Differential Privacy in Kernelized Contextual Bandits via Random Projections

Nikola Pavlovic
Cornell University

Sudeep Salgia
Carnegie Mellon University

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. We study this problem under an additional constraint of Differential Privacy, where the agent 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 achieves the state-of-the-art cumulative regret of $\widetilde{\mathcal{O}}(\sqrt{\gamma_T T}+\frac{\gamma_T}{\varepsilon_{\mathrm{DP}}})$ and $\widetilde{\mathcal{O}}(\sqrt{\gamma_T T}+\frac{\gamma_T\sqrt{T}}{\varepsilon_{\mathrm{DP}}})$ over a time horizon of T in the joint and local models of differential privacy, respectively, where γ_T is the effective dimension of the kernel and $\varepsilon_{\mathrm{DP}} > 0$ is the privacy parameter. The key ingredient of the proposed algorithm is a novel private kernel-ridge regression estimator which is based on a combination of private covariance estimation and private random projections. It offers a significantly reduced sensitivity compared to its classical counterpart while maintaining a high prediction accuracy, allowing our algorithm to achieve the state-of-the-art performance guarantees.

1 INTRODUCTION

We study the contextual bandit problem, where a learning agent aims to maximize an unknown reward function f based on noisy observations of the function at sequentially queried points. Specifically, at each time instant t, the agent is presented with a context $c_t \in \mathcal{C}$, which is drawn i.i.d. from a context distribution κ. Based on the context, the agent takes an action $x_t \in \mathcal{X}$ and observes y_t, a noisy value of the reward $f(c_t, x_t)$. We measure the performance of a learning algorithm $\mathscr{A}$ using its cumulative regret over a time horizon of T, which is defined as follows:

$$\mathrm{R}_T(\mathscr{A}) := \mathop{\mathbb{E}}_{c_t\sim\kappa}\left[\sum_{t=1}^{T}\max_{x\in\mathcal{X}} f(c_t,x) - f(c_t,x_t)\right] \quad (1)$$

where the expectation is taken over the context sequence. In this work, we focus on the kernelized setting, i.e., the reward function $f : \mathcal{C}\times\mathcal{X}\to\mathbb{R}$ is assumed to belong to a Reproducing Kernel Hilbert space (RKHS) of a known kernel $k : \mathcal{W}\times\mathcal{W}\to\mathbb{R}$, where $\mathcal{W} = \mathcal{C}\times\mathcal{X}$. The kernel-based modeling offers a highly expressive framework for the reward function. In particular, it is known that the RKHS of the Matérn family of kernels, can approximate almost all continuous functions on compact subsets of $\mathbb{R}^d$ (Srinivas et al., 2010). As a result, kernelized bandits have received significant attention in recent years.

1.1 Private Kernel Bandits

In several applications of contextual bandits, the contexts and rewards carry sensitive information that needs to be protected. For example, in recommendation systems, the contexts correspond to the information related to the user to whom the algorithm needs to recommend the products (the action) and the reward captures the user behavior, i.e., whether the user interacted with the recommended products. Evidently, the contexts and rewards contain user-related private information that needs to be systematically protected. This need for systematic guarantees on privacy in contextual bandit applications has fueled the study of contextual bandits under privacy constraints.

For the problem of contextual bandits, there are two notions of privacy that are widely considered — Joint Differential Privacy (JDP) (Shariff and Sheffet, 2018) and Local Differential Privacy (LDP) (Han et al., 2021). At a high level, in the JDP setting, the algorithm directly observes the sensitive data, i.e. the context-reward pair, and needs to ensure that the actions taken

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

by the algorithm are differentially private w.r.t. to the data. On the other hand, in the LDP setting, the algorithm gets to observe only a privatized version of the data and use that for future predictions. Evidently, LDP is a stricter notion of privacy than JDP.

The problems of contextual multi-armed bandits and contextual linear bandits have been studied in detail under both JDP and LDP. On the other hand, our understanding of the problem of private contextual kernel bandits is relatively limited. Existing studies either consider weaker notions of privacy and utility, e.g., only reward-level privacy and/or focus only on simple regret, or limit their attention to a narrow family of structured and highly smooth kernels that do not reflect the inherent hardness of the kernelized setting. The objective of this work is to develop an algorithm for private contextual kernelized bandits that offers state-of-art cumulative regret performance across a wide family of kernels under both JDP and LDP.

1.2 Main Results

We propose a new private contextual kernel bandit algorithm called CAPRI, that achieves state-of-the-art cumulative regret performance across a wide range of kernels under both JDP and LDP. In particular, we demonstrate that CAPRI achieves a cumulative regret of $\widetilde{\mathcal{O}}\left(\sqrt{T\gamma_T} + \frac{\gamma_T}{\varepsilon_{\mathrm{DP}}}\right)$ and $\widetilde{\mathcal{O}}\left(\frac{\gamma_T\sqrt{T}}{\varepsilon_{\mathrm{DP}}} + \sqrt{T\gamma_T}\right)$ under the JDP and LDP settings respectively. Here, γ_T refers to the effective dimensionality of the kernel and is a measure of the complexity of the kernel and $\varepsilon_{\mathrm{DP}}$ is the privacy parameter. Thus, our guarantees hold for all kernel families with polynomially decaying eigenvalues, which include the widely used Matérn and Squared-Exponential families of kernels. To the best of our knowledge, this is the first work to provide cumulative regret guarantees for the Matérn family of kernels, even under the less stringent constraint of JDP.

Notably, our results recover the order optimal cumulative regret for contextual kernel bandits as $\varepsilon_{\mathrm{DP}} \to \infty$, i.e., in the non-private setting (Scarlett et al., 2017). Moreover, for the simplest case of linear kernels, i.e., linear bandits, where $\gamma_T = \mathcal{O}(d)$, our regret bounds match the lower bound under JDP and achieve the best known bound under LDP (He and Zhang, 2022; Han et al., 2021).

The kernel-based modeling of the reward function, which allows for infinite-dimensional feature vectors, presents several challenges over the simpler linear reward function model for which the problem is well-understood. Firstly, in any sequential learning problem, it is imperative to add a small layer of privacy at each time step due to the dynamic nature of the dataset. In absence of a careful control, this can result in trivial guarantees on the utility of the algorithm. Under the premise of JDP, the common approach to address this hurdle in linear bandits (Shariff and Sheffet, 2018; Dubey and Pentland, 2020) is to use the tree-based mechanism (Dwork and Roth, 2014). The tree-based mechanism enables one to construct a private version of the covariance matrix, which is a sum of rank one matrices, and consequently a private predictor of the reward function whose error differs from the non-private version only by a factor growing logarithmically in the number of observed points. On the other hand, the requirement of privatization of the data at the source under LDP precludes the use of tree-based mechanism to construct a private covariance matrix. The challenge of privatizing the covariance matrix in the LDP framework is resolved by lower bounding the minimum eigenvalue of the covariance matrix (Han et al., 2021; Ruiquan et al., 2024). Specifically, the authors establish that the smallest eigenvalue of the covariance matrix grows as $\Omega(T/d)$ and thus addition of privatization noise of standard deviation $\mathcal{O}(\sqrt{T})$ has only a small effect on the predictive performance.

In contextual kernel bandits, these techniques are no longer applicable due to the infinite-dimensional nature of the feature vectors. In particular, the tree-based mechanism leverages the feature space representation of the covariance matrix, which in this case would be infinite-dimensional and hence impractical to work with. On the other hand, the finite-dimensional representation of the covariance matrix using the kernel trick no longer allows for a representation through a sum of rank-one matrices, which prevents the use of the tree-based mechanism. Similarly, an eigenvalue based argument cannot be applied for kernel bandits under the LDP framework because the decaying spectrum of the kernel operator implies that the smallest eigenvalue of the covariance matrix is bounded as $\mathcal{O}(1)$ and does not grow with the size of the dataset. Another subtle challenge that is unique to the kernel-based setting is that the predictor used to estimate the reward function is also an infinite-dimensional element in the RKHS. In the linear setting, the predictor is a finite-dimensional vector which can be privatized by adding a random vector from the same vector space. Such a technique cannot be extended to the kernel-based setting. Moreover, privatization of all possible outputs would scale the privacy-based error by a factor of $|\mathcal{W}|$, rendering it trivial.

Our proposed approach presents a significant departure from the commonly used methodology in private linear bandits. The key contribution of this work is a novel, private predictor for the reward function that offers a similar predictive power as its non-private counterpart. In particular, we privatize the covariance matrix by re-

placing it with a statistically identical and independent copy of the matrix. The independence allows us to obtain privacy for free, while the statistical similarity enables achieving a comparable predictive performance. An important aspect of our algorithm that enables such a construction is the use of reward-independent sampling, i.e., we use actions that do not depend upon the previous rewards. This allows us to draw i.i.d.. copies of context-action pairs to construct the required covariance matrix without requiring feedback from the function. Secondly, to privatize the predictor as an RKHS function, we project the function onto the subspace spanned by finitely many feature vectors, which are drawn at random and are treated as basis vectors for the subspace. This helps approximate the predictor through a finite-dimensional vector consisting of the coefficients the predictor when represented in the randomly chosen basis. The private predictor is obtained by privatizing the coefficients of the original predictor. We show that the 'random basis', if chosen correctly, can not only provide a faithful approximation of the original predictor but also ensure that the coefficients have low sensitivity to the data points. This enables privatization at only a small cost to the predictive performance. We show that these two approaches can be seamlessly integrated to yield our final estimator that has low sensitivity to the underlying dataset and high predictive power. The CAPRI algorithm builds upon this new private estimator by integrating it into an explore-then-eliminate framework.

1.3 Related Work

Kernel-based bandits. The problem of kernel-based bandit optimization has been extensively studied in the non-private setting. The study of the problem was initiated by the seminal work of Srinivas et al. (2010). Since then, numerous algorithms for kernel-based bandits have been proposed in both contextual (Valko et al., 2013) and non-contextual settings (Chowdhury and Gopalan, 2017; 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. Both multi-armed bandits and linear bandits under privacy constraints have received considerable attention in the literature. In multi-armed bandits, Azize and Basu (2020); Azize et al. (2024) study the problem of private best arm-identification and Tenenbaum et al. (2021) consider cumulative regret minimization under shuffle model of privacy. Shariff and Sheffet (2018) were the first to study the problem of contextual linear bandits with adversarial contexts under the JDP model. They propose a new algorithm that achieves a cumulative regret of $\mathcal{O}(\sqrt{T}d^{3/4}/\varepsilon_{\mathrm{DP}})$ under a finite arm set assumption. Zheng et al. (2020) consider the analogous problem under the LDP setting and propose an algorithm with regret $\mathcal{O}((Td)^{3/4}/\varepsilon_{\mathrm{DP}})$. He and Zhang (2022) study linear bandits with stochastically generated contexts, under both JDP and LDP. They establish lower bounds under both the privacy models, specifically for JDP they outline the algorithm[1] that achieves $\mathcal{O}\left(\sqrt{dT} + \frac{d\log(1/\delta_{\mathrm{DP}})}{\varepsilon_{\mathrm{DP}}}\right)$. Han et al. (2021) propose an algorithm for linear and non-linear contextual bandits under the LDP model that incurs a regret of $\mathcal{O}(\sqrt{T}d^{3/2}/\varepsilon_{\mathrm{DP}})$ under a finite-arm setting.

The problem of private kernel bandits was first studied by Kusner et al. (2015) in the context of hyperparameter tuning. Kharkovskii et al. (2020) studied private kernel bandits for Squared Exponential kernels under a weaker notion of LDP where the selected actions need to be locally private. Under the restrictive assumption of a diagonally dominant covariance matrix, they propose an algorithm with a simple regret of $\tilde{\mathcal{O}}((\varepsilon_{\mathrm{DP}}^{-2} + \log T/T)^{1/2})$. Zhou and Tan (2021) consider kernel bandits with heavy-tailed noise, under a weaker notion of privacy where only the rewards need to be private.

The two studies that are closest to ours are those by Dubey (2021) and Pavlovic et al. (2025). Dubey (2021) study the problem of private kernel bandits with adversarial contexts under both JDP and LDP models. They approximate the functions in RKHS using a low-dimensional embedding which allows them to reduce the original problem of that of private linear bandits and circumvent the challenges of kernelized bandits. However, the existence of such a low-dimensional embedding that can approximate all functions in the RKHS is guaranteed only for highly smooth kernels like the Squared Exponential kernel. Consequently, the approach in Dubey (2021) does not extend beyond the Squared Exponential kernel. The results and methodology proposed in this work hold for a much wider class of kernels, e.g. to the Matérn family of kernels. Pavlovic et al. (2025) consider private kernel bandits with stochastic contexts under JDP. They propose an algorithm that achieves a simple regret of $\widetilde{\mathcal{O}}(\sqrt{\frac{\gamma_T}{T}} + \frac{\gamma_T}{T\varepsilon_{\mathrm{DP}}})$. In this work, we also consider stochastic contextual kernel bandits under both JDP and LDP and focus on developing an algorithm with low cumulative regret which is more challenging than the metric of simple regret considered in that work. Moreover, our cumulative regret bound implies the

[1]Authors however do not offer a rigorous theoretical derivation proving this upper bound.

simple regret bound obtained in Pavlovic et al. (2025).

The wide improvement over the results obtained in Pavlovic et al. (2025) is made possible by the novel approach of using random projections in the Hilbert Space. The entire analysis in the online setting for both JDP and LDP essentially hinges on using this idea of private change of basis via random projections. Specifically, the random projection decouples the mean estimator from its data-dependent basis by projecting it onto a data-independent private basis, which then allows us to privatize each basis coefficient individually. This step is especially helpful for the case of LDP. In particular, the careful design of our random projection not only decouples the estimator from a data-dependent basis but also isolate the contribution of each user into a single coefficient. This allows us to easily localize and privatize the contribution from each user and provide the required LDP guarantees. For more details on please see Sec.(3)

2 PROBLEM FORMULATION AND PRELIMINARIES

2.1 RKHS, Mercer Theorem, and GP models

Consider a positive definite kernel $k : \mathcal{W} \times \mathcal{W} \to \mathbb{R}$, where $\mathcal{W} = \mathcal{C} \times \mathcal{X}$ is the product of the context and action 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, given by $\|f\|^2_{\mathcal{H}_k} = \langle f, f \rangle_{\mathcal{H}_k}$. 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 $w, w' \in \mathcal{W}$.*

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

Definition 2.2. Vakili et al. (2021) 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 assumptions on the kernel k and the eigenfunctions $\{\psi_j\}_{j \in \mathbb{N}}$, that have also been adopted recent results for kernelized bandits (Salgia et al., 2023a; Pavlovic et al., 2025).

Let $\mathcal{M}_{\mathcal{X}} = \{\mathcal{X}' \subseteq \mathcal{X} | \text{Unif}(\mathcal{X}')\}$ be a collection of measures. Here $\text{Unif}(\mathcal{X}')$ denotes the uniform measure over the set $\mathcal{X}'$. Let $\mathcal{M}_{\mathcal{W}}$ denote the family of measures over subsets of $\mathcal{W}$, defined as $\mathcal{M}_{\mathcal{W}} := \{\varrho(c, x) = \kappa(c) p(x|c) \mid p(x|c) \in \mathcal{M}_{\mathcal{X}}\ \forall c \in \mathcal{C}\}$.

Assumption 2.3. The kernel k satisfies the polynomial eigendecay condition with parameter $\beta_p > 1$. Let $\varrho \in \mathcal{M}_{\mathcal{W}}$ and $\{\psi^{\varrho}_j\}_{j \in \mathbb{N}}$ denote the corresponding eigenfunctions of the kernel operator corresponding to ϱ. There exists $F > 0$ such that $\|\psi^{\varrho}_j(w)\|_\infty \leq F$ for all $j \in \mathbb{N}$ and $\varrho \in \mathcal{M}_{\mathcal{W}}$.

We stress that the assumption (2.3) bounds eigenfunctions over the support of the measure ϱ , and not the set $\mathcal{W}$. Assuming bounded eigen-functions with respect to a measure is standard practice in both kernel bandit (Salgia et al., 2024; Pavlovic et al., 2025; Vakili et al., 2021; Chatterji et al., 2019; Riutort-Mayol et al., 2023) and functional estimation (Maggioni and Mhaskar, 2008; Mhaskar et al., 2001; Grochenig, 2020; Filbir and Mhaskar, 2010) literature. Moreover, we emphasize that the boundedness of the eigen-functions is a crucial assumption in obtaining meaningful information gain bounds for a kernel k (Vakili et al., 2021).

Next, we briefly introduce Gaussian Processes following a standard exposition in Srinivas et al. (2010); Chowdhury and Gopalan (2017); Kanagawa et al. (2018). 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 $[k(w_i, w_j)]^T_{i,j=1}$. 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 $\{\mathbf{W}_T, \mathbf{y}_T\} = \{(w_i, y_i)\}^T_{i=1}$

from the underlying process, the expression for posterior mean and variance of GP model is given as follows:

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

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

In the above expressions, $k_{\mathbf{W}_T}(w) := [k(w_1,w), k(w_2,w)\dots k(w_T,w)]^\top$, $\mathbf{K}_{\mathbf{W}_T,\mathbf{W}_T} := [k(w_i,w_j)]_{i,j=1}^T$, $\mathbf{I}_T$ 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 modeling 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 $\mathbf{W}_T = \{w_1, w_2, \dots, w_T\} \in \mathcal{W}$, 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 helps characterize the regret performance of kernel bandit algorithms (Srinivas et al., 2010; Chowdhury and Gopalan, 2017). For the kernels satisfying polynomial eigendecay condition with parameter β_p, γ_T is known to satisfy $\mathcal{O}(T^{\frac{1}{\beta}p}\log^{1-\frac{1}{\beta p}}T)$(Vakili et al., 2021).

2.2 Problem Statement

We consider the problem of private contextual kernel bandits which consists of sequential interaction between a learning algorithm and nature. Under this setting, at the beginning of each time instant t, the learning algorithm $\mathscr{A}$, builds a query strategy $\mathcal{Q}_t$, which is a (possibly randomized) mapping from the context set $\mathcal{C}$ to the action set $\mathcal{X}$. During the time instant t, nature draws a context $c_t \in \mathcal{C}$, i.i.d. from a distribution κ, plays the action $x_t = \mathcal{Q}_t(c_t)$, generates a noisy reward $y_t = f(c_t, x_t) + \eta_t$, and reveals d_t, a possibly privatized function of (c_t, y_t), to the learning algorithm. Here, η_t denotes the zero-mean noise. The algorithm uses the collection $\mathcal{D}_t = \{d_s\}_{s=1}^t$ to design the query strategy $\mathcal{Q}_{t+1}$ for the next time instant.

We assume that both the action set $\mathcal{X} \subset \mathbb{R}^d$ and the context set $\mathcal{C} \subset \mathbb{R}^{d'}$ are finite sets with possibly very large cardinality that scales polynomially with T and exponentially with d and d'. The reward function f belongs to an RKHS $\mathcal{H}_k$ corresponding to a known kernel k defined over $\mathcal{W} := \mathcal{C}\times\mathcal{X}$ and satisfies $\|f\|_{\mathcal{H}_k} \le B$.

We make the following assumptions that are commonly adopted in the literature.

Assumption 2.4. The rewards $\{y_t\}_{t=1}^T$ are bounded in absolute value, i.e, $|y_t| < B, \forall t \le T$ for some $B > 0$.

Assumption 2.5. We have a context generator that can generate i.i.d. contexts according to the distribution κ.

Assumption 2.4 is commonly adopted in private bandit literature (Dubey and Pentland, 2020; Shariff and Sheffet, 2018; Han et al., 2021; Zheng et al., 2020). This ensures that an adversary cannot probe an unbounded reward as an input to the algorithm. Assumption 2.5 is a commonly adopted assumption in several studies (Gergely et al., 2024; Zierahn et al., 2023; Gergely and Olkhovskaya, 2020; Pavlovic et al., 2025). Existing studies on stochastic contextual bandits (Amani et al., 2023; Hanna et al., 2022, 2023) often assume complete knowledge of context distribution. Assumption 2.5 is much milder than having complete knowledge of the context distribution. For possible relaxations of assumption(2.5) please see Sec.(4).

2.3 Joint and Local Differential Privacy

As mentioned earlier, we consider two different notions of privacy in this work — Joint Differential Privacy (JDP) and Local Differential Privacy (LDP).

Joint differential privacy. This model of differential privacy was first introduced in Shariff and Sheffet (2018) and extends the classical notion of central differential privacy in offline setting to online setting. We first define the concept of t-neighbouring datasets and then use it to formally define JDP.

Definition 2.6. Two datasets $\mathcal{D}_T = \{d_s\}_{s=1}^T, \mathcal{D}'_T = \{d'_s\}_{s=1}^T$ are said to be t-neighbors if $d_t \ne d'_t$ and $d_s = d'_s$ for all $s = \{1, 2, \dots, T\} \setminus \{t\}$.

Definition 2.7. Let $\{x_t\}_{t=1}^T$ and $\{x'_t\}_{t=1}^T$ respectively denote the actions taken by a randomized contextual bandit algorithm $\mathscr{A}$ based on datasets $\mathcal{D}_T = \{d_s\}_{s=1}^T$ and $\mathcal{D}'_T = \{d'_s\}_{s=1}^T$. The algorithm $\mathscr{A}$ is said to be $(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}})$-joint differentially private (JDP) if the relation

$$\Pr(\{x_s\}_{s=t+1}^T = \mathcal{B}) \le \exp(\varepsilon_{\mathrm{DP}})\Pr(\{x'_s\}_{s=t+1}^T = \mathcal{B}) + \\ + \delta_{\mathrm{DP}}$$

holds for all pairs of t-neighboring datasets $\mathcal{D}_T, \mathcal{D}'_T$, all subsets $\mathcal{B} \subset \mathcal{X}^{T-t}$ and all $t \in \{1, 2\dots, T\}$. Here, the probability is taken over the random coins generated by the algorithm.

Local Differential Privacy. The notion of LDP for online learning problem was introduced by Han et al. (2021) for the problem of linear bandits and is akin to its counterpart in the offline setting (Kairouz et al., 2015).

Definition 2.8. (He and Zhang, 2022) A randomized contextual bandit algorithm $\mathscr{A}$ is said to be $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$ locally differentially private if the dataset $\mathcal{D}_T = \{d_s\}_{s=1}^T$ satisfies the following relations for all $t \in \{1, 2, \ldots, T\}$:

- d_t is conditionally independent of $\{(c_s, y_s)\}_{s=1}^{t-1}$ given $\{d_s\}_{s=1}^{t-1}$ and (c_t, y_t)
- For all $u, v, w \in \mathcal{C} \times \mathbb{R}$:

$$\Pr(d_t = w \mid \{d_s\}_{s=1}^{t-1}, (c_t, y_t) = u) \leq \delta_{\text{DP}} + \\ + \exp(\varepsilon_{\text{DP}}) \cdot \Pr(d_t = w \mid \{d_s\}_{s=1}^{t-1}, (c_t, y_t) = v)$$

  As before, the probability is taken over the random coins generated by the algorithm.

LDP ensures that data (c_t, y_t) is privatized immediately (as d_t) before being uploaded and used by the algorithm $\mathscr{A}$. In other words, LDP enforces the privatization of data at the *time it is generated*. In contrast, JDP relaxes this assumption and allows data be privatized at a *time of its use* by the algorithm $\mathscr{A}$.

3 ALGORITHM DESCRIPTION

In this section, we describe our proposed algorithm, CAPRI, (***C**ovariance **A**pproximation with **P**rojected **R**egress**I**on*), in both JDP and LDP settings. We first describe the construction of our novel private posterior mean estimator, which is a key component in our algorithm, followed by a detailed description of the algorithm.

Privatizing the Posterior Mean As mentioned earlier, the posterior mean based on GP model is a powerful predictor of the function values and is widely used in kernel bandit algorithms to design the query strategy $\mathcal{Q}_t$. Despite its high predictive power and ubiquitous use in the non-private setting (Chowdhury and Gopalan, 2017; Srinivas et al., 2010; Li and Scarlett, 2022), the use of posterior mean as outlined in Eqn. (2), presents several layers of challenges in the private setting. Firstly, if μ_t is used to design $\mathcal{Q}_{t+1}$, the query strategy at $t+1$, then such a choice results in highly intricate and non-linear relationship between the actions and the observations across different time instants. This cascading effect can make the actions taken by an algorithm highly sensitive to the past ones and consequently challenging to privatize. Secondly, even in the absence of such an intricate relationship between the actions and the observations, the sensitivity of the posterior mean is given by

$$\begin{aligned} &\sup_w |\mu_t(w) - \mu_t'(w)| = \\ &= \sup_w |\mu_t(w) - f(w) + f(w) - \mu_t'(w)| \\ &\approx \sup_{w \in \mathcal{W}} \sigma_t(w) + \sup_{w' \in \mathcal{W}} \sigma_t'(w). \end{aligned}$$

Here μ_t, σ_t denote the posterior mean and standard deviation corresponding to dataset $\mathcal{D}_t$ and μ_t', σ_t' denote the respective values corresponding to $\mathcal{D}_t'$, a t-neighbour of $\mathcal{D}_t$. Thus, the sensitivity implicitly depends on the rest of the dataset and in the worst-case can be as large as $\Omega(1)$. Lastly, if one were to privatize μ by separately privatizing $\{\mu(w)\}_{w \in \mathcal{W}}$, the privacy budget for each point would need to shrink by a factor of $|\mathcal{W}|$ to maintain the overall privacy guarantees, thereby resulting in trivial utility guarantees. Thus, it is essential to privatize μ directly as an element of RKHS. As discussed earlier, the infinite-dimensional nature of μ makes it particularly challenging.

In this work, we develop a novel, privatized version of the posterior mean by systematically addressing each of the three challenges outlined above. Inspired by Li and Scarlett (2022); Salgia et al. (2023a), we also adopt a non-adaptive query strategy, where the actions do not depend on past observations. This allows us to alleviate the first concern. To address the second challenge, we use a modified version of the covariance matrix to build our estimator. In particular, given a dataset $\mathcal{D}_t = (\mathbf{W}_t, \mathbf{y}_t)$, we use the estimator

$$\tilde{\mu}_t = (\Phi_{\mathcal{R}_t} \Phi_{\mathcal{R}_t}^\top + \tau \mathbf{Id})^{-1} \Phi_{\mathbf{W}_t} \mathbf{y}_t,$$

where $\mathcal{R}_t$ is an i.i.d. copy of the original dataset $\mathbf{W}_t$. The use of a statistically identical copy to estimate the mean offers two advantages. Firstly, it reduces the dependence on the estimator on the dataset. Secondly, the sensitivity of the above estimator depends on the variance corresponding to the points in $\mathcal{R}_t$ and thus is independent of the original data. More importantly, the statistical similarity between $\mathcal{R}_t$ and $\mathbf{W}_t$ ensures that $\sigma_{\mathcal{R}_t} \approx \sigma_{\mathbf{W}_t}$. This implies that such an estimator has low sensitivity whenever the original dataset has high predictive power — a scenario that constitutes a typical use case in our algorithm.

A similar technique was also employed in Pavlovic et al. (2025). Lastly, to privatize μ as an RKHS function, which is infinite-dimensional, we note that the posterior mean (as an element of the RKHS) can also be written as

$$\mu_t = \sum_{i=1}^t \alpha_s \phi(w_i),$$

where α_i is the ith coordinate of the vector $\boldsymbol{\alpha} = (\mathbf{K}_{\mathbf{W}_T,\mathbf{W}_T} + \tau \mathbf{I}_T)^{-1} \mathbf{y}_t$. Since $\{\phi(w_i)\}_{i=1}^t$ are linearly independent, they form a basis of the subspace $\mathcal{H}_k^t := \text{span}\{\phi(w_1), \phi(w_2), \ldots, \phi(w_t)\}$. Thus, μ_t can be represented by a t-dimensional vector $\boldsymbol{\alpha}$. If the basis were data-independent, then μ_t can be privatized by adding noise to $\boldsymbol{\alpha}$. However, we cannot directly use this idea to privatize μ_t since our basis is also data-dependent. To resolve this bottleneck, we approximate μ by projecting it onto $\overline{\mathcal{H}}_k^t = \text{span}\{\phi(s_1), \phi(s_2), \ldots, \phi(s_t)\}$, where $\mathcal{S}_t = \{s_1, s_2, \ldots, s_t\}$ is a random set generated independently from the private dataset . Since the basis elements are now data-independent, we can privatize μ by simply privatizing the coefficients. We set $\mathcal{S}_t$ to be a private i.i.d. copy of $\mathbf{W}_t$. Such a choice naturally satisfies the independence constraint and also offers a faithful approximation to μ_t, which is essential to ensure high utility for our estimator.

Based on a combination of all the approaches described above, our proposed private estimator corresponding to the dataset $\mathcal{D}_t = (\mathbf{W}_t, \mathbf{y}_t)$ is given as:

$$\underline{\widehat{\mu}}_t(\cdot) := \mathbf{k}_{\mathcal{S}_t}^\top(\cdot) \left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau \mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2} \cdot \tag{4}$$
$$\cdot \left(\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau \mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2} \mathbf{K}_{\mathcal{S}_r,\mathbf{W}_t}\mathbf{y}_t + \mathbf{Z}_{\text{priv}}\right)$$

where $\mathbf{Z}_{\text{priv}}$ denotes the noise used to privatize the coefficients α_i's. Under JDP, $\mathbf{Z}_{\text{priv}} \sim \mathcal{N}(0, \sigma_0^2 \cdot \mathbf{I}_t)$ and under LDP $\mathbf{Z}_{\text{priv}} \sim \mathcal{N}(0, t\sigma_0^2 \cdot \mathbf{I}_t)$, where $\mathcal{N}(\cdot)$ is the gaussian measure and

$$\sigma_0 := \widetilde{\sigma}_{\max} \cdot \frac{4B \log T}{\varepsilon_{\text{DP}}} \cdot \sqrt{\log\left(\frac{1.25 \log T}{\delta_{\text{DP}}}\right)}, \tag{5}$$

In the above definition,

$$\tau \widetilde{\sigma}_{\max}^2 := \max_{w \in \text{supp}(\rho)} k(w,w) - k_{\mathcal{S}}^\top(w) \mathbf{V} k_{\mathcal{S}}(w), \text{ where} \tag{6}$$

$$\mathbf{V} = \mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}} \left(\tau \mathbf{I} + \mathbf{K}_{\mathcal{R},\mathcal{S}}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}\right)^{-1} \mathbf{K}_{\mathcal{R},\mathcal{S}}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}$$

denotes the maximum posterior variance of points $\mathcal{R}$ projected onto the set $\mathcal{S}$ over the support of the sampling measure ϱ that generates $\mathbf{W}_t$. For clarity of notation, we use $\underline{\widehat{\mu}}_t$ and $\underline{\widehat{\mu}}_{t,\text{LDP}}$ to denote the estimators under JDP and LDP setting. The following lemma characterizes the privacy and predictive performance of our proposed estimators.

Lemma 3.1. *Let $\mathcal{D}_t = \{\mathbf{W}_t, \mathbf{y}_t\}$ be a dataset consisting of t points $\mathbf{W}_t$ and their corresponding noisy observations $\mathbf{y}_t$ from the underlying reward function f, such that $\mathbf{W}_t$ is independent of the noise vector corresponding to $\mathbf{y}_t$. Let $\mathcal{R}_t$ and $\mathcal{S}_t$ be two i.i.d. copies of $\mathbf{W}_t$ that are independent of both $\mathbf{W}_t$ and $\mathbf{y}_t$. If μ and μ_{LDP} denote the estimators constructed using $\mathcal{D}_t, \mathcal{R}_t$ and $\mathcal{S}_t$ as outlined in Eqn. (4), then μ is $(\varepsilon_{DP}/\log T, \delta_{DP}/\log T)$-JDP and μ_{LDP} is $(\varepsilon_{DP}, \delta_{DP})$-LDP w.r.t. to the dataset $\mathcal{D}_t$. Moreover, the relations*

$$\max_{w \in \text{supp}\varrho} |f(w) - \underline{\widehat{\mu}}_t(w)| \leq \widetilde{\mathcal{O}}\left(\frac{\widetilde{\sigma}_{\max}^2}{\varepsilon_{DP}} + \widetilde{\sigma}_{\max}\right)$$
$$\max_{w \in \text{supp}\varrho} |f(w) - \underline{\widehat{\mu}}_{t,LDP}(w)| \leq \widetilde{\mathcal{O}}\left(\frac{\sqrt{T}\widetilde{\sigma}_{\max}^2}{\varepsilon_{DP}} + \widetilde{\sigma}_{\max}\right)$$

hold with probability $1-\delta$ over the randomness of noise sequence, where $\widetilde{\sigma}_{\max}$ is as defined in Eqn. (6)

Notably, in the proof of Theorem 4.1 we show that $\max_{w \in \text{supp}\varrho} \widetilde{\sigma}_t^2(w)$ can be bounded as $\mathcal{O}(\frac{\gamma_T}{T})$. This implies that our error bound scales as $\mathcal{O}\left(\sqrt{\frac{\gamma_T}{T}}\varepsilon_{\text{DP}}\right)$ in the JDP setting and $\mathcal{O}\left(\frac{\gamma_T}{\sqrt{T}}\varepsilon_{\text{DP}}\right)$ in the LDP setting.

3.1 The CAPRI algorithm

Algorithm 1 CAPRI

Input: Error probability δ_{ERR}, Privacy parameters $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$,
$r \leftarrow 1, g \leftarrow 0, t_{\text{prev}} \leftarrow 0, \mathcal{S}, \mathcal{R} \leftarrow \emptyset, T_1 \leftarrow \lceil \sqrt{T} \rceil$
$\mathcal{X}_1(c) = \mathcal{X}\ \forall c \in \mathcal{C}$
for $s = 1, 2, \ldots, T_1$ **do**
  Sample $c, c' \sim \kappa$
  Sample $x \sim \text{Unif}(\mathcal{X}_1(c)), x' \sim \text{Unif}(\mathcal{X}_1(c'))$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{(c,x)\}, \mathcal{R} \leftarrow \mathcal{R} \cup \{(c',x')\}$
end for
Compute $\sigma_{\max}$ using Eqn. (6)
for $t = 1, 2, \ldots, T$ **do**
  Receive context c_t
  Take the action $x_t \sim \text{Unif}(\mathcal{X}_r(c_t))$ and observe reward y_t
  $g^{(t)} \leftarrow y_t(\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}} + \tau \mathbf{K}_{\mathcal{S},\mathcal{S}})^{-1/2} k_{\mathcal{S}}((c_t, x_t))$
  $g^{(t)} \leftarrow g^{(t)} + \mathcal{N}(0, \sigma_0^2 \cdot \mathbf{I}_{T_r})$ // Only for LDP
  $g \leftarrow g + g^{(t)}$
  if $t - t_{\text{prev}} = T_r$ **then**
    $g \leftarrow g + \mathcal{N}(0, \sigma_0^2 \cdot \mathbf{I}_{T_r})$ // Only for JDP
    $\widehat{\mu}_r(\cdot) := k_{\mathcal{S}}^\top(\cdot)(\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}} + \tau \mathbf{K}_{\mathcal{S},\mathcal{S}})^{-1/2} g$
    Obtain $\mathcal{X}_{r+1}(c)$ for all $c \in \mathcal{C}$ using Eqn. (7)
    $\mathcal{S}, \mathcal{R} \leftarrow \emptyset, g \leftarrow 0, t_{\text{prev}} \leftarrow t, T_{r+1} \leftarrow 2T_r, r \leftarrow r+1$
    for $s = 1, 2, \ldots, T_r$ **do**
      Sample $c, c' \sim \kappa$
      Sample $x \sim \text{Unif}(\mathcal{X}_r(c)), x' \sim \text{Unif}(\mathcal{X}_r(c'))$
      $\mathcal{S} \leftarrow \mathcal{S} \cup \{(c,x)\}, \mathcal{R} \leftarrow \mathcal{R} \cup \{(c',x')\}$
    end for
    Compute $\widetilde{\sigma}_{\max}$ using Eqn. (6)
  end if
end for

The CAPRI algorithm is based on the *explore-and-eliminate* philosophy. It operates in epochs that pro-

gressively double in length. For each epoch r, CAPRI maintains the set of active actions, $\mathcal{X}_r(c) \subseteq \mathcal{X}$, corresponding to each context $c \in \mathcal{C}$. These sets are initialized to $\mathcal{X}_1(c) = \mathcal{X}$ for all $c \in \mathcal{C}$. At the beginning of each epoch r, the algorithm has access to two sets $\mathcal{R}_r$ and $\mathcal{S}_r$ consisting of T_r context-action pairs, where T_r is the length of the rth epoch. During the rth epoch and time instant t, the algorithm observes the context c_t and samples x_t uniformly from the set $\mathcal{X}_r(c_t)$. It is straightforward to note that the points generated during each epoch are independent and identically distributed. It uses $\mathcal{R}_r$ and $\mathcal{S}_r$ to construct $\underline{\widehat{\mu}}_r$ as outlined in Eqn. (4) and update the action sets for the $r+1$-th epoch:

$$\mathcal{X}_{r+1}(c) = \mathcal{X}_r(c) \cap \left\{ \underline{\widehat{\mu}}_r(c,x) \geq \max_{x \in \mathcal{X}_r(c)} \underline{\widehat{\mu}}_r(c,x) - 4\Delta_r \right\}$$

$$\Delta_r := \beta_1 \left(\varepsilon_{\text{DP}}, \delta_{\text{DP}}, \delta'\right) \widetilde{\sigma}^2_{\max} + \beta\left(\delta'\right) \sigma_{\max} \tag{7}$$

where $\beta_1 = \frac{\log T \sqrt{8B \log(\log T |\mathcal{W}|/\delta) \log(1.25 \log T/\delta_{\text{DP}})}}{\varepsilon_{\text{DP}}}$, $\beta_{1,\text{LDP}} = \sqrt{T_r}\beta_1$, $\delta' = \frac{\delta}{|\mathcal{W}| T \log T}$ and $\beta(\cdot)$ is the confidence parameter [2]. $\widetilde{\sigma}^2_{\max}$ is computed from eq.(6) with the maximum taken over all context-action pairs remaining in the rth epoch. At the end of the rth epoch, CAPRI constructs the collections $\mathcal{R}_{r+1}$ and $\mathcal{S}_{r+1}$, to be used in the next epoch. Both sets consist of i.i.d. T_{r+1} context-action pairs obtained by first drawing $c \sim \kappa$ from the context generator and then drawing the corresponding action x uniformly from the set $\mathcal{X}_{r+1}(c)$.

Computational complexity of CAPRI. As is commonly the case in kernel learning, the computational bottleneck of CAPRI is the inversion of the $|\mathcal{S}_r|$-dimensional Gramian $\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}} + \tau \mathbf{K}_{\mathcal{S},\mathcal{S}}$. In r^{th} epoch, the projection set $\mathcal{S}_r$ is of size T_r and hence the computational complexity of calculating the approximate posterior $\widehat{\mu}_r$ is $\mathcal{O}(T_r^3)$. CAPRI operates in batched regime, where the posterior needs to be update only once an epoch and hence the total computational complexity is $\mathcal{O}(T^3 \log T)$. This complexity matches the complexity of non-private kernel algorithms such as MPV (Li and Scarlett, 2022) and UCB (Srinivas et al., 2010) with infrequent posterior mean updates .

4 PERFRORMANCE ANALYSIS

The following theorem characterizes the regret and privacy guarantees of the CAPRI Algorithm. We refer the reader to the supplementary material for a detailed proof of the theorem.

Theorem 4.1. *Consider the problem of contextual kernelized bandits described in the Section 2 where Assumptions 2.3-2.5 hold. If CAPRI is run for a time horizon of T with $T \geq T_0$, and privacy parameters $\varepsilon_{DP} > 0$ and $\delta_{DP} \in (0,1)$ under JDP (resp. LDP), then*

- *CAPRI satisfies $(\varepsilon_{DP}, \delta_{DP})$-JDP (resp. LDP)*
- *The regret incurred by the algorithm satisfies:*

$$R_{T,\text{JDP}}(\mathscr{A}) \leq \widetilde{\mathcal{O}}\left(\sqrt{T\gamma_T} + \frac{\gamma_T \sqrt{\log(1/\delta_{DP})}}{\varepsilon_{DP}}\right)$$

$$R_{T,\text{LDP}}(\mathscr{A}) \leq \widetilde{\mathcal{O}}\left(\sqrt{T\gamma_T} + \frac{\gamma_T \sqrt{T\log(1/\delta_{DP})}}{\varepsilon_{DP}}\right)$$

with probability $1-\delta_{ERR}$ [3] over the randomness in the algorithm and the noise sequence. Here T_0 is a constant that depends only on the kernel k and the context distribution.

As shown in the above theorem, CAPRI guarantees $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$ differential privacy under both JDP and LDP. Moreover, it achieves a cumulative regret of $\widetilde{\mathcal{O}}\left(\sqrt{\gamma_T T} + \frac{\gamma_T}{\varepsilon_{\text{DP}}}\right)$ and $\widetilde{\mathcal{O}}\left(\sqrt{\gamma_T T} + \frac{\gamma_T \sqrt{T}}{\varepsilon_{\text{DP}}}\right)$ under JDP and LDP respectively. We would like to point out that for the special case of linear bandits ($\gamma_T = d$), our regret under JDP matches the lower bound derived in He and Zhang (2022) and under LDP matches the best known upper bound (Han et al., 2021). The above result also partly resolves the conjecture posed in Dubey (2021). Specifically, Dubey (2021) poses the question about the additional factor of $\mathcal{O}(T^{1/4})$ in the cumulative regret for the LDP setting when compared to the JDP, for adversarially generated contexts. As shown by the above theorem, this gap in performance between JDP and LDP settings is not present for stochastically generated contexts, unlike in the case of adversarially generated contexts. This difference in performance under stochastic and adversarial contextual settings was previously seen in both linear (He and Zhang, 2022) and generalized-linear bandits (Han et al., 2021).

Possible relaxations of assumption(2.5). The knowledge of the context distribution can be relaxed in 2 directions. Firstly, we can work with a misspecified context distribution while keeping the order of our regret bounds. Namely for a misspecified measure κ', for which $\forall c, \frac{\kappa'(c)}{\kappa(c)} \in (1-u, 1+u)$ for some $u = \mathcal{O}(\sqrt{\gamma_T/T})$, Theorem 4.1 still remains true.
The second approach is to allow for a context sampler but privatize the contexts drawn from this distribution. Through a more rigorous analysis similar to Calandriello et al. (2019), we can reduce the number

[2] Please refer to the supplementary material for a full expression

[3] The regret exhibits a logarithmic dependence on δ_{ERR}, for the exact expression please see Theorem(A.7) in the Supplemental Material.

of contexts used for projection to $|\mathcal{S}| = \mathcal{O}(\gamma_T)$. The remaining contexts, used to simulate the covariance operator, can be privatized by averaging. In more detail, suppose we are given $T\gamma_T$ contexts $\{c_i\}_{i=1}^{T\gamma_T}$ drawn from κ and the set of contexts used for projection $\mathcal{S}$, we can estimate the projected covariance operator as:

$$\widehat{\mathrm{Cov}} = \sum_{i=1}^{T\gamma_T} \frac{1}{\gamma_T} \mathcal{P}_{\mathcal{S}} \phi(c_i, \pi(c_i)) \otimes \phi^\top(c_i, \pi(c_i)) \mathcal{P}_{\mathcal{S}} + \lambda \mathbf{Id}$$

Notice that the sensitivity of $\widehat{\mathrm{Cov}}$ with respect to each c_i is simply $\mathcal{O}(1/\gamma_T)$ and thus the Gramian used in CAPRI can be privatized by adding a spectrally small[4] Gaussian matrix via the same methods as in Shariff and Sheffet (2018). Crucially, only the contexts used to construct the projection operator $\mathcal{P}_{\mathcal{S}}$ remain non-private, and this only requires $\mathcal{O}(\gamma_T)$ samples. We can thus keep the regret order in Theorem 4.1 while using only $\mathcal{O}(\gamma_T)$ "public" contexts .
The authors believe that the assumption 2.5 is not fundamental and could be entirely removed by developing privacy mechanisms that operate natively in the Hilbert space, without requiring finite-dimensional projections.

5 CONCLUSION

In this work, we study the problem of private contextual kernel bandits with stochastically generated contexts. We propose a novel private estimator for RKHS functions that is based on a careful combination of covariance approximation and random projection in the RKHS space. Our estimator simultaneously offers low sensitivity to the underlying dataset and order-optimal prediction accuracy. Building upon our novel estimator, we also proposed a new algorithm called CAPRI that achieves state-of-the-art cumulative regret performance across a wide range of kernels under both JDP and LDP setting. We conjecture that our proposed algorithm offer optimal regret performance under both JDP and LDP settings. Establishing the lower bounds on regret in private settings is an interesting future direction to explore.

6 ACKNOWLEDGEMENTS

The work of N. Pavlovic and Q. Zhao was supported by the National Science Foundation under the 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*, 2020.

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.

F. Filbir and H. N. Mhaskar. A quadrature formula for diffusion polynomials corresponding to a generalized heat kernel. *J. Fourier Anal. Appl., 16(5):629–657*, 2010.

N. Gergely and J. Olkhovskaya. Efficient and robust algorithms for adversarial linear contextual bandits. *Conference on Learning Theory, pp. 3049-3068. PMLR,*, 2020.

N. Gergely, J. Olkhovskaya, and S. Vakili. Adversarial contextual bandits go kernelized. *In International Conference on Algorithmic Learning Theory, pp. 907-929. PMLR*, 2024.

K. Grochenig. Sampling, marcinkiewicz-zygmund inequalities, approximation, and quadrature rules. *J. Approx. Theory, 257,*, 2020.

Y. Han, L. Zhipeng, W. Yang, and J. Zhang. Generalized linear bandits with local differential privacy. *Advances in Neural Information Processing Systems*, 2021.

[4] $\mathcal{O}(1)$ spectral norm

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.

J. Z. He, Jiahao and R. Q. Zhang. A reduction from linear contextual bandits lower bounds to estimations lower bounds. *In International Conference on Machine Learning, pp. 8660-8677. PMLR*, 2022.

P. Kairouz, S. Oh, and P. Viswanath. The composition theorem for differential privacy. In *Proceedings of the 32nd International Conference on Machine Learning, ICML*, pages 1376–1385. PMLR, 2015.

M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur. Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences, 2018.

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.

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.

M. Maggioni and H. N. Mhaskar. Diffusion polynomial frames on metric measure spaces. *Appl. Comput. Harmon. Anal., 24(3):329–353*, 2008.

H. N. Mhaskar, F. J. Narcowich, and J. D. Ward. Spherical marcinkiewicz-zygmund inequalities and positive quadrature. *Math. Comp., 70(235):1113–1130*, 2001.

N. Pavlovic, S. Salgia, and Q. Zhao. Differentially private kernelized contextual bandits. *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

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, 2023a.

S. Salgia, S. Vakili, and Q. Zhao. Collaborative learning in kernel-based bandits for distributed users. *IEEE Transactions on Signal Processing*, 71:3956–3967, 2023b.

S. Salgia, S. Vakili, and Q. Zhao. Random exploration in bayesian optimization: Order-optimal regret and computational efficiency. *Forty-first International Conference on Machine Learning*, 2024.

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.

M. Stanislav. On some extensions of bernstein's inequality for self-adjoint operators. *Statistics & Probability Letters*, 2017.

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, 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*, 2021.

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

M. J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint. Vol. 48.* Cambridge university press, 2019.

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.

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.

L. Zierahn, D. van der Hoeven, N. Cesa-Bianchi, and G. Neu. Nonstochastic contextual combinatorial bandits. *In International Conference on Artificial Intelligence and Statistics, pp. 8771-8813. PMLR,*, 2023.

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. All assumptions and the problem statement are outlined in the Sec.2 of our paper.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm.
   Yes. We state and prove the bound on the regret of CAPRI in Sec. 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 assumptions are stated in Sec. 2.
   (b) Complete proofs of all theoretical results.
   Yes. Both privacy and regret claims are rigorously shown in Sec A of the Supplemental Material.
   (c) Clear explanations of any assumptions.
   Yes. All assumptions are explained and motivated in Sec. 2 of our work.

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).
   Not Applicable
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen).
   Not Applicable
   (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).
   Not Applicable
   (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

Supplemental Material

A Appendix A. Performance and Privacy Guarantees

We first state several results frequently utilized in kernel-based optimization literature.

Lemma A.1. *Consider a measure ϱ supported on a $\mathcal{W} = \mathcal{C} \times \mathcal{X}$ defined in assumption 2.3 and let $\mathbf{Z} = T\mathbf{\Lambda} + \tau\mathbf{Id}$ where $\Lambda = \mathbb{E}_{w\sim\varrho}[\phi(w)\phi^\top(w)]$. For $T > \overline{T}$:*

$$\sup_{x\in \mathrm{supp}\varrho} \phi^\top(w)\mathbf{Z}^{-1}\phi(w) \leq \frac{54F^2}{13}\frac{\gamma_T}{T}$$

Where $\overline{T}$ is a measure ϱ and kernel k dependent constant defined in Salgia et al. (2023a).

The result above is a direct consequence of Lemma 3.4 in Salgia et al. (2023a). Assumption 2.3 ensures that the eigenfunctions are bounded over the support of ϱ , and thus results from Salgia et al. (2023a) can be applied.

Lemma A.2. *(Adapted after Calandriello et al. (2019)) Consider a covariance matrix $\Phi_\mathcal{S}\Phi_\mathcal{S}^\top + \tau\mathbf{Id}$, where $\mathcal{S} \subset \mathcal{W}$ and a linear operator $\mathbf{A} + \tau\mathbf{Id} : \mathcal{H}_k \to \mathcal{H}_k$. Suppose for some $\alpha \in (0, 1)$ we have $(1-\alpha)(\mathbf{A} + \tau\mathbf{Id}) \preceq \mathbf{\Phi}_\mathcal{S}\mathbf{\Phi}_\mathcal{S}^\top + \tau\mathbf{Id} \preceq (\mathbf{1}+\alpha)(\mathbf{A} + \tau\mathbf{Id})$. The following holds for the projected operator $\mathcal{P}_\mathcal{S}\mathbf{A}\mathcal{P}_\mathcal{S}$:*

$$\frac{1-\alpha}{1+\alpha}(\mathbf{A} + \tau\mathbf{Id}) \preceq \mathcal{P}_\mathcal{S}\mathbf{A}\mathcal{P}_\mathcal{S} + \tau\mathbf{Id} \preceq \frac{1+\alpha}{1-\alpha}(\mathbf{A} + \tau\mathbf{Id})$$

Here, the projection operator is defined as $\mathcal{P}_\mathcal{S} = \Phi_\mathcal{S}(\Phi_\mathcal{S}^\top\Phi_\mathcal{S})^{-1}\Phi_\mathcal{S}^\top$.

Proof. We can write:

$$\begin{aligned}\mathcal{P}_\mathcal{S}\mathbf{A}\mathcal{P}_\mathcal{S} &\succeq \frac{1}{1+\alpha}\mathcal{P}_\mathcal{S}(\mathbf{\Phi}_\mathcal{S}\mathbf{\Phi}_\mathcal{S}^\top - \alpha\tau\mathbf{Id})\mathcal{P}_\mathcal{S} = \frac{\mathbf{1}}{\mathbf{1}+\alpha}(\mathbf{\Phi}_\mathcal{S}\mathbf{\Phi}_\mathcal{S}^\top - \alpha\tau\mathcal{P}_\mathcal{S}) \succeq \\ &\succeq \frac{1}{1+\alpha}((1-\alpha)(\mathbf{A} + \tau\mathbf{Id}) - (\alpha+1)\tau\mathbf{Id}) = \frac{1-\alpha}{1+\alpha}(\mathbf{A} + \tau\mathbf{Id}) - \tau\mathbf{Id}\end{aligned}$$

Adding $\alpha\mathbf{Id}$ to both sides we have $\mathcal{P}_\mathcal{S}\mathbf{A}\mathcal{P}_\mathcal{S} + \tau\mathbf{Id} \succeq \frac{1-\alpha}{1+\alpha}(\mathbf{A} + \tau\mathbf{Id})$. Symmetric derivation holds for the right-hand side

□

Lemma A.3. *(Adapted after Calandriello et al. (2019)) Consider the projection operator $\mathcal{P}_\mathcal{S}$ induced by the columns of $\Phi_\mathcal{S}$ where $\mathcal{S} \subset \mathcal{W}$, and a linear operator $\mathbf{A} + \tau\mathbf{Id} : \mathcal{H}_k \to \mathcal{H}_k$. If for some $\alpha \in (0, 1)$ we have $(1-\alpha)(\mathbf{A} + \tau\mathbf{Id}) \preceq \Phi_\mathcal{S}\Phi_\mathcal{S}^\top + \tau\mathbf{Id} \preceq (1+\alpha)(\mathbf{A} + \tau\mathbf{Id})$, then the following identity holds:*

$$(\mathbf{Id} - \mathcal{P}_\mathcal{S}) \preceq \frac{\tau}{1-\alpha}(\mathbf{A} + \mathbf{Id})^{-1}$$

Proof.

$$\begin{aligned}&(\mathbf{Id} - \mathcal{P}_\mathcal{S}) \preceq \mathbf{Id} - \Phi_\mathcal{S}(\Phi_\mathcal{S}^\top\Phi_\mathcal{S} + \tau\mathbf{I}_\mathcal{S})^{-1}\Phi_\mathcal{S}^\top \preceq \mathbf{Id} - \Phi_\mathcal{S}\Phi_\mathcal{S}^\top(\Phi_\mathcal{S}\Phi_\mathcal{S}^\top + \tau\mathbf{Id})^{-1} \preceq \\ &\preceq \tau(\Phi_\mathcal{S}\Phi_\mathcal{S}^\top + \tau\mathbf{Id})^{-1} \preceq \frac{\tau}{1-\alpha}(\mathbf{A} + \tau\mathbf{Id})^{-1}\end{aligned}$$

The first inequality follows from $(\Phi_\mathcal{S}^\top\Phi_\mathcal{S} + \tau\mathbf{Id})^{-1} \preceq (\Phi_\mathcal{S}^\top\Phi_\mathcal{S})^{-1}$. □

Lemma A.4. *(Kernel trick matrix identity) Consider two matrices $\mathbf{A}, \mathbf{B}$ so that $\mathbf{AB} + \tau\mathbf{Id}$ is invertible. The following identity holds:*

$$\mathbf{A}(\mathbf{BA} + \tau\mathbf{Id})^{-1} = (\mathbf{AB} + \tau\mathbf{Id})^{-1}\mathbf{A}$$

Proof. Note that:

$$\mathbf{A}(\mathbf{BA} + \tau\mathbf{Id}) = (\mathbf{AB} + \tau\mathbf{Id})\mathbf{A}$$

$\mathbf{AB} + \tau\mathbf{Id}$ is invertible and as $\mathbf{AB}, \mathbf{BA}$ have the same spectrum, so is $\mathbf{BA} + \tau\mathbf{Id}$. Multiplying both sides of the equality by $(\mathbf{BA} + \tau\mathbf{Id})^{-1}$ from the right, and $(\mathbf{AB} + \tau\mathbf{Id})^{-1}$ from the left gives the desired identity. □

We next show that a randomly drawn covariance matrix is sufficiently close to the expected operator $\mathbf{Z} = T\mathbf{\Lambda} + \tau\mathbf{Id}$. This result will be essential in showing that projection onto an in-distribution "copy" of the data-set, keeps the same order of utility. We will also use Lemma(A.5) to show that that the covariance of the data-set can be estimated by an in distribution copy.

Lemma A.5. *Suppose $T > \max\left(\overline{T}, \left(\frac{180\log(14T/\delta)}{F^2}\right)^{\beta_p/\beta_p - 1}\right)$ points $\mathcal{R} = \{r_1, r_2, \dots r_T\}$ are sampled i.i.d from $\mathcal{W} = \mathcal{C} \times \mathcal{X}$ according to the measure ϱ, satisfying assumption 2.3. Consider the operator $\mathbf{Z} = \tau\mathbf{Id} + T\mathbf{\Lambda}$ where $\Lambda = \mathbb{E}_{x\sim\varrho}[\phi(w)\phi^\top(w)]$ and let $\mathbf{R} = \sum_{i=1}^T \phi(r_i)\phi(r_i)^\top + \tau\mathbf{Id}$. We claim with probability at least $1-\delta$:*

$$\left\|\mathbf{Z}^{-1/2}\mathbf{R}\mathbf{Z}^{-1/2} - \mathbf{Id}\right\|_2 \leq 2\sqrt{C_0\log(14T/\delta)} + \frac{4}{3}\log(14T/\delta)C_0$$

Where $C_0 = \sup_{w\in\text{supp}\rho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)$ is the maximum variance of the expected covariance operator over the support of the measure ϱ.

Proof. We will use Bernstein concentration inequality for self-adjoint Hilbert Space operators. We start by re-writing the expression as:

$$\begin{aligned}\mathbf{Z}^{-1/2}\mathbf{R}\mathbf{Z}^{-1/2} - \mathbf{Id} &= \sum_{i=1}^T \left(\mathbf{Z}^{-1/2}\phi(r_i)\phi^\top(r_i)\mathbf{Z}^{-1/2} - \mathbf{Z}^{-1/2}\mathbf{\Lambda}\mathbf{Z}^{-1/2}\right) = \\ &= \sum_{i=1}^T \left(V_i - \mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2}\right)\end{aligned}$$

First note that $\mathbb{E}\left[V_i - \mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2}\right] = 0$. We can bound the spectrum as:

$$U = \sup_{r_i}\|V_i - \mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2}\|_2 \leq \max(\sup_{r_i}\|V_i\|_2, \|\mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2}\|_2) \leq \max\left(\sup_r \phi^\top(r)\mathbf{Z}^{-1}\phi(r), \frac{1}{T}\right)$$

Where the first inequality follows as both matrices as symmetric positive semi-definite. Introduce the shorthand $C_0 = \sup_r \phi^\top(r)\mathbf{Z}^{-1}\phi(r)$. We next bound the sum of variances:

$$\begin{aligned}\sigma^2 &= \left\|\sum_{i=1}^T \mathbb{E}\left[(V_i - \mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2})^2\right]\right\|_2 = \\ &= \left\|\sum_{i=1}^T \mathbb{E}\left[V_i^2\right] - (\mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2})^2\right\|_2 \leq T\left\|\mathbb{E}_{r\sim\varrho}\left[\mathbf{Z}^{-1/2}\phi(r)\phi^\top(r)\mathbf{Z}^{-1}\phi(r)\phi^\top(r)\mathbf{Z}^{-1/2}\right]\right\|_2 \leq \\ &\leq TC_0\left\|\mathbb{E}_{r\sim\varrho}\left[\mathbf{Z}^{-1/2}\phi(r)\phi^\top(r)\mathbf{Z}^{-1/2}\right]\right\|_2 = TC_0\left\|\mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2}\right\|_2 \leq C_0\end{aligned}$$

The first inequality follows as V_i's are iid and the operators $\mathbb{E}\left[\left(V_i - \mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2}\right)^2\right]$ are positive semi-definite, while the second is simply the result of upper bounding $\phi^\top(r)\mathbf{Z}^{-1}\phi(r) \leq C_0$.

The final inequality follows as $\mathbf{Z} \succeq T\Lambda$. We now apply operator Bernstein inequality(Stanislav, 2017; Calandriello et al., 2019):

$$P\left(\left\|\sum_{i=1}^{T}(V_i - \mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2})\right\|_2 \geq t\right) \leq 14T\exp\left(-\frac{t^2/2}{\sigma^2 + tU/3}\right)$$

Choosing $t = 2\sqrt{C_0\log(14T/\delta)} + \frac{4}{3}\log(14T/\delta)\max\left(C_0, \frac{1}{T}\right)$ we have w.p at least $1-\delta$:

$$\left\|\sum_{i=1}^{T}(V_i - \mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2})\right\|_2 \leq 2\sqrt{C_0\log(14T/\delta)} + \frac{4}{3}\log(14T/\delta)\max\left(C_0, \frac{1}{T}\right)$$

By using Lemma A.1 and the assumed bound on T this expression can be simplified as:

$$\left\|\sum_{i=1}^{T}(V_i - \mathbf{Z}^{-1/2}\Lambda\mathbf{Z}^{-1/2})\right\|_2 \leq 2\sqrt{C_0\log(14T/\delta)} + \frac{4}{3}\log(14T/\delta)C_0$$

□

We note that the threshold $T > \max\left(\overline{T}, \left(\frac{180\log(14T/\delta)}{F^2}\right)^{\beta/\beta-1}\right)$ was chosen so that $2\sqrt{C_0\log(14T/\delta)} > \frac{4}{3}\log(14T/\delta)C_0$ and $2\sqrt{C_0\log(14T/\delta)} + \frac{4}{3}\log(14T/\delta)C_0 < \frac{1}{9}$. This was namely done for the ease of presentation and the only key requirement here is that $T > \overline{T}$.

Next, we introduce a novel estimator that builds upon the posterior mean by both estimating the covariance matrix and projecting the data set onto randomly sampled feature directions. We show that this estimator, despite having a significantly weaker dependence on the data-set, still keeps the order of approximation error of the posterior-mean.

Lemma A.6. *Let $\mathcal{D}_t = \{\mathbf{W}_t, \mathbf{y}_t\}$ be a dataset of consisting of t points $\mathbf{W}_t$ generated via the measure ϱ and their corresponding noisy observations $\mathbf{y}_t$ from the underlying reward function f, such that $\mathbf{W}_t$ is independent of the noise vector corresponding to $\mathbf{y}_t$. Let $\mathcal{R}_t$ and $\mathcal{S}_t$ be two i.i.d. copies of $\mathbf{W}_t$ that are independent of both $\mathbf{W}_t$ and $\mathbf{y}_t$. If μ and μ_{LDP} denote the estimators constructed using $\mathcal{D}_t, \mathcal{R}_t$ and $\mathcal{S}_t$ as :*

$$\underline{\widehat{\mu}}_t(\cdot) := \mathbf{k}_{\mathcal{S}_t}^\top(\cdot)\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}\cdot\left(\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}\mathbf{K}_{\mathcal{S}_r,\mathbf{W}_t}\mathbf{y}_t + \mathbf{Z}_{\text{priv}}\right)$$

$$\underline{\widehat{\mu}}_{t,LDP}(\cdot) := \mathbf{k}_{\mathcal{S}_t}^\top(\cdot)\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}\cdot\left(\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}\mathbf{K}_{\mathcal{S}_r,\mathbf{W}_t}\mathbf{y}_t + \mathbf{Z}_{\text{priv,LDP}}\right)$$

where $\mathbf{Z}_{\text{priv}} \sim \mathcal{N}(0, \sigma_0^2\cdot\mathbf{I}_t)$ and under LDP $\mathbf{Z}_{\text{priv,LDP}} \sim \mathcal{N}(0, t\sigma_0^2\cdot\mathbf{I}_t)$,

$$\sigma_0 := \widetilde{\sigma}_{\max}\cdot\frac{4B\log T}{\varepsilon_{DP}}\cdot\sqrt{\log\left(\frac{1.25\log T}{\delta_{DP}}\right)}, \tag{8}$$

and $\widetilde{\sigma}_{\max} = \max_{w\in\text{supp}\varrho}\phi^\top(w)\left(\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}^\top\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{I}\right)^{-1}\phi(w)$ then μ is $(\varepsilon_{DP}/\log T, \delta_{DP}/\log T)$-JDP and μ_{LDP} is $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$-LDP w.r.t.the dataset $\mathcal{D}_t$. Moreover w.p. at least $1-\delta$ the following relations hold:

$$\max_{w\in\text{supp}\varrho}|f(w) - \underline{\widehat{\mu}}_t(w)| \leq \beta\left(\frac{\delta}{|\mathcal{W}|}\right)\widetilde{\sigma}_{\max} + \beta_1\left(\varepsilon_{\text{DP}}, \delta_{\text{DP}}, \frac{\delta}{|\mathcal{W}|}\right)\widetilde{\sigma}_{\max}^2$$

$$\max_{w\in\text{supp}\varrho}|f(w) - \underline{\widehat{\mu}}_{t,LDP}(w)| \leq \beta\left(\frac{\delta}{|\mathcal{W}|}\right)\widetilde{\sigma}_{\max} + \beta_{1,\text{LDP}}\left(\varepsilon_{\text{DP}}, \delta_{\text{DP}}, \frac{\delta}{|\mathcal{W}|}\right)\widetilde{\sigma}_{\max}^2$$

where $\beta(\delta) = \left(90B\sqrt{\log\left(\frac{168T}{\delta}\right)} + \frac{52B\sqrt{\log\left(\frac{168T}{\delta}\right)\log\left(\frac{12}{\delta}\right)}}{\sqrt{\tau}} + 3B\sqrt{2\log\left(\frac{6}{\delta}\right)} + \sqrt{24\tau}\right)$, $\beta_1(\varepsilon_{\text{DP}}, \delta_{\text{DP}}, \delta) = \frac{8B\log T}{\varepsilon_{DP}}\log\left(\frac{3}{\delta}\right)\sqrt{\log\left(\frac{1.25\log T}{\delta_{DP}}\right)}$ and $\beta_{1,\text{LDP}} = \sqrt{t}\beta_{1,\text{LDP}}$.

Proof. Note that by assumption 2.4 noise instances $\{\eta_i\}_{i=1}^T$ are in $[-B, B]$ and thus η_i is B^2-sub Gaussian random variable (Wainwright, 2019). We will divide the proof in several steps. Note that the analysis for LDP and JDP is entirely the same, the only exception being the scaling of the error function due to the extra privatization noise $\mathbf{Z}_{\text{priv,LDP}}$. Having this in mind, we will present the derivation for the JDP estimator $\underline{\widehat{\mu}}_t(\cdot)$ and outline the steps for $\underline{\widehat{\mu}}_{t,\text{LDP}}(\cdot)$

The total approximation error can be written as:

$$|f(w) - \underline{\widehat{\mu}}_t(w)| \leq \underbrace{|f(w) - \widetilde{\mu}_t(w)|}_{E_1} + \underbrace{|\widetilde{\mu}_t(w) - \widehat{\mu}_t(w)|}_{E_2} + \underbrace{|\underline{\widehat{\mu}}_t - \widehat{\mu}_t(w)|}_{E_3}$$

Where $\widetilde{\mu}_t, \widehat{\mu}_t$ are "intermediate" estimators to $\underline{\widehat{\mu}}_t$:

$$\widetilde{\mu}_t(\cdot) = \phi^\top(\cdot)(\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\mathcal{P}_\mathcal{S} + \tau\mathbf{Id})^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\mathbf{y}_t \tag{9}$$

$$\widehat{\mu}_t(\cdot) = \phi^\top(\cdot)(\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S} + \tau\mathbf{Id})^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\mathbf{y}_t \tag{10}$$

We will separately bound E_1, E_2, E_3 and then finally prove the privacy claim in the end. For the ease of presentation, We drop the t subscript from the in-distribution copies of the data-set, $\mathcal{S}_t, \mathcal{R}_t$.

Step 1: Bounding E_1.

In bounding E_1 we closely follow the derivation in Salgia et al. (2023b) with appropriate changes to account for different approach to exploration and the design of the projection set. We first separate the error in the noise and the approximation component.

For notational convenience introduce the short-hand $\widehat{\mathbf{Z}}_\mathcal{S} = \sum_{i=1}^t \mathcal{P}_\mathcal{S}\phi(w_i)\phi^\top(w_i)\mathcal{P}_\mathcal{S} + \tau\mathbf{Id}$.

$$|f(w) - \phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W_t}}y_T| \leq |\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}f| + |\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\eta_{1:t}|$$

We further bound the first term as :

$$|\phi^\top(w)(\mathbf{Id} - \widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top)f| \leq \|\phi^\top(w)(\mathbf{Id} - \widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top)\|_{\mathcal{H}_k} \tag{11}$$

$$\leq \|\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}(\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\mathcal{P}_\mathcal{S} + \tau\mathbf{Id} - \mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W_t}}^\top)\| \leq \tag{12}$$

$$\leq \tau\|\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\| + \|\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top(\mathcal{P}_\mathcal{S} - \mathbf{Id})\| \leq \tag{13}$$

$$\leq \tau\|\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\| + \sqrt{\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top(\mathbf{Id} - \mathcal{P}_\mathcal{S})\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\mathcal{P}_\mathcal{S}\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\phi(w)} \tag{14}$$

From Lemma A.5 it follows with probability at least $1 - \delta/2$ that $\max(\|\mathbf{Z}^{-1/2}(\Phi_\mathcal{S}\Phi_\mathcal{S}^\top + \tau\mathbf{Id})\mathbf{Z}^{-1/2} - \mathbf{Id}\|_2, \|\mathbf{Z}^{-1/2}\widehat{\mathbf{Z}}\mathbf{Z}^{-1/2} - \mathbf{Id}\|) \leq \frac{1}{9}$. This implies that all of the eigen-values of the two operators $\mathbf{Z}^{-1/2}(\Phi_\mathcal{S}\Phi_\mathcal{S}^\top + \tau\mathbf{Id})\mathbf{Z}^{-1/2}, \mathbf{Z}^{-1/2}\widehat{\mathbf{Z}}\mathbf{Z}^{-1/2}$ are in range $[\frac{8}{9}, \frac{10}{9}]$. By multiplying both operators with $\mathbf{Z}^{1/2}$ from left and right we can write:

$$\frac{6}{8}\widehat{\mathbf{Z}} \preceq \frac{8}{10}\widehat{\mathbf{Z}} \preceq \frac{8}{9}\mathbf{Z} \preceq \Phi_\mathcal{S}\Phi_\mathcal{S}^\top + \tau\mathbf{Id} \preceq \frac{10}{9}\mathbf{Z} \preceq \frac{10}{8}\widehat{\mathbf{Z}}$$

By Lemma A.3 applied on operators $\widehat{\mathbf{Z}}, \Phi_\mathcal{S}\Phi_\mathcal{S}^\top + \tau\mathbf{Id}$ we have $\mathbf{Id} - \mathcal{P}_\mathcal{S} \preceq \frac{4}{3}\widehat{\mathbf{Z}}^{-1}$. By Lemma A.2 , $\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\phi(w) \leq \frac{5}{3}\phi^\top(w)\widehat{\mathbf{Z}}^{-1}\phi(w) \leq \frac{50}{27}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)$. Now applying the same derivation as in Salgia et al. (2023b) we have:

$$|\phi^\top(w)(\mathbf{Id} - \widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top)f| \leq \sqrt{\frac{8\tau}{3}}\sqrt{\phi^\top(w)\mathbf{Z}_\mathcal{S}^{-1}\phi^\top(w)} \leq \sqrt{\frac{8\tau}{3}}\sqrt{\frac{50}{27}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)}$$

To bound the noise component, it suffices to bound the $\|\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\|_2$. Vector $\{\eta_{1:t}\}$ is B^2-sub Gaussian, and the rest of the derivation will follow from Chernoff-Hoeffding inequality (Wainwright, 2019).
Note that by identical argument as in Salgia et al. (2023b), we have:

$$\|\phi^\top(w)\widehat{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_t}\|_2 \leq \sqrt{\phi^\top(w)\mathbf{Z}_\mathcal{S}^{-1}\phi(w)} \leq \sqrt{\frac{50}{27}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)}$$

Thus w.p at least $1-\delta/2$

$$\phi^\top(w)\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\eta_{1:t} \le B\sqrt{2\log(2/\delta)\frac{50}{27}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)}$$

Finally w.p. probability at least $1-\delta$:

$$E_1 \le \sqrt{\frac{50}{27}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)}\left(B\sqrt{2\log(2/\delta)}+\sqrt{\frac{8\tau}{3}}\right)$$

Step 2: Bounding E_2.

For notational convenience let $\widetilde{\mathbf{Z}} = \Phi_{\mathcal{R}}\Phi_{\mathcal{R}}^\top + \tau\mathbf{Id}$ and in general add a subscript $\mathcal{S}$ to denote a space $\Phi_{\mathcal{S}}$ projected operator. Specifically let $\mathbf{Z}_{\mathcal{S}} = \mathcal{P}_{\mathcal{S}}T\mathbf{\Lambda}\mathcal{P}_{\mathcal{S}} + \tau\mathbf{Id}$ and $\widetilde{\mathbf{Z}}_{\mathcal{S}} = \mathcal{P}_{\mathcal{S}}\Phi_{\mathcal{R}}\Phi_{\mathcal{R}}^\top\mathcal{P}_{\mathcal{S}} + \tau\mathbf{Id}$.

$$|\widetilde{\mu}_t(w) - \widehat{\mu}_t(w)| \le |\phi^\top(w)\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\mathbf{y}_t - \phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\mathbf{y}_t| \le \tag{15}$$

$$\le |\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} - \widehat{\mathbf{Z}}_{\mathcal{S}})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\mathbf{y}_t| + |\phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}} - \mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\mathbf{y}_t| \tag{16}$$

We will bound each of the two summands in Eq.(15) separately. By noting that $y_t = \phi^\top(w_t)f + \eta_t$, we can separate the approximation and the noise related error:

$$|\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} - \widehat{\mathbf{Z}}_{\mathcal{S}})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}y_t| \le \tag{17}$$

$$\le |\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} - \widehat{\mathbf{Z}}_{\mathcal{S}})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f| + |\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} - \widehat{\mathbf{Z}}_{\mathcal{S}})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\eta_{1:t}| \tag{18}$$

After some algebraic manipulation on the first term we arrive to:

$$|\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} - \widehat{\mathbf{Z}}_{\mathcal{S}})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f| = \tag{19}$$

$$= |\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}(\mathbf{Z} - \widehat{\mathbf{Z}})P_{\mathcal{S}}\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f| \le \tag{20}$$

$$\le \|\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}(\mathbf{Z} - \widehat{\mathbf{Z}})P_{\mathcal{S}}\|_2\|\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f\|_2 = \tag{21}$$

$$= \|P_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\mathbf{Z}(\mathbf{Id} - \mathbf{Z}^{-1}\widehat{\mathbf{Z}})\|_2\|\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f\|_2 \le \tag{22}$$

$$\le \|P_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\mathbf{Z}\|_2\|(\mathbf{Id} - \mathbf{Z}^{-1}\widehat{\mathbf{Z}})\|_2\|\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f\|_2 \le \tag{23}$$

$$\le B\|\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\mathbf{Z}\mathcal{P}_{\mathcal{S}}\|_2\|\mathbf{Id} - \mathbf{Z}^{-1/2}\widehat{\mathbf{Z}}\mathbf{Z}^{-1/2}\|_2\|\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\| \tag{24}$$

Notice that $\|\mathbf{Z}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}\mathcal{P}_{\mathcal{S}}\|_2 = \|\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} + \tau(\mathcal{P}_{\mathcal{S}} - \mathbf{Id}))\|_2 \le 2$, because $0 \preceq \mathcal{P}_{\mathcal{S}} \preceq 1$. In a similar fashion, based on Lemma A.2 we have $\|\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\|_2 \le \|\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}\|_2 = \|(\widehat{\mathbf{Z}} - \tau\mathbf{Id})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}\| \le \|\widehat{\mathbf{Z}}\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}\|_2 + 1 \le \frac{8}{3}$. Utilizing Lemma A.5 we have w.p. at least $1-\delta/4$:

$$|\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} - \widehat{\mathbf{Z}}_{\mathcal{S}})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f| \le \frac{64B}{3}\sqrt{\max_{w\in\text{supp}\varrho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)\log(56T/\delta)}$$

To bound the second term of Eq.(17), we use standard concentration results for sub-Gaussian random variables:

$$\begin{aligned}
&\|\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} - \widehat{\mathbf{Z}}_{\mathcal{S}})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\|_2 = \|\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}(\mathbf{Z} - \widehat{\mathbf{Z}})P_{\mathcal{S}}\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\|_2 \le \\
&\le \frac{1}{\sqrt{\tau}}\|\mathcal{P}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}\|_2\|(\mathbf{Id} - \mathbf{Z}^{-1}\widehat{\mathbf{Z}})\|_2\|\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1/2}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\|_2 \le \\
&\le \frac{4}{\sqrt{\tau}}\sqrt{\max_{w\in\text{supp}\varrho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)\log(56T/\delta)}\sqrt{\left\|\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1/2}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\mathcal{P}_{\mathcal{S}}\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1/2}\right\|_2} \le \\
&\le \frac{4}{\sqrt{\tau}}\sqrt{\max_{w\in\text{supp}\varrho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)\log(56T/\delta)}
\end{aligned}$$

$$\left|\phi^\top(w)\mathbf{Z}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}} - \widehat{\mathbf{Z}}_{\mathcal{S}})\widehat{\mathbf{Z}}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\eta_{1:t}\right| \le \frac{8B\sqrt{\log(4/\delta)}}{\sqrt{\tau}}\sqrt{\max_{w\in\text{supp}\varrho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)\log(56T/\delta)}$$

We now turn our attention to the second term of Eq.(15). We use the symmetric derivation as for the first term:

$$|\phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}y_t| \leq \tag{25}$$

$$\leq |\phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f| + |\phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\eta_{1:t}| \tag{26}$$

We again start by bounding the first term:

$$|\phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f| \leq B\|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\|_2 = \tag{27}$$

$$= B\|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}(\mathbf{Z}^{-1}\widetilde{\mathbf{Z}}-\mathbf{Id})\mathcal{P}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\|_2 \leq \tag{28}$$

$$\leq B\|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}(\mathbf{Z}^{-1}\widetilde{\mathbf{Z}}-\mathbf{Id})\mathcal{P}_{\mathcal{S}}\|_2\|\mathcal{P}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\|_2 \leq \tag{29}$$

$$\leq B\|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}\mathcal{P}_{\mathcal{S}}\|_2\|\mathbf{Z}^{-1}\widetilde{\mathbf{Z}}-\mathbf{Id}\|_2\|\mathcal{P}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\|_2 \tag{30}$$

In the above derivation we used that exchanging the places of two operators keeps the spectrum the same, and the identity $\mathcal{P}_{\mathcal{S}}^2 = \mathcal{P}_{\mathcal{S}}$.
Note that $\|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}\mathcal{P}_{\mathcal{S}}\|_2 = \|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\mathbf{Z}_{\mathcal{S}}-\tau(\mathbf{Id}-\mathcal{P}_{\mathcal{S}}))\|_2 \leq \|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathbf{Z}_{\mathcal{S}}\|_2 + 1$. By using Lemmas (A.2,A.5) we have with probability at least $1-\delta/4$[5], $\frac{8}{9}\mathbf{Z} \preceq \widetilde{\mathbf{Z}}, \Phi_{\mathcal{S}}\Phi_{\mathcal{S}}^\top + \tau\mathbf{Id} \preceq \frac{10}{9}\mathbf{Z}$ and $\frac{72}{100}\widetilde{\mathbf{Z}} \preceq \frac{8}{10}\mathbf{Z} \preceq \mathbf{Z}_{\mathcal{S}} \preceq \frac{10}{8}\mathbf{Z} \preceq \frac{90}{64}\widetilde{\mathbf{Z}} \preceq \frac{900}{512}\widetilde{\mathbf{Z}}_{\mathcal{S}}$. Multiplying both sides with $\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1/2}$ we have $\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1/2}\mathbf{Z}_{\mathcal{S}}\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1/2} \preceq \frac{900}{512}\mathbf{Id}$, and as $\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1/2}\mathbf{Z}_{\mathcal{S}}\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1/2}, \widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathbf{Z}_{\mathcal{S}}$ have the same spectrum we finally have $\|\widetilde{\mathbf{Z}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}\mathcal{P}_{\mathcal{S}}\|_2 \leq \frac{353}{128} < 3$.

We adopt the same approach in bounding $\|\mathcal{P}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\|_2 = \|\mathbf{Z}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\mathcal{P}_{\mathcal{S}}\|_2 \leq \|\mathbf{Z}_{\mathcal{S}}^{-1}\widehat{\mathbf{Z}}_{\mathcal{S}}\|_2 + 1$. Again using Lemmas (A.5,A.2) we have $\frac{72}{125}\widehat{\mathbf{Z}}_{\mathcal{S}} \preceq \frac{72}{100}\widehat{\mathbf{Z}} \preceq \frac{8}{10}\mathbf{Z} \preceq \mathbf{Z}_{\mathcal{S}} \implies \|\mathbf{Z}_{\mathcal{S}}^{-1/2}\widehat{\mathbf{Z}}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1/2}\|_2 \leq \frac{125}{72} < 3$. Combining these inequalities we finally have w.p. at least $1-\delta/4$:

$$|\phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top f| < 9B\sqrt{\max_{w\in\text{supp}\varrho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)\log(56T/\delta)}$$

To bound the second term of Eq.(25) we will use the B^2-sub-Gaussianity of the noise $\eta_{1:t}$. Note that:

$$\begin{aligned}
&\|\phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\|_2 \leq \|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\|_2 \leq \\
&\leq \|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}(\mathbf{Z}^{-1}\widetilde{\mathbf{Z}}-\mathbf{Id})\mathcal{P}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\|_2 \leq \frac{1}{\sqrt{\tau}}\|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}(\mathbf{Z}^{-1}\widetilde{\mathbf{Z}}-\mathbf{Id})\mathcal{P}_{\mathcal{S}}\|_2\|\mathbf{Z}_{\mathcal{S}}^{-1/2}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\|_2 \leq \\
&\leq \frac{1}{\sqrt{\tau}}\|\mathcal{P}_{\mathcal{S}}\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}\|_2\|(\mathbf{Z}^{-1}\widetilde{\mathbf{Z}}-\mathbf{Id})\|_2\|\mathbf{Z}_{\mathcal{S}}^{-1/2}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\|_2 = \\
&= \frac{1}{\sqrt{\tau}}\|\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\mathcal{P}_{\mathcal{S}}\mathbf{Z}\mathcal{P}_{\mathcal{S}}\|_2\|(\mathbf{Z}^{-1/2}\widetilde{\mathbf{Z}}\mathbf{Z}^{-1/2}-\mathbf{Id})\|_2\sqrt{\|\mathbf{Z}_{\mathcal{S}}^{-1/2}\mathcal{P}_{\mathcal{S}}\Phi_{\mathbf{W}_t}\Phi_{\mathbf{W}_t}^\top\mathcal{P}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1/2}\|_2} \leq \\
&\leq \frac{1}{\sqrt{\tau}}\frac{353}{64}\sqrt{\max_{w\in\text{supp}\varrho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)\log(56T/\delta)}\sqrt{\|\mathbf{Z}_{\mathcal{S}}^{-1/2}\widehat{\mathbf{Z}}_{\mathcal{S}}\mathbf{Z}_{\mathcal{S}}^{-1/2}\|_2} \leq \\
&< \frac{6}{\sqrt{\tau}}\sqrt{2\max_{w\in\text{supp}\varrho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)\log(56T/\delta)}
\end{aligned}$$

Now by sub-Gaussianity we have w.p at least $1-\delta/4$:

$$|\phi^\top(w)\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}(\widetilde{\mathbf{Z}}_{\mathcal{S}}-\mathbf{Z}_{\mathcal{S}})\mathbf{Z}_{\mathcal{S}}^{-1}P_{\mathcal{S}}\Phi_{\mathbf{W}_t}\eta_{1:t}| \leq \frac{12B\sqrt{\log(4/\delta)}}{\sqrt{\tau}}\sqrt{2\max_{w\in\text{supp}\varrho}\phi^\top(w)\mathbf{Z}^{-1}\phi(w)\log(56T/\delta)}$$

To express the final error bound in terms of the variance with respect to the projected operator $\widetilde{\mathbf{Z}}^{-1}$ we use Lemmas(A.2, A.5): $\mathbf{Z}^{-1} \preceq \frac{10}{9}\widetilde{\mathbf{Z}}^{-1} \preceq \frac{50}{27}\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}$. Finally combining all the bounds in this step we can write w.p. at least $1-\delta$:

$$\begin{aligned}
E_2 < \left(90B\log\left(\frac{56T}{\delta}\right) + \frac{52B\log\left(\frac{4}{\delta}\right)}{\sqrt{\tau}}\right)\widetilde{\sigma}(w) \leq \max_{w\in\mathcal{W}}\left(90B\log\left(\frac{56T}{\delta}\right) + \frac{52B\log\left(\frac{56T}{\delta}\right)}{\sqrt{\tau}}\right)\widetilde{\sigma}(w) = \\
= \left(90B\sqrt{\log\left(\frac{56T}{\delta}\right)} + \frac{52B\sqrt{\log\left(\frac{56T}{\delta}\right)\log\left(\frac{4}{\delta}\right)}}{\sqrt{\tau}}\right)\widetilde{\sigma}_{\max}
\end{aligned}$$

[5]The parameter δ affect the lower bound on the number of action necessary, in Lemma A.5.

Step 3: Bounding E_3.

We note that based on Lemma B.1 we have:

$$|\widehat{\underline{\mu}}_t(w) - \widehat{\mu}_t(w)| = \left|\mathbf{k}_{\mathcal{S}_t}^\top(w)\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}\mathbf{Z}_{\text{priv}}\right|$$

$\mathbf{Z}_{\text{priv}}$ is a noise vector, drawn independently of $\mathbf{k}_{\mathcal{S}_t}^\top(\cdot)\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}$. To bound the dot product, we will use Chernoff-Hoefding along with the sub-Gaussianity of $\mathbf{Z}_{\text{priv}}$. By definition, $\mathbf{Z}_{\text{priv}}$ is σ_0^2-sub Gaussian, thus to bound the dot product it suffices to bound:

$$\|\mathbf{k}_{\mathcal{S}_t}^\top(w)\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}\|_2 \leq \tag{31}$$

$$\leq \sqrt{\mathbf{k}_{\mathcal{S}_t}^\top(w)\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1}\mathbf{k}_{\mathcal{S}_t}^\top} \leq \tag{32}$$

$$\leq \sqrt{\mathbf{k}_{\mathcal{S}_t}^\top(w)\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1}\mathbf{k}_{\mathcal{S}_t}(w)} = \tag{33}$$

$$= \sqrt{\mathbf{k}_{\mathcal{S}_t}^\top(w)\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}^{-1/2}\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}^{-1/2}\mathbf{K}_{\mathcal{S}_t\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t}\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}^{-1/2} + \tau\mathbf{I}_{\mathcal{S}_t}\right)^{-1}\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}^{-1/2}\mathbf{k}_{\mathcal{S}}(w)} = \tag{34}$$

$$= \sqrt{\phi^\top(w)\Phi_{\mathcal{S}_t}\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}^{-1/2}\left(\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}^{-1/2}\Phi_{\mathcal{S}_t}^\top\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top\Phi_{\mathcal{S}_t}\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}^{-1/2} + \tau\mathbf{I}_{\mathcal{S}_t}\right)^{-1}\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}^{-1/2}\Phi_{\mathcal{S}_t}^\top\phi(w)} = \tag{35}$$

$$= \sqrt{\phi^\top(w)\mathcal{P}_{\mathcal{S}_t}\left(\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{I}_{\mathcal{S}_t}\right)^{-1}\phi(w)} = \tag{36}$$

$$= \sqrt{\phi^\top(w)\mathcal{P}_{\mathcal{S}_t}\left(\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top\mathcal{P}_{\mathcal{S}_t}^2 + \tau\mathbf{I}_{\mathcal{S}_t}\right)^{-1}\phi(w)} = \tag{37}$$

$$= \sqrt{\phi^\top(w)\left(\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{I}_{\mathcal{S}_t}\right)^{-1}\mathcal{P}_{\mathcal{S}_t}\phi(w)} \tag{38}$$

We next show that $(\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}^\top\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{I}_{\mathcal{S}_t})^{-1}, \mathcal{P}_{\mathcal{S}_t}$ commute. The proof once again relies on the Lemma A.4:

$$\begin{aligned}\mathcal{P}_{\mathcal{S}_t}(\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}^\top\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{I}_{\mathcal{S}_t})^{-1} &= \mathcal{P}_{\mathcal{S}_t}(\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}^\top\mathcal{P}_{\mathcal{S}_t}^2 + \tau\mathbf{I}_{\mathcal{S}_t})^{-1} = \\ &= (\mathcal{P}_{\mathcal{S}_t}^2\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}^\top\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{I}_{\mathcal{S}_t})^{-1}\mathcal{P}_{\mathcal{S}} = (\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}^\top\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{I}_{\mathcal{S}_t})^{-1}\mathcal{P}_{\mathcal{S}}\end{aligned}$$

As $\mathbf{I} - \mathcal{P}_{\mathcal{S}_t} \succeq 0, (\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{Id}) \succeq 0$ from commutativity we have $(\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{S}_t}\Phi_{\mathcal{S}_t}\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{Id})^{-1}(\mathbf{I} - \mathcal{P}_{\mathcal{S}_t}) \succeq 0$ and thus $\phi^\top(w)(\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{I}_{\mathcal{S}_t})^{-1}\phi(w) \leq \phi^\top(w)(\mathcal{P}_{\mathcal{S}_t}\Phi_{\mathcal{R}_t}\Phi_{\mathcal{R}_t}\mathcal{P}_{\mathcal{S}_t} + \tau\mathbf{Id})^{-1}\mathcal{P}_{\mathcal{S}_t}\phi(w) = \widetilde{\sigma}^2(w)$

We now apply Chernof-Hoefding inequality on the dot product. The following bound holds with probability at least $1-\delta$:

$$E_3 \leq 2\log\left(\frac{1}{\delta}\right)\widetilde{\sigma}_0\widetilde{\sigma}(w) \leq \frac{8B\log T}{\varepsilon_{\text{DP}}}\log\left(\frac{1}{\delta}\right)\sqrt{\log\left(\frac{1.25\log T}{\delta_{\text{DP}}}\right)}\widetilde{\sigma}_{\max}^2$$

Step 4: Total approximation error bound .

We can now finally add the errors E_1, E_2, E_3 and scale the error probability by 3 to obtain w.p. at least $1-\delta$. Using Lemmas(A.2, A.5) once more $\left(\mathbf{Z}^{-1} \preceq \frac{10}{9}\widetilde{\mathbf{Z}}^{-1} \preceq \frac{50}{27}\widetilde{\mathbf{Z}}_{\mathcal{S}}^{-1}\right)$ we have:

$$\begin{aligned}&|f(w) - \widehat{\underline{\mu}}_t(w)| \leq E_1 + E_2 + E_3 \leq \\ &\leq \widetilde{\sigma}_{\max}\underbrace{\left(90B\sqrt{\log\left(\frac{168T}{\delta}\right)} + \frac{52B\sqrt{\log\left(\frac{168T}{\delta}\right)\log\left(\frac{12}{\delta}\right)}}{\sqrt{\tau}} + 3B\sqrt{2\log\left(\frac{6}{\delta}\right)} + \sqrt{24\tau}\right)}_{\beta(\delta)} + \\ &+ \widetilde{\sigma}_{\max}^2\underbrace{\frac{8B\log T}{\varepsilon_{\text{DP}}}\log\left(\frac{3}{\delta}\right)\sqrt{\log\left(\frac{1.25\log T}{\delta_{\text{DP}}}\right)}}_{\beta_1(\varepsilon_{\text{DP}},\delta_{\text{DP}},\delta)}\end{aligned}$$

To further enforce the approximation error holds uniformly over $\text{supp}\varrho$ we use the union bound over the $\text{supp}\varrho \subset \mathcal{W}$

$$\max_{w \in \text{supp}\varrho} |f(w) - \underline{\widehat{\mu}}_t(w)| \leq \beta\left(\frac{\delta}{|\mathcal{W}|}\right)\widetilde{\sigma}_{\max} + \beta_1\left(\varepsilon_{\text{DP}}, \delta_{\text{DP}}, \frac{\delta}{|\mathcal{W}|}\right)\widetilde{\sigma}^2_{\max}$$

The only difference in the approximation error for the LDP estimator is that variance of the privatization noise, $\mathbf{Z}_{\text{priv,LDP}}$ is $\sqrt{t}\sigma_0$. Thus, E_1, E_2 remain unchanged and the upper bound for E_3 is scaled up by $\sqrt{t}$:

$$\max_{w \in \text{supp}\varrho} |f(w) - \underline{\widehat{\mu}}_{t,\text{LDP}}(w)| \leq \beta\left(\frac{\delta}{|\mathcal{W}|}\right)\widetilde{\sigma}_{\max} + \beta_{1,\text{LDP}}\left(\varepsilon_{\text{DP}}, \delta_{\text{DP}}, \frac{\delta}{|\mathcal{W}|}\right)\widetilde{\sigma}^2_{\max}$$

where $\beta_{1,\text{LDP}} = \sqrt{t}\beta_1$.

Step 5: JDP and LDP privacy guarantees.

To show that $\underline{\widehat{\mu}}_t$ is $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$ with respect to the the data-set $\mathcal{D}_t$ it suffices to show that

$$g := \left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}\mathbf{K}_{\mathcal{S}_r,}\mathbf{w}_t\mathbf{y}_t + \mathbf{Z}_{\text{priv}}$$

is $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$ wrt $\mathcal{D}_t$ as the approximating and the projection set $\mathcal{S}_t, \mathcal{R}_t$ are data-invariant. To this end, we use Gaussian Mechanism (Dwork and Roth, 2014). It suffices to bound that the ℓ_2 sensitivity of $h(\mathcal{D}_t) := (\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t})^{-1/2}\mathbf{K}_{\mathcal{S}_t,}\mathbf{w}_t\mathbf{y}_t$:

$$\begin{aligned}\Delta_2 h = \sup_{\mathcal{D}_t,\mathcal{D}'_t} \|h(\mathcal{D}_t) - h(\mathcal{D}_t)\|_2 &\leq 2 \sup_{w \in \text{supp}\varrho, y_t \in [-B,B]} \|(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t})^{-1/2}\mathbf{k}_{\mathcal{S}_t}(w)y_t\|_2 \leq \\ &\leq 2B \sup_{w \in \text{supp}\varrho} \|(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t})^{-1/2}\mathbf{k}_{\mathcal{S}_t}(w)\|_2\end{aligned}$$

However note that this exactly the expression we bounded in Eq.(31) and thus $\Delta_2 h \leq 2B\widetilde{\sigma}_{\max}$. By Gaussian Mechanism, adding noise distributed as $\mathbf{Z}_{\text{priv}} \sim \mathcal{N}\left(0, \left(\widetilde{\sigma}_{\max}\frac{4B\log T}{\varepsilon_{\text{DP}}} \cdot \sqrt{\log\left(\frac{1.25\log T}{\delta_{\text{DP}}}\right)}\right)^2 \mathbf{I}_t\right)$ indeed guarantees that g is private wrt $\mathcal{D}_t$ and thus by post-processing property of differential privacy so is $\underline{\widehat{\mu}}_t$.

$(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$-LDP guarantee of $\underline{\widehat{\mu}}_t$ follows in a similar fashion. We have to show that:

$$g_{\text{LDP}} := \left(\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t}\right)^{-1/2}\mathbf{k}_{\mathcal{S}_r}(w_t)y_t + \mathbf{Z}_{\text{priv}}$$

is $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$-private with respect to the single data-point $w_t = (c_t, x_t)$. It suffices to bound the ℓ_2 sensitivity of $h_{\text{LDP}}(w_t) := (\mathbf{K}_{\mathcal{S}_t,\mathcal{R}_t}\mathbf{K}_{\mathcal{R}_t,\mathcal{S}_t} + \tau\mathbf{K}_{\mathcal{S}_t,\mathcal{S}_t})^{-1/2}\mathbf{k}_{\mathcal{S}_t}(w_t)y_t$. By the exactly the same argument as in the JDP setting we have $\Delta_2 h_{\text{LDP}} \leq 2B\widetilde{\sigma}_{\max}$. Utilizing Gaussian Mechanism we have that g_{LDP} is $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$-differentially private with respect to the data-point (c_t, x_t). In thus follows that $\underline{\widehat{\mu}}_{t,\text{LDP}}$ is $(\varepsilon_{\text{LDP}}, \delta_{\text{LDP}})$. □

We now state and prove the main result upper bounding the cumulative regret of our algorithm.

δ_{ERR}

Theorem A.7. *Consider the problem of contextual kernelized bandits described in the Section 2 where Assumptions 2.3-2.5 hold. If CAPRI is run for a time horizon of T with $T > T_0(\delta_{\text{ERR}}) := \max\left(\overline{T}, \left(\frac{180\log(14T/\delta_{\text{ERR}})}{F^2}\right)^{\beta_p/\beta_p - 1}\right)$, and privacy parameters $\varepsilon_{\text{DP}} > 0$ and $\delta_{\text{DP}} \in (0,1)$ under JDP (resp. LDP), then*

- *CAPRI satisfies $(\varepsilon_{\text{DP}}, \delta_{\text{DP}})$-JDP (resp. LDP)*

- *The regret incurred by the algorithm satisfies:*

$$R_{T,\mathrm{JDP}}(\mathscr{A}) \leq \sqrt{T}\left(4B + \frac{135\log TF^2}{26}\beta\left(\frac{\delta_{ERR}}{|\mathcal{W}|T\log T}\right)\sqrt{2\gamma_T}\right) + \frac{135\log TF^2}{13}\beta_1\left(\varepsilon_{DP}, \delta_{DP}, \frac{\delta_{ERR}}{T\log|\mathcal{W}|}\right)\gamma_T$$

$$R_{T,\mathrm{LDP}}(\mathscr{A}) \leq \sqrt{T}\left(4B\frac{135\log TF^2}{26}\beta\left(\frac{\delta_{ERR}}{|\mathcal{W}|T\log T}\right)\sqrt{2\gamma_T} + \frac{135\log TF^2}{13}\beta_1\left(\varepsilon_{DP}, \delta_{DP}, \frac{\delta_{ERR}}{T\log T|\mathcal{W}|}\right)\gamma_T\right)$$

Regret bound holds with probability $1-\delta_{\mathrm{ERR}}$ over the randomness in the algorithm and the noise sequence. Here $\overline{T}$ is a constant that depends only on the kernel k and the context distribution, and β, β_1 are confidence parameters introduced in Lemma A.6 :

$$\beta(\delta_{ERR}) := \left(90B\sqrt{\log\left(\frac{168T}{\delta_{ERR}}\right)} + \frac{52B\sqrt{\log\left(\frac{168T}{\delta_{ERR}}\right)\log\left(\frac{12}{\delta_{ERR}}\right)}}{\sqrt{\tau}} + 3B\sqrt{2\log\left(\frac{6}{\delta_{ERR}}\right)} + \sqrt{24\tau}\right),$$

$$\beta_1(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}}, \delta_{ERR}) := \frac{8B\log T}{\varepsilon_{DP}}\log\left(\frac{3}{\delta_{ERR}}\right)\sqrt{\log\left(\frac{1.25\log T}{\delta_{DP}}\right)}$$

Proof. We will proceed in several steps. We will first bound the simple regret in the rth epoch as a function of maximum projected variance (see Lemma A.6) in $(r-1)^{\mathrm{th}}$ epoch. We will then establish a bound on the maximum projected variance in terms of maximal information gain, and then finally bound the cumulative regret. In terms of privacy guarantees, the proof will largely follow from the privacy guarantees in Lemma A.6 along with invariance to post-processing with respect to the subsequent epochs. For the sake of noatational overhead we introduce a shorthand $\delta \equiv \delta_{\mathrm{ERR}}$.

Step 1: Bounding the simple regret in r^{th} epoch.

We define the simple regret in the r-th epoch as:

$$\overline{\Delta}_r = \max_{(c,x)\in\mathbf{W}_r}\max_{x_c\in\mathcal{X}} f(c,x_c) - f(c,x)$$

In other words, it is defined as the largest error between the reward function at the maximum point and the point played by the agent, taken over all context-point pairs in the rth epoch.

We bound the simple regret in the first epoch as $\overline{\Delta}_1 \leq \max_{w\in\mathcal{W}} f(w) - \min_{w\in\mathcal{W}} f(w) \leq 2B$.

For $r \geq 2$, w.p. at least $1-\delta/T$ all over epochs we will show a bound $\overline{\Delta}_r \leq 6\Delta_{r-1}$ where $\Delta_r = \beta\left(\frac{\delta}{|\mathcal{W}|T\log T}\right)\widetilde{\sigma}_{r,\max} + \beta_1\left(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}}, \frac{\delta}{|\mathcal{W}|T\log T}\right)\widetilde{\sigma}^2_{r,\max}$[6] is the confidence parameter used in CAPRI.

We will show that $\max_{x\in\mathcal{X}} f(c_t,x) - f(c_t,x_t) \leq 6\Delta_{r-1}$ holds over all time instances in the r-th epoch(in probability). In particular, we will show that this is the case for all elements of the set $\mathcal{X}_r(c_t) := \{x \in \mathcal{X}_{r-1}(c_t) | \widehat{\underline{\mu}}_{r-1}(c_t,x) \geq \max_{x\in\mathcal{X}_{r-1}(c_t)} \widehat{\underline{\mu}}_{r-1}(c_t,x) - 4\Delta_{r-1}\}$. Suppose that, by the way of contradiction, for a $z \in \mathcal{X}_r(c_t)$ we have $f(c_t,x_t^*) - f(c_t,z) \geq 6\Delta_{r-1}$ where $x_t^* = \mathrm{argmax}_{x\in\mathcal{X}} f(c_t,x)$ and let $z^* = \mathrm{argmax}_{z\in\mathcal{X}_{r-1}(c_t)} \widehat{\underline{\mu}}_{r-1}(c_t,z)$.

Clearly $z^* \in \mathcal{X}_r(c_t)$, we will show that $x_t^* \in \mathcal{X}_r(c_t)$. We will show this by inducting on r. For $r=1$, $\mathcal{X}_1(c_t) \equiv \mathcal{X}$ thus the claim is clearly true. Suppose $x_t^* \in \mathcal{X}_{r-1}(c_t)$ we will show $x_t^* \in \mathcal{X}_r(c_t)$. Note that by the result of Lemma A.6 with probability at least $1 - \frac{\delta}{T\log T}$, $\widehat{\underline{\mu}}_{r-1}(c_t,z^*) - \widehat{\underline{\mu}}_{r-1}(c_t,x_t^*) \leq |f(c_t,z^*) - \widehat{\underline{\mu}}_{r-1}(c_t,z^*)| + |f(c_t,x_t^*) - \widehat{\underline{\mu}}_{r-1}(c_t,x_t^*)| + (f(c_t,z^*) - f(c_t,x_t^*)) \leq 2\Delta_{r-1}$ and thus indeed $x_t^* \in \mathcal{X}_r(c_t)$.

[6] The expression for the LDP confidence bound is the same with β_1 being replaced with $\beta_{1,\mathrm{LDP}}$

We can now write:

$$f(c_t, x_t^*) - f(c_t, z) \le |f(c_t, x_t^*) - \underline{\widehat{\mu}}_{r-1}(c_t, x_t^*)| + |f(c_t, z) - \underline{\widehat{\mu}}_{r-1}(c_t, z)| + |\underline{\widehat{\mu}}_{r-1}(c_t, x_t^*) - \underline{\widehat{\mu}}_{r-1}(c_t, z)| \le$$
$$\le 2\Delta_r + \max_{x \in \mathcal{X}_{r-1}(c_t)} \underline{\widehat{\mu}}_{r-1}(c_t, x) - \underline{\widehat{\mu}}_{r-1}(c_t, z) \le 6\Delta_{r-1}$$

contradicting the original assumption and proving the claim.

Taking the union bound over all epochs we have with probability least $1 - \delta/T$:

$$\overline{\Delta}_i \le \begin{cases} 2B & i = 1 \\ 6\Delta_{i-1} & i > 1 \end{cases} \qquad \overline{\Delta}_{i,\text{LDP}} \le \begin{cases} 2B & i = 1 \\ 6\Delta_{i-1,\text{LDP}} & i > 1 \end{cases}$$

Step 2: Bounding $\widetilde{\sigma}^2_{r,\max}$.

In this section $\mathbf{Z}$ reffers to the expected covariance matrix of the sample measure ϱ_r, in the rth epoch. For notational convenience, we drop the subscript r.
Note that by Lemma A.5 w.p. $1 - \delta$ [7] we have $\frac{8}{9}\mathbf{Z} \preceq \Phi_{\mathcal{S}_r}\Phi_{\mathcal{S}_r}^\top + \tau\mathbf{I} \preceq \frac{10}{9}\mathbf{Z}$ where $\mathbf{Z}$ is the expected covariance estimator(for more details please see Lemma A.5). Using the before mentioned lemma also gives us $\frac{8}{9}\mathbf{Z} \preceq \Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top + \tau\mathbf{I} \preceq \frac{10}{9}\mathbf{Z}$. We can now write:

$$\frac{8}{10}\left(\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top + \tau\mathbf{Id}\right) \preceq \frac{8}{9}\mathbf{Z} \preceq \Phi_{\mathcal{S}_r}\Phi_{\mathcal{S}_r}^\top + \tau\mathbf{I} \preceq \frac{10}{9}\mathbf{Z} \preceq \frac{10}{8}\left(\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top + \tau\mathbf{Id}\right)$$

By Lemma A.2 we have:

$$\frac{3}{5}\left(\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top + \tau\mathbf{Id}\right) \preceq \mathcal{P}_{\mathcal{S}}\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top\mathcal{P}_{\mathcal{S}_r} + \tau\mathbf{I} \preceq \frac{5}{3}\left(\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top + \tau\mathbf{Id}\right) \tag{39}$$

Recall that $\widetilde{\sigma}^2_{\max,r} = \max_{w \in \text{supp}\varrho} \phi^\top(w)\left(\mathcal{P}_{\mathcal{S}_r}\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top\mathcal{P}_{\mathcal{S}_r} + \tau\mathbf{I}\right)^{-1}\phi(w)$. Using Eq(39) we have:

$$\widetilde{\sigma}^2_{\max,r} \le \frac{5}{3}\max_{w \in \text{supp}\varrho} \phi^\top(w)\left(\Phi_{\mathcal{R}_r}\Phi_{\mathcal{R}_r}^\top + \tau\mathbf{Id}\right)^{-1}\phi(w) \le \frac{5}{24}\max_{w \in \text{supp}\varrho} \phi^\top(w)\mathbf{Z}^{-1}\phi(w) \le \frac{90F^2}{104}\frac{\gamma_T}{T}$$

, where the last inequality is the direct consequence of Lemma.A.1.

Step 3: Finalizing the cumulative regret bound.

We can re-write the regret as:

$$R_T(\mathscr{A}) = \mathbb{E}_{c_1, c_2, \ldots c_T \sim \kappa}\left[\sum_{t=1}^T \max_{x \in \mathcal{X}} f(c_t, x) - f(c_t, x_t)\right] = \sum_{t=1}^T \mathbb{E}_{c_1, c_2, \ldots c_t}\left[\max_{x \in \mathcal{X}} f(c_t, x) - f(c_t, x_t)\right]$$

For a fixed epoch, the bound from Step 1 holds w.p. at least $1 - \frac{\delta}{T\log T}$ over the previously drawn contexts. Conditioning on this event, we can bound the expected simple regret w.p. at least $1 - \frac{\delta}{\log T}$ over the reward sequence:

$$t \text{ is in } r\text{-th epoch}, r > 1 \text{ , } \mathbb{E}_{\{c_i\}_{i=1}^t \sim \kappa}\left[\max_{x \in \mathcal{X}} f(c_t, x) - f(c_t, x_t)\right] \le 6\Delta_{r-1} + \frac{2B\delta}{T\log T} \tag{40}$$

[7] This dependence on the probability is captured in $T \ge T_0(\delta)$

Summing Eq.(40) over all time instance(and thus over all epochs) we have with probability at least $1-\delta$:

$$
\begin{aligned}
R_{T,\mathrm{JDP}}(\mathscr{A}) &\le 2B\sqrt{T} + \frac{2B\delta}{\log T} + \sum_{r=2}^{\log T} 6\Delta_{r-1}T_r \\
&\le 4B\sqrt{T} + \sum_{r=2}^{\log T} 6T_r \left(\beta\left(\frac{\delta}{|\mathcal{W}|T\log T}\right)\widetilde{\sigma}_{r-1,\max} + \beta_1\left(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}}, \frac{\delta}{|\mathcal{W}|T\log T}\right)\widetilde{\sigma}^2_{r-1,\max}\right) \le \\
&\le 4B\sqrt{T} + \sum_{r=2}^{\log T} \frac{135F^2}{26}\beta\left(\frac{\delta}{|\mathcal{W}|T\log T}\right)\sqrt{2\gamma_T T_{r-1}} + \frac{135F^2}{13}\beta_1\left(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}}, \frac{\delta}{T\log T|\mathcal{W}|}\right)\gamma_T \le \\
&\le 4B\sqrt{T} + \frac{135\log TF^2}{26}\beta\left(\frac{\delta}{|\mathcal{W}|T\log T}\right)\sqrt{2\gamma_T T} + \frac{135\log TF^2}{13}\beta_1\left(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}}, \frac{\delta}{T\log T|\mathcal{W}|}\right)\gamma_T
\end{aligned}
$$

Using the analogous result of Step 1 for the LDP setting we have with probability at least $1-\delta$:

$$
\begin{aligned}
R_{T,\mathrm{JDP}}(\mathscr{A}) &\le 4B\sqrt{T} + \frac{135\log TF^2}{26}\beta\left(\frac{\delta}{|\mathcal{W}|T\log T}\right)\sqrt{2\gamma_T T} + \frac{135\log TF^2}{13}\beta_{1,\mathrm{LDP}}\left(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}}, \frac{\delta}{T\log|\mathcal{W}|}\right)\gamma_T \le \\
&\le 4B\sqrt{T} + \frac{135\log TF^2}{26}\beta\left(\frac{\delta}{|\mathcal{W}|T\log T}\right)\sqrt{2\gamma_T T} + \frac{135\log TF^2}{13}\beta_1\left(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}}, \frac{\delta}{T\log T|\mathcal{W}|}\right)\gamma_T\sqrt{T}
\end{aligned}
$$

Step 4: Privacy Guarantees.

JDP guarantee.

In *Rand-Proj-USCA*, the points in the r-th epoch are are queried uniformly from the prunned domain set . It is thus clear that the distribution of the points in r-th epoch *does not* depend on the (context, reward) pairs in the same epoch. Hence, it is sufficient to show the points in the r-th epoch are differentially private with respect to the data-set accumulated in the previous $r-1$ epochs, $\cup_{i=1}^{r-1}\{\mathbf{C}_i, \mathbf{Y}_i\}$ [8] As we previously concluded, the points in r-th epoch only depend on the dataset $\cup_{i=1}^{r-1}\{\mathbf{C}_i, \mathbf{Y}_i\}$ through sampling of the domain, which is in turn a random function of the privatized estimators $\{\underline{\widehat{\mu}}_i\}_{i=1}^{r-1}$. We will first prove an auxiliary claim :

Claim A.8. Privatized estimator $\underline{\widehat{\mu}}_i$, calculated in the i-th epoch is $(i\varepsilon_{\mathrm{DP}}/\log T, i\delta_{\mathrm{DP}}/\log T)$-DP with respect to the dataset $\{\mathbf{C}_j, \mathbf{Y}_j\}_{j=1}^{i-1}$

Proof. We proceed by induction on i. The case for $i=1$ follows immediately from the privacy guarantee shown in Lemma A.6. Now suppose the hypothesis holds true for $i-1$. $\underline{\widehat{\mu}}_i$ can be seen as a post-processing of the composition of two private mappings. The first one is $\{\mathbf{W}_j, \mathbf{C}_j\}_{j=1}^{i-1} \to \{\mathbf{W}_i, \mathbf{Y}_i\}$, by induction hypothesis and post-processing invariance ($\{\mathbf{W}_i, \mathbf{Y}_i\}$ only depends on $\{\underline{\widehat{\mu}}_j\}_{j=1}^{i-1}$) it is $((i-1)\varepsilon_{\mathrm{DP}}/\log T, (i-1)\delta_{\mathrm{DP}}/\log T)$-DP. The second mapping is $\{\mathbf{W}_i, \mathbf{Y}_i\} \to \underline{\widehat{\mu}}_i$. More specifically $\widehat{\mu}_i$ is a function of queried-points in the r-th epoch, and by the result of Lemma A.6$(\varepsilon_{\mathrm{DP}}/\log T, \delta_{\mathrm{DP}}/\log T)$-DP. The result of the claim now follows from composition theorem of Dwork and Roth (2014). □

As shown in the claim the set $\{\underline{\widehat{\mu}}_i\}_{i=1}^{r-1}$ is $((r-1)\varepsilon_{\mathrm{DP}}/\log T, (r-1)\delta_{\mathrm{DP}}/\log T)$-DP wrt $\{\mathbf{C}_j, \mathbf{Y}_j\}_{j=1}^{r-1}$. The query points in the r-th epoch, are a result of post-processing from $\{\underline{\widehat{\mu}}_i\}_{i=1}^{r-1}$(they are uniformly sampled from the prunned domain) and thus the desired conclusion follows.

LDP guarantee.

At time instant t , in epoch r only the statistic $g_{\mathrm{LDP}}((c_t, x_t), y_t)) := y_t(\mathbf{K}_{\mathcal{S}_r,\mathcal{R}_r}\mathbf{K}_{\mathcal{R}_r,\mathcal{S}_r} + \tau\mathbf{K}_{\mathcal{S}_r,\mathcal{S}_r})^{-1}\mathbf{k}_{\mathcal{S}_r}(c_t, y_t) + \mathbf{Z}_{\mathrm{priv}}$ is uploaded to the algorithm. As shown in Lemma A.6(Step 5), g_{LDP} is $(\varepsilon_{\mathrm{DP}}, \delta_{\mathrm{DP}})$differentially private with

[8] Here we define $\mathbf{C}_i = \{(x, c) \in \mathbf{W}_i | c\}$ as the context set in the ith epoch.

respect to the context-reward pair (c_t, y_t). This holds over all time instances and all epochs , hence the algorithm is indeed locally differentially private.

$\square$

B Appendix B. Kernel Trick

In this section we show how the estimator $\widehat{\mu}_T$ introduced in Lemma A.6 can be efficiently calculated.

Lemma B.1. *The projected estimator and variance:*

$$\widehat{\mu}_T(w) := \phi^\top(w)(\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S} + \tau\mathbf{Id})^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_T}\mathbf{y}_T$$
$$\widetilde{\sigma}_T^2(w) := \phi^\top(w)(\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S} + \tau\mathbf{Id})^{-1}\phi(w)$$

introduced in Lemma (A.6) can be equivalently written as:

$$\widehat{\mu}_T(w) = \mathbf{k}_\mathcal{S}(w)(\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}} + \tau\mathbf{K}_{\mathcal{S},\mathcal{S}})^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T$$
$$\widetilde{\sigma}_T(w) = \frac{1}{\tau}\left(k(w,w) - \mathbf{k}_\mathcal{S}^\top(w)\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}\left(\mathbf{K}_{\mathcal{R},\mathcal{S}}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}} + \tau\mathbf{Id}\right)^{-1}\mathbf{K}_{\mathcal{R},\mathcal{S}}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{k}_\mathcal{S}(w)\right)$$

Proof.

$$\begin{aligned}
\widehat{\mu}_T(w) &= \phi^\top(w)(\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S} + \tau\mathbf{Id})^{-1}\mathcal{P}_\mathcal{S}\Phi_{\mathbf{W}_T}\mathbf{y}_T = && (41)\\
&= \frac{1}{\tau}\phi^\top(w)\widetilde{\mathbf{Z}}_\mathcal{S}^{-1}(\widetilde{\mathbf{Z}}_\mathcal{S} - \mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S})\Phi_\mathcal{S}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T = && (42)\\
&= \frac{1}{\tau}\left(\mathbf{k}_\mathcal{S}(w) - \phi^\top(w)\widetilde{\mathbf{Z}}_\mathcal{S}^{-1}\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S}\Phi_\mathcal{S}\right)\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T = && (43)\\
&= \frac{1}{\tau}\left(\mathbf{k}_\mathcal{S}(w) - \phi^\top(w)\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}(\mathbf{K}_{\mathcal{R},\mathcal{S}}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}} + \tau\mathbf{I}_\mathcal{R})^{-1}\mathbf{K}_{\mathcal{R},\mathcal{S}}\right)\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T = \\
&= \frac{1}{\tau}\mathbf{k}_\mathcal{S}(w)\left(\mathbf{I} - \mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}(\mathbf{K}_{\mathcal{R},\mathcal{S}}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}} + \tau\mathbf{I}_\mathcal{R})^{-1}\mathbf{K}_{\mathcal{R},\mathcal{S}}\right)\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T = \\
&= \frac{1}{\tau}\mathbf{k}_\mathcal{S}(w)\left(\mathbf{I}_\mathcal{S} - \mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}}(\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}} + \tau\mathbf{I}_\mathcal{S})^{-1}\right)\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T = \\
&= \frac{1}{\tau}\mathbf{k}_\mathcal{S}(w)\left(\mathbf{I}_\mathcal{S} - \mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}}(\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}} + \tau\mathbf{I}_\mathcal{S})^{-1}\right)\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T = \\
&= \mathbf{k}_\mathcal{S}(w)(\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}} + \tau\mathbf{I}_\mathcal{S})^{-1}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T = && (44)\\
&= \mathbf{k}_\mathcal{S}(w)(\mathbf{K}_{\mathcal{S},\mathcal{R}}\mathbf{K}_{\mathcal{R},\mathcal{S}} + \tau\mathbf{K}_{\mathcal{S},\mathcal{S}})^{-1}\mathbf{K}_{\mathcal{S},\mathbf{W}_T}\mathbf{y}_T && (45)\\
\widetilde{\sigma}_T^2(w) &= \phi^\top(w)(\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S} + \tau\mathbf{Id})^{-1}\phi(w) = && (46)\\
&= \frac{1}{\tau}(k(w,w) - \phi^\top(w)\left(\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S} + \tau\mathbf{Id}\right)^{-1}\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S}\phi(w)) = && (47)\\
&= \frac{1}{\tau}\left(k(w,w) - \phi^\top(w)\mathcal{P}_\mathcal{S}\Phi_\mathcal{R}\left(\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S}\Phi_\mathcal{R} + \tau\mathbf{Id}\right)^{-1}\Phi_\mathcal{R}^\top\mathcal{P}_\mathcal{S}\phi(w)\right) = && (48)\\
&= \frac{1}{\tau}\left(k(w,w) - \mathbf{k}_\mathcal{S}^\top(w)\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}}\left(\mathbf{K}_{\mathcal{R},\mathcal{S}}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S},\mathcal{R}} + \tau\mathbf{Id}\right)^{-1}\mathbf{K}_{\mathcal{R},\mathcal{S}}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\mathbf{k}_\mathcal{S}(w)\right)
\end{aligned}$$

In the above derivation we used Lemma A.4 along with the parametric form of the projection operator $\mathcal{P}_\mathcal{S} = \Phi_\mathcal{S}\mathbf{K}_{\mathcal{S},\mathcal{S}}^{-1}\Phi_\mathcal{S}^\top$ □