Nearly Optimal Algorithms for Adversarial Kernelized Bandits

Shogo Iwazaki
LY Corporation
Tokyo, Japan
siwazaki@lycorp.co.jp

Abstract

This paper studies kernelized bandits (also known as Gaussian process bandits) in an adversarial environment, where the reward functions in a known reproducing kernel Hilbert space (RKHS) may be adversarially chosen at each round. We show that the exponential-weight algorithm achieves $\widetilde{O}(\sqrt{T\gamma_T})$ adversarial regret, where T and γ_T denote the number of total rounds and the maximum information gain, respectively. For squared exponential (SE) and ν-Matérn kernels, we also show algorithm-independent lower bounds that guarantee the optimality of our algorithm up to polylogarithmic factors. Furthermore, we present a computationally efficient variant of our algorithm using Nyström approximation while maintaining nearly optimal regret guarantees.

1 Introduction

The kernelized bandit (KB) problem, also known as the Gaussian process (GP) bandit, is an effective sequential decision-making framework for nonlinear reward functions. Its applications are extensive, e.g., robotics [39], recommendation systems [69], and hyperparameter tuning [58]. In prior research on KB, most existing works focus on the *stochastic setting* in which the learner makes decisions from noisy observations of a common *fixed* reward function over rounds. A key desideratum in the application of KB is to develop algorithms that bypass the restrictive assumptions in a stochastic setting, which are often invalid in practice. For example, in recommendation systems, the presence of a malicious user may critically degrade the performance of the stochastic KB algorithm [23]. Decision-making in the stock market may also encounter adversarial behavior from other market participants. To overcome the limitations of the stochastic setting, several existing works [17, 61] have studied the *adversarial* KB problem, aiming to design algorithms under minimal smoothness assumptions encoded via the kernel.

This paper addresses a significant open problem in adversarial KB: the construction of a nearly-optimal algorithm. Specifically, to our knowledge, existing algorithms fail to attain even a no-regret guarantee for the commonly used ν-Matérn kernels. Our algorithm and theoretical findings substantially enhance the existing theoretical guarantees in adversarial KB.

Contributions. Our contributions are described below.

- In Sec. 3, we propose **kernelized Exp3**, which is the kernelized extension of the exponential-weight algorithm for exploration and exploitation (Exp3). We provide an $\widetilde{O}(\sqrt{T\gamma_T})$ regret upper bound for kernelized Exp3, where T and γ_T represent the total number of rounds and the *maximum information gain* (MIG), respectively (the definition is in Sec. 2)[1]. In addition, this is the first guarantee for adversarial KB under an *adaptive* adversary.

[1]The notations $\widetilde{O}(\cdot)$ and $\widetilde{\Omega}(\cdot)$ ignore the polylogarithmic factors. A comprehensive notation table is in Appendix A.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

- In Sec. 4, we propose **RLS-kernelized Exp3**, which is a variant of kernelized Exp3 that alleviates the heavy computational burden of the original kernelized Exp3. Specifically, by utilizing the existing ridge-leverage-score (RLS)-based Nyström method [41], the per-round computational complexity of RLS-kernelized Exp3 becomes $\widetilde{O}(|\mathcal{X}|\gamma_T^2 + \gamma_T^3)$, while that of the original kernelized Exp3 is $O(|\mathcal{X}|^3)$. Here, $|\mathcal{X}|$ denotes the cardinality of the input domain $\mathcal{X} \subset \mathbb{R}^{d}$[2]. Furthermore, RLS-kernelized Exp3 also achieves an $\widetilde{O}(\sqrt{T\gamma_T})$ regret.
- Finally, in Sec. 5, we show the algorithm-independent lower bounds for the adversarial KB problem with SE or ν-Matérn kernels. Our lower bounds guarantee that the proposed algorithms are optimal up to polylogarithmic factors for SE and ν-Matérn kernels with $\nu > 1$.

Tab. 1 summarizes the existing algorithm and our results.

1.1 Related works

The KB algorithms in the stochastic setting have been widely studied, including e.g., GP-upper confidence bound (GP-UCB) [60], GP-Thompson sampling (GP-TS) [18], and theoretically more advanced non-adaptive sampling-based algorithms [16, 38, 49, 67]. In addition to the standard stochastic KB settings, several variants are also studied, e.g., the contextual model [35, 67], the multi-objective setting [19, 73], the parallel setting [21], the noise-free setting [9, 26, 29, 64], and the Bayesian setting [27, 47, 46, 51, 63, 62]. In addition to these settings, some existing studies have examined more robust or general variants of the standard stochastic KB model, which are orthogonal to the adversarial KB model in this paper. For example, robust objective settings [4, 31, 34, 43, 48], adversarial corruptions of rewards [5, 6, 23], adversarial contextual information [53, 54], and non-stationary rewards [11, 20, 30, 71] have been considered.

Regarding the adversarial KB problem, to our knowledge, [17] is the first work to tackle it. However, their analysis is only valid under the quite restrictive rank 1 assumption; specifically, the reward function f_t at each round t must have the form $f_t(\cdot) = k(z_t, \cdot)$. In addition, their lower bounds (Theorem 43 in [17]) only apply to an unrealistic infinite-dimensional linear kernel and do not imply lower bounds for an adversarial KB model with practically used SE or Matérn kernels.[3] Takemori and Sato [61] proposed the algorithm, which broke the limitation of [17], called APG-Exp3. The core idea of [61] is to reduce the adversarial KB problem to an approximated (misspecified) linear bandit (LB) problem. However, due to the non-negligible effect of the linear approximation, the current analysis of APG-Exp3 fails to provide even a no-regret guarantee under the commonly-used ν-Matérn kernel with $d \geq \nu$ (see Tab. 1). Another related work for adversarial KB is [42], which studies the K-armed adversarial contextual KB problem and remains distinct from our setting. Their model considers independent K reward functions on the RKHS defined over the context space (with $K \ll T$). In contrast, [17, 61] and our analysis assume that the reward function is an element of RKHS defined on *entire* decision space $\mathcal{X}$, whose cardinality may be $|\mathcal{X}| \gg T$ or infinite.

From a technical perspective, our algorithm construction and several parts of our analysis leverage the existing results in the adversarial LB and stochastic KB. Firstly, from the high-level view, our algorithm construction is interpreted as an extension of the existing Exp3-based algorithms for adversarial LB [3, 36, 72]. The differences from the existing Exp3 are the one-sample estimator of the reward function and the exploration distribution, which are redesigned by adapting the existing theoretical tools in the stochastic KB problem [12, 16, 25, 65]. Secondly, in our analysis of the regret upper bound, the existing information gain-based arguments by Srinivas et al. [60] play a central role in quantifying the regret. To our knowledge, our paper is the first to bridge the adversarial KB theory with the information-gain-based arguments commonly used in stochastic KB analysis. Thirdly, regarding the RLS-kernelized Exp3 proposed in Sec. 4, we construct our algorithm by combining our kernelized Exp3 with the existing RLS-based Nyström approximation [41]. Furthermore, several parts of our regret analysis are motivated by the techniques in [15], which studies the stochastic KB algorithm with Nyström approximation. Finally, regarding the lower bound, our analysis reduces the adversarial KB lower bound to that of the stochastic KB analyzed in [10, 52]. Specifically, our hard problem instance in the proof of the lower bound is interpreted as the kernelized extension of the parameter noise model in LB (see Chap. 29 in [36]), which is used to study the lower bound for adversarial LB [55].

[2] As described in Sec. 3, our analysis is applicable to any compact input domain via the discretization.

[3] The concurrent work by Zhang et al. [70] explicitly discusses this point. See Appendix B in [70] for details.

Table 1: The summary of our results and the existing algorithm under a compact input domain $\mathcal{X} \subset \mathbb{R}^d$. Here, D_T denotes the number of basis functions returned by the sub-routine in APG-Exp3. For the ν-Matérn kernel, the reported bound $D_T = \widetilde{O}\left(T^{\frac{d}{\nu}}\right)$ [61] has a larger exponent of T than the bound on the MIG, $\gamma_T = \widetilde{O}\left(T^{\frac{d}{2\nu+d}}\right)$ [27, 66]. Regarding the per-round computational cost in the table, we assume $|\mathcal{X}| < \infty$. For the continuous input domain, the per-round computational cost can be obtained by substituting $|\mathcal{X}|$ in the table with $O(T^{d/2})$ (the detail is in Sec. 3).

Algorithm	Regret (General kernel)	Regret (ν-Matérn)	Per-round computational cost	Adaptive adversary
APG-Exp3 (Takemori and Sato [61])	$\widetilde{O}(\sqrt{TD_T})$	$\widetilde{O}\left(T^{\frac{\nu+d}{2\nu}}\right)$	$O\left(\lvert\mathcal{X}\rvert D_T^2 + D_T^3\right)$	No
Kernelized Exp3 (**Ours**, in Sec. 3)	$\widetilde{O}(\sqrt{T\gamma_T})$	$\widetilde{O}\left(T^{\frac{\nu+d}{2\nu+d}}\right)$	$O\left(\lvert\mathcal{X}\rvert^3\right)$	Yes
RLS-kernelized Exp3 (**Ours**, in Sec. 4)	$\widetilde{O}(\sqrt{T\gamma_T})$	$\widetilde{O}\left(T^{\frac{\nu+d}{2\nu+d}}\right)$	$\widetilde{O}\left(\lvert\mathcal{X}\rvert\gamma_T^2 + \gamma_T^3\right)$	Yes
Lower bound (**Ours**, in Sec. 5)	N/A	$\widetilde{\Omega}\left(T^{\frac{\nu+d}{2\nu+d}}\right)$		

Concurrent work. The concurrent work by Zhang et al. [70] studies an algorithm nearly identical to our kernelized Exp3 in Sec. 3. Indeed, our kernelized Exp3 becomes identical to theirs by replacing our MVR-based exploration distribution with a G-optimal design-based one (defined in Appendix I). Furthermore, their regret analysis and ours are almost identical. The summary of the minor differences between our algorithm and theirs can be found in Appendix B of [70]. We would like to note that our proposal of the computationally efficient nearly-optimal algorithm (RLS-kernelized Exp3 in Sec. 4) and the lower bounds for SE and Matérn kernels (in Sec. 5) do not overlap with the contents in [70].

2 Preliminaries

Problem setting. We study the KB problem in an adversarial environment. Let $\mathcal{X} \subset [0,1]^d$ be the input domain of the problem. At each round t, the learner sequentially draws a query point $\boldsymbol{x}_t \in \mathcal{X}$ based on the history up to round $t-1$; subsequently, the bandit feedback of the reward $f_t(\boldsymbol{x}_t)$ is revealed, where $f_t : \mathcal{X} \to \mathbb{R}$ is the reward function drawn by the environment. This paper assumes the environment to be *adaptive*. Namely, the environment can draw f_t based on the learner's past history $\mathcal{H}_{t-1} \coloneqq \{\boldsymbol{x}_1, f_1(\boldsymbol{x}_1), \ldots, \boldsymbol{x}_{t-1}, f_{t-1}(\boldsymbol{x}_{t-1})\}$, while $(f_t, \boldsymbol{x}_t)$ must be conditionally independent given $\mathcal{H}_{t-1}$. Under this setup, the learner's goal is to minimize the following expected (pseudo) regret $\bar{R}_T$:

$$\bar{R}_T = \sup_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}\left[R_T(\boldsymbol{x})\right], \text{ where } R_T(\boldsymbol{x}) = \sum_{t \in [T]} f_t(\boldsymbol{x}) - \sum_{t \in [T]} f_t(\boldsymbol{x}_t) \text{ and } [T] = \{1, \ldots, T\}. \quad (1)$$

In this paper, we adopt the following standard assumption of KB, which demands that the reward functions lie in a known RKHS.

Assumption 2.1. For some known positive definite kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and some known bounded constant $B > 0$, the reward function f_t drawn by the environment satisfies $f_t \in \mathcal{H}_k$ and $\|f_t\|_k \leq B$ for any $t \in \mathbb{N}_+$. Here, $\mathcal{H}_k$ and $\|\cdot\|_k$ denote the RKHS endowed with the kernel k and its RKHS norm, respectively. Furthermore, we assume that $k(\boldsymbol{x}, \boldsymbol{x}) \leq 1$ for all $\boldsymbol{x} \in \mathcal{X}$.

As examples of the kernel, the following SE and ν-Matérn kernels are commonly studied:

$$k_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2}{2\ell^2}\right), \; k_{\mathrm{Mat}}(\boldsymbol{x}, \boldsymbol{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\|\boldsymbol{x} - \boldsymbol{x}'\|_2}{\ell}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}\|\boldsymbol{x} - \boldsymbol{x}'\|_2}{\ell}\right),$$

where $\ell > 0$ and $\nu > 0$ are the lengthscale and smoothness parameters, respectively. Furthermore, $\Gamma(\nu)$ and $K_\nu(\cdot)$ denote the Gamma function and the modified Bessel function of the second kind, respectively. Here, our analysis assumes that the underlying kernel function is known beforehand as in the vast majority of existing analyses [16, 18, 38, 60, 65, 67]. Addressing more practical scenarios with unknown kernels is beyond the scope of this work and is future work.

Gaussian process (GP) model. In a stochastic setting of KB, the GP model [44] is a commonly used tool for guiding the learner. In the adversarial setting, although GP is not directly used in our algorithm construction, it plays an important role in quantifying regret via GP's information gain in the analysis. Let us consider a function f that follows a mean-zero GP prior $f \sim \mathcal{GP}(0, k)$ characterized by a covariance kernel k. Then, for some training inputs $\mathbf{X} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(t)})$ and outputs $\boldsymbol{y} = (y^{(1)}, \ldots, y^{(t)})^\top$, the GP model assumes that outputs are generated as $y^{(i)} = f(\boldsymbol{x}^{(i)}) + \epsilon^{(i)}$, where $\epsilon^{(i)} \sim \mathcal{N}(0, \lambda)$ is the Gaussian noise with variance $\lambda > 0$, and is independent across $i \in [t]$. Under the aforementioned modeling assumptions, the posterior distribution of f given $(\mathbf{X}, \boldsymbol{y})$ is also a GP [44]. Specifically, its posterior variance $\sigma^2(\boldsymbol{x}; \mathbf{X}, \lambda)$ of $f(\boldsymbol{x})$ is often used to upper-bound the regret of the KB problem:

$$\sigma^2(\boldsymbol{x}; \mathbf{X}, \lambda) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x}, \mathbf{X})^\top \left(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \lambda \boldsymbol{I}_t\right)^{-1} \boldsymbol{k}(\boldsymbol{x}, \mathbf{X}), \tag{2}$$

where $\boldsymbol{k}(\boldsymbol{x}, \mathbf{X}) \coloneqq [k(\boldsymbol{x}, \boldsymbol{x}^{(i)})]_{i \in [t]} \in \mathbb{R}^t$, $\mathbf{K}(\mathbf{X}, \mathbf{X}) \coloneqq [k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})]_{i,j \in [t]} \in \mathbb{R}^{t \times t}$, and $\boldsymbol{I}_t$ are the kernel vector, the kernel matrix, and the $t \times t$ identity matrix, respectively.

Maximum information gain (MIG). The MIG of a GP is a widely-used complexity parameter for the stochastic KB problem [60]. Given some inputs $\mathbf{X} = [\boldsymbol{x}^{(i)}]_{i \in [t]}$, let $I(\boldsymbol{f}(\mathbf{X}); \boldsymbol{y}) \coloneqq \frac{1}{2} \log \det(\boldsymbol{I}_t + \lambda^{-1} \mathbf{K}(\mathbf{X}, \mathbf{X}))$ be the mutual information between random vectors $\boldsymbol{f}(\mathbf{X}) = (f(\boldsymbol{x}^{(1)}), \ldots, f(\boldsymbol{x}^{(t)}))^\top \sim \mathcal{N}(\boldsymbol{0}, \mathbf{K}(\mathbf{X}, \mathbf{X}))$ and $\boldsymbol{y} \coloneqq \boldsymbol{f}(\mathbf{X}) + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \lambda \boldsymbol{I}_t)$. Namely, $I(\boldsymbol{f}(\mathbf{X}); \boldsymbol{y})$ represents the mutual information between underlying function values and outputs under the Bayesian assumption of GP. Then, the MIG $\gamma_t(\lambda, \mathcal{X})$ is defined as the maximum amount of $I(\boldsymbol{f}(\mathbf{X}); \boldsymbol{y})$ over all possible selections of t inputs $\mathbf{X} \in \mathcal{X}^t$: $\gamma_t(\lambda, \mathcal{X}) \coloneqq \sup_{\mathbf{X} \in \mathcal{X}^t} I(\boldsymbol{f}(\mathbf{X}); \boldsymbol{y})$. For simplicity, we use the notation γ_t by omitting the dependence on λ and $\mathcal{X}$. For commonly used kernels, the growth rates of MIG are known. For example, for the SE kernel, $\gamma_t = O((\log t)^{d+1} / (\log \log t)^d)$ holds [28]. Regarding ν-Matérn kernels, it is known that $\gamma_t = \widetilde{O}(t^{\frac{d}{2\nu + d}})$ [27, 66]. Here, one of the well-known useful results for regret analysis of stochastic KB is the relation between MIG and the posterior variance of GP. Specifically, it is known that the cumulative posterior variance at the observed points is bounded from above by the MIG [Lemma 5.3 and 5.4 in 60]:

$$\sum_{i \in [t]} \sigma^2(\boldsymbol{x}^{(i)}; \mathbf{X}^{(i-1)}, \lambda) \leq \frac{2\gamma_t}{\log(1 + \lambda^{-1})}, \tag{3}$$

where we define $\mathbf{X}^{(i-1)}$ as $\mathbf{X}^{(i-1)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(i-1)})$ for any i. Even in the adversarial KB setting, the above inequality plays a central role in upper-bounding the regret, as highlighted in Sec. 3.

Feature representation of kernel and GP. In our algorithm construction and analysis, we leverage the feature representation of kernels. In general, it is known that a positive definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ cannot always be represented by a finite-dimensional feature map for a continuous input domain $\mathcal{X}$; however, for a finite input domain $\mathcal{X}$, we can always construct an $|\mathcal{X}|$-dimensional feature map $\psi : \mathbb{R}^d \to \mathbb{R}^{|\mathcal{X}|}$, which satisfies $k(\boldsymbol{x}, \boldsymbol{x}') = \psi(\boldsymbol{x})^\top \psi(\boldsymbol{x}')$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$. For example, for kernel matrix $\mathbf{K}(\mathcal{X}, \mathcal{X}) \coloneqq [k(\boldsymbol{x}, \boldsymbol{x}')]_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}}$, the Cholesky decomposition $\mathbf{K}(\mathcal{X}, \mathcal{X}) \coloneqq \boldsymbol{L}\boldsymbol{L}^\top$ induces the feature map by defining $\psi(\boldsymbol{x}^{(i)})$ as the transpose of the i-th row of $\boldsymbol{L} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$.

3 Exponential-weight algorithm for adversarial kernelized bandits

The pseudocode for our first algorithm, kernelized Exp3, is presented in Alg. 1. While Alg. 1 assumes that $\mathcal{X}$ is a finite set, the extension to a continuous input domain is obtained by discretizing $\mathcal{X}$, as discussed later in this section. From a high-level perspective, Alg. 1 is constructed based on the Exp3 algorithm for adversarial LB [3, 72], using a kernel feature map. The primary differences lie in (i) the choice of the estimator for $f_t(\cdot)$ (Lines 6–7) and (ii) the choice of the exploration distribution (Line 2).

Estimator. To use the Exp3 algorithm, at each round t, we require an estimator $\widehat{f}_t(\cdot)$ of the reward function. The standard Exp3 algorithm in the adversarial LB adopts an unbiased estimator based on inverse propensity score (IPS); however, the IPS estimator suffers from high variance as the feature map dimension increases, making it unsuitable for a kernelized setting. To address this issue, we adopt the following regularized version of the IPS estimator, whose effectiveness has been demonstrated in

Algorithm 1 Kernelized Exp3 for adversarial kernelized bandits

Require: Finite input domain $\mathcal{X}$, kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, learning rate $\eta > 0$, regularization parameter $\lambda \geq 1$, mixing ratio $\alpha \in (0, 1)$, confidence width parameter $\beta > 0$, total budget $T \in \mathbb{N}_+$.

1: Prepare feature map $(\psi(\boldsymbol{x}))_{\boldsymbol{x} \in \mathcal{X}}$ (e.g., by Cholesky decomposition).

2: Prepare the exploration distribution $\pi(\boldsymbol{x})$ by using MVR-input sequence $(\boldsymbol{x}_t^{(\mathrm{MVR})})_{t \leq \lceil T\alpha \rceil}$:

$$\pi(\boldsymbol{x}) = \frac{1}{\lceil T\alpha \rceil} \sum_{t=1}^{\lceil T\alpha \rceil} \mathbb{1}\left\{\boldsymbol{x}_t^{(\mathrm{MVR})} = \boldsymbol{x}\right\}, \tag{4}$$

where $\boldsymbol{x}_t^{(\mathrm{MVR})} \in \mathrm{argmax}_{\boldsymbol{x} \in \mathcal{X}} \sigma^2\left(\boldsymbol{x}; \mathbf{X}_{t-1}^{(\mathrm{MVR})}, \lambda\right)$ and $\mathbf{X}_{t-1}^{(\mathrm{MVR})} = \left(\boldsymbol{x}_1^{(\mathrm{MVR})}, \ldots, \boldsymbol{x}_{t-1}^{(\mathrm{MVR})}\right)$.

3: Initialize sampling distribution: $P_1(\boldsymbol{x}) \leftarrow \alpha\pi(\boldsymbol{x}) + (1-\alpha)\widetilde{P}_1(\boldsymbol{x})$, where $\widetilde{P}_1(\boldsymbol{x}) = \frac{1}{|\mathcal{X}|}$.

4: **for** $t = 1, 2, \ldots, T$ **do**

5: The environment draws f_t, and the learner draws $\boldsymbol{x}_t \sim P_t$ and receives the reward $f_t(\boldsymbol{x}_t)$.

6: Calculate the reward estimator $\widehat{f}_t$ as in Eq. (6).

7: Calculate optimistic estimator u_t as $u_t(\boldsymbol{x}) = \widehat{f}_t(\boldsymbol{x}) + \beta\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}}$.

8: Update the sampling distribution:

$$P_{t+1}(\boldsymbol{x}) = \alpha\pi(\boldsymbol{x}) + (1-\alpha)\widetilde{P}_{t+1}(\boldsymbol{x}), \text{ where } \widetilde{P}_{t+1}(\boldsymbol{x}) = \frac{\exp\left(\eta \sum_{i=1}^{t} u_i(\boldsymbol{x})\right)}{\sum_{\boldsymbol{x}' \in \mathcal{X}} \exp\left(\eta \sum_{i=1}^{t} u_i(\boldsymbol{x}')\right)}. \tag{5}$$

9: **end for**

the stochastic KB literature [16, 25, 40]:

$$\widehat{f}_t(\boldsymbol{x}) = \psi(\boldsymbol{x})^\top G_t(\lambda)^{-1} \psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t), \text{ where } G_t(\lambda) = \sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}. \tag{6}$$

Here, $P_t(\cdot)$ is the sampling distribution of the learner at round t, and $\lambda > 0$ is the regularization parameter. Note that, when $\lambda = 0$, $\widehat{f}_t(\cdot)$ reduces to the standard IPS estimator used in LB [see, e.g., Chap. 27 in 36]. Due to the regularization, the estimator $\widehat{f}_t(\boldsymbol{x})$ is not unbiased when $\lambda > 0$. To prevent the exploration behavior from being undermined by the bias of the estimator, we use the optimistic estimator $u_t(\cdot) \coloneqq \widehat{f}_t(\cdot) + \beta\|\psi(\cdot)\|_{G_t(\lambda)^{-1}}$ (Line 7) in the exponential-weight procedures by adding the optimistic term $\beta\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}}$, where $\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}} \coloneqq \sqrt{\psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x})}$. Note that the addition of such a term is the common algorithm construction strategy in the adversarial LB [3, 72].

Exploration distribution. As with Exp3 for adversarial LB, we adopt an exploration distribution $\pi(\boldsymbol{x})$ to prevent the estimator $\widehat{f}_t(\cdot)$ from becoming unstable due to the extremely small sampling probabilities. As we will see later in Lemma 3.3, the variance (the second moment) of the estimator $\widehat{f}_t(\cdot)$ is dominated by the posterior variance of a GP, whose inputs are sampled from the distribution P_t. Based on this fact, we propose using the empirical distribution of the input sequence generated by *maximum variance reduction* (MVR) [65] (see, Alg. 2 in Appendix B), which greedily minimizes the maximum posterior variance of the GP.

Regret upper bound. We provide the following regret upper bound for kernelized Exp3.

Theorem 3.1 (General regret upper bound for kernelized Exp3)**.** *Suppose that Assumption 2.1 holds. Assume that $\mathcal{X}$ is finite. Furthermore, set $\lambda \geq 1$, $\beta = B\sqrt{\lambda/T}$, $\eta \in (0, \overline{\eta})$, and $\alpha = 2\eta(2B\gamma_T + \beta\sqrt{\gamma_T})$, where $\overline{\eta} = 2^{-1}(2B\gamma_T + \beta\sqrt{\gamma_T})^{-1}$. Then, Alg. 1 satisfies*

$$\bar{R}_T \leq 2\alpha BT + \frac{\log|\mathcal{X}|}{\eta} + 8\eta(\lambda + T)B^2\gamma_T + 4B\sqrt{\lambda T\gamma_T}. \tag{7}$$

The full proof is provided in Appendix D. The desired regret upper bound is obtained by tuning the learning rate η depending on T and $|\mathcal{X}|$, as formally stated in the following Corollary 3.2.

Corollary 3.2 (Regret upper bound for kernelized Exp3 with tuned learning rate η)**.** *Fix $\lambda \geq 1$ and set $\beta = B\sqrt{\lambda/T}$, $\eta = \Theta(\sqrt{\log|\mathcal{X}|/T\gamma_T})$, and $\alpha = 2\eta(2B\gamma_T + \beta\sqrt{\gamma_T})$. Assume $|\mathcal{X}| = o(\exp(T\gamma_T^{-1}))$[4]. Furthermore, suppose that $B = \Theta(1)$ and $d = \Theta(1)$ hold. Then, we have $\bar{R}_T = O(\sqrt{T\gamma_T \log|\mathcal{X}|})$.*

Proof overview of Theorem 3.1. The basic proof strategy follows from that of Exp3. Compared with existing analyses in adversarial LB, the primary technical challenge is how to address the high dimensionality of the feature map $\psi(\boldsymbol{x})$. Specifically, by following the standard Exp3 analysis, we can observe that the quadratic form $\|\psi(\boldsymbol{x})\|^2_{G_t(\lambda)^{-1}}$ is the central quantity for governing the analysis. For example, the second moment $\sum_{\boldsymbol{x}\in\mathcal{X}} P_t(\boldsymbol{x})\widehat{f}_t^2(\boldsymbol{x})$ and the supremum of the estimator, which are crucial quantities for Exp3 analysis [e.g., see Chaps. 11 and 27 in 36], are dominated by this quadratic form. Without regularization $\lambda = 0$, in general, the worst-case value of this quantity depends on the dimensionality of the feature map, i.e., $\|\psi(\boldsymbol{x})\|^2_{G_t(0)^{-1}} = \Theta(|\mathcal{X}|)$. However, under the regularization, we can observe a simple yet useful connection among the quadratic form, the posterior variance of the GP, and the MIG. We introduce the following lemma, which is used in the various parts of our analysis to handle the quadratic form $\|\psi(\boldsymbol{x})\|^2_{G_t(\lambda)^{-1}}$ via MIG.

Lemma 3.3 (Quadratic form of feature map, posterior variance, and MIG)**.** *Fix any $T \in \mathbb{N}_+$ and $\lambda > 0$. Fix any finite input domain $\mathcal{X}$, any kernel $k : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$, and any probability mass function $P : \mathcal{X} \to [0,1]$ on $\mathcal{X}$. Let $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(T)}, \boldsymbol{x}^{(T+1)} \sim P$ be an i.i.d. input sequence drawn from P. Furthermore, let $\psi(\cdot) : \mathcal{X} \to \mathbb{R}^{|\mathcal{X}|}$ be any feature map of k, i.e., $k(\boldsymbol{x},\boldsymbol{x}') = \psi(\boldsymbol{x})^\top\psi(\boldsymbol{x}')$ for any $\boldsymbol{x},\boldsymbol{x}' \in \mathcal{X}$. Then, for any $\boldsymbol{z} \in \mathcal{X}$, the inequality: $\|\psi(\boldsymbol{z})\|^2_{G(T,\lambda,P)^{-1}} \leq \frac{T}{\lambda}\mathbb{E}_{\mathbf{X}^{(T)}}\left[\sigma^2(\boldsymbol{z};\mathbf{X}^{(T)},\lambda)\right]$ holds, where $G(T,\lambda,P) = \sum_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$ and $\mathbf{X}^{(T)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(T)})$. Furthermore, if $\forall \boldsymbol{x} \in \mathcal{X},\ k(\boldsymbol{x},\boldsymbol{x}) \leq 1$, we have*

$$\mathbb{E}_{\boldsymbol{x}^{(T+1)}}\left[\|\psi(\boldsymbol{x}^{(T+1)})\|^2_{G(T,\lambda,P)^{-1}}\right] \leq \frac{T}{\lambda}\mathbb{E}_{\mathbf{X}^{(T)},\boldsymbol{x}^{(T+1)}}\left[\sigma^2(\boldsymbol{x}^{(T+1)};\mathbf{X}^{(T)},\lambda)\right] \leq \frac{2\gamma_T}{\lambda\log(1+\lambda^{-1})}. \quad (8)$$

In addition, if $\lambda \geq 1$, the rightmost quantity in Eq. (8) is bounded from above as $\frac{2\gamma_T}{\lambda\log(1+\lambda^{-1})} \leq 4\gamma_T$.

Lemma 3.3 is derived by combining the direct consequence of matrix Jensen's inequality with information gain inequality in Eq. (3) from [60]. We give the full proof in Appendix C. Our proof for Theorem 3.1 is obtained by carefully combining Lemma 3.3 with the existing Exp3 analysis and the existing MVR analysis, in e.g., [12, 65].

Remark 3.4 (Comparison with [16, 25]). Results similar to those in Lemma 3.3 have already been reported in [16, 25] using the generalized notion of MIG. Specifically, Lemma C.4 in [25] claims $\mathbb{E}_{\boldsymbol{x}^{(T+1)}}[\|\psi(\boldsymbol{x}^{(T+1)})\|^2_{G(T,\lambda,P)^{-1}}] = O(\gamma_{\psi,T})$, where $\gamma_{\psi,T} = \max_{P\in\mathcal{P}_\mathcal{X}} \log\det(T\lambda^{-1}\sum_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \boldsymbol{I}_{|\mathcal{X}|})$, and $\mathcal{P}_\mathcal{X}$ is the set of all distributions over $\mathcal{X}$. Here, $\gamma_{\psi,T}$ is greater than or equal to the original MIG: $\gamma_T \leq \gamma_{\psi,T}$ from the definition [25], and the converse inequality $\gamma_{\psi,T} = O(\gamma_T)$ is provided by the concurrent work [70]. Zhang et al. [70] provide the proof for the inequality $\gamma_{\psi,T} \leq 4\gamma_T$ when $\lambda \geq 1$. See Eq. (16) in the proof of Lemma 8 in [70] for details. The combination of this inequality with the existing results by Camilleri et al. [16], Hong et al. [25] also leads to the same guarantees as our Eq. (8) when $\lambda \geq 1$.

Extension to a continuous domain. The extension to a continuous domain[5] can be directly obtained by relying on the discretization argument used in the stochastic KB literature [e.g., 18, 38]. The only additional requirement is the following uniform Lipschitz condition of RKHS.

Assumption 3.5 (Uniform Lipschitz condition)**.** The function class $\mathcal{F}_k(B) = \{f \in \mathcal{H}_k \mid \|f\|_k \leq B\}$, from which the environment chooses (f_t), satisfies $\forall f \in \mathcal{F}_k(B),\ \forall \boldsymbol{x},\boldsymbol{x}' \in \mathcal{X},\ |f(\boldsymbol{x}) - f(\boldsymbol{x}')| \leq L(B)\|\boldsymbol{x} - \boldsymbol{x}'\|_2$ for some constant $L(B) \in (0,\infty)$, which may depend on B.

The above Lipschitz condition is relatively mild. For example, SE and Matérn kernels with $\nu > 1$ satisfy this condition with $L(B) = CB$ for some constant $C > 0$[6] [e.g., Proposition 1 and Remark 5

[4]The condition $|\mathcal{X}| = o(\exp(T\gamma_T^{-1}))$ is required to guarantee $\eta < \overline{\eta}$, where $\overline{\eta}$ is defined in Theorem 3.1. Note that, if $|\mathcal{X}| = \Omega(\exp(T\gamma_T^{-1}))$, the upper bound $O(\sqrt{T\gamma_T\log|\mathcal{X}|})$ becomes meaningless since $\sqrt{T\gamma_T\log|\mathcal{X}|} = \Omega(T)$.

[5]For an exponentially large finite input domain $|\mathcal{X}| = \Omega(\exp(T))$, we can also obtain a tighter regret than that in Corollary 3.2 by the discretization argument.

[6]This constant $C > 0$ may depend on d, ℓ, and ν.

in 56]. By Assumption 3.5, we can ignore the regret incurred from a sufficiently fine discretization. Specifically, we use a discretized input set $\widetilde{\mathcal{X}} \subset \mathcal{X} \subset [0,1]^d$ as a $\Theta(1/\sqrt{T})$-net of $\mathcal{X}$ with respect to $\|\cdot\|_2$, as with [38][7]. Then, the existence of $\widetilde{\mathcal{X}}$ with $|\widetilde{\mathcal{X}}| = O(T^{\frac{d}{2}})$ is guaranteed (e.g., Corollary 4.2.13 in [68]), and the regret incurred from the cumulative discretization error is at most $O(\sqrt{T})$, which does not affect the growth rate of regret. Thus, we obtain the following corollary of Theorem 3.1.

Corollary 3.6 (Regret upper bound for kernelized Exp3 with general compact input domain)**.** *Suppose that Assumptions 2.1 and 3.5 hold. Assume $\mathcal{X} \subset [0,1]^d$. Furthermore, let $\widetilde{\mathcal{X}} \subset \mathcal{X}$ be a $\Theta(1/\sqrt{T})$-net of $\mathcal{X}$. Then, when running Alg. 1 on $\widetilde{\mathcal{X}}$ with $\beta = B\sqrt{\lambda/T}$, $\eta = \Theta(\sqrt{\log T/T\gamma_T})$, and $\alpha = 2\eta(2B\gamma_T + \beta\sqrt{\gamma_T})$, we have $\bar{R}_T = O(\sqrt{T\gamma_T \log T})$.*

Corollary 3.6 shows that kernelized Exp3 achieves an $\widetilde{O}(\sqrt{T\gamma_T})$ regret, which is identical to the nearly-optimal regret obtained in the stochastic KB literature [38, 49, 67]. By substituting the explicit upper bounds of the MIG, we can also obtain kernel-specific regret results. For the SE and ν-Matérn kernels, $\gamma_T = O((\log T)^{d+1}(\log\log T)^{-d})$ [28] and $\gamma_T = \widetilde{O}(T^{\frac{d}{2\nu+d}})$ [27, 66], respectively. Thus, for SE and ν-Matérn kernels, we have $\bar{R}_T = O(\sqrt{T(\log T)^{d+2}(\log\log T)^{-d}})$ and $\bar{R}_T = \widetilde{O}(T^{\frac{\nu+d}{2\nu+d}})$, respectively. In addition, for the linear kernel, we have $\bar{R}_T = O(\sqrt{T}\log(T))$ since $\gamma_T = O(\log T)$ [60]. This result matches the regret of Exp3 in the adversarial LB up to a logarithmic factor [3, 36, 72].

Computational cost. The primary drawback of kernelized Exp3 is its prohibitive computational cost, stemming from the manipulation of the $|\mathcal{X}|$-dimensional feature map. To obtain the MVR sequence used in the exploration distribution, we require $O(|\mathcal{X}|\sum_{t=1}^{\lceil T\alpha\rceil} t^2) = O(|\mathcal{X}|(T\alpha)^3)$ computation via rank-one update of the matrix inverse. With the parameter settings in Corollary 3.2, this is $\widetilde{O}(|\mathcal{X}|(T\gamma_T)^{3/2})$ for fixed λ. Regarding the per-round computational cost, we need to calculate the inverse of the $|\mathcal{X}| \times |\mathcal{X}|$ matrix $G_t(\lambda)$, which requires $O(|\mathcal{X}|^3)$ computation[8]. Since the KB problem mainly focuses on the regime where $|\mathcal{X}|$ is extremely large (e.g., as in Corollary 3.6, we must consider the case $|\mathcal{X}| = \Theta(T^{d/2})$ for a continuous domain), $O(|\mathcal{X}|^3)$ computation per round is computationally infeasible. We tackle this efficiency issue in the next section.

4 Exponential-weight algorithm with RLS-based Nyström approximation

Our second algorithm is a computationally efficient variant of kernelized Exp3, called RLS-kernelized Exp3. Due to space limitations, the pseudocode is provided in Alg. 5 in Appendix B. The huge computational cost in kernelized Exp3 stems from calculating the inverse of the $|\mathcal{X}| \times |\mathcal{X}|$-matrix $G_t(\lambda)$. We mitigate this burden via a low-rank approximation of the feature map weighted by the distribution P_t. Specifically, the resulting approximation is built on a Nyström approximation for the *weighted* kernel function $\widetilde{k}_t(\boldsymbol{x}, \boldsymbol{x}') \coloneqq \sqrt{P_t(\boldsymbol{x})P_t(\boldsymbol{x}')}k(\boldsymbol{x}, \boldsymbol{x}')$[9].

At each round t, let $\mathbf{S}_t \in \mathbb{R}^{|\mathcal{X}|\times s_t}$ be a matrix with exactly one strictly positive entry in each column and zeros elsewhere. Here, $s_t \in [|\mathcal{X}|]$ is determined within the algorithm as described later. Note that we can use $\mathbf{S}_t$ to extract the s_t weighted columns of some matrix $\boldsymbol{M} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ by multiplying from the right as $\boldsymbol{M}\mathbf{S}_t$. Thus, as with [41], we call $\mathbf{S}_t$ a *sampling matrix*. We also define the *weighted* feature matrix $\bar{\boldsymbol{\Psi}}_t = (\sqrt{P_t(\boldsymbol{x}^{(1)})}\psi(\boldsymbol{x}^{(1)}), \ldots, \sqrt{P_t(\boldsymbol{x}^{(|\mathcal{X}|)})}\psi(\boldsymbol{x}^{(|\mathcal{X}|)})) \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$, where we arbitrarily index the elements of $\mathcal{X}$ as $\mathcal{X} = \{\boldsymbol{x}^{(i)}\}_{i\in[|\mathcal{X}|]}$. Note that $G_t(\lambda) = \bar{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$ holds from the definition. Furthermore, note that $\bar{\boldsymbol{\Psi}}_t$ corresponds to the feature matrix of the weighted kernel function $\widetilde{k}_t(\boldsymbol{x}, \boldsymbol{x}')$. In our algorithm, we approximate $G_t(\lambda)$ by projecting $\bar{\boldsymbol{\Psi}}_t$ into the column span of $\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t$. To be more specific, let us define the orthogonal projection matrix $\mathbf{Q}_t \coloneqq (\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t)[(\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t)^\top(\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t)]^\dagger(\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t)^\top \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$, where $\boldsymbol{M}^\dagger$ represents the pseudo-inverse of matrix $\boldsymbol{M}$. Then, RLS-kernelized Exp3 is constructed by replacing quantities $\widehat{f}_t$ and $\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}}$ in kernelized Exp3 with their low-rank approximated versions: $\widetilde{f}_t(\boldsymbol{x}) \coloneqq \psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}(\mathbf{Q}_t\psi(\boldsymbol{x}_t))f_t(\boldsymbol{x}_t)$ and $\widetilde{\sigma}_t(\boldsymbol{x}) \coloneqq \|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}$,

[7] $\boldsymbol{\epsilon}$-net $\widetilde{\mathcal{X}}$ of $\mathcal{X}$ with respect to the norm $\|\cdot\|_2$ is a subset of $\mathcal{X}$ such that $\forall\boldsymbol{x} \in \mathcal{X},\ \exists\boldsymbol{z} \in \widetilde{\mathcal{X}}, \|\boldsymbol{x} - \boldsymbol{z}\|_2 \le \boldsymbol{\epsilon}$ holds.

[8] Note that the difference $G_t(\lambda) - G_{t-1}(\lambda)$ is generally full-rank since $P_t(\boldsymbol{x})$ is updated from $P_{t-1}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}$; thus, we cannot update $G_t(\lambda)^{-1}$ efficiently by relying on standard low-rank update procedures.

[9] For readers unfamiliar with the Nyström approximation of kernels, we refer to the related KB or kernel approximation literature [13, 14, 15, 19, 41].

respectively, where $\widetilde{G}_t(\lambda) = (\mathbf{Q}_t\bar{\mathbf{\Psi}}_t)(\mathbf{Q}_t\bar{\mathbf{\Psi}}_t)^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$. After elementary matrix calculations, we can obtain the following identities, which can be efficiently calculated (See Lemma H.6 for the detailed derivation):

$$\widetilde{f}_t(\boldsymbol{x}) = \frac{f_t(\boldsymbol{x}_t)}{\sqrt{P_t(\boldsymbol{x})P_t(\boldsymbol{x}_t)}}\phi_t(\boldsymbol{x})^\top\left(\sum_{\boldsymbol{x}'\in\mathcal{X}}\phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_{s_t}\right)^{-1}\phi_t(\boldsymbol{x}_t), \tag{9}$$

$$\begin{aligned}\widetilde{\sigma}_t^2(\boldsymbol{x}) = {}& \frac{T}{\lambda P_t(\boldsymbol{x})}\left(\widetilde{k}_t(\boldsymbol{x},\boldsymbol{x}) - \phi_t(\boldsymbol{x})^\top\phi_t(\boldsymbol{x})\right)\\ &+ \frac{1}{P_t(\boldsymbol{x})}\phi_t(\boldsymbol{x})^\top\left(\sum_{\boldsymbol{x}'\in\mathcal{X}}\phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_{s_t}\right)^{-1}\phi_t(\boldsymbol{x}),\end{aligned} \tag{10}$$

where $\phi_t(\boldsymbol{x}) = [(\mathbf{S}_t^\top\widetilde{\mathbf{K}}_t\mathbf{S}_t)^{1/2}]^\dagger\mathbf{S}_t^\top\widetilde{\boldsymbol{k}}_t(\boldsymbol{x}) \in \mathbb{R}^{s_t}$ with the weighted kernel matrix $\widetilde{\mathbf{K}}_t \coloneqq [\widetilde{k}_t(\boldsymbol{x}^{(i)},\boldsymbol{x}^{(j)})]_{i,j\in[|\mathcal{X}|]} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ and the weighted kernel vector $\widetilde{\boldsymbol{k}}_t(\boldsymbol{x}) \coloneqq [\widetilde{k}_t(\boldsymbol{x},\boldsymbol{x}^{(i)})]_{i\in[|\mathcal{X}|]} \in \mathbb{R}^{|\mathcal{X}|}$. The quantity $\phi_t(\boldsymbol{x})$ is referred to as the *Nyström embedding* [15, 19] associated with the weighted kernel function $\widetilde{k}_t$. From the above expressions, we can easily confirm that the calculation of $[\widetilde{f}_t(\boldsymbol{x})]_{\boldsymbol{x}\in\mathcal{X}}$ and $[\widetilde{\sigma}_t^2(\boldsymbol{x})]_{\boldsymbol{x}\in\mathcal{X}}$ requires only $O(|\mathcal{X}|s_t^2 + s_t^3)$-computation (see the first paragraph in Appendix E for details). This drastically improves on the $O(|\mathcal{X}|^3)$-computational cost of the original kernelized Exp3 when $s_t \ll |\mathcal{X}|$.

Selection of the sampling matrix. A remaining question is how to define the sampling matrix $\mathbf{S}_t \in \mathbb{R}^{|\mathcal{X}|\times s_t}$ such that the effect of the approximation does not severely degrade the algorithm performance, while maintaining $s_t \ll |\mathcal{X}|$. To do so, we leverage the RLS-based Nyström approximation methods [14, 41], in which we can guarantee both the approximation accuracy of the kernel and small s_t. Specifically, we adopt the recursive RLS-Nyström algorithm proposed in [41] (Alg. 3) as a subroutine to obtain $\mathbf{S}_t$ (Line 7 in Alg. 5).

Regret and computational cost for RLS-kernelized Exp3. By combining Lemma 3.3 with the existing guarantees in [41], we observe that the recursive RLS-Nyström subroutine attains sufficient approximation accuracy to ensure $\widetilde{O}(\sqrt{T\gamma_T})$ regret with $s_t = \widetilde{O}(\gamma_T)$ and requires only $\widetilde{O}(|\mathcal{X}|\gamma_T^2)$-computation with high probability. As a result, we can obtain the following Theorem 4.1, which formally describes the regret and computational cost guarantees of RLS-kernelized Exp3.

Theorem 4.1 (Regret upper bound and computational cost for RLS-kernelized Exp3)**.** *Suppose that Assumption 2.1 holds. Assume that $\mathcal{X}$ is finite and $|\mathcal{X}| = o(\exp(T\gamma_T^{-1}))$ holds. Furthermore, suppose that d and B are $\Theta(1)$. Set algorithm parameters as $\beta = B(1+\sqrt{2})\sqrt{\lambda/T}$, $\lambda \geq 1$, $\eta = \Theta(\sqrt{\log|\mathcal{X}|}/\sqrt{T\gamma_T})$, $\alpha = 2\eta\sqrt{3}(2(\sqrt{3}+\sqrt{2})B\gamma_T + \beta\sqrt{\gamma_T})$, and $\delta = 1/T^2$. Then, when running Alg. 5, we have $\bar{R}_T = O(\sqrt{T\gamma_T\log|\mathcal{X}|})$. Furthermore, with probability at least $1 - 1/T^2$, for any $t \in [T]$, the computational cost at round t is $\widetilde{O}(|\mathcal{X}|\gamma_T^2 + \gamma_T^3)$.*

The full proof is provided in Appendix E. The above theorem confirms that RLS-kernelized Exp3 is nearly optimal while offering improved computational complexity. Compared with the regret guarantees of the original kernelized Exp3 (Corollary 3.2), the price of the improved computational efficiency is only a constant factor in the regret upper bound, which is hidden in the implied constant of $\bar{R}_T = O(\sqrt{T\gamma_T\log|\mathcal{X}|})$.

Proof sketch. We first consider a high-probability good event where the recursive RLS-Nyström subroutine accurately approximates the true weighted feature matrix (Eq. (86) in the appendix, derived from the existing RLS-Nyström guarantees in Lemma H.7). Under this event, we derive some useful spectral properties (Lemmas H.9 and H.10) to control the approximation errors of both the inverse projected matrix $\widetilde{G}_t(\lambda)^{-1}$ and the projected feature map $\mathbf{Q}_t\bar{\mathbf{\Psi}}_t$. Leveraging these properties within the standard Exp3 analysis framework demonstrates that the additional regrets due to approximation errors are negligible up to a constant factor, thereby yielding the same regret rate as in Theorem 3.1. Finally, by crudely bounding the regret under the complementary bad event, we confirm that setting the confidence parameter to $\delta = 1/T^2$ is sufficient to achieve near-optimal regret.

Remarks on Nyström approximation. Although we leverage the recursive RLS-Nyström algorithm in [41] as a subroutine of our algorithm, other approximation methods may be applicable as long as (i) they have the spectral guarantees of the projection matrix (Eq. (235) in the appendix), and (ii) their computational cost guarantees are quantified by *effective dimension* of the weighted kernel $\widetilde{k}_t$ (see the proof of Lemma H.7). However, to our knowledge, no existing method provides both strictly better computational efficiency and tighter regret bounds than those in Theorem 4.1.

Remarks on exploration distribution. Note that an $O(|\mathcal{X}|(T\alpha)^3)$ pre-processing cost for the preparation of the exploration distribution is required even for RLS-kernelized Exp3 (see the last paragraph in Sec. 3). If this overhead is unacceptable, one can approximate the MVR-sequence itself via recursive RLS-Nyström, which alleviates $O(|\mathcal{X}|(T\alpha)^3)$ computation while maintaining the nearly-optimal guarantees of regret. This variant of RLS-kernelized Exp3 is given in Appendix G. Another approach is to replace the MVR-based distribution with a simpler one, e.g., the uniform distribution. We discuss such possible choices for the exploration distribution in Appendix I; however, based on the known theoretical tools for KB, we are not aware of any choice that strictly reduces the computational cost of MVR-based exploration distributions without imposing further restrictive assumptions on the kernel.

Finally, although this paper mainly focuses on the theoretical side, the simulation experiments for supporting our superior guarantees are provided in Appendix J.

5 Lower bounds

The following Theorem 5.1 provides the lower bounds for adversarial KB under $k = k_{\mathrm{SE}}$ or $k = k_{\mathrm{Mat}}$.

Theorem 5.1 (Lower bounds for adversarial KB)**.** *Fix $\mathcal{X} = [0,1]^d$. Suppose that $k = k_{\mathrm{SE}}$ or $k = k_{\mathrm{Mat}}$ holds. Suppose that ν, ℓ, d, and B are $\Theta(1)$. Then, for any learner's algorithm, there exists a strategy of the environment that satisfies Assumption 2.1 and*

$$\bar{R}_T = \begin{cases} \Omega\left(\sqrt{T(\log T)^{\frac{d}{2}-1}}\right) & \text{if } k = k_{\mathrm{SE}}, \\ \Omega\left(T^{\frac{\nu+d}{2\nu+d}}(\log T)^{-\frac{\nu}{2\nu+d}}\right) & \text{if } k = k_{\mathrm{Mat}}. \end{cases} \tag{11}$$

The full proof is in Appendix F. From Theorem 5.1, we confirm that kernelized Exp3 and RLS-kernelized Exp3 are optimal up to polylogarithmic factors for the SE and Matérn kernels with $\nu > 1$. Compared with the lower bounds for the stochastic setting [52], there are additional multiplicative logarithmic factors $(\log T)^{-1/2}$ and $(\log T)^{-\frac{\nu}{2\nu+d}}$ in the lower bounds for the SE and ν-Matérn kernels, respectively. As clarified in the proof sketch below, these logarithmic factors arise because our proof reduces the adversarial problem to a stochastic KB lower bound under Gaussian noise.

Proof sketch. We consider an environment that reduces the lower bound for adversarial KB to that for stochastic KB by setting: $f_t(\cdot) = f(\cdot) + \widetilde{\eta}_t k(\mathbf{0}, \cdot)$, where $\widetilde{\eta}_t \sim \mathcal{N}(0, \sigma^2)$ for some variance $\sigma^2 > 0$, and $\|f\|_k \leq B/2$. Through a slight modification of the lower bounds for stochastic KB [12, 52], we can show that $\sup_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}[\sum_{t=1}^T f_t(\boldsymbol{x}) - f_t(\boldsymbol{x}_t)] = \Omega(\underline{R}(T, \sigma^2))$, where $\underline{R}(T, \sigma^2) = \sqrt{\sigma^2 T(\log(T/\sigma^2))^{d/2}}$ and $\underline{R}(T, \sigma^2) = \sigma^{\frac{2\nu}{2\nu+d}} T^{\frac{\nu+d}{2\nu+d}}$ for $k = k_{\mathrm{SE}}$ and $k = k_{\mathrm{Mat}}$, respectively. While this constitutes a lower bound on the expected regret, the above problem instance does not satisfy Assumption 2.1 since $\|f_t\|_k \leq B$ may be violated due to the unboundedness of the Gaussian noise $\widetilde{\eta}_t$. Therefore, the only lower bound we can rigorously obtain from the above problem instance is the looser lower bound: $\Omega(\underline{R}(T, \sigma^2)) - \sup_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}[\mathbb{1}\{\exists t \in [T], \|f_t\|_k > B\}(\sum_{t=1}^T f_t(\boldsymbol{x}) - f_t(\boldsymbol{x}_t))]$, where the contribution of the event $\{\exists t \in [T], \|f_t\|_k > B\}$ in the original lower bound is explicitly eliminated (rigorous argument is in Appendix F). To ensure the second term remains negligible, it is sufficient to set σ^2 as $\sigma^2 = \Theta((\log T)^{-1})$ based on the tail properties of the Gaussian noise. Then, the lower bounds $\Omega(\underline{R}(T, \Theta((\log T)^{-1})))$ match Eq. (11).

6 Discussions

In this section, we discuss both the limitations of our work and potential avenues for future research.

Algorithm without discretization. For a continuous input domain, our algorithms require reducing the original problem to a finite input domain via discretization. Such discretization is commonly used in the stochastic KB analysis [29, 38]; however, a limitation of this approach is that the algorithm becomes computationally infeasible as the dimension d increases. This drawback motivates the development of algorithms that operate directly on continuous domains. For example, the kernelized extension of continuous exponential weight and follow-the-regularized leader in the adversarial LB setting [e.g., Chaps. 27 and 28 in 36] may be a promising direction for future research.

Refining lower bounds. Our lower bound has the undesirable dependence on the logarithmic factors. As described in Sec. 5, the source of this worse dependence is that we use the lower bound for stochastic KB under the Gaussian noise model in the proof. Thus, this gap could be addressed in future research by investigating lower bounds for other noise models, such as bounded noise.

Kernelized variants for further advanced topics. Finally, it is interesting to study the other kernelized variants of further advanced topics in the adversarial bandit literature, such as first-order bounds [1, 2], variation-bounded regret [8, 24], and best-of-both-worlds algorithms [7, 37]. We believe that the theoretical tools leveraged in Secs. 3 and 4 are helpful for these future directions.

7 Conclusion

We investigated a nearly-optimal algorithm for adversarial KB. We showed that the kernelized version of the Exp3 algorithm achieves nearly-optimal regret by bridging the gap between the existing information-gain-based arguments for GP and the analysis of Exp3. Furthermore, we proposed a computationally efficient algorithm that leverages RLS-based Nyström approximation. As noted in the discussions in the previous section, we believe that this research opens several avenues for future research on the adversarial KB problem.

Acknowledgment

We appreciate the helpful comments from Jonathan Scarlett and Yu-Jie Zhang on revising the paper.

References

[1] J. D. Abernethy, E. Hazan, and A. Rakhlin. Interior-point methods for full-information and bandit online learning. *IEEE Transactions on Information Theory*, 58(7):4164–4175, 2012.

[2] C. Allenberg, P. Auer, L. Györfi, and G. Ottucsák. Hannan consistency in on-line learning in case of unbounded losses under partial monitoring. In *International Conference on Algorithmic Learning Theory*, pages 229–243. Springer, 2006.

[3] P. Bartlett, V. Dani, T. Hayes, S. Kakade, A. Rakhlin, and A. Tewari. High-probability regret bounds for bandit online linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory-COLT 2008*, 2008.

[4] I. Bogunovic, J. Scarlett, S. Jegelka, and V. Cevher. Adversarially robust optimization with Gaussian processes. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2018.

[5] I. Bogunovic, A. Krause, and J. Scarlett. Corruption-tolerant Gaussian process bandit optimization. In *International Conference on Artificial Intelligence and Statistics*, 2020.

[6] I. Bogunovic, Z. Li, A. Krause, and J. Scarlett. A robust phased elimination algorithm for corruption-tolerant Gaussian process bandits. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2022.

[7] S. Bubeck and A. Slivkins. The best of both worlds: Stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 42.1–42.23. JMLR Workshop and Conference Proceedings, 2012.

[8] S. Bubeck, M. Cohen, and Y. Li. Sparsity, variance and curvature in multi-armed bandits. In *Algorithmic Learning Theory*, pages 111–127. PMLR, 2018.

[9] A. D. Bull. Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 2011.

[10] X. Cai and J. Scarlett. On lower bounds for standard and robust Gaussian process bandit optimization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 1216–1226. PMLR, 2021.

[11] X. Cai and J. Scarlett. Lower bounds for time-varying kernelized bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 73–81. PMLR, 2025.

[12] X. Cai, S. Gomes, and J. Scarlett. Lenient regret and good-action identification in Gaussian process bandits. In *International Conference on Machine Learning*, pages 1183–1192. PMLR, 2021.

[13] D. Calandriello and L. Rosasco. Statistical and computational trade-offs in kernel k-means. *Advances in Neural Information Processing Systems*, 31, 2018.

[14] D. Calandriello, A. Lazaric, and M. Valko. Distributed adaptive sampling for kernel matrix approximation. In *Artificial Intelligence and Statistics*, 2017.

[15] D. Calandriello, L. Carratino, A. Lazaric, M. Valko, and L. Rosasco. Gaussian process optimization with adaptive sketching: Scalable and no regret. In *Conference on Learning Theory*, 2019.

[16] R. Camilleri, K. Jamieson, and J. Katz-Samuels. High-dimensional experimental design and kernel bandits. In *Proc. International Conference on Machine Learning (ICML)*, 2021.

[17] N. Chatterji, A. Pacchiano, and P. Bartlett. Online learning with kernel losses. In *International Conference on Machine Learning*, 2019.

[18] S. R. Chowdhury and A. Gopalan. On kernelized multi-armed bandits. In *Proc. International Conference on Machine Learning (ICML)*, 2017.

[19] S. R. Chowdhury and A. Gopalan. No-regret algorithms for multi-task Bayesian optimization. In *International Conference on Artificial Intelligence and Statistics*, 2021.

[20] Y. Deng, X. Zhou, B. Kim, A. Tewari, A. Gupta, and N. Shroff. Weighted Gaussian process bandits for non-stationary environments. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

[21] T. Desautels, A. Krause, and J. W. Burdick. Parallelizing exploration-exploitation tradeoffs in Gaussian process bandit optimization. *Journal of Machine Learning Research*, 2014.

[22] V. V. Fedorov. *Theory of optimal experiments*. Elsevier, 2013.

[23] E. Han and J. Scarlett. Adversarial attacks on Gaussian process bandits. In *International Conference on Machine Learning*, 2022.

[24] E. Hazan and S. Kale. A simple multi-armed bandit algorithm with optimal variation-bounded regret. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 817–820. JMLR Workshop and Conference Proceedings, 2011.

[25] K. Hong, Y. Li, and A. Tewari. An optimization-based algorithm for non-stationary kernel bandits without prior knowledge. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.

[26] S. Iwazaki. Gaussian process upper confidence bound achieves nearly-optimal regret in noise-free Gaussian process bandits. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[27] S. Iwazaki. Improved regret bounds for Gaussian process upper confidence bound in Bayesian optimization. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[28] S. Iwazaki. Tighter regret lower bound for Gaussian process bandits with squared exponential kernel in hypersphere. *arXiv preprint arXiv:2602.17940*, 2026.

[29] S. Iwazaki and S. Takeno. Improved regret analysis in Gaussian process bandits: Optimality for noiseless reward, RKHS norm, and non-stationary variance. In *International Conference on Machine Learning*, pages 26642–26672. PMLR, 2025.

[30] S. Iwazaki and S. Takeno. Near-optimal algorithm for non-stationary kernelized bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 406–414. PMLR, 2025.

[31] S. Iwazaki, Y. Inatsu, and I. Takeuchi. Mean-variance analysis in Bayesian optimization under uncertainty. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

[32] D. Janz. *Sequential decision making with feature-linear models*. PhD thesis, University of Cambridge, 2021.

[33] M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.

[34] J. Kirschner, I. Bogunovic, S. Jegelka, and A. Krause. Distributionally robust Bayesian optimization. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

[35] A. Krause and C. Ong. Contextual Gaussian process bandit optimization. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2011.

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

[37] C.-W. Lee, H. Luo, C.-Y. Wei, M. Zhang, and X. Zhang. Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously. In *International Conference on Machine Learning*, 2021.

[38] Z. Li and J. Scarlett. Gaussian process bandit optimization with few batches. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

[39] D. Lizotte, T. Wang, M. Bowling, and D. Schuurmans. Automatic gait optimization with Gaussian process regression. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[40] B. Mason, L. Jain, S. Mukherjee, R. Camilleri, K. Jamieson, and R. Nowak. Nearly optimal algorithms for level set estimation. In *International Conference on Artificial Intelligence and Statistics*, 2022.

[41] C. Musco and C. Musco. Recursive sampling for the Nyström method. *Advances in Neural Information Processing Systems*, 30, 2017.

[42] G. Neu, J. Olkhovskaya, and S. Vakili. Adversarial contextual bandits go kernelized. In *International Conference on Algorithmic Learning Theory*, 2024.

[43] Q. P. Nguyen, Z. Dai, B. K. H. Low, and P. Jaillet. Value-at-risk optimization with Gaussian processes. In *Proc. International Conference on Machine Learning (ICML)*, 2021.

[44] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2006.

[45] G. Riutort-Mayol, P.-C. Bürkner, M. R. Andersen, A. Solin, and A. Vehtari. Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming. *Statistics and Computing*, 33(1):17, 2023.

[46] D. Russo and B. Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

[47] D. Russo and B. Van Roy. Learning to optimize via information-directed sampling. *Advances in Neural Information Processing Systems*, 27, 2014.

[48] A. Saday, Y. C. Yıldırım, and C. Tekin. Robust Bayesian satisficing. *Advances in Neural Information Processing Systems*, 2023.

[49] S. Salgia, S. Vakili, and Q. Zhao. A domain-shrinking based Bayesian optimization algorithm with order-optimal regret performance. In *Advances in Neural Information Processing Systems*, volume 34, pages 28836–28847. Curran Associates, Inc., 2021.

[50] S. Salgia, S. Vakili, and Q. Zhao. Random exploration in Bayesian optimization: Order-optimal regret and computational efficiency. In *Proc. International Conference on Machine Learning (ICML)*, 2024.

[51] J. Scarlett. Tight regret bounds for Bayesian optimization in one dimension. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4500–4508. PMLR, 2018.

[52] J. Scarlett, I. Bogunovic, and V. Cevher. Lower bounds on regret for noisy Gaussian process bandit optimization. In *Proc. Conference on Learning Theory (COLT)*, 2017.

[53] P. G. Sessa, I. Bogunovic, M. Kamgarpour, and A. Krause. No-regret learning in unknown games with correlated payoffs. *Advances in Neural Information Processing Systems*, 32, 2019.

[54] P. G. Sessa, I. Bogunovic, M. Kamgarpour, and A. Krause. Learning to play sequential games versus unknown opponents. *Advances in Neural Information Processing Systems*, 33:8971–8981, 2020.

[55] O. Shamir. On the complexity of bandit linear optimization. In *Conference on Learning Theory*, 2015.

[56] S. Shekhar and T. Javidi. Multi-scale zero-order optimization of smooth functions in an RKHS. *arXiv preprint arXiv:2005.04832*, 2020.

[57] S. Shekhar and T. Javidi. Instance dependent regret analysis of kernelized bandits. In *International Conference on Machine Learning*, 2022.

[58] J. Snoek, H. Larochelle, and R. P. Adams. Practical Bayesian optimization of machine learning algorithms. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2012.

[59] A. Solin and S. Särkkä. Hilbert space methods for reduced-rank Gaussian process regression. *Statistics and Computing*, 2020.

[60] N. Srinivas, A. Krause, S. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proc. International Conference on Machine Learning (ICML)*, 2010.

[61] S. Takemori and M. Sato. Approximation theory based methods for RKHS bandits. In *International Conference on Machine Learning*, 2021.

[62] S. Takeno and S. Iwazaki. On regret bounds of Thompson sampling for Bayesian optimization. *arXiv preprint arXiv:2603.09276*, 2026.

[63] S. Takeno, Y. Inatsu, and M. Karasuyama. Randomized Gaussian process upper confidence bound with tighter Bayesian regret bounds. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 33490–33515. PMLR, 2023.

[64] S. Vakili. Open problem: Regret bounds for noise-free kernel-based bandits. In *Proc. Conference on Learning Theory (COLT)*, 2022.

[65] S. Vakili, N. Bouziani, S. Jalali, A. Bernacchia, and D. shan Shiu. Optimal order simple regret for Gaussian process bandits. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2021.

[66] S. Vakili, K. Khezeli, and V. Picheny. On information gain and regret bounds in Gaussian process bandits. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

[67] M. Valko, N. Korda, R. Munos, I. Flaounas, and N. Cristianini. Finite-time analysis of kernelised contextual bandits. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, UAI'13, page 654–663. AUAI Press, 2013.

[68] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press, 2018.

[69] M. Yadav, D. Sheldon, and C. Musco. Gaussian process bandits for top-k recommendations. *Advances in Neural Information Processing Systems*, 2024.

[70] Y.-J. Zhang, H. Qiu, J. Scarlett, and K. Jamieson. Near-optimal regret in adversarial kernel bandits, 2026. URL `https://arxiv.org/abs/2605.26585`.

[71] X. Zhou and N. Shroff. No-regret algorithms for time-varying Bayesian optimization. In *2021 55th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2021.

[72] J. Zimmert and T. Lattimore. Return of the bias: Almost minimax optimal high probability bounds for adversarial linear bandits. In *Conference on Learning Theory*, 2022.

[73] M. Zuluaga, A. Krause, and M. Püschel. e-PAL: An active learning approach to the multi-objective optimization problem. *Journal of Machine Learning Research*, 2016.

Table 2: Notation table.

Symbol	Definition/Description
$\mathbb{N}_+$	Set of natural numbers without 0: $\mathbb{N}_+ \coloneqq \{1,\ 2, \ldots\}$.
$[t]$	Set of elements of $\mathbb{N}_+$ up to $t \in \mathbb{N}_+$: $[t] \coloneqq \{1, \ldots, t\}$.
$\|\boldsymbol{x}\|_{\boldsymbol{M}}$	Mahalanobis seminorm with respect to a positive semi-definite matrix $\boldsymbol{M}$: $\|\boldsymbol{x}\|_{\boldsymbol{M}} \coloneqq \sqrt{\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x}}$.
$\|\boldsymbol{x}\|_2$	L2-norm of $\boldsymbol{x} \in \mathbb{R}^m$: $\|\boldsymbol{x}\|_2 \coloneqq \sqrt{\sum_{i=1}^m x_i^2}$.
$\|\boldsymbol{M}\|$	Spectral norm of a matrix $\boldsymbol{M} \in \mathbb{R}^{n \times m}$: $\|\boldsymbol{M}\| \coloneqq \sup_{\boldsymbol{x} \in \mathbb{R}^m \setminus \{\boldsymbol{0}\}} \|\boldsymbol{M}\boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2$.
$\preceq$	Loewner order, i.e., $\boldsymbol{M}_1 \preceq \boldsymbol{M}_2$ means $\boldsymbol{M}_2 - \boldsymbol{M}_1$ is a positive semi-definite matrix.
$\boldsymbol{M}^\dagger$	Pseudo-inverse of matrix $\boldsymbol{M}$.
$\widetilde{O}(\cdot)$	$g_1(T) = \widetilde{O}(g_2(T))$ denotes $g_1(T) = O(g_2(T)(\log T)^c)$ for an absolute constant $c \geq 0$.
$\widetilde{\Omega}(\cdot)$	$g_1(T) = \widetilde{\Omega}(g_2(T))$ denotes $g_1(T) = \Omega(g_2(T)(\log T)^{-c})$ for an absolute constant $c \geq 0$.
$\psi(\cdot)$	Feature map of k on a finite input domain $\mathcal{X}$, i.e., $\psi : \mathcal{X} \to \mathbb{R}^{\|\mathcal{X}\|}$ satisfies $k(\boldsymbol{x}, \boldsymbol{x}') = \psi(\boldsymbol{x})^\top \psi(\boldsymbol{x}')$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$.
$\boldsymbol{\Psi}$	Feature matrix on finite input domain $\mathcal{X} \coloneqq \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(\|\mathcal{X}\|)}\}$: $\boldsymbol{\Psi} \coloneqq (\psi(\boldsymbol{x}^{(1)}), \ldots, \psi(\boldsymbol{x}^{(\|\mathcal{X}\|)})) \in \mathbb{R}^{\|\mathcal{X}\| \times \|\mathcal{X}\|}$.
$\boldsymbol{\theta}_t$	Feature representation of f_t, i.e., $\boldsymbol{\theta}_t \in \mathbb{R}^{\|\mathcal{X}\|}$ satisfies $f_t(\cdot) = \psi(\cdot)^\top \boldsymbol{\theta}_t$ (see Lemma H.1).
$\mathcal{H}_t$	Learner's history of kernelized Exp3 up to round t: $\mathcal{H}_t \coloneqq \{(\boldsymbol{x}_i, f_i(\boldsymbol{x}_i))\}_{i \in [t]}$.
$\widetilde{\mathcal{H}}_t$	Learner's history of RLS-kernelized Exp3 up to round t: $\widetilde{\mathcal{H}}_t \coloneqq \{(\boldsymbol{x}_i, \mathbf{S}_i, f_i(\boldsymbol{x}_i))\}_{i \in [t]}$.
$\pi(\cdot)$	Exploration distribution $\pi(\boldsymbol{x}) \coloneqq \frac{1}{\lceil T\alpha \rceil} \sum_{t=1}^{\lceil T\alpha \rceil} \mathbb{1}\{\boldsymbol{x} = \boldsymbol{x}_t^{(\mathrm{MVR})}\}$. $(\boldsymbol{x}_t^{(\mathrm{MVR})})$ is generated by Alg. 2.
$\widetilde{\pi}(\cdot)$	Approximation of π: $\widetilde{\pi}(\boldsymbol{x}) \coloneqq \frac{1}{\lceil T\alpha \rceil} \sum_{t=1}^{\lceil T\alpha \rceil} \mathbb{1}\{\boldsymbol{x} = \widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})}\}$. $(\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})})$ is generated by Alg. 4.
$\widetilde{P}_t(\cdot)$	Exponential-weight distribution: $\widetilde{P}_t(\cdot) \coloneqq \exp(\eta \sum_{i=1}^{t-1} u_i(\cdot)) / \sum_{\boldsymbol{x}' \in \mathcal{X}} \exp(\eta \sum_{i=1}^{t-1} u_i(\boldsymbol{x}'))$ or $\widetilde{P}_t(\cdot) \coloneqq \exp(\eta \sum_{i=1}^{t-1} \widetilde{u}_i(\cdot)) / \sum_{\boldsymbol{x}' \in \mathcal{X}} \exp(\eta \sum_{i=1}^{t-1} \widetilde{u}_i(\boldsymbol{x}'))$.
$P_t(\cdot)$	Sampling distribution of the learner at round t: $P_t(\cdot) \coloneqq \alpha\pi(\cdot) + (1-\alpha)\widetilde{P}_t(\cdot)$.
γ_T	Maximum information gain up to T-inputs: $\gamma_T \coloneqq \sup_{\mathbf{X} \in \mathcal{X}^T} \frac{1}{2} \log\det(\boldsymbol{I}_T + \lambda^{-1}\mathbf{K}(\mathbf{X}, \mathbf{X}))$.
s_t	Rank determined in recursive RLS-Nyström at round t in RLS-kernelized Exp3.
$\mathbf{S}_t$	$(\|\mathcal{X}\| \times s_t)$-sampling matrix at round t of RLS-kernelized Exp3.
$\mathbf{Q}_t$	Orthogonal projection matrix into the subspace spanned by column vectors of $\bar{\boldsymbol{\Psi}}_t \mathbf{S}_t$, i.e., $\mathbf{Q}_t \coloneqq (\bar{\boldsymbol{\Psi}}_t \mathbf{S}_t)[(\bar{\boldsymbol{\Psi}}_t \mathbf{S}_t)^\top (\bar{\boldsymbol{\Psi}}_t \mathbf{S}_t)]^\dagger (\bar{\boldsymbol{\Psi}}_t \mathbf{S}_t)^\top$.
$\widetilde{k}_t$	Weighted kernel at round t of RLS-kernelized Exp3: $\widetilde{k}_t(\boldsymbol{x}, \boldsymbol{x}') \coloneqq \sqrt{P_t(\boldsymbol{x}) P_t(\boldsymbol{x}')} k(\boldsymbol{x}, \boldsymbol{x}')$.
$\bar{\psi}_t(\cdot)$	Weighted feature map $\bar{\psi}_t(\boldsymbol{x}) \coloneqq \sqrt{P_t(\boldsymbol{x})}\psi(\boldsymbol{x})$.
$\bar{\boldsymbol{\Psi}}_t$	Weighted feature matrix on finite input domain $\mathcal{X} \coloneqq \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(\|\mathcal{X}\|)}\}$: $\bar{\boldsymbol{\Psi}}_t \coloneqq (\bar{\psi}_t(\boldsymbol{x}^{(1)}), \ldots, \bar{\psi}_t(\boldsymbol{x}^{(\|\mathcal{X}\|)})) \in \mathbb{R}^{\|\mathcal{X}\| \times \|\mathcal{X}\|}$.
$\widetilde{\psi}_t(\cdot)$	Projected feature map: $\widetilde{\psi}_t(\boldsymbol{x}) \coloneqq \mathbf{Q}_t \psi(\boldsymbol{x})$.
$G_t(\lambda)$	Inverse weight matrix for estimator: $G_t(\lambda) \coloneqq \bar{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T} \boldsymbol{I}_{\|\mathcal{X}\|}$.
$\widetilde{G}_t(\lambda)$	Approximation of G_t: $\widetilde{G}_t(\lambda) \coloneqq (\mathbf{Q}_t \bar{\boldsymbol{\Psi}}_t)(\mathbf{Q}_t \bar{\boldsymbol{\Psi}}_t)^\top + \frac{\lambda}{T} \boldsymbol{I}_{\|\mathcal{X}\|}$.
$\widehat{f}_t(\cdot)$	Reward estimator for kernelized Exp3: $\widehat{f}_t(\boldsymbol{x}) \coloneqq \psi(\boldsymbol{x})^\top G_t(\lambda)^{-1} \psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t)$.
$\widetilde{f}_t(\cdot)$	Approximated reward estimator for RLS-kernelized Exp3: $\widetilde{f}_t(\boldsymbol{x}) \coloneqq \psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1} \widetilde{\psi}_t(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t)$.
$u_t(\cdot)$	Optimistic estimator for kernelized Exp3: $u_t(\cdot) \coloneqq \widehat{f}_t(\cdot) + \beta \|\psi(\cdot)\|_{G_t(\lambda)^{-1}}$.
$\widetilde{u}_t(\cdot)$	Approximated optimistic estimator for RLS-kernelized Exp3: $\widetilde{u}_t(\cdot) \coloneqq \widetilde{f}_t(\cdot) + \beta \|\psi(\cdot)\|_{\widetilde{G}_t(\lambda)^{-1}}$.
$\sigma^2(\cdot; \mathbf{X}, \lambda)$	Posterior variance of GP: $\sigma^2(\cdot; \mathbf{X}, \lambda) \coloneqq k(\cdot, \cdot) - \boldsymbol{k}(\cdot, \mathbf{X})^\top (\mathbf{K}(\mathbf{X}, \mathbf{X}) + \lambda \boldsymbol{I})^{-1} \boldsymbol{k}(\cdot, \mathbf{X})$.
$\widetilde{\sigma}_t^2(\cdot)$	Approximation of $\|\psi(\cdot)\|^2_{G_t(\lambda)^{-1}}$: $\widetilde{\sigma}_t^2(\cdot) \coloneqq \|\psi(\cdot)\|^2_{\widetilde{G}_t(\lambda)^{-1}}$.
A_t	Event: $\frac{1}{2}\left(\bar{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T} \boldsymbol{I}_{\|\mathcal{X}\|}\right) \preceq \bar{\boldsymbol{\Psi}}_t \mathbf{S}_t \mathbf{S}_t^\top \bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T} \boldsymbol{I}_{\|\mathcal{X}\|} \preceq \frac{3}{2}\left(\bar{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T} \boldsymbol{I}_{\|\mathcal{X}\|}\right)$ (see Lemma H.7).
A	$A \coloneqq \cap_{t \in [T]} A_t$.

A Notation table

Tab. 2 summarizes the notations used in this paper.

B Pseudocode

In this section, we provide the pseudocode for the algorithms that could not be included in the main text due to space limitations. Alg. 2 is the pseudocode of the MVR algorithm leveraged in the preparation of the exploration distribution π in our algorithm. Algs. 3 and 5 are the pseudocode of the recursive RLS-Nyström algorithm and the RLS-kernelized Exp3 algorithm described in Sec. 4, respectively. Note that Alg. 3 is identical to the original Alg. 2 in [41] except for some changes of

notations. Furthermore, Alg. 6 represents the pseudocode of the variant of RLS-kernelized Exp3, which uses an approximated MVR sequence for the exploration distribution as described in the paragraph "Remarks on exploration distribution" of Sec. 4 and in Appendix G. Finally, Alg. 4 is the pseudocode of the algorithm for generating an approximated MVR sequence leveraged in Alg. 6.

Algorithm 2 Maximum variance reduction (MVR)

input: Input domain $\mathcal{X}$, kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, regularization parameter $\lambda > 0$, total budget M.
output: MVR-sequence $(\boldsymbol{x}_t^{(\mathrm{MVR})})_{t\in[M]}$.

$\mathbf{X}_0^{(\mathrm{MVR})} \leftarrow ()$.
for $t = 1, 2, \ldots, M$ **do**
  $\boldsymbol{x}_t^{(\mathrm{MVR})} \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \sigma^2\left(\boldsymbol{x}; \mathbf{X}_{t-1}^{(\mathrm{MVR})}, \lambda\right)$.
  Update training input: $\mathbf{X}_t^{(\mathrm{MVR})} \leftarrow (\boldsymbol{x}_1^{(\mathrm{MVR})}, \ldots, \boldsymbol{x}_t^{(\mathrm{MVR})})$.
end for
return $(\boldsymbol{x}_t^{(\mathrm{MVR})})_{t\in[M]}$.

Algorithm 3 RecursiveRLS-Nyström ($\mathcal{X}$, k, λ, δ) [41]

input: Finite input set $\mathcal{X} \coloneqq \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(|\mathcal{X}|)}\}$, kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, regularization parameter $\lambda > 0$, confidence level parameter $\delta \in (0, 1/32)$
output: weighted sampling matrix $\mathbf{S} \in \mathbb{R}^{|\mathcal{X}|\times s}$

if $|\mathcal{X}| \le 192\log(1/\delta)$ **then**
  return $\mathbf{S} \coloneqq \mathbf{I}_{|\mathcal{X}|\times|\mathcal{X}|}$.
end if
Let $\bar{\mathcal{S}}$ be a random subset of $\{1, \ldots, |\mathcal{X}|\}$, with each $i \in \{1, \ldots, |\mathcal{X}|\}$ included independently with probability $\frac{1}{2}$.
  ▷ Let $\bar{\mathcal{X}} = \{\boldsymbol{x}^{(i_1)}, \boldsymbol{x}^{(i_2)}, \ldots, \boldsymbol{x}^{(i_{|\bar{\mathcal{S}}|})}\}$ for $i_j \in \bar{\mathcal{S}}$ be the data sample corresponding to $\bar{\mathcal{S}}$.
  ▷ Let $\bar{\mathbf{S}} = [\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_{|\bar{\mathcal{S}}|}}] \in \mathbb{R}^{|\mathcal{X}|\times|\bar{\mathcal{S}}|}$ be the sampling matrix corresponding to $\bar{\mathcal{S}}$.
    ▷ $\mathbf{e}_i \in \mathbb{R}^{|\mathcal{X}|}$ denotes the unit vector whose i-th element is only 1.
$\widetilde{\mathbf{S}} \leftarrow$ RecursiveRLS-Nyström$(\bar{\mathcal{X}}, k, \lambda, \delta/3)$.
$\widehat{\mathbf{S}} \coloneqq \bar{\mathbf{S}}\widetilde{\mathbf{S}}$.
Set $\widetilde{l}_i^{\lambda} \coloneqq \frac{3}{2\lambda}\left(\mathbf{K} - \mathbf{K}\widehat{\mathbf{S}}\left(\widehat{\mathbf{S}}^T\mathbf{K}\widehat{\mathbf{S}} + \lambda\mathbf{I}\right)^{-1}\widehat{\mathbf{S}}^T\mathbf{K}\right)_{i,i}$ for each $i \in \{1, \ldots, |\mathcal{X}|\}$.
Set $p_i \coloneqq \min\{1, \widetilde{l}_i^{\lambda} \cdot 16\log(\sum \widetilde{l}_i^{\lambda}/\delta)\}$ for each $i \in \{1, \ldots, |\mathcal{X}|\}$.
Initially set weighted sampling matrix $\mathbf{S}$ to be empty. For each $i \in \{1, \ldots, |\mathcal{X}|\}$, with probability p_i, append the column $\frac{1}{\sqrt{p_i}}\mathbf{e}_i$ onto $\mathbf{S}$.
return $\mathbf{S}$.

C Proof of Lemma 3.3

Proof of Lemma 3.3. Firstly, note that the following identity of the posterior variance holds (Lemma H.3):

$$\sigma^2(\boldsymbol{x}; \mathbf{X}^{(T)}, \lambda) = \lambda\psi(\boldsymbol{x})^\top\left(\sum_{t=1}^{T}\psi(\boldsymbol{x}^{(t)})\psi(\boldsymbol{x}^{(t)})^\top + \lambda\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\psi(\boldsymbol{x}). \tag{21}$$

Here, from the definition of $\|\psi(\boldsymbol{z})\|^2_{G(T,\lambda,P)^{-1}}$, we have

$$\|\psi(\boldsymbol{z})\|^2_{G(T,\lambda,P)^{-1}} = T\psi(\boldsymbol{z})^\top\left(T\sum_{\boldsymbol{x}\in\mathcal{X}}P(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \lambda\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\psi(\boldsymbol{z}) \tag{22}$$

$$= T\psi(\boldsymbol{z})^\top\left(\mathbb{E}_{\mathbf{X}^{(T)}}\left[\sum_{t=1}^{T}\psi(\boldsymbol{x}^{(t)})\psi(\boldsymbol{x}^{(t)})^\top + \lambda\boldsymbol{I}_{|\mathcal{X}|}\right]\right)^{-1}\psi(\boldsymbol{z}). \tag{23}$$

Algorithm 4 Approximated-MVR($\mathcal{X}$, k, λ, M, δ)

input: Input domain $\mathcal{X}$, kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, regularization parameter $\lambda > 0$, total budget $M \in \mathbb{N}_+$, confidence level $\delta \in (0, \min\{1, M/32\})$.

output: approximated MVR-sequence $(\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})})_{t\in[M]}$.

$\widetilde{\mathbf{X}}_0^{(\mathrm{MVR})} \leftarrow \emptyset$.
for $t = 1, 2, \ldots, M$ **do**
  Calculate sampling matrix: $\widetilde{\mathbf{S}}_t \leftarrow \mathrm{RecursiveRLS\text{-}Nyström}(\widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, k, \lambda, \delta/M)$
  Calculate approximated posterior variance $\bar{\sigma}^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right)$:

$$\bar{\sigma}^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right) = k(\boldsymbol{x}, \boldsymbol{x}) - \bar{\boldsymbol{\phi}}_t(\boldsymbol{x})^\top \bar{\boldsymbol{\phi}}_t(\boldsymbol{x}) + \lambda \bar{\boldsymbol{\phi}}_t(\boldsymbol{x})^\top \left(\sum_{i=1}^{t-1} \bar{\boldsymbol{\phi}}_t(\widetilde{\boldsymbol{x}}_i^{(\mathrm{MVR})}) \bar{\boldsymbol{\phi}}_t(\widetilde{\boldsymbol{x}}_i^{(\mathrm{MVR})})^\top + \lambda \boldsymbol{I}_{\widetilde{s}_t}\right)^{-1} \bar{\boldsymbol{\phi}}_t(\boldsymbol{x}). \tag{12}$$

    ▷ $\bar{\boldsymbol{\phi}}_t(\boldsymbol{x}) \coloneqq [(\widetilde{\mathbf{S}}_t^\top \mathbf{K}_t \widetilde{\mathbf{S}}_t)^{1/2}]^\dagger \widetilde{\mathbf{S}}_t^\top \boldsymbol{k}_t(\boldsymbol{x}) \in \mathbb{R}^{\widetilde{s}_t}$ with $\widetilde{\mathbf{S}}_t \in \mathbb{R}^{(t-1)\times \widetilde{s}_t}$.
    ▷ $\mathbf{K}_t \coloneqq \mathbf{K}\left(\widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}\right) \in \mathbb{R}^{(t-1)\times(t-1)}$, $\boldsymbol{k}_t(\boldsymbol{x}) \coloneqq \boldsymbol{k}\left(\boldsymbol{x}, \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}\right) \in \mathbb{R}^{t-1}$.
  Calculate $\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})} \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \bar{\sigma}^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right)$.
  Update training input: $\widetilde{\mathbf{X}}_t^{(\mathrm{MVR})} \leftarrow (\widetilde{\boldsymbol{x}}_1^{(\mathrm{MVR})}, \ldots, \widetilde{\boldsymbol{x}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})})$.
end for
return $(\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})})_{t\in[M]}$.

Applying Jensen's inequality for a positive definite matrix by noting the convexity of the matrix inversion, we obtain

$$\|\psi(\boldsymbol{z})\|^2_{G(T,\lambda,P)^{-1}} \le T\mathbb{E}_{\mathbf{X}^{(T)}}\left[\psi(\boldsymbol{z})^\top \left(\sum_{t=1}^{T} \psi(\boldsymbol{x}^{(t)})\psi(\boldsymbol{x}^{(t)})^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \psi(\boldsymbol{z})\right] \tag{24}$$

$$= \frac{T}{\lambda}\mathbb{E}_{\mathbf{X}^{(T)}}\left[\sigma^2(\boldsymbol{z}; \mathbf{X}^{(T)}, \lambda)\right], \tag{25}$$

where the last equality follows from the identity in Eq. (21). Next, by taking the expectation, we immediately obtain the first inequality $\mathbb{E}_{\boldsymbol{x}^{(T+1)}}\left[\|\psi(\boldsymbol{x}^{(T+1)})\|^2_{G(T,\lambda,P)^{-1}}\right] \le \frac{T}{\lambda}\mathbb{E}_{\mathbf{X}^{(T)},\boldsymbol{x}^{(T+1)}}\left[\sigma^2(\boldsymbol{x}^{(T+1)}; \mathbf{X}^{(T)}, \lambda)\right]$ in Eq. (8) from Eq. (25). Regarding the second inequality in Eq. (8), we have

$$T\mathbb{E}_{\mathbf{X}^{(T)},\boldsymbol{x}^{(T+1)}}\left[\sigma^2(\boldsymbol{x}^{(T+1)}; \mathbf{X}^{(T)}, \lambda)\right] = \mathbb{E}_{\mathbf{X}^{(T)},\boldsymbol{x}^{(T+1)}}\left[\sum_{t=1}^{T} \sigma^2(\boldsymbol{x}^{(T+1)}; \mathbf{X}^{(T)}, \lambda)\right] \tag{26}$$

$$\le \mathbb{E}_{\mathbf{X}^{(T)},\boldsymbol{x}^{(T+1)}}\left[\sum_{t=1}^{T} \sigma^2(\boldsymbol{x}^{(T+1)}; \mathbf{X}^{(t-1)}, \lambda)\right], \tag{27}$$

where the last inequality follows from the monotonicity of the posterior variance with respect to training input data. Since $\boldsymbol{x}^{(T+1)}$ and $\boldsymbol{x}^{(t)}$ are identically distributed, we have $\mathbb{E}_{\mathbf{X}^{(T)},\boldsymbol{x}^{(T+1)}}\left[\sum_{t=1}^{T} \sigma^2(\boldsymbol{x}^{(T+1)}; \mathbf{X}^{(t-1)}, \lambda)\right] = \mathbb{E}_{\mathbf{X}^{(T)},\boldsymbol{x}^{(T+1)}}\left[\sum_{t=1}^{T} \sigma^2(\boldsymbol{x}^{(t)}; \mathbf{X}^{(t-1)}, \lambda)\right] \le 2\gamma_T/\log(1 + \lambda^{-1})$, where the final inequality follows from Eq. (3)[10]. This concludes the second inequality in Eq. (8). Finally, the inequality $2\lambda^{-1}\gamma_T/\log(1 + \lambda^{-1}) \le 4\gamma_T$ for $\lambda \ge 1$ is the direct consequence of the elementary inequality $\forall a \in [0, 1], \log(1 + a) \ge a/2$. □

D Proof of Theorem 3.1

From Lemma H.1, each reward function f_t can be rewritten as $f_t(\cdot) = \psi(\cdot)^\top \boldsymbol{\theta}_t$, where $\boldsymbol{\theta}_t$ is drawn by the environment. Note that $\boldsymbol{\theta}_t$ and $\boldsymbol{x}_t$ are conditionally independent given $\mathcal{H}_{t-1}$ as defined in

[10]Note that Eq. (3) requires the condition $\forall \boldsymbol{x} \in \mathcal{X}$, $k(\boldsymbol{x}, \boldsymbol{x}) \le 1$. See Lemma 5.4 in [60].

Algorithm 5 RLS-kernelized Exp3 for adversarial kernelized bandits

Require: Finite input domain $\mathcal{X}$, kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, learning rate $\eta > 0$, regularization parameter $\lambda \geq 1$, mixing ratio $\alpha \in (0, 1)$, confidence width parameter $\beta > 0$, total budget $T \in \mathbb{N}_+$, RKHS norm upper bound $B \in (0, \infty)$, confidence level parameter $\delta \in (0, \min\{1, 3T/32\})$.

Prepare the exploration distribution $\pi(\boldsymbol{x})$ by using MVR-input sequence $(\boldsymbol{x}_t^{(\mathrm{MVR})})_{t \leq \lceil T\alpha \rceil}$:

$$\pi(\boldsymbol{x}) = \frac{1}{\lceil T\alpha \rceil} \sum_{t=1}^{\lceil T\alpha \rceil} \mathbb{1}\left\{\boldsymbol{x}_t^{(\mathrm{MVR})} = \boldsymbol{x}\right\}, \tag{13}$$

where $\boldsymbol{x}_t^{(\mathrm{MVR})} \in \mathrm{argmax}_{\boldsymbol{x} \in \mathcal{X}} \sigma^2\left(\boldsymbol{x}; \mathbf{X}_{t-1}^{(\mathrm{MVR})}, \lambda\right)$ and $\mathbf{X}_{t-1}^{(\mathrm{MVR})} = \left(\boldsymbol{x}_1^{(\mathrm{MVR})}, \ldots, \boldsymbol{x}_{t-1}^{(\mathrm{MVR})}\right)$.

Initialize sampling distribution: $P_1(\boldsymbol{x}) \leftarrow \alpha\pi(\boldsymbol{x}) + (1-\alpha)\widetilde{P}_1(\boldsymbol{x})$, where $\widetilde{P}_1(\boldsymbol{x}) = \frac{1}{|\mathcal{X}|}$.

for $t = 1, 2, \ldots, T$ **do**

  Draw $\boldsymbol{x}_t \sim P_t$.

  Receive reward $f_t(\boldsymbol{x}_t)$.

  Construct weighted kernel function $\widetilde{k}_t(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{P_t(\boldsymbol{x})P_t(\boldsymbol{x}')}k(\boldsymbol{x}, \boldsymbol{x}')$.

  Calculate the sampling matrix $\boldsymbol{S}_t$ by running Alg. 3: $\boldsymbol{S}_t \leftarrow$ RecursiveRLS-Nyström$(\mathcal{X}, \widetilde{k}_t, \frac{\lambda}{T}, \frac{\delta}{3T})$.

  Calculate the reward estimator $\widetilde{f}_t(\cdot)$ as follows:

$$\widetilde{f}_t(\boldsymbol{x}) = \frac{f_t(\boldsymbol{x}_t)}{\sqrt{P_t(\boldsymbol{x})P_t(\boldsymbol{x}_t)}}\phi_t(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}' \in \mathcal{X}} \phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_{s_t}\right)^{-1} \phi_t(\boldsymbol{x}_t). \tag{14}$$

  Calculate optimistic estimate $\widetilde{u}_t$ as $\widetilde{u}_t(\boldsymbol{x}) = \widetilde{f}_t(\boldsymbol{x}) + \beta\widetilde{\sigma}_t(\boldsymbol{x})$, where

$$\begin{aligned}\widetilde{\sigma}_t^2(\boldsymbol{x}) = {} & \frac{T}{\lambda P_t(\boldsymbol{x})}\left(\widetilde{k}_t(\boldsymbol{x}, \boldsymbol{x}) - \phi_t(\boldsymbol{x})^\top\phi_t(\boldsymbol{x})\right) \\ & + \frac{1}{P_t(\boldsymbol{x})}\phi_t(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}' \in \mathcal{X}} \phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_{s_t}\right)^{-1} \phi_t(\boldsymbol{x}).\end{aligned} \tag{15}$$

  Update the sampling distribution:

$$P_{t+1}(\boldsymbol{x}) = \alpha\pi(\boldsymbol{x}) + (1-\alpha)\widetilde{P}_{t+1}(\boldsymbol{x}), \text{ where } \widetilde{P}_{t+1}(\boldsymbol{x}) = \frac{\exp\left(\eta \sum_{i=1}^t \widetilde{u}_i(\boldsymbol{x})\right)}{\sum_{\boldsymbol{x}' \in \mathcal{X}} \exp\left(\eta \sum_{i=1}^t \widetilde{u}_i(\boldsymbol{x}')\right)}. \tag{16}$$

end for

the problem setting. Here, we first recall the following three basic facts used in various parts of our proof: (i) $\mathbb{E}[g(\boldsymbol{x}_t)|\mathcal{H}_{t-1}] = \sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x})g(\boldsymbol{x})$ holds for any (measurable) fixed function g, (ii) $\mathbb{E}[g(\boldsymbol{x}_t)|\mathcal{H}_{t-1}] = \mathbb{E}[g(\boldsymbol{x}_t)|\mathcal{H}_{t-1}, \boldsymbol{\theta}_t]$ holds from the conditional independence of $\boldsymbol{x}_t$ and $\boldsymbol{\theta}_t$ given $\mathcal{H}_{t-1}$, and (iii) $\mathbb{E}[h(\mathcal{H}_{t-1}, \boldsymbol{\theta}_t)g(\boldsymbol{x}_t)|\mathcal{H}_{t-1}, \boldsymbol{\theta}_t] = h(\mathcal{H}_{t-1}, \boldsymbol{\theta}_t)\mathbb{E}[g(\boldsymbol{x}_t)|\mathcal{H}_{t-1}, \boldsymbol{\theta}_t]$ for any (measurable) fixed functions g and h.

Next, from the tower property of conditional expectation, we have $\mathbb{E}[\sum_{t=1}^T f_t(\boldsymbol{x}_t)] = \sum_{t=1}^T \mathbb{E}[\mathbb{E}[f_t(\boldsymbol{x}_t) \mid \mathcal{H}_{t-1}, \boldsymbol{\theta}_t]] = \sum_{t=1}^T \mathbb{E}[\sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x})f_t(\boldsymbol{x})]$, where the second equality follows from

$$\begin{aligned}\mathbb{E}[f_t(\boldsymbol{x}_t) \mid \mathcal{H}_{t-1}, \boldsymbol{\theta}_t] &= \mathbb{E}[\boldsymbol{\theta}_t^\top\psi(\boldsymbol{x}_t) \mid \mathcal{H}_{t-1}, \boldsymbol{\theta}_t] = \boldsymbol{\theta}_t^\top\mathbb{E}[\psi(\boldsymbol{x}_t) \mid \mathcal{H}_{t-1}, \boldsymbol{\theta}_t] &(28)\\ &= \boldsymbol{\theta}_t^\top\mathbb{E}[\psi(\boldsymbol{x}_t) \mid \mathcal{H}_{t-1}] = \sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x})\boldsymbol{\theta}_t^\top\psi(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x})f_t(\boldsymbol{x}). &(29)\end{aligned}$$

Thus, we have

$$\bar{R}_T = \sum_{t=1}^T \mathbb{E}\left[f_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x})f_t(\boldsymbol{x})\right], \tag{30}$$

Algorithm 6 RLS-kernelized Exp3 for adversarial kernelized bandits with approximated MVR

Require: Finite input domain $\mathcal{X}$, kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, learning rate $\eta > 0$, regularization parameter $\lambda \geq 1$, mixing ratio $\alpha \in (0, 1)$, confidence width parameter $\beta > 0$, total budget $T \in \mathbb{N}_+$, RKHS norm upper bound $B \in (0, \infty)$, confidence level parameter $\delta \in (0, \min\{1, 3\lceil T\alpha \rceil / 16\})$.

Prepare the exploration distribution $\widetilde{\pi}(\boldsymbol{x})$ by using approximated MVR-sequence as follows:

$$\widetilde{\pi}(\boldsymbol{x}) = \frac{1}{\lceil T\alpha \rceil} \sum_{t=1}^{\lceil T\alpha \rceil} \mathbb{1}\left\{\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})} = \boldsymbol{x}\right\}, \tag{17}$$

where $(\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})})_{t \in [\lceil T\alpha \rceil]} \leftarrow$ Approximated-MVR$(\mathcal{X}, k, \lambda, \lceil T\alpha \rceil, \delta/6)$ (Alg. 4).

Initialize sampling distribution: $P_1(\boldsymbol{x}) \leftarrow \alpha\widetilde{\pi}(\boldsymbol{x}) + (1-\alpha)\widetilde{P}_1(\boldsymbol{x})$, where $\widetilde{P}_1(\boldsymbol{x}) = \frac{1}{|\mathcal{X}|}$.

for $t = 1, 2, \ldots, T$ **do**

Draw $\boldsymbol{x}_t \sim P_t$.

Receive reward $f_t(\boldsymbol{x}_t)$.

Construct weighted kernel function $\widetilde{k}_t(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{P_t(\boldsymbol{x})P_t(\boldsymbol{x}')}k(\boldsymbol{x}, \boldsymbol{x}')$.

Calculate the sampling matrix $\boldsymbol{S}_t$ by running Alg. 3: $\boldsymbol{S}_t \leftarrow$ RecursiveRLS-Nyström$(\mathcal{X}, \widetilde{k}_t, \frac{\lambda}{T}, \frac{\delta}{6T})$.

Calculate the reward estimator $\widetilde{f}_t(\cdot)$ as follows:

$$\widetilde{f}_t(\boldsymbol{x}) = \frac{f_t(\boldsymbol{x}_t)}{\sqrt{P_t(\boldsymbol{x})P_t(\boldsymbol{x}_t)}}\phi_t(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}' \in \mathcal{X}} \phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_{s_t}\right)^{-1} \phi_t(\boldsymbol{x}_t). \tag{18}$$

Calculate optimistic estimate $\widetilde{u}_t$ as $\widetilde{u}_t(\boldsymbol{x}) = \widetilde{f}_t(\boldsymbol{x}) + \beta\widetilde{\sigma}_t(\boldsymbol{x})$, where

$$\begin{aligned}\widetilde{\sigma}_t^2(\boldsymbol{x}) = {} & \frac{T}{\lambda P_t(\boldsymbol{x})}\left(\widetilde{k}_t(\boldsymbol{x}, \boldsymbol{x}) - \phi_t(\boldsymbol{x})^\top\phi_t(\boldsymbol{x})\right) \\ & + \frac{1}{P_t(\boldsymbol{x})}\phi_t(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}' \in \mathcal{X}} \phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_{s_t}\right)^{-1} \phi_t(\boldsymbol{x}).\end{aligned} \tag{19}$$

Update the sampling distribution:

$$P_{t+1}(\boldsymbol{x}) = \alpha\widetilde{\pi}(\boldsymbol{x}) + (1-\alpha)\widetilde{P}_{t+1}(\boldsymbol{x}), \text{ where } \widetilde{P}_{t+1}(\boldsymbol{x}) = \frac{\exp\left(\eta \sum_{i=1}^t \widetilde{u}_i(\boldsymbol{x})\right)}{\sum_{\boldsymbol{x}' \in \mathcal{X}} \exp\left(\eta \sum_{i=1}^t \widetilde{u}_i(\boldsymbol{x}')\right)}. \tag{20}$$

end for

where $\boldsymbol{x}^* \in \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}[\sum_{t=1}^T f_t(\boldsymbol{x})]$. Note that $\boldsymbol{x}^*$ is always well-defined since $|\mathcal{X}| < \infty$. Furthermore,

$$\mathbb{E}\left[f_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x})f_t(\boldsymbol{x})\right] \tag{31}$$

$$= \alpha\mathbb{E}\left[f_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x} \in \mathcal{X}} \pi(\boldsymbol{x})f_t(\boldsymbol{x})\right] + (1-\alpha)\mathbb{E}\left[f_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_t(\boldsymbol{x})f_t(\boldsymbol{x})\right] \tag{32}$$

$$\leq 2\alpha B + (1-\alpha)\mathbb{E}\left[f_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_t(\boldsymbol{x})f_t(\boldsymbol{x})\right], \tag{33}$$

where the last inequality follows from $\|f_t\|_\infty \leq \|f_t\|_k \leq B$. Here, we further decompose the second term of the above equation by using the optimistic estimate $u_t(\boldsymbol{x})$ as follows:

$$\mathbb{E}\left[f_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) f_t(\boldsymbol{x})\right] \tag{34}$$

$$= \underbrace{\mathbb{E}\left[f_t(\boldsymbol{x}^*) - u_t(\boldsymbol{x}^*)\right]}_{\text{(i)}} + \underbrace{\mathbb{E}\left[u_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) u_t(\boldsymbol{x})\right]}_{\text{(ii)}} + \underbrace{\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(u_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right]}_{\text{(iii)}}. \tag{35}$$

Upper bound for the first term (i) in Eq. (35)**.** From the definition of $u_t(\boldsymbol{x}^*)$, we have

$$f_t(\boldsymbol{x}^*) - u_t(\boldsymbol{x}^*) = f_t(\boldsymbol{x}^*) - \widehat{f}_t(\boldsymbol{x}^*) - \beta\|\psi(\boldsymbol{x}^*)\|_{G_t(\lambda)^{-1}}. \tag{36}$$

By noting the definition of $\widehat{f}_t$ and $f_t(\cdot) = \psi(\cdot)^\top \boldsymbol{\theta}_t$ (Lemma H.1), the error term $f_t(\boldsymbol{x}^*) - \widehat{f}_t(\boldsymbol{x}^*)$ in the above equation can be rewritten as follows:

$$f_t(\boldsymbol{x}^*) - \widehat{f}_t(\boldsymbol{x}^*) = \psi(\boldsymbol{x}^*)^\top \boldsymbol{\theta}_t - \psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-1} \psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t) \tag{37}$$

$$= \psi(\boldsymbol{x}^*)^\top \boldsymbol{\theta}_t - \psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-1} \psi(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top \boldsymbol{\theta}_t \tag{38}$$

$$= \psi(\boldsymbol{x}^*)^\top \left(\boldsymbol{I}_{|\mathcal{X}|} - G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top\right)\boldsymbol{\theta}_t. \tag{39}$$

By noting that $\boldsymbol{\theta}_t$ is conditionally independent of $\boldsymbol{x}_t$ given $\mathcal{H}_{t-1}$, we have

$$\mathbb{E}[f_t(\boldsymbol{x}^*) - \widehat{f}_t(\boldsymbol{x}^*) \mid \mathcal{H}_{t-1}, \boldsymbol{\theta}_t] = \psi(\boldsymbol{x}^*)^\top \left(\boldsymbol{I}_{|\mathcal{X}|} - G_t(\lambda)^{-1}\mathbb{E}[\psi(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top \mid \mathcal{H}_{t-1}, \boldsymbol{\theta}_t]\right)\boldsymbol{\theta}_t \tag{40}$$

$$= \psi(\boldsymbol{x}^*)^\top \left[G_t(\lambda)^{-1}\left(G_t(\lambda) - \sum_{\boldsymbol{x}\in\mathcal{X}} P_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top\right)\right]\boldsymbol{\theta}_t \tag{41}$$

$$= \frac{\lambda}{T}\psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-1}\boldsymbol{\theta}_t, \tag{42}$$

where the last line follows from the definition of $G_t(\lambda) \coloneqq \sum_{\boldsymbol{x}\in\mathcal{X}} P_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$. From the Cauchy–Schwarz inequality and $\|\boldsymbol{\theta}_t\|_2 = \|f_t\|_k \leq B$ (Lemma H.1), we have

$$\frac{\lambda}{T}\psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-1}\boldsymbol{\theta}_t \leq \frac{\lambda}{T}\|\psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-1}\|_2\|\boldsymbol{\theta}_t\|_2 \tag{43}$$

$$\leq B\frac{\lambda}{T}\sqrt{\psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-2}\psi(\boldsymbol{x}^*)} \tag{44}$$

$$\leq B\sqrt{\frac{\lambda}{T}}\sqrt{\psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}^*)} \tag{45}$$

$$= B\sqrt{\frac{\lambda}{T}}\|\psi(\boldsymbol{x}^*)\|_{G_t(\lambda)^{-1}}, \tag{46}$$

where the last inequality follows from $\psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-2}\psi(\boldsymbol{x}^*) \leq \frac{T}{\lambda}\psi(\boldsymbol{x}^*)^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}^*)$, which is implied by the fact that the maximum eigenvalue of $G_t(\lambda)^{-1}$ is bounded from above by T/λ. From Eqs. (42) and (46), we have

$$\mathbb{E}[f_t(\boldsymbol{x}^*) - u_t(\boldsymbol{x}^*)] = \mathbb{E}\left[f_t(\boldsymbol{x}^*) - \widehat{f}_t(\boldsymbol{x}^*) - \beta\|\psi(\boldsymbol{x}^*)\|_{G_t(\lambda)^{-1}}\right] \tag{47}$$

$$= \mathbb{E}[\mathbb{E}[f_t(\boldsymbol{x}^*) - \widehat{f}_t(\boldsymbol{x}^*) \mid \mathcal{H}_{t-1}, \boldsymbol{\theta}_t]] - \mathbb{E}[\beta\|\psi(\boldsymbol{x}^*)\|_{G_t(\lambda)^{-1}}] \tag{48}$$

$$\leq \mathbb{E}\left[\left(B\sqrt{\frac{\lambda}{T}} - \beta\right)\|\psi(\boldsymbol{x}^*)\|_{G_t(\lambda)^{-1}}\right] \tag{49}$$

$$= 0. \tag{50}$$

Thus, the first term (i) in Eq. (35) is bounded from above by 0.

Upper bound for the second term (ii) in Eq. (35)**.** For any $\boldsymbol{x} \in \mathcal{X}$, the definition of the exploration distribution implies the following upper bound of $u_t(\boldsymbol{x})$ (Lemma H.5):

$$u_t(\boldsymbol{x}) \leq \frac{2}{\alpha}(2B\gamma_T + \beta\sqrt{\gamma_T}). \tag{51}$$

By combining the above inequality with the definition of α, we can observe that $\eta u_t(\boldsymbol{x}) \leq 1$ holds for any $\boldsymbol{x} \in \mathcal{X}$; therefore, by following the standard proof for Exp3 (Lemma H.4), we obtain

$$\sum_{t=1}^{T} u_t(\boldsymbol{x}^*) - \sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) u_t(\boldsymbol{x}) \leq \frac{\log|\mathcal{X}|}{\eta} + \eta \sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) u_t^2(\boldsymbol{x}). \tag{52}$$

Regarding the term $\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) u_t^2(\boldsymbol{x})$, we further obtain the upper bound as follows:

$$\begin{aligned}
\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) u_t^2(\boldsymbol{x}) &= \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) \left(\widehat{f}_t(\boldsymbol{x}) + \beta\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}}\right)^2 && (53)\\
&\leq 2\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) \widehat{f}_t^2(\boldsymbol{x}) + 2\beta^2 \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) \|\psi(\boldsymbol{x})\|^2_{G_t(\lambda)^{-1}}, && (54)
\end{aligned}$$

where we used $\forall a, b \in \mathbb{R},\ (a+b)^2 \leq 2a^2 + 2b^2$ in the last inequality. Here, note the following relation:

$$(1-\alpha)\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|} \preceq \sum_{\boldsymbol{x}\in\mathcal{X}} P_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|} \coloneqq G_t(\lambda). \tag{55}$$

Thus, for $M_t(\lambda) \coloneqq \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{\lceil T(1-\alpha)\rceil}\boldsymbol{I}_{|\mathcal{X}|}$, we can confirm

$$G_t(\lambda)^{-1} \preceq \frac{1}{1-\alpha}\left(\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{T(1-\alpha)}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \preceq \frac{1}{1-\alpha} M_t(\lambda)^{-1}. \tag{56}$$

The above relation further implies

$$\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{G_t(\lambda)^{-1}} \leq \frac{1}{1-\alpha}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{M_t(\lambda)^{-1}} \leq \frac{4\gamma_{\lceil T(1-\alpha)\rceil}}{1-\alpha} \leq \frac{4\gamma_T}{1-\alpha}, \tag{57}$$

where the second inequality follows from Lemma 3.3. Therefore, by noting the definition of β, the second term in Eq. (54) is bounded from above as follows:

$$2\beta^2 \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{G_t(\lambda)^{-1}} \leq \frac{8\lambda B^2\gamma_T}{(1-\alpha)T}. \tag{58}$$

Next, regarding the first term $\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\widehat{f}_t^2(\boldsymbol{x})$ in Eq. (54), we have

$$\begin{aligned}
\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\widehat{f}_t^2(\boldsymbol{x}) &= \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left[\psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t)\right]^2 && (59)\\
&\leq B^2 \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left[\psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t)\right]^2 && (60)\\
&= \frac{B^2}{1-\alpha}\psi(\boldsymbol{x}_t)^\top G_t(\lambda)^{-1}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} (1-\alpha)\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top\right] G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t) && (61)\\
&\leq \frac{B^2}{1-\alpha}\psi(\boldsymbol{x}_t)^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t). && (62)
\end{aligned}$$

Thus, by aggregating Eqs. (52), (54), (58), and (62), we have

$$\mathbb{E}\left[\sum_{t=1}^{T} u_t(\boldsymbol{x}^*) - \sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) u_t(\boldsymbol{x})\right] \tag{63}$$

$$\leq \frac{\log|\mathcal{X}|}{\eta} + \frac{8\eta\lambda B^2\gamma_T}{1-\alpha} + \frac{2\eta B^2}{1-\alpha}\sum_{t=1}^{T}\mathbb{E}\left[\psi(\boldsymbol{x}_t)^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t)\right] \tag{64}$$

$$= \frac{\log|\mathcal{X}|}{\eta} + \frac{8\eta\lambda B^2\gamma_T}{1-\alpha} + \frac{2\eta B^2}{1-\alpha}\sum_{t=1}^{T}\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} P_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{G_t(\lambda)^{-1}}\right] \tag{65}$$

$$\leq \frac{\log|\mathcal{X}|}{\eta} + \frac{8\eta\lambda B^2\gamma_T}{1-\alpha} + \frac{8\eta B^2 T\gamma_T}{1-\alpha}, \tag{66}$$

where Eq. (65) follows from the tower property of conditional expectation, and Eq. (66) follows from Lemma 3.3.

Upper bound for the third term (iii) in Eq. (35)**.** By noting the definition of u_t, we have

$$\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(u_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right) = \beta\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}} + \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widehat{f}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right). \tag{67}$$

Regarding the first term in Eq. (67), the applications of Jensen's inequality for the square root $\sqrt{x}$ and Eq. (57) imply

$$\beta\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}} \leq \beta\sqrt{\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{G_t(\lambda)^{-1}}} \tag{68}$$

$$\leq 2B\sqrt{\frac{\lambda\gamma_T}{(1-\alpha)T}}. \tag{69}$$

Regarding the second term in Eq. (67), we have

$$\widehat{f}_t(\boldsymbol{x}) - f_t(\boldsymbol{x}) = \psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}) \tag{70}$$

$$= \psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top\boldsymbol{\theta}_t - \psi(\boldsymbol{x})^\top\boldsymbol{\theta}_t \tag{71}$$

$$= \psi(\boldsymbol{x})^\top\left(G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top - \boldsymbol{I}_{|\mathcal{X}|}\right)\boldsymbol{\theta}_t \tag{72}$$

$$= \psi(\boldsymbol{x})^\top\left[G_t(\lambda)^{-1}\left(\psi(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top - G_t(\lambda)\right)\right]\boldsymbol{\theta}_t. \tag{73}$$

Thus, we have the following inequality from the above equation:

$$\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widehat{f}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] \tag{74}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widehat{f}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\middle|\,\mathcal{H}_{t-1}, \boldsymbol{\theta}_t\right]\right] \tag{75}$$

$$= \mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\left[G_t(\lambda)^{-1}\left(\mathbb{E}\left[\psi(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top\,\middle|\,\mathcal{H}_{t-1}, \boldsymbol{\theta}_t\right] - G_t(\lambda)\right)\right]\boldsymbol{\theta}_t\right] \tag{76}$$

$$\leq \frac{\lambda}{T}\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} |\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\boldsymbol{\theta}_t|\right]. \tag{77}$$

The upper bound of the quantity in the above expectation can be further obtained as

$$\sum_{\boldsymbol{x}\in\mathcal{X}} |\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\boldsymbol{\theta}_t| \le \sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\sqrt{\psi(\boldsymbol{x})^\top G_t(\lambda)^{-2}\psi(\boldsymbol{x})}\|\boldsymbol{\theta}_t\|_2 \tag{78}$$

$$\le B\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\sqrt{\psi(\boldsymbol{x})^\top G_t(\lambda)^{-2}\psi(\boldsymbol{x})} \tag{79}$$

$$\le B\sqrt{\frac{T}{\lambda}}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\sqrt{\psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x})} \tag{80}$$

$$\le B\sqrt{\frac{T}{\lambda}}\sqrt{\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x})} \tag{81}$$

$$\le 2B\sqrt{\frac{T\gamma_T}{(1-\alpha)\lambda}}. \tag{82}$$

Thus, by aggregating Eqs. (67), (69), (77), and (82), the upper bound of the third term (iii) in Eq. (35) is obtained as follows:

$$\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(u_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] \le 4B\sqrt{\frac{\lambda\gamma_T}{(1-\alpha)T}}. \tag{83}$$

Aggregate all the upper bounds. From Eqs. (30), (33), (35), (50), (66), and (83), we finally obtain

$$\bar{R}_T \le 2\alpha BT + \frac{(1-\alpha)\log|\mathcal{X}|}{\eta} + 8\eta\lambda B^2\gamma_T + 8\eta B^2 T\gamma_T + 4B\sqrt{(1-\alpha)\lambda T\gamma_T} \tag{84}$$

$$\le 2\alpha BT + \frac{\log|\mathcal{X}|}{\eta} + 8\eta(\lambda+T)B^2\gamma_T + 4B\sqrt{\lambda T\gamma_T}. \tag{85}$$

□

E Proof of Theorem 4.1

Proof for computational complexity. Here, for any $t \in [T]$, we define the event A_t as follows:

$$\frac{1}{2}\left(\bar{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right) \preceq \bar{\boldsymbol{\Psi}}_t\mathbf{S}_t\mathbf{S}_t^\top\bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|} \preceq \frac{3}{2}\left(\bar{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right). \tag{86}$$

From Lemma H.7, we can confirm that $\mathbb{P}(A_t \mid \widetilde{\mathcal{H}}_{t-1}) \ge 1 - 1/T^3$ for any $t \in [T]$ and $\widetilde{\mathcal{H}}_{t-1} \coloneqq \{(\mathbf{S}_i, \boldsymbol{x}_i, f_i(\boldsymbol{x}_i))\}_{i\in[t-1]}$, which implies $\mathbb{P}(A_t) \ge 1 - 1/T^3$ due to the tower property of conditional expectation. By taking the union bound, we can obtain $\mathbb{P}(A) \ge 1 - 1/T^2$, where $A = \bigcap_{t\in[T]} A_t$. Furthermore, Lemma H.7 suggests that, with probability at least $1 - 1/T^2$, the following two statements simultaneously also hold:

1. The number s_t satisfies $s_t = O(\gamma_T \log(T\gamma_T)) = O(\gamma_T \log(T))$ for any $t \in [T]$[11].
2. The computational cost of recursive RLS-Nyström is $O(|\mathcal{X}|\gamma_T^2(\log(T\gamma_T))^2) = O(|\mathcal{X}|\gamma_T^2(\log T)^2)$.

Here, we confirm the computational cost of $(\widetilde{f}_t(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$ and $(\widetilde{\sigma}_t^2(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$ is $O(|\mathcal{X}|s_t^2 + s_t^3)$. Firstly, the computation of Nyström embedding $\phi_t(\cdot)$ requires the calculation of pseudo-inverse $[(\mathbf{S}_t^\top\widetilde{\mathbf{K}}_t\mathbf{S}_t)^{1/2}]^\dagger$, which requires $O(s_t^3)$-computation. Once $[(\mathbf{S}_t^\top\widetilde{\mathbf{K}}_t\mathbf{S}_t)^{1/2}]^\dagger$ is computed, the calculation of $\phi_t(\boldsymbol{x}) \coloneqq [(\mathbf{S}_t^\top\widetilde{\mathbf{K}}_t\mathbf{S}_t)^{1/2}]^\dagger\mathbf{S}_t^\top\widetilde{\boldsymbol{k}}_t(\boldsymbol{x})$ can be conducted within $O(s_t^2)$ for some given $\boldsymbol{x} \in \mathcal{X}$. Thus, the computational cost for the preparation of $(\phi_t(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$ is $O(|\mathcal{X}|s_t^2 + s_t^3)$. After calculating $(\phi_t(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$, we require $O(|\mathcal{X}|s_t^2)$ and $O(s_t^3)$ computations for calculating $\sum_{\boldsymbol{x}'\in\mathcal{X}}\phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top$ and $(\sum_{\boldsymbol{x}'\in\mathcal{X}}\phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \lambda/T\boldsymbol{I}_{s_t})^{-1}$, respectively. Once we have computed the inverse $(\sum_{\boldsymbol{x}'\in\mathcal{X}}\phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \lambda/T\boldsymbol{I}_{s_t})^{-1}$, the calculation of the quadratic form

[11] Note that $\gamma_T = O(T)$ always holds.

$\phi_t(\boldsymbol{x})^\top(\sum_{\boldsymbol{x}'\in\mathcal{X}}\phi_t(\boldsymbol{x}')\phi_t(\boldsymbol{x}')^\top + \lambda/T\boldsymbol{I}_{s_t})^{-1}\phi_t(\boldsymbol{x})$ for all $\boldsymbol{x}\in\mathcal{X}$ requires $O(|\mathcal{X}|s_t^2)$. In summary, given the sampling matrix $\mathbf{S}_t$, the overall computational cost for round t is $O(|\mathcal{X}|s_t^2 + s_t^3)$. By combining this with the above two high probability statements, we can confirm that the total computational cost at each round t is $\widetilde{O}(|\mathcal{X}|\gamma_T^2 + \gamma_T^3)$ with high probability.

Proof for regret upper bound. We first define the following notations, while some of them are already defined in Sec. 4:

- Let $\mathbf{S}_t \in \mathbb{R}^{|\mathcal{X}|\times s_t}$ be the sampling matrix at round t returned by recursive RLS-Nyström.
- Let $\bar{\boldsymbol{\Psi}}_t \coloneqq (\bar{\psi}_t(\boldsymbol{x}^{(1)}), \ldots, \bar{\psi}_t(\boldsymbol{x}^{(|\mathcal{X}|)})) \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ be the weighted feature matrix, where $\bar{\psi}_t(\boldsymbol{x}) \coloneqq \sqrt{P_t(\boldsymbol{x})}\psi(\boldsymbol{x})$ is the weighted feature map.
- Let $\mathbf{Q}_t \coloneqq \bar{\boldsymbol{\Psi}}_t\mathbf{S}_t\left[(\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t)^\top(\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t)\right]^\dagger(\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t)^\top$ be the projection matrix into the subspace spanned by the column vectors of $\bar{\boldsymbol{\Psi}}_t\mathbf{S}_t$. We also define $\widetilde{\psi}_t(\boldsymbol{x})$ as $\widetilde{\psi}_t(\boldsymbol{x}) = \mathbf{Q}_t\psi(\boldsymbol{x})$.
- Let $\widetilde{G}_t(\lambda)$ be $\widetilde{G}_t(\lambda) = \mathbf{Q}_t\bar{\boldsymbol{\Psi}}_t(\mathbf{Q}_t\bar{\boldsymbol{\Psi}}_t)^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$, which is the approximated version of $G_t(\lambda) \coloneqq \bar{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|} = \sum_{\boldsymbol{x}\in\mathcal{X}}P_t(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$.
- Let $\widetilde{\boldsymbol{\Psi}}_t \coloneqq \mathbf{Q}_t\bar{\boldsymbol{\Psi}}_t$ be the projected version of $\bar{\boldsymbol{\Psi}}_t$.

In the proof of this section, we define $\widetilde{\mathcal{H}}_t$ as the learner's history up to round t: $\widetilde{\mathcal{H}}_t = \{(\mathbf{S}_i, \boldsymbol{x}_i, f_i(\boldsymbol{x}_i))\}_{i\in[t]}$. Here, from the algorithm construction, note that $\boldsymbol{\theta}_t$, $\boldsymbol{x}_t$, and $\mathbf{S}_t$ are conditionally independent given $\widetilde{\mathcal{H}}_{t-1}$. As described in Sec. 4, note that the following identities hold (see also Lemma H.6):

$$\widetilde{f}_t(\boldsymbol{x}) = \psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top\boldsymbol{\theta}_t, \tag{87}$$

$$\widetilde{u}_t(\boldsymbol{x}) = \widetilde{f}_t(\boldsymbol{x}) + \beta\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}. \tag{88}$$

To prove Theorem 4.1, as with the proof of Theorem 3.1, we decompose $f_t(\boldsymbol{x}^*) - f_t(\boldsymbol{x}_t)$ as follows:

$$\mathbb{E}\left[f_t(\boldsymbol{x}^*) - f_t(\boldsymbol{x}_t)\right] \le 2\alpha B + (1-\alpha)\mathbb{E}\left[f_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})f_t(\boldsymbol{x})\right], \tag{89}$$

where $\boldsymbol{x}^* \in \operatorname{argmax}_{\boldsymbol{x}\in\mathcal{X}}\mathbb{E}[\sum_{t=1}^T f_t(\boldsymbol{x})]$. Now, we decompose the second term in Eq. (89) into the following three terms similarly to the proof of Theorem 3.1:

$$\mathbb{E}\left[f_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})f_t(\boldsymbol{x})\right] \tag{90}$$

$$= \underbrace{\mathbb{E}\left[f_t(\boldsymbol{x}^*) - \widetilde{u}_t(\boldsymbol{x}^*)\right]}_{\text{(i)}} + \underbrace{\mathbb{E}\left[\widetilde{u}_t(\boldsymbol{x}^*) - \sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t(\boldsymbol{x})\right]}_{\text{(ii)}} + \underbrace{\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\left(\widetilde{u}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right]}_{\text{(iii)}}. \tag{91}$$

Below, we consider the upper bound of terms (i), (ii), and (iii).

Upper bound for the first term (i) in Eq. (91)**.** By noting the definition of $\widetilde{f}_t$ and $f_t(\cdot) = \psi(\cdot)^\top\boldsymbol{\theta}_t$, the error term $f_t(\boldsymbol{x}^*) - \widetilde{f}_t(\boldsymbol{x}^*)$ in the above equation can be rewritten as follows:

$$f_t(\boldsymbol{x}^*) - \widetilde{f}_t(\boldsymbol{x}^*) = \psi(\boldsymbol{x}^*)^\top\boldsymbol{\theta}_t - \psi(\boldsymbol{x}^*)^\top\widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t)f_t(\boldsymbol{x}_t) \tag{92}$$

$$= \psi(\boldsymbol{x}^*)^\top\boldsymbol{\theta}_t - \psi(\boldsymbol{x}^*)^\top\widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top\boldsymbol{\theta}_t \tag{93}$$

$$= \psi(\boldsymbol{x}^*)^\top\left(\boldsymbol{I}_{|\mathcal{X}|} - \widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top\right)\boldsymbol{\theta}_t. \tag{94}$$

Then, we have

$$\mathbb{E}[f_t(\boldsymbol{x}^*) - \widetilde{f}_t(\boldsymbol{x}^*) \mid \widetilde{\mathcal{H}}_{t-1}, \mathbf{S}_t, \boldsymbol{\theta}_t] \tag{95}$$

$$= \psi(\boldsymbol{x}^*)^\top \left(\boldsymbol{I}_{|\mathcal{X}|} - \widetilde{G}_t(\lambda)^{-1} \mathbb{E}[\widetilde{\psi}_t(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top \mid \widetilde{\mathcal{H}}_{t-1}, \mathbf{S}_t, \boldsymbol{\theta}_t] \right) \boldsymbol{\theta}_t \tag{96}$$

$$= \psi(\boldsymbol{x}^*)^\top \left[\widetilde{G}_t(\lambda)^{-1} \left(\widetilde{G}_t(\lambda) - \sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x}) \widetilde{\psi}_t(\boldsymbol{x}) \psi(\boldsymbol{x})^\top \right) \right] \boldsymbol{\theta}_t \tag{97}$$

$$= \psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1} \left(\widetilde{\boldsymbol{\Psi}}_t \widetilde{\boldsymbol{\Psi}}_t^\top - \widetilde{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right) \boldsymbol{\theta}_t \tag{98}$$

$$= \frac{\lambda}{T} \psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1} \boldsymbol{\theta}_t + \psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1} \left(\widetilde{\boldsymbol{\Psi}}_t \widetilde{\boldsymbol{\Psi}}_t^\top - \widetilde{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top \right) \boldsymbol{\theta}_t. \tag{99}$$

Regarding the first term in Eq. (99), we have

$$\frac{\lambda}{T} \psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1} \boldsymbol{\theta}_t \leq \frac{\lambda}{T} \|\psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1}\|_2 \|\boldsymbol{\theta}_t\|_2 \tag{100}$$

$$\leq B \frac{\lambda}{T} \sqrt{\psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-2} \psi(\boldsymbol{x}^*)} \tag{101}$$

$$\leq B \sqrt{\frac{\lambda}{T}} \sqrt{\psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1} \psi(\boldsymbol{x}^*)} \tag{102}$$

$$= B \sqrt{\frac{\lambda}{T}} \|\psi(\boldsymbol{x}^*)\|_{\widetilde{G}_t(\lambda)^{-1}}. \tag{103}$$

Regarding the second term in Eq. (99), under $(\widetilde{\mathcal{H}}_{t-1}, \boldsymbol{\theta}_t, \boldsymbol{S}_t)$ such that event A_t is true, we have

$$\psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1} \left(\widetilde{\boldsymbol{\Psi}}_t \widetilde{\boldsymbol{\Psi}}_t^\top - \widetilde{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top \right) \boldsymbol{\theta}_t \tag{104}$$

$$= \psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1} \left(\widetilde{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top \mathbf{Q}_t^\top - \widetilde{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top \right) \boldsymbol{\theta}_t \tag{105}$$

$$= \psi(\boldsymbol{x}^*)^\top \widetilde{G}_t(\lambda)^{-1} \widetilde{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top \left(\mathbf{Q}_t - \boldsymbol{I}_{|\mathcal{X}|} \right) \boldsymbol{\theta}_t \tag{106}$$

$$\leq \|\psi(\boldsymbol{x}^*)\|_{\widetilde{G}_t(\lambda)^{-1}} \|\widetilde{G}_t(\lambda)^{-1/2} \widetilde{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top \left(\mathbf{Q}_t - \boldsymbol{I}_{|\mathcal{X}|} \right) \boldsymbol{\theta}_t\|_2 \tag{107}$$

$$\leq B \|\psi(\boldsymbol{x}^*)\|_{\widetilde{G}_t(\lambda)^{-1}} \|\widetilde{G}_t(\lambda)^{-1/2} \widetilde{\boldsymbol{\Psi}}_t\| \|\bar{\boldsymbol{\Psi}}_t^\top \left(\mathbf{Q}_t - \boldsymbol{I}_{|\mathcal{X}|} \right)\| \tag{108}$$

$$\leq B \|\psi(\boldsymbol{x}^*)\|_{\widetilde{G}_t(\lambda)^{-1}} \|\bar{\boldsymbol{\Psi}}_t^\top \left(\mathbf{Q}_t - \boldsymbol{I}_{|\mathcal{X}|} \right)\| \tag{109}$$

$$\leq B \sqrt{\frac{2\lambda}{T}} \|\psi(\boldsymbol{x}^*)\|_{\widetilde{G}_t(\lambda)^{-1}}, \tag{110}$$

where Eq. (109) follows from $\|\widetilde{G}_t(\lambda)^{-1/2} \widetilde{\boldsymbol{\Psi}}_t\| = \sqrt{\|\widetilde{\boldsymbol{\Psi}}_t^\top \widetilde{G}_t(\lambda)^{-1} \widetilde{\boldsymbol{\Psi}}_t\|} \leq 1$, and Eq. (110) follows from $\|\bar{\boldsymbol{\Psi}}_t^\top \left(\mathbf{Q}_t - \boldsymbol{I}_{|\mathcal{X}|} \right)\| = \sqrt{\|\bar{\boldsymbol{\Psi}}_t^\top \left(\mathbf{Q}_t - \boldsymbol{I}_{|\mathcal{X}|} \right) \bar{\boldsymbol{\Psi}}_t\|} \leq \sqrt{\frac{2\lambda}{T}}$, which is implied by the fact that $\bar{\boldsymbol{\Psi}}_t^\top \left(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t \right) \bar{\boldsymbol{\Psi}}_t \preceq \frac{2\lambda}{T} \bar{\boldsymbol{\Psi}}_t^\top (\bar{\boldsymbol{\Psi}}_t \bar{\boldsymbol{\Psi}}_t^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|})^{-1} \bar{\boldsymbol{\Psi}}_t \preceq \frac{2\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|}$ holds from Lemma H.10. Thus, we have

$$\mathbb{E}[f_t(\boldsymbol{x}^*) - \widetilde{u}_t(\boldsymbol{x}^*)] \tag{111}$$

$$\leq \mathbb{E}\left[\mathbb{1}\{A_t\} \left(f_t(\boldsymbol{x}^*) - \widetilde{u}_t(\boldsymbol{x}^*) \right) \right] + \mathbb{P}\left(A_t^c \right) \left(B + \frac{BT}{\lambda} + \beta \sqrt{\frac{T}{\lambda}} \right) \tag{112}$$

$$\leq \mathbb{E}\left[\mathbb{1}\{A_t\} \mathbb{E}\left[f_t(\boldsymbol{x}^*) - \widetilde{u}_t(\boldsymbol{x}^*) \mid \widetilde{\mathcal{H}}_{t-1}, \boldsymbol{\theta}_t, \mathbf{S}_t \right] \right] + \frac{1}{T^3} \left(B + \frac{BT}{\lambda} + \beta \sqrt{\frac{T}{\lambda}} \right) \tag{113}$$

$$\leq \mathbb{E}\left[\mathbb{1}\{A_t\} \left(B \sqrt{\frac{\lambda}{T}} \left(\sqrt{2} + 1 \right) - \beta \right) \|\psi(\boldsymbol{x}^*)\|_{\widetilde{G}_t(\lambda)^{-1}} \right] + \frac{1}{T^3} \left(B + \frac{BT}{\lambda} + \beta \sqrt{\frac{T}{\lambda}} \right) \tag{114}$$

$$= \frac{B}{T^3} \left(\frac{T}{\lambda} + \sqrt{2} + 2 \right). \tag{115}$$

Upper bound for the second term (ii) in Eq. (91)**.** For any $\boldsymbol{x} \in \mathcal{X}$, the definition of the exploration distribution implies the following upper bound of $\widetilde{u}_t(\boldsymbol{x})$ under event A (Lemma H.8):

$$\widetilde{u}_t(\boldsymbol{x}) \leq \frac{2\sqrt{3}}{\alpha}\left(2(\sqrt{3}+\sqrt{2})B\gamma_T + \beta\sqrt{\gamma_T}\right). \tag{116}$$

By combining the above inequality with the definition of α, we can observe that $\eta\widetilde{u}_t(\boldsymbol{x}) \leq 1$ holds for any $\boldsymbol{x} \in \mathcal{X}$ under A; therefore, by following the standard proof for Exp3 (Lemma H.4), we obtain the following inequality under event A:

$$\sum_{t=1}^{T}\widetilde{u}_t(\boldsymbol{x}^*) - \sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t(\boldsymbol{x}) \leq \frac{\log|\mathcal{X}|}{\eta} + \eta\sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t^2(\boldsymbol{x}). \tag{117}$$

Regarding the term $\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t^2(\boldsymbol{x})$, we further obtain the upper bound as follows:

$$\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t^2(\boldsymbol{x}) = \sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\left(\widetilde{f}_t(\boldsymbol{x}) + \beta\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}\right)^2 \tag{118}$$

$$\leq 2\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{f}_t^2(\boldsymbol{x}) + 2\beta^2\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{\widetilde{G}_t(\lambda)^{-1}}, \tag{119}$$

As with the proof of Theorem 3.1 (Eqs. (55)–(58)), under event A, the second term in Eq. (119) is bounded from above as follows:

$$2\beta^2\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{\widetilde{G}_t(\lambda)^{-1}} \leq 6\beta^2\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{G_t(\lambda)^{-1}} \leq \frac{24(1+\sqrt{2})^2\lambda B^2\gamma_T}{(1-\alpha)T}, \tag{120}$$

where the first inequality follows from Lemma H.9 and event A_t. Thus, we have

$$\mathbb{E}\left[\mathbb{1}\{A\}\left(\sum_{t=1}^{T}\widetilde{u}_t(\boldsymbol{x}^*) - \sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t(\boldsymbol{x})\right)\right] \tag{121}$$

$$\leq \mathbb{E}\left[\mathbb{1}\{A\}\left(\frac{\log|\mathcal{X}|}{\eta} + \eta\sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t^2(\boldsymbol{x})\right)\right] \tag{122}$$

$$\leq \frac{\log|\mathcal{X}|}{\eta} + \mathbb{E}\left[\mathbb{1}\{A\}\left(2\eta\sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{f}_t^2(\boldsymbol{x}) + 2\eta\beta^2\sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|^2_{\widetilde{G}_t(\lambda)^{-1}}\right)\right] \tag{123}$$

$$\leq \frac{\log|\mathcal{X}|}{\eta} + \frac{24\eta(1+\sqrt{2})^2\lambda B^2\gamma_T}{1-\alpha} + 2\eta\sum_{t=1}^{T}\mathbb{E}\left[\mathbb{1}\{A\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{f}_t^2(\boldsymbol{x})\right]. \tag{124}$$

Furthermore, we have

$$\mathbb{E}\left[\mathbb{1}\{A\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{f}_t^2(\boldsymbol{x})\right] \tag{125}$$

$$\leq \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{f}_t^2(\boldsymbol{x})\right] \tag{126}$$

$$\leq B^2\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\left[\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t)\right]^2\right] \tag{127}$$

$$= B^2\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t)\widetilde{\psi}_t(\boldsymbol{x}_t)^\top\widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x})\right] \tag{128}$$

$$= B^2\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t)\widetilde{\psi}_t(\boldsymbol{x}_t)^\top\widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x}) \mid \widetilde{\mathcal{H}}_{t-1}, \mathbf{S}_t\right]\right] \tag{129}$$

$$= B^2\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\mathbb{E}\left[\widetilde{\psi}_t(\boldsymbol{x}_t)\widetilde{\psi}_t(\boldsymbol{x}_t)^\top \mid \widetilde{\mathcal{H}}_{t-1}, \mathbf{S}_t\right]\widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x})\right] \tag{130}$$

$$\leq B^2\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x})\right] \tag{131}$$

$$\leq 3B^2\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x})\right] \tag{132}$$

$$\leq \frac{12B^2\gamma_T}{1-\alpha}, \tag{133}$$

where Eq. (130) follows from the fact that A_t, $\widetilde{P}_t(\cdot)$, $\widetilde{G}_t(\lambda)$, and $\widetilde{\psi}_t(\cdot)$ are deterministic given $\widetilde{\mathcal{H}}_{t-1}$, $\mathbf{S}_t$. Furthermore, Eq. (131) follows from $\|\widetilde{G}_t(\lambda)^{-1/2}\mathbb{E}\left[\widetilde{\psi}_t(\boldsymbol{x}_t)\widetilde{\psi}_t(\boldsymbol{x}_t)^\top \mid \widetilde{\mathcal{H}}_{t-1}, \mathbf{S}_t\right]\widetilde{G}_t(\lambda)^{-1/2}\| \leq 1$, which is implied by $\mathbb{E}\left[\widetilde{\psi}_t(\boldsymbol{x}_t)\widetilde{\psi}_t(\boldsymbol{x}_t)^\top \mid \widetilde{\mathcal{H}}_{t-1}, \mathbf{S}_t\right] = \sum_{\boldsymbol{x}\in\mathcal{X}} P_t(\boldsymbol{x})\widetilde{\psi}_t(\boldsymbol{x})\widetilde{\psi}_t(\boldsymbol{x})^\top$ and $\widetilde{G}_t(\lambda) = \sum_{\boldsymbol{x}\in\mathcal{X}} P_t(\boldsymbol{x})\widetilde{\psi}_t(\boldsymbol{x})\widetilde{\psi}_t(\boldsymbol{x})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$. In addition, Eq. (132) is implied by leveraging event A_t with Lemma H.9. Hence, from Eqs. (124) and (133), we have

$$\mathbb{E}\left[\sum_{t=1}^{T}\widetilde{u}_t(\boldsymbol{x}^*) - \sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t(\boldsymbol{x})\right] \tag{134}$$

$$\leq \mathbb{E}\left[\mathbb{1}\{A^c\}\left(\sum_{t=1}^{T}\widetilde{u}_t(\boldsymbol{x}^*) - \sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\widetilde{u}_t(\boldsymbol{x})\right)\right] + \frac{\log|\mathcal{X}|}{\eta} + \frac{24\eta(1+\sqrt{2})^2\lambda B^2\gamma_T}{1-\alpha} + \frac{24\eta B^2 T\gamma_T}{1-\alpha} \tag{135}$$

$$\leq 2T\left(\frac{BT}{\lambda} + \beta\sqrt{\frac{T}{\lambda}}\right)\mathbb{P}(A^c) + \frac{\log|\mathcal{X}|}{\eta} + \frac{24\eta(1+\sqrt{2})^2\lambda B^2\gamma_T}{1-\alpha} + \frac{24\eta B^2 T\gamma_T}{1-\alpha} \tag{136}$$

$$\leq \frac{2B}{T}\left(\frac{T}{\lambda} + \sqrt{2} + 1\right) + \frac{\log|\mathcal{X}|}{\eta} + \frac{24\eta B^2((1+\sqrt{2})^2\lambda + T)\gamma_T}{1-\alpha}, \tag{137}$$

where Eq. (136) follows from Lemma H.11, and Eq. (137) follows from $\mathbb{P}(A^c) \leq 1/T^2$ and the definition of β.

Upper bound for the third term (iii) in Eq. (91)**.** From Lemma H.11 and the inequality $\mathbb{P}(A_t^c) \leq 1/T^3$, we have

$$\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widetilde{u}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] \tag{138}$$

$$= \mathbb{E}\left[\mathbb{1}\{A_t^c\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widetilde{u}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] + \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widetilde{u}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] \tag{139}$$

$$\leq \frac{1}{T^3}\left(B + \frac{BT}{\lambda} + \beta\sqrt{\frac{T}{\lambda}}\right) + \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widetilde{u}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] \tag{140}$$

$$\leq \frac{B}{T^3}\left(\frac{T}{\lambda} + \sqrt{2} + 2\right) + \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widetilde{u}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right]. \tag{141}$$

Here, we can rewrite the second term as

$$\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widetilde{u}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] \tag{142}$$

$$= \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widetilde{f}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] + \beta\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}\right]. \tag{143}$$

Then, we have

$$\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\left(\widetilde{f}_t(\boldsymbol{x}) - f_t(\boldsymbol{x})\right)\right] \tag{144}$$

$$= \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\left(\widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top - \boldsymbol{I}_{|\mathcal{X}|}\right)\boldsymbol{\theta}_t\right] \tag{145}$$

$$= \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\left[\widetilde{G}_t(\lambda)^{-1}\left(\widetilde{\psi}_t(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top - \widetilde{G}_t(\lambda)\right)\right]\boldsymbol{\theta}_t\right] \tag{146}$$

$$= \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\left[\widetilde{G}_t(\lambda)^{-1}\left(\mathbb{E}\left[\widetilde{\psi}_t(\boldsymbol{x}_t)\psi(\boldsymbol{x}_t)^\top \mid \widetilde{\mathcal{H}}_{t-1}, \mathbf{S}_t, \boldsymbol{\theta}_t\right] - \widetilde{G}_t(\lambda)\right)\right]\boldsymbol{\theta}_t\right] \tag{147}$$

$$= \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\left[\widetilde{G}_t(\lambda)^{-1}\left(\widetilde{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Psi}}_t^\top - \widetilde{\boldsymbol{\Psi}}_t\widetilde{\boldsymbol{\Psi}}_t^\top - \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)\right]\boldsymbol{\theta}_t\right] \tag{148}$$

$$\begin{aligned} &= \mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\widetilde{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Psi}}_t^\top\left(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t\right)\boldsymbol{\theta}_t\right] \\ &\quad - \frac{\lambda}{T}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\boldsymbol{\theta}_t\right]. \end{aligned} \tag{149}$$

Similar to the derivation for the upper bound of the first term (i) in Eq. (91), we examine the upper bound of the above two terms. Regarding the first term, we have

$$\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\widetilde{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Psi}}_t^\top\left(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t\right)\boldsymbol{\theta}_t\right] \tag{150}$$

$$\leq B\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}\left\|\widetilde{G}_t(\lambda)^{-1/2}\widetilde{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Psi}}_t^\top\left(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t\right)\right\|\right] \tag{151}$$

$$\leq B\sqrt{\frac{2\lambda}{T}}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}} \widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}\right]. \tag{152}$$

Regarding the second term $-\frac{\lambda}{T}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\boldsymbol{\theta}_t\right]$, we have

$$-\frac{\lambda}{T}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\boldsymbol{\theta}_t\right] \leq \frac{\lambda}{T}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})|\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\boldsymbol{\theta}_t|\right] \tag{153}$$

$$\leq \frac{B\lambda}{T}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})^\top\widetilde{G}_t(\lambda)^{-1}\|_2\right] \tag{154}$$

$$\leq B\sqrt{\frac{\lambda}{T}}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}\right]. \tag{155}$$

Hence, by aggregating Eqs. (141), (143), (149), (152), and (155), we have

$$\mathbb{E}\left[\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\left(\widetilde{u}_t(\boldsymbol{x})-f_t(\boldsymbol{x})\right)\right] \tag{156}$$

$$\leq \frac{B}{T^3}\left(\frac{T}{\lambda}+\sqrt{2}+2\right)+2(1+\sqrt{2})B\sqrt{\frac{\lambda}{T}}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}\right] \tag{157}$$

$$\leq \frac{B}{T^3}\left(\frac{T}{\lambda}+\sqrt{2}+2\right)+2\sqrt{3}(1+\sqrt{2})B\sqrt{\frac{\lambda}{T}}\mathbb{E}\left[\mathbb{1}\{A_t\}\sum_{\boldsymbol{x}\in\mathcal{X}}\widetilde{P}_t(\boldsymbol{x})\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}}\right] \tag{158}$$

$$\leq \frac{B}{T^3}\left(\frac{T}{\lambda}+\sqrt{2}+2\right)+4\sqrt{3}(1+\sqrt{2})B\sqrt{\frac{\lambda\gamma_T}{(1-\alpha)T}}. \tag{159}$$

Aggregate all the upper bounds. From Eqs. (89), (91), (115), (137), and (159), we finally obtain

$$\bar{R}_T \leq 2\alpha BT+(1-\alpha)\left[\frac{B}{T^2}\left(\frac{T}{\lambda}+\sqrt{2}+2\right)+\frac{2B}{T}\left(\frac{T}{\lambda}+\sqrt{2}+1\right)\right.$$
$$\left.+\frac{\log|\mathcal{X}|}{\eta}+\frac{24\eta B^2((1+\sqrt{2})^2\lambda+T)\gamma_T}{1-\alpha}+\frac{B}{T^2}\left(\frac{T}{\lambda}+\sqrt{2}+2\right)+4\sqrt{3}(1+\sqrt{2})B\sqrt{\frac{\lambda T\gamma_T}{1-\alpha}}\right] \tag{160}$$

$$\leq 2\alpha BT+\frac{2B}{T^2}\left(\frac{T}{\lambda}+\sqrt{2}+2\right)+\frac{2B}{T}\left(\frac{T}{\lambda}+\sqrt{2}+1\right)$$
$$+\frac{\log|\mathcal{X}|}{\eta}+24\eta B^2((1+\sqrt{2})^2\lambda+T)\gamma_T+4\sqrt{3}(1+\sqrt{2})B\sqrt{\lambda T\gamma_T}. \tag{161}$$

In the above upper bound, the additional term $\frac{2B}{T^2}\left(\frac{T}{\lambda}+\sqrt{2}+2\right)+\frac{2B}{T}\left(\frac{T}{\lambda}+\sqrt{2}+1\right)$ is $O(1)$ and does not affect the growth rate of regret. Furthermore, the remaining term: $2\alpha BT+\frac{\log|\mathcal{X}|}{\eta}+24\eta B^2((1+\sqrt{2})^2\lambda+T)\gamma_T+4\sqrt{3}(1+\sqrt{2})B\sqrt{\lambda T\gamma_T}$ is the same as the upper bound in Theorem 3.1 up to constant factors. Thus, we complete the proof. □

F Proof of Theorem 5.1

We define the noise function g_t as $g_t(\cdot)=\frac{\eta_t B}{2\sqrt{2\log(2T^2)}}k(\mathbf{0},\cdot)$, where $\eta_t\sim\mathcal{N}(0,1)$. We assume that $(\eta_t)_{t\in\mathbb{N}_+}$ is mutually-independent. Then, we consider the following two problem instances (i.e., strategies of the environment) based on some common function $f\in\mathcal{F}_k(B/2)\coloneqq\{f\in\mathcal{H}_k\mid\|f\|_k\leq B/2\}$ and $(\eta_t)_{t\in\mathbb{N}_+}$:

- **Problem instance 1**. The environment defines f_t as $f_t(\cdot)=f(\cdot)+g_t(\cdot)$.
- **Problem instance 2**. The environment defines f_t as $f_t(\cdot)=f(\cdot)+\mathbb{1}\{|\eta_t|<\sqrt{2\log(2T^2)}\}g_t(\cdot)+\mathbb{1}\{|\eta_t|\geq\sqrt{2\log(2T^2)}\}\mathrm{sign}(\eta_t)\frac{B}{2}k(\mathbf{0},\cdot)$.

In the above definition, the problem instance 1 does not always satisfy $\|f_t\|_k \leq B$. On the other hand, for any realization, problem instance 2 always satisfies $\|f_t\|_k \leq B$. Next, we derive a lower bound for problem instance 2 using problem instance 1. Here, we use the notations $\mathbb{E}^{(1)}[\cdot]$ and $\mathbb{P}^{(1)}(\cdot)$ as the expectation and probability under problem instance 1. Similarly, we also define $\mathbb{E}^{(2)}[\cdot]$ and $\mathbb{P}^{(2)}(\cdot)$. Then, for any algorithm, we have

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}^{(2)}[R_T(\boldsymbol{x})] = \sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}^{(2)}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}) - f_t(\boldsymbol{x}_t)\right] \tag{162}$$

$$= \sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}^{(2)}\left[\sum_{t=1}^{T} f(\boldsymbol{x}) - f(\boldsymbol{x}_t)\right] \tag{163}$$

$$\geq \sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}^{(2)}\left[\mathbb{1}\left\{\forall i \in [T], |\eta_i| < \sqrt{2\log(2T^2)}\right\} \sum_{t=1}^{T} (f(\boldsymbol{x}) - f(\boldsymbol{x}_t))\right] \tag{164}$$

$$= \sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}^{(1)}\left[\mathbb{1}\left\{\forall i \in [T], |\eta_i| < \sqrt{2\log(2T^2)}\right\} \sum_{t=1}^{T} (f(\boldsymbol{x}) - f(\boldsymbol{x}_t))\right] \tag{165}$$

$$\geq \sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}^{(1)}\left[\sum_{t=1}^{T} f(\boldsymbol{x}) - f(\boldsymbol{x}_t)\right] - BT\mathbb{P}^{(1)}\left(\exists t \in [T],\ |\eta_t| \geq \sqrt{2\log(2T^2)}\right), \tag{166}$$

where:

- Eq. (163) follows from the symmetry of the Gaussian distribution. Namely, $\mathbb{E}[\mathbb{1}\{|\eta_t| < \sqrt{2\log(2T^2)}\}g_t(\boldsymbol{x}) + \mathbb{1}\{|\eta_t| \geq \sqrt{2\log(2T^2)}\}\mathrm{sign}(\eta_t)\frac{B}{2}k(\boldsymbol{0},\boldsymbol{x})] = 0$ for any fixed $\boldsymbol{x} \in \mathcal{X}$. Furthermore, the independence between $\boldsymbol{x}_t$ and η_t also implies $\mathbb{E}[\mathbb{1}\{|\eta_t| < \sqrt{2\log(2T^2)}\}g_t(\boldsymbol{x}_t) + \mathbb{1}\{|\eta_t| \geq \sqrt{2\log(2T^2)}\}\mathrm{sign}(\eta_t)\frac{B}{2}k(\boldsymbol{0},\boldsymbol{x}_t)] = 0$.
- Eq. (164) follows from the fact that $\sum_{t=1}^{T} f(\boldsymbol{x}^*) - f(\boldsymbol{x}_t) \geq 0$ always holds for the maximizer $\boldsymbol{x}^*$ of f. (Note that such $\boldsymbol{x}^*$ always exists since $\mathcal{X}$ is compact and f is continuous under either the SE or Matérn kernel.)
- Eq. (165) uses the fact that the expectation of $\mathbb{1}\left\{\forall i \in [T], |\eta_i| < \sqrt{2\log(2T^2)}\right\} \sum_{t=1}^{T}(f(\boldsymbol{x}) - f(\boldsymbol{x}_t))$ under instance 1 equals that under instance 2 from the definitions. To see this rigorously, we first assume that the learner's algorithm is deterministic. Then, we can rewrite $\boldsymbol{x}_t$ as the function of $\boldsymbol{\eta}_{t-1} \coloneqq (\eta_1, \ldots, \eta_{t-1})$ such as $\boldsymbol{x}_t = \boldsymbol{x}_t(\boldsymbol{\eta}_{t-1})$. Then, we can write

$$\mathbb{E}^{(2)}\left[\mathbb{1}\left\{\forall i \in [T], |\eta_i| < \sqrt{2\log(2T^2)}\right\} \sum_{t=1}^{T} (f(\boldsymbol{x}) - f(\boldsymbol{x}_t))\right] \tag{167}$$

$$= \int_{\boldsymbol{\eta}_T \in (-\sqrt{2\log(2T^2)}, \sqrt{2\log(2T^2)})^T} \sum_{t=1}^{T} [f(\boldsymbol{x}) - f(\boldsymbol{x}_t(\boldsymbol{\eta}_{t-1}))]\, \mathbb{P}_{\boldsymbol{\eta}_T}(\mathrm{d}\boldsymbol{\eta}_T). \tag{168}$$

  Note that the distribution of $\mathbb{P}_{\boldsymbol{\eta}_{T-1}}(\cdot)$ is the same across instances 1 and 2, and $\boldsymbol{x}_t(\boldsymbol{\eta}_{t-1})$ is also common under event $\boldsymbol{\eta}_{T-1} \in (-\sqrt{2\log(2T^2)}, \sqrt{2\log(2T^2)})^{T-1}$, since the learner's history $\mathcal{H}_{t-1} \coloneqq (\boldsymbol{x}_1, f_1(\boldsymbol{x}_1), \ldots, \boldsymbol{x}_{t-1}, f_{t-1}(\boldsymbol{x}_{t-1}))$ is the same in the two instances on this event. Thus, we can confirm $\mathbb{E}^{(2)}\left[\mathbb{1}\left\{\forall i \in [T], |\eta_i| < \sqrt{2\log(2T^2)}\right\} \sum_{t=1}^{T}(f(\boldsymbol{x}) - f(\boldsymbol{x}_t))\right] = \mathbb{E}^{(1)}\left[\mathbb{1}\left\{\forall i \in [T], |\eta_i| < \sqrt{2\log(2T^2)}\right\} \sum_{t=1}^{T}(f(\boldsymbol{x}) - f(\boldsymbol{x}_t))\right]$. Finally, for a stochastic algorithm of the learner, the same argument is valid by rewriting $\boldsymbol{x}_t$ as the function form: $\boldsymbol{x}_t(\mathcal{H}_{t-1}, (z_i)_{i\in[t]})$, where $z_i \sim \mathrm{Unif}(\mathcal{X})$ is used to represent additional randomness (see, e.g., Definition 1 in [57]).
- Eq. (166) follows from $\mathbb{E}^{(1)}[\mathbb{1}\{\exists t \in [T],\ |\eta_t| \geq \sqrt{2\log(2T^2)}\} \sum_{t=1}^{T}(f(\boldsymbol{x}^*) - f(\boldsymbol{x}_t))] \leq BT\mathbb{E}^{(1)}[\mathbb{1}\{\exists t \in [T],\ |\eta_t| \geq \sqrt{2\log(2T^2)}\}]$, which is implied by $\|f\|_\infty \leq \|f\|_k \leq B/2$.

Regarding the first term in Eq. (166), there exists a function $f \in \mathcal{F}_k(B/2)$ such that $\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}^{(1)}\left[\sum_{t=1}^{T} f(\boldsymbol{x}) - f(\boldsymbol{x}_t)\right] = \Omega(\underline{R}(T, \sigma_T^2))$, where $\sigma_T^2 = \frac{B^2}{8\log(2T^2)}$ (see Lemma H.16).

Here, we defined $\underline{R}(T, \sigma^2)$ as $\underline{R}(T, \sigma^2) = \sqrt{\sigma^2 T (\log(T/\sigma^2))^{d/2}}$ and $\underline{R}(T, \sigma^2) = \sigma^{\frac{2\nu}{2\nu+d}} T^{\frac{\nu+d}{2\nu+d}}$ for $k = k_{\mathrm{SE}}$ and $k = k_{\mathrm{Mat}}$, respectively. Regarding the second term, we use the union bound with the fact $\mathbb{P}(|\eta_t| \geq \sqrt{2\log(2T^2)}) \leq 1/T^2$, which follows from the basic concentration inequality: $\forall m \geq 0, \mathbb{P}(|\eta_t| \geq m) \leq 2\exp(-m^2/2)$. Then, we obtain $BT\mathbb{P}^{(1)}\left(\exists t \in [T],\ |\eta_t| \geq \sqrt{2\log(2T^2)}\right) \leq B$. Thus, the second term is negligible compared to the first term since $B = \Theta(1)$. Consequently, we conclude that there exists a problem instance (strategy of the environment) such that

$$\sup_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}^{(2)}[R_T(\boldsymbol{x})] = \Omega(\underline{R}(T, \sigma_T^2)). \tag{169}$$

By simplifying the constant factors in the above equation, we obtain the desired results. □

G RLS-kernelized Exp3 with approximated MVR-sequence

As described in the paragraph "Remarks on exploration distribution" of Sec. 4, by applying recursive RLS-Nyström, we can alleviate the computational cost of preparation of the MVR-based exploration distribution. Alg. 6 provides this variant of RLS-kernelized Exp3. The difference from the original RLS-kernelized Exp3 is the definition of the exploration distribution $\widetilde{\pi}$, which is calculated by the approximated MVR algorithm in Alg. 4. With proper modification of the parameters and proof, we can obtain the following theorem, which provides the improved computational cost for the preparation of the exploration distribution, while maintaining the same per-round computational cost and regret upper bound in Theorem 4.1.

Theorem G.1 (Regret upper bound and computational cost for RLS-kernelized Exp3 with approximated MVR sequence)**.** *Suppose that Assumption 2.1 holds. Assume that $\mathcal{X}$ is finite and $|\mathcal{X}| = o(\exp(T\gamma_T^{-1}))$ holds. Furthermore, suppose that d and B are $\Theta(1)$. Set algorithm parameters as $\beta = B(1+\sqrt{2})\sqrt{\lambda/T}$, $\lambda \geq 1$, $\eta = \Theta(\sqrt{\log|\mathcal{X}|/T\gamma_T})$, $\alpha = \eta\left(36(3+\sqrt{6})B\gamma_T + \beta\sqrt{108\gamma_T}\right)$, and $\delta = 1/T^2$. Then, when running Alg. 6, we have $\bar{R}_T = O(\sqrt{T\gamma_T \log|\mathcal{X}|})$. Furthermore, with probability at least $1 - 1/T^2$, the following two statements hold simultaneously:*

- *For any $t \in [T]$, the computational cost at round t is $\widetilde{O}(|\mathcal{X}|\gamma_T^2 + \gamma_T^3)$.*
- *The computational cost for calculating approximated MVR sequence-based exploration distribution $\widetilde{\pi}$ is $\widetilde{O}\left(\sum_{t=1}^{\lceil T\alpha \rceil}(t\gamma_t^2 + \gamma_t^3 + |\mathcal{X}|\gamma_t^2)\right)$.*

Computational cost. To compare the costs of preparing the exploration distribution, fix $\lambda = \Theta(1)$ and assume $T \ll |\mathcal{X}| \leq T^c$ for some fixed $c > 1$. Suppose also that $\gamma_t = \widetilde{O}(t^a)$ for some fixed $a \in [0, 1)$; this holds for the SE kernel with $a = 0$ and for the ν-Matérn kernel with $a = d/(2\nu + d)$. Under these assumptions and the parameter settings in Theorem G.1, $\alpha = \Theta\left(\sqrt{\gamma_T \log|\mathcal{X}|/T}\right)$. Let $M := \lceil T\alpha \rceil = \Theta\left(\sqrt{T\gamma_T \log|\mathcal{X}|}\right)$. Since $M \leq T$ for sufficiently large T and $\gamma_M = O(M)$, we obtain

$$\widetilde{O}\left(\sum_{t=1}^{M}(t\gamma_t^2 + \gamma_t^3 + |\mathcal{X}|\gamma_t^2)\right) = \widetilde{O}\left(|\mathcal{X}|M\gamma_M^2\right) = \widetilde{O}\left(|\mathcal{X}|M^{1+2a}\right) = o\left(|\mathcal{X}|(T\gamma_T)^{3/2}\right). \tag{170}$$

The last relation follows from $a < 1$ and $\log|\mathcal{X}| = O(\log T)$. Thus, under the stated assumptions, the resulting computational cost bound is smaller than the $O(|\mathcal{X}|M^3)$ bound for preparing the exploration distribution using exact MVR.

Summary of the proof. The core lemma for the proof of Theorem G.1 is Lemma H.14, which guarantees that the posterior variance under the approximated MVR sequence diminishes at the same rate as that under the exact MVR sequence. See Lemma H.14 and its proof for details. Once we establish Lemma H.14, Theorem G.1 follows by applying the same proof as that of Theorem 4.1 with slight modifications. Thus, we omit the proof of Theorem G.1, while we summarize the required modifications below:

- In addition to events (A_t) in Eq. (86), we consider additional event A_0 such that the three events in Corollary H.12 with $\widetilde{\mathbf{S}} = \widetilde{\mathbf{S}}_t$, $\boldsymbol{\Psi} = \boldsymbol{\Psi}_{t-1}^{(\mathrm{MVR})}$, and $\widetilde{s} = \widetilde{s}_t$ hold for all rounds

$t \in [\lceil T\alpha \rceil]]$ in the approximated MVR procedures (Line 1 in Alg. 6). Then, we also define the modified event $\bar{A} \coloneqq \cap_{t\in[T]\cup\{0\}} A_t$ as the replacement of event A in the original proof. Then, note that $\mathbb{P}(A_0) \geq 1 - 1/(2T^2)$ and $\mathbb{P}(\cap_{t\in[T]} A_t) \geq 1 - 1/(2T^2)$ hold due to Lemma H.7, Corollary H.12, and the settings of the confidence level parameters (Lines 1 and 7 in Alg. 6). Thus, the union bound implies $\mathbb{P}(\bar{A}^c) \leq 1/T^2$, which matches the upper bound $\mathbb{P}(A^c)$ used in the proof of the original RLS-kernelized Exp3.

- The proof for the upper bounds for terms (i) and (iii) in Eq. (91) remains unchanged from the original proof except for the replacement of the constant factor arising from the difference in $\mathbb{P}(A_t^c)$. (In the original proof, we have $\mathbb{P}(A_t^c) \leq 1/T^3$, while $\mathbb{P}(A_t^c) \leq 1/(2T^3)$ under Alg. 6.)
- Regarding the upper bounds for term (ii) in Eq. (91), we can leverage Lemma H.15 under event $\bar{A}$ instead of using Lemma H.8 under event A. Note that event: $\left\{\max_{\boldsymbol{x}\in\mathcal{X}} \sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}^{(\mathrm{MVR})}_{\lceil T\alpha\rceil}, \lambda\right) \leq \frac{36\lambda\gamma_{\lceil T\alpha\rceil}}{\lceil T\alpha\rceil}\right\}$, which is required for the application of Lemma H.15, is true under event A_0. Then, from Lemma H.15, for guaranteeing $\eta\widetilde{u}_t(\boldsymbol{x}) \leq 1$, we must set α such that $\frac{\eta}{\alpha}\left(36(3+\sqrt{6})B\gamma_T + \beta\sqrt{108\gamma_T}\right) \leq 1$ holds. The setting of α in Theorem G.1 satisfies this condition; thus, we can apply the standard Exp3 analysis in Lemma H.4 under event $\bar{A}$. After that, we can directly follow the original proof of the upper bound for term (ii). By aggregating the upper bounds for terms (i), (ii), and (iii), we can obtain the desired regret guarantees.
- Finally, we prove the statements regarding the computational costs. As with the original RLS-kernelized Exp3, the upper bound for the per-round computational cost (the first statement in Theorem G.1) directly follows from Lemma H.7. Regarding the computational cost for the preparation of the exploration distribution, under event A_0, we can conclude the following two facts: (i) the computational cost of recursive RLS-Nyström is $\widetilde{O}(t\gamma_t^2)$, and (ii) the dimension of the Nyström embedding $\widetilde{s}_t$ is $\widetilde{O}(\gamma_t)$, which implies the computation of Eq. (12) for all $\boldsymbol{x} \in \mathcal{X}$ is $\widetilde{O}(|\mathcal{X}|\gamma_t^2 + \gamma_t^3)$. Thus, the computational cost for approximated MVR procedures at round t is $\widetilde{O}(t\gamma_t^2 + |\mathcal{X}|\gamma_t^2 + \gamma_t^3)$, which implies the desired statement.

H Helper lemmas

H.1 Basic properties of feature map, RKHS, and GP

Overview. The lemmas in this subsection are well-known, useful properties for handling the feature representation of the kernel [19, 67], while we provide the proof for completeness.

Lemma H.1 (Feature map and RKHS norm)**.** *Let $k : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ be a positive definite kernel function on a finite input domain $\mathcal{X}$. Let $\psi : \mathcal{X} \to \mathbb{R}^{|\mathcal{X}|}$ be any feature map that satisfies $k(\boldsymbol{x}, \boldsymbol{x}') = \psi(\boldsymbol{x})^\top\psi(\boldsymbol{x}')$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$. Fix any $f \in \mathcal{H}_k$. Then, there exists a vector $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{X}|}$ such that $f(\cdot) = \psi(\cdot)^\top\boldsymbol{\theta}$. Furthermore, $\|\boldsymbol{\theta}\|_2 = \|f\|_k$ holds.*

Proof. Since $f \in \mathcal{H}_k$ and $|\mathcal{X}| < \infty$, the function f is represented as $f(\cdot) = \sum_{\boldsymbol{x}\in\mathcal{X}} c_{\boldsymbol{x}} k(\boldsymbol{x}, \cdot)$ with some coefficients $c_{\boldsymbol{x}} \in \mathbb{R}$. In addition, the RKHS norm $\|f\|_k$ is defined as $\|f\|_k = \sqrt{\sum_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}} c_{\boldsymbol{x}} c_{\boldsymbol{x}'} k(\boldsymbol{x}, \boldsymbol{x}')}$ [e.g., see Chap. 2.3 in 33]. By using the feature map, we have $f(\cdot) = \psi(\cdot)^\top(\sum_{\boldsymbol{x}\in\mathcal{X}} c_{\boldsymbol{x}}\psi(\boldsymbol{x}))$. This implies that we can set $\boldsymbol{\theta} = \sum_{\boldsymbol{x}\in\mathcal{X}} c_{\boldsymbol{x}}\psi(\boldsymbol{x})$. Furthermore, we have

$$\|\boldsymbol{\theta}\|_2 = \sqrt{\sum_{j=1}^{|\mathcal{X}|}\left(\sum_{\boldsymbol{x}\in\mathcal{X}} c_{\boldsymbol{x}}\psi_j(\boldsymbol{x})\right)^2} = \sqrt{\sum_{j=1}^{|\mathcal{X}|}\sum_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}} c_{\boldsymbol{x}} c_{\boldsymbol{x}'}\psi_j(\boldsymbol{x})\psi_j(\boldsymbol{x}')} \tag{171}$$

$$= \sqrt{\sum_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}} c_{\boldsymbol{x}} c_{\boldsymbol{x}'}\left(\sum_{j=1}^{|\mathcal{X}|}\psi_j(\boldsymbol{x})\psi_j(\boldsymbol{x}')\right)} = \sqrt{\sum_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}} c_{\boldsymbol{x}} c_{\boldsymbol{x}'}\psi(\boldsymbol{x})^\top\psi(\boldsymbol{x}')} \tag{172}$$

$$= \sqrt{\sum_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}} c_{\boldsymbol{x}} c_{\boldsymbol{x}'} k(\boldsymbol{x}, \boldsymbol{x}')} = \|f\|_k. \tag{173}$$

Thus, we complete the proof. □

Lemma H.2 (Basic matrix identities)**.** *Fix any $n, m \in \mathbb{N}_+$ and $\lambda > 0$. Then, for any matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, the following two identities hold:*

$$(\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}_n)^{-1}\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_m)^{-1}, \tag{174}$$

$$\boldsymbol{I}_m - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}_n)^{-1}\boldsymbol{A} = \lambda(\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_m)^{-1}. \tag{175}$$

Proof. The first identity is implied by multiplying $(\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_m)^{-1}$ and $(\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}_n)^{-1}$ from both sides of the equation below:

$$\boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_m) = \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{A} = (\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}_n)\boldsymbol{A}. \tag{176}$$

Regarding the second identity, we have

$$\boldsymbol{I}_m - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}_n)^{-1}\boldsymbol{A} = (\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_m)^{-1}(\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_m) - (\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_m)^{-1}\boldsymbol{A}^\top \boldsymbol{A} \tag{177}$$

$$= \lambda(\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_m)^{-1}, \tag{178}$$

where the first equality follows from Eq. (174). □

Lemma H.3 (Feature map and posterior variance of GP)**.** *Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive definite kernel function on a finite input domain $\mathcal{X}$. Let $\psi : \mathcal{X} \to \mathbb{R}^{|\mathcal{X}|}$ be any feature map that satisfies $k(\boldsymbol{x}, \boldsymbol{x}') = \psi(\boldsymbol{x})^\top \psi(\boldsymbol{x}')$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$. Furthermore, let $\mathbf{X} \coloneqq (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(t)})$ be any training inputs of the GP, which satisfies $\boldsymbol{x}^{(i)} \in \mathcal{X}$ for any $i \in [t]$. Assume that the variance parameter λ of the GP is strictly positive. Then, for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, the posterior variance in Eq.* (2) *is represented as*

$$\sigma^2(\boldsymbol{x}; \mathbf{X}, \lambda) = \lambda \psi(\boldsymbol{x})^\top \left(\sum_{i=1}^{t} \psi(\boldsymbol{x}^{(i)}) \psi(\boldsymbol{x}^{(i)})^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \psi(\boldsymbol{x}) \tag{179}$$

$$= \lambda \psi(\boldsymbol{x})^\top \left(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \psi(\boldsymbol{x}), \tag{180}$$

where $\boldsymbol{\Psi} = (\psi(\boldsymbol{x}^{(1)}), \ldots, \psi(\boldsymbol{x}^{(t)})) \in \mathbb{R}^{|\mathcal{X}| \times t}$.

Proof. The desired statement can be obtained by the following basic matrix calculus:

$$\sigma^2(\boldsymbol{x}; \mathbf{X}, \lambda) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x}, \mathbf{X})^\top (\mathbf{K}(\mathbf{X}, \mathbf{X}) + \lambda \boldsymbol{I}_t)^{-1} \boldsymbol{k}(\boldsymbol{x}, \mathbf{X}) \tag{181}$$

$$= \psi(\boldsymbol{x})^\top \psi(\boldsymbol{x}) - (\boldsymbol{\Psi}^\top \psi(\boldsymbol{x}))^\top (\boldsymbol{\Psi}^\top \boldsymbol{\Psi} + \lambda \boldsymbol{I}_t)^{-1} \boldsymbol{\Psi}^\top \psi(\boldsymbol{x}) \tag{182}$$

$$= \psi(\boldsymbol{x})^\top \psi(\boldsymbol{x}) - \psi(\boldsymbol{x})^\top \boldsymbol{\Psi} (\boldsymbol{\Psi}^\top \boldsymbol{\Psi} + \lambda \boldsymbol{I}_t)^{-1} \boldsymbol{\Psi}^\top \psi(\boldsymbol{x}) \tag{183}$$

$$= \psi(\boldsymbol{x})^\top \psi(\boldsymbol{x}) - \psi(\boldsymbol{x})^\top (\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|})^{-1} \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \psi(\boldsymbol{x}) \tag{184}$$

$$= \psi(\boldsymbol{x})^\top \left[\boldsymbol{I}_{|\mathcal{X}|} - (\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|})^{-1} \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \right] \psi(\boldsymbol{x}) \tag{185}$$

$$= \psi(\boldsymbol{x})^\top (\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|})^{-1} \left[(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}) - \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \right] \psi(\boldsymbol{x}) \tag{186}$$

$$= \lambda \psi(\boldsymbol{x})^\top (\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|})^{-1} \psi(\boldsymbol{x}). \tag{187}$$

In the above expressions, Eq. (184) follows from Lemma H.2. □

H.2 Helper lemmas for Kernelized Exp3

Overview. The lemmas in this subsection are leveraged in the proof of kernelized Exp3 in Appendix D. Lemma H.4 is an adaptation of the standard Exp3 analysis. For completeness, we provide the full proof. Lemma H.5 is the upper bound for the optimistic estimator, which is used to guarantee the required condition to apply Lemma H.4. In the proof of Lemma H.5, we leverage the known property of posterior variance regarding the MVR sequence, which can be found in the existing stochastic KB literature [12, 38, 65].

Lemma H.4 (Standard Exp3 analysis, adapted from e.g., the proof of Theorem 11.1 in [36])**.** *Fix any $T \in \mathbb{N}_+$, $\eta > 0$, and finite input domain $\mathcal{X}$. Furthermore, for any $t \in [T]$, let $u_t : \mathcal{X} \to \mathbb{R}$ be a function that satisfies $\eta u_t(\boldsymbol{x}) \leq 1$ for any $\boldsymbol{x} \in \mathcal{X}$. Then, for any $\boldsymbol{z} \in \mathcal{X}$, the following inequality holds:*

$$\sum_{t=1}^{T} u_t(\boldsymbol{z}) - \sum_{t=1}^{T} \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) u_t(\boldsymbol{x}) \leq \frac{\log |\mathcal{X}|}{\eta} + \eta \sum_{t=1}^{T} \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_t(\boldsymbol{x}) u_t^2(\boldsymbol{x}), \tag{188}$$

where $\widetilde{P}_1(\boldsymbol{x}) = 1/|\mathcal{X}|$ *and*

$$\widetilde{P}_{t+1}(\boldsymbol{x}) = \frac{\exp\left(\eta \sum_{i=1}^{t} u_i(\boldsymbol{x})\right)}{\sum_{\boldsymbol{x}' \in \mathcal{X}} \exp\left(\eta \sum_{i=1}^{t} u_i(\boldsymbol{x}')\right)} \tag{189}$$

for any $t \in [T-1]$.

Proof. Let us define w_j as $w_j = \sum_{\boldsymbol{x} \in \mathcal{X}} \exp\left(\eta \sum_{t=1}^{j} u_t(\boldsymbol{x})\right)$ for any $j \in [T]$. Furthermore, we define $w_0 = |\mathcal{X}|$. Then, we have

$$\exp\left(\eta \sum_{t=1}^{T} u_t(\boldsymbol{z})\right) \leq \sum_{\boldsymbol{x} \in \mathcal{X}} \exp\left(\eta \sum_{t=1}^{T} u_t(\boldsymbol{x})\right) = w_T = w_0 \prod_{j=1}^{T} \frac{w_j}{w_{j-1}}. \tag{190}$$

Regarding the term $\frac{w_j}{w_{j-1}}$ in the product, by noting the definition of $\widetilde{P}_t(\cdot)$, we have

$$\begin{aligned}
\frac{w_j}{w_{j-1}} &= \sum_{\boldsymbol{x} \in \mathcal{X}} \frac{\exp\left(\eta \sum_{t=1}^{j} u_t(\boldsymbol{x})\right)}{w_{j-1}} && (191)\\
&= \sum_{\boldsymbol{x} \in \mathcal{X}} \frac{\exp\left(\eta \sum_{t=1}^{j-1} u_t(\boldsymbol{x})\right)}{w_{j-1}} \exp\left(\eta u_j(\boldsymbol{x})\right) && (192)\\
&= \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) \exp\left(\eta u_j(\boldsymbol{x})\right). && (193)
\end{aligned}$$

Note that $\eta u_j(\boldsymbol{x}) \leq 1$ holds. Thus, by leveraging the elementary inequalities $\forall a \leq 1, \exp(a) \leq 1 + a + a^2$ and $\forall a \in \mathbb{R},\ 1 + a \leq \exp(a)$, we have

$$\begin{aligned}
\sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) \exp\left(\eta u_j(\boldsymbol{x})\right) &\leq \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) \left(1 + \eta u_j(\boldsymbol{x}) + \eta^2 u_j^2(\boldsymbol{x})\right) && (194)\\
&= 1 + \eta \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) u_j(\boldsymbol{x}) + \eta^2 \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) u_j^2(\boldsymbol{x}) && (195)\\
&\leq \exp\left(\eta \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) u_j(\boldsymbol{x}) + \eta^2 \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) u_j^2(\boldsymbol{x})\right). && (196)
\end{aligned}$$

By aggregating Eqs. (190), (193), and (196), we have

$$\exp\left(\eta \sum_{t=1}^{T} u_t(\boldsymbol{z})\right) \leq |\mathcal{X}| \exp\left(\eta \sum_{j=1}^{T} \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) u_j(\boldsymbol{x}) + \eta^2 \sum_{j=1}^{T} \sum_{\boldsymbol{x} \in \mathcal{X}} \widetilde{P}_j(\boldsymbol{x}) u_j^2(\boldsymbol{x})\right). \tag{197}$$

By taking the logarithm and dividing by the factor $\eta > 0$ on both sides of the above inequality, we obtain the desired result. □

Lemma H.5 (Uniform upper bound of the optimistic estimator u_t)**.** *Fix any* $T \in \mathbb{N}_+$, $\alpha \in (0, 1)$, $\eta > 0$, $\beta > 0$, *and* $\lambda \geq 1$. *Fix any finite input domain* $\mathcal{X}$. *Suppose that Assumption 2.1 holds. Then, when running Alg. 1, the following upper bound of the optimistic estimator* u_t *holds for any* $t \in [T]$ *and* $\boldsymbol{x} \in \mathcal{X}$*:*

$$u_t(\boldsymbol{x}) \leq \frac{2}{\alpha}(2B\gamma_T + \beta\sqrt{\gamma_T}). \tag{198}$$

Proof. From the definition of $\widehat{f}_t$, we have

$$\begin{aligned}
\widehat{f}_t(\boldsymbol{x}) &= \psi(\boldsymbol{x})^\top G_t(\lambda)^{-1} \psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t) && (199)\\
&\leq \|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}} \|G_t(\lambda)^{-1/2} \psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t)\|_2 && (200)\\
&\leq B \|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}} \|\psi(\boldsymbol{x}_t)\|_{G_t(\lambda)^{-1}} && (201)\\
&\leq B \max_{\boldsymbol{x}' \in \mathcal{X}} \|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}}. && (202)
\end{aligned}$$

Thus, the optimistic estimator u_t satisfies the following upper bound:

$$u_t(\boldsymbol{x}) = \widehat{f}_t(\boldsymbol{x}) + \beta \|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}} \tag{203}$$

$$\leq B \max_{\boldsymbol{x}' \in \mathcal{X}} \|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}} + \beta \max_{\boldsymbol{x}' \in \mathcal{X}} \|\psi(\boldsymbol{x}')\|_{G_t(\lambda)^{-1}}. \tag{204}$$

Furthermore, we obtain the upper bound of $\|\psi(\boldsymbol{x}')\|_{G_t(\lambda)^{-1}}$ as follows:

$$\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}} = \psi(\boldsymbol{x}')^\top \left(\sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x}) \psi(\boldsymbol{x}) \psi(\boldsymbol{x})^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \psi(\boldsymbol{x}') \tag{205}$$

$$\leq \psi(\boldsymbol{x}')^\top \left(\alpha \sum_{\boldsymbol{x} \in \mathcal{X}} \pi(\boldsymbol{x}) \psi(\boldsymbol{x}) \psi(\boldsymbol{x})^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \psi(\boldsymbol{x}') \tag{206}$$

$$= \frac{1}{\alpha} \psi(\boldsymbol{x}')^\top \left(\frac{1}{\lceil T\alpha \rceil} \sum_{j=1}^{\lceil T\alpha \rceil} \psi(\boldsymbol{x}_j^{(\mathrm{MVR})}) \psi(\boldsymbol{x}_j^{(\mathrm{MVR})})^\top + \frac{\lambda}{T\alpha} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \psi(\boldsymbol{x}') \tag{207}$$

$$\leq \frac{1}{\alpha} \psi(\boldsymbol{x}')^\top \left(\frac{1}{\lceil T\alpha \rceil} \sum_{j=1}^{\lceil T\alpha \rceil} \psi(\boldsymbol{x}_j^{(\mathrm{MVR})}) \psi(\boldsymbol{x}_j^{(\mathrm{MVR})})^\top + \frac{\lambda}{\lceil T\alpha \rceil} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \psi(\boldsymbol{x}') \tag{208}$$

$$= \frac{\lceil T\alpha \rceil}{\alpha} \psi(\boldsymbol{x}')^\top \left(\sum_{j=1}^{\lceil T\alpha \rceil} \psi(\boldsymbol{x}_j^{(\mathrm{MVR})}) \psi(\boldsymbol{x}_j^{(\mathrm{MVR})})^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \psi(\boldsymbol{x}') \tag{209}$$

$$= \frac{\lceil T\alpha \rceil}{\alpha\lambda} \sigma^2 \left(\boldsymbol{x}'; \mathbf{X}^{(\mathrm{MVR})}_{\lceil T\alpha \rceil}, \lambda \right) \tag{210}$$

$$\leq \frac{4\gamma_T}{\alpha}, \tag{211}$$

where:

- Eq. (206) follows from $\sum_{\boldsymbol{x} \in \mathcal{X}} \alpha\pi(\boldsymbol{x}) \psi(\boldsymbol{x}) \psi(\boldsymbol{x})^\top \preceq \sum_{\boldsymbol{x} \in \mathcal{X}} P_t(\boldsymbol{x}) \psi(\boldsymbol{x}) \psi(\boldsymbol{x})^\top$, which is implied by the definition $P_t(\boldsymbol{x}) = \alpha\pi(\boldsymbol{x}) + (1-\alpha)\widetilde{P}_t(\boldsymbol{x})$.
- Eqs. (208) and (210) follow from $\frac{\lambda}{\lceil T\alpha \rceil} \boldsymbol{I}_{|\mathcal{X}|} \preceq \frac{\lambda}{T\alpha} \boldsymbol{I}_{|\mathcal{X}|}$ and the feature representation of the posterior variance in Eq. (21), respectively.
- Eq. (211) follows from $\sigma^2 \left(\boldsymbol{x}'; \mathbf{X}^{(\mathrm{MVR})}_{\lceil T\alpha \rceil}, \lambda \right) \leq 4\lambda\gamma_{\lceil T\alpha \rceil} / \lceil T\alpha \rceil$ and $\gamma_{\lceil T\alpha \rceil} \leq \gamma_T$. Indeed, from the definition of $\mathbf{X}^{(\mathrm{MVR})}_{\lceil T\alpha \rceil}$ and monotonicity with respect to the training inputs of the posterior variance, we have

  $$\sigma^2 \left(\boldsymbol{x}'; \mathbf{X}^{(\mathrm{MVR})}_{\lceil T\alpha \rceil}, \lambda \right) \leq \frac{1}{\lceil T\alpha \rceil} \sum_{t=1}^{\lceil T\alpha \rceil} \sigma^2 \left(\boldsymbol{x}_t^{(\mathrm{MVR})}; \mathbf{X}^{(\mathrm{MVR})}_{t-1}, \lambda \right) \leq \frac{2\gamma_{\lceil T\alpha \rceil}}{\lceil T\alpha \rceil \log(1+\lambda^{-1})}. \tag{212}$$

  See [12, 38, 65] for details. Since $\forall a \in [0,1], \log(1+a) \geq a/2$ and $\lambda \geq 1$, we can confirm $\sigma^2 \left(\boldsymbol{x}'; \mathbf{X}^{(\mathrm{MVR})}_{\lceil T\alpha \rceil}, \lambda \right) \leq 4\lambda\gamma_{\lceil T\alpha \rceil} / \lceil T\alpha \rceil$.

Hence, by combining Eq. (211) with Eq. (204), we obtain

$$u_t(\boldsymbol{x}) \leq \frac{4B\gamma_T}{\alpha} + \beta \sqrt{\frac{4\gamma_T}{\alpha}} \leq \frac{2}{\alpha} (2B\gamma_T + \beta\sqrt{\gamma_T}), \tag{213}$$

which is the desired inequality. The last inequality in Eq. (213) follows from $1/\sqrt{\alpha} \leq 1/\alpha$ due to $\alpha \in (0,1)$. □

H.3 Helper lemmas for RLS-Kernelized Exp3

Overview. The lemmas in this subsection are leveraged in the proof of RLS-kernelized Exp3 in Appendix E. Lemma H.6 provides the identities of the approximated estimators and alternative representations based on Nyström embedding. We would like to note that, in a multi-task setting under an infinite-dimensional feature map, Lemma 8 in [19] already provides a similar result to that in Lemma H.6. Lemma H.7 provides the theoretical property of the computational cost and accuracy of the recursive RLS-Nyström subroutine. The proof of Lemma H.7 is obtained by combining the existing effective dimension-based result by [41] with Lemma 3.3. Lemma H.8 provides the upper bound of approximated optimistic estimator $\widetilde{u}_t$, which serves as the counterpart of Lemma H.5 for the proof of the original kernelized Exp3 algorithm. Lemmas H.9 and H.10 provide the spectral results. Roughly speaking, under the sufficient accuracy of the sampling matrix $\mathbf{S}_t$, these results are useful for connecting the approximated quantities in RLS-kernelized Exp3 with non-approximated quantities in the proof. The proofs of Lemmas H.9 and H.10 are almost directly obtained by adapting the proof of [13] and [15], which study the kernel approximation for k-means clustering and the computationally efficient KB algorithm, respectively. Finally, Lemma H.11 is used to obtain the upper bound on the regret arising from the unfavorable rare event, where the accuracy of recursive RLS-Nyström is not sufficient. The proof of Lemma H.11 is obtained with elementary matrix calculations.

Lemma H.6 (Identities in Eqs. (9) and (10))**.** *Fix any finite input domain $\mathcal{X} \coloneqq \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(|\mathcal{X}|)}\}$, any probability mass function $P : \mathcal{X} \to (0, 1]$, any positive definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, any feature map $\psi : \mathcal{X} \to \mathbb{R}^{|\mathcal{X}|}$ of k such that $k(\boldsymbol{x}, \boldsymbol{x}') = \psi(\boldsymbol{x})^\top \psi(\boldsymbol{x}')$ holds for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$. Fix any $T \in \mathbb{N}_+$, $\lambda > 0$, and $s \in [|\mathcal{X}|]$. Furthermore, define a sampling matrix $\mathbf{S} \in \mathbb{R}^{|\mathcal{X}| \times s}$ with exactly one strictly positive entry in each column and zeros elsewhere. In addition, define $\widetilde{G}(\lambda)$ as $\widetilde{G}(\lambda) = (\mathbf{Q}\bar{\mathbf{\Psi}})(\mathbf{Q}\bar{\mathbf{\Psi}})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$, where $\bar{\mathbf{\Psi}} = (\bar{\psi}(\boldsymbol{x}^{(1)}), \ldots, \bar{\psi}(\boldsymbol{x}^{(|\mathcal{X}|)})) \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$, $\bar{\psi}(\boldsymbol{x}) = \sqrt{P(\boldsymbol{x})}\psi(\boldsymbol{x})$, and $\mathbf{Q} = \bar{\mathbf{\Psi}}\mathbf{S}[(\bar{\mathbf{\Psi}}\mathbf{S})^\top(\bar{\mathbf{\Psi}}\mathbf{S})]^\dagger(\bar{\mathbf{\Psi}}\mathbf{S})^\top$. Then, the following identities hold:*

$$\psi(\boldsymbol{x})^\top \widetilde{G}(\lambda)^{-1}\mathbf{Q}\psi(\boldsymbol{z}) = \frac{1}{\sqrt{P(\boldsymbol{x})P(\boldsymbol{z})}}\phi(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}' \in \mathcal{X}} \phi(\boldsymbol{x}')\phi(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_s\right)^{-1} \phi(\boldsymbol{z}), \tag{214}$$

$$\begin{aligned}\|\psi(\boldsymbol{x})\|^2_{\widetilde{G}(\lambda)^{-1}} = {} & \frac{T}{\lambda P(\boldsymbol{x})}\left(\widetilde{k}(\boldsymbol{x}, \boldsymbol{x}) - \phi(\boldsymbol{x})^\top\phi(\boldsymbol{x})\right) \\ & + \frac{1}{P(\boldsymbol{x})}\phi(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}' \in \mathcal{X}} \phi(\boldsymbol{x}')\phi(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_s\right)^{-1} \phi(\boldsymbol{x}),\end{aligned} \tag{215}$$

where $\phi(\boldsymbol{x}) = [(\mathbf{S}^\top\widetilde{\mathbf{K}}\mathbf{S})^{1/2}]^\dagger\mathbf{S}^\top\widetilde{\boldsymbol{k}}(\boldsymbol{x}) \in \mathbb{R}^s$. Here, $\widetilde{k}(\boldsymbol{x}, \boldsymbol{x}')$, $\widetilde{\boldsymbol{k}}(\boldsymbol{x}) \in \mathbb{R}^{|\mathcal{X}|}$, and $\widetilde{\mathbf{K}} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$ are defined as $\widetilde{k}(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{P(\boldsymbol{x})P(\boldsymbol{x}')}k(\boldsymbol{x}, \boldsymbol{x}')$, $\widetilde{\boldsymbol{k}}(\boldsymbol{x}) = [\widetilde{k}(\boldsymbol{x}, \boldsymbol{x}^{(i)})]_{i \in [|\mathcal{X}|]}$, and $\widetilde{\mathbf{K}} = [\widetilde{k}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})]_{i,j \in [|\mathcal{X}|]}$, respectively.

Proof. For any $\boldsymbol{z}, \boldsymbol{x} \in \mathcal{X}$, we first confirm the two identities: $\phi(\boldsymbol{z})^\top\phi(\boldsymbol{x}) = (\mathbf{Q}\bar{\psi}(\boldsymbol{z}))^\top\bar{\psi}(\boldsymbol{x})$ and $\phi(\boldsymbol{z})^\top\phi(\boldsymbol{x}) = (\mathbf{Q}\bar{\psi}(\boldsymbol{x}))^\top(\mathbf{Q}\bar{\psi}(\boldsymbol{z}))$. From the definition of $\phi(\cdot)$, we have

$$\begin{aligned}\phi(\boldsymbol{z})^\top\phi(\boldsymbol{x}) &= \widetilde{\boldsymbol{k}}(\boldsymbol{z})^\top\mathbf{S}[(\mathbf{S}^\top\widetilde{\mathbf{K}}\mathbf{S})^{1/2}]^\dagger[(\mathbf{S}^\top\widetilde{\mathbf{K}}\mathbf{S})^{1/2}]^\dagger\mathbf{S}^\top\widetilde{\boldsymbol{k}}(\boldsymbol{x}) && (216)\\ &= \widetilde{\boldsymbol{k}}(\boldsymbol{z})^\top\mathbf{S}(\mathbf{S}^\top\widetilde{\mathbf{K}}\mathbf{S})^\dagger\mathbf{S}^\top\widetilde{\boldsymbol{k}}(\boldsymbol{x}) && (217)\\ &= \bar{\psi}(\boldsymbol{z})^\top\bar{\mathbf{\Psi}}\mathbf{S}(\mathbf{S}^\top\bar{\mathbf{\Psi}}^\top\bar{\mathbf{\Psi}}\mathbf{S})^\dagger\mathbf{S}^\top\bar{\mathbf{\Psi}}^\top\bar{\psi}(\boldsymbol{x}) && (218)\\ &= \bar{\psi}(\boldsymbol{z})^\top\mathbf{Q}\bar{\psi}(\boldsymbol{x}) && (219)\\ &= \bar{\psi}(\boldsymbol{z})^\top\mathbf{Q}^\top\bar{\psi}(\boldsymbol{x}) && (220)\\ &= (\mathbf{Q}\bar{\psi}(\boldsymbol{z}))^\top\bar{\psi}(\boldsymbol{x}). && (221)\end{aligned}$$

Furthermore, since the orthogonal projection matrix $\mathbf{Q}$ satisfies $\mathbf{Q}^\top = \mathbf{Q}$ and $\mathbf{Q}^2 = \mathbf{Q}$, we also obtain

$$\phi(\boldsymbol{z})^\top\phi(\boldsymbol{x}) = \bar{\psi}(\boldsymbol{z})^\top\mathbf{Q}\bar{\psi}(\boldsymbol{x}) = \bar{\psi}(\boldsymbol{z})^\top\mathbf{Q}^2\bar{\psi}(\boldsymbol{x}) = (\mathbf{Q}\bar{\psi}(\boldsymbol{z}))^\top(\mathbf{Q}\bar{\psi}(\boldsymbol{x})). \tag{222}$$

Here, we define $\widetilde{\mathbf{\Psi}} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$ and $\mathbf{\Phi} \in \mathbb{R}^{s \times |\mathcal{X}|}$ as $\widetilde{\mathbf{\Psi}} = \mathbf{Q}\bar{\mathbf{\Psi}}$ and $\mathbf{\Phi} = (\phi(\boldsymbol{x}^{(1)}), \ldots, \phi(\boldsymbol{x}^{(|\mathcal{X}|)}))$, respectively. By leveraging the identities $\phi(\boldsymbol{z})^\top\phi(\boldsymbol{x}) = (\mathbf{Q}\bar{\psi}(\boldsymbol{z}))^\top\bar{\psi}(\boldsymbol{x})$ and $\phi(\boldsymbol{z})^\top\phi(\boldsymbol{x}) = (\mathbf{Q}\bar{\psi}(\boldsymbol{x}))^\top(\mathbf{Q}\bar{\psi}(\boldsymbol{z}))$ shown in Eqs. (221) and (222), we have

$$\mathbf{\Phi}^\top\phi(\boldsymbol{x}) = \widetilde{\mathbf{\Psi}}^\top\bar{\psi}(\boldsymbol{x}) \;\text{ and }\; \mathbf{\Phi}^\top\mathbf{\Phi} = \widetilde{\mathbf{\Psi}}^\top\widetilde{\mathbf{\Psi}}. \tag{223}$$

Using the above identities, we prove Eqs. (214) and (215). Firstly, regarding Eq. (214), by using the unit vector $\boldsymbol{e}_{\boldsymbol{z}} \in \mathbb{R}^{|\mathcal{X}|}$ such that $(\boldsymbol{e}_{\boldsymbol{z}})_i = \mathbb{1}\{\boldsymbol{z} = \boldsymbol{x}^{(i)}\}$, we have

$$\phi(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}' \in \mathcal{X}} \phi(\boldsymbol{x}')\phi(\boldsymbol{x}')^\top + \frac{\lambda}{T} \boldsymbol{I}_s \right)^{-1} \phi(\boldsymbol{z}) = \phi(\boldsymbol{x})^\top \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \frac{\lambda}{T} \boldsymbol{I}_s \right)^{-1} \boldsymbol{\Phi} \boldsymbol{e}_{\boldsymbol{z}} \tag{224}$$

$$= \phi(\boldsymbol{x})^\top \boldsymbol{\Phi} \left(\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \boldsymbol{e}_{\boldsymbol{z}} \tag{225}$$

$$= \bar{\psi}(\boldsymbol{x})^\top \widetilde{\boldsymbol{\Psi}} \left(\widetilde{\boldsymbol{\Psi}}^\top \widetilde{\boldsymbol{\Psi}} + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \boldsymbol{e}_{\boldsymbol{z}} \tag{226}$$

$$= \bar{\psi}(\boldsymbol{x})^\top \left(\widetilde{\boldsymbol{\Psi}} \widetilde{\boldsymbol{\Psi}}^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \widetilde{\boldsymbol{\Psi}} \boldsymbol{e}_{\boldsymbol{z}} \tag{227}$$

$$= \bar{\psi}(\boldsymbol{x})^\top \left((\mathbf{Q}\bar{\boldsymbol{\Psi}})(\mathbf{Q}\bar{\boldsymbol{\Psi}})^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \mathbf{Q} \bar{\psi}(\boldsymbol{z}), \tag{228}$$

where Eqs. (225) and (227) follow from Lemma H.2, and Eq. (226) uses Eq. (223). Since $\widetilde{G}(\lambda) = (\mathbf{Q}\bar{\boldsymbol{\Psi}})(\mathbf{Q}\bar{\boldsymbol{\Psi}})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}$ and $\bar{\psi}(\boldsymbol{x}) = \sqrt{P(\boldsymbol{x})}\psi(\boldsymbol{x})$, the above equation implies Eq. (214). Regarding Eq. (215), we have

$$\widetilde{k}(\boldsymbol{x}, \boldsymbol{x}) - \phi(\boldsymbol{x})^\top \phi(\boldsymbol{x}) + \frac{\lambda}{T} \phi(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}' \in \mathcal{X}} \phi(\boldsymbol{x}')\phi(\boldsymbol{x}')^\top + \frac{\lambda}{T} \boldsymbol{I}_s \right)^{-1} \phi(\boldsymbol{x}) \tag{229}$$

$$= \widetilde{k}(\boldsymbol{x}, \boldsymbol{x}) - \phi(\boldsymbol{x})^\top \left[\boldsymbol{I}_s - \frac{\lambda}{T} \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \frac{\lambda}{T} \boldsymbol{I}_s \right)^{-1} \right] \phi(\boldsymbol{x}) \tag{230}$$

$$= \widetilde{k}(\boldsymbol{x}, \boldsymbol{x}) - \phi(\boldsymbol{x})^\top \boldsymbol{\Phi} \left(\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \boldsymbol{\Phi}^\top \phi(\boldsymbol{x}) \tag{231}$$

$$= \bar{\psi}(\boldsymbol{x})^\top \bar{\psi}(\boldsymbol{x}) - \bar{\psi}(\boldsymbol{x})^\top \widetilde{\boldsymbol{\Psi}} \left(\widetilde{\boldsymbol{\Psi}}^\top \widetilde{\boldsymbol{\Psi}} + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \widetilde{\boldsymbol{\Psi}}^\top \bar{\psi}(\boldsymbol{x}) \tag{232}$$

$$= \bar{\psi}(\boldsymbol{x})^\top \left(\boldsymbol{I}_{|\mathcal{X}|} - \widetilde{\boldsymbol{\Psi}} \left(\widetilde{\boldsymbol{\Psi}}^\top \widetilde{\boldsymbol{\Psi}} + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \widetilde{\boldsymbol{\Psi}}^\top \right) \bar{\psi}(\boldsymbol{x}) \tag{233}$$

$$= \frac{\lambda}{T} \bar{\psi}(\boldsymbol{x})^\top \left(\widetilde{\boldsymbol{\Psi}} \widetilde{\boldsymbol{\Psi}}^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right)^{-1} \bar{\psi}(\boldsymbol{x}), \tag{234}$$

where Eqs. (231) and (234) follow from Lemma H.2, and Eq. (232) uses Eq. (223). By aligning the above equation, we obtain Eq. (215). □

Lemma H.7 (Accuracy and computational cost of recursive RLS-Nyström with weighted kernel, adapted from Theorem 8 in [41])**.** *Fix any $T \in \mathbb{N}_+$, $\lambda \geq 1$, $\delta \in (0, 1/32)$, any finite input domain $\mathcal{X} \coloneqq \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(|\mathcal{X}|)}\}$, any probability mass function $P : \mathcal{X} \to [0, 1]$, and any positive definite kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $\forall \boldsymbol{x} \in \mathcal{X},\ k(\boldsymbol{x}, \boldsymbol{x}) \leq 1$. Furthermore, let $\psi(\cdot) : \mathcal{X} \to \mathbb{R}^{|\mathcal{X}|}$ be a feature map of k such that $k(\boldsymbol{x}, \boldsymbol{x}') = \psi(\boldsymbol{x})^\top \psi(\boldsymbol{x}')$ holds for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$. We also define the weighted kernel $\widetilde{k}$ as $\widetilde{k}(\boldsymbol{x}, \boldsymbol{x}') = \bar{\psi}(\boldsymbol{x})^\top \bar{\psi}(\boldsymbol{x}')$, where $\bar{\psi}(\boldsymbol{x}) = \sqrt{P(\boldsymbol{x})}\psi(\boldsymbol{x})$. Then, when running Alg. 3 with the inputs $(\mathcal{X},\ \widetilde{k},\ \lambda/T,\ \delta)$, the following three statements simultaneously hold with probability at least $1 - 3\delta$:*

1. *The number of sampled columns $s \in [|\mathcal{X}|]$ satisfies $s \leq s_{\max}(4\gamma_T, \delta)$, where $s_{\max}(w, z) = 384(w + 1)\log((w + 1)/z)$.*
2. *The computational time of Alg. 3 is at most $O(|\mathcal{X}| s^2_{\max}(4\gamma_T, \delta))$.*
3. *The weighted sampling matrix $\mathbf{S} \in \mathbb{R}^{|\mathcal{X}| \times s}$ returned by Alg. 3 satisfies the following:*

$$\frac{1}{2} \left(\bar{\boldsymbol{\Psi}} \bar{\boldsymbol{\Psi}}^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right) \preceq \bar{\boldsymbol{\Psi}} \mathbf{S}\mathbf{S}^\top \bar{\boldsymbol{\Psi}}^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \preceq \frac{3}{2} \left(\bar{\boldsymbol{\Psi}} \bar{\boldsymbol{\Psi}}^\top + \frac{\lambda}{T} \boldsymbol{I}_{|\mathcal{X}|} \right), \tag{235}$$

where $\bar{\boldsymbol{\Psi}} = (\bar{\psi}(\boldsymbol{x}^{(1)}), \ldots, \bar{\psi}(\boldsymbol{x}^{(|\mathcal{X}|)})) \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$.

In statements 1 and 2, $\gamma_T \coloneqq \gamma_T(\lambda, \mathcal{X})$ denotes the maximum information gain of kernel k on $\mathcal{X}$ with the variance parameter λ.

Proof. Let $\bar{\mathbf{K}} \coloneqq [\widetilde{k}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})]_{i,j \in [|\mathcal{X}|]} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$ be the Gram matrix of the weighted kernel function $\widetilde{k}$. Let us define the following quantity $d_{\mathrm{eff}}(\bar{\mathbf{K}})$:

$$d_{\mathrm{eff}}(\bar{\mathbf{K}}) = \mathrm{Tr}\left(\bar{\mathbf{K}}\left(\bar{\mathbf{K}} + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\right). \tag{236}$$

Here, note that $\bar{\mathbf{K}} = \bar{\boldsymbol{\Psi}}^\top \bar{\boldsymbol{\Psi}}$ holds. Then, with probability at least $1 - 3\delta$, statement 3 and the following two statements simultaneously hold as the direct consequence of Theorem 8 in [41]:

- The number of sampled columns $s \in [|\mathcal{X}|]$ satisfies $s \leq s_{\max}(d_{\mathrm{eff}}(\bar{\mathbf{K}}), \delta)$.
- The computational time of Alg. 3 is at most $O(|\mathcal{X}| s_{\max}^2(d_{\mathrm{eff}}(\bar{\mathbf{K}}), \delta))$.

Thus, it remains to prove the inequality $d_{\mathrm{eff}}(\bar{\mathbf{K}}) \leq 4\gamma_T$. By leveraging the feature representation $\bar{\mathbf{K}} = \bar{\boldsymbol{\Psi}}^\top \bar{\boldsymbol{\Psi}}$, we obtain the desired inequality as follows:

$$\begin{aligned}
d_{\mathrm{eff}}(\bar{\mathbf{K}}) &= \mathrm{Tr}\left(\bar{\mathbf{K}}\left(\bar{\mathbf{K}} + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\right) && (237)\\
&= \mathrm{Tr}\left(\bar{\boldsymbol{\Psi}}^\top\bar{\boldsymbol{\Psi}}\left(\bar{\boldsymbol{\Psi}}^\top\bar{\boldsymbol{\Psi}} + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\right) && (238)\\
&= \mathrm{Tr}\left(\bar{\boldsymbol{\Psi}}^\top\left(\bar{\boldsymbol{\Psi}}\bar{\boldsymbol{\Psi}}^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\bar{\boldsymbol{\Psi}}\right) && (239)\\
&= \sum_{\boldsymbol{x}\in\mathcal{X}} \bar{\psi}(\boldsymbol{x})^\top\left(\sum_{\boldsymbol{x}'\in\mathcal{X}} \bar{\psi}(\boldsymbol{x}')\bar{\psi}(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\bar{\psi}(\boldsymbol{x}) && (240)\\
&= \sum_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{x})\psi(\boldsymbol{x})^\top\left(\sum_{\boldsymbol{x}'\in\mathcal{X}} P(\boldsymbol{x}')\psi(\boldsymbol{x}')\psi(\boldsymbol{x}')^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\psi(\boldsymbol{x}) && (241)\\
&\leq 4\gamma_T, && (242)
\end{aligned}$$

where Eq. (239) follows from Lemma H.2, and Eq. (242) follows from Lemma 3.3 with $\lambda \geq 1$. Thus, we complete the proof. □

Lemma H.8 (Uniform upper bound of the approximated optimistic estimator $\widetilde{u}_t$)**.** *Fix any $T \in \mathbb{N}_+$, $\alpha \in (0, 1)$, $\eta > 0$, $\beta > 0$, and $\lambda \geq 1$. Fix any finite input domain $\mathcal{X}$. Suppose that Assumption 2.1 holds. Then, when running Alg. 5, under event $\cap_{t\in[T]} A_t$ (Eq.* (86)*), the following upper bound of the approximated optimistic estimator $\widetilde{u}_t$ holds for any $t \in [T]$ and $\boldsymbol{x} \in \mathcal{X}$:*

$$\widetilde{u}_t(\boldsymbol{x}) \leq \frac{4(3+\sqrt{6})B\gamma_T}{\alpha} + 2\beta\sqrt{\frac{3\gamma_T}{\alpha}} \leq \frac{2\sqrt{3}}{\alpha}\left(2(\sqrt{3}+\sqrt{2})B\gamma_T + \beta\sqrt{\gamma_T}\right). \tag{243}$$

Proof. From the definition of $\widetilde{f}_t$, we have

$$\begin{aligned}
\widetilde{f}_t(\boldsymbol{x}) &= \psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}\widetilde{\psi}_t(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t) && (244)\\
&= \psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t) - \psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}\left(\psi(\boldsymbol{x}_t) - \mathbf{Q}_t\psi(\boldsymbol{x}_t)\right) f_t(\boldsymbol{x}_t) && (245)\\
&= \psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t) - \psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)\psi(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t) && (246)\\
&\leq B|\psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x}_t)| + B|\psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)\psi(\boldsymbol{x}_t)|. && (247)
\end{aligned}$$

The first term of the above inequality is bounded from above as

$$B|\psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x}_t)| \leq B\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}\|\psi(\boldsymbol{x}_t)\|_{\widetilde{G}_t(\lambda)^{-1}} \leq 3B\max_{\boldsymbol{x}'\in\mathcal{X}}\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}}, \tag{248}$$

where the last inequality follows from Lemma H.9. Regarding the second term $B|\psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)\psi(\boldsymbol{x}_t)|$, we have

$$
\begin{aligned}
&B|\psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1}(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)\psi(\boldsymbol{x}_t)| && (249)\\
&\leq B\|\widetilde{G}_t(\lambda)^{-1}\psi(\boldsymbol{x})\|_2 \|(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)\psi(\boldsymbol{x}_t)\|_2 && (250)\\
&= B\sqrt{\psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-2}\psi(\boldsymbol{x})}\sqrt{\psi(\boldsymbol{x}_t)^\top(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)^2\psi(\boldsymbol{x}_t)} && (251)\\
&\leq B\sqrt{\frac{T}{\lambda}}\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}}\sqrt{\psi(\boldsymbol{x}_t)^\top(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)\psi(\boldsymbol{x}_t)} && (252)\\
&\leq B\sqrt{\frac{3T}{\lambda}}\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}}\sqrt{\frac{2\lambda}{T}}\|\psi(\boldsymbol{x}_t)\|_{G_t(\lambda)^{-1}} && (253)\\
&\leq \sqrt{6}B\max_{\boldsymbol{x}'\in\mathcal{X}}\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}}, && (254)
\end{aligned}
$$

where:

- Eq. (250) follows from the Cauchy–Schwarz inequality.
- Eq. (252) follows from $\psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-2}\psi(\boldsymbol{x}) \leq \|\widetilde{G}_t(\lambda)^{-1}\|\|\psi(\boldsymbol{x})\|^2_{\widetilde{G}_t(\lambda)^{-1}} \leq \frac{T}{\lambda}\|\psi(\boldsymbol{x})\|^2_{\widetilde{G}_t(\lambda)^{-1}}$. Furthermore, note that $(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)^2 = (\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)$ holds since the matrix $\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t$ is also a projection matrix.
- Eq. (253) follows from $\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}} \leq \sqrt{3}\|\psi(\boldsymbol{x})\|_{G_t(\lambda)^{-1}}$ and $\psi(\boldsymbol{x}_t)^\top(\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q}_t)\psi(\boldsymbol{x}_t) \leq \frac{2\lambda}{T}\psi(\boldsymbol{x}_t)^\top G_t(\lambda)^{-1}\psi(\boldsymbol{x}_t)$, which are implied by Lemma H.9 and Lemma H.10, respectively.

Thus, the approximated optimistic estimator $\widetilde{u}_t$ satisfies the following upper bound:

$$
\begin{aligned}
\widetilde{u}_t(\boldsymbol{x}) &= \widetilde{f}_t(\boldsymbol{x}) + \beta\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}} && (255)\\
&\leq (3+\sqrt{6})B\max_{\boldsymbol{x}'\in\mathcal{X}}\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}} + \sqrt{3}\beta\max_{\boldsymbol{x}'\in\mathcal{X}}\|\psi(\boldsymbol{x}')\|_{G_t(\lambda)^{-1}}. && (256)
\end{aligned}
$$

Furthermore, as with the proof of Lemma H.5 (Eqs. (205)–(211)), we obtain the upper bound of $\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}}$ as $\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}} \leq 4\gamma_T/\alpha$. Hence, by combining Eq. (256) with this upper bound of $\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}}$, we obtain

$$
\widetilde{u}_t(\boldsymbol{x}) \leq \frac{4(3+\sqrt{6})B\gamma_T}{\alpha} + 2\beta\sqrt{\frac{3\gamma_T}{\alpha}} \leq \frac{2\sqrt{3}}{\alpha}\left(2(\sqrt{3}+\sqrt{2})B\gamma_T + \beta\sqrt{\gamma_T}\right). \tag{257}
$$

□

Lemma H.9 (Approximation accuracy of multiplicative form for the weighted kernel matrix, adapted from the proof of Lemma 6 in [15])**.** *Assume that the sampling matrix $\mathbf{S} \in \mathbb{R}^{|\mathcal{X}|\times s}$ satisfies Eq.* (235)*. Then, under the same notations in Lemma H.7, the following relation holds:*

$$
\sum_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|} \preceq 3\left(\mathbf{Q}\bar{\boldsymbol{\Psi}}(\mathbf{Q}\bar{\boldsymbol{\Psi}})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right), \tag{258}
$$

where $\mathbf{Q} \coloneqq \bar{\boldsymbol{\Psi}}\mathbf{S}\left[(\bar{\boldsymbol{\Psi}}\mathbf{S})^\top\bar{\boldsymbol{\Psi}}\mathbf{S}\right]^\dagger(\bar{\boldsymbol{\Psi}}\mathbf{S})^\top$ is the projection matrix into the subspace spanned by the column vectors of $\bar{\boldsymbol{\Psi}}\mathbf{S}$.

Proof. Since $\mathbf{Q}$ is the orthogonal projection matrix into the column span of $\bar{\boldsymbol{\Psi}}\mathbf{S}$, the identity $\mathbf{Q}\bar{\boldsymbol{\Psi}}\mathbf{S} = \bar{\boldsymbol{\Psi}}\mathbf{S}$ holds. Furthermore, by aligning Eq. (235), we can confirm

$$
\bar{\boldsymbol{\Psi}}\bar{\boldsymbol{\Psi}}^\top \succeq \frac{2}{3}\bar{\boldsymbol{\Psi}}\mathbf{S}\mathbf{S}^\top\bar{\boldsymbol{\Psi}}^\top - \frac{\lambda}{3T}\boldsymbol{I}_{|\mathcal{X}|} \quad\text{and}\quad \bar{\boldsymbol{\Psi}}\mathbf{S}\mathbf{S}^\top\bar{\boldsymbol{\Psi}}^\top \succeq \frac{1}{2}\bar{\boldsymbol{\Psi}}\bar{\boldsymbol{\Psi}}^\top - \frac{\lambda}{2T}\boldsymbol{I}_{|\mathcal{X}|}. \tag{259}
$$

Thus, we have

$$\mathbf{Q}\bar{\mathbf{\Psi}}(\mathbf{Q}\bar{\mathbf{\Psi}})^\top = \mathbf{Q}\bar{\mathbf{\Psi}}\bar{\mathbf{\Psi}}^\top\mathbf{Q}^\top \succeq \frac{2}{3}\mathbf{Q}\bar{\mathbf{\Psi}}\mathbf{S}\mathbf{S}^\top\bar{\mathbf{\Psi}}^\top\mathbf{Q}^\top - \frac{\lambda}{3T}\mathbf{Q}\mathbf{Q}^\top \tag{260}$$

$$= \frac{2}{3}\bar{\mathbf{\Psi}}\mathbf{S}\mathbf{S}^\top\bar{\mathbf{\Psi}}^\top - \frac{\lambda}{3T}\mathbf{Q} \tag{261}$$

$$\succeq \frac{1}{3}\bar{\mathbf{\Psi}}\bar{\mathbf{\Psi}}^\top - \frac{\lambda}{3T}\boldsymbol{I}_{|\mathcal{X}|} - \frac{\lambda}{3T}\mathbf{Q} \tag{262}$$

$$\succeq \frac{1}{3}\bar{\mathbf{\Psi}}\bar{\mathbf{\Psi}}^\top - \frac{2\lambda}{3T}\boldsymbol{I}_{|\mathcal{X}|}, \tag{263}$$

where the last line follows from $\boldsymbol{I}_{|\mathcal{X}|} \succeq \mathbf{Q}$, which is implied by the fact that the eigenvalue of the orthogonal projection matrix is 1 or 0. By aligning the above relation, we obtain

$$\bar{\mathbf{\Psi}}\bar{\mathbf{\Psi}}^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|} \preceq 3\left(\mathbf{Q}\bar{\mathbf{\Psi}}(\mathbf{Q}\bar{\mathbf{\Psi}})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right). \tag{264}$$

Note that $\sum_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top$ can be rewritten as $\sum_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{x})\psi(\boldsymbol{x})\psi(\boldsymbol{x})^\top = \bar{\mathbf{\Psi}}\bar{\mathbf{\Psi}}^\top$. Thus, the above inequality is the desired statement. □

Lemma H.10 (Approximation accuracy of the difference form for the weighted kernel matrix, adapted from the proof of Lemma 1 in [13])**.** *Assume that the sampling matrix $\mathbf{S} \in \mathbb{R}^{|\mathcal{X}|\times s}$ satisfies Eq.* (235)*. Then, under the same notations in Lemma H.7, the following relation holds:*

$$\boldsymbol{I}_{|\mathcal{X}|} - \mathbf{Q} \preceq \frac{2\lambda}{T}\left(\bar{\mathbf{\Psi}}\bar{\mathbf{\Psi}}^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}, \tag{265}$$

where $\mathbf{Q} = \bar{\mathbf{\Psi}}\mathbf{S}\left[(\bar{\mathbf{\Psi}}\mathbf{S})^\top\bar{\mathbf{\Psi}}\mathbf{S}\right]^\dagger(\bar{\mathbf{\Psi}}\mathbf{S})^\top$.

Proof. By leveraging the definition of matrix inverse, we have

$$\boldsymbol{I}_{|\mathcal{X}|} = \left(\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)\left(\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \tag{266}$$

$$= \bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top\left(\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} + \frac{\lambda}{T}\left(\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \tag{267}$$

$$\preceq \bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top\left(\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top\right)^\dagger + \frac{\lambda}{T}\left(\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \tag{268}$$

$$\preceq \bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top\left(\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top\right)^\dagger + \frac{2\lambda}{T}\left(\bar{\mathbf{\Psi}}\bar{\mathbf{\Psi}}^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \tag{269}$$

$$\preceq \mathbf{Q} + \frac{2\lambda}{T}\left(\bar{\mathbf{\Psi}}\bar{\mathbf{\Psi}}^\top + \frac{\lambda}{T}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}, \tag{270}$$

where Eq. (268) follows from the basic fact that $\boldsymbol{M}\boldsymbol{M}^\top(\boldsymbol{M}\boldsymbol{M}^\top + \lambda\boldsymbol{I})^{-1} \preceq \boldsymbol{M}\boldsymbol{M}^\top(\boldsymbol{M}\boldsymbol{M}^\top)^\dagger$ is valid for any matrix $\boldsymbol{M}$ and $\lambda > 0$, and Eq. (269) follows from Eq. (235). In addition, Eq. (270) follows from the uniqueness of the orthogonal projection matrix by noting that $\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top\left(\bar{\mathbf{\Psi}}\mathbf{S}(\bar{\mathbf{\Psi}}\mathbf{S})^\top\right)^\dagger$ is also an orthogonal projection matrix into the column span of $\bar{\mathbf{\Psi}}\mathbf{S}$ from the definition of the pseudo inverse. The above expression is the desired statement. □

Lemma H.11 (Upper bound for sup-norm of $\widetilde{u}_t(\cdot)$)**.** *Suppose that Assumption 2.1 holds. Then, for any $\boldsymbol{x} \in \mathcal{X}$ and $t \in [T]$, the approximated optimistic estimator $\widetilde{u}_t$ used in RLS-kernelized Exp3 (Alg. 5) or its variant (Alg. 6) satisfies*

$$|\widetilde{u}_t(\boldsymbol{x})| \leq \frac{BT}{\lambda} + \beta\sqrt{\frac{T}{\lambda}}. \tag{271}$$

Proof. From the definition of $\widetilde{f}_t(\boldsymbol{x})$, we have

$$|\widetilde{f}_t(\boldsymbol{x})| = \left|\psi(\boldsymbol{x})^\top \widetilde{G}_t(\lambda)^{-1} \widetilde{\psi}_t(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t)\right| \tag{272}$$

$$\leq B\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}} \|\widetilde{\psi}_t(\boldsymbol{x}_t)\|_{\widetilde{G}_t(\lambda)^{-1}} \tag{273}$$

$$\leq B\|\psi(\boldsymbol{x})\|_2 \|\widetilde{G}_t(\lambda)^{-1}\| \|\widetilde{\psi}_t(\boldsymbol{x}_t)\|_2 \tag{274}$$

$$\leq \frac{BT}{\lambda}, \tag{275}$$

where the first inequality is the consequence of the Cauchy–Schwarz inequality and $\|f_t\|_\infty \leq \|f_t\|_k \leq B$, and the second inequality follows from the definition of the matrix operator norm. Furthermore, the third inequality follows from $\|\psi(\boldsymbol{x})\|_2 = \sqrt{k(\boldsymbol{x},\boldsymbol{x})} \leq 1$, $\|\widetilde{\psi}_t(\boldsymbol{x})\|_2 = \|\mathbf{Q}_t\psi(\boldsymbol{x})\|_2 \leq \|\mathbf{Q}_t\| \|\psi(\boldsymbol{x})\|_2 \leq \sqrt{k(\boldsymbol{x},\boldsymbol{x})} \leq 1$, and the fact that the minimum eigenvalue of $\widetilde{G}_t(\lambda)$ is not smaller than λ/T. In addition, similarly to the above upper bound, we have

$$\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}} \leq \sqrt{\|\widetilde{G}_t(\lambda)^{-1}\|} \|\psi(\boldsymbol{x})\|_2 \leq \sqrt{\frac{T}{\lambda}}. \tag{276}$$

Thus, for any $\boldsymbol{x} \in \mathcal{X}$ and $t \in [T]$, the approximated optimistic estimator $\widetilde{u}_t$ satisfies

$$|\widetilde{u}_t(\boldsymbol{x})| \leq |\widetilde{f}_t(\boldsymbol{x})| + \beta\|\psi(\boldsymbol{x})\|_{\widetilde{G}_t(\lambda)^{-1}} \leq \frac{BT}{\lambda} + \beta\sqrt{\frac{T}{\lambda}}, \tag{277}$$

which is the desired inequality. □

H.4 Helper lemmas for RLS-Kernelized Exp3 with approximated MVR sequence

Overview. The lemmas in this subsection are leveraged in the proof of a variant of RLS-kernelized Exp3 using the approximated MVR sequence in Appendix G. Corollary H.12 and Lemma H.13 provide the approximation properties related to the standard kernel functions. We omit the proof of Corollary H.12 since it directly follows by noting the fact that the effective dimension of the kernel is upper bounded by MIG (see, e.g., [15]). Furthermore, Lemma H.13 also directly follows by adapting the proof of Lemma H.9 for the standard kernel instead of the weighted kernel; thus, we omit the proof of Lemma H.13. Lemma H.14 shows that the maximum of the posterior variance is also upper-bounded by MIG even under the approximated MVR sequence. This lemma serves as a replacement for the upper bound on the posterior variance in the exact MVR sequence in the proof. Finally, based on the result of Lemma H.13, we can prove Lemma H.15, which serves as a replacement for Lemma H.8 in the proof of original RLS-kernelized Exp3.

Corollary H.12 (Accuracy and computational cost of recursive RLS-Nyström with standard kernel, adapted from Theorem 8 in [41])**.** *Fix any $T \in \mathbb{N}_+$, $\lambda \geq 1$, $\delta \in (0, 1/32)$, any finite input set $\mathbf{X} \coloneqq \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(T)}\}$, and any positive definite kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Furthermore, let $\psi(\cdot) : \mathcal{X} \to \mathbb{R}^{|\mathcal{X}|}$ be a feature map of k such that $k(\boldsymbol{x},\boldsymbol{x}') = \psi(\boldsymbol{x})^\top\psi(\boldsymbol{x}')$ holds for any $\boldsymbol{x},\boldsymbol{x}' \in \mathcal{X}$. Then, when running Alg. 3 with the inputs $(\mathbf{X},\ k,\ \lambda,\ \delta)$, the following three statements simultaneously hold with probability at least $1 - 3\delta$:*

1. *The number of sampled columns $\widetilde{s} \in [T]$ satisfies $\widetilde{s} \leq s_{\max}(2\gamma_T, \delta)$, where $s_{\max}(w, z) = 384(w + 1)\log((w + 1)/z)$.*
2. *The computational time of Alg. 3 is at most $O(Ts^2_{\max}(2\gamma_T, \delta))$.*
3. *The weighted sampling matrix $\widetilde{\mathbf{S}} \in \mathbb{R}^{T\times\widetilde{s}}$ returned by Alg. 3 satisfies the following:*

$$\frac{1}{2}\left(\mathbf{\Psi}\mathbf{\Psi}^\top + \lambda\boldsymbol{I}_{|\mathcal{X}|}\right) \preceq \mathbf{\Psi}\widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^\top\mathbf{\Psi}^\top + \lambda\boldsymbol{I}_{|\mathcal{X}|} \preceq \frac{3}{2}\left(\mathbf{\Psi}\mathbf{\Psi}^\top + \lambda\boldsymbol{I}_{|\mathcal{X}|}\right), \tag{278}$$

where $\mathbf{\Psi} = (\psi(\boldsymbol{x}^{(1)}), \ldots, \psi(\boldsymbol{x}^{(T)})) \in \mathbb{R}^{|\mathcal{X}|\times T}$.

In statements 1 and 2, $\gamma_T \coloneqq \gamma_T(\lambda, \mathcal{X})$ denotes the maximum information gain of kernel k on $\mathcal{X}$ with the variance parameter λ.

Lemma H.13 (Approximation accuracy of multiplicative form for the standard kernel matrix, adapted from the proof of Lemma 6 in [15])**.** *Assume that the sampling matrix $\widetilde{\mathbf{S}} \in \mathbb{R}^{T \times \widetilde{s}}$ satisfies Eq.* (278)*. Then, under the same notations in Corollary H.12, the following relation holds:*

$$\frac{1}{3}\left(\widetilde{\mathbf{Q}}\mathbf{\Psi}(\widetilde{\mathbf{Q}}\mathbf{\Psi})^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}\right) \preceq \sum_{i=1}^{T} \psi(\boldsymbol{x}^{(i)})\psi(\boldsymbol{x}^{(i)})^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|} \preceq 3\left(\widetilde{\mathbf{Q}}\mathbf{\Psi}(\widetilde{\mathbf{Q}}\mathbf{\Psi})^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}\right), \tag{279}$$

where $\widetilde{\mathbf{Q}} \coloneqq \mathbf{\Psi}\widetilde{\mathbf{S}}\left[(\mathbf{\Psi}\widetilde{\mathbf{S}})^\top \mathbf{\Psi}\widetilde{\mathbf{S}}\right]^\dagger (\mathbf{\Psi}\widetilde{\mathbf{S}})^\top$ is the projection matrix into the subspace spanned by the column vectors of $\mathbf{\Psi}\widetilde{\mathbf{S}}$.

Lemma H.14 (Property of the approximated MVR-sequence)**.** *Fix any finite input domain $\mathcal{X}$. Fix any $T \in \mathbb{N}_+$, $\lambda \geq 1$, $\delta \in (0, \min\{1, T/32\})$, any positive definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with $\forall \boldsymbol{x} \in \mathcal{X}, k(\boldsymbol{x}, \boldsymbol{x}) \leq 1$. Then, when running Alg. 4 with the input $(\mathcal{X}, k, \lambda, T, \delta)$, the following event holds with probability at least $1 - 3\delta$:*

$$\max_{\boldsymbol{x} \in \mathcal{X}} \sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_T^{(MVR)}, \lambda\right) \leq \frac{36\lambda\gamma_T}{T}. \tag{280}$$

Proof. Let $\mathbf{\Psi}_t = (\psi(\widetilde{\boldsymbol{x}}_1^{(\mathrm{MVR})}), \ldots, \psi(\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})})) \in \mathbb{R}^{|\mathcal{X}| \times t}$ be a feature matrix defined by the approximated MVR-sequence. Furthermore, let $\widetilde{\mathbf{S}}_t \in \mathbb{R}^{(t-1) \times \widetilde{s}_t}$ be the sampling matrix returned by the recursive RLS-Nyström algorithm at round t of Alg. 4 (Line 3). Then, from Corollary H.12 and the union bound, we can confirm that the following statement holds with probability at least $1 - 3\delta$:

$$\forall t \in [T],\ \frac{1}{2}\left(\mathbf{\Psi}_{t-1}\mathbf{\Psi}_{t-1}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}\right) \preceq \mathbf{\Psi}_{t-1}\widetilde{\mathbf{S}}_t\widetilde{\mathbf{S}}_t^\top\mathbf{\Psi}_{t-1}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|} \preceq \frac{3}{2}\left(\mathbf{\Psi}_{t-1}\mathbf{\Psi}_{t-1}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}\right). \tag{281}$$

Therefore, it is enough to show the desired statement under the above event. Henceforth, we suppose that event (281) holds in the following arguments. By using Lemma H.13, we can obtain the following inequality:

$$\begin{aligned}
\bar{\sigma}^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right) &= \lambda\psi(\boldsymbol{x})^\top\left(\widetilde{\mathbf{Q}}_t\mathbf{\Psi}_{t-1}(\widetilde{\mathbf{Q}}_t\mathbf{\Psi}_{t-1})^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\psi(\boldsymbol{x}) && (282)\\
&\leq 3\lambda\psi(\boldsymbol{x})^\top\left(\mathbf{\Psi}_{t-1}\mathbf{\Psi}_{t-1}^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\psi(\boldsymbol{x}) && (283)\\
&= 3\sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \lambda\right), && (284)
\end{aligned}$$

where $\widetilde{\mathbf{Q}}_t = \mathbf{\Psi}_{t-1}\widetilde{\mathbf{S}}_t\left[(\mathbf{\Psi}_{t-1}\widetilde{\mathbf{S}}_t)^\top\mathbf{\Psi}_{t-1}\widetilde{\mathbf{S}}_t\right]^\dagger(\mathbf{\Psi}_{t-1}\widetilde{\mathbf{S}}_t)^\top$. Note that the identity of $\lambda\psi(\boldsymbol{x})^\top\left(\widetilde{\mathbf{Q}}_t\mathbf{\Psi}_{t-1}(\widetilde{\mathbf{Q}}_t\mathbf{\Psi}_{t-1})^\top + \lambda \boldsymbol{I}_{|\mathcal{X}|}\right)^{-1}\psi(\boldsymbol{x})$ and Eq. (12) is straightforwardly obtained by the proof of Lemma H.6 as the special case. Similar to the above inequalities, we can also obtain the converse inequality $\sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \lambda\right) \leq 3\bar{\sigma}^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right)$ by using Lemma H.13. Thus, for any $\boldsymbol{x} \in \mathcal{X}$ and $t \in [T]$, we have

$$\frac{1}{3}\sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \lambda\right) \leq \bar{\sigma}^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right) \leq 3\sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \lambda\right). \tag{285}$$

Here, note that the following statement also holds from the definition of $\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})}$:

$$\forall t \in [T],\ \max_{\boldsymbol{x} \in \mathcal{X}} \bar{\sigma}^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right) \leq \bar{\sigma}^2\left(\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right). \tag{286}$$

By combining Eq. (285) with Eq. (286), we have

$$\max_{\boldsymbol{x}\in\mathcal{X}} \sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_T^{(\mathrm{MVR})}, \lambda\right) \leq \frac{1}{T}\sum_{t=1}^{T} \max_{\boldsymbol{x}\in\mathcal{X}} \sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \lambda\right) \tag{287}$$

$$\leq \frac{3}{T}\sum_{t=1}^{T} \max_{\boldsymbol{x}\in\mathcal{X}} \bar{\sigma}^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right) \tag{288}$$

$$\leq \frac{3}{T}\sum_{t=1}^{T} \bar{\sigma}^2\left(\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \widetilde{\mathbf{S}}_t, \lambda\right) \tag{289}$$

$$\leq \frac{9}{T}\sum_{t=1}^{T} \sigma^2\left(\widetilde{\boldsymbol{x}}_t^{(\mathrm{MVR})}; \widetilde{\mathbf{X}}_{t-1}^{(\mathrm{MVR})}, \lambda\right) \tag{290}$$

$$\leq \frac{18\gamma_T}{T\log(1+\lambda^{-1})} \tag{291}$$

$$\leq \frac{36\lambda\gamma_T}{T}, \tag{292}$$

where Eq. (287) uses the monotonicity of the posterior variance with respect to the training input of the GP, Eq. (291) follows from Eq. (3), and Eq. (292) follows from the elementary inequality $\forall a \in [0,1], \log(1+a) \geq a/2$. □

Lemma H.15 (Uniform upper bound of the approximated optimistic estimator $\widetilde{u}_t$ with approximated MVR sequence)**.** *Fix any $T \in \mathbb{N}_+$, $\alpha \in (0,1)$, $\eta > 0$, $\beta > 0$, and $\lambda > 0$. Fix any finite input domain $\mathcal{X}$. Suppose that Assumption 2.1 holds. Then, when running Alg. 6, under event $\cap_{t\in[T]} A_t$ (see Eq.* (86)*) and $\max_{\boldsymbol{x}\in\mathcal{X}} \sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{\lceil T\alpha\rceil}^{(MVR)}, \lambda\right) \leq \frac{36\lambda\gamma_{\lceil T\alpha\rceil}}{\lceil T\alpha\rceil}$ (see Eq.* (280)*), the following upper bound of the approximated optimistic estimator $\widetilde{u}_t$ holds for any $t \in [T]$ and $\boldsymbol{x} \in \mathcal{X}$:*

$$\widetilde{u}_t(\boldsymbol{x}) \leq \frac{36(3+\sqrt{6})B\gamma_T}{\alpha} + \beta\sqrt{\frac{108\gamma_T}{\alpha}} \leq \frac{1}{\alpha}\left(36(3+\sqrt{6})B\gamma_T + \beta\sqrt{108\gamma_T}\right). \tag{293}$$

Proof. By applying the proof of Lemma H.8, we can confirm the following inequality (see Eq. (256)).

$$\widetilde{u}_t(\boldsymbol{x}) \leq (3+\sqrt{6})B \max_{\boldsymbol{x}'\in\mathcal{X}} \|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}} + \sqrt{3}\beta \max_{\boldsymbol{x}'\in\mathcal{X}} \|\psi(\boldsymbol{x}')\|_{G_t(\lambda)^{-1}}. \tag{294}$$

Then, we obtain the upper bound of $\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}}$ as

$$\|\psi(\boldsymbol{x}')\|^2_{G_t(\lambda)^{-1}} \leq \frac{\lceil T\alpha\rceil}{\alpha\lambda}\sigma^2\left(\boldsymbol{x}'; \widetilde{\mathbf{X}}_{\lceil T\alpha\rceil}^{(\mathrm{MVR})}, \lambda\right) \tag{295}$$

$$\leq \frac{36\gamma_T}{\alpha}, \tag{296}$$

where the first inequality follows from the same arguments provided in Eqs. (205)–(210), and the second inequality follows from event $\max_{\boldsymbol{x}\in\mathcal{X}} \sigma^2\left(\boldsymbol{x}; \widetilde{\mathbf{X}}_{\lceil T\alpha\rceil}^{(\mathrm{MVR})}, \lambda\right) \leq \frac{36\lambda\gamma_{\lceil T\alpha\rceil}}{\lceil T\alpha\rceil}$. Hence, we obtain

$$\widetilde{u}_t(\boldsymbol{x}) \leq \frac{36(3+\sqrt{6})B\gamma_T}{\alpha} + \beta\sqrt{\frac{108\gamma_T}{\alpha}} \leq \frac{1}{\alpha}\left(36(3+\sqrt{6})B\gamma_T + \beta\sqrt{108\gamma_T}\right). \tag{297}$$

□

H.5 Helper lemma for lower bounds

Lemma H.16 (Application of the lower bounds for stochastic setting, adapted from Theorem 2 in [10])**.** *Fix any $d \in \mathbb{N}_+$, $B, \ell, \nu, \sigma^2 > 0$ and $\mathcal{X} = [0,1]^d$. Fix kernel function $k : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$. Assume that B, d, ℓ, and ν are $\Theta(1)$. Furthermore, let us consider the adversarial KB problem with $f_t(\cdot) \coloneqq f(\cdot) + \widetilde{\eta}_t k(\mathbf{0}, \cdot)$, where $f \in \mathcal{F}_k(B) \coloneqq \{f \in \mathcal{H}_k \mid \|f\|_k \leq B\}$ and $\widetilde{\eta}_t \sim \mathcal{N}(0, \sigma^2)$. In addition,*

we assume $\sigma^2 = O(T)$ with a sufficiently small implied constant. Then, for any algorithm, there exists a function $f \in \mathcal{F}_k(B)$ such that

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}\left[\sum_{t=1}^{T} f(\boldsymbol{x}) - f(\boldsymbol{x}_t)\right] = \begin{cases} \Omega\left(\sqrt{T\sigma^2(\log(T/\sigma^2))^{d/2}}\right) & \text{if } k = k_{\mathrm{SE}}, \\ \Omega\left(\sigma^{\frac{2\nu}{2\nu+d}} T^{\frac{\nu+d}{2\nu+d}}\right) & \text{if } k = k_{\mathrm{Mat}}. \end{cases} \tag{298}$$

The above lemma is obtained with a subtle modification of the proof of Theorem 2 in [10]. Thus, we omit the full proof while providing a description of the required modification. The only part of the proof of Theorem 2 in [10] that requires modification is the upper bound of Kullback–Leibler (KL) divergence in the change of measure argument. In this section, hereafter, we use the same notations as those in [10]. When we apply the proof of [10], the maximum KL divergence term: $\bar{D}^j_{f,f'}$ under the two functions f, f' and some partition $\mathcal{R}_j \subset \mathcal{X}$ indexed with j becomes different from the original proof. Since they study the Gaussian noise model whose variance σ^2 is homogeneous over the input domain, the term $\bar{D}^j_{f,f'}$ is given as

$$\bar{D}^j_{f,f'} = \max_{\boldsymbol{x}\in\mathcal{R}_j} \frac{(f(\boldsymbol{x}) - f'(\boldsymbol{x}))^2}{2\sigma^2}. \tag{299}$$

See Lemma 1 and Eqs. (9) and (19) in [10]. On the other hand, the term $\bar{D}^j_{f,f'}$ in our model takes the following form due to the input dependence of the noise term $\widetilde{\eta}_t k(\mathbf{0}, \cdot)$:

$$\bar{D}^j_{f,f'} = \max_{\boldsymbol{x}\in\mathcal{R}_j} \frac{(f(\boldsymbol{x}) - f'(\boldsymbol{x}))^2}{2\sigma^2 k^2(\mathbf{0}, \boldsymbol{x})}. \tag{300}$$

The upper bound of the above term $\bar{D}^j_{f,f'}$ can be further obtained as

$$\bar{D}^j_{f,f'} \leq \max_{\boldsymbol{x}\in\mathcal{R}_j} \frac{(f(\boldsymbol{x}) - f'(\boldsymbol{x}))^2}{2\sigma^2 k^2(\mathbf{0}, \mathbf{1})}, \tag{301}$$

where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$. By using the above upper bound instead of the exact value of the KL divergence $\frac{(f(\boldsymbol{x})-f'(\boldsymbol{x}))^2}{2\sigma^2}$, we can obtain the desired lower bound by directly following the proof of [10]. Specifically, the resulting lower bound matches the lower bound in [10] with the smaller noise variance $\sigma^2 k^2(\mathbf{0}, \mathbf{1})$. Note that $k^2(\mathbf{0}, \mathbf{1})$ depends on only d, ν, and ℓ, which are $\Theta(1)$; thus, we can recover the stochastic KB lower bound up to (d, ν, ℓ)-dependent constant factors.

I Additional discussion about exploration distribution

Our algorithm uses the empirical distribution of the MVR sequence as the exploration distribution π. In this section, we discuss the other possible choices of π.

Exploration distribution based on generalized G-optimal design. The role of exploration distribution in our proof is to obtain Lemmas H.5 and H.8, which are required for applying Exp3 analysis by guaranteeing $\forall \boldsymbol{x} \in \mathcal{X}, \eta u_t(\boldsymbol{x}) \leq 1$ and $\forall \boldsymbol{x} \in \mathcal{X}, \eta \widetilde{u}_t(\boldsymbol{x}) \leq 1$ (see Lemma H.4). Although we used the MVR-based empirical distribution to obtain Lemmas H.5 and H.8, this is not the only choice. The requirement of $\pi(\cdot)$ for extending our proof of Lemmas H.5 and H.8 is that the upper bound of the following supremum of the quadratic form increases with $O(\gamma_{\lceil T\alpha \rceil})$:

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \psi(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}'\in\mathcal{X}} \pi(\boldsymbol{x}')\psi(\boldsymbol{x}')\psi(\boldsymbol{x}')^\top + \frac{\lambda}{T\alpha}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \psi(\boldsymbol{x}) = O(\gamma_{\lceil T\alpha \rceil}). \tag{302}$$

See Eqs. (205)–(211) in the proof of Lemma H.5. From the above fact, the natural choice is to choose the following distribution π^* by minimizing the left-hand side of the above expression:

$$\pi^* \in \operatorname*{argmin}_{P\in\mathcal{P}_\mathcal{X}} \sup_{\boldsymbol{x}\in\mathcal{X}} \psi(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}'\in\mathcal{X}} P(\boldsymbol{x}')\psi(\boldsymbol{x}')\psi(\boldsymbol{x}')^\top + \frac{\lambda}{T\alpha}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \psi(\boldsymbol{x}), \tag{303}$$

where $\mathcal{P}_{\mathcal{X}} \coloneqq \{P : \mathcal{X} \to [0,1] \mid \sum_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{x}) = 1\}$ denotes the set of all probability mass functions on $\mathcal{X}$. This π^* is the generalization of G-optimal design [e.g., Chap. 21 in 36] and is used in stochastic KB literature [16]. Even if we replace π with π^* in Algs. 1 and 5, we can obtain the same regret guarantees as those in Theorems 3.1 and 4.1, and Corollary 3.2. Then, the resulting algorithm with π^* can be interpreted as the kernelized extension of Exp3 with Kiefer-Wolfowitz exploration in adversarial LB (Alg. 3 in [72]). The drawback of this approach is the computational cost of π^*. To our knowledge, even in the calculation of the near-optimal approximation of π^*, the total computational cost is $\Omega(|\mathcal{X}|^3)$. For example, the well-known iterative method based on the Frank-Wolfe algorithm [e.g., Chap. 21 in 36] requires $O(|\mathcal{X}|\log\log|\mathcal{X}|)$ iterations with $O(|\mathcal{X}|^3)$ per-iteration complexity[12]. For the Frank–Wolfe implementation described above, the resulting bound is $O(|\mathcal{X}|^4\log\log|\mathcal{X}|)$. In the typical KB regime $|\mathcal{X}| \gg T$, this bound is asymptotically larger than the $O(|\mathcal{X}|\lceil T\alpha\rceil^3)$ bound for preparing the MVR-based exploration distribution π.

Uniform distribution. Another possible choice is the uniform distribution $\pi_{\mathrm{unif}}(\cdot) \coloneqq 1/|\mathcal{X}|$. If we can use π_{unif}, the computational cost for exploration distribution prior to the algorithm run is completely omitted. In this case, by applying Lemma 3.3 to the quadratic term in Eq. (302), we observe that

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \psi(\boldsymbol{x})^\top \left(\sum_{\boldsymbol{x}'\in\mathcal{X}} \pi_{\mathrm{unif}}(\boldsymbol{x}')\psi(\boldsymbol{x}')\psi(\boldsymbol{x}')^\top + \frac{\lambda}{T\alpha}\boldsymbol{I}_{|\mathcal{X}|}\right)^{-1} \psi(\boldsymbol{x}) \le \frac{\lceil T\alpha\rceil}{\lambda} \sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbf{X}_{\mathrm{unif}}}[\sigma^2(\boldsymbol{x};\mathbf{X}_{\mathrm{unif}},\lambda)], \tag{304}$$

where $\mathbf{X}_{\mathrm{unif}} \coloneqq (\boldsymbol{x}_{\mathrm{unif}}^{(1)},\ldots,\boldsymbol{x}_{\mathrm{unif}}^{(\lceil T\alpha\rceil)})$ with $\boldsymbol{x}_{\mathrm{unif}}^{(1)},\ldots,\boldsymbol{x}_{\mathrm{unif}}^{(\lceil T\alpha\rceil)} \sim_{\text{i.i.d.}} \pi_{\mathrm{unif}}$. Thus, if $\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbf{X}_{\mathrm{unif}}}[\sigma^2(\boldsymbol{x};\mathbf{X}_{\mathrm{unif}},\lambda)] = O\left(\frac{\lambda\gamma_{\lceil T\alpha\rceil}}{\lceil T\alpha\rceil}\right)$, we can verify the condition (302). The examination for the upper bound $\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbf{X}_{\mathrm{unif}}}[\sigma^2(\boldsymbol{x};\mathbf{X}_{\mathrm{unif}},\lambda)] = O\left(\frac{\lambda\gamma_{\lceil T\alpha\rceil}}{\lceil T\alpha\rceil}\right)$ is found in the stochastic KB literature [50]. Salgia et al. [50] show that $\sup_{\boldsymbol{x}\in\mathcal{X}} \sigma^2(\boldsymbol{x};\mathbf{X}_{\mathrm{unif}},\lambda) = O\left(\frac{\lambda\gamma_{\lceil T\alpha\rceil}}{\lceil T\alpha\rceil}\right)$ asymptotically holds with high probability under the *uniform boundedness assumption* on the eigenfunctions of the kernel with respect to π_{unif}. However, to our knowledge, the validity of the uniform boundedness assumption for commonly-used kernels in KB (such as SE or ν-Matérn kernels) is unclear[13]. For example, see discussion in Chap. 4.4 in [32]. Thus, we believe that further careful theoretical examination is desired for the utilization of π_{unif}.

J Simulation experiments

We conduct simulation experiments to confirm the empirical behavior of our algorithms.

Input domain, kernel, and strategy of the environment. We conduct experiments using the following four input domains $\mathcal{X}$:

- **Setting 1.** We set $\mathcal{X}$ as 400 uniformly spaced grid points of $[0,1]$.
- **Setting 2.** We set $\mathcal{X}$ as 20×20 uniformly spaced grid points of $[0,1]^2$.
- **Setting 3.** We set $\mathcal{X}$ as 10000 uniformly spaced grid points of $[0,1]$.
- **Setting 4.** We set $\mathcal{X}$ as 100×100 uniformly spaced grid points of $[0,1]^2$.

For each setting of input domain, we conduct experiments with three kernel functions: (i) the SE kernel, (ii) the ν-Matérn kernel with $\nu = 5/2$, and (iii) the ν-Matérn kernel with $\nu = 3/2$. For each kernel, we set the lengthscale parameter ℓ as $\ell = 0.3\sqrt{d}$. To define f_t in the experiment, we consider the fully-adversarial environment. Specifically, with some pre-defined candidate set $\mathcal{F} \subset \mathcal{F}_k(B)$ and the learner's sampling distribution P_t, we assume that the environment chooses f_t so that the learner's conditional expected instantaneous regret becomes highest, as

[12] Specifically, even if we rely on the rank-one update formula of the matrix inverse, the calculation of the maximum of the quadratic term per-iteration requires $O(|\mathcal{X}|^2)$-calculation of the quadratic term for all input $\boldsymbol{x}\in\mathcal{X}$, which leads to $O(|\mathcal{X}|^3)$ per-iteration complexity.

[13] To our knowledge, the examples of the kernels for which the uniform boundedness assumption is rigorously verified are $1/2$-Matérn kernel with $d = 1$ (discussed in [32]) and Hilbert-space approximation of SE and ν-Matérn kernels [45, 59].

follows: $f_t \in \mathrm{argmax}_{f\in\mathcal{F}}(\max_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}) - \sum_{\boldsymbol{x}'\in\mathcal{X}} P_t(\boldsymbol{x}')f(\boldsymbol{x}'))$. We define the candidate set $\mathcal{F}$ as $\mathcal{F} = (f^{(i)})_{i\in[1000]}$ with $f^{(i)}(\cdot) = \min\{1, B/\|\widetilde{f}^{(i)}\|_k\}\widetilde{f}^{(i)}(\cdot)$ and $\widetilde{f}^{(i)}(\cdot) = \sum_{m=1}^{100} c^{(m,i)}k(\boldsymbol{x}^{(m,i)}, \cdot)$, where $\boldsymbol{x}^{(m,i)} \sim \mathrm{Uniform}(\mathcal{X})$ and $c^{(m,i)} \sim \mathrm{Uniform}(-1, 1)$ are independent random variables. Note that we can calculate $\|\widetilde{f}^{(i)}\|_k$ as $\|\widetilde{f}^{(i)}\|_k = \sqrt{\sum_{m,m'} c^{(m,i)}c^{(m',i)}k(\boldsymbol{x}^{(m,i)}, \boldsymbol{x}^{(m',i)})}$ under the definition of $\widetilde{f}^{(i)}$. Regarding the RKHS norm upper bound B, we set $B = 2$.

Algorithms. We compare the following algorithms:

- **Random.** At each round t, this algorithm draws $\boldsymbol{x}_t$ uniformly at random from $\mathcal{X}$.
- **APG-Exp3.** This is a variant of the Exp3 algorithm based on a linear approximation, proposed in [61]. We use the G-optimal design $\widetilde{\pi}^* \in \mathrm{argmin}_{P\in\mathcal{P}_\mathcal{X}} \max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{N}(\boldsymbol{x})\|_{(\sum_{\boldsymbol{x}'\in\mathcal{X}} P(\boldsymbol{x}')\boldsymbol{N}(\boldsymbol{x}')\boldsymbol{N}(\boldsymbol{x}')^\top)^{-1}}$ as the exploration distribution, where $\boldsymbol{N}(\boldsymbol{x}) \in \mathbb{R}^{D_T}$ is the feature vector based on the Newton basis functions. To calculate $\widetilde{\pi}^*$, we use the Frank-Wolfe algorithm [22]. The basis function $\boldsymbol{N}(\boldsymbol{x})$ and D_T are determined in the P-greedy algorithm as described in [61]. We fix the admissible error $\mathfrak{e}$ of the P-greedy algorithm as the theoretically suggested value $\mathfrak{e} = (\log|\mathcal{X}|)/T$. Regarding the choice of the learning rate $\eta > 0$ and mixing ratio $\alpha \in (0, 1)$[14], we set $\eta = c_1/\sqrt{D_T T}$ and $\alpha = c_2\eta D_T$ for some constants $c_1 > 0$ and $c_2 > 0$. See [61] for details. Here, these settings are based on the theoretically suggested diminishing rate of η and α for achieving $\widetilde{O}(\sqrt{TD_T})$ regret proved in [61]. For each kernel, we fix the constants c_1 and c_2 based on the average performance over 10 different seeds under setting 1 described above, with $T = 50$. Specifically, for each kernel, we choose the constants c_1 and c_2 by grid search across $c_1, c_2 \in \{0.1, 0.5, 1.0, 5.0, 10.0\}$. The selected settings of c_1, c_2 are $(c_1, c_2) = (0.5, 0.1)$, $(c_1, c_2) = (0.5, 0.1)$, and $(c_1, c_2) = (0.5, 0.1)$ for the SE, 5/2-Matérn, and 3/2-Matérn kernels, respectively.
- **Kernelized Exp3.** This is the kernelized Exp3 algorithm proposed in Sec. 3. Regarding the choice of the parameters, based on the theoretically suggested values in Theorem 3.1, we set $\eta = c_3/\sqrt{T\bar{\gamma}_T}$ and $\alpha = c_4\eta\bar{\gamma}_T$ with some constants $c_3 > 0$ and $c_4 > 0$. Here, $\bar{\gamma}_T$ denotes a known growth rate upper bound on MIG, which is defined as $\bar{\gamma}_T = (\log T)^{d+1}(\log\log T)^{-d}$ and $\bar{\gamma}_T = T^{\frac{d}{2\nu+d}}(\log T)^{\frac{4\nu+d}{2\nu+d}}$ for $k = k_{\mathrm{SE}}$ and $k = k_{\mathrm{Mat}}$, respectively [27, 28]. For each kernel, to choose constants c_3, c_4, and the regularization parameter $\lambda \geq 1$, we conduct the grid search across $c_3, c_4 \in \{0.1, 0.5, 1.0, 5.0, 10.0\}$ and $\lambda \in \{1.0, 5.0, 10.0\}$ as with the parameter selections of APG-Exp3. Here, we set the confidence width parameter β as the theoretically suggested value $\beta := B\sqrt{\lambda/T}$ in Theorem 3.1. Here, we omit the full experiments on kernelized Exp3 for the large-scale input domain setting (settings 3 and 4) due to its high computational cost; however, as a baseline for the computational time, we report the estimated per-round computational time for kernelized Exp3 by running the algorithm for only 1 round. The selected settings of c_3, c_4, λ are $(c_3, c_4, \lambda) = (5.0, 0.1, 1.0)$, $(c_3, c_4, \lambda) = (10.0, 0.1, 10.0)$, and $(c_3, c_4, \lambda) = (5.0, 0.1, 5.0)$ for the SE, 5/2-Matérn, and 3/2-Matérn kernels, respectively.
- **RLS-Kernelized Exp3.** This is the RLS-based kernelized Exp3 algorithm proposed in Sec. 4. Regarding the choice of η, α, and λ, we use the same settings as used in kernelized Exp3. We set the confidence width parameter β as $\beta = B(1 + \sqrt{2})\sqrt{\lambda/T}$ as suggested in Theorem 4.1. Furthermore, we fix the confidence level of RLS-Nyström as 0.1 for each iteration [15].

Resource for computation. Our experiments are conducted on an AMD EPYC 7702P 64-core processor with 16 GB RAM.

Results. We report the average regret and per-round computational time over 10 trials with different seeds. Tab. 3 shows the results. Firstly, we confirm that the regret of our proposed algorithms is

[14]The parameter α is defined as notation γ in [61].

[15]This setting formally corresponds to $\delta = 0.3T$ in Alg. 5 and is not covered by our theoretical guarantees. We adopt it as a practical choice to limit the sampling overhead associated with the horizon-dependent confidence allocation. In Tab. 3, we observe broadly similar average regret to kernelized Exp3 in settings 1 and 2, and lower per-round runtimes in settings 3 and 4 compared with the kernelized Exp3 runtime estimates obtained from a single round.

substantially lower than that of other baseline methods. Secondly, we confirm that the computational time of our algorithms is worse than that of the baseline algorithms. However, in the large input domain $|\mathcal{X}| = 10000$ (settings 3 and 4), the per-round cost of RLS-kernelized Exp3 is improved compared with the estimated per-round cost of the original kernelized Exp3.

In our experiments, we confirm that the computational time of RLS-kernelized Exp3 for SE kernel is worse than that of $5/2$-Matérn and $3/2$-Matérn. This phenomenon appears to contradict our theoretical guarantees, where the kernel with the small information gain leads to small computational cost of RLS-kernelized Exp3. However, as described in the algorithm setting, the regularization parameter $\lambda = 1.0$ chosen by the grid search for SE-kernel is smaller than $\lambda = 10.0$ and $\lambda = 5.0$ for $5/2$-Matérn and $3/2$-Matérn kernels. Therefore, RLS-kernelized Exp3 takes longer to run with the SE kernel than with the $5/2$-Matérn and $3/2$-Matérn kernels in our experiments, because the smaller regularization parameter λ leads to a larger information gain γ_T. We also found the computational cost of RLS-kernelized Exp3 exceeds that of APG-Exp3. We observe that the reason for this phenomenon is the large constant factor for $\widetilde{O}(|\mathcal{X}|\gamma_T^2)$-computation of the recursive Nyström algorithm provided in Lemma H.7.

Table 3: Experimental results for the average regret and per-round computational cost with $T \in \{50, 100, 150\}$. The abbreviations K-Exp3 and RLS-K-Exp3 in the table denote our kernelized Exp3 and RLS-kernelized Exp3 algorithms, respectively. All numerical values are rounded to one decimal place. Two standard errors are shown in parentheses. Note that we do not conduct experiments of kernelized Exp3 in settings 3 and 4, while we report the estimated per-round computational time by running the algorithm for only one round.

			SE Kernel			5/2-Matérn			3/2-Matérn		
			$T = 50$	$T = 100$	$T = 150$	$T = 50$	$T = 100$	$T = 150$	$T = 50$	$T = 100$	$T = 150$
Regret	Setting 1	Random	89.2 (2.8)	178.0 (4.3)	268.1 (4.4)	84.4 (2.6)	168.3 (3.7)	253.6 (4.1)	78.1 (4.0)	156.4 (7.6)	234.4 (7.2)
		APG-Exp3	28.2 (7.7)	42.9 (9.9)	60.7 (10.7)	28.0 (5.5)	48.2 (8.0)	69.3 (4.5)	23.5 (4.3)	42.9 (4.9)	60.6 (6.4)
		K-Exp3	18.8 (3.0)	22.4 (10.3)	34.9 (13.9)	16.1 (4.7)	32.7 (8.5)	41.5 (6.1)	14.8 (4.7)	28.2 (7.9)	42.9 (11.7)
		RLS-K-Exp3	20.4 (5.8)	27.9 (4.8)	43.7 (11.3)	19.1 (4.7)	31.0 (5.6)	51.9 (6.5)	14.7 (6.7)	24.5 (7.9)	41.4 (13.2)
	Setting 2	Random	79.6 (2.3)	158.2 (2.8)	235.1 (3.0)	75.0 (2.1)	149.2 (2.6)	221.5 (2.8)	72.4 (2.0)	144.0 (2.4)	213.7 (2.7)
		APG-Exp3	35.8 (7.6)	50.8 (7.5)	69.0 (10.2)	29.1 (7.5)	55.5 (13.6)	73.1 (12.3)	36.6 (9.3)	80.5 (12.6)	129.6 (19.1)
		K-Exp3	18.4 (4.0)	25.6 (4.5)	40.3 (7.7)	12.9 (3.2)	22.6 (5.9)	26.3 (8.2)	13.3 (4.4)	25.0 (8.5)	24.2 (8.2)
		RLS-K-Exp3	16.1 (3.1)	32.6 (5.5)	50.3 (10.4)	11.1 (4.5)	21.5 (6.0)	20.9 (5.7)	16.8 (4.0)	26.0 (8.2)	29.9 (6.5)
	Setting 3	Random	89.2 (2.8)	178.1 (4.3)	268.2 (4.4)	84.5 (2.6)	168.4 (3.7)	253.7 (4.1)	81.8 (2.5)	163.0 (3.5)	245.5 (3.9)
		APG-Exp3	27.4 (6.4)	45.4 (6.7)	65.9 (8.2)	31.0 (4.8)	47.1 (8.0)	67.7 (11.4)	23.5 (5.8)	41.5 (6.4)	62.6 (11.1)
		K-Exp3					N/A				
		RLS-K-Exp3	20.3 (5.4)	26.0 (10.3)	37.4 (13.0)	26.7 (5.2)	37.6 (8.4)	57.8 (11.6)	21.9 (5.7)	44.8 (11.2)	44.7 (9.1)
	Setting 4	Random	76.4 (2.5)	152.7 (4.8)	229.2 (5.5)	72.9 (3.4)	145.3 (3.4)	219.4 (3.3)	70.4 (3.3)	140.4 (3.3)	211.9 (3.2)
		APG-Exp3	35.8 (3.9)	73.0 (7.3)	95.3 (14.7)	33.1 (5.1)	61.9 (7.6)	116.6 (12.1)	38.8 (5.7)	87.3 (7.7)	132.4 (13.0)
		K-Exp3					N/A				
		RLS-K-Exp3	18.7 (5.2)	30.1 (8.8)	48.9 (10.8)	18.3 (4.5)	35.6 (7.2)	38.0 (5.4)	18.0 (5.6)	24.0 (4.0)	28.4 (7.7)
Per-round time cost [s]	Setting 1	Random	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		APG-Exp3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		K-Exp3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		RLS-K-Exp3	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)
	Setting 2	Random	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		APG-Exp3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		K-Exp3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		RLS-K-Exp3	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)
	Setting 3	Random	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		APG-Exp3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		K-Exp3					39.5 (0.0)				
		RLS-K-Exp3	0.9 (0.0)	1.1 (0.0)	1.3 (0.0)	0.2 (0.0)	0.4 (0.0)	0.6 (0.0)	0.5 (0.0)	0.7 (0.0)	1.1 (0.0)
	Setting 4	Random	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		APG-Exp3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		K-Exp3					40.6 (0.0)				
		RLS-K-Exp3	1.5 (0.0)	2.2 (0.0)	3.4 (0.0)	0.3 (0.0)	0.5 (0.0)	1.0 (0.0)	0.7 (0.0)	1.4 (0.0)	2.5 (0.0)

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The "contribution" paragraph in Sec. 1 summarize the paper's contributions. The abstract reflects these contributions.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Discussions provided in Sec. 6 include the limitations of this paper. Specifically, in Sec. 6, we clearly state the scalability issue in higher dimensions d due to the algorithm's discretization.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The full proofs of our theoretical results are provided in the appendix. Furthermore, the theoretical assumptions are clearly stated in Sec. 2.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: In Appendix J, we provide the comprehensive information for our experiments, including the problem parameters and the algorithms' parameters.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Our main contributions are not built on the experimental results. Furthermore, we believe that reproducing our experimental results based on the information in Appendix J is straightforward given their simple, low-dimensional settings.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: In Appendix J, we provide all the parameter settings for experiments, including definitions of the input domain, algorithms' parameters, and problem instances.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: All the experimental results in Appendix J are provided with the error bars that represent one standard error.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: In Appendix J, the compute resource information used in simulation experiments is provided.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We confirmed that the research conducted in the paper conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: Due to the theoretical nature of research work, we believe that there are no societal impacts.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper does not provide a release of data or models that have a high risk for misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: This paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.