Near-Optimal Algorithm for Non-Stationary Kernelized Bandits

Shogo Iwazaki
MI-6 Ltd.

Shion Takeno
Nagoya University, RIKEN AIP

Abstract

This paper studies a non-stationary kernelized bandit (KB) problem, also called time-varying Bayesian optimization, where one seeks to minimize the regret under an unknown reward function that varies over time. In particular, we focus on a near-optimal algorithm whose regret upper bound matches the regret lower bound. For this goal, we show the first algorithm-independent regret lower bound for non-stationary KB with squared exponential and Matérn kernels, which reveals that an existing optimization-based KB algorithm with slight modification is near-optimal. However, this existing algorithm suffers from feasibility issues due to its huge computational cost. Therefore, we propose a novel near-optimal algorithm called restarting phased elimination with random permutation (R-PERP), which bypasses the huge computational cost. A technical key point is the simple permutation procedures of query candidates, which enable us to derive a novel tighter confidence bound tailored to the non-stationary problems.

1 INTRODUCTION

Kernelized bandit (KB) problem [Srinivas et al., 2010], also called Gaussian process bandit or Bayesian optimization, is one of the important sequential decision-making problems where one seeks to minimize the regret under an unknown reward function via sequentially acquiring function evaluations. As the name suggests, in the KB problem, the underlying reward function is assumed to be an element of reproducing kernel Hilbert space (RKHS) induced by a known fixed kernel function. KB has been applied in many applications, such as materials discovery [Ueno et al., 2016], drug discovery [Korovina et al., 2020], and robotics [Berkenkamp et al., 2023]. In addition, the near-optimal KB algorithms, whose regret upper bound matches the regret lower bound derived in Scarlett et al. [2017], have been shown [Camilleri et al., 2021, Salgia et al., 2021, Li and Scarlett, 2022, Salgia et al., 2024].

Non-stationary KB [Bogunovic et al., 2016] considers the optimization under a non-stationary environment; that is, the reward function may change over time within some RKHS. This modification is crucial in many practical applications where an objective function varies over time, such as financial markets [Heaton and Lucas, 1999] and recommender systems [Hariri et al., 2015]. For example, Zhou and Shroff [2021], Deng et al. [2022] have proposed upper confidence bound (UCB)-based algorithms for the non-stationary KB problem and derived the upper bound of the cumulative regret. Recently, Hong et al. [2023] have proposed an optimization-based KB (OPKB[1]) algorithm, which achieves a tighter regret upper bound than that of Zhou and Shroff [2021], Deng et al. [2022]. This result implies that the OPKB algorithm is near-optimal for a linear kernel.

However, there are still two open problems with the non-stationary KB problem. First, although the OPKB algorithm is near-optimal for the linear kernel, the optimality for squared exponential (SE) and the Matérn kernels has not been revealed. Since SE and Matérn kernels are widely used in practice and of interest in theory [Shahriari et al., 2016], revealing the optimality for these two kernels is valuable. Second, although the OPKB algorithm achieves the known-best regret upper bound, the OPKB suffers from feasibility issues when the cardinality of the input set is huge. The OPKB algorithm requires running two costly procedures: (i) the construction of an explicit feature map of the kernel and (ii) the optimization whose dimension is $|\mathcal{X}|$, where $\mathcal{X}$ is the input set of the reward function. In particular, the procedure (ii) requires $O(|\mathcal{X}|^3)$ computation for every steps in the $|\mathcal{X}|$-dimensional optimization. Therefore, when the input set $\mathcal{X}$ is huge, even running the OPKB algorithm can be unrealistic.

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

[1] Hong et al. [2023] originally uses the terminology OPKB only for the algorithm under the stationary KB problems. For simplicity, in this paper, we use the terminology OPKB for the general KB algorithms constructed on optimization-based procedures.

Table 1: The comparison of existing and our algorithms for regrets and computational costs under a finite input set $\mathcal{X} \subset \mathbb{R}^d$. We denote V_T as the upper bound of the total variation of the sequence of underlying reward functions (precise definition is in Assumption 3) and M as the total iteration to solve $|\mathcal{X}|$-dimensional optimization problem in the OPKB. For the regret upper bound described in the table, we assume that V_T satisfies $V_T > c$, where $c > 0$ is any fixed constant.

Algorithm	Regret (SE)	Regret (Matérn)	Computational cost at step $t \leq T$ [2][3]
R/SW-GP-UCB	$\widetilde{\mathcal{O}}(T^{\frac{3}{4}} V_T^{\frac{1}{4}})$	$\widetilde{\mathcal{O}}(T^{\frac{12\nu+13d}{16\nu+8d}} V_T^{\frac{1}{4}})$	$\mathcal{O}(\lvert\mathcal{X}\rvert t^2)$
WGP-UCB	$\widetilde{\mathcal{O}}(T^{\frac{3}{4}} V_T^{\frac{1}{4}})$	$\widetilde{\mathcal{O}}(T^{\frac{12\nu+13d}{16\nu+8d}} V_T^{\frac{1}{4}})$	$\mathcal{O}(\lvert\mathcal{X}\rvert t^2)$
OPKB	$\widetilde{\mathcal{O}}(T^{\frac{2}{3}} V_T^{\frac{1}{3}})$	$\widetilde{\mathcal{O}}(T^{\frac{4\nu+3d}{6\nu+3d}} V_T^{\frac{1}{3}})$	$\mathcal{O}(M\lvert\mathcal{X}\rvert^3)$
R-PERP (Ours)	$\widetilde{\mathcal{O}}(T^{\frac{2}{3}} V_T^{\frac{1}{3}})$	$\widetilde{\mathcal{O}}(T^{\frac{2\nu+d}{3\nu+d}} V_T^{\frac{\nu}{3\nu+d}})$	$\mathcal{O}(\lvert\mathcal{X}\rvert t^2)$
Lower bounds (Ours)	$\widetilde{\Omega}(T^{\frac{2}{3}} V_T^{\frac{1}{3}})$	$\Omega(T^{\frac{2\nu+d}{3\nu+d}} V_T^{\frac{\nu}{3\nu+d}})$	

This paper tackles the above two open problems. Our contributions are summarized as follows:

- We show the first algorithm-independent lower bounds for the non-stationary KB problem for SE and Matérn kernel. Our results shows that any algorithms suffer $\widetilde{\Omega}(V_T^{\frac{1}{3}} T^{\frac{2}{3}})$ and $\widetilde{\Omega}(V_T^{\frac{\nu}{3\nu+d}} T^{\frac{2\nu+d}{3\nu+d}})$ worst-case regret for SE and Matérn kernel, respectively. Here, V_T denotes the upper bound of the total variation of the sequence of underlying reward functions. (See Assumption 3 for the formal definition.)

- Based on our lower bound, we confirm that the existing OPKB algorithm achieves nearly optimal regret for the SE kernel. In addition, we show that the OPKB algorithm with a slight modification achieves near-optimal regret for the Matérn kernels under known V_T.

- We further propose a novel phased elimination (PE) [Li and Scarlett, 2022] based algorithm, which bypasses the computationally hard procedures of OPKB. Our proposed algorithm, restarting PE with random permutation (R-PERP), combines the existing PE-based algorithm with simple permutation procedures of query candidates. The key to our regret analysis is the derivation of the tighter confidence bound (CB) tailored to the non-stationary setting based on such a permutation procedure.

In Table 1, we summarized the computational cost and the regret of the existing and our algorithms for SE and Matérn kernels.

[2]Strictly speaking, the dependence of t can become smaller in R/SW-GP-UCB and our R-PERP algorithms since these algorithms discard past training data of surrogate models at some intervals. Furthermore, OPKB and R-PERP choose query points using batch-based calculations; therefore, several time steps are skipped, and the computational costs of such time steps are zero.

[3]Note that KB problem usually focuses on the regime of $T \ll |\mathcal{X}|$.

Limitations The main limitation of this paper is that our near-optimal guarantees of OPKB and R-PERP rely on the prior knowledge of the upper bound V_T. However, it is worth noting that, even if V_T is unknown, our R-PERP algorithm achieves tighter regret than existing non-stationary KB algorithms [Zhou and Shroff, 2021, Deng et al., 2022] except for OPKB (see, Appendix D). Specifically, except for OPKB, R-PERP is the only algorithm whose regret upper bound is always sub-linear in Matérn kernels for fixed V_T. Furthermore, in contrast to our proposed R-PERP algorithm, OPKB is hard to apply the problem whose input set $\mathcal{X}$ is huge; therefore, the proposal of R-PERP is also valuable for situations whose V_T is unknown.

1.1 Related Works

The KB problems under stationary environments are extensively studied, and several algorithms are proposed, including Gaussian process UCB (GP-UCB) and Thompson sampling (TS)-based algorithm [Srinivas et al., 2010, Chowdhury and Gopalan, 2017]. Furthermore, Scarlett et al. [2017] derive the regret lower bound of stationary KB for SE and Matérn kernels and shows that existing regret upper bounds of UCB or TS-based algorithms are strictly sub-optimal for Matérn kernel. Recently, several works have tackled constructing an algorithm whose regret nearly matches lower bounds [Camilleri et al., 2021, Salgia et al., 2021, Li and Scarlett, 2022, Salgia et al., 2024].

As for the non-stationary KB problem, Bogunovic et al. [2016] first propose the UCB-based algorithms, called resetting GP-UCB (R-GP-UCB) and time-varying GP-UCB (TV-GP-UCB), which is based on the restart and reset strategy and the smoothly forgetting strategy for past observations, respectively. The analysis of Bogunovic et al. [2016] is based on Bayesian assumption whose reward functions follow some Gaussian process. The frequentist analysis of R-GP-UCB was later shown in Zhou and Shroff [2021]. Zhou and Shroff [2021] further proposed sliding window GP-UCB (SW-GP-UCB), which uses the training data in sliding window. Deng et al. [2022] have proposed another type of UCB-based algorithm called weighted GP-

UCB (WGP-UCB), which is based on the modified version of the GP whose past observed outputs are weighted in a time-dependent manner. The regret of these UCB-based algorithms has been shown to become $O(\gamma_T^{\frac{7}{8}} T^{\frac{3}{4}} V_T^{\frac{1}{4}})$, where γ_T is the maximum information gain, which represents the complexity of the problem (precise definition is described in Section 2). Recently, the OPKB algorithm proposed by Hong et al. [2023] has been shown to achieve $\widetilde{O}(\gamma_T^{\frac{1}{3}} T^{\frac{2}{3}} V_T^{\frac{1}{3}})$ regret, which nearly matches the lower bound for the linear kernel [Cheung et al., 2019]. As mentioned above, the computational cost of OPKB may become huge when the learner uses SE or Matérn kernels. UCB-based algorithms do not have such computational issues, while the regret guarantees are worse than that of OPKB.

Finally, a concurrent work [Cai and Scarlett, 2024] shows the same worst-case lower bound as ours. Furthermore, their work also considers the lower bound for RKHS norm variation settings, which is an interesting direction for our future research. On the other hand, their work does not show the near-optimal algorithm, which is resolved in this paper.

2 PRELIMINARIES

Problem Setting We consider the reward maximization problem under a non-stationary environment. Let $f_t : \mathcal{X} \to \mathbb{R}$ be an unknown reward function at step t. The input domain $\mathcal{X}$ is a compact subset of $\mathbb{R}^d$. The learner sequentially chooses the input $\boldsymbol{x}_t$ at step t; then, the environment reveals the noisy observation $y_t \coloneqq f_t(\boldsymbol{x}_t) + \epsilon_t$, where ϵ_t is a zero-mean noise random variable. In this paper, we assume the functions $(f_t)_{t \in \mathbb{N}_+}$ are determined by the environment *obliviously*; namely, all functions $(f_t)_{t \in \mathbb{N}_+}$ are fixed by the environment before the learner chooses the first input $\boldsymbol{x}_1$. In the aforementioned setup, the learner's goal is to minimize the following cumulative regret R_T:

$$R_T = \sum_{t \in [T]} f_t(\boldsymbol{x}_t^*) - f_t(\boldsymbol{x}_t), \tag{1}$$

where $\boldsymbol{x}_t^* \in \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}} f_t(\boldsymbol{x})$ and $[T] = \{1, \dots, T\}$.

For our theory, we make the following assumptions.

Assumption 1 (Assumption for noise)**.** *The noise sequence $(\epsilon_t)_{t \in \mathbb{N}_+}$ is mutually independent. Furthermore, the noise ϵ_t is ρ-sub-Gaussian random variable for any $t \in \mathbb{N}_+$; namely, $\mathbb{E}[\exp(\eta \epsilon_t)] \leq \exp(\eta^2 \rho^2 / 2)$ for all $\eta \in \mathbb{R}$.*

Assumption 2 (Assumption for reward functions)**.** *Each function f_t is an element of known RKHS with the bounded Hilbert norm. Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $\mathcal{H}_k$ be a known positive definite kernel and its corresponding RKHS. Then, we assume that $f_t \in \mathcal{H}_k$ and $\|f_t\|_{\mathcal{H}_k} \leq B < \infty$ hold for any $t \in \mathbb{N}_+$, where $\|\cdot\|_{\mathcal{H}_k}$ is the Hilbert norm on $\mathcal{H}_k$. Furthermore, suppose that $k(\boldsymbol{x}, \boldsymbol{x}) \leq 1$ holds for all $\boldsymbol{x} \in \mathcal{X}$.*

Assumption 3 (Assumption for non-stationarity)**.** *The total drift $\sum_{t=2}^T \|f_t - f_{t-1}\|_\infty$ of functions $(f_t)_{t \in [T]}$ is bounded as $\sum_{t=2}^T \|f_t - f_{t-1}\|_\infty \leq V_T$, where $\|f_t - f_{t-1}\|_\infty = \sup_{\boldsymbol{x} \in \mathcal{X}} |f_t(\boldsymbol{x}) - f_{t-1}(\boldsymbol{x})|$.*

Assumptions 1 and 2 are standard for ordinary KB literature [Srinivas et al., 2010, Chowdhury and Gopalan, 2017, Vakili et al., 2021a]. Specifically, as for Assumption 2, we focus on the following SE kernel k_{SE} and Matérn kernel $k_{\text{Matérn}}$:

$$k_{\mathrm{SE}}(\boldsymbol{x}, \widetilde{\boldsymbol{x}}) = \exp\left(-\frac{\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}\|_2^2}{2\ell^2}\right), \tag{2}$$

$$k_{\text{Matérn}}(\boldsymbol{x}, \widetilde{\boldsymbol{x}}) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}\|_2}{\ell}\right) J_\nu \left(\frac{\sqrt{2\nu}\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}\|_2}{\ell}\right), \tag{3}$$

where $\ell > 0$ and $\nu > 0$ denotes the length-scale and smoothness parameter, respectively. Furthermore, $\Gamma(\cdot)$ and $J_\nu(\cdot)$ are the Gamma and modified Bessel functions, respectively. Assumption 3 is also standard for non-stationary KB [Zhou and Shroff, 2021, Deng et al., 2022, Hong et al., 2023], and the increasing speed of the upper bound of the total variation V_T has an important role in characterizing both regret lower and upper bounds.

Gaussian Process Model Gaussian process (GP) model [Rasmussen and Williams, 2005] is a useful tool for estimating the underlying function while quantifying the uncertainty of its estimate; thus, GP is commonly used to construct an algorithm for the KB problem. Here, let us consider the Bayesian modeling of f whose prior is $\mathcal{GP}(0, k)$, where $\mathcal{GP}(0, k)$ denotes the mean-zero GP with covariance function k. Given the input data $\mathbf{X}_t \coloneqq (\boldsymbol{x}_1, \dots, \boldsymbol{x}_t)^\top$ and the corresponding outputs $\boldsymbol{y}_t \coloneqq (y_1, \dots, y_t)^\top$, the posterior is again the GP, whose posterior mean $\mu(\boldsymbol{x}; \mathbf{X}_t, \boldsymbol{y}_t)$ and variance $\sigma^2(\boldsymbol{x}; \mathbf{X}_t)$ of $f(\boldsymbol{x})$ are defined as follows:

$$\mu(\boldsymbol{x}; \mathbf{X}_t, \boldsymbol{y}_t) = \boldsymbol{k}(\boldsymbol{x}, \mathbf{X}_t)^\top (\mathbf{K}(\mathbf{X}_t, \mathbf{X}_t) + \lambda \boldsymbol{I}_t)^{-1} \boldsymbol{y}_t, \tag{4}$$

$$\begin{aligned} \sigma^2(\boldsymbol{x}; \mathbf{X}_t) = k(\boldsymbol{x}, \boldsymbol{x}) \\ - \boldsymbol{k}(\boldsymbol{x}, \mathbf{X}_t)^\top (\mathbf{K}(\mathbf{X}_t, \mathbf{X}_t) + \lambda \boldsymbol{I}_t)^{-1} \boldsymbol{k}(\boldsymbol{x}, \mathbf{X}_t), \end{aligned} \tag{5}$$

where $\lambda > 0$ is a noise variance parameter, and $\boldsymbol{k}(\boldsymbol{x}, \mathbf{X}_t) \coloneqq [k(\boldsymbol{x}, \boldsymbol{x}_i)]_{i \in [t]} \in \mathbb{R}^t$ and $\mathbf{K}(\mathbf{X}_t, \mathbf{X}_t) \coloneqq [k(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j \in [t]} \in \mathbb{R}^{t \times t}$ are the kernel vector and matrix, respectively. Furthermore, $\boldsymbol{I}_t \in \mathbb{R}^{t \times t}$ denotes the identity matrix.

Finally, we define the kernel-dependent quantity γ_T as $\gamma_T = \frac{1}{2} \sup_{\mathbf{X}_t} \ln \det (\boldsymbol{I}_t + \lambda^{-1} \mathbf{K}(\mathbf{X}_t, \mathbf{X}_t))$. If we suppose that f follows GP, γ_T is equal to the maximum amount of the information gain of f up to T observations [Srinivas et al., 2010]; thus, the quantity γ_T is called *maximum information gain* (MIG). MIG characterizes the complexity of the KB problem, and its upper bound increases sublinearly in several commonly used kernels. For example,

$\gamma_T = O(\ln^{d+1} T)$ in SE kernel, and $\gamma_T = O(T^{\frac{d}{2\nu+d}} \ln^{\frac{2\nu}{2\nu+d}} T)$ in Matérn kernel with $\nu > 1/2$ [Vakili et al., 2021b].

Phased Elimination PE [Li and Scarlett, 2022, Bogunovic et al., 2022], referred to as batched pure exploration in Li and Scarlett [2022], is a near-optimal KB algorithm in the stationary environment for the SE and Matérn kernels. PE divides the entire time horizons T into some *batches* whose sizes are exponentially increasing. In each batch, PE performs the (non-adaptive) maximum variance reduction algorithm [Vakili et al., 2021a] in *potential maximizers*, which is a candidate input whose UCB exceeds the maximum of lower CB. By the above procedures, PE achieves the regret upper bound $\widetilde{O}(\sqrt{T\gamma_T})$ while the regret upper bound of classical GP-UCB and TS is $\widetilde{O}(\gamma_T\sqrt{T})$. We combine this PE algorithm, the restart and reset strategy, and the random permutation of the query candidates for the non-stationary KB problem.

3 LOWER BOUND FOR NON-STATIONARY KERNELIZED BANDITS

Our first main result is the following Theorem 1, which shows the algorithm-independent lower bound of the non-stationary KB problem with SE or Matérn kernel.

Theorem 1 (Lower bound)**.** *Fix any $T \in \mathbb{N}_+$, $V_T > 0$, $B > 0$, and $\rho > 0$. Furthermore, assume $\mathcal{X} = [0,1]^d$, and $(\epsilon_t)_{t\in\mathbb{N}_+}$ is the noise sequence of independent Gaussian random variables $\epsilon_t \sim \mathcal{N}(0, \rho^2)$ for all $t \in \mathbb{N}_+$.*

- *Suppose $k = k_{\mathrm{SE}}$, $\widetilde{C}_{\mathrm{SE}} T^{-\frac{1}{2}} \ln^{\frac{d}{4}} T \le V_T < \widetilde{C}_{\mathrm{SE}} T \ln^{\frac{d}{4}} T$, and $V_T \le \overline{C} T^{\overline{c}}$; then, for any algorithm, there exists reward functions $(f_t)_{t\in\mathbb{N}_+}$ such that Assumptions 2, 3 and $\mathbb{E}[R_T] \ge C_{\mathrm{SE}} T^{\frac{2}{3}} V_T^{\frac{1}{3}} \ln^{\frac{d}{6}} T$ hold for sufficiently large T. Here, $\overline{C} > 0$ and $\overline{c} \in (0,1)$ are any absolute constants.*
- *Suppose $k = k_{\text{Matérn}}$ and $\widetilde{C}_{\mathrm{Mat}} T^{-\frac{\nu}{2\nu+d}} \le V_T \le 2^{\frac{3\nu+d}{2\nu+d}} \widetilde{C}_{\mathrm{Mat}} T$; then, for any algorithm, there exists reward functions $(f_t)_{t\in\mathbb{N}_+}$ such that Assumption 2, 3 and $\mathbb{E}[R_T] \ge C_{\mathrm{Mat}} T^{\frac{2\nu+d}{3\nu+d}} V_T^{\frac{\nu}{3\nu+d}}$ hold.*

Here, $\widetilde{C}_{\mathrm{SE}}, C_{\mathrm{SE}}, \widetilde{C}_{\mathrm{Mat}}, C_{\mathrm{Mat}} > 0$ are constants that only depend on ℓ, ν, B, ρ, $\overline{C}$, $\overline{c}$, and d.

The full proof of Theorem 1 is shown in Appendix A.1.

Proof Sketch The proof of Theorem 1 is derived by combining the idea of the lower bound of non-stationary linear bandits [Cheung et al., 2019] with the existing lower bound of stationary KB [Scarlett et al., 2017]. By following the idea of Cheung et al. [2019], we first separate the time step set $[T]$ into the $\lceil T/H \rceil$ intervals whose lengths are $H \in [T]$ (except for the last interval). Next, we consider to assign the function f that achieves the existing lower bound of stationary KB [Scarlett et al., 2017] for each interval. From this construction of $(f_t)_{t\in\mathbb{N}_+}$, $\Omega(\lceil T/H \rceil \mathrm{LB}(H))$ regret incurs even if the algorithm knows the interval length H. Here, $\mathrm{LB}(H)$ denotes the lower bound of stationary KB for total step size H. For example, $\mathrm{LB}(H) = \widetilde{\Omega}(\sqrt{H})$ for $k = k_{\mathrm{SE}}$ and $\mathrm{LB}(H) = \widetilde{\Omega}(H^{\frac{\nu+d}{2\nu+d}})$ for $k = k_{\text{Matérn}}$. The remaining interest is to choose the length H as small as possible to maximize $\lceil T/H \rceil \mathrm{LB}(H)$ under $\sum_{t=2}^{T} \|f_t - f_{t-1}\|_\infty \le V_T$. To choose such H, we leverage the precise characterization of the functions that achieve the lower bound provided by Scarlett et al. [2017]. Fortunately, with a slight modification of the proof of Scarlett et al. [2017], we can obtain the following lemma that connects the total variation of the functions with lower bounds.

Lemma 2. *Suppose the same conditions as those of Theorem 1. Furthermore, for any $T \in \mathbb{N}_+$, set $\varepsilon_{\mathrm{SE}}(T) = C_{\mathrm{SE}}^{(1)} \sqrt{(\ln T)^{\frac{d}{2}}/T}$ and $\varepsilon_{\text{Matérn}}(T) = C_{\mathrm{Mat}}^{(1)} T^{-\frac{\nu}{2\nu+d}}$, where $C_{\mathrm{SE}}^{(1)}$, $C_{\mathrm{Mat}}^{(1)} > 0$ are constants that only depends on ℓ, ν, ρ, B, and d. Then, there exists the function set $\mathcal{F} \subset \mathcal{H}_k$ such that $\forall f \in \mathcal{F}, \|f\|_{\mathcal{H}_k} \le B$ and the following hold:*

- *For any algorithm, there exist $f \in \mathcal{F}$ such that $\mathbb{E}[R_T] \ge T\varepsilon(T)$ under $f_t = f$ for all $t \in \mathbb{N}_+$. Here, we set $\varepsilon(T) = \varepsilon_{\mathrm{SE}}(T)$ and $\varepsilon(T) = \varepsilon_{\text{Matérn}}(T)$ for $k = k_{\mathrm{SE}}$ and $k = k_{\text{Matérn}}$, respectively.*
- *For any $f, \widetilde{f} \in \mathcal{F}$, $\|f - \widetilde{f}\|_\infty \le 4\varepsilon(T)$.*

From Lemma 2, the total drift $\sum_{t=2}^{T} \|f_t - f_{t-1}\|_\infty$ is bounded from above by $4\lceil T/H \rceil \varepsilon(H)$ under the aforementioned construction of $(f_t)_{t\in\mathbb{N}_+}$. Finally, by selecting the smallest H such that $4\lceil T/H \rceil \varepsilon(H) \le V_T$ holds, the lower bounds $\Omega(\lceil T/H \rceil \mathrm{LB}(H))$ matches the results of Theorem 1.

Comparison with the Lower Bounds for Stationary KB For standard stationary KB problems, Scarlett et al. [2017] shows $\widetilde{\Omega}(\sqrt{T})$ and $\Omega(T^{\frac{\nu+d}{2\nu+d}})$ lower bounds in SE and Matérn family of kernels, respectively. By comparing our lower bounds, we can confirm that the learner suffers from at least $T^{\frac{1}{6}}$ and $T^{\frac{\nu^2}{(3\nu+d)(2\nu+d)}}$ additional polynomial factors at the cost of non-stationarity for SE and Matérn family of kernels, respectively. Note that the degeneration of the worst-case lower bounds is also observed in the existing finite-armed and linear bandit problems [Besbes et al., 2014, Cheung et al., 2019]. Our lower bounds show that such degeneration also occurs in KB problems.

4 MODIFIED NEAR-OPTIMAL OPKB ALGORITHM FOR MATÉRN KERNEL

Hong et al. [2023] have proposed adapting OPKB (ADA-OPKB) for the non-stationary environment without prior information of V_T. Since V_t is unknown, ADA-OPKB leverages the adaptive scheduling of restart-reset procedures. Then, the ADA-OPKB algorithm [Hong et al., 2023] achieves $\widetilde{O}(\gamma_T^{1/3} T^{2/3} V_T^{1/3})$ regret. Therefore, the regret upper bounds for the SE and Matérn kernels are $\widetilde{O}(T^{\frac{2}{3}} V_T^{\frac{1}{3}})$ and $\widetilde{O}(T^{\frac{4\nu+3d}{6\nu+3d}} V_T^{\frac{1}{3}})$, respectively. Hence, for the SE kernel, the ADA-OPKB is near-optimal even though V_T is unknown. However, for the Matérn kernels, the regret upper bound is worse compared with our $\Omega(T^{\frac{2\nu+d}{3\nu+d}} V_T^{\frac{\nu}{3\nu+d}})$ lower bound. Therefore, no near-optimal regret upper bound for the Matérn kernel is known.

In contrast to the unknown V_T setting where Hong et al. [2023] focuses on, this paper focuses on the known V_T setting. In the known V_T setting, by combining OPKB procedures with a restart-reset strategy whose reset interval is carefully chosen by depending on d, ν, and V_T, we can achieve near-optimal $\widetilde{O}(T^{\frac{2\nu+d}{3\nu+d}} V_T^{\frac{\nu}{3\nu+d}})$ regret for the Matérn kernel:

Theorem 3 (The modified version of OPKB algorithm with the restart-reset strategy.)**.** *Assume that the underlying kernel is Matérn kernel with smoothness parameter $\nu > 1/2$. Furthermore, suppose that Assumptions 1–3 and $V_T \geq T^{-\frac{\nu}{2\nu+d}} \ln^{\frac{\nu+d}{2\nu+d}} T$ hold. Then, if we set the restarting interval H as $H = \lceil T^{\frac{2\nu+d}{3\nu+d}} V_T^{-\frac{2\nu+d}{3\nu+d}} \ln^{\frac{4\nu+d}{6\nu+2d}} T \rceil$, the OPKB algorithm (Algorithm 2 in Hong et al. [2023]) with the restart-reset strategy achieve $R_T = \widetilde{O}(V_T^{\frac{\nu}{3\nu+d}} T^{\frac{2\nu+d}{3\nu+d}})$ with probability at least $1-\delta$.*

See Appendix C for details. This result also justifies the tightness of our lower bounds. Namely, our $\Omega(V_T^{\frac{\nu}{3\nu+d}} T^{\frac{2\nu+d}{3\nu+d}})$ lower bound for the Matérn kernel has no room for improvement in polynomial factors.

On the other hand, as discussed in Section 1, the OPKB algorithm suffers from the huge computational complexity $O(M|\mathcal{X}|^3)$. Therefore, in the next section, we propose yet another near-optimal algorithm, which enjoys both ease-computation and the near-optimal regret upper bound.

Algorithm 1 Restarting Phased Elimination with Random Permutation (R-PERP)

Require: Total step size T, confidence width parameter $\beta_T^{1/2} > 0$, reset interval $H \in [T]$, finite input set $\mathcal{X}$.

for $i = 1, \ldots, \lceil \frac{T}{H} \rceil$ **do**
  Compute the i-th interval size: $T^{(i)} \leftarrow \min\{H, T - (i-1)H\}$.
  Initialize the potential maximizer $\mathcal{X}_1^{(i)} \leftarrow \mathcal{X}$ and $N_0^{(i)} \leftarrow 1$.
  for $j = 1, 2, \ldots$ **do**
    Compute the j-th batch size $N_j^{(i)} \leftarrow \min\left\{\left\lceil \sqrt{T^{(i)} N_{j-1}^{(i)}} \right\rceil, T^{(i)} - \sum_{\widetilde{j}=1}^{j-1} N_{\widetilde{j}}^{(i)}\right\}$.
    Initialize the query candidate set $\mathcal{S}_j^{(i)} \leftarrow \emptyset$.
    for $m = 1, \ldots, N_j^{(i)}$ **do**
      $\widetilde{\boldsymbol{x}}_{j,m}^{(i)} \leftarrow \arg\max_{\boldsymbol{x} \in \mathcal{X}_j^{(i)}} \sigma^2(\boldsymbol{x}; \mathcal{S}_j^{(i)})$.
      $\mathcal{S}_j^{(i)} \leftarrow \mathcal{S}_j^{(i)} \cup \left\{\widetilde{\boldsymbol{x}}_{j,m}^{(i)}\right\}$.
    end for
    Obtain $\left(\boldsymbol{x}_{j,1}^{(i)}, \ldots, \boldsymbol{x}_{j,N_j^{(i)}}^{(i)}\right)$ as an uniform permutation of $\mathcal{S}_j^{(i)}$.
    for $m = 1, \ldots, N_j^{(i)}$ **do**
      Observe $y_{j,m}^{(i)} = f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right) + \epsilon_{j,m}^{(i)}$.
    end for
    if $\sum_{\widetilde{j}=1}^{j} N_{\widetilde{j}}^{(i)} = T^{(i)}$ **then**
      Move to next $(i+1)$-th interval.
    end if
    $\boldsymbol{y}_j^{(i)} \leftarrow \left[y_{j,m}^{(i)}\right]_{m \leq N_j^{(i)}}$, $\boldsymbol{X}_j^{(i)} \leftarrow \left[\boldsymbol{x}_{j,m}^{(i)}\right]_{m \leq N_j^{(i)}}$
    Calculate $\mathrm{lcb}_j^{(i)}(\cdot)$ and $\mathrm{ucb}_j^{(i)}(\cdot)$ as

$$\mathrm{lcb}_j^{(i)}(\boldsymbol{x}) = \mu(\boldsymbol{x}; \boldsymbol{X}_j^{(i)}, \boldsymbol{y}_j^{(i)}) - \beta_T^{1/2} \sigma(\boldsymbol{x}; \boldsymbol{X}_j^{(i)}), \tag{6}$$

$$\mathrm{ucb}_j^{(i)}(\boldsymbol{x}) = \mu(\boldsymbol{x}; \boldsymbol{X}_j^{(i)}, \boldsymbol{y}_j^{(i)}) + \beta_T^{1/2} \sigma(\boldsymbol{x}; \boldsymbol{X}_j^{(i)}). \tag{7}$$

    $\mathcal{X}_{j+1}^{(i)} \leftarrow \left\{\boldsymbol{x} \in \mathcal{X}_j^{(i)} \,\middle|\, \mathrm{ucb}_j^{(i)}(\boldsymbol{x}) \geq \max_{\widetilde{\boldsymbol{x}} \in \mathcal{X}_j^{(i)}} \mathrm{lcb}_j^{(i)}(\widetilde{\boldsymbol{x}})\right\}$
  end for
end for

5 PHASED ELIMINATION WITH RANDOM PERMUTATION

In this section, we show our R-PERP algorithm. For simplicity, we assume that the input set $\mathcal{X}$ is finite without loss of generality. That is, by relying on the discretization arguments [Li and Scarlett, 2022] of the input space with the Lipschitz assumption of reward functions[4], the same guarantees described in this section are obtained under compact continuous input set $\mathcal{X}$.

[4]Specifically, Assumption 2 implies Lipschitz assumption when the underlying RKHS is induced by SE or Matérn kernel with $\nu > 1$. See, e.g., Lemma 2 in Li and Scarlett [2022].

Algorithm Construction The pseudo-code of the R-PERP algorithm is shown in Algorithm 1. The algorithm construction relies on the restarting strategy, which reset some base KB algorithm with the prespecified interval H. For the construction of the base algorithm, we leverage the PE-based algorithm [Li and Scarlett, 2022], which successively keeps and eliminates the potential maximizer based on the CBs of the reward functions.

Hereafter, we denote $\boldsymbol{x}_{j,m}^{(i)}$ as the query point corresponding to m-th observation in j-th batch of PE. Furthermore, the superscript index i indicate that $\boldsymbol{x}_{j,m}^{(i)}$ is in i-th interval of restarting strategy. Namely, $\boldsymbol{x}_{j,m}^{(i)}$ denotes the query point at time step $(i-1)H + \sum_{\widetilde{j}=1}^{j-1} N_{\widetilde{j}}^{(i)} + m$, where $N_{\widetilde{j}}^{(i)}$ denotes the $\widetilde{j}$-th batch size of interval i (Line 5 in Algorithm 1). We define $y_{j,m}^{(i)}$, $\epsilon_{j,m}^{(i)}$, and $f_{j,m}^{(i)}$ as the similar way to $\boldsymbol{x}_{j,m}^{(i)}$.

5.1 Confidence Bounds with Random Permutation

The main challenge in adapting the PE-based algorithm is the construction of the CBs under non-stationary rewards. As pointed out in Remark 1 of Hong et al. [2023], the existing regression-based CBs (e.g., Zhou and Shroff [2021]) have additional $\sqrt{H\gamma_H}$ factor compared with the inverse propensity score-based estimate proposed in Hong et al. [2023]. This $\sqrt{H\gamma_H}$ degeneration leads to sub-optimal regret; therefore, constructing tighter regression-based CB under non-stationarity is essential. Our key idea is to construct CBs for the average function $\overline{f}_j^{(i)}(\cdot) \coloneqq \sum_{m=1}^{N_j^{(i)}} f_{j,m}^{(i)}(\cdot)/N_j^{(i)}$ by leveraging the random permutation of the query candidate set, instead of directly constructing CBs for some future reward functions.

Here, let $\mathcal{S}_j^{(i)}$ be the query candidate set at batch j of interval i. As with the standard stationary PE, R-PERP chooses the query candidate based on the posterior variance (Line 7–9). After that, R-PERP chooses the order of observation within the query candidate set by uniformly permutating the elements of $\mathcal{S}_j^{(i)}$ (Line 11). Intuitively, this permutation procedure makes it robust against the environment's worst-case reward choice. Figure 1 shows the illustrative image.

Based on such construction of R-PERP, the following theorem shows the tighter CB for the average function without $\sqrt{H\gamma_H}$ additional factor.

Theorem 4 (Confidence bounds for average functions.)**.** *Fix any $T \in \mathbb{N}_+$, $H \in [T] \setminus \{1\}$, and $\delta \in (0,1)$. Suppose that Assumptions 1 and 2 hold. Furthermore, set the confidence width parameter $\beta_T^{1/2}$ as*

$$\beta_T^{1/2} = B\left(\frac{C}{\sqrt{\lambda}}\sqrt{\ln\frac{4|\mathcal{X}|Q_{T,H}}{\delta}} + 1\right) + \frac{\rho}{\sqrt{\lambda}}\sqrt{2\ln\frac{4|\mathcal{X}|Q_{T,H}}{\delta}}, \tag{8}$$

where $Q_{T,H} = \lceil T/H \rceil (1 + \log_2 \log_2 H)$, and $C > 0$ is an absolute constant. Then, when running Algorithm 1, the following event holds with probability at least $1-\delta$:

$$\forall i \le \left\lceil \frac{T}{H} \right\rceil, \forall j \le Q^{(i)} - 1,\ \forall \boldsymbol{x} \in \mathcal{X}, \\ \mathrm{lcb}_j^{(i)}(\boldsymbol{x}) \le \overline{f}_j^{(i)}(\boldsymbol{x}) \le \mathrm{ucb}_j^{(i)}(\boldsymbol{x}), \tag{9}$$

where $\mathrm{lcb}_j^{(i)}(\cdot)$ and $\mathrm{ucb}_j^{(i)}(\cdot)$ are defined in Eq. (6) and Eq. (7), respectively, and $\overline{f}_j^{(i)}(\cdot) \coloneqq \sum_{m=1}^{N_j^{(i)}} f_{j,m}^{(i)}(\cdot)/N_j^{(i)}$ is the average function in j-th batch at i-th interval. Furthermore, $Q^{(i)}$ represents the total number of batches over i-th interval.

The full proof of Theorem 4 is given in Appendix B.

Proof Sketch of Theorem 4 We decompose the error $|\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{y}_j^{(i)})|$ as

$$|\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{y}_j^{(i)})| \le \\ |\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)})| + |\mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{\epsilon}_j^{(i)})|, \tag{10}$$

where $\boldsymbol{f}_j^{(i)} = (f_{j,1}^{(i)}(\boldsymbol{x}_{j,1}^{(i)}), \ldots, f_{j,N_j^{(i)}}^{(i)}(\boldsymbol{x}_{j,N_j^{(i)}}^{(i)}))^\top$ and $\boldsymbol{\epsilon}_j^{(i)} = (\epsilon_{j,1}^{(i)}, \ldots, \epsilon_{j,N_j^{(i)}}^{(i)})$. The second term in r.h.s. represents the effect of noise and is bounded from above as with the proof of Theorem 1 in Vakili et al. [2021a]. The remaining interest is the first term. Here, let us denote $\widetilde{\boldsymbol{x}}_{j,m}^{(i)}$ as the m-th element of query candidate set $\mathcal{S}_j^{(i)}$ (see Line 8 of Algorithm 1). Furthermore, we denote permutation index $\psi(m)$ as the natural number such that $\widetilde{\boldsymbol{x}}_{j,m}^{(i)} = \boldsymbol{x}_{j,\psi(m)}^{(i)}$ holds. Then, we obtain the equivalent expression of $\mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)})$ as $\mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}) = \mu(\boldsymbol{x}; \mathcal{S}_j^{(i)}, \widetilde{\boldsymbol{f}}_j^{(i)})$, where $\widetilde{\boldsymbol{f}}_j^{(i)} = \left[f_{j,\psi(m)}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j,m}^{(i)}\right)\right]_{m \in [N_j^{(i)}]}$. From this representation and the linearity of the expectation, we observe that $\mathbb{E}[\mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}) \mid H_{j-1}^{(i)}] = \mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \overline{\boldsymbol{f}}_j^{(i)})$ holds, where $\overline{\boldsymbol{f}}_j^{(i)} = [\overline{f}_j^{(i)}(\widetilde{\boldsymbol{x}}_{j,m}^{(i)})]_{m \in [N_j^{(i)}]}$, and $H_{j-1}^{(i)}$ is the history up to $(j-1)$-th batch at interval i. Finally, by carefully combining the above fact with the concentration inequality of random permutation (Lemma 7 in the appendix) and Proposition 1 in Vakili et al. [2021a], we obtain the high-probability upper bound of the first term in Eq. (10).

5.2 Regret Analysis

By combining Theorem 4 with the analysis of standard PE and the gap between average function $\overline{f}_j^{(i)}$ and the true non-stationary reward functions $f_{j,m}^{(i)}$, we obtain the following regret upper bound of R-PERP.

Theorem 5 (Regret upper bound of R-PERP)**.** *Fix any $\delta \in (0,1)$. Suppose that Assumptions 1–3 hold. Furthermore,*

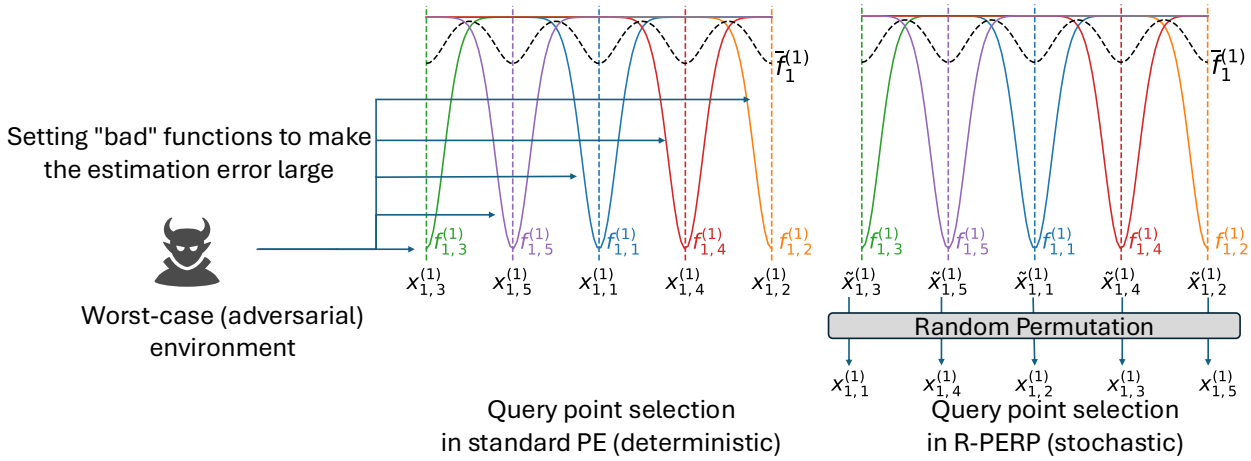

Figure 1: Illustrative image of the R-PERP algorithm in the first batch at the first interval with $N_1^{(1)} = 5$. The standard PE algorithm (left) chooses the query points deterministically within each batch. Thus, intuitively, the environment can arbitrarily choose the reward function $f_{1,j}^1$ such that the learner's estimation error becomes large. The R-PERP algorithm (right) alleviates the effect of such worst-case selection of the reward functions by randomly permutating the query candidates of the standard PE.

set confidence width $\beta_T^{1/2}$ as in Eq. (8)*. Then, the following statements hold with probability at least $1-\delta$:*

- *If $k = k_{\mathrm{SE}}$ and $V_T \geq T^{-\frac{1}{2}} \ln^{\frac{d+2}{3}} T$, the regret of R-PERP satisfies $R_T = \widetilde{O}(V_T^{\frac{1}{3}} T^{\frac{2}{3}})$ by setting $H = \lceil T^{\frac{2}{3}} V_T^{-\frac{2}{3}} \ln^{\frac{d+2}{3}} T \rceil$.*
- *If $k = k_{\text{Matérn}}$ with $\nu > 1/2$ and $V_T \geq T^{-\frac{\nu}{2\nu+d}} \ln^{\frac{\nu+d}{2\nu+d}} T$, the regret of R-PERP satisfies $R_T = \widetilde{O}(V_T^{\frac{\nu}{3\nu+d}} T^{\frac{2\nu+d}{3\nu+d}})$ by setting $H = \lceil T^{\frac{2\nu+d}{3\nu+d}} V_T^{-\frac{2\nu+d}{3\nu+d}} \ln^{\frac{4\nu+d}{6\nu+2d}} T \rceil$.*

Comparing our lower bounds (Theorem 1) with Theorem 5, we can confirm that R-PERP achieves nearly optimal regret for SE and Matérn RKHS.

Remark 1. High-probability regret guarantees provided in Theorem 5 imply the expected regret guarantees of the same order. This can be easily confirmed by setting $\delta = 1/T$ and noting that $f_t(\boldsymbol{x}^*) - f_t(\boldsymbol{x}_t) \leq 2B$ holds from Assumption 2.

Proof Sketch of Theorem 5 To leverage the regret analysis technique of PE, we decompose the instantaneous regret $r_{j,m}^{(i)} \coloneqq f_{j,m}^{(i)}(\boldsymbol{x}_{j,m}^{(i)*}) - f_{j,m}^{(i)}(\boldsymbol{x}_{j,m}^{(i)})$ as follows:

$$\begin{aligned} r_{j,m}^{(i)} =& f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)*}\right) - \overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right) \\ &+ \overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right) - \overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right) \\ &+ \overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right) - f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right), \end{aligned} \tag{11}$$

where $\widetilde{\boldsymbol{x}}_{j-1}^{(i)*} \in \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}_{j-1}^{(i)}} \overline{f}_{j-1}^{(i)}(\boldsymbol{x})$ is the maximum over the potential maximizer $\mathcal{X}_{j-1}^{(i)}$ of $(j-1)$-th batch of R-PERP. The cumulative regret that arises from the second term is bounded from above by combining our CB for the average function (Theorem 4) with the analysis of the standard PE. In each interval, the order of this term becomes $\widetilde{O}(\sqrt{H\gamma_H})$. As for the cumulative regret that arises from the first and third terms, we can obtain $\widetilde{O}(V_T^{(i)} H)$ upper bound in any i-th interval. Here, $V_T^{(i)}$ represents the total variation of the sequence of the underlying reward functions on the i-th interval. Therefore, aggregating the cumulative regret upper bounds of each interval, we have $R_T \leq \sum_{i=1}^{\lceil T/H \rceil} \widetilde{O}(V_T^{(i)} H + \sqrt{H\gamma_H}) = \widetilde{O}(V_T H + T\sqrt{\gamma_H / H})$. By setting H to balance $V_T H$ and $T\sqrt{\gamma_H / H}$, we obtain the desired results.

6 NUMERICAL EXPERIMENTS

In this section, we confirm the empirical performance of our R-PERP algorithm. We would like to emphasize that we will not claim the state-of-the-art empirical performance of our R-PERP algorithm. Specifically, the existing works report that the empirical performance of PE-based algorithms tends to be inferior to UCB-like algorithms due to the large constant factor of the regret upper bound [Li and Scarlett, 2022, Bogunovic et al., 2022]. In our numerical experiments, we find that R-PERP inherits such deterioration of the empirical performance compared with existing UCB-like algorithms such as R-GP-UCB. Filling the gap between practical and empirical performance is an important direction for our future work; however, we believe that the worse practical results of R-PERP do not diminish our theoretical contributions.

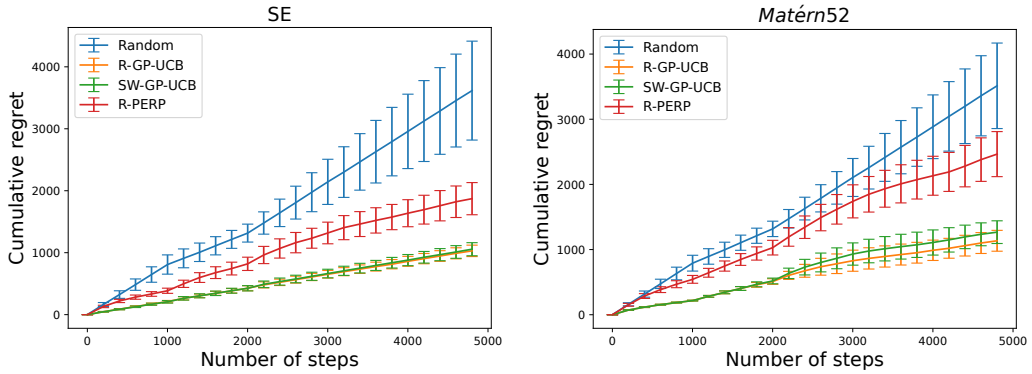

Figure 2: Numerical experiment results with 5 different seeds. The plots show the average cumulative regret, and the error bars represent one standard error. The left and right plots show the results with SE kernel and Matérn kernel with $\nu = 5/2$, respectively.

Setting We conduct the experiments with synthetic functions used in Deng et al. [2022], which considers abrupt changes in the reward functions. To construct such synthetic functions, we first generate the base objective function $f_{\text{base}} \in \mathcal{H}_k$ as $f_{\text{base}}(\boldsymbol{x}) = \sum_{i=1}^{U} \alpha_i k(\cdot, \overline{\boldsymbol{x}}_i)$, where $\alpha_i \in [-1, 1]$ and $\overline{\boldsymbol{x}}_i \in [0, 1]^2$ are generated uniformly at random. Then, we define the underlying reward functions $(f_t)_{t\in[T]}$ as $f_t = f_{\text{base}}^{(1)}$ for $t \in [T/5]$; $f_t = f_{\text{base}}^{(2)}$ for $t \in [2T/5] \setminus [T/5]$; otherwise, $f_t = f_{\text{base}}^{(3)}$. Here, $(f_{\text{base}}^{(i)})_{i\in[3]}$ are the base functions generated by the aforementioned procedures with different seeds. Our experiments were conducted with SE and Matérn kernels with the lengthscale parameter $\ell = 0.5$ and the smoothness parameter $\nu = 5/2$. Furthermore, we set the input domain $\mathcal{X}$ as the 30×30 grid points obtained by evenly splitting $[0, 1]^2$. Finally, we set $U = 10$ and $T = 5000$.

Algorithms We consider the following four algorithms:

- **Random**: Baseline algorithm whose query points are chosen uniformly at random on $\mathcal{X}$.
- **R-GP-UCB**: The GP-UCB algorithm with restart and resetting strategy [Zhou and Shroff, 2021]. We set the restart interval H based on the theoretically suggested value $H = \widetilde{\gamma}_T^{1/4}(T/V_T)^{1/2}$, where $\widetilde{\gamma}_T = \ln^{d+1} T$ for $k = k_{\text{SE}}$ and $\widetilde{\gamma}_T = T^{\frac{d}{2\nu+d}} \ln^{\frac{2\nu}{2\nu+d}} T$ for $k = k_{\text{Matérn}}$.
- **SW-GP-UCB**: The sliding-window GP-UCB algorithm proposed in [Zhou and Shroff, 2021]. SW-GP-UCB constructs the GP model in each step using only past W input-output pairs, where W is the pre-specified window width. Then, SW-GP-UCB chooses the next query point based on the UCB score calculated by the GP model described above. As with the R-GP-UCB algorithm, we set W based on theoretically suggested value $W = \widetilde{\gamma}_T^{1/4}(T/V_T)^{1/2}$.
- **R-PERP**: Our proposed algorithm. We set the reset interval H and confidence width parameter β_T as described in Theorem 5.

In all methods, we use the theoretically suggested confidence width $\beta_t^{1/2}$ with confidence level $\delta = 0.1$. Here, the theoretically suggested confidence width of R-GP-UCB and SW-GP-UCB require the MIG γ_T. In our experiments, we compute the upper bound of the MIG by relying on the greedy sampling rule proposed in Section. 5.1 of Srinivas et al. [2010]; then, we use this upper bound as the proxy of the MIG. Here, we would like to note that running the OPKB algorithm in our experimental setup of $\mathcal{X}$ requires at least approximately 10^8–10^9 order of computations; therefore, we exclude OPKB from our numerical experiments.

Results Figure 2 shows the results. We confirm that our R-PERP algorithm achieves superior performance to a simple random sampling baseline. Specifically, the average regret of R-PERP increases sub-linearly between 2000 and 5000 steps, even if two abrupt changes of reward functions occur at 1000 and 2000 steps. These results indicate that PE-based algorithms such as R-PERP work in non-stationary environments. On the other hand, R-PERP has a worse practical performance than that of UCB-based algorithms. These practical performance gaps between PE-based algorithms and UCB-based algorithms in non-stationary environments are consistent with the results in the existing stationary environment settings (e.g., [Bogunovic et al., 2022, Li and Scarlett, 2022]).

7 CONCLUSION

In this paper, we study the near-optimal algorithms of the non-stationary kernelized bandit problem. We show the first lower bounds for the problem with the SE and Matérn kernels. Furthermore, we propose a novel nearly-optimal

PE-based algorithm for a non-stationary environment. Our proposed algorithm is based on simple random permutation procedures on the query candidate sets of PE, which enables us to derive tighter confidence bounds tailored to non-stationary settings.

The important future direction of our research is to extend our method to the unknown total drift setting. Our works rely on the prior knowledge of the upper bound of the total variation V_T of the underlying reward functions. Some existing works tackle the unknown V_T setting by adaptively scheduling the restart-reset interval of the algorithm (e.g., Wei and Luo [2021], Hong et al. [2023]). Another interesting future direction is developing a near-optimal non-stationary kernelized bandit algorithm for the RKHS norm variation setting, where the room for improvement is implied by the concurrent work [Cai and Scarlett, 2024].

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number (JP23K19967 and JP24K20847) and RIKEN Center for Advanced Intelligence Project.

References

Rados law Adamczak, Djalil Chafaï, and Pawe l Wolff. Circular law for random matrices with exchangeable entries. *Random Structures & Algorithms*, 2016.

Felix Berkenkamp, Andreas Krause, and Angela P Schoellig. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. *Machine Learning*, 112(10):3713–3747, 2023.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. *Proc. Neural Information Processing Systems (NeurIPS)*, 2014.

Ilija Bogunovic, Jonathan Scarlett, and Volkan Cevher. Time-varying Gaussian process bandit optimization. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2016.

Ilija Bogunovic, Zihan Li, Andreas Krause, and Jonathan Scarlett. A robust phased elimination algorithm for corruption-tolerant Gaussian process bandits. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2022.

Xu Cai and Jonathan Scarlett. On lower bounds for standard and robust Gaussian process bandit optimization. In *Proc. International Conference on Machine Learning (ICML)*, 2021.

Xu Cai and Jonathan Scarlett. Lower bounds for time-varying kernelized bandits. *arXiv preprint arXiv:2410.16692*, 2024.

Romain Camilleri, Kevin Jamieson, and Julian Katz-Samuels. High-dimensional experimental design and kernel bandits. In *Proc. International Conference on Machine Learning (ICML)*, 2021.

Wang Chi Cheung, David Simchi-Levi, and Ruihao Zhu. Learning to optimize under non-stationarity. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proc. International Conference on Machine Learning (ICML)*, 2017.

Yuntian Deng, Xingyu Zhou, Baekjin Kim, Ambuj Tewari, Abhishek Gupta, and Ness Shroff. Weighted Gaussian process bandits for non-stationary environments. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

Negar Hariri, Bamshad Mobasher, and Robin Burke. Adapting to user preference changes in interactive recommendation. In *Proceedings of the 24th International Conference on Artificial Intelligence*, page 4268–4274. AAAI Press, 2015.

John Heaton and Deborah Lucas. Stock prices and fundamentals. *NBER macroeconomics annual*, 14:213–242, 1999.

Kihyuk Hong, Yuhang Li, and Ambuj Tewari. An optimization-based algorithm for non-stationary kernel bandits without prior knowledge. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.

Ksenia Korovina, Sailun Xu, Kirthevasan Kandasamy, Willie Neiswanger, Barnabás Poczos, Jeff Schneider, and Eric Xing. ChemBO: Bayesian optimization of small organic molecules with synthesizable recommendations. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 3393–3403, 2020.

Zihan Li and Jonathan Scarlett. Gaussian process bandit optimization with few batches. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

Sudeep Salgia, Sattar Vakili, and Qing Zhao. A domain-shrinking based Bayesian optimization algorithm with order-optimal regret performance. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2021.

Sudeep Salgia, Sattar Vakili, and Qing Zhao. Random exploration in Bayesian optimization: Order-optimal regret and computational efficiency. In *Proc. International Conference on Machine Learning (ICML)*, 2024.

Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy Gaussian process bandit optimization. In *Proc. Conference on Learning Theory (COLT)*, 2017.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proc. International Conference on Machine Learning (ICML)*, 2010.

Tsuyoshi Ueno, Trevor David Rhone, Zhufeng Hou, Teruyasu Mizoguchi, and Koji Tsuda. COMBO: An efficient Bayesian optimization library for materials science. *Materials discovery*, 4:18–21, 2016.

Sattar Vakili, Nacime Bouziani, Sepehr Jalali, Alberto Bernacchia, and Da shan Shiu. Optimal order simple regret for Gaussian process bandits. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2021a.

Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in Gaussian process bandits. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021b.

Chen-Yu Wei and Haipeng Luo. Non-stationary reinforcement learning without prior knowledge: An optimal black-box approach. In *Proc. Conference on Learning Theory (COLT)*, 2021.

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

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes. We describe the mathematical setting, assumptions, and algorithm in Section 2 and Section 5.]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes. Table 1 included time complexity of the algorithms.]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes. We describe the comprehensive explanations for our numerical experiments in Section 6]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [No]

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

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

Supplementary Materials for the Submission: "Near-Optimal Algorithm for Non-Stationary Kernelized Bandits"

A PROOF OF SECTION 3

A.1 Proof of Theorem 1

Proof. For any $H \in [T]$, we separate $[T]$ into the $\lceil T/H \rceil$ intervals. Then, suppose that the length of the interval $i \in [\lceil T/H \rceil - 1]$ and the last interval are chosen as H and $T - H(\lceil T/H \rceil - 1)$, respectively. For the notational convenience, we denote $\widetilde{H}_i$ as the length of the i-th interval. From Lemma 2, given any algorithm, the sequence $(f_t)_{t\in[T]}$ that satisfies 2 and the following properties exist:

1. For any interval i, there exist $f \in \mathcal{F}$ such that $f_t = f$ for all $\sum_{j=1}^{i-1} \widetilde{H}_j < t \leq \sum_{j=1}^{i} \widetilde{H}_j$, where $\mathcal{F}$ is the function set defined in Lemma 2.

2. The expected cumulative regret on the interval i is at least $\widetilde{H}_i \varepsilon(\widetilde{H}_i)$, where $\varepsilon(\cdot)$ is defined in Lemma 2.

Under such $(f_t)_{t\in[T]}$, we have

$$\mathbb{E}[R_T] \geq \sum_{i=1}^{\lceil T/H \rceil} \widetilde{H}_i \varepsilon(\widetilde{H}_i) \geq \lceil T/H \rceil H \varepsilon(H) \geq T\varepsilon(H). \tag{12}$$

Furthermore, since the f_t only changes $\lceil T/H \rceil - 1$ times from property 1, the following inequality holds from Lemma 2:

$$\sum_{t=2}^{T} \|f_t - f_{t-1}\|_\infty \leq 4\varepsilon(H) \left(\left\lceil \frac{T}{H} \right\rceil - 1 \right) \leq 4\varepsilon(H) \frac{T}{H}. \tag{13}$$

The results in Theorem 1 are obtained by choosing H such that the upper bound of (13) is equal or less than V_T.

Lower bound for the SE kernel From Lemma 2, $\varepsilon(H) = C_{\mathrm{SE}}^{(1)} \sqrt{(\ln H)^{d/2}/H}$ when $k = k_{\mathrm{SE}}$. Then,

$$\sum_{t=2}^{T} \|f_t - f_{t-1}\|_\infty \leq 4C_{\mathrm{SE}}^{(1)} (\ln T)^{d/4} H^{-3/2} T. \tag{14}$$

Here, we consider the setting $H = \lceil [4C_{\mathrm{SE}}^{(1)} (\ln T)^{d/4}]^{2/3} T^{2/3} V_T^{-2/3} \rceil$ and check if the condition $4C_{\mathrm{SE}}^{(1)} (\ln T)^{d/4} H^{-3/2} T \leq V_T$ is satisfied for such H. By aligning the condition $\widetilde{C}_{\mathrm{SE}} T^{-1/2} \ln^{d/4} T \leq V_T < \widetilde{C}_{\mathrm{SE}} T \ln^{d/4} T$ with $\widetilde{C}_{\mathrm{SE}} = 4C_{\mathrm{SE}}^{(1)}$, we can easily confirm that $1 < [4C_{\mathrm{SE}}^{(1)} (\ln T)^{d/4}]^{2/3} T^{2/3} V_T^{-2/3}$ and $H \leq T$ hold. Furthermore, we have

$$[4C_{\mathrm{SE}}^{(1)} (\ln T)^{d/4}]^{2/3} T^{2/3} V_T^{-2/3} \leq H \tag{15}$$

$$\Leftrightarrow 4C_{\mathrm{SE}}^{(1)} (\ln T)^{d/4} T V_T^{-1} \leq H^{3/2} \tag{16}$$

$$\Leftrightarrow 4C_{\mathrm{SE}}^{(1)} (\ln T)^{d/4} H^{-3/2} T \leq V_T. \tag{17}$$

Therefore, for sufficiently large T, we have

$$\begin{aligned}
\mathbb{E}[R_T] &\geq T\varepsilon(H) && (18)\\
&= TC_{\mathrm{SE}}^{(1)}(\ln H)^{d/4}H^{-1/2} && (19)\\
&\geq TC_{\mathrm{SE}}^{(1)}(\ln H)^{d/4}\{2[4C_{\mathrm{SE}}^{(1)}(\ln T)^{d/4}]^{2/3}T^{2/3}V_T^{-2/3}\}^{-1/2} && (20)\\
&\geq TC_{\mathrm{SE}}^{(1)}\left(\ln\left\{[4C_{\mathrm{SE}}^{(1)}(\ln T)^{d/4}]^{2/3}\overline{C}^{-2/3}T^{2(1-\overline{c})/3}\right\}\right)^{d/4}\{2[4C_{\mathrm{SE}}^{(1)}(\ln T)^{d/4}]^{2/3}T^{2/3}V_T^{-2/3}\}^{-1/2} && (21)\\
&\geq TC_{\mathrm{SE}}^{(1)}\left(-\frac{2}{3}\ln\overline{C}+\frac{2(1-\overline{c})}{3}\ln T\right)^{d/4}2^{-5/6}{C_{\mathrm{SE}}^{(1)}}^{-1/3}(\ln T)^{-d/12}T^{-1/3}V_T^{1/3} && (22)\\
&\geq 2^{-5/6}{C_{\mathrm{SE}}^{(1)}}^{2/3}C_{\overline{C},\overline{c},d}(\ln T)^{d/6}T^{2/3}V_T^{1/3}, && (23)
\end{aligned}$$

where $C_{\overline{C},\overline{c},d} > 0$ denote the constant that only depends on $\overline{C}, \overline{c}$, and d. In the above inequalities,

- the third line follows from $H \leq 2[4C_{\mathrm{SE}}^{(1)}(\ln T)^{d/4}]^{2/3}T^{2/3}V_T^{-2/3}$ since $1 < [4C_{\mathrm{SE}}^{(1)}(\ln T)^{d/4}]^{2/3}T^{2/3}V_T^{-2/3}$.
- the fourth line follows from $H \geq [4C_{\mathrm{SE}}^{(1)}(\ln T)^{d/4}]^{2/3}T^{2/3}V_T^{-2/3} \geq [4C_{\mathrm{SE}}^{(1)}(\ln T)^{d/4}]^{2/3}\overline{C}^{-2/3}T^{2(1-\overline{c})/3}$.
- the fifth line follows from the fact that $\ln[4C_{\mathrm{SE}}^{(1)}(\ln T)^{d/4}]^{2/3} \geq 0$ holds for sufficiently large T.
- the last line follows from the fact that there exists $C_{\overline{C},\overline{c},d} > 0$ such that $\left(-\frac{2}{3}\ln\overline{C}+\frac{2(1-\overline{c})}{3}\ln T\right)^{d/4} \geq C_{\overline{C},\overline{c},d}(\ln T)^{d/4}$ for sufficiently large $T \in \mathbb{N}_+$.

Finally, defining the constant C_{SE} as $C_{\mathrm{SE}} = 2^{-5/6}{C_{\mathrm{SE}}^{(1)}}^{2/3}C_{\overline{C},\overline{c},d}$, we obtain the desired result.

Lower bound for the Matérn kernel From Lemma 2, $\varepsilon(H) = C_{\mathrm{Mat}}^{(1)}H^{-\frac{\nu}{2\nu+d}}$ when $k = k_{\mathrm{Mat}}$. Then,

$$\sum_{t=2}^{T}\|f_t - f_{t-1}\|_\infty \leq 4C_{\mathrm{Mat}}^{(1)}H^{-\frac{3\nu+d}{2\nu+d}}T. \tag{24}$$

Here, we consider the setting $H = \lceil(4C_{\mathrm{Mat}}^{(1)}TV_T^{-1})^{\frac{2\nu+d}{3\nu+d}}\rceil$ and check if the condition $4C_{\mathrm{Mat}}^{(1)}H^{-\frac{3\nu+d}{2\nu+d}}T \leq V_T$ is satisfied for such H. By aligning the condition $\widetilde{C}_{\mathrm{Mat}}T^{-\frac{\nu}{2\nu+d}} \leq V_T \leq 2^{\frac{3\nu+d}{2\nu+d}}\widetilde{C}_{\mathrm{Mat}}T$ with $\widetilde{C}_{\mathrm{Mat}} = 4C_{\mathrm{Mat}}^{(1)}$, we can easily confirm that $1/2 \leq (4C_{\mathrm{Mat}}^{(1)}TV_T^{-1})^{\frac{2\nu+d}{3\nu+d}}$ and $H \leq T$ hold. Furthermore, we have

$$\begin{aligned}
&\frac{H}{2} \leq (4C_{\mathrm{Mat}}^{(1)}TV_T^{-1})^{\frac{2\nu+d}{3\nu+d}} \leq H && (25)\\
&\Leftrightarrow 2^{-\frac{3\nu+d}{2\nu+d}}H^{\frac{3\nu+d}{2\nu+d}} \leq 4C_{\mathrm{Mat}}^{(1)}TV_T^{-1} \leq H^{\frac{3\nu+d}{2\nu+d}} && (26)\\
&\Leftrightarrow 2^{-\frac{3\nu+d}{2\nu+d}}V_T \leq 4C_{\mathrm{Mat}}^{(1)}H^{-\frac{3\nu+d}{2\nu+d}}T \leq V_T. && (27)
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathbb{E}[R_T] &\geq T\varepsilon(H) && (28)\\
&= TC_{\mathrm{Mat}}^{(1)}H^{-\frac{\nu}{2\nu+d}} && (29)\\
&= \frac{H}{4}\cdot 4C_{\mathrm{Mat}}^{(1)}H^{-\frac{3\nu+d}{2\nu+d}}T && (30)\\
&\geq 2^{-2-\frac{3\nu+d}{2\nu+d}}V_TH && (31)\\
&\geq 2^{-2-\frac{3\nu+d}{2\nu+d}}(4C_{\mathrm{Mat}}^{(1)})^{\frac{2\nu+d}{3\nu+d}}T^{\frac{2\nu+d}{3\nu+d}}V_T^{\frac{\nu}{3\nu+d}}. && (32)
\end{aligned}$$

Defining the constant C_{Mat} as $C_{\mathrm{Mat}} = 2^{-2-\frac{3\nu+d}{2\nu+d}}(4C_{\mathrm{Mat}}^{(1)})^{\frac{2\nu+d}{3\nu+d}}$, we obtain the desired result. □

A.2 Proof of Lemma 2

Lemma 2 can be obtained with minor modification of the proof of Theorem 2 in Scarlett et al. [2017]. The required modifications are as follows:

- The upper bound of the total reward (Eq. (70) on p16 in Scarlett et al. [2017]). In the original paper, the upper bound of Eq. (70) is given as the worst-case total reward for the elements over $\mathcal{H}_k(B) \coloneqq \{f \in \mathcal{H}_k \mid \|f\|_{\mathcal{H}_k} \leq B\}$. This upper bound is derived from the upper bound of the average reward of a finite function family $\mathcal{F}$, which is constructed by the Fourier transform of bump functions (for details on $\mathcal{F}$, refer to Sec. 3 of Scarlett et al. [2017] or Sec. 4 of Cai and Scarlett [2021]). Since the upper bound on the average reward over $\mathcal{F}$ implies that the upper bound holds for some element in $\mathcal{F}$, we can prove the lower bound of regret for such f in the same manner as in the original paper.
- The output range of an element of aforementioned $\mathcal{F}$ is, due to its construction, bounded in $[-2\varepsilon(T), 2\varepsilon(T)]$. From this property, it is trivial that for any $f, \widetilde{f} \in \mathcal{F}$, $\|f - \widetilde{f}\| \leq 4\varepsilon(T)$ holds. Moreover, the constructions of $\varepsilon(T)$ for SE kernels and Matérn kernels are described on $p17$ in Scarlett et al. [2017].

B PROOF OF SECTION 5

B.1 Proof of Theorem 4

Lemma 6. *Fix any $i \in \lceil T/H \rceil$, $j \leq Q^{(i)} - 1$, and $\boldsymbol{x} \in \mathcal{X}$, where $Q^{(i)}$ is the total number of batch on i-th interval. Under Assumption 2, the following inequality holds when running Algorithm 1:*

$$\left| \mathbb{E}\left[\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) \,\middle|\, H_{j-1}^{(i)} \right] \right| \leq B\sigma\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}\right), \tag{33}$$

where $H_{j-1}^{(i)} \coloneqq \bigcup_{\widetilde{i} \leq i, \widetilde{j} \leq j-1} \{(\boldsymbol{x}_{\widetilde{j},m}^{(\widetilde{i})}, y_{\widetilde{j},m}^{(\widetilde{i})})\}_{m \leq N_{\widetilde{j}}^{\widetilde{i}}}$ represents the history up to $j-1$-th batch on i-th interval.

Proof. Note that the query candidate points $\widetilde{\boldsymbol{x}}_{j,m}^{(i)}$ are fixed given $H_{j-1}^{(i)}$. Furthermore, since the permutation index is chosen uniformly, we have $\mathbb{E}[f_{j,\psi(m)}^{(i)}(\widetilde{\boldsymbol{x}}_{j,m}^{(i)})] = \overline{f}_j^{(i)}(\widetilde{\boldsymbol{x}}_{j,m}^{(i)})$. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) \,\middle|\, H_{j-1}^{(i)} \right] && (34)\\
&= \overline{f}_j^{(i)}(\boldsymbol{x}) - \mathbb{E}\left[\mu\left(\boldsymbol{x}; \mathcal{S}_j^{(i)}, \widetilde{\boldsymbol{f}}_j^{(i)}\right) \,\middle|\, H_{j-1}^{(i)} \right] && (35)\\
&= \overline{f}_j^{(i)}(\boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x}, \mathcal{S}_j^{(i)})^\top (\mathbf{K}(\mathcal{S}_j^{(i)}, \mathcal{S}_j^{(i)}) + \lambda \boldsymbol{I}_t)^{-1} \mathbb{E}\left[\widetilde{\boldsymbol{f}}_j^{(i)} \,\middle|\, H_{j-1}^{(i)} \right] && (36)\\
&= \overline{f}_j^{(i)}(\boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x}, \mathcal{S}_j^{(i)})^\top (\mathbf{K}(\mathcal{S}_j^{(i)}, \mathcal{S}_j^{(i)}) + \lambda \boldsymbol{I}_t)^{-1} \overline{\boldsymbol{f}}_j^{(i)} && (37)\\
&= \overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathcal{S}_j^{(i)}, \overline{\boldsymbol{f}}_j^{(i)}\right), && (38)
\end{aligned}$$

where $\widetilde{\boldsymbol{f}}_j^{(i)} = \left[f_{j,\psi(m)}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j,m}^{(i)}\right) \right]_{m \in [N_j^{(i)}]}$ and $\overline{\boldsymbol{f}}_j^{(i)} = [\overline{f}_j^{(i)}(\widetilde{\boldsymbol{x}}_{j,m}^{(i)})]_{m \in [N_j^{(i)}]}$. Finally, by combining Proposition 1 in Vakili et al. [2021a] with $\|\overline{f}_j^{(i)}\|_{\mathcal{H}_k} \leq B$, we obtain

$$\left| \mathbb{E}\left[\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) \,\middle|\, H_{j-1}^{(i)} \right] \right| \leq B\sigma\left(\boldsymbol{x}; \mathcal{S}_j^{(i)}\right) = B\sigma\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}\right). \tag{39}$$

□

Lemma 7 (Theorem 3.1 in Adamczak et al. [2016])**.** *Fix any sequence $a_1, \ldots, a_n \in [0,1]$. Suppose that $h : [0,1]^n \to \mathbb{R}$ is an L-Lipschitz convex function. Then, the following inequality holds for any $\eta \geq 0$:*

$$\mathbb{P}(|h(x_{\psi(1)}, \ldots, x_{\psi(n)}) - \mathbb{E}[h(x_{\psi(1)}, \ldots, x_{\psi(n)})]| \geq \eta) \leq 2\exp\left(-\frac{c\eta^2}{L^2}\right), \tag{40}$$

where $c > 0$ is an absolute constant. Furthermore, $\psi(\cdot)$ represent the uniform permutation indices on the set $[n]$.

Lemma 8. *Fix any $i \in \lceil T/H \rceil$, $j \leq Q^{(i)} - 1$, $\boldsymbol{x} \in \mathcal{X}$, and $\delta \in (0,1)$. Then, under Assumptions 1, 2, the following inequality holds with probability at least $1-\delta$ when running Algorithm 1:*

$$\left|\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right)\right| \leq B\left(1 + \frac{C}{\sqrt{\lambda}}\sqrt{\ln\frac{2}{\delta}}\right)\sigma\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}\right), \tag{41}$$

where $C > 0$ is an absolute constant.

Proof. Let us define the function $h : [0,1]^{N_j^{(i)}} \to \mathbb{R}$ as $h(a_1, \ldots, a_{N_j^{(i)}}) = \boldsymbol{k}(\boldsymbol{x}, \mathcal{S}_j^{(i)})^\top(\mathbf{K}(\mathcal{S}_j^{(i)}, \mathcal{S}_j^{(i)}) + \lambda \boldsymbol{I}_t)^{-1}(a_1, \ldots, a_{N_j^{(i)}})^\top$. Then, given the history $H_{j-1}^{(i)}$, the function h is a fixed convex function. Furthermore, from Proposition 1 in Vakili et al. [2021a], the following inequality holds for any $\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)} \in [0,1]^{N_j^{(i)}}$:

$$\begin{aligned}
|h(\boldsymbol{a}^{(1)}) - h(\boldsymbol{a}^{(2)})| &= |\boldsymbol{k}(\boldsymbol{x}, \mathcal{S}_j^{(i)})^\top(\mathbf{K}(\mathcal{S}_j^{(i)}, \mathcal{S}_j^{(i)}) + \lambda \boldsymbol{I}_t)^{-1}(\boldsymbol{a}^{(1)} - \boldsymbol{a}^{(2)})| && (42)\\
&\leq \|\boldsymbol{k}(\boldsymbol{x}, \mathcal{S}_j^{(i)})^\top(\mathbf{K}(\mathcal{S}_j^{(i)}, \mathcal{S}_j^{(i)}) + \lambda \boldsymbol{I}_t)^{-1}\|_2 \|\boldsymbol{a}^{(1)} - \boldsymbol{a}^{(2)}\|_2 && (43)\\
&\leq \lambda^{-1/2}\sigma(\boldsymbol{x}; \mathbf{X}_j^{(i)})\|\boldsymbol{a}^{(1)} - \boldsymbol{a}^{(2)}\|_2. && (44)
\end{aligned}$$

Therefore, given the history $H_{j-1}^{(i)}$, h is an $\lambda^{-1/2}\sigma(\boldsymbol{x}; \mathbf{X}_j^{(i)})$-Lipschitz convex function. Here, for any $H_{j-1}^{(i)}$ and $\eta \geq 0$, we have

$$\begin{aligned}
&\left|\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) - \mathbb{E}\left[\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) \,\middle|\, H_{j-1}^{(i)}\right]\right| \geq \eta && (45)\\
&\Leftrightarrow \left|\mu\left(\boldsymbol{x}; \mathcal{S}_j^{(i)}, \widetilde{\boldsymbol{f}}_j^{(i)}\right) - \mathbb{E}\left[\mu\left(\boldsymbol{x}; \mathcal{S}_j^{(i)}, \widetilde{\boldsymbol{f}}_j^{(i)}\right) \,\middle|\, H_{j-1}^{(i)}\right]\right| \geq \eta && (46)\\
&\Leftrightarrow \left|\mu\left(\boldsymbol{x}; \mathcal{S}_j^{(i)}, \frac{\widetilde{\boldsymbol{f}}_j^{(i)} + B}{2B}\right) - \mathbb{E}\left[\mu\left(\boldsymbol{x}; \mathcal{S}_j^{(i)}, \frac{\widetilde{\boldsymbol{f}}_j^{(i)} + B}{2B}\right) \,\middle|\, H_{j-1}^{(i)}\right]\right| \geq \frac{\eta}{2B} && (47)\\
&\Leftrightarrow \left|h\left(\frac{\widetilde{\boldsymbol{f}}_j^{(i)} + B}{2B}\right) - \mathbb{E}\left[h\left(\frac{\widetilde{\boldsymbol{f}}_j^{(i)} + B}{2B}\right) \,\middle|\, H_{j-1}^{(i)}\right]\right| \geq \frac{\eta}{2B}. && (48)
\end{aligned}$$

By noting $\frac{\widetilde{f}_{j,m}^{(i)} + B}{2B} \in [0,1]$, we have the following inequality from Lemma 7:

$$\begin{aligned}
&\mathbb{P}\left(\left|\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) - \mathbb{E}\left[\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) \,\middle|\, H_{j-1}^{(i)}\right]\right| \geq \eta \,\middle|\, H_{j-1}^{(i)}\right) && (49)\\
&\leq 2\exp\left(-\frac{c\eta^2}{4B^2\lambda^{-1}\sigma^2(\boldsymbol{x}; \mathbf{X}_j^{(i)})}\right). && (50)
\end{aligned}$$

Setting η as $\eta = 2B\lambda^{-1/2}\sigma(\boldsymbol{x}; \mathbf{X}_j^{(i)})\sqrt{c^{-1}\ln(2/\delta)}$ in the above inequality, we obtain

$$\mathbb{P}\left(\left|\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) - \mathbb{E}\left[\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) \,\middle|\, H_{j-1}^{(i)}\right]\right| < \eta \,\middle|\, H_{j-1}^{(i)}\right) \geq 1 - \delta. \tag{51}$$

Here, by combining the above inequality with Lemma 6, we have

$$\mathbb{P}\left(\left|\overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right)\right| < B\left(1 + 2\lambda^{-1/2}\sqrt{c^{-1}\ln(2/\delta)}\right)\sigma(\boldsymbol{x}; \mathbf{X}_j^{(i)}) \,\middle|\, H_{j-1}^{(i)}\right) \geq 1 - \delta. \tag{52}$$

From the tower property of the conditional expectation, we obtain the desired result by setting the absolute constant C as $C = 2\sqrt{c^{-1}}$. □

Now, we describe the proof of Theorem 4.

Proof of Theorem 4. From Lemma 8 and the union bound, with probability at least $1 - \delta/2$, the following inequality holds for any $i \leq \lceil T/H \rceil$, $j \leq Q^{(i)} - 1$, and $\boldsymbol{x} \in \mathcal{X}$:

$$\left| \overline{f}_j^{(i)}(\boldsymbol{x}) - \mu\left(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{f}_j^{(i)}\right) \right| < B\left(1 + C\lambda^{-1/2}\sqrt{\ln(4|\mathcal{X}|\widetilde{Q}_{T,H}/\delta)}\right)\sigma(\boldsymbol{x}; \mathbf{X}_j^{(i)}), \tag{53}$$

where $\widetilde{Q}_{T,H} = \sum_{i=1}^{\lceil T/H \rceil}(Q^{(i)} - 1)$. Furthermore, as with the proof of Theorem 1 in Vakili et al. [2021a], with probability at least $1 - \delta/2$, the following inequality holds for any $i \leq \lceil T/H \rceil$, $j \leq Q^{(i)} - 1$, and $\boldsymbol{x} \in \mathcal{X}$:

$$\left| \mu(\boldsymbol{x}; \mathbf{X}_j^{(i)}, \boldsymbol{\epsilon}_j^{(i)}) \right| < \frac{\rho}{\sqrt{\lambda}}\sqrt{2\ln\frac{4|\mathcal{X}|\widetilde{Q}_{T,H}}{\delta}}\sigma(\boldsymbol{x}; \mathbf{X}_j^{(i)}), \tag{54}$$

Finally, by noting that $\widetilde{Q}_{T,H} \leq Q_{T,H}$ holds from Proposition 1 in Li and Scarlett [2022], we have the desired result.

□

B.2 Proof of Theorem 5

Lemma 9. *For any $i \leq \lceil T/H \rceil$, $j \leq Q^{(i)}$, and $m \leq N_j^{(i)}$, the following inequality holds:*

$$f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)*}\right) - f_{j,m}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) \leq 2V_T^{(i)}, \tag{55}$$

where $\boldsymbol{x}_{j,m}^{(i)*} \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} f_{j,m}^{(i)}(\boldsymbol{x})$, $\overline{\boldsymbol{x}}_j^{(i)*} \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} \overline{f}_j^{(i)}(\boldsymbol{x})$, and $V_T^{(i)} = \sum_{t=(i-1)H+1}^{iH-1} \|f_{t+1} - f_t\|_\infty$

Proof. We prove this by contradiction. Assume that $m^* \in \arg\max_{m\in\left[N_j^{(i)}\right]}\left[f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)*}\right) - f_{j,m}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right)\right]$ and that $f_{j,m^*}^{(i)}\left(\boldsymbol{x}_{j,m^*}^{(i)*}\right) - f_{j,m^*}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) > 2V_T^{(i)}$ holds. Then, for any $m \in \left[N_j^{(i)}\right]$, we have:

$$\begin{aligned} f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m^*}^{(i)*}\right) &\geq f_{j,m^*}^{(i)}\left(\boldsymbol{x}_{j,m^*}^{(i)*}\right) - V_T^{(i)} && (56)\\ &> f_{j,m^*}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) + V_T^{(i)} && (57)\\ &\geq f_{j,m}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right). && (58) \end{aligned}$$

Therefore,

$$\overline{f}_j^{(i)}\left(\boldsymbol{x}_{j,m^*}^{(i)*}\right) > \overline{f}_j^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right). \tag{59}$$

This contradicts the definition of $\overline{\boldsymbol{x}}_j^{(i)*}$. □

Lemma 10. *For any $i \leq \lceil T/H \rceil$, $j \leq Q^{(i)}$, and $m \leq N_j^{(i)}$, the following inequality holds:*

$$f_{j,m}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) - \overline{f}_j^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) \leq V_T^{(i)}. \tag{60}$$

Proof.

$$\begin{aligned} f_{j,m}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) - \overline{f}_j^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) &= \frac{1}{N_j^{(i)}}\sum_{\widetilde{m}=1}^{N_j^{(i)}}\left[f_{j,m}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) - f_{j,\widetilde{m}}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right)\right] && (61)\\ &\leq \frac{1}{N_j^{(i)}}\sum_{\widetilde{m}=1}^{N_j^{(i)}} V_T^{(i)} && (62)\\ &= V_T^{(i)}. && (63) \end{aligned}$$

□

Lemma 11. *For any $i \le \lceil T/H \rceil$ and j $\left(2 \le j \le Q^{(i)}\right)$, the following inequality holds:*

$$\overline{f}_j^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) - \overline{f}_{j-1}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) \le V_T^{(i)}. \tag{64}$$

Proof. Let $m^* = \arg\max_{m \in [N_j^{(i)}]} f_{j,m}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right)$, then

$$\begin{aligned}
\overline{f}_j^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) - \overline{f}_{j-1}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) &\le \overline{f}_j^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) - \overline{f}_{j-1}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) && (65)\\
&= \frac{1}{N_{j-1}^{(i)}} \sum_{\widetilde{m}=1}^{N_{j-1}^{(i)}} \left[\overline{f}_j^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) - f_{j-1,\widetilde{m}}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right)\right] && (66)\\
&\le \frac{1}{N_{j-1}^{(i)}} \sum_{\widetilde{m}=1}^{N_{j-1}^{(i)}} \left[f_{j,m^*}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right) - f_{j-1,\widetilde{m}}^{(i)}\left(\overline{\boldsymbol{x}}_j^{(i)*}\right)\right] && (67)\\
&\le \frac{1}{N_{j-1}^{(i)}} \sum_{\widetilde{m}=1}^{N_{j-1}^{(i)}} V_T^{(i)} && (68)\\
&= V_T^{(i)}. && (69)
\end{aligned}$$

□

Lemma 12. *Suppose the following event holds:*

$$\forall i \le \left\lceil \frac{T}{H} \right\rceil, \forall j \le Q^{(i)} - 1, \forall \boldsymbol{x} \in \mathcal{X}, \mathrm{lcb}_j^{(i)}(\boldsymbol{x}) \le \overline{f}_j^{(i)}(\boldsymbol{x}) \le \mathrm{ucb}_j^{(i)}(\boldsymbol{x}). \tag{70}$$

Then, for any $i \le \lceil T/H \rceil$, j $\left(2 \le j \le Q^{(i)}\right)$, and $m \le N_j^{(i)}$, we have

$$\overline{f}_{j-1}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) - \overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right) \le \left(\log_2 \log_2 T^{(i)} + 1\right) V_T^{(i)}, \tag{71}$$

where $\widetilde{\boldsymbol{x}}_j^{(i)*} \in \arg\max_{\boldsymbol{x} \in \mathcal{X}_j^{(i)}} \overline{f}_j^{(i)}(\boldsymbol{x})$.

Proof. The case where $\overline{\boldsymbol{x}}_{j-1}^{(i)*} = \widetilde{\boldsymbol{x}}_{j-1}^{(i)*}$ is trivial. When $\overline{\boldsymbol{x}}_{j-1}^{(i)*} \ne \widetilde{\boldsymbol{x}}_{j-1}^{(i)*}$, there exists some $\widetilde{j} < j-1$ such that

$$\overline{\boldsymbol{x}}_{j-1}^{(i)*} \in \mathcal{X}_{\widetilde{j}}^{(i)} \quad \text{and} \quad \overline{\boldsymbol{x}}_{j-1}^{(i)*} \notin \mathcal{X}_{\widetilde{j}+1}^{(i)}. \tag{72}$$

In this case, we have

$$\begin{aligned}
&\overline{f}_{j-1}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) - \overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right) && (73)\\
&= \overline{f}_{j-1}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) - \overline{f}_{\widetilde{j}}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) + \overline{f}_{\widetilde{j}}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) - \overline{f}_{\widetilde{j}}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widetilde{j}}^{(i)*}\right) + \overline{f}_{\widetilde{j}}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widetilde{j}}^{(i)*}\right) - \overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right) && (74)\\
&\le V_T^{(i)} + \sum_{\widehat{j}=\widetilde{j}}^{j-2} \left[\overline{f}_{\widehat{j}}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widehat{j}}^{(i)*}\right) - \overline{f}_{\widehat{j}+1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widehat{j}+1}^{(i)*}\right)\right]. && (75)
\end{aligned}$$

In the final line, we use the fact that $\overline{f}_{j-1}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) - \overline{f}_{\widetilde{j}}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) \le V_T^{(i)}$, and since $\overline{\boldsymbol{x}}_{j-1}^{(i)*} \in \mathcal{X}_{\widetilde{j}}^{(i)}$, it follows that $\overline{f}_{\widetilde{j}}^{(i)}\left(\overline{\boldsymbol{x}}_{j-1}^{(i)*}\right) \le$

$\overline{f}_{\widetilde{j}}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widetilde{j}}^{(i)*}\right)$. Under the event (70), for any $\widehat{j}$, we have $\widetilde{\boldsymbol{x}}_{\widehat{j}}^{(i)*} \in \mathcal{X}_{\widehat{j}+1}^{(i)}$. Therefore,

$$\sum_{\widehat{j}=\widetilde{j}}^{j-2}\left[\overline{f}_{\widehat{j}}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widehat{j}}^{(i)*}\right)-\overline{f}_{\widehat{j}+1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widehat{j}+1}^{(i)*}\right)\right] \tag{76}$$

$$=\sum_{\widehat{j}=\widetilde{j}}^{j-2}\left[\overline{f}_{\widehat{j}}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widehat{j}}^{(i)*}\right)-\overline{f}_{\widehat{j}+1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widehat{j}}^{(i)*}\right)+\overline{f}_{\widehat{j}+1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widehat{j}}^{(i)*}\right)-\overline{f}_{\widehat{j}+1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{\widehat{j}+1}^{(i)*}\right)\right] \tag{77}$$

$$\leq \sum_{\widehat{j}=\widetilde{j}}^{j-2} V_T^{(i)} \tag{78}$$

$$\leq (Q^{(i)}-2)V_T^{(i)} \tag{79}$$

$$\leq V_T^{(i)}\log_2\log_2 T^{(i)}. \tag{80}$$

The final line uses Proposition 1 from Li and Scarlett [2022]. □

Lemma 13. *For any $i \leq \lceil T/H \rceil$, j $(2 \leq j \leq Q^{(i)})$, and $m \leq N_j^{(i)}$, the following inequality holds:*

$$\overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)-f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)\leq V_T^{(i)}. \tag{81}$$

Proof.

$$\overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)-f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)\leq N_{j-1}^{(i)-1}\sum_{\widetilde{m}=1}^{N_{j-1}^{(i)}}\left[f_{j,\widetilde{m}}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)-f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)\right] \tag{82}$$

$$\leq N_{j-1}^{(i)-1}\sum_{\widetilde{m}=1}^{N_{j-1}^{(i)}} V_T^{(i)} \tag{83}$$

$$\leq V_T^{(i)}. \tag{84}$$

□

Lemma 14. *Suppose that the event* (70) *holds. Then, for any $i \leq \lceil T/H \rceil$, the following inequality holds:*

$$\sum_{j=2}^{Q^{(i)}}\sum_{m=1}^{N_j^{(i)}}\left[\overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)-\overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)\right]\leq 4\left(\log_2\log_2 T^{(i)}+1\right)\left(\sqrt{T^{(i)}}+T^{(i)-1/4}\right)\sqrt{C_1\gamma_{T^{(i)}}\beta_T}, \tag{85}$$

where $C_1 = 8/\ln(1+\sigma^{-2})$.

Proof. Under the event (70),

$$\overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)-\overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right) \tag{86}$$

$$\leq \mathrm{ucb}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)-\mathrm{lcb}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right) \tag{87}$$

$$=\mathrm{ucb}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)-\mathrm{ucb}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)+2\beta_T^{1/2}\sigma\left(\boldsymbol{x}_{j,m}^{(i)};\mathbf{X}_{j-1}^{(i)}\right) \tag{88}$$

$$\leq \mathrm{ucb}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)-\max_{\boldsymbol{x}\in\mathcal{X}_{j-1}^{(i)}}\mathrm{lcb}_{j-1}^{(i)}(\boldsymbol{x})+2\beta_T^{1/2}\sigma\left(\boldsymbol{x}_{j,m}^{(i)};\mathbf{X}_{j-1}^{(i)}\right) \tag{89}$$

$$\leq \mathrm{ucb}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)-\mathrm{lcb}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)+2\beta_T^{1/2}\sigma\left(\boldsymbol{x}_{j,m}^{(i)};\mathbf{X}_{j-1}^{(i)}\right) \tag{90}$$

$$\leq 4\beta_T^{1/2}\max_{\boldsymbol{x}\in\mathcal{X}_{j-1}^{(i)}}\sigma\left(\boldsymbol{x};\mathbf{X}_{j-1}^{(i)}\right) \tag{91}$$

$$\leq 4\sqrt{\frac{C_1\beta_T\gamma_{N_{j-1}^{(i)}}}{N_{j-1}^{(i)}}} \tag{92}$$

$$\leq 4\sqrt{\frac{C_1\beta_T\gamma_{T^{(i)}}}{N_{j-1}^{(i)}}}. \tag{93}$$

Therefore,

$$\sum_{j=2}^{Q^{(i)}}\sum_{m=1}^{N_j^{(i)}}\left[\overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)-\overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)\right] \tag{94}$$

$$\leq 4\sum_{j=2}^{Q^{(i)}}N_j^{(i)}\sqrt{\frac{C_1\beta_T\gamma_{T^{(i)}}}{N_{j-1}^{(i)}}} \tag{95}$$

$$\leq 4\left(Q^{(i)}-1\right)\left(\sqrt{T^{(i)}}+T^{(i)-1/4}\right)\sqrt{C_1\beta_T\gamma_{T^{(i)}}} \tag{96}$$

$$\leq 4\left(\log_2\log_2 T^{(i)}+1\right)\left(\sqrt{T^{(i)}}+T^{(i)-1/4}\right)\sqrt{C_1\beta_T\gamma_{T^{(i)}}}. \tag{97}$$

In the second line from the bottom, we used the fact that $\frac{N_j^{(i)}}{\sqrt{N_{j-1}^{(i)}}}\leq\sqrt{T^{(i)}}+T^{(i)-1/4}$ holds (for example, see the proof of Theorem 1 in Li and Scarlett [2022]). The last line follows from Proposition 1 in Li and Scarlett [2022]. □

Proof of Theorem 5. Let $R_T^{(i)}\coloneqq\sum_{j=1}^{Q^{(i)}}\sum_{m=1}^{N_j^{(i)}}f_{j,m}^{(i)}(\boldsymbol{x}_{j,m}^{(i)*})-f_{j,m}^{(i)}(\boldsymbol{x}_{j,m}^{(i)})$ be the regret incurred over i-th interval. Then, we decompose the regret $R_T^{(i)}$ as

$$R_T^{(i)}=\sum_{m=1}^{N_1^{(i)}}f_{1,m}^{(i)}\left(\boldsymbol{x}_{1,m}^{(i)*}\right)-f_{1,m}^{(i)}\left(\boldsymbol{x}_{1,m}^{(i)}\right)+\sum_{j=2}^{Q^{(i)}}\sum_{m=1}^{N_j^{(i)}}f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)*}\right)-\overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right) \tag{98}$$

$$+\sum_{j=2}^{Q^{(i)}}\sum_{m=1}^{N_j^{(i)}}\overline{f}_{j-1}^{(i)}\left(\widetilde{\boldsymbol{x}}_{j-1}^{(i)*}\right)-\overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)+\sum_{j=2}^{Q^{(i)}}\sum_{m=1}^{N_j^{(i)}}\overline{f}_{j-1}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right)-f_{j,m}^{(i)}\left(\boldsymbol{x}_{j,m}^{(i)}\right). \tag{99}$$

By using the aforementioned lemmas, we can obtain the upper bound of each term as follows:

- Since $N_1^{(i)}\leq\sqrt{T^{(i)}}+1$, the first term is bounded from above by $2B(\sqrt{T^{(i)}}+1)$.
- From Lemmas 9–12, the second term is bounded from above by $(6+\log_2\log_2 T^{(i)})T^{(i)}V_T^{(i)}$.
- From Lemma 14, the third term is bounded from above by $4\left(\log_2\log_2 T^{(i)}+1\right)\left(\sqrt{T^{(i)}}+T^{(i)-1/4}\right)\sqrt{C_1\gamma_{T^{(i)}}\beta_T}$.

- From Lemma 13, the fourth term is bounded from above by $T^{(i)} V_T^{(i)}$.

Combining the above upper bounds with $T^{(i)} \leq H$ and $V_T = \sum_{i=1}^{\lceil T/H \rceil} V_T^{(i)}$, we have

$$R_T = \sum_{i=1}^{\lceil T/H \rceil} R_T^{(i)} \tag{100}$$

$$\leq (7 + \log_2 \log_2 H) V_T H + \left\lceil \frac{T}{H} \right\rceil \left[2B \left(\sqrt{H} + 1 \right) + 4 \left(\log_2 \log_2 H + 1 \right) \left(\sqrt{H} + 1 \right) \sqrt{C_1 \gamma_H \beta_T} \right]. \tag{101}$$

The desired results are obtained by choosing the proper H in the above inequality.

For SE kernel When $k = k_{\mathrm{SE}}$, $\gamma_H = O\left(\ln^{d+1} H \right)$. Therefore, we have

$$R_T = O\left(V_T H \log_2 \log_2 T + T \left(\log_2 \log_2 T \right) (\ln T)^{1/2} \sqrt{\frac{\ln^{d+1} H}{H}} \right). \tag{102}$$

Here,

$$H \geq T^{2/3} V_T^{-2/3} (\ln T)^{(d+2)/3} \tag{103}$$

$$\Leftrightarrow H^{3/2} \geq V_T^{-1} T (\ln T)^{1/2} (\ln T)^{(d+1)/2} \tag{104}$$

$$\Rightarrow H^{3/2} \geq V_T^{-1} T (\ln T)^{1/2} (\ln H)^{(d+1)/2} \tag{105}$$

$$\Rightarrow V_T H \log_2 \log_2 T \geq T \left(\log_2 \log_2 T \right) (\ln T)^{1/2} \sqrt{\frac{\ln^{d+1} H}{H}}. \tag{106}$$

Therefore, by setting $H = \left\lceil T^{2/3} V_T^{-2/3} (\ln T)^{(d+2)/3} \right\rceil$,

$$V_T H \log_2 \log_2 T + T \left(\log_2 \log_2 T \right) (\ln T)^{1/2} \sqrt{\frac{\ln^{d+1} H}{H}} \tag{107}$$

$$\leq 2 V_T \left(T^{2/3} V_T^{-2/3} (\ln T)^{(d+2)/3} + 1 \right) \left(\log_2 \log_2 T \right) \tag{108}$$

$$= \widetilde{O}\left(V_T^{\frac{1}{3}} T^{\frac{2}{3}} \right). \tag{109}$$

For Matérn kernel When $k = k_{\text{Matérn}}$, $\gamma_H = O\left(H^{\frac{d}{2\nu+d}} \ln^{\frac{2\nu}{2\nu+d}} H \right)$. Therefore,

$$R_T = O\left(V_T H \log_2 \log_2 T + T \left(\log_2 \log_2 T \right) (\ln T)^{1/2} \sqrt{\frac{H^{\frac{d}{2\nu+d}} \ln^{\frac{2\nu}{2\nu+d}} H}{H}} \right). \tag{110}$$

Now,

$$H \geq V_T^{-\frac{2\nu+d}{3\nu+d}} T^{\frac{2\nu+d}{3\nu+d}} (\ln T)^{\frac{4\nu+d}{2(3\nu+d)}} \tag{111}$$

$$\Leftrightarrow H^{\frac{3\nu+d}{2\nu+d}} \geq V_T^{-1} T (\ln T)^{1/2} (\ln T)^{\frac{2\nu}{2(2\nu+d)}} \tag{112}$$

$$\Rightarrow H^{\frac{3\nu+d}{2\nu+d}} \geq V_T^{-1} T (\ln T)^{1/2} (\ln H)^{\frac{2\nu}{2(2\nu+d)}} \tag{113}$$

$$\Leftrightarrow V_T H \log_2 \log_2 T \geq T \left(\log_2 \log_2 T \right) (\ln T)^{1/2} \sqrt{\frac{H^{\frac{d}{2\nu+d}} \ln^{\frac{2\nu}{2\nu+d}} H}{H}}. \tag{114}$$

Therefore, by setting $H = \left\lceil V_T^{-\frac{2\nu+d}{3\nu+d}} T^{\frac{2\nu+d}{3\nu+d}} (\ln T)^{\frac{4\nu+d}{6\nu+2d}} \right\rceil$, we have

$$V_T H \log_2 \log_2 T + T \left(\log_2 \log_2 T\right) (\ln T)^{1/2} \sqrt{\frac{H^{\frac{d}{2\nu+d}} \ln^{\frac{2\nu}{2\nu+d}} H}{H}} \tag{115}$$

$$\leq 2 V_T H \log_2 \log_2 T \tag{116}$$

$$= \widetilde{O}\left(T^{\frac{2\nu+d}{3\nu+d}} V_T^{\frac{\nu}{3\nu+d}}\right). \tag{117}$$

□

C NEAR-OPTIMAL VERSION OF OPKB FOR MATÉRN RKHS

We show that the OPKB algorithm with the restart-reset strategy can achieve near-optimal regret upper bound even in the Matérn RKHS by properly selecting the restarting interval. The formal statement is described in the following Theorem 15.

Theorem 15 (The modified version of OPKB algorithm with the restart-reset strategy.)**.** *Assume that the underlying kernel is Matérn kernel with smoothness parameter $\nu > 1/2$. Furthermore, suppose that Assumptions 1–3 and $V_T \geq T^{-\frac{\nu}{2\nu+d}} \ln^{\frac{\nu+d}{2\nu+d}} T$ hold. Then, if we set the restarting interval H as $H = \lceil T^{\frac{2\nu+d}{3\nu+d}} V_T^{-\frac{2\nu+d}{3\nu+d}} \ln^{\frac{4\nu+d}{6\nu+2d}} T \rceil$, the OPKB algorithm (Algorithm 2 in Hong et al. [2023]) with the restart-reset strategy achieve $R_T = \widetilde{O}(V_T^{\frac{\nu}{3\nu+d}} T^{\frac{2\nu+d}{3\nu+d}})$ with probability at least $1-\delta$.*

To show the above statement, it is enough to show that $R_T^{(i)} = \widetilde{O}(V_T^{(i)} H + \sqrt{\gamma_H H})$, where $R_T^{(i)}$ denote the cumulative regret among i-th interval (also defined in the proof of Theorem 5). This is because that $\widetilde{O}(V_T^{(i)} H + \sqrt{\gamma_H H})$ interval regret implies the same order of the cumulative regret as that of R-PERP (as shown in Eq. (101)); therefore, if we once obtain the $\widetilde{O}(V_T^{(i)} H + \sqrt{\gamma_H H})$ interval regret, we can obtain the $\widetilde{O}(V_T^{\frac{\nu}{3\nu+d}} T^{\frac{2\nu+d}{3\nu+d}})$ cumulative regret by following final part of the proof of Theorem 5.

Note that the stationary base algorithm (without restart-reset strategy) with $\widetilde{O}(V_T T + \sqrt{\gamma_T T})$ cumulative regret under non-stationary environment achieve $\widetilde{O}(V_T^{(i)} H + \sqrt{\gamma_H H})$ interval regret when restart-reset strategy is applied. Therefore, we show that $\widetilde{O}(V_T T + \sqrt{\gamma_T T})$ cumulative regret of the standard OPKB algorithm under Assumptions 1–3.

Hereafter, we use the same notation as those of Hong et al. [2023] for our proof unless otherwise specified.

Proof of Theorem 15. As with the proof of Theorem 4.6 in Hong et al. [2023], we have the following inequality by leveraging Azuma-Hoeffding inequality[5]:

$$R_{\mathcal{B}(j)} \leq \widetilde{O}(\sqrt{2^j E}) + \sum_{t \in \mathcal{B}(j)} f_t(\boldsymbol{x}_t^*) - \mathbb{E}_t[f_t(\boldsymbol{x}_t)], \tag{118}$$

where $\mathcal{B}(j) \in [T]$ is the j-th batch index set of OPKB, and $R_{\mathcal{B}(j)}$ denote the cumulative regret incurred over $\mathcal{B}(j)$. Furthermore, we decompose the second term into the following four terms:

$$f_t(\boldsymbol{x}_t^*) - \mathbb{E}_t[f_t(\boldsymbol{x}_t)] \tag{119}$$

$$= \underbrace{f_t(\boldsymbol{x}_t^*) - f_t(\overline{\boldsymbol{x}}^*_{C(j-1)})}_{A_1} + \underbrace{f_t(\overline{\boldsymbol{x}}^*_{C(j-1)}) - \overline{f}_{C(j-1)}(\overline{\boldsymbol{x}}^*_{C(j-1)})}_{A_2} \tag{120}$$

$$+ \underbrace{\mathbb{E}_t[\overline{f}_{C(j-1)}(\overline{\boldsymbol{x}}^*_{C(j-1)}) - \overline{f}_{C(j-1)}(\boldsymbol{x}_t)]}_{A_3} + \underbrace{\mathbb{E}_t[\overline{f}_{C(j-1)}(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t)]}_{A_4}, \tag{121}$$

where $C(j) \coloneqq \bigcup_{\widetilde{j} \leq j} \mathcal{B}(\widetilde{j})$ is the cumulative time step set until the end of the batch j. Furthermore, $\overline{f}_{C(j-1)}(\boldsymbol{x}) \coloneqq \sum_{t \in C(j-1)} f_t(\boldsymbol{x}) / |C(j-1)|$ and $\overline{\boldsymbol{x}}^*_{C(j-1)}$ represent the average function over $C(j-1)$ and its maximizer, respectively. As with the proof of stationary OPKB (Theorem 4.6 in Hong et al. [2023]), the upper bound of A_3 is obtained as $O(V_{C(j)} + \mu_j)$

[5]The setting of Hong et al. [2023] assumes the boundness assumption $f_t(\boldsymbol{x}) \in [0,1]$ of the reward function. On the other hand, Assumption 2 implies $f_t(\boldsymbol{x}) \in [-B, B]$ in our setting. This difference in the magnitude of the reward function does not break the validity of the proof and appears as the difference of the constant factor.

Table 2: The comparison of existing and our algorithms for regrets and computational costs under the setting where V_T is unknown. For the regret upper bound described in the table, we assume that V_T satisfies $V_T > c$, where $c > 0$ is any fixed constant.

Algorithm	Regret (SE)	Regret (Matérn)	Computational cost at step $t \leq T$
R/SW-GP-UCB	$\widetilde{O}(T^{\frac{3}{4}} V_T)$	$\widetilde{O}(T^{\frac{12\nu+13d}{16\nu+8d}} V_T)$	$O(\lvert\mathcal{X}\rvert t^2)$
WGP-UCB	$\widetilde{O}(T^{\frac{3}{4}} V_T)$	$\widetilde{O}(T^{\frac{12\nu+13d}{16\nu+8d}} V_T)$	$O(\lvert\mathcal{X}\rvert t^2)$
OPKB	$\widetilde{O}(T^{\frac{2}{3}} V_T^{\frac{1}{3}})$	$\widetilde{O}(T^{\frac{4\nu+3d}{6\nu+3d}} V_T^{\frac{1}{3}})$	$O(M\lvert\mathcal{X}\rvert^3)$
R-PERP (Ours)	$\widetilde{O}(T^{\frac{2}{3}} V_T)$	$\widetilde{O}(T^{\frac{2\nu+d}{3\nu+d}} V_T)$	$O(\lvert\mathcal{X}\rvert t^2)$

using Lemma 4.5 in Hong et al. [2023]. As for A_1, A_2, and A_4, we can easily confirm that the sum of those terms is $O(V_T)$ (e.g., by relying on the same arguments as those of Lemma 9–12). By aggregating and arranging the above upper bounds, the cumulative regret $R_{\mathcal{B}(j)}$ over $\mathcal{B}(j)$ is bounded from above by $O(V_T|\mathcal{B}(j)| + E\sqrt{2^j})$. Therefore, we have

$$R_T = \sum_j R_{\mathcal{B}(j)} \tag{122}$$

$$= O(V_T T) + \sum_j \widetilde{O}(E\sqrt{2^j}) \tag{123}$$

$$= \widetilde{O}(V_T T + E\sqrt{T/E}) \tag{124}$$

$$= \widetilde{O}(V_T T + \sqrt{\gamma_T T \ln |\mathcal{X}|}), \tag{125}$$

where the third line follows from Schwarz's inequality. As described before the proof, the above upper bound implies the $\widetilde{O}(V_T^{(i)} H + \sqrt{\gamma_H H})$ interval regret when the OPKB algorithm is used as the base algorithm of the restart-reset-based procedures. Therefore, the proof is completed. □

D REGRET UPPER BOUND FOR UNKNOWN V_T

In this section, we derive the regret upper bound for R-PERP in a setting where the total drift upper bound V_T is unknown. Importantly, in the proof in Theorem 5, the prior information of V_T is used only for tightening the regret upper bound. That is, Lemmas 9-14 hold for any $H \in [T] \setminus \{1\}$. Therefore, the derivation of Eqs. (98)-(101) can be obtained even if H is determined independently of V_T. Hence, the following regret upper bound can be obtained by choosing H independently of V_T and slightly modifying the derivation of Eqs. (102)-(109) and Eqs. (110)-(117) for SE and Matérn kernels, respectively:

- **SE Kernel**: By setting $H = \lceil T^{2/3} (\ln T)^{(d+2)/3} \rceil$, it can be shown in a similar manner to Eq. (103)–(106) that

$$H \log_2 \log_2 T \geq T \left(\log_2 \log_2 T\right) (\ln T)^{1/2} \sqrt{\frac{\ln^{d+1} H}{H}}. \tag{126}$$

  Therefore, $R_T = \widetilde{O}(T^{2/3}(V_T + 1))$ is satisfied by R-PERP.

- **Matérn Kernel**: By setting $H = \lceil T^{\frac{2\nu+d}{3\nu+d}} (\ln T)^{\frac{4\nu+d}{6\nu+2d}} \rceil$, we have

$$H \log_2 \log_2 T \geq T \left(\log_2 \log_2 T\right) (\ln T)^{1/2} \sqrt{\frac{H^{\frac{d}{2\nu+d}} \ln^{\frac{2\nu}{2\nu+d}} H}{H}}$$

  in a similar manner to Eq. (111)–(114); thus, combining with Eq. (110), it can be shown that R-PERP satisfies the regret upper bound $R_T = \widetilde{O}(T^{\frac{2\nu+d}{3\nu+d}} (V_T + 1))$.

Table 2 lists the regret and computational complexity of each method in the setting where V_T is unknown. In the setting where V_T is unknown, OPKB shows better theoretical performance than R-PERP in terms of the dependence of V_T by utilizing an appropriate adaptive reset scheduling. Extending R-PERP based on the ideas of adaptive reset scheduling of OPKB is an important future research direction. On the other hand, the computational complexity issue with OPKB exists regardless of whether V_T is known or unknown. Therefore, R-PERP shows the best regret guarantees among the algorithms applicable in scenarios where $\mathcal{X}$ is huge, such as R/SW-GP-UCB and W-GP-UCB.