-1}\tilde{\Phi}_n^T f_n - \tilde{\theta}_{n,k^*}\|_{\tilde{V}_n}\tilde{\sigma}_n(x), \tag{7}$$

where $f_n = [f(x_1), \ldots, f(x_n)]^T$, i.e., a vector containing f's evaluations up to epoch n. Now, we need to focus on the term $\|\tilde{V}_n^{-1}\tilde{\Phi}_n^T f_n - \tilde{\theta}_{n,k^*}\|_{\tilde{V}_n}$. To this end, we first establish the following result regarding the k LSEs. In particular, for $j \in [k]$, we have

$$\mathbb{P}\left(\|\tilde{V}_n^{-1}\tilde{\Phi}_n^T f_n - \tilde{\theta}_{n,j}\|_{\tilde{V}_n} \le \gamma\right) \ge \frac{3}{4},$$

where $\gamma := (9m_n c)^{\frac{1}{1+\alpha}} n^{\frac{1-\alpha}{2(1+\alpha)}}$. Based on this result, by the choice of k^*, we obtain that if $k = \lceil 24\ln\left(\frac{eT}{\delta}\right)\rceil$, then for all $n \in [N]$, with probability at least $1-\delta$,

$$\|\tilde{V}_n^{-1}\tilde{\Phi}_n^T f_n - \tilde{\theta}_{n,k^*}\|_{\tilde{V}_n} \le 3\gamma. \tag{8}$$

Combining Eqs. (8) and (7), yields that, under the event that the kernel approximation is ε-accurate, with probability at least $1-\delta$,

$$|f(x) - \tilde{\mu}_n(x)| \le \left(B(1+\frac{1}{\sqrt{1-\varepsilon}}) + \lambda^{-1/2}3\gamma\right)\tilde{\sigma}_n(x)$$

Algorithm 3 MoMA-GP-UCB

1: **Input:** Parameters λ, δ, $\{\beta_t\}_{t\ge1}$, and q.
2: **Set:** $\tilde{\mu}_0(x) = 0$ and $\tilde{\sigma}_0(x) = \tilde{k}(x,x)$ for all $x \in \mathcal{D}$.
3: **Set:** $k = \lceil 24\ln\left(\frac{4eT}{\delta}\right)\rceil$ and $N = \lfloor \frac{T}{k} \rfloor$.
4: **for** $n = 1, 2, 3, \ldots, N$ **do**
5: $\quad x_n = \arg\max_{x\in\mathcal{D}} \tilde{\mu}_{n-1}(x) + \beta_n(x)\tilde{\sigma}_{n-1}(x)$
6: $\quad$ Play x_n with k times and observe rewards $y_{n,1}, y_{n,2}, \ldots, y_{n,k}$.
7: $\quad \tilde{\varphi}_n(x) = \text{NyströmEmbed}(\{(x_i, \tilde{\sigma}_{n-1}(x_i))\}_{i=1}^n, q)$
8: $\quad$ Set m_n as the dimension of $\tilde{\varphi}_n$
9: $\quad$ Set $\tilde{\Phi}_n^T = [\tilde{\varphi}_n(x_1), \ldots, \tilde{\varphi}_n(x_n)]$ and $\tilde{V}_n = \tilde{\Phi}_n^T\tilde{\Phi}_n + \lambda I_{m_n}$
10: $\quad$ For $j \in [k]$, $\tilde{\theta}_{n,j} = \tilde{V}_n^{-1}\sum_{i=1}^n y_{i,j}\tilde{\varphi}_n(x_i)$
11: $\quad$ For $j \in [k]$, $r_j = \text{median}(\{\|\tilde{\theta}_{n,j} - \tilde{\theta}_{n,s}\|_{\tilde{V}_n} : s \in [k]\setminus j\})$
12: $\quad$ Set $k^* = \arg\min_{j\in[k]} r_j$
13: $\quad$ Set $\tilde{\mu}_n(x) = \tilde{\varphi}_n(x)^T\tilde{\theta}_{n,k^*}$
14: $\quad$ Set $\tilde{\sigma}_n^2(x) = k(x,x) - \tilde{\varphi}_n(x)^T\tilde{\varphi}_n(x) + \lambda\tilde{\varphi}_n(x)^T\tilde{V}_n^{-1}\tilde{\varphi}_n(x)$
15: **end for**

for all $n \in [N]$ when $k = \lceil 24\ln\left(\frac{eT}{\delta}\right)\rceil$. Since for any δ, the kernel approximation (under given parameters) is ε-accurate with probability at least $1-\delta$, by the virtue of union bound, we have that when $k = \lceil 24\ln\left(\frac{2eT}{\delta}\right)\rceil$,

$$|f(x) - \tilde{\mu}_n(x)| \le \beta_{n+1}\tilde{\sigma}_n(x) \tag{9}$$

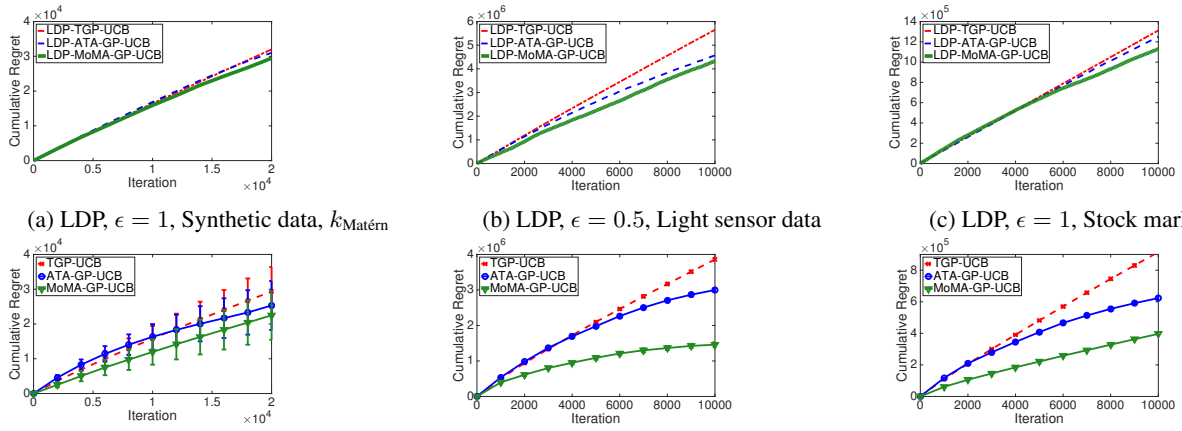
for all $n \in [N]$, where $\beta_{n+1} := B(1+\frac{1}{\sqrt{1-\varepsilon}}) + \lambda^{-1/2}3\gamma$. Finally, by the nice properties of Nyström approximation, we can obtain that with probability at least $1-\delta$, both $m_n = O(\frac{\rho^2}{\varepsilon^2}\gamma_n \ln(T/\delta))$ and $\tilde{\sigma}_{n-1}(x_n) \le \rho\sigma_{n-1}(x_n)$. Then, the regret bound follows from that $R_T = 2k\sum_{n=1}^N \beta_n\tilde{\sigma}_{n-1}(x)$ along with standard GP-UCB analysis. The full proof is relegated to Appendix. □

Now, we can easily design the private version of it, called LDP-MoMA-GP-UCB. The only difference is line 6, in which LDP-MoMA-GP-UCB received private response from CTL(ϵ). Thus, we can directly obtain the regret bound of LDP-MoMA-GP-UCB as a corollary of Theorem 4.

Corollary 1. *Let* $f \in \mathcal{H}_k(\mathcal{D})$ *with* $\|f\|_{\mathcal{H}} \le B$ *for all* $x \in \mathcal{D}$ *and noise* η_t *is bounded by* R*. Fix* $\epsilon > 0$, $\varepsilon \in (0,1)$ *and set* $\rho = \frac{1+\varepsilon}{1-\varepsilon}$*, and* $c = R^2 + 8(B+R)^2/\epsilon^2$*. Then, for any* $\delta \in (0,1]$*, LDP-MoMA-GP-UCB with parameters* $q = 6\rho\ln(4T/\delta)/\varepsilon^2$*,* $k = \lceil 24\ln\left(\frac{4eT}{\delta}\right)\rceil$ *and* $\beta_{n+1} = B(1+\frac{1}{\sqrt{1-\varepsilon}}) + 3\left((9m_nc)^{\frac{1}{1+\alpha}} n^{\frac{1-\alpha}{2(1+\alpha)}}\right)$*, with probability at least* $1-\delta$*, has regret bound*

$$R_T = O\left(\hat{B}\sqrt{\gamma_T T\ln\frac{T}{\delta}} + Z\ln\frac{T}{\delta}\frac{B+R}{\epsilon}\gamma_T\sqrt{T}\right),$$

where $\hat{B} = \rho B(1+\frac{1}{\sqrt{1-\varepsilon}})$ *and* $Z = \left(\frac{\rho^{3+\alpha}}{\varepsilon^2}\right)^{\frac{1}{1+\alpha}}$.



(a) LDP, $\epsilon = 1$, Synthetic data, $k_{\text{Matérn}}$ (b) LDP, $\epsilon = 0.5$, Light sensor data (c) LDP, $\epsilon = 1$, Stock market data.

(d) Non-private, Synthetic data, k_{SE} (e) Non-private, Light sensor data (f) Non-private, Stock market data

Figure 2: (a)-(c) Cumulative regrets for three LDP algorithms; (d)-(f) Cumulative regrets (and standard variance) for non-private versions on heavy-tailed data.

Unbounded Noise Case

In the case of unbounded noise, we show that Laplace mechanism can ensure (ϵ, δ)-LDP (weaker than ϵ-LDP), see Appendix.

Experiments

We conduct experiments to compare the performance of the three private algorithms (i.e., LDP-ATA-GP-UCB, LDP-TGP-UCB, LDP-MoMA-GP-UCB) and the performance of three non-private BO methods for general heavy-tailed payoffs (i.e., ATA-GP-UCB, TGP-UCB in (Chowdhury and Gopalan 2019) and MoMA-GP-UCB proposed in this paper). As in (Chowdhury and Gopalan 2019), the parameters used for each algorithm are set order-wise similar to those recommended by the theorems. We run each algorithm for 10 independent trials and plot the average of cumulative regret along with time evolution.

Datasets and Settings

Synthetic data. The domain $\mathcal{D}$ is generated by discretizing $[0, 1]$ uniformly into 100 points. The black-box function $f = \sum_{i=1}^{p} a_i k(\cdot, x_i)$ is generated by uniformly sampling $a_i \in [-1, 1]$ and support points $x_i \in \mathcal{D}$ with $p = 100$. The parameters for the kernel function are $l = 0.2$ for k_{SE} and $l = 0.2$, $\nu = 2.5$ for $k_{\text{Matérn}}$. We set $B = \max_{x \in \mathcal{D}} |f(x)|$ and $y(x) = f(x) + \eta$. For the LDP case, the noise η is uniformly sampled in $[-1, 1]$ and hence $R = 1$. For the non-private heavy-tailed case, the noise η are samples from the Student's t-distribution with 3 degrees of freedom. Hence, $v = B^2 + 3$ and $c = 3$.

Light sensor data. This data is collected in the CMU Intelligent Workplace in Nov 2005, which is available online as Matlab structure[4] and contains locations of 41 sensors, 601 train samples and 192 test samples. We use it in the context of finding the maximum average reading of the sensors. For fair comparison, the settings for this dataset follow from (Chowdhury and Gopalan 2019),which has shown that the payoffs are heavy-tailed. In particular, f is set as empirical average of the test samples, with B set as its maximum, and k is set as the empirical covariance of the normalized train samples. The noise is estimated by taking the difference between the test samples and its empirical mean (i.e., f), and R is set as the maximum. Here, we consider $\alpha = 1$, set v as the empirical mean of the squared readings of test samples, and c is the empirical mean of the squared noise.

Stock market data. This dataset is the adjusted closing price of 29 stocks from January 4th, 2016 to April 10th, 2019. We use it in the context of identifying the most profitable stock in a given pool of stocks. As verified in (Chowdhury and Gopalan 2019), the rewards follows from heavy-tailed distribution. We take the empirical mean of stock prices as our objective function f and empirical covariance of the normalized stock prices as our kernel function k. The noise is estimated by taking the difference between the raw prices and its empirical mean (i.e., f), with R set as the maximum. Consider $\alpha = 1$, with v set as the empirical mean of the squared prices and c set as the empirical mean of squared noise.

Results

From Figure 2, we can see that MoMA-GP-UCB (or LDP-MoMA-GP-UCB) tends to empirically outperform the other two algorithms in both non-private and private settings. We also conduct additional experiments (relegated to Appendix), and similar observations are obtained. Note that similar to (Chowdhury and Gopalan 2019), the high error bar in (d) is because a different f is chosen for each trial.

Conclusion

We derived regret lower bounds for LDP BO and presented three almost optimal algorithms. We also proposed MoMA-GP-UCB. It complements previous BO algorithms for heavy-tailed payoffs and has superior performance with a reduced complexity.

[4] http://www.cs.cmu.edu/~guestrin/Class/10708-F08/projects

References

Abbasi-Yadkori, Y.; Pál, D.; and Szepesvári, C. 2011. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, 2312–2320.

Basu, D.; Dimitrakakis, C.; and Tossou, A. 2019. Differential Privacy for Multi-armed Bandits: What Is It and What Is Its Cost? *arXiv preprint arXiv:1905.12298* .

Bubeck, S.; Cesa-Bianchi, N.; and Lugosi, G. 2013. Bandits with heavy tail. *IEEE Transactions on Information Theory* 59(11): 7711–7717.

Calandriello, D.; Carratino, L.; Lazaric, A.; Valko, M.; and Rosasco, L. 2019. Gaussian process optimization with adaptive sketching: Scalable and no regret. *arXiv preprint arXiv:1903.05594* .

Chowdhury, S. R.; and Gopalan, A. 2017. On kernelized multi-armed bandits. *arXiv preprint arXiv:1704.00445* .

Chowdhury, S. R.; and Gopalan, A. 2019. Bayesian optimization under heavy-tailed payoffs. In *Advances in Neural Information Processing Systems*, 13790–13801.

Cormode, G.; Jha, S.; Kulkarni, T.; Li, N.; Srivastava, D.; and Wang, T. 2018. Privacy at scale: Local differential privacy in practice. In *Proceedings of the 2018 International Conference on Management of Data*, 1655–1658.

Duchi, J. C.; Jordan, M. I.; and Wainwright, M. J. 2013. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 429–438. IEEE.

Durand, A.; Maillard, O.-A.; and Pineau, J. 2018. Streaming kernel regression with provably adaptive mean, variance, and regularization. *The Journal of Machine Learning Research* 19(1): 650–683.

Dwork, C.; Roth, A.; et al. 2014. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science* 9(3–4): 211–407.

Gajane, P.; Urvoy, T.; and Kaufmann, E. 2018. Corrupt bandits for preserving local privacy. In *Algorithmic Learning Theory*, 387–412. PMLR.

Kasiviswanathan, S. P.; Lee, H. K.; Nissim, K.; Raskhodnikova, S.; and Smith, A. 2011. What can we learn privately? *SIAM Journal on Computing* 40(3): 793–826.

Kusner, M.; Gardner, J.; Garnett, R.; and Weinberger, K. 2015. Differentially private Bayesian optimization. In *International conference on machine learning*, 918–927.

Lai, T. L.; and Robbins, H. 1985. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics* 6(1): 4–22.

Li, L.; Chu, W.; Langford, J.; and Schapire, R. E. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, 661–670.

Mutny, M.; and Krause, A. 2018. Efficient high dimensional bayesian optimization with additivity and quadrature fourier features. In *Advances in Neural Information Processing Systems*, 9005–9016.

Rasmussen, C. E. 2003. Gaussian processes in machine learning. In *Summer School on Machine Learning*, 63–71. Springer.

Ren, W.; Zhou, X.; Liu, J.; and Shroff, N. B. 2020. Multi-Armed Bandits with Local Differential Privacy. *arXiv preprint arXiv:2007.03121* .

Scarlett, J.; Bogunovic, I.; and Cevher, V. 2017. Lower bounds on regret for noisy gaussian process bandit optimization. *arXiv preprint arXiv:1706.00090* .

Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; and De Freitas, N. 2015. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE* 104(1): 148–175.

Shao, H.; Yu, X.; King, I.; and Lyu, M. R. 2018. Almost optimal algorithms for linear stochastic bandits with heavy-tailed payoffs. In *Advances in Neural Information Processing Systems*, 8420–8429.

Srinivas, N.; Krause, A.; Kakade, S. M.; and Seeger, M. 2009. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995* .

Vladimirova, M.; Girard, S.; Nguyen, H.; and Arbel, J. 2019. Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions. *arXiv preprint arXiv:1905.04955* .

Zheng, K.; Cai, T.; Huang, W.; Li, Z.; and Wang, L. 2020. Locally Differentially Private (Contextual) Bandits Learning. *arXiv preprint arXiv:2006.00701* .