Differentially Private Kernelized Contextual Bandits

Nikola Pavlovic
Cornell University

Sudeep Salgia
Carnegie Mellon

Qing Zhao
Cornell University

Abstract

We consider the problem of contextual kernel bandits with stochastic contexts, where the underlying reward function belongs to a known Reproducing Kernel Hilbert Space (RKHS). We study this problem under the additional constraint of joint differential privacy, where the agents needs to ensure that the sequence of query points is differentially private with respect to both the sequence of contexts and rewards. We propose a novel algorithm that improves upon the state of the art and achieves an error rate of $\mathcal{O}\left(\sqrt{\frac{\gamma_T}{T}+\frac{\gamma_T}{T\varepsilon}}\right)$ after T queries for a large class of kernel families, where γ_T represents the effective dimensionality of the kernel and $\varepsilon > 0$ is the privacy parameter. Our results are based on a novel estimator for the reward function that simultaneously enjoys high utility along with a low-sensitivity to observed rewards and contexts, which is crucial to obtain an order optimal learning performance with improved dependence on the privacy parameter.

1 INTRODUCTION

We study the problem of contextual kernel bandits, where an agent aims to maximize an unknown reward function based on noisy observations of the function at sequentially queried points. Specifically, at each time instant t, the agent is presented with a context $c_t \in \mathcal{C}$, based on which it takes an action $x_t \in \mathcal{X}$ and receives a noisy value of the reward $f(x_t, c_t)$. In this work, we consider the case where the reward function $f : \mathcal{X} \times \mathcal{C} \to \mathbb{R}$ belongs to a Reproducing Kernel Hilbert space (RKHS) of a known kernel k and the contexts c_t are drawn i.i.d. from a context distribution κ. Kernel bandits offer significantly more modelling capabilities compared to their linear counter-parts . In particular, it is known that the RKHS of typical kernels, such as the Matérn family of kernels, can approximate almost all continuous functions on compact subsets of $\mathbb{R}^d$ [Srinivas et al., 2010]. We measure the performance of an agent using expected predictive error rate, which is akin to the popular notion of simple regret, adapted for the contextual setting. In particular, after T rounds of interaction, let $\widehat{x}_T(c_{T+1})$ denote the output of the algorithm $\mathscr{A}$ for an observed context c_{T+1}. Then the error rate of the algorithm is defined as

$$\mathrm{ER}(\mathscr{A}) = \tag{1}$$
$$= \mathop{\mathbb{E}}_{c_{T+1}\sim\kappa}\left[\sup_{x\in\mathcal{X}} f(c_{T+1}, x) - f(c_{T+1}, \widehat{x}_T(c_{T+1}))\right]$$

The learning objective of the agent is to minimize the worst-case error rate over the class functions with a given bounded RKHS norm.

1.1 Private Kernel Bandits

In many applications, the contexts and the rewards may carry sensitive information, that might be inadvertently revealed by the algorithm through its choice of query points. For example, consider the problem of learning a recommendation system for an online shopping platform. At each time instant, the learning agent observes a random user along with their associated information, e.g., their search and purchase history, and then chooses a product to recommend and subsequently observes whether the user interacts with the recommended item. The objective for the learning agent is to design a predictor with a low error rate on a newly observed user from the distribution. The analogy to the contextual bandit setting is almost immediate — the user associated information serves as a context, the recommended product is the action and the user feedback is the

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

reward. While the user related information, and user feedback, provide valuable data for learning a good predictor, they contain private information about the user that needs to be protected.

This motivates the problem of contextual kernel bandits under privacy constraints. We adopt the framework of joint differential privacy (JDP)[Shariff and Sheffet, 2018, Dubey and Pentland, 2020], where we seek to design algorithms that are differentially private with respect to both the context and the reward sequence (See Section 2.2 for a precise definition). The primary challenge in designing differentially private learning algorithms is to balance the privacy-utility trade-off, i.e., to ensure meaningful learning while guaranteeing the privacy of the dataset. While there has been some effort towards designing differentially algorithm for multi-armed and linear bandits, the setting of kernelized bandits remains largely unexplored. Existing results on differentially private kernel bandits either apply only to a small class of kernel families or adopt a weaker notion of privacy (See Section 1.3 for additional discussion). In particular, there are no differentially private algorithms that achieve diminishing error rate under JDP for the commonly used kernel families, e.g., Matérn kernels.

1.2 Main Results

We propose the first algorithm for contextual kernel bandits that is jointly differentially private with respect to the contexts and the rewards and theoretically guarantees a diminishing simple regret for all kernels with polynomially decaying eigen-values. Kernels with polynomial eigen decay include the class of commonly used kernels like Matérn and Square exponential kernels. In particular, we establish a worst case error rate of $\mathcal{O}\left(\sqrt{\gamma_T/T} + \gamma_T/(T\varepsilon)\right)$, where γ_T is the information gain and represents the effective dimensionality of the kernel and ε is the privacy parameter. In the non-private setting i.e., $\varepsilon \to \infty$, this reduces to an error rate of $\mathcal{O}(\sqrt{\gamma_T/T})$, which is known to be order-optimal [Scarlett et al., 2017]. The best current simple regret upper bound In the private setting, derived from the bound for the cumulative regret, is $\mathcal{O}(\sqrt{\gamma_T/T} + \sqrt{\gamma_T/(T\varepsilon)})$. Notably this bound only applies to Square exponential kernels.

In sequential learning problems, the dataset with respect to which the algorithm needs to guarantee privacy continues to expand as new data points arrive over time. Consequently, this forces the algorithms to add a small additional layer of privacy for each query point. This injection of additional privacy induced noise at each time instant, in absence of a careful control, compounds over time and leads to poor utility. For the problem of private linear bandits Shariff and Sheffet [2018], Dubey and Pentland [2020], existing studies avoid this pitfall through the use of the tree-based mechanism Dwork and Roth [2014]. Tree-based mechanism allows for online release of prefix-sums, where the noise added to the sums does not scale linearly but rather logarithmically; thereby alleviating the noise compounding effect. For the problem of linear bandits, the key challenge is to privatize the covariance matrix for the ridge regression estimate. This is solved by noting that the covariance has an additive structure (sum of rank-one matrices) in the feature space, thereby allowing the use of tree-based mechanism. Taking this approach ensures that the privacy based error grows logarithmically, instead of linearly, resulting in meaningful utility bounds. However, in kernel based bandits, the features belong to an infinite-dimensional space which renders the use of this technique infeasible.

The work of Dubey [2021], which is the current state of the art, circumvents this issue by approximating the kernel with a low-dimensional surrogate, which allows them to reduce the problem to that of a finite-dimensional linear bandit. However, this approach is limited to Square Exponential kernels. It is not clear how such an approach can be applied to other kernels, e.g. the Matérn family of kernels.

The proposed technique in this work offers a departure from prevailing approaches to resolve the pitfalls of ensuring privacy. The common approach adopted by existing algorithms to address the privacy-utility trade-off is to privatize a high utility estimator. Namely, Shariff and Sheffet [2018], Dubey [2021] take a UCB-based estimator with order optimal non-private performance, and through careful addition of noise ensure privacy constraints while not significantly degrading the performance of the algorithm.

This utility first approach leads to a query strategy that is highly adaptive to the observed history in order to maximize the utility. However, high adaptivity results in high sensitivity to data points which makes preserving privacy much harder.

In this work, we adopt a different approach, where we put the privacy constraint at the forefront and then optimize utility. To this effect, there are two key components to our algorithm design that ensure the privacy constraint. The first component is the uniform sampling of the query points from the action set $\mathcal{X}$, independent of the context-reward observations. This *data-independent*, *non-adaptive* query strategy, decouples the query points from the context-reward

pairs and immediately guarantees privacy during the learning process. Our approach solves the problem of compounding noise during learning by simply ensuring the query points are private by design. There is thus no need to add any noise during the learning stage.

The second component is the design of a novel low -sensitivity estimator to be used for the final prediction. In current kernel bandit literature the posterior-mean estimator (see eq.(2)) is used almost exclusively. Although the posterior-mean offers order optimal approximation error of the reward function, it carries a strong dependence on the dataset through the reward and feature vector $k_{\mathbf{W}_T}(\cdot)$ as well as the Gramian matrix $\mathbf{K}_{\mathbf{W}_T,\mathbf{W}_T}$. As the prediction of the algorithm also needs to be private, it is necessary for the estimator to have both high utility and small sensitivity with respect to the dataset. This is seemingly an infeasible requirement as the small error requires adaptivity while the small sensitivity requires a weak dependence on the dataset.

By combining the recent advancements in non-private kernel bandits [Salgia et al., 2023] along with a novel technique for covariance estimation we design an estimator that only depends on the dataset through the feature and reward vector (please see eq.(4)) while retaining the approximation error of the posterior mean estimator. The essence of our proposed approach is to replace the covariance corresponding to a set of randomly sampled actions with that of an independently drawn set of actions. We use concentration results to establish that our new estimator offers the same order of approximation error as the posterior mean. At the same time, the independence between the set of samples and the dataset allows it to also enjoy low sensitivity.

1.3 Related Work

Kernel-based bandits. The problem of kernel-based bandit optimization has been extensively studied in the non-private setting. Starting with the seminal work of Srinivas et al. [2010], numerous algorithms for kernel-based have been proposed in both contextualValko et al. [2013] and non-contextual settings [Li and Scarlett, 2022, Salgia et al., 2021]. The optimal performance in the non-private setting is well-understood where several algorithms [Li and Scarlett, 2022, Valko et al., 2013, Salgia et al., 2021] are known to achieve the order-optimal performance that matches the lower bound [Scarlett et al., 2017].

Private Bandit optimization. The problem of differentially private bandit optimization has received considerable attention for both multi-armed and linear bandits. The problem of linear bandits with JDP was first introduced by Shariff and Sheffet [2018] where they propose an algorithm that achieves a cumulative regret of $\tilde{\mathcal{O}}(\sqrt{dT}/\varepsilon)$. Dubey and Pentland [2020] extend their results to the distributed setting where they achieve a cumulative regret of $\mathcal{O}(M^{3/4}\sqrt{T/\varepsilon})$[1]. Garcelon et al. [2022] consider the shuffle model of privacy as midway point between the central (JDP) privacy and (LDP) local privacy. The LDP notion of privacy is a more challenging setup of differential privacy , where the users do not trust the algorithm and all the data needs to be privatized before leaving the user. In contrast (JDP), requires the data to be privatized only before being used, allowing for processing in batches as in Shariff and Sheffet [2018]. Zheng et al. [2020] studies the problem of generalized linear bandits under an LDP constraint and propose an algorithm with a cumulative regret $\tilde{\mathcal{O}}(T^{3/4}/\varepsilon)$. Ruiquan et al. [2024] study differential privacy for distributed contextual linear bandits in both central and local setting. They provide lower bounds for both settings, with a matching upper bound in the case of central differential privacy. Hanna et al. [2024] study the LDP, CDP and shuffle model for linear non-contextual bandits. Only the rewards are considered to be sensitive data.
Privacy constraints have also been studied in the multi-arm-bandit framework(MAB). Azize and Basu [2022], Azize et al. [2024] explore globally private best arm-indentification while Tenenbaum et al. [2021] studied cumulative regret optimization under shuffle model of privacy.

The problem of private kernel bandits was first studied by Kusner et al. [2015] in the context of hyper-parameter tuning where the authors focus only on privatizing the final query point as opposed to all of them. Kharkovskii et al. [2020] study private kernel bandits for square exponential kernels under the setting where the algorithm and user are separate entities. The query points are required to be locally differentially private but not the rewards. Under the additional assumption of covariance matrix being diagonally dominant, they establish a simple regret of $\tilde{\mathcal{O}}((\varepsilon^{-2}+\gamma_T/T)^{1/2})$. Zhou and Tan [2021] consider kernel bandits with heavy tailed noise and only rewards are required

[1]This is rectified regret bound from Zhou and Chowdhury [2023]

to be private. Zhongxiang et al. [2021] study privacy for Thompson sampling in the problem of distributed kernel bandits where the reward functions are assumed to be heterogeneous over users.

The work that is closest to ours is by Dubey [2021] where the authors consider private contextual bandits. As mentioned earlier, they approximate the kernel using a low dimensional surrogate after which they use techniques from private linear bandits to design an algorithm for the kernel-based problem. However, the result in Dubey [2021] crucially depends on the assumption that underlying kernel can be approximated by a features whose dimension is at most polylogarithmic in T and has a separable Fourier transform. Such an assumption is only satisfied for the family of Squared Exponential kernels and it is not obvious how to extend this approach for more general kernels.

2 PROBLEM FORMULATION AND PRELIMINARIES

2.1 RKHS, Mercer's Theorem and GP Models

Consider a positive definite kernel $k : \mathcal{W} \times \mathcal{W} \to \mathbb{R}$, where $\mathcal{W}$ is a compact set in a given metric space. A Hilbert space $\mathcal{H}_k$ of functions on $\mathcal{W}$ equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$ is called a Reproducing Kernel Hilbert Space (RKHS) with reproducing kernel k if the following conditions are satisfied: (i) $\forall\ w \in \mathcal{W}$, $k(\cdot, w) \in \mathcal{H}_k$; (ii) $\forall\ w \in \mathcal{W}$, $\forall\ f \in \mathcal{H}_k$, $f(w) = \langle f, k(\cdot, w) \rangle_{\mathcal{H}_k}$. The inner product induces the RKHS norm, $\|f\|^2_{\mathcal{H}_k} = \langle f, f \rangle_{\mathcal{H}_k}$. We use $\phi(w)$ to denote $k(\cdot, w)$ and WLOG assume that $k(w, w) = \|\phi(w)\|^2_{\mathcal{H}_k} \leq 1$.

Let ζ be a finite Borel probability measure supported on $\mathcal{W}$ and let $L_2(\zeta, \mathcal{W})$ denote the Hilbert space of functions that are square-integrable w.r.t. ζ. Mercer's Theorem provides an alternative representation for RKHS through the eigenvalues and eigenfunctions of a kernel integral operator defined over $L_2(\zeta, \mathcal{W})$ using the kernel k.

Theorem 2.1. *Steinwart and Christmann [2008] Let $\mathcal{W}$ be a compact metric space and $k : \mathcal{W} \times \mathcal{W} \to \mathbb{R}$ be a continuous kernel. Furthermore, let ζ be a finite Borel probability measure supported on $\mathcal{W}$. Then, there exists an orthonormal system of functions $\{\psi_j\}_{j \in \mathbb{N}}$ in $L_2(\zeta, \mathcal{W})$ and a sequence of non-negative values $\{\lambda_j\}_{j \in \mathbb{N}}$ satisfying $\lambda_1 \geq \lambda_2 \cdots \geq 0$, such that $k(w, w') = \sum_{j \in \mathbb{N}} \lambda_j \psi_j(w) \psi_j(w')$ holds for all $w, w' \in \mathcal{W}$ and the convergence is absolute and uniform over $w, w' \in \mathcal{W}$.*

Consequently, the Mercer representation[Steinwart and Christmann, 2008, Thm. 4.51] of the RKHS of k is given as

$$\mathcal{H}_k = \left\{ f := \sum_{j \in \mathbb{N}} \alpha_j \lambda_j^{\frac{1}{2}} \psi_j : \|f\|^2_{\mathcal{H}_k} = \sum_{j \in \mathbb{N}} \alpha_j^2 < \infty \right\}.$$

A commonly used technique to characterize a class of kernels is through their eigendecay profile.

Definition 2.2. Let $\{\lambda_j\}_{j \in \mathbb{N}}$ denote the eigenvalues of a kernel k arranged in the descending order. The kernel k is said to satisfy the polynomial eigendecay condition with a parameter $\beta_p > 1$ if, for some universal constant $C_p > 0$, the relation $\lambda_j \leq C_p j^{-\beta_p}$ holds for all $j \in \mathbb{N}$.

We make the following assumption on the kernel k and the eigenfunctions $\{\psi_j\}_{j \in \mathbb{N}}$, which is commonly adopted in kernel-based optimization literature [Vakili et al., 2021b, Chatterji et al., 2019, Riutort-Mayol et al., 2023, Whitehouse et al., 2023].

Assumption 2.3. We assume that the kernel k satisfies the polynomial eigendecay condition with parameter $\beta_p > 1$. The eigen-functions $\{\psi_j\}_{j \in \mathbb{N}}$ corresponding to the kernel k are continous and hence bounded on $\mathcal{W}$ i.e $\exists\ F > 0$, such that $\sup_{w \in \mathcal{W}} |\psi_j(w)| \leq F$ for all $j \in \mathbb{N}$.

A Gaussian Process (GP) is a random process G indexed by $\mathcal{W}$ and is associated with a mean function $\mu : \mathcal{W} \to \mathbb{R}$ and a positive definite kernel $k : \mathcal{W} \times \mathcal{W} \to \mathbb{R}$. The random process G is defined such that for all finite subsets of $\mathcal{W}$, $\{w_1, w_2, \ldots, w_T\} \subset \mathcal{W}$, $T \in \mathbb{N}$, the random vector $[G(w_1), G(w_2), \ldots, G(w_T)]^\top$ follows a multivariate Gaussian distribution with mean vector $[\mu(w_1), \ldots, \mu(w_T)]^\top$ and covariance matrix $[k(w_i, w_j)]_{i,j=1}^T$. Throughout the work, we consider GPs with $\mu \equiv 0$. When used as a prior for a data generating process under Gaussian noise, the conjugate property provides closed form expressions for the posterior mean and covariance of the GP model. Specifically, given a set of observations $\{\mathbf{W}_T, \mathbf{Y}_T\} = \{(w_i, y_i)\}_{i=1}^T$ from the underlying process, the expression for posterior mean and variance of GP model is given as follows:

$$\mu_T(w) = k_{\mathbf{Z}_T}(w)^\top (\tau \mathbf{I}_T + \mathbf{K}_{\mathbf{W}_T, \mathbf{W}_T})^{-1} \mathbf{Y}_T, \tag{2}$$

$$\sigma_T^2(w) = k(w, w) - \\ - k_{\mathbf{W}_T}^\top(w)(\tau \mathbf{I}_T + \mathbf{K}_{\mathbf{W}_T, \mathbf{W}_T})^{-1} k_{\mathbf{W}_T}(w). \tag{3}$$

In the above expressions, $k_{\mathbf{W}_T}(w) = [k(w_1, w), k(w_2, w) \ldots k(w_T, w)]^\top$, $\mathbf{K}_{\mathbf{W}_T, \mathbf{W}_T} =$

$[k(w_i, w_j)]_{i,j=1}^T$, $\mathbf{I}_T$ is the $T \times T$ identity matrix and τ is the variance of the Gaussian noise.

Following a standard approach in the literature [Srinivas et al., 2010], we model the data corresponding to observations from the unknown f, which belongs to the RKHS of a positive definite kernel k, using a GP with the same covariance kernel k. In particular, we assume a *fictitious* GP prior over the fixed, unknown function f along with *fictitious* Gaussian distribution for the noise. Such a modelling allows us to predict the values of f and characterize the prediction error through the posterior mean and variance of the GP model.

Lastly, given a set of points $\mathbf{W}_T = \{w_1, w_2, \ldots, w_T\} \in \mathcal{W}$, the information gain of the set $\mathbf{W}_T$ is defined as $\gamma_{\mathbf{W}_T} := \frac{1}{2}\log(\det(\mathbf{I}_T + \tau^{-1}\mathbf{K}_{\mathbf{W}_T,\mathbf{W}_T}))$. Using this, we can define the maximal information gain of a kernel as $\gamma_T := \sup_{\mathbf{W}_T \in \mathcal{W}^T} \gamma_{\mathbf{W}_T}$. Maximal information gain is closely related to the effective dimension of a kernel [Calandriello et al., 2019] and helps characterize the regret performance of kernel bandit algorithms [Srinivas et al., 2010, Chowdhury and Gopalan, 2017]. γ_T depends only the kernel and τ and has been shown to be an increasing sublinear function of T [Srinivas et al., 2010, Vakili et al., 2021b].

2.2 Joint Differential Privacy

We adopt the framework of Joint Differential Privacy presented in Shariff and Sheffet [2018]. Let $\mathcal{S}_T = \{(c_1, y_1), (c_2, y_2), \ldots (c_T, y_T), c_{T+1}\}$, referred to as a database, denote the collection of all contexts and rewards seen in the duration of the algorithm. Here, c_{T+1} denotes the contexts drawn at the evaluation instant $T+1$.

Definition 2.4. Two databases $\mathcal{S}_T, \mathcal{S}'_T$ are said to be t-neighbours if they only differ in the context and reward at time t . Specifically, $\mathcal{S}_T = \{(c_1, y_1), (c_2, y_2), \ldots (c_t, y_t), \ldots (c_T, y_T), c_{T+1}\}$ and $\mathcal{S}'_T = \{(c_1, y_1), (c_2, y_2), \ldots (c'_t, y'_t), \ldots (c_T, y_T), c_{T+1}\}$ are considered to be t-neighbours.

JDP ensures that a malicious adversary cannot confidently differentiate between the agent database and any of its neighbours. This constraint can be mathematically presented as:

Definition 2.5. A randomized algorithm $\mathscr{A}$ is (ε, δ)-joint differentially private (JDP) under continual observation if for all $t \leq T$ and all pairs of t-neighboring databases $\mathcal{S}$ and $\mathcal{S}'$ and any subset $\mathcal{P}_{>t} \subset \mathcal{X}^{T-t+1}$ of sequence of points ranging from day $t+1$ to $T+1$ [2], it holds that:

$$\Pr\left(\mathscr{A}(\mathcal{S}) \in \mathcal{P}_{>t}\right) \leq e^{\varepsilon} \cdot \Pr\left(\mathscr{A}(\mathcal{S}') \in \mathcal{P}_{>t}\right) + \delta,$$

where the probability is taken over the random coins generated by the algorithm.

Note that JDP does not require the query point x_t to be private with respect to the current context and reward (c_t, y_t), but only with respect to previously seen contexts and rewards $\mathcal{S}_{<t}$. As shown in [Shariff and Sheffet, 2018, Claim 13] requiring privacy with respect to current context and reward would lead to provably $\mathcal{O}(1)$ simple regret performance.

2.3 Problem statement

We consider the problem of contextual kernel bandits, where at each time instant t, the learning agent is presented with a context $c_t \in \mathcal{C}$ based on which it queries a point $x_t \in \mathcal{X}$ and receives a noisy reward $y_t = f(x_t, c_t) + \eta_t$, where η_t denotes the noise. We assume that the sets $\mathcal{X} \in \mathbb{R}^d$ and $\mathcal{C} \in \mathbb{R}^{d'}$ are compact and convex. The reward function f belongs to a RKHS corresponding to a kernel k defined over $\mathcal{W} := \mathcal{X} \times \mathcal{C}$. We consider the setting where the contexts are drawn i.i.d. across time according to some context distribution κ. The contextual bandits with stochastic contexts has widely studied in the literature [Ruiquan et al., 2024, Han et al., 2021, Amani et al., 2023, Hanna et al., 2022, 2023]. We make the following assumptions that are commonly adopted in the literature.

Assumption 2.6. The noise term η_t is assumed to be i.i.d across all time instances and is a zero-mean, R sub-Gaussian random variable i.e., it satisfies the relation $\mathbb{E}[\exp(q\eta)] \leq \exp(q^2R^2/2)$ for all $q \in \mathbb{R}$.

Assumption 2.7. The rewards $\{y_t\}_{t=1}^T$ in the duration of the algorithm are bounded in absolute values, $|y_t| < B, \forall t \leq T$

Assumption(2.7) is adopted across privacy literature [Dubey and Pentland, 2020, Shariff and Sheffet, 2018, Han et al., 2021, Zheng et al., 2020] , and ensures that an adversary cannot probe an unbounded reward as an input to the algorithm. We note that that this assumption could be removed by simple clipping the rewards that have modulus higher than $B + R\log(T/\delta)$. By sub-gaussian assumption on the rewards all rewards would remain unchanged, with probability $1 - \delta$.

[2] counting the final $\widehat{x}_T(c_{T+1})$ for which there is no feedback

Assumption 2.8. We assume the reward function f is L_f-Lipschitz, i.e., the following relation holds for all $w, w' \in \mathcal{W} = \mathcal{X} \times \mathcal{C}$

$$|f(w) - f(w')| \leq L_f \|w - w'\|_2$$

Assumption 2.9. For each $r \in \mathbb{N}$, there exists a discretization $\mathcal{U}_r$ of $\mathcal{W}$ with $|\mathcal{U}_r| = \text{poly}(r)$[3] such that, for any $f \in \mathcal{H}_k$, we have $|f(w) - f([w]_{\mathcal{U}_r})| \leq \frac{\|f\|_{\mathcal{H}_k}}{r}$, where $[w]_{\mathcal{U}_r} = \arg\min_{w' \in \mathcal{U}_r} \|w - w'\|_2$.

Assumption 2.10. We have a context generator that is able to generate contexts i.i.d according to the distribution κ.

Assumptions 2.6 and 2.8 are mild assumptions that are commonly adopted in the literature [Srinivas et al., 2010, Chowdhury and Gopalan, 2017, Li and Scarlett, 2022, Salgia et al., 2021, Lee et al., 2022]. For commonly used kernels kernels like Squared Exponential and Matérn kernels elements of its RKHS are known to be Lipschitz continuous Lee et al. [2022]. Assumption 2.9 is nearly universally used to apply the confidence bounds on the continuous domain [Vakili et al., 2022, 2021a, Li and Scarlett, 2022, Salgia et al., 2023]

Assumption 2.10 is milder than the assumption of having complete knowledge of context distribution, an assumption that has been commonly adopted in several existing studies on stochastic contextual bandits[Amani et al., 2023, Hanna et al., 2022, 2023]

In our analogy to the shopping platform in the introduction, these contexts are not recorded from the arrivals of any "real" users but is rather an assumption that the platform can learn its user demographic(location, common searches etc.). In other words these contexts are not the input to the algorithm but are rather generated by the algorithm's random coin. We also emphasize these "artificially generated" contexts are not a result of interaction with the environment. Consequently, no feedback is associated with the generated contexts and thus they carry no information on the reward function and cannot offer any trivial advantages in the learning.

[3]The notation $g(x) = \text{poly}(x)$ is equivalent to $g(x) = \mathcal{O}(x^k)$ for some $k \in \mathbb{N}$.

3 ALGORITHM DESCRIPTION

Algorithm 1 USCA

Input: error probability δ, privacy budget ε,
Initialize $\mathbf{W}_T, \mathbf{Y}_T, \mathcal{Z} \leftarrow \emptyset$
`// Learning stage`
for $t = 1, 2, \ldots, T$ **do**
  Receive the context vector c_t
  Query x_t from $\mathcal{X}$ uniformly at random and observe the reward y_t
  $\mathbf{W}_T \leftarrow \mathbf{W}_T \cup \{(x_t, c_t)\}, \mathbf{Y}_T \leftarrow \mathbf{Y}_T \cup \{y_t\}$
end for
for $k = 1, 2, \ldots, TK$ **do**
  Sample c_k from the context measure κ
  Sample x_k from $\mathcal{X}$ uniformly at random
  $\mathcal{Z} \leftarrow \mathcal{Z} \cup \{(x_k, c_k)\}$
end for
Construct the posterior mean $\overline{\mu}_T, \overline{\sigma}$ with an approximating set $\mathcal{Z}$ and dataset $\mathbf{W}_T$ from eq.(4, 5)
`// Prediction stage`
Observe the context set c_{T+1}
Sample

$$\widehat{x}_T(c_{T+1}) \sim \mathcal{E}(\overline{\mu}_T(\cdot, c_{T+1}), \varepsilon, 2B \sup \overline{\sigma}^2)$$

where $\mathcal{E}$ is as defined in Def.(3.2)
Output $\widehat{x}_T(c_{T+1})$

In this section we present our algorithm, USCA. The main features of the algorithm are **U**niform **S**ampling with **C**ovariance **A**pproximation. We seperate USCA into two stages, reflecting the nature of the agents interaction with the environment and not a design philosphy. In the learning stage agent collects the information on the reward function through the collected feedback, $\mathbf{Y}_T$. In the prediction stage, the agent uses the collected data to project the best performing point for a given context c_{T+1}.

During learning stage, our algorithm adopts a data-independent, random sampling approach. After being presented with a context vector c_t agent draws a query point x_t from $\mathcal{X}$ uniformly at random, independently from previous contexts and rewards. This, data invariant, sampling method ensures that there is no loss of privacy during learning and thus no noise injection is necessary at this stage.

A widely accepted approach in kernel bandit learning is to sample points from the domain based on the posterior statistics in eqs.(2, 3). USCA abandons this approach and utilizes novel estimator $\overline{\mu}_T$ and

surrogate variance $\overline{\sigma}^2$. Estimator $\overline{\mu}_T$ is obtained by approximating the covariance matrix of the sampled points (lines 4-7) by an empirical average. The empirical average is calculated based on the points in the approximating set $\mathcal{Z}$. In the context of our algorithm $\overline{\sigma}^2$, is seen as the posterior variance of a GP process after sampling the approximating set $\mathcal{Z}$(see eq.3) . As such, $\overline{\sigma}$ is independent of any contexts or reward seen by the agent, and only depends on the points of the approximating set $\mathcal{Z}$.

To construct $\overline{\sigma}, \overline{\mu}_T$ we need a sample of TK context vectors drawn from the context distribution κ where $K = \lceil T/\gamma_T \rceil$. We emphasize that these contexts are sampled independently from the context space $\mathcal{C}$ and are not an input to the algorithm but are rather generated by the algorithm's random coin. Following the generation of the context vector c_k the agent samples x_k from the domain $\mathcal{X}$ uniformly at random and the sample $(x_k, c_k) \in \mathcal{W}$ is added to the approximating set $\mathcal{Z}$ that we use to calculate $\overline{\mu}_T, \overline{\sigma}^2$.

Next, we calculate the statistics $\overline{\sigma}, \overline{\mu}_T : \mathcal{W} \to \mathbb{R}$ from the sampled approximating set(lines 11-15) $\mathcal{Z}$ and the data set $(\mathbf{W}_T, \mathbf{Y}_T)$ accrued during learning (lines 4-7). We introduce the parametric form of the estimator $\overline{\mu}_T$ in Lemma A.2, proving it sufficiently approximates the posterior mean μ_T(see eq.(2)). However in this form $\overline{\mu}_T$ is computationally intractable . We apply the celebrated kernel trick [Steinwart and Christmann, 2008] to obtain a computationally more favorable form:

Lemma 3.1. *The statistics $\overline{\mu}_T, \overline{\sigma}$ in USCA can be calculated as:*

$$\overline{\mu}_T(w) = \frac{1}{\tau} k_{\mathbf{W}_T}(w)^\top \mathbf{Y}_T - \frac{1}{\tau} k_{\mathcal{Z}}(w)^\top \left(\mathbf{K}_{\mathcal{Z},\mathcal{Z}} + K\tau \mathbf{I}_{\mathcal{Z}}\right)^{-1} \mathbf{K}_{\mathcal{Z},\mathbf{W}_T} \mathbf{Y}_T \tag{4}$$

$$\overline{\sigma}^2(w) = \frac{1}{\tau} k(w,w) - \frac{1}{\tau} k_{\mathcal{Z}}(w)^\top \left(\mathbf{K}_{\mathcal{Z},\mathcal{Z}} + K\tau \mathbf{I}_{\mathcal{Z}}\right)^{-1} k_{\mathcal{Z}}(w) \tag{5}$$

Where $\mathbf{K}_{\mathcal{Z},\mathbf{W}_T} = \{k(a,b)\}_{a\in\mathcal{Z}, b\in\mathbf{W}_T}, \mathbf{K}_{\mathcal{Z},\mathcal{Z}} = \{k(a,b)\}_{(a,b)\in\mathcal{Z}^2}$ *and* $K = \lceil T/\gamma_T \rceil$.

Proof. For a proof please see Lemma D.1. □

After T-time instances the agent is provided with a final context vector c_{T+1} and has to output the point $\widehat{x}_T(c_{T+1})$ that should maximize the reward for c_{T+1}. To privatize the final output agent samples $\widehat{x}_T(c_{T+1})$ according to exponential measure, with the exponent proportional to $\overline{\mu_T}$. More specifically:

Definition 3.2. Define the measure $\mathcal{E}(\overline{\mu}_T, \varepsilon, m)$ on $\mathcal{X}$ as :

$$\mathcal{E}(\overline{\mu}_T(r), \varepsilon, m) = \frac{\exp(\overline{\mu}_T(r)\varepsilon/(2m))}{\int_{\mathcal{X}} \exp(\overline{\mu}_T(r)\varepsilon/(2m))\nu_0(dr)}$$

Where ν_0 is the Lebesgue measure over $\mathcal{X}$.

In privacy literature the method of sampling from $\mathcal{E}$ is known as the exponential mechanism [McSherry and Talwar, 2007, Dwork and Roth, 2014]. Using the closed form expression for the estimator $\overline{\mu}_T$ and surrogate variance $\overline{\sigma}$ in eqs.(4,5) agent samples $\widehat{x}_T(c_{T+1})$ from $\mathcal{X}$ according to $\mathcal{E}\left(\overline{\mu}_T(\cdot, c_{T+1}), \varepsilon, 2B \sup \overline{\sigma}^2\right)$.

4 PERFORMANCE ANALYSIS

The following theorem characterizes the performance of our proposed algorithm, USCA.

Theorem 4.1. *Consider the contextual kernelized bandits problem described in Sec. 2.3 where the underlying kernel function satisfies Assumption 2.2 with parameter $\beta_p > 1$ and the reward function f is L_f-Lipschitz. If the USCA algorithm is run for T steps with a privacy parameter ε, then for all $\varepsilon > 0$, $\delta \in (0,1)$ and $T > T_0$,*

- *USCA is ε-JDP;*
- *The error rate of USCA satisfies the following relation with probability $1-\delta$*

$$ER(USCA) = \mathcal{O}\left(\sqrt{\frac{\gamma_T}{T}} + \frac{1}{\varepsilon}\frac{\gamma_T}{T}\right).$$

Here T_0 is the constant that depends on the kernel and the context distribution[4] and probability in the error bound is taken over all the contexts, rewards and random coins of the algorithm.

As shown by the above theorem, USCA retains ε-JDP privacy while achieving diminising regret rate of $\mathcal{O}(\sqrt{\gamma_T/T} + \gamma_T/(T\varepsilon))$. We emphasize that although we state the final result in terms of expectation over the context vector c_{T+1}, as proven in the Theorem C.2, the claim holds uniformly over the entire context set $\mathcal{C}$.

The key ingredient that allows USCA to achieve the performance guarantees outlined in Theorem C.2 is the use of a novel reward estimator $\overline{\mu}_T$. The classical

[4]Please refer to Theorem C.2 for an exact expression.

posterior mean estimate μ_T offers powerful predictive performance. However, characterizing the sensitivity of the estimator μ_T is challenging due to the non-linear relationship between the contexts, rewards and the predicted value. The problem is exacerbated by the fact that the sensitivity is defined in an adversarial sense, i.e., the differing context can be any value from the context set and are not necessarily drawn from the context distribution. This results in trivial bounds on the sensitivity of μ_T which prevents us from obtaining any meaningful utility guarantees. Our estimator $\overline{\mu}_T$ alleviates this issue by having a far weaker dependency on the the dataset $\mathcal{S}_T$. Specifically, $\overline{\mu}_T$ is constructed by choosing a feature covariance matrix that is *independent* of the context and reward sequence, which helps us significantly decrease the impact of a single point on the output and hence obtain meaningful sensitivity bounds. This idea is formalized in the following lemma:

Lemma 4.2. *Let $\mathcal{S}_T, \mathcal{S}_T^{'}$ be two t-neighbouring databases, and let $\overline{\mu}_T, \overline{\mu}_T'$ be the 2 estimator constructed for each of the databases . For $T > \overline{N}_1(\delta)$, we can bound the sensitivity of an estimator $\overline{\mu}_T$ in USCA as :*

$$\Delta\overline{\mu}_T = \sup_{w\in\mathcal{W}} \sup_{\mathcal{S}_T,\mathcal{S}_T' \text{are t neigbours}} |\overline{\mu}_T(w) - \overline{\mu}_T'(w)| \leq$$
$$\leq 2B \sup_{w\in\mathcal{W}} \overline{\sigma}^2(w)$$

What makes $\overline{\mu}_T$ a particularly powerful tool is the fact that in addition to guaranteeing low-sensitivity it also offers high predictive performance, as shown in the following lemma.

Lemma 4.3. *For the estimator $\overline{\mu}_T$ introduced in eq. (4), under the condition $T > \max(\overline{N}(\delta/4), \overline{N}_1(\delta/4))$ we claim with probability at least $1-\delta$:*

$$\sup_{w\in\mathcal{W}} |\overline{\mu}_T(w) - f(w)| = \mathcal{O}\left(\sqrt{\frac{\gamma_T}{T}}\right)$$

Where $\overline{N}_1(\delta), \overline{N}(\delta)$ are a δ-dependent constants.

In particular, Lemma 4.3 states that $\overline{\mu}_T$ retains the order of approximation error of the posterior mean obtained in Vakili et al. [2021a]. This implies that predictive performance of $\overline{\mu}_T$ is the same as that of μ_T while offering reduced sensitivity. Note that even though $\overline{\mu}_T$ is constructed using a feature covariance matrix that is *independent* of the context and reward sequence, the covariance matrix in $\overline{\mu}_T$ and μ_T are identically distributed. The independence allows us to obtain strong sensitivity bounds while the concentration properties lead to similar predictive performance. We formalize this idea in the following novel spectral bound for empirical covariance matrix which may be of independent interest:

Lemma 4.4. *Suppose TK points $\{z_1, z_2, \dots z_{TK}\}$ are sampled i.i.d from $\mathcal{W}$ according to a Borel measure ϱ. Let $\mathbf{Z} = T\mathbf{\Lambda} + \tau\mathbf{Id}$ where $\mathbf{\Lambda} = \mathbb{E}_{w\sim\varrho}[\phi(w)\phi(w)^\top]$ and define the operator:*

$$\widetilde{\mathbf{Z}} = \frac{1}{K}\sum_{i=1}^{TK} \phi(z_i)\phi(z_i)^\top + \tau\mathbf{Id}$$

By choosing parameter $K = \lceil T/\gamma_T \rceil$ under the condition that $T > \overline{N}_1(\delta) = (\frac{\log(1/\delta)}{16C_pF^2})^{2\beta_p/(\beta_p-1)}$ we claim with probability at least $1-\delta$:

$$\|\widetilde{\mathbf{Z}}^{-1}\mathbf{Z} - \mathbf{Id}\|_2 \leq \frac{28}{17}\cdot\sqrt{\frac{\gamma_T}{T}\frac{\log(1/\delta)}{\tau}}$$

Next we briefly explain how the presented lemmas are utilized in the proof of Theorem 4.1.
With the sensitivity bounds obtained in Lemma 4.2, we use a similar approach to [McSherry and Talwar, 2007, Lemma 7] to ensure that the final prediction $f(\widehat{x}_T(c_{T+1}))$ is close to optimal value $\sup_{x\in\mathcal{X}} f(x, c_{T+1})$. In order to ensure the utility of continuous exponential mechanism the usual hurdle to overcome is lowed-bounding the volume of well performing points [McSherry and Talwar, 2007, Dwork and Roth, 2014]. To this end, we use a geometric Lemma C.1 that establishes a bound on volume, polynomial in the distance to the optimum.
Our approach to ensuring privacy follows the approach given in McSherry and Talwar [2007], with the necessary modification for random functions. For a full proof please see Theorem B.3.

5 NUMERICAL EXPERIMENTS

In this section we demonstrate the empirical performance of USCA. For the benchmark we adopt the widely used Hartmann-4D function [Salgia et al., 2023, Picheny et al., 2013] and choose k to be the Square exponential kernel. Contexts are taken to be uniformly distributed over $\mathcal{C} = [0,1]^2$ and query points are chosen from the $\mathcal{X} = [0,1]^2$. We run USCA for $T = 50$ time instances and plot the expected regret for various privacy budgets $\varepsilon \in \{0.1, 0.2, 2, 5, 10, 30, 50\}$.

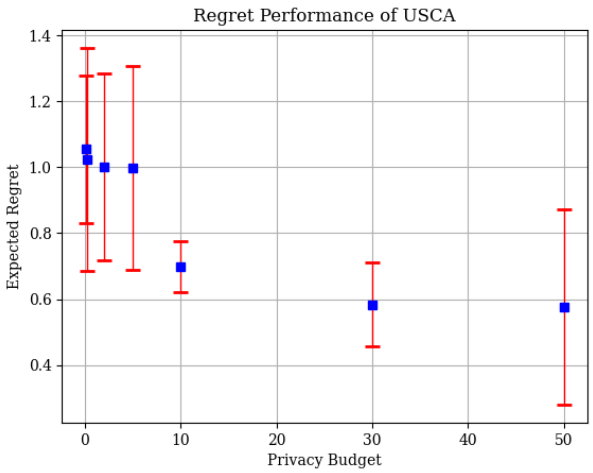

Figure 1: Regret peformance of USCA

The expected regret with respect to the final context c_{T+1} is estimated as a 5 sample average, and the entire algorithm run is repeated 5 times.
In the Figure 1. we plot the expected regret for varying privacy budget with a confidence interval of 2 standard deviations. As seen from the plot, the regret decreases with ε and and as $\varepsilon \to \infty$ USCA obtains regret performance in line with the non-private algotihms on the same benchmark [Salgia et al., 2023].

6 CONCLUSION

We propose the first algorithm for contextual kernel bandits that is continually differentially private with respect to the contexts and the rewards and theoretically guarantees a diminishing simple regret for all commonly used kernels. In particular, we improve on the state of the art while greatly expanding the set of admissible kernel families.
Key aspects of our approach are random sampling during learning and a novel high-utility, low-sensitivity estimator. As a theoretical contribution, we propose a novel concentration result for the covariance matrices of RKHS elements, that could be of independent interest.

Although our work provides state of the art performance in the simple regret setting , the ideas developed here are not readily applicable to the cumulative regret setting. The design of regret efficient, private algorithms in the cumulative setting for kernels beyond the Square Exponential kernel remains an interesting direction for future work.

Acknowledgments

The work of N. Pavlovic and Q. Zhao was supported by the National Science Foundation under Grant CCF-2419622.

References

S. Amani, L. Tor, G. András, and L. Yang. Distributed contextual linear bandits with minimax optimal communication cost. *In International Conference on Machine Learning,pp. 691-717. PMLR, 2023*, 2023.

A. Azize and D. Basu. When privacy meets partial information: A refined analysis of differentially private bandits. *Advances in Neural Information Processing Systems 35*, 2022.

A. Azize, J. Marc, A. A. Marjani, and D. Basu. On the complexity of differentially private best-arm identification with fixed confidence. *Advances in Neural Information Processing Systems 36*, 2024.

D. Calandriello, L. Carratino, A. Lazaric, M. Valko, and L. Rosasco. Gaussian Process Optimization with Adaptive Sketching: Scalable and No Regret. *Proceedings of Machine Learning Research*, 99:1–25, 2019.

N. Chatterji, A. Pacchiano, and P. Bartlett. Online learning with kernel losses. *In International Conference on Machine Learning, pp. 971-980. PMLR*, 2019.

S. R. Chowdhury and A. Gopalan. On kernelized multi-armed bandits. *In International Conference on Machine Learning, pp. 844-853. PMLR,*, 2017.

A. Dubey. No-regret algorithms for private gaussian process bandit optimization. *In International Conference on Artificial Intelligence and Statistics, pp. 2062-2070. PMLR*, 2021.

A. Dubey and A. Pentland. Differentially-private federated linear bandits. *Advances in Neural Information Processing Systems 33*, 2020.

C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 2014.

E. Garcelon, C. Kamalika, P. Vianney, and P. Matteo. Privacy amplification via shuffling for linear contextual bandits. *In International Conference on Algorithmic Learning Theory, pp. 381-407. PMLR,*, 2022.

Y. Han, L. Zhipeng, W. Yang, and J. Zhang. Generalized linear bandits with local differential privacy.

Advances in Neural Information Processing Systems, 2021.

O. Hanna, A. M. Girgis, F. Christina, and S. Diggavi. Differentially private stochastic linear bandits:(almost) for free. *IEEE Journal on Selected Areas in Information Theory*, 2024.

O. A. Hanna, Y. Lin, and C. Fragouli. Learning from distributed users in contextual linear bandits without sharing the context. *Advances in Neural Information Processing Systems 35*, 2022.

O. A. Hanna, L. Yang, and C. Fragouli. Contexts can be cheap: Solving stochastic contextual bandits with linear bandit algorithms. *Thirty Sixth Annual Conference on Learning Theory pp. 1791-1821. PMLR*, 2023.

D. Kharkovskii, D. Zhongxiang, and B. K. H. Low. Private outsourced bayesian optimization. *In International Conference on Machine Learning, pp. 5231-5242*, 2020.

M. Kusner, G. Jacob, G. Roman, and W. Kilian. Differentially private bayesian optimization. *In International conference on machine learning, pp. 918-927. PMLR*, 2015.

M. Lee, S. Shubhanshu, and T. Javidi. Multi-scale zero-order optimization of smooth functions in an rkhs. *2022 IEEE International Symposium on Information Theory (ISIT), pp. 288-293. IEEE*, 2022.

Z. Li and J. Scarlett. Gaussian process bandit optimization with few batches. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, AISTATS*, 2022.

F. McSherry and K. Talwar. Mechanism design via differential privacy. *IEEE Symposium on Foundations of Computer Science (FOCS'07)*, 2007.

V. Picheny, T. Wagner, and D. Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626, 2013. ISSN 1615147X.

G. Riutort-Mayol, B. Paul-Christian, A. MichaelR., S. Arno, and A. Vehtari. Practical hilbert space approximate bayesian gaussian processes for probabilistic programming. *Statistics and Computing 33, no. 1 (2023): 17*, 2023.

H. Ruiquan, H. Zhang, L. Melis, M. Shen, M. Hejazinia, and J. Yang. Federated linear contextual bandits with user-level differential privacy. *In International Conference on Machine Learning, pp. 14060-14095. PMLR*, 2024.

S. Salgia, S. Vakili, and Q. Zhao. A domain-shrinking based Bayesian optimization algorithm with order-optimal regret performance. In *Proceedings of the 35th Annual Conference on Neural Information Processing Systems*, volume 34, 2021.

S. Salgia, S. Vakili, and Q. Zhao. Random exploration in bayesian optimization: Order-optimal regret and computational efficiency, 2023.

J. Scarlett, I. Bogunovic, and V. Cevher. Lower bounds on regret for noisy gaussian process bandit optimization. *In Conference on Learning Theory, pp. 1723-1742. PMLR,*, 2017.

R. Shariff and O. Sheffet. Differentially private contextual linear bandits. *Advances in Neural Information Processing Systems 31*, 2018.

N. Srinivas, A. Krause, S. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning, ICML*, pages 1015–1022, 2010.

I. Steinwart and A. Christmann. *Support Vector Machines*. Springer, 2008.

J. Tenenbaum, K. Haim, M. Yishay, and S. Uri. Differentially private multi-armed bandits in the shuffle model. *Advances in Neural Information Processing Systems 34*, 2021.

S. Vakili, N. Bouziani, S. Jalali, A. Bernacchia, and D.-s. Shiu. Optimal order simple regret for Gaussian process bandits. *Proceedings of the 35th Annual Conference on Neural Information Processing Systems*, 2021a.

S. Vakili, K. Khezeli, and V. Picheny. On information gain and regret bounds in Gaussian process bandits. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics, AISTATS*, 2021b.

S. Vakili, J. Scarlett, D.-s. Shiu, and A. Bernacchia. Improved convergence rates for sparse approximation methods in kernel-based learning. In *Proceedings of the 39th International Conference on Machine Learning, ICML*, pages 21960–21983. PMLR, 2022.

M. Valko, N. Korda, R. Munos, I. Flaounas, and N. Cristianini. Finite-time analysis of kernelised contextual bandits. *arXiv preprint arXiv:1309.6869*, 2013.

J. Whitehouse, R. Aaditya, and S. Z. Wu. On the sublinear regret of gp-ucb. *Advances in Neural Information Processing Systems 36 (2023): 35266-35276*, 2023.

K. Zheng, C. Tianle, H. Weiran, L. Zhenguo, and W. Liwei. Locally differentially private (contextual) bandits learning. *Advances in Neural Information Processing Systems 33*, 2020.

D. Zhongxiang, B. K. H. Low, and P. Jaillet. Differentially private federated bayesian optimization with distributed exploration. *Advances in Neural Information Processing Systems 34*, 2021.

X. Zhou and S. R. Chowdhury. On differentially private federated linear contextual bandits. *arXiv preprint arXiv:2302.13945*, 2023.

X. Zhou and J. Tan. Local differential privacy for bayesian optimization. *In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 12, pp. 11152-11159*, 2021.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model.
   Yes. A clear description is provided throughout Sections 2,3,4.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm.
   Yes, all properties of our algorithm are proved in Section 4.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries.
   Not Applicable.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results.
   Yes, all used assumptions are provided in the the Section 2.
   (b) Complete proofs of all theoretical results.
   Yes, all theoretical claims are proven in the appendices.
   (c) Clear explanations of any assumptions.
   Yes, all assumptions are clearly explained in Section 2.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL).
   Yes
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen).
   Yes
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times).
   Yes
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider).
   Not Applicable

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets.
   Not Applicable
   (b) The license information of the assets, if applicable.
   Not Applicable
   (c) New assets either in the supplemental material or as a URL, if applicable.
   Not Applicable
   (d) Information about consent from data providers/curators.
   Not Applicable
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content
   . Not Applicable

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots.
   Not Applicable
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable.
   Not Applicable
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation.
   Not Applicable

A Spectral bounds

Lemma A.1. *Suppose TK points $\{z_1, z_2, \dots z_{TK}\}$ are sampled i.i.d from $\mathcal{W}$ according to a Borel measure ϱ. Let $\mathbf{Z} = T\mathbf{\Lambda} + \tau\mathbf{Id}$ where $\mathbf{\Lambda} = \mathbb{E}_{w\sim\varrho}[\phi(w)\phi(w)^\top]$ and :*

$$\widetilde{\mathbf{Z}} = \frac{1}{K}\sum_{i=1}^{TK}\phi(z_i)\phi(z_i)^\top + \tau\mathbf{Id}$$

By choosing parameter $K = \lceil T/\gamma_T \rceil$ under the condition that $T > \overline{N}_1(\delta) = \left(\frac{\log(1/\delta)}{C_p F^2 16}\right)^{2\beta_p/(\beta_p-1)}$ we claim with probability at least $1-\delta$:

$$\|\widetilde{\mathbf{Z}}^{-1}\mathbf{Z} - \mathbf{Id}\|_2 \leq \frac{28}{17}\cdot\sqrt{\frac{\gamma_T}{T}\frac{\log(1/\delta)}{\tau}}$$

Proof. Note that we will first bound the spectral norm of $\|\mathbf{Z}^{-1}\widetilde{\mathbf{Z}} - \mathbf{Id}\|$. Consider a $g \in \mathcal{H}_k$ with $\|g\| = 1$. We have:

$$g^\top(\mathbf{Z}^{-1}\widetilde{\mathbf{Z}} - \mathbf{Id})g = \sum_{j=1}^{KT}\frac{1}{K}g^\top\mathbf{Z}^{-1}\phi(x_i)\phi(x_i)^\top g - g^\top\mathbf{Z}^{-1}(T\mathbf{\Lambda})g.$$

Define $V_i := \frac{1}{K}g^\top\mathbf{Z}^{-1}\phi(z_i)\phi(z_i)^\top g$. We can now write:

$$\begin{aligned}
\mathbb{E}[V_i] &= \frac{1}{K}g^\top\mathbf{Z}^{-1}\mathbf{\Lambda}g \\
|V_i| &= \sup_{z,z_i\in\mathcal{W}}\sup_{g\in\mathcal{H}_k}\frac{1}{K}g^\top\mathbf{Z}^{-1}\phi(z_i)\phi(z_i)^\top g \leq \sup_{z,z_i\in\mathcal{W}}\sup_{g\in\mathcal{H}_k}\frac{g(z)}{K}\sqrt{g^\top\mathbf{Z}^{-1}g}\sqrt{\phi(z_i)^\top\mathbf{Z}^{-1}\phi(z_i)} \leq \\
&\leq \frac{1}{K}\max\{\sup_g g^\top\mathbf{Z}^{-1}g, \sup_z \phi^\top(z)\mathbf{Z}^{-1}\phi(z)\} = C_1 \\
\mathbb{E}[V_i^2] &\leq \frac{1}{K^2}\mathbb{E}\left[\left(g^\top\mathbf{Z}^{-1}\phi(z_i)\phi(z_i)^\top g\right)^2\right] = \frac{1}{K^2}\mathbb{E}\left[g(z)^2 g^\top\mathbf{Z}^{-1}\phi(z_i)\phi(z_i)^\top\mathbf{Z}^{-1}g\right] \leq \frac{1}{K^2}g^\top\mathbf{Z}^{-1}\mathbf{\Lambda}\mathbf{Z}^{-1}g \implies \\
\sum_{i=1}^{KT}\mathbb{E}[V_i^2] &\leq \frac{1}{K}g^T\mathbf{Z}^{-1}T\mathbf{\Lambda}\mathbf{Z}^{-1}g \leq \frac{1}{K}g^\top\mathbf{Z}^{-1}g = C_0
\end{aligned}$$

Note that in deriving the second and the third inequality we used $\sup_z g(z) \leq 1$, which follows as $g(z) = \langle g, \phi(z)\rangle_{\mathcal{H}_k} \leq \|g\|_{\mathcal{H}_k}\|\phi(z)\|_{\mathcal{H}_k} = 1$

Applying Bernstien inequality to the collection of random variables $\{V_i\}_{i=1}^{KT}$ we obtain:

$$P\left(\left|\sum_{i=1}^{KT}V_i - g^\top\mathbf{Z}^{-1}T\mathbf{\Lambda}g\right| > r\right) \leq 2\exp\left(-\frac{r^2}{2\left(C_0 + C_1 r/3\right)}\right)$$

Taking $r = \sqrt{\frac{1}{K}\cdot g^\top\mathbf{Z}^{-1}g\cdot\log(1/\delta)} + \frac{2\log(1/\delta)}{3K}\max\{\sup_g g^\top\mathbf{Z}^{-1}g, \sup_z \phi^\top(z)\mathbf{Z}^{-1}\phi(z)\}$ we have with probability at least $1-\delta$:

$$\|\mathbf{Z}^{-1}\widetilde{\mathbf{Z}} - \mathbf{Id}\|_2 \leq \sqrt{\frac{1}{K}\cdot g^\top\mathbf{Z}^{-1}g\cdot\log(1/\delta)} + \frac{2\log(1/\delta)}{3K}\max\{\sup_g g^\top\mathbf{Z}^{-1}g, \sup_z \phi^\top(z)\mathbf{Z}^{-1}\phi(z)\}.$$

Substituting $K = \lceil \frac{T}{\gamma_T} \rceil$ and noting that $\mathbf{Z}^{-1} \prec \tau^{-1}\mathbf{Id}, K \geq T/\gamma_T$ we obtain:

$$\|\mathbf{Z}^{-1}\widetilde{\mathbf{Z}} - \mathbf{Id}\|_2 \leq \sqrt{\frac{\gamma_T}{T}\frac{1}{\tau}\log(1/\delta)} + \frac{\gamma_T}{T}\frac{2}{3\tau}\log(1/\delta)$$

To obtain a bound on the spectral norm of $\widetilde{\mathbf{Z}}^{-1}\mathbf{Z} - \mathbf{Id}$ note that if $\|\mathbf{Z}^{-1}\widetilde{\mathbf{Z}} - \mathbf{Id}\|_2 \leq b$ the eigen- values of $\mathbf{Z}^{-1}\widetilde{\mathbf{Z}}$ belong to the interval $(1-b, 1+b)$. Thus the eigen-values of $\widetilde{\mathbf{Z}}^{-1}\mathbf{Z}$ are contained in an interval $((1+b)^{-1}, (1-b)^{-1})$. We can finally conclude with probability at least $1-\delta$:

$$\|\widetilde{\mathbf{Z}}^{-1}\mathbf{Z} - \mathbf{Id}\|_2 \leq \frac{b}{1-b} \leq \left(\sqrt{\frac{\gamma_T}{T}\frac{1}{\tau}\log(1/\delta)} + \frac{\gamma_T}{T}\frac{2}{3\tau}\log(1/\delta)\right)\frac{1}{1-\left(\sqrt{\gamma_T/T\tau\log(1/\delta)} + 2\gamma_T/3T\tau\log(1/\delta)\right)}$$

Along with the condition that $T > (\frac{\log(1/\delta)}{C_p F^2 16})^{2\beta_p/(\beta_p-1)} = \overline{N}_1(\delta)$ which implies Vakili et al. [2021b] $\sqrt{\frac{\gamma_T}{T}\frac{1}{\tau}\log(1/\delta)} \leq 1/4$, we can simplify the expression as:

$$\|\widetilde{\mathbf{Z}}^{-1}\mathbf{Z} - \mathbf{Id}\|_2 \leq \frac{28}{17}\cdot\sqrt{\frac{\gamma_T}{T}\frac{\log(1/\delta)}{\tau}}$$

□

Lemma A.2. *Consider an T i.i.d point $\{w_i\}_{i=1}^T$ sampled according to a Borel measure ϱ from the domain $\mathcal{X}$ and introduce an estimator $\mu_T(w) = \phi(w)^\top \widehat{\mathbf{Z}}^{-1}\Phi_T y_T$, where $\widehat{\mathbf{Z}} = \Phi_T\Phi_T^\top + \tau\mathbf{Id}$. Let $\{z_i\}_{i=1}^{KT}$ be another set of i.i.d sampled point according to ϱ and let $\widetilde{\mathbf{Z}} = \frac{1}{K}\sum_{i=1}^{KT}\phi(z_i)\phi(z_i)^\top + \tau\mathbf{Id}$ be the same operator introduced in Lemma A.1. Define a new estimator:*

$$\overline{\mu}_T(w) = \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\Phi_T y_T$$

Fix an arbitrary $w \in \mathcal{W}$, then with probability at least $1-\delta$:

$$|\mu_T(w) - \overline{\mu}_T(w)| \leq \beta_1(\delta)\sqrt{\frac{\gamma_T}{T}}$$

where $\beta_1(\delta) = \left(\sqrt{162B^3/13F^2} + 81F^2/52\right)\log(8/\delta) + 28/17\frac{\log(4/\delta)}{\tau} + 2B\sqrt{108F^2\tau/13} + 4R\log(4/\delta)\sqrt{243F^2/26}$

Proof. We can re-write the estimator difference as:

$$\begin{aligned}|\mu_T(w) - \overline{\mu}_T(w)| &= |\phi(x)^\top(\widetilde{\mathbf{Z}}^{-1} - \mathbf{Z}^{-1})\Phi_T y_T| \leq \\ &\leq |\phi(w)^\top\widehat{\mathbf{Z}}^{-1}\Phi_T\varepsilon_{1:T}| + |\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\Phi_T\varepsilon_{1:T}| + |\phi(w)^\top(\widetilde{\mathbf{Z}}^{-1} - \widehat{\mathbf{Z}}^{-1})\Phi_T\Phi_T^\top f| \qquad (6)\end{aligned}$$

We start by bounding the first two terms of eq.(6). We will first bound the norms of the vectors $\|\phi(w)^\top\widehat{\mathbf{Z}}^{-1}\Phi_T\|, \|\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\Phi_T\|$ and then utilize the fact $\varepsilon_{1:T}$ is an R-sub-Gaussian independent from both vectors.

$$\begin{aligned}\|\phi(w)^\top\widehat{\mathbf{Z}}^{-1}\Phi_T\|_2^2 &\leq \sup_{w\in\mathcal{W}}\phi^\top(w)\widehat{\mathbf{Z}}^{-1}\Phi_T\Phi_T^\top\widehat{\mathbf{Z}}^{-1}\phi(w) \leq \\ &\leq \sup_{w\in\mathcal{W}}\phi^\top(w)\widehat{\mathbf{Z}}^{-1}\phi(w)\end{aligned}$$

Where, in the second line we use the identity $\Phi_T\Phi_T^\top = \widehat{\mathbf{Z}} - \tau\mathbf{Id} \prec \widehat{\mathbf{Z}}$. Next not that $\varepsilon_{1:T}$ is as an R-sub-Gaussian vector and thus after using [Salgia et al., 2023, Lemma 3.3,Lemma 3.4] we can bound the dot product with probability at least $1-\delta$:

$$|\phi(w)^\top \widehat{\mathbf{Z}}^{-1}\Phi_T\varepsilon_{1:T}| \leq 2R\log(2/\delta)\sqrt{\sup_w \phi^\top(w)\widehat{\mathbf{Z}}^{-1}\phi(w)} \leq 2R\log(1/\delta)\sqrt{108F^2/13}\sqrt{\frac{\gamma_T}{T}} \tag{7}$$

We continue by bounding the norm $\|\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\Phi_T\|_2$. We will utilize the result of [Salgia et al., 2023, Lemma 3.2] that allows to say for $T > \overline{N}(\delta/4)$ w.p $1-\delta/4$ $\|\mathbf{Z}^{-0.5}\widehat{\mathbf{Z}}\mathbf{Z}^{-0.5} - \mathbf{Id}\|_2 \leq 1/9$. Note that $\mathbf{Z}^{-0.5}\widehat{\mathbf{Z}}\mathbf{Z}^{-0.5}, \widehat{\mathbf{Z}}\mathbf{Z}^{-1}$ have the same spectrum's and hence $8/9\mathbf{Id} \prec \widehat{\mathbf{Z}}\mathbf{Z}^{-1} \prec 10/9\mathbf{Id}$. As both $\mathbf{Z}, \widehat{\mathbf{Z}}$ are positivite-definite matrices we can conclude $8/9\mathbf{Z} \prec \widehat{\mathbf{Z}} \prec 10/9\mathbf{Z}$. Using this result we can write:

$$\begin{aligned}
\|\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\Phi_T\|_2^2 &\leq \sup_{w\in\mathcal{W}} \phi(w)\widetilde{\mathbf{Z}}^{-1}\Phi_T\Phi_T^\top\widetilde{\mathbf{Z}}^{-1}\phi(w) \leq \sup_{w\in\mathcal{W}} \phi^\top(w)\widetilde{\mathbf{Z}}^{-1}\Phi\Phi^\top\widetilde{\mathbf{Z}}^{-1}\phi(w) \\
&\leq \sup_{w\in\mathcal{W}} \phi^\top(w)\widetilde{\mathbf{Z}}^{-1}\widehat{\mathbf{Z}}\widetilde{\mathbf{Z}}^{-1}\phi(w) \\
&\leq 10/9\cdot\sup_w \phi^\top(w)\widetilde{\mathbf{Z}}^{-1}\mathbf{Z}\widetilde{\mathbf{Z}}^{-1}\phi(w) \\
&\leq 10/9\cdot\sup_w \phi^\top(w)\widetilde{\mathbf{Z}}^{-1}\phi(w)\|\widetilde{\mathbf{Z}}^{-1/2}\mathbf{Z}\widetilde{\mathbf{Z}}^{-1/2}\|_2
\end{aligned}$$

To bound $\|\widetilde{\mathbf{Z}}^{-1/2}\mathbf{Z}\widetilde{\mathbf{Z}}^{-1/2}\|_2$ we will use Lemma A.1 that provides the bound $\|\widetilde{\mathbf{Z}}^{-1}\mathbf{Z}\|_2 < 3/2$, w.p $1-\delta/4$ for $T \geq \overline{N_1}(\delta/4)$. Note that $\widetilde{\mathbf{Z}}^{-1/2}\mathbf{Z}\widetilde{\mathbf{Z}}^{-1/2}, \widetilde{\mathbf{Z}}^{-1}\mathbf{Z}$ have the same spectrum's and thus $\|\widetilde{\mathbf{Z}}^{-1/2}\mathbf{Z}\widetilde{\mathbf{Z}}^{-1/2}\|_2 \leq 3/2$. We will once again use Lemma A.1 in bounding $\sup_w \phi^\top(w)\widetilde{\mathbf{Z}}^{-1}\phi(w)$. Namely note that $\|\widetilde{\mathbf{Z}}^{-1}\mathbf{Z} - \mathbf{Id}\|_2 \leq \frac{1}{2}$ and thus $1/2 \prec \widetilde{\mathbf{Z}}^{-1}\mathbf{Z} \prec 3/2$. As both $\mathbf{Z}, \widetilde{\mathbf{Z}}$ are positive definite we can conclude $1/2\widetilde{\mathbf{Z}} \prec \mathbf{Z} \prec 3/2\widetilde{\mathbf{Z}} \implies 2/3\widetilde{\mathbf{Z}}^{-1} \prec \mathbf{Z}^{-1} \prec 2\widetilde{\mathbf{Z}}^{-1}$ and hence:

$$\sup_w \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w) \leq 3/2\sup\phi(w)^\top\mathbf{Z}^{-1}\phi(w) \leq \frac{\gamma_T}{T}81F^2/13$$

Thus finally:

$$|\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\Phi_T\|_2^2 \leq 243F^2/26\frac{\gamma_T}{T}$$

Repeating the previous argument for the R-sub-Gaussian vector $\varepsilon_{1:T}$ we can write w.p at least $1-\delta$:

$$|\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\Phi_T\varepsilon_{1:T}| \leq 2R\log(1/\delta)\sqrt{243F^2/26}\sqrt{\frac{\gamma_T}{T}} \tag{8}$$

We now turn our attention to the first term of eq.(6):

$$\begin{aligned}
|\phi^\top(w)\widetilde{\mathbf{Z}}^{-1}\Phi\Phi^\top f - \phi^\top(w)\widehat{\mathbf{Z}}^{-1}\Phi\Phi^\top f| &= |\phi^\top(w)(\widetilde{\mathbf{Z}}^{-1} - \widehat{\mathbf{Z}}^{-1})(\widehat{\mathbf{Z}} - \tau\mathbf{Id})| \leq \\
&\leq |\phi^\top(w)(\widetilde{\mathbf{Z}}^{-1}\widehat{\mathbf{Z}} - \mathbf{Id})f| + \tau|\phi^\top(w)\widehat{\mathbf{Z}}^{-1}f| + \tau|\phi^\top(w)\widetilde{\mathbf{Z}}^{-1}f|
\end{aligned} \tag{9}$$

We will first bound the last two terms of eq.(9):

$$\begin{aligned}
\tau|\phi^\top(w)\widehat{\mathbf{Z}}^{-1}f| &\leq \tau\sqrt{\phi^\top(w)\widehat{\mathbf{Z}}^{-1}\phi(w)}\sqrt{f^\top\widehat{\mathbf{Z}}^{-1}f} \leq \\
&\leq B\sqrt{108F^2\tau/13}\sqrt{\frac{\gamma_T}{T}}
\end{aligned} \tag{10}$$

The first inequality is Cauchy-Schwartz while the second follows from [Salgia et al., 2023, Lemma 3.3,Lemma 3.4] and $\widehat{\mathbf{Z}}^{-1} \prec \tau^{-1}\mathbf{Id}$. We follow the same methodology in bounding the second term:

$$\tau|\phi^\top(w)\widetilde{\mathbf{Z}}^{-1}f| \leq \tau\sqrt{\phi^\top(w)\widetilde{\mathbf{Z}}^{-1}\phi(w)}\sqrt{f^\top\widetilde{\mathbf{Z}}^{-1}f} \leq$$
$$\leq B\sqrt{81F^2\tau/13}\sqrt{\frac{\gamma_T}{T}} \tag{11}$$

Here we once again used the previously derived inequality, $3/2\mathbf{Z}^{-1} \succ \widehat{\mathbf{Z}}^{-1}$.

To bound the first term of eq.(9), $|\phi^\top(w)(\widetilde{\mathbf{Z}}^{-1}\widehat{\mathbf{Z}} - \mathbf{Id})f|$ we will utilize Bernstein inequality. Note that $\widetilde{\mathbf{Z}}, \widehat{\mathbf{Z}}$ are independent and hence by conditioning on $\widetilde{\mathbf{Z}}$ we may assume it is a fixed matrix. We can re-write $\phi^\top(w)(\widetilde{\mathbf{Z}}^{-1}\widehat{\mathbf{Z}} - \mathbf{Id})f$ as:

$$\phi^\top(w)(\widetilde{\mathbf{Z}}^{-1}\widehat{\mathbf{Z}} - \mathbf{Id})f = \sum_{i=1}^{T}\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\phi(w_i)^\top f + \tau\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}f - \phi(w)^\top f =$$
$$= \sum_{i=1}^{T}\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\phi(w_i)^\top f - \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\mathbf{A}f$$

Where for notational convenience we introduce the notation $\mathbf{A} = \widetilde{\mathbf{Z}} - \tau\mathbf{Id}$. Let $Q_i = \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\phi(w_i)^\top f$. We first bound the moments of $\{Q_i\}_{i=1}^{KT}$:

$$\mathbb{E}[Q_i] = \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\mathbf{\Lambda}f \tag{12}$$
$$|Q_i| = \sup_y \phi(y)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\phi(w_i)^\top f \leq \sup_y\sqrt{\phi(y)^\top\widetilde{\mathbf{Z}}^{-1}\phi(y)^\top}\sqrt{\phi(y)^\top\widetilde{\mathbf{Z}}^{-1}\phi(y)}B \leq$$
$$\leq B\sup_y \phi(y)^\top\widetilde{\mathbf{Z}}^{-1}\phi(y)^\top := V_0 \tag{13}$$
$$\mathbb{E}[Q_i^2] \leq \mathbb{E}[(\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i))^2B^2] = B^2\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\mathbf{\Lambda}\widetilde{\mathbf{Z}}^{-1}\phi(w) \implies$$
$$\implies \sum_{i=1}^{T}\mathbb{E}[Q_i^2] = B^2\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}T\mathbf{\Lambda}\widetilde{\mathbf{Z}}^{-1}\phi(w) \leq 3/2B^2\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w) \leq BV_0 \tag{14}$$

In the last step we again used the inequality $3/2\mathbf{Z}^{-1} \succ \widetilde{\mathbf{Z}}^{-1}$

Applying Bernstein inequality with the moment bounds from (12-14) we have:

$$P\left(\left|\sum_{i=1}^{T}\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\phi(w_i)^\top f - \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}T\mathbf{\Lambda}f\right| \geq r\right) \leq 2\exp\left(-\frac{r^2}{2(BV_0 + rV_0/3)}\right)$$

Choosing $r = (\sqrt{2BV_0} + 2V_0/3)\log(2/\delta)$ ensures that w.p $1-\delta$:

$$\left|\sum_{i=1}^{T}\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\phi(w_i)^\top f - \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}T\mathbf{\Lambda}f\right| \leq (\sqrt{2BV_0} + 2V_0/3)\log(2/\delta)$$

We can now bound the original expression as:

$$\begin{aligned}
&\left|\phi^\top(w)(\widetilde{\mathbf{Z}}^{-1}\widehat{\mathbf{Z}} - \mathbf{Id})f\right| = \left|\sum_{i=1}^{T}\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\phi(w_i)^\top f - \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\mathbf{A}f\right| \leq \\
&\leq \left|\sum_{i=1}^{T}\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\phi(w_i)^\top f - \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}T\mathbf{\Lambda}f\right| + \left|\phi(w)^\top\widehat{\mathbf{Z}}^{-1}(\mathbf{A} - T\mathbf{\Lambda})f\right| \leq \\
&\leq (\sqrt{2BV_0} + 2V_0/3)\log(2/\delta) + \left|\phi(w)^\top\widetilde{\mathbf{Z}}^{-1}(\widetilde{\mathbf{Z}} - \mathbf{Z})f\right| \leq \\
&\leq \sqrt{\frac{\gamma_T}{T}}\left(\sqrt{162B^3/13F^2} + 81F^2/13\sqrt{\frac{\gamma_T}{T}}\right)\log(2/\delta) + 28/17\sqrt{\frac{\gamma_T}{T}\frac{\log(1/\delta)}{\tau}} \qquad (15)
\end{aligned}$$

In the last line we used the previously introduced inequality $\sup_w \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w)^\top \leq 81F^2/13\frac{\gamma_T}{T}$ and the result of Lemma A.1.

Finally adding all inequalities from eqs.(15,7,8,11,10) we obtain with probability $1-\delta$:

$$\begin{aligned}
&|\mu_T(w) - \overline{\mu}_T(w)| \leq \\
&\leq \sqrt{\frac{\gamma_T}{T}}\left(\left(\sqrt{162B^3/13F^2} + 81F^2/52\right)\log(8/\delta) + 28/17\sqrt{\frac{\log(4/\delta)}{\tau}} + 2B\sqrt{108F^2\tau/13} + 4R\log(4/\delta)\sqrt{243F^2/26}\right)
\end{aligned}$$

For notational convinience we introduce the shorthand notation:

$$\beta_1(\delta) = \left(\left(\sqrt{162B^3/13F^2} + 81F^2/52\right)\log(8/\delta) + 28/17\sqrt{\frac{\log(4/\delta)}{\tau}} + 2B\sqrt{108F^2\tau/13} + 4R\log(4/\delta)\sqrt{243F^2/26}\right)$$

□

We next prove that the estimator $\overline{\mu}_T$ approximates the reward function f sufficiently well over the entire domain $\mathcal{X}$. To this end we use the assumption 2.9 condition to construct a grid and apply the confidence bounds on each point in the gid.

Lemma A.3. *For the estimator $\overline{\mu}_T$ introduced in Lemma A.2 under the condition that $T > \max(\overline{N}(\delta/4), \overline{N}_1(\delta/4))$ we claim with probability at least $1-\delta$:*

$$\sup_{w\in\mathcal{W}}|\overline{\mu}_T(w) - f(w)| \leq \frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + (\beta(\delta/4|\mathcal{U}_T|) + \beta_1(\delta/4|\mathcal{U}_T|))\sqrt{\frac{\gamma_T}{T}}$$

Where $\beta(\delta) = \frac{2R}{\tau}\log\left(\frac{1}{\delta}\right)\sqrt{\frac{108F^2}{13}}$

Proof. We use the standard discretization argument from Vakili et al. [2021a]. Let $\mathcal{U}_T$ be the discretization described in the Assumption2.9. By the argument following used in Lemma A.2 we have $\widetilde{\mathbf{Z}}^{-1} \prec 3/2\mathbf{Z}^{-1} \prec 5/3\widehat{\mathbf{Z}}^{-1}$ and hence :

$$\|\overline{\mu}_T\|_{\mathcal{H}_k} \leq \|\widetilde{\mathbf{Z}}^{-1}\Phi_T\Phi_T^\top f\|_{\mathcal{H}_k} + \|\widetilde{\mathbf{Z}}^{-1}\Phi_T\varepsilon_T\|_{\mathcal{H}_K} \leq$$
$$\leq 5/3B + \sup_{g\in\mathcal{H}_K} g^\top\widetilde{\mathbf{Z}}^{-1}\Phi_T\varepsilon_{1:T} = 5/3B + \sup_{g\in\mathcal{H}_k}\sum_{i=1}^{T} g^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)\varepsilon_i \leq$$
$$\leq 5/3B + \sup\sum_{i=1}^{T}\sqrt{g^\top\widetilde{\mathbf{Z}}^{-1}g}\sqrt{\phi(w_i)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)}\varepsilon_i \leq$$
$$\leq 5/3B + \sqrt{1/\tau}\sum_{i=1}^{T}\sqrt{\phi(w_i)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)}\varepsilon_i$$

Once again $\widetilde{\mathbf{Z}}^{-1} \prec 3/2\mathbf{Z}^{-1}$ and thus w.p at least $1-\delta$: $\phi(w_i)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i) \leq 81F^2/13\gamma_T/T$. $\{\varepsilon_i\}_{i=1}^T$ are R-sub-Gaussian random variables and hence $\sum_{i=1}^T\sqrt{\phi(w_i)^\top\widetilde{\mathbf{Z}}^{-1}\phi(w_i)}\varepsilon_i$ is $R\sqrt{81F^2/13\gamma_T}$-sub-Gaussian. We now finally have with probability at least $1-\delta/4$:

$$\|\overline{\mu}_T\|_{\mathcal{H}_k} \leq 5/3B + 2R\sqrt{(81F^2/13\tau)\gamma_T\log(4/\delta)}$$

Repeating the same argument for $\widehat{\mathbf{Z}}$ we have with probability at least $1-\delta/4$

$$\|\mu_T\|_{\mathcal{H}_k} \leq B + 2R\sqrt{(81F^2/13\tau)\gamma_T\log(4/\delta)}$$

Recall that $[w]_T = \arg\min_{y\in\mathcal{U}_T}\|w-y\|_2$. We can now write w.p $1-3\delta/4$:

$$\forall w\in\mathcal{W}, |\overline{\mu}_T(w)-\mu_T(w)| \leq$$
$$\leq |\overline{\mu}_T([w]_T)-\overline{\mu}_T(w)| + |\mu_T([w]_T)-\mu_T(w)| + |\overline{\mu}_T([w]_T)-\mu_T([w]_T)| \leq$$
$$\leq \frac{8/3B+\sqrt{\gamma_T}2R\sqrt{81F^2/13\log(2/\delta)}}{T} + \beta_1(\delta/4|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}$$

The last line follows by applying the Lemma A.2 over the entire discrete grid $\mathcal{U}_T$ and utilizing Assumption(2.9).

By using [Vakili et al., 2021a, Theorem 1.] and noting that by [Salgia et al., 2023, Lemma 3.3, 3.4] $\sup_{w\in\mathcal{W}}\tau\phi(w)^\top\widehat{\mathbf{Z}}^{-1}\phi(w) = \sigma_T^2(w) \leq 108F^2/13\frac{\gamma_T}{T}$ we can now write w.p $1-\delta$:

$$\forall w\in\mathcal{W}, |f(w)-\overline{\mu}_T(w)| \leq$$
$$\leq |f(w)-\mu_T(w)| + |\mu_T(w)-\overline{\mu}_T(w)| \leq$$
$$\leq |f([w]_T)-\mu_T([w]_T)| + |f([w]_T)-f(w_T)| + |\mu_T([w]_T)-\mu_T(w_T)| + |\mu_T(w)-\overline{\mu}_T(w)| \leq$$
$$\leq \frac{11/3B+3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + (\beta(\delta/4|\mathcal{U}_T|)+\beta_1(\delta/4|\mathcal{U}_T|))\sqrt{\frac{\gamma_T}{T}}$$

Where $\beta(\delta) = \frac{2R}{\tau}\log\left(\frac{1}{\delta}\right)\sqrt{\frac{108F^2}{13}}$ □

B Privacy Constraints

Here we proved our algorithm achives ε-JDP privacy. First we bound the sensitivity of the estimator $\overline{\mu}_T$:

Lemma B.1. *Let $\mathcal{S}_T, \mathcal{S}_T^{'}$ be two t-neighbouring databases, and let $\overline{\mu}_T, \overline{\mu}_T'$ be the 2 estimator constructed for each of the databases . For $T > \max(\overline{N}_1(\delta/4), \overline{N}(\delta/4))$, we can bound the sensitivity of an estimator $\overline{\mu}_T$ in Algortihm(1) as :*

$$\Delta\overline{\mu}_T = \sup_{w\in\mathcal{W}} \sup_{\mathcal{S}_T, \mathcal{S}_T' \text{are } t \text{ neigbours}} |\overline{\mu}_T(w) - \overline{\mu}_T'(w)| \leq 2B \sup_{w\in\mathcal{W}} \overline{\sigma}^2(w)$$

Furthermore we can bound the posterior variance w.p. $1-\delta$ as:

$$\sup_{w\in\mathcal{W}} \overline{\sigma}^2(w) \leq \frac{81F^2\gamma_T}{13T}$$

Proof. Recall the parametric form of $\overline{\mu}_T$ introduced in the Lemma A.2

$$\overline{\mu}_T(w) = \phi(w)^\top \widetilde{\mathbf{Z}}^{-1} \Phi_{\mathbf{W}_T} \mathbf{Y}_T$$

Where $\widetilde{\mathbf{Z}} = 1/\lceil T/\gamma_T \rceil \Phi_{\mathcal{Z}} \Phi_{\mathcal{Z}}^\top + \tau \mathbf{Id}$. Note that in USCA query points $\{x_i\}_{i=1}^T$ are chosen only depending on the domain $\mathcal{X}$ and are independent from previous rewards and contexts. We can thus conclude that the dataset, $\mathcal{D}_T(\mathcal{S}_T) = \{w_1, w_2 \dots w_T\}, w_i = (x_i, c_i)$, consisting of (point, context) pairs only differs at time t i.e $\mathcal{D}_T(\mathcal{S}_T) \triangle \mathcal{D}_T(\mathcal{S}_T') = \{(x_t, c_t), (x_t, c_t')\}$. Introduce shorthand notation $w = (c, x)$ we can now write:

$$\begin{aligned}
|\overline{\mu}_T(w) - \overline{\mu}_T'(w)| &= \left| \sum_{w_i \in \mathcal{D}_T(\mathcal{S}_T)} \phi(w)^\top \widetilde{\mathbf{Z}}^{-1} \phi(w_i) y_i - \sum_{(w_i) \in \mathcal{D}_T(\mathcal{S}_T')} \phi(w)^\top \widetilde{\mathbf{Z}}^{-1} \phi(w_i) y_i \right| \leq \\
&\leq \left| \phi(w)^\top \widetilde{\mathbf{Z}}^{-1} \phi(w_t) y_t \right| + \left| \phi(w)^\top \widetilde{\mathbf{Z}}^{-1} \phi(w_t') y_t' \right| \leq \\
&\leq 2B \sup_{w\in\mathcal{W}} \phi(w)^\top \widetilde{\mathbf{Z}}^{-1} \phi(w)
\end{aligned}$$

Here the last line stems from the assumption that reward are bounded by B(see assumption(2.7)). By the result of Lemma A.1 for $T > \max(\overline{N}_1(\delta/2), \overline{N}(\delta/2))$ with probability at least $1-\delta/2$ $\|\widetilde{\mathbf{Z}}^{-1}\mathbf{Z}\|_2 < 3/2 \implies 3/2\mathbf{Z}^{-1} \succ \widetilde{\mathbf{Z}}^{-1}$. Thus, by [Salgia et al., 2023, Lemma 3.4] we have with probability at least $1-\delta$, $\sup_{w\in\mathcal{W}} \phi(w)^\top \widetilde{\mathbf{Z}}^{-1} \phi(w) < 3/2 \sup_{g\in\mathcal{W}} \phi(w)^\top \mathbf{Z}^{-1} \phi(w) < 81F^2/13 \frac{\gamma_T}{T}$ .Hence with probability at least $1-\delta$:

$$\sup_{w\in\mathcal{W}} \overline{\sigma}^2(w) \leq \frac{81F^2\gamma_T}{13T}$$

□

We next show the output of $\widehat{X}_T(c_{T+1})$ is ε-DP with respect to the previously seen history. The proof closely follows the Theorem 6 of McSherry and Talwar [2007].

Lemma B.2. *$\widehat{x}_T(c_{T+1})$ is ε -DP with respect to the database $\mathcal{S}_T \setminus \{c_{T+1}\}$*

Proof. As usual denote $\mathcal{S}_T, \mathcal{S}_T'$ to be two t-neighbouring databases. For the sake of space, introduce the shorthand notation $c_{T+1} \equiv c$. Recall that $\mathcal{Z}$ is the aproximating set of the algorithm. We can now write:

$$\frac{P(\widehat{x}_T(c) = r)}{P(\widehat{x}_T'(c) = r)} = \frac{\int_{\mathcal{Z}} \frac{\exp(\varepsilon \overline{\mu}_T(r,c)/(4B \sup \overline{\sigma}^2)}{\int_{\mathcal{X}} \exp(\varepsilon \overline{\mu}_T(r,c)/(4B \sup \overline{\sigma}^2) \nu_0(dr)} d\mathcal{Z}}{\int_{\mathcal{Z}} \frac{\exp(\varepsilon \overline{\mu}_T'(r,c)/(4B \sup \overline{\sigma}^2)}{\int_{\mathcal{X}} \exp(\varepsilon \overline{\mu}_T'(r,c)/(4B \sup \overline{\sigma}^2) \nu_0(dr)} d\mathcal{Z}}$$

By using the definition of sensitivity we can bound the ratio :

$$\frac{\frac{\exp(\varepsilon\overline{\mu}_T(r,c)/(4B\sup\overline{\sigma}^2)}{\int_{\mathcal{X}}\exp(\varepsilon\overline{\mu}_T(r,c)/(4B\sup\overline{\sigma})\nu_0(dr)}}{\frac{\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)}{\int_{\mathcal{X}}\exp(\varepsilon\overline{\mu}'_T(z,c)/(4B\sup\overline{\sigma}^2)\nu_0(dr)}} \leq \exp(\varepsilon\Delta\overline{\mu}_T/(4B\sup\overline{\sigma}^2))\frac{\int_{\mathcal{X}}\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)\nu_0(dr)}{\int_{\mathcal{X}}\exp(\varepsilon\overline{\mu}_T(r,c)/(4B\sup\overline{\sigma}^2)\nu_0(dr)} \leq$$
$$\leq \exp(2\varepsilon\Delta\mu_T/(4B\sup\overline{\sigma}^2))$$

Note that $\overline{\sigma}^2$ is only a function of $\mathcal{Z}$ and does not depend on the database $\mathcal{S}_T$. We can hence write:

$$\frac{P(\widehat{x}_T(c)=r)}{P(\widehat{x}'_T(c)=r)} \leq \frac{\int_{\mathcal{Z}}\exp(2\varepsilon\Delta\overline{\mu}_T/(4B\sup\overline{\sigma}^2))\frac{\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)}{\int_{\mathcal{X}}\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)\nu_0(dr)}d\mathcal{Z}}{\int_{\mathcal{Z}}\frac{\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)}{\int_{\mathcal{X}}\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)\nu_0(dr)}d\mathcal{Z}}$$

Recall that lemma B.1 ensures that $\Delta\overline{\mu}(T) \leq 2B\sup\overline{\sigma}^2$ with probability 1 over the randomness generated by $\mathcal{Z}$. We can hence write:

$$\frac{P(\widehat{x}_T(c)=r)}{P(\widehat{x}'_T(c)=r)} \leq \frac{\int_{\mathcal{Z}}\exp(2\varepsilon\Delta\mu_T/(4B\sup\overline{\sigma}^2))\frac{\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)}{\int_{\mathcal{X}}\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)\nu_0(dr)}d\mathcal{Z}}{\int_{\mathcal{Z}}\frac{\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)}{\int_{\mathcal{X}}\exp(\varepsilon\overline{\mu}'_T(r,c)/(4B\sup\overline{\sigma}^2)\nu_0(dr)}d\mathcal{Z}} \leq \exp(\varepsilon)$$

Which is what was originally claimed.

□

Theorem B.3. *USCA is ε-JDP.*

Proof. By previos lemma $\widehat{x}_T(c_{T+1})$ is ε-DP wrt the database $\mathcal{S}_T \setminus c_{T+1}$. To now formally show that USCA satisfies the ε-JDP, fix two t-neighbouring databases $\mathcal{S}_T, \mathcal{S}_T$, i.e $\mathcal{S}_T \triangle \mathcal{S}_T^{'} = \{(c_t, y_t), (c'_t, y'_t)\}$.
If $t = T+1$ there is nothing to show , as before time $T+1$ the points are chosen non-adaptively so their distribution does not depend on the contexts or rewards. Assume $t \leq T$, we have:

$$\frac{P\left((x_{t+1}, x_{t+2}, \dots x_T, \widehat{x}_T(c_{T+1}) = (r_{t+1}, r_{t+2}, \dots r_T, r_{T+1})\right)}{P\left((x'_{t+1}, x'_{t+2}, \dots x'_T, \widehat{x}'_T(c_{T+1})) = (r_{t+1}, r_{t+2}, \dots r_T, r_{T+1})\right)} = \frac{P(\widehat{x}_T(c_{T+1}) = r_{T+1} | \cap_{j=t+1}^{T}(x_j = r_j, y_j, c_j))}{P(\widehat{x}'_T(c_{T+1}) = r'_{T+1} | \cap_{j=t+1}^{T}(x_j = r_j, y_j, c_j)))} \tag{16}$$

The second equality follows as before time T al points are queried uniformly from the domain $\mathcal{X}$. They thus have the same distribution, independent of previous contexts and rewards.
To utilize the previously established result on ε-DP , we need to first fix the previous points $\{x_i\}_{i=1}^{t-1}$. We do this by applying the law of total probability to the numerator and the denominator of eq.(16):
:

$$\frac{P(\widehat{x}_T(c_{T+1}) = r_{T+1} | \cap_{j=t+1}^{T}(x_j = r_j, y_j, c_j))}{P(\widehat{x}'_T(c_{T+1}) = r_{T+1} | \cap_{j=t+1}^{T}(x_j = r_j, y_j, c_j)))} =$$
$$= \frac{\int_{\mathcal{M}_{t-1}} P(x_T(c_{T+1}) = r_{T+1} | \cap_{j=1, j\neq t}^{T}(x_j = r_j, y_j, c_j)) P(\cap_{j=1}^{t-1}(x_j = r_j, y_j, c_j))}{\int_{\mathcal{M}_{t-1}} P(x'_T(c_{T+1}) = r_{T+1} | \cap_{j=1, j\neq t}^{T}(x_j = r_j, y_j, c_j)) P(\cap_{j=1}^{t-1}(x_j = r_j, y_j, c_j))}$$

Where the randomness in the integration the first $t-1$ query points which is succinctly denoted as $\mathcal{M}_{t-1}$. Note that x_{T+1} only depends on $\mathcal{S}_T$ through previous (point, reward, context) triples. We can hence use the previously derived ε-DP of $\widehat{x}_T(c_{T+1})$ with respect to $\mathcal{S}_T$ to now write:

$$\frac{P(\widehat{x}_T(c_{T+1}) = r_{T+1} | \cap_{j=t+1}^{T} (x_j = r_j, y_j, c_j))}{P(\widehat{x}'_T(c_T+1) = r_{T+1} | \cap_{j=t+1}^{T} (x_j = r_j, y_j, c_j)))} =$$
$$= \frac{\int_{\mathcal{M}_{t-1}} P(\widehat{x}_T(c_{T+1}) = r_{T+1} | \cap_{j=1, j\neq t}^{T} (x_j = r_j, y_j, c_j)) P(\cap_{j=1}^{t-1}(x_j = r_j, y_j, c_j))}{\int_{\mathcal{M}_{t-1}} P(\widehat{x}'_T(c_{T+1}) = r_{T+1} | \cap_{j=1, j\neq t}^{T} (x_j = r_j, y_j, c_j)) P(\cap_{j=1}^{t-1}(x_j = r_j, y_j, c_j))} \leq$$
$$\leq \frac{\int_{\mathcal{M}_{t-1}} \exp(\varepsilon) P(\widehat{x}'_T(c_{T+1}) = r_{T+1} | \cap_{j=1, j\neq t}^{T} (x_j = r_j, y_j, c_j)) P(\cap_{j=1}^{t-1}(x_j = r_j, y_j, c_j))}{\int_{\mathcal{M}_{t-1}} P(\widehat{x}'_T(c_{T+1}) = r_{T+1} | \cap_{j=1, j\neq t}^{T} (x_j = r_j, y_j, c_j)) P(\cap_{j=1}^{t-1}(x_j = r_j, y_j, c_j))} = \exp(\varepsilon)$$

Plugging this result back into eq.(16) we finally have:

$$P\left((x_{t+1}, x_{t+2}, \dots \widehat{x}_T(c_{T+1})) = (r_{t+1}, r_{t+2}, \dots r_{T+1})\right) \leq$$
$$\leq \exp(\varepsilon) P\left((x'_{t+1}, x'_{t+2}, \dots \widehat{x}'_T(c_{T+1})) = (r_{t+1}, r_{t+2}, \dots r_{T+1})\right)$$

For every two neighbouring data-bases which is what was desired. □

C Utility Analysis

The following Lemma will be useful in applying the Exponential mechanism:

Lemma C.1. *Let ν_0 be the Lebesgue on $\mathbb{R}^d$. Consider a convex set $\mathcal{X} \subset \mathbb{R}^d$ with diameter bounded by $\mathrm{diam}\mathcal{X} \leq D_0$ and a L_g-Lipschitz function $g : \mathcal{X} \to \mathbb{R}$. For an arbitrary $r > 0$, we have:*

$$\frac{\nu_0 g^{-1}\left([g(x^*) - r, g(x^*)]\right)}{\nu_0 \mathcal{X}} \geq \min(1, (r/D_0 L_g)^d)$$

Where $x^* = \mathrm{argsup}_{x\in\mathcal{X}} g(x)$.

Proof. Note that if $r > D_0 L_g$ then $r > \sup_{x,y\in\mathcal{X}} \|g(x) - g(y)\| \leq L_g \sup \|x - y\|_2 \leq L_g D_0$ and thus $g^{-1}[g(x^*) - r, g(x^*)] \equiv \mathcal{X}$, giving the desired inequality.

Assume now $r < D_0 L_g$. Consider the image of $\mathcal{X}$ under the homothety centered at x^*, $\mathbf{H} : x \to x^*(1-\eta_0) + x\eta_0$ where $\eta_0 = (r/D_0 L_g)$. Denote by $\mathcal{Y} = \mathbf{H}(\mathcal{X})$, we will show that $\mathcal{Y} \subseteq g^{-1}[g(x^*) - r, g(x^*)]$ from which the desired inequality will follow, as $\nu_0(\mathcal{Y})/\nu_0(\mathcal{X}) = \eta_0^d$.
Note that by convexity, clearly $\mathcal{Y} \subset \mathcal{X}$. Fix an arbitrary point $z \in \mathcal{Y}$. Note that $\mathbf{H}$ scales distances by η_0. Indeed :

$$\|\mathbf{H}(x_1) - \mathbf{H}(x_2)\|_2 = \|\eta_0(x_1 - x_2)\|_2 = \eta_0 \|x_1 - x_2\|_2$$

It thus follows that $\mathrm{diam}\mathcal{Y} = \eta_0 \mathrm{diam}\mathcal{X} \leq \eta_0 D_0$. Hence we can bound $\|x^* - z\|_2 \leq \eta_0 D$, as $x^* \in \mathcal{Y}$. By Lipschitz condition we can further write:

$$|g(x^*) - g(z)| \leq L_g \eta_0 D_0 = r$$

Thus clearly $z \in g^{-1}[g(x^*) - r, g(x^*)]$. Note that this holds for all $z \in \mathcal{Y}$ and thus we have the desired $\mathcal{Y} \subseteq g^{-1}[g(x^*) - r, g(x^*)]$. □

We can now present a result characterizing the simple regret performance of USCA.

Theorem C.2. *Assume the kernel function satisfies the polynomial eigen-decay condition(2.2) for $\beta_p > 1$. For L_f-Lipshitz reward function and $T > \max(\overline{N}(\delta/4), \overline{N}_1(\delta/4))$ where $\overline{N}(\delta), \overline{N_1}(\delta)$ are δ-dependent constants introduced in Salgia et al. [2023] and Lemma A.1 respectively. We can bound the average simple regret of the output points $\widehat{x}_T(c_{T+1})$ w.p $1-\delta$ as :*

$$\mathop{\mathbb{E}}_{c_{T+1}\sim\kappa}\left[\sup_{x\in\mathcal{X}} f(x, c_{T+1}) - f(\widehat{x}_T(c_{T+1}), c_{T+1})\right] \leq$$
$$\leq 10\left(\frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(6/\delta)}}{T} + \beta_2(\delta/12|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right) +$$
$$+ \frac{\gamma_T}{T}\frac{1}{\varepsilon}\frac{648BF^2}{13}\left(d\log\left(\frac{2592BF^2TL_f}{13\gamma_T}\varepsilon\right) + \log(3/\delta)\right)$$

Here $\beta_2(\delta) = \beta_1(\delta) + \beta(\delta)$ where $\beta(\delta) = \frac{2R}{\tau}\sqrt{\frac{108F^2}{13}}\log\left(\frac{1}{\delta}\right)$ and β_1 is a δ-dependant constant introduced in Lemma A.2. In the above expression the randomness is over contexts, rewards and random coins of the algorithm

Proof. We will first show that the sample $\widehat{x}_T(c_{T+1}) \sim \mathcal{E}(\overline{\mu}(\cdot, c_{T+1}), \varepsilon, 2B\sup_{w\in\mathcal{W}}\overline{\sigma}^2(w))$ is close to $\sup_{x\in\mathcal{X}}\overline{\mu}_T(c_{T+1}, x)$. The full argument will then follow from the confidence bounds derived for the estimator $\overline{\mu}_T$ in Lemma A.3.
Define $\mathcal{A}_r = \{x \in \mathcal{X} | \overline{\mu}_T(x, c_{T+1}) \geq \sup_{x\in\mathcal{X}}\overline{\mu}_T(x, c_{T+1}) - r\}$. Consider 2 events in the sigma-algebra spanned by $\{\mathbf{W}_T, \mathbf{Y}_T, \mathcal{Z}\}$:

$$\Gamma_1 = \left\{\sup_{w\in\mathcal{W}}\overline{\sigma}^2(w) \leq \frac{81F^2\gamma_T}{13T}\right\}$$
$$\Gamma_2 = \left\{\sup_{w\in\mathcal{W}}|\overline{\mu}_T(w) - f(w)| \leq \frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + (\beta(\delta/4|\mathcal{U}_T|) + \beta_1(\delta/4|\mathcal{U}_T|))\sqrt{\frac{\gamma_T}{T}}\right\}$$

By Lemma A.3 $P(\Gamma_2) \geq 1-\delta$ and by Lemma A.1 $P(\Gamma_1) \geq 1-\delta$ by union bound we have $P(\Gamma_1, \Gamma_2) \geq 1-2\delta$. Note that:

$$P(\widehat{x}_T(c_{T+1}) \in \overline{\mathcal{A}}_r) \leq P(\widehat{x}_T(c_{T+1}) \in \overline{\mathcal{A}}_r | \Gamma_1, \Gamma_2) + 2\delta \cdot 1 \tag{17}$$

We thus only need to bound $P(\widehat{x}_T(c_{T+1}) \in \overline{\mathcal{A}}_r | \Gamma_1, \Gamma_2)$. In further writing we drop the conditioning notation in the interest of space. We can now write:

$$P(\widehat{x}_T(c_{T+1}) \in \overline{\mathcal{A}}_r) \leq \frac{P(\widehat{x}_T(c_{T+1}) \in \overline{\mathcal{A}}_r)}{P(\widehat{x}_T(c_{T+1}) \in \mathcal{A}_{r/2})} =$$
$$= \frac{\int_{\mathbf{W}_T,\mathbf{Y}_T,\mathcal{Z}}\int_{x\in\overline{\mathcal{A}}_r}\frac{\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}{\int_{x\in\mathcal{X}}\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}}{\int_{\mathbf{W}_T,\mathbf{Y}_T,\mathcal{Z}}\int_{x\in\mathcal{A}_{r/2}}\frac{\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}{\int_{x\in\mathcal{X}}\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}}$$

We can bound the ratio inside the outer integral as:

$$\frac{\int_{x\in\overline{\mathcal{A}}_r}\frac{\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}{\int_{x\in\mathcal{X}}\exp(\varepsilon\overline{\mu}_T(c,c_{T+1})/4B\sup\overline{\sigma}^2)}}{\int_{x\in\mathcal{A}_{r/2}}\frac{\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}{\int_{x\in\mathcal{X}}\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}} = \frac{\int_{x\in\overline{\mathcal{A}}_r}\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}{\int_{x\in\mathcal{A}_{r/2}}\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2} \leq$$
$$\leq \frac{\exp(\varepsilon(\sup\overline{\mu}_T(x,c_{T+1}) - r)/4B\sup\overline{\sigma}^2\nu_0(\overline{\mathcal{A}}_r)}{\exp(\varepsilon(\sup\overline{\mu}_T(x,c_{T+1}) - r/2)/4B\sup\overline{\sigma}^2\nu_0(\mathcal{A}_{r/2})} \leq \exp(-\varepsilon r/(8B\sup\overline{\sigma}^2))\frac{\nu_0(\overline{\mathcal{A}}_r)}{\nu_0(\mathcal{A}_{r/2})} \leq$$
$$\leq \exp(-\varepsilon r/(8B\sup\overline{\sigma}^2))\frac{\nu_0(\mathcal{X})}{\nu_0(\mathcal{A}_{r/2})} \leq$$

Plugging this back into the previous equation we have:

$$P(\widehat{x}_T(c_{T+1}) \in \overline{\mathcal{A}}_r) \leq \frac{\int_{\mathbf{W}_T,\mathbf{Y}_T,\mathcal{Z}} \exp(-\varepsilon r/(8B\sup\overline{\sigma}^2))\frac{\nu_0(\mathcal{X})}{\nu_0(\mathcal{A}_{r/2})}\int_{x\in\mathcal{A}_{r/2}}\frac{\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}{\int_{x\in\mathcal{X}}\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}}{\int_{\mathbf{W}_T,\mathbf{Y}_T,\mathcal{Z}}\int_{x\in\mathcal{A}_{r/2}}\frac{\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}{\int_{x\in\mathcal{X}}\exp(\varepsilon\overline{\mu}_T(x,c_{T+1})/4B\sup\overline{\sigma}^2)}} \tag{18}$$

We now bound the ratio $\frac{\nu_0(\mathcal{X})}{\nu_0(\mathcal{A}_{r/2})}$ conditioned on Γ_1, Γ_2. To this end we use Lemma C.1 along with confidence bounds in Lemma A.3 , implied by the event Γ_2. Consider the set:

$$\mathcal{A}' = \left\{ x \in \mathcal{X} | f(x, c_{T+1}) \geq \sup_{x\in\mathcal{X}} f(x, c_{T+1}) - r/2 + 2\left(\frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + \beta_2(\delta/4|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right)\right\}$$

Fix an arbitrary $z \in \mathcal{A}'$. By the definition of the event Γ_2 and the definition of $\mathcal{A}'$ we can now write:

$$\begin{aligned}
&\sup_{x\in\mathcal{X}} \overline{\mu}_T(x, c_{T+1}) - \overline{\mu}_T(z, c_{T+1}) \leq \\
&\leq f(x^*_{T+1}, c_{T+1}) - f(z, c_{T+1}) + 2\left(\frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + \beta_2(\delta/4|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right) = \\
&= \sup_{x\in\mathcal{X}} f(x, c_{T+1}) - f(z, c_{T+1}) + 2\left(\frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + \beta_2(\delta/4|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right) \leq r/2
\end{aligned}$$

It thus follows that $\forall z \in \mathcal{A}'$ we have $z \in \mathcal{A}_{r/2}$ and thus $\mathcal{A}' \subseteq \mathcal{A}_{r/2}$. We can now directly bound $\nu_0(\mathcal{A}_{r/2})/\nu_0(\mathcal{X})$ from Lemma C.1 :

$$\begin{aligned}
&\frac{\nu_0(\mathcal{A}_{r/2})}{\nu_0(\mathcal{X})} \geq \frac{\nu_0(\mathcal{A}')}{\nu_0(\mathcal{X})} \geq \\
&\min\left(1, \left(r/2 - 2\left(\frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + \beta_2(\delta/4|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right)\right)^d (1/D_0 L_f)^d\right)
\end{aligned}$$

From the event Γ_1 we also know $\sup_{w\in\mathcal{W}}\overline{\sigma}^2(w) \leq \frac{81F^2\gamma_T}{13T}$. Plugging both of these bounds in eq.(18) and integrating out the $\{\mathbf{W}_T, \mathbf{Y}_T, \mathcal{Z}\}$ we have:

$$\begin{aligned}
&P(\widehat{x}_T(c_{T+1}) \in \overline{\mathcal{A}}_r | \Gamma_1, \Gamma_2) \\
&\leq \frac{\exp\left(-\frac{13rT}{648F^2\gamma_T}\right)}{\min\left(1, \left(r/2 - 2\left(\frac{11/3B+3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + \beta_2(\delta/4|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right)\right)^d (1/(D_0 L_f))^d\right)}
\end{aligned}$$

By choosing r as:

$$\begin{aligned}
r(\delta) =& 8\left(\frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + \beta_2(\delta/2|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right) + \\
&+ \frac{\gamma_T}{T}\frac{1}{\varepsilon}\frac{648BF^2}{13}\left(d\log\left(\frac{2592BF^2TL_fD_0}{13\gamma_T}\varepsilon\right) + \log(1/\delta)\right)
\end{aligned} \tag{19}$$

and using eq.(17) we have with probability at least $1 - 3\delta$:

$$\sup_{x\in\mathcal{X}} \overline{\mu}_T(x, c_{T+1}) - \overline{\mu}_T(x_{T+1}, c_{T+1}) \leq r(\delta) \tag{20}$$

Recall the notation $x^*_{T+1} = \text{argsup}_{x\in\mathcal{X}} f(x, c_{T+1})$. To finish the proof we once again use Lemma 4.2 to guarantee w.p at least $1-\delta$

$$
\begin{aligned}
&f(x^*_{T+1}, c_{T+1}) - f(x_{T+1}, c_{T+1}) = \\
&= f(x^*_{T+1}, c_{T+1}) - \overline{\mu}_T(x^*_{T+1}, c_{T+1}) - (f(x_{T+1}, c_{T+1}) - \overline{\mu}_T(x_{T+1}, c_{T+1})) + (\overline{\mu}_T(x^*_{T+1}, c_{T+1}) - \overline{\mu}_T(x_{T+1}, c_{T+1})) \leq \\
&\leq 2 \sup_{w\in\mathcal{W}} |f(w) - \overline{\mu}_T(w)| + \sup_{x\in\mathcal{X}} \overline{\mu}_T(x, c_{T+1}) - \overline{\mu}_T(x_{T+1}, c_{T+1}) \leq \\
&\leq r(\delta/3) + 2\left(\frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(2/\delta)}}{T} + \beta_2(\delta/4|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right)
\end{aligned}
$$

Note that in the above expression the randomness is over the previous contexts,rewards and random coins of the algorithm and not over the final context c_{T+1}. It thus follows that the claim holds uniformly over the entire context set $\mathcal{C}$. We can thus write:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{c_{T+1}\sim\kappa}\left[\sup_{x\in\mathcal{X}} f(x, c_{T+1}) - f(\widehat{x}_T(c_{T+1}), c_{T+1})\right] \leq \\
&\leq 10\left(\frac{11/3B + 3R\sqrt{\gamma_T 81F^2/13\log(6/\delta)}}{T} + \beta_2(\delta/12|\mathcal{U}_T|)\sqrt{\frac{\gamma_T}{T}}\right) + \\
&+ \frac{\gamma_T}{T}\frac{1}{\varepsilon}\frac{648BF^2}{13}\left(d\log\left(\frac{2592BF^2TL_fD_0}{13\gamma_T}\varepsilon\right) + \log(3/\delta)\right)
\end{aligned}
$$

□

D Kernel Trick

In this section we present an efficient way to calculate the posterior statistics $\overline{\mu}_T, \overline{\sigma}$.

Lemma D.1. *The estimator given in parametric form in LemmaA.2 can be equivalently written as:*

$$
\overline{\mu}_T(w) = \frac{1}{\tau}k_{\mathbf{W}_T}(w)^\top \mathbf{Y}_T - \frac{1}{\tau}k_{\mathcal{Z}}(w)^\top \left(\mathbf{K}_{\mathcal{Z},\mathcal{Z}} + K\tau\mathbf{I}_{\mathcal{Z}}\right)^{-1}\mathbf{K}_{\mathcal{Z},\mathbf{W}_T}\mathbf{Y}_T \tag{21}
$$

$$
\overline{\sigma}(w) = \frac{1}{\tau}\left(k(w,w) - k_{\mathcal{Z}}(w)^\top \left(K_{\mathcal{Z},\mathcal{Z}} + K\tau\mathbf{I}_{\mathcal{Z}}\right)^{-1}k_{\mathcal{Z}}(w)\right) \tag{22}
$$

Where $\mathbf{K}_{\mathcal{Z},\mathbf{W}_T} = \{k(a,b)\}_{a\in\mathcal{Z},b\in\mathbf{W}_T}, \mathbf{K}_{\mathcal{Z},\mathcal{Z}} = \{k(a,b)\}_{(a,b)\in\mathcal{Z}^2}$ *and* $K = \lceil T/\gamma_T\rceil$.

Proof. We introduce the shorthand notation $K = \lceil T/\gamma_T\rceil$. The parametric expression for $\overline{\mu}_T$ expression can be re-written as:

$$
\begin{aligned}
\overline{\mu}_T(w) &= \phi(w)^\top\widetilde{\mathbf{Z}}^{-1}\Phi_{\mathbf{W}_T}\mathbf{Y}_T = K\phi(w)^\top(\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top + K\tau\mathbf{Id})^{-1}\Phi_{\mathbf{W}_T}\mathbf{Y}_T = \\
&= \frac{1}{\tau}\phi(w)^\top\left(\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top + K\tau\mathbf{Id}\right)^{-1}(\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top + K\tau\mathbf{Id} - \Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top)\Phi_{\mathbf{W}_T}\mathbf{Y}_T = \\
&= \frac{1}{\tau}k_{\mathbf{W}_T}(w)^\top\mathbf{Y}_T - \frac{1}{\tau}\phi(w)^\top\left(\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top + K\tau\mathbf{Id}\right)^{-1}\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top\Phi_{\mathbf{W}_T}\mathbf{Y}_T
\end{aligned}
$$

Where The second line follows from the reproducing property. We next utilize a commonly applied identity Valko et al. [2013], Vakili et al. [2021a] $\left(\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top + K\tau\mathbf{Id}\right)^{-1}\Phi_{\mathcal{Z}} = \Phi_{\mathcal{Z}}\left(\Phi_{\mathcal{Z}}^\top\Phi_{\mathcal{Z}} + K\tau\mathbf{Id}\right)^{-1}$. We can now write:

$$
\begin{aligned}
\overline{\mu}_T(w) &= \frac{1}{\tau}k_{\mathbf{W}_T}(w)^\top\mathbf{Y}_T - \frac{1}{\tau}\phi(w)^\top\Phi_{\mathcal{Z}}\left(\Phi_{\mathcal{Z}}^\top\Phi_{\mathcal{Z}} + K\tau\mathbf{Id}\right)^{-1}\Phi_{\mathcal{Z}}^\top\Phi_{\mathbf{W}_T}\mathbf{Y}_T = \\
&= \frac{1}{\tau}k_{\mathbf{W}_T}(w)^\top\mathbf{Y}_T - \frac{1}{\tau}\phi(w)^\top\Phi_{\mathcal{Z}}\left(\Phi_{\mathcal{Z}}^\top\Phi_{\mathcal{Z}} + K\tau\mathbf{Id}\right)^{-1}\Phi_{\mathcal{Z}}^\top\Phi_{\mathbf{W}_T}\mathbf{Y}_T = \\
&= \frac{1}{\tau}k_{\mathbf{W}_T}(w)^\top\mathbf{Y}_T - \frac{1}{\tau}k_{\mathcal{Z}}(w)^\top\left(\mathbf{K}_{\mathcal{Z},\mathcal{Z}} + K\tau\mathbf{I}_{\mathcal{Z}}\right)^{-1}\mathbf{K}_{\mathcal{Z},\mathbf{W}_T}\mathbf{Y}_T
\end{aligned}
$$

Where $\mathbf{K}_{\mathcal{Z},\mathbf{W}_T} = \{k(a,b)\}_{a\in\mathcal{Z},b\in\mathbf{W}_T}$ and $\mathbf{K}_{\mathcal{Z},\mathcal{Z}} = \{k(a,b)\}_{(a,b)\in\mathcal{Z}^2}$. We use a similar approach in calculating $\overline{\sigma}(w)$:

$$\begin{aligned}
\overline{\sigma}(w) &= \phi(w)^\top \widetilde{\mathbf{Z}}^{-1}\phi(w) = \\
&= \frac{1}{\tau}\phi(w)^\top \left(\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top + K\tau\mathbf{Id}\right)^{-1}(\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top + K\tau\mathbf{Id} - \Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top)\phi(w) = \\
&= \frac{1}{\tau}k(w,w) - \frac{1}{\tau}\phi(w)^\top \left(\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top + K\tau\mathbf{Id}\right)^{-1}\Phi_{\mathcal{Z}}\Phi_{\mathcal{Z}}^\top\phi(w) = \\
&= \frac{1}{\tau}k(w,w) - \frac{1}{K^2\tau}\phi(w)^\top\Phi_{\mathcal{Z}}\left(\Phi_{\mathcal{Z}}^\top\Phi_{\mathcal{Z}} + K\tau\mathbf{Id}\right)^{-1}\Phi_{\mathcal{Z}}^\top\phi(w) = \\
&= \frac{1}{\tau}\left(k(w,w) - k_{\mathcal{Z}}(w)^\top\left(K_{\mathcal{Z},\mathcal{Z}} + K\tau\mathbf{I}_{\mathcal{Z}}\right)^{-1}k_{\mathcal{Z}}(w)\right)
\end{aligned}$$

□