Risk Seeking Bayesian Optimization under Uncertainty for Obtaining Extremum

Shogo Iwazaki
MI-6 Ltd.

Tomohiko Tanabe
MI-6 Ltd.

Mitsuru Irie
MI-6 Ltd.

Shion Takeno
RIKEN

Yu Inatsu
Nagoya Institute of Technology

Abstract

Real-world black-box optimization tasks often focus on obtaining the best reward, which includes an intrinsic random quantity from uncontrollable environmental factors. For this problem, we formulate a novel risk-seeking optimization problem whose goal is to obtain the best possible reward within a fixed budget under uncontrollable factors. We consider two settings: (1) environmental model setting for the case that uncontrollable environmental variables can be combined with surrogate models for optimization and (2) heteroscedastic model setting for the case that uncontrollable environmental variables are hard to model. We propose a novel Bayesian optimization method called kernel explore-then-commit (kernel-ETC) and provide the regret upper bound for both settings. We demonstrate the effectiveness of kernel-ETC through several numerical experiments, including the hyperparameter tuning task and the simulation function derived from polymer synthesis real data.

1 INTRODUCTION

Black-box optimization problems with costly objective functions frequently arise in a wide range of real-world problems, including robotics (Lizotte et al., 2007), experimental design (González et al., 2015), and hyperparameter tuning of machine learning models (Snoek et al., 2012). Bayesian optimization (BO) (Frazier, 2018) is a powerful framework for solving black-box optimization tasks efficiently.

Real-world optimization under the uncertainty of uncontrollable factors is also an important task. For example, in materials development, a researcher controls experimental parameters to create materials with the desired physical properties. However, uncontrollable experimental conditions and errors influence the resulting materials. In such cases, one typical formulation is optimizing expectations over uncontrollable factors by assuming a *risk-neutral* attitude of the learner (Toscano-Palmerin and Frazier, 2018; Kirschner et al., 2020). Another well-studied formulation is the *risk-averse* setting, in which some risk measures, e.g., mean-variance measures (Iwazaki et al., 2021b; Makarova et al., 2021), are optimized.

In contrast, some real-world applications seek one single best reward in the presence of uncontrollable factors within limited budgets. In the above example of materials development, it may suffice for the researcher to obtain the desired material only once for scientific discovery. Another example is hyperparameter tuning with stochastic algorithms, in which we only pursue a model with the smallest validation error under uncontrollable factors, such as dropout, stochastic gradient descent, and random initializations. To achieve this goal, the learner should query the input that can potentially yield a high reward, even if most of the obtained rewards are low. Therefore, existing BO methods are unsuitable for these problems; thus, developing a BO algorithm for a *risk-seeking* setting is desired.

A key consideration in the above examples is how to model the rewards and the uncontrollable factors. The uncontrollable experimental conditions in materials development can often be incorporated as input into a surrogate model, while the randomness in the stochastic learning algorithms is difficult to incorpo-

Proceedings of the 27$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

rate as input into a model. Thus, we formulate the two types of BO problems under uncertainty: environmental model setting (Kirschner et al., 2020; Iwazaki et al., 2021b; Inatsu et al., 2022) and heteroscedastic model setting (Kirschner and Krause, 2018; Makarova et al., 2021). In the environmental model setting, the learner models reward generation by using a function of the form: $f(\boldsymbol{x}, \boldsymbol{W})$, where $\boldsymbol{x}$ and $\boldsymbol{W}$ are called *controllable* and *uncontrollable* variables, respectively. In the heteroscedastic model setting, the learner assumes the reward is generated from an unknown distribution, with mean $f(\boldsymbol{x})$, which can be heterogeneous over an input space.

Our Contributions We study risk-seeking BO problems in the environmental and heteroscedastic model settings. In both settings, we propose novel kernel-explore-then-commit (ETC) algorithms. We analyze the performance of kernel-ETC through *extreme regret* (the precise definitions are in Sec. 2 and Sec. 3) and prove that the convergence of the extreme regret is guaranteed in kernel-ETC. Finally, we demonstrate the effectiveness of kernel-ETC via numerical experiments, including the risk-seeking BO problems of a polymer synthesis simulation function and a hyperparameter tuning task, which are derived from real-world data.

Related Works Various strategies in standard BO settings have been extensively studied in the past few decades (Močkus, 1975; Srinivas et al., 2010; Wang and Jegelka, 2017). Moreover, a lot of extended settings are considered, such as parallel (Desautels et al., 2014), constrained (Gardner et al., 2014), high-dimensional (Kandasamy et al., 2015), and multi-fidelity optimization (Kandasamy et al., 2019).

BO problems with uncontrollable environmental variables, which are related to our environmental model setting in Sec. 2, are extensively studied. One standard formulation is to optimize the risk-neutral expected function of environmental variables (Toscano-Palmerin and Frazier, 2018). Furthermore, Kirschner et al. (2020) considers optimizing the distributionally robust variants of the expected function. Other works focus on risk-averse settings that seek to optimize some risk measures, which are designed to avoid the uncertainty of environmental variables. For example, value-at-risk, conditional value-at-risk, and mean-variance risk measures are considered in Nguyen et al. (2021b), Nguyen et al. (2021a), and Iwazaki et al. (2021b), respectively. Based on the assumption for environmental variables in the optimization phase, we can categorize the settings of aforementioned works into two settings (Inatsu et al., 2022). The first setting is the *simulator-based setting*, which assumes the environmental variables are controllable in the optimization phase and become uncontrollable after deploying the identified controllable input during the optimization phase in the real system. The second setting is *uncontrollable setting*, which assumes the environmental variables are uncontrollable in the optimization phase. Note that our work considers uncontrollable settings and is fundamentally different from existing works that only consider simulator-based settings, such as the works of Toscano-Palmerin and Frazier (2018); Bogunovic et al. (2018); Iwazaki et al. (2021a); Nguyen et al. (2021b,a).

Another related setting is heteroscedastic BO problems, whose formulation is similar to our heteroscedastic model setting in Sec. 3. Kirschner and Krause (2018) give the theoretical analysis of the heteroscedastic BO problems under the risk-neutral formulation, which focuses on the expected function of the heterogeneous noise. Makarova et al. (2021) consider the risk-averse formulation by considering the mean-variance objective function in the heteroscedastic noise model. In particular, the algorithm of the exploration phase in our kernel-ETC algorithm in Sec. 3 can be interpreted as the special case of the algorithm of Makarova et al. (2021), which adaptively determines the weight of mean-variance objectives.

Our work can be interpreted as the kernelized extension of *max K-armed bandit problem* (sometimes referred to as *extreme bandits*) in the multi-armed bandits field (Cicirello and Smith, 2005; Carpentier and Valko, 2014). In particular, Achab et al. (2017); Baudry et al. (2022) consider the ETC-based algorithms in the finite-armed setting; some parts of our analysis are inspired by their proofs. However, to extend the finite-armed problem to infinite action space and correlated rewards, many non-trivial treatments are required in the algorithm design and theoretical analysis.

2 ENVIRONMENTAL MODEL SETTING

Problem Setup Let $f : \mathcal{X} \times \mathcal{W} \to \mathbb{R}$ be an unknown function whose input domain is defined as the product of a controllable parameter set $\mathcal{X} \subset \mathbb{R}^{d_1}$ and an uncontrollable parameter set $\mathcal{W} \subset \mathbb{R}^{d_2}$. We also assume that $\mathcal{W}$ is a finite set, on which a *known* probability mass function $p(\boldsymbol{w})$ is defined such that $p(\boldsymbol{w}) > 0$ for all $\boldsymbol{w} \in \mathcal{W}$. At each step t, a learner chooses the controllable parameter $\boldsymbol{x}_t \in \mathcal{X}$, whereas the environment provides the uncontrollable parameter $\boldsymbol{w}_t \sim p(\boldsymbol{w})$. Then, the learner obtains the noisy observation $y_t = f(\boldsymbol{x}_t, \boldsymbol{w}_t) + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. We further assume that the noises $(\epsilon_t)_{t \in \mathbb{N}_+}$ and the uncontrollable parameters $(\boldsymbol{w}_t)_{t \in \mathbb{N}_+}$ are independent.

As the regularity assumptions, we assume that f lies on some known reproducing kernel Hilbert space (RKHS). Let $k : (\mathcal{X} \times \mathcal{W}) \times (\mathcal{X} \times \mathcal{W}) \to \mathbb{R}$ and $\mathcal{H}(k)$ be a positive-definite kernel with $\forall(\boldsymbol{x}, \boldsymbol{w}) \in (\mathcal{X} \times \mathcal{W}), k((\boldsymbol{x}, \boldsymbol{w}), (\boldsymbol{x}, \boldsymbol{w})) \leq 1$ and its corresponding RKHS, respectively. We assume that f is an element of $\mathcal{H}(k)$ and has the bounded RKHS norm $\|f\|_{\mathcal{H}(k)} \leq B$.

Learner's Goal and Regret Under the uncertainty of the uncontrollable parameter $\boldsymbol{w}$, the learner's goal is to observe the value of $f(\boldsymbol{x}, \boldsymbol{w})$ that is as high as possible within a *known* total step size T, which is specified *a priori* by the learner. As the analogous to the finite-armed extreme bandit literature (Carpentier and Valko, 2014), we define the following *extreme regret* $\Delta(T)$ as the performance metric:

$$\Delta(T) = \mathbb{E}\left[\max_{t\in[T]} f(\boldsymbol{x}^*, \boldsymbol{w}_t)\right] - \mathbb{E}\left[\max_{t\in[T]} f(\boldsymbol{x}_t, \boldsymbol{w}_t)\right],$$

where $\boldsymbol{x}^* \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}}\mathbb{E}[\max_{t\in[T]} f(\boldsymbol{x}, \boldsymbol{w}_t)]$ and $[T] = \{1, \ldots, T\}$. It should be noted that the expectations in the above definition are taken with respect to the randomness of $(\boldsymbol{w}_t)_{t\in[T]}$ and $(\boldsymbol{\epsilon}_t)_{t\in[T]}$.

Failures of Existing Methods To minimize the extreme regret, an algorithm must focus on querying the input whose objective function value is high, regardless of the value of $p(\boldsymbol{w})$. Thus, existing works that focus on maximizing other measures (such as expected function) generally do not lead to the minimization of $\Delta(T)$. Figure 1 shows an illustrative example.

Gaussian Process Modeling Our algorithm uses the modeling information of Gaussian process (GP). We assume $\mathcal{GP}(0, k)$ as the prior of f, where $\mathcal{GP}(0, k)$ denotes the zero-mean GP defined by the kernel k. At each step t, given the data $\{(\boldsymbol{x}_i, \boldsymbol{w}_i), y_i\}_{i\in[t]}$ obtained by the learner, the posterior distribution of $f(\boldsymbol{x}, \boldsymbol{w})$ becomes a Gaussian distribution whose mean $\mu_t(\boldsymbol{x}, \boldsymbol{w})$ and variance $\sigma_t^2(\boldsymbol{x}, \boldsymbol{w})$ are given respectively as

$$\begin{aligned}\mu_t(\boldsymbol{x}, \boldsymbol{w}) &= \boldsymbol{k}_t(\boldsymbol{x}, \boldsymbol{w})^\top (\boldsymbol{K}_t + \sigma^2 \boldsymbol{I}_t)^{-1} \boldsymbol{y}_t, \\ \sigma_t^2(\boldsymbol{x}, \boldsymbol{w}) &= k((\boldsymbol{x}, \boldsymbol{w}), (\boldsymbol{x}, \boldsymbol{w})) \\ &\quad - \boldsymbol{k}_t(\boldsymbol{x}, \boldsymbol{w})^\top (\boldsymbol{K}_t + \sigma^2 \boldsymbol{I}_t)^{-1} \boldsymbol{k}_t(\boldsymbol{x}, \boldsymbol{w}),\end{aligned}$$

where $\boldsymbol{y}_t = (y_1, \ldots, y_t)^\top$ and $\boldsymbol{k}_t(\boldsymbol{x}, \boldsymbol{w})$ is a t-dimensional vector whose i-th element is $k((\boldsymbol{x}, \boldsymbol{w}), (\boldsymbol{x}_i, \boldsymbol{w}_i))$. Furthermore, $\boldsymbol{K}_t$ is a $t \times t$ kernel matrix whose (i, j)-th element is $k((\boldsymbol{x}_i, \boldsymbol{w}_i), (\boldsymbol{x}_j, \boldsymbol{w}_j))$.

We further define the following quantity $\gamma(t)$, which is called the *maximum information gain* (MIG):

$$\gamma(t) = \frac{1}{2} \max_{(\boldsymbol{x}_1, \boldsymbol{w}_1), \ldots, (\boldsymbol{x}_t, \boldsymbol{w}_t)} \ln \det(\boldsymbol{I}_t + \sigma^{-2} \boldsymbol{K}_t).$$

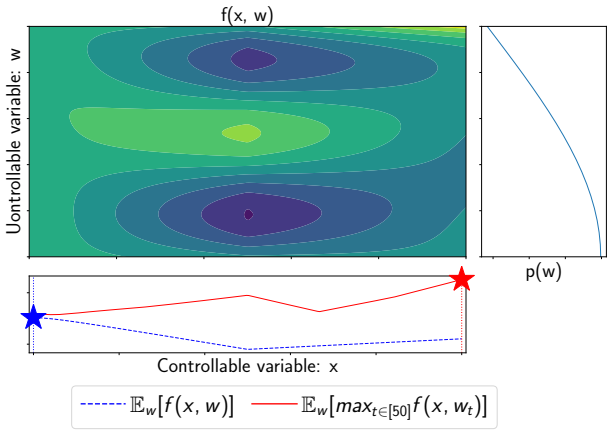

Figure 1: A two-dimensional example problem illustrating that existing methods for seeking the maximum of expected function may not work to minimize $\Delta(T)$ with $T = 50$. The red line and star represent the value of $\mathbb{E}[\max_{t\in[50]} f(\boldsymbol{x}, \boldsymbol{w}_t)]$ and $\boldsymbol{x}^*$, respectively. In this problem, to minimize $\Delta(T)$, we need to focus on querying in the right-hand side area that rarely yields the highest rewards. However, the algorithms that seek the maximum of the expected function (blue star) focus on querying in the left-hand side area.

The quantity $\gamma(t)$ is often used to characterize the confidence bounds or the regret in the standard BO literature (Srinivas et al., 2010; Chowdhury and Gopalan, 2017). In this paper, we leverage the following Lemma 2.1 to construct confidence bounds using $\gamma(t)$.

Lemma 2.1 (Theorem 3.11 in Abbasi-Yadkori (2013))**.** *Fix $f \in \mathcal{H}(k)$ with $\|f\|_{\mathcal{H}(k)} \leq B$ and $\delta \in (0, 1)$. Let us assume that the noise term ϵ_t independently follows $\mathcal{N}(0, \sigma^2)$. Then, with probability at least $1-\delta$, the following inequality holds for any $t \in \mathbb{N}_+$ and $(\boldsymbol{x}, \boldsymbol{w}) \in \mathcal{X} \times \mathcal{W}$:*

$$\begin{aligned}&|f(\boldsymbol{x}, \boldsymbol{w}) - \mu_{t-1}(\boldsymbol{x}, \boldsymbol{w})| \\ &\leq (B + \sqrt{2(\gamma(t) + \ln \delta^{-1})})\sigma_{t-1}(\boldsymbol{x}, \boldsymbol{w}).\end{aligned}$$

Proposed Algorithm Our proposed algorithm is based on an ETC strategy. The ETC algorithm purely explores the search space $\mathcal{X}$ until some fixed amount of step size $\tilde{T} \leq T$ is reached. After this exploration period, the ETC algorithm exploits the knowledge collected so far to obtain high values of $f(\boldsymbol{x}_t, \boldsymbol{w}_t)$. Our main challenge is devising how to explore the vast $\mathcal{X}$ by leveraging the kernel-based smoothness assumption.

Algorithm 1 is a pseudo-code of our proposed algorithm: kernel-ETC. Here, let $\hat{\boldsymbol{x}}^*$ be the point that is chosen after the exploration period. We design the exploration and exploitation strategy based on the fol-

Algorithm 1 The kernel-ETC algorithm for environmental model setting.

Input: GP prior $\mathcal{GP}(0,\ k)$, exploration ratio $\alpha \in (0, 1]$, width of confidence bound $\{\beta(t)\}_{t\in\mathbb{N}_+}$.

$\tilde{T} \leftarrow \lceil \alpha(T-1) \rceil$.
for $t = 1$ to $\tilde{T}$ **do**
  $\boldsymbol{x}_t \leftarrow \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{j\in[T]} \mathrm{ucb}_t(\boldsymbol{x}, W_j)\right]$.
  Observe y_t and update GP posterior.
end for
$\tilde{t} \leftarrow \max_{t\in[\tilde{T}]} \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{j\in[T]} \mathrm{lcb}_t(\boldsymbol{x}_t, W_j)\right]$.
$\hat{\boldsymbol{x}}^* \leftarrow \boldsymbol{x}_{\tilde{t}}$.
for $t = \tilde{T}+1$ to T **do**
  Set $\boldsymbol{x}_t$ as $\boldsymbol{x}_t = \hat{\boldsymbol{x}}^*$ and observe y_t.
end for

lowing decomposition of the upper bound of $\Delta(T)$:

$$\Delta(T) \le \mathbb{E}[\Delta_1(T)] + \mathbb{E}[\Delta_2(T)], \tag{1}$$

where $\Delta_1(T)$ and $\Delta_2(T)$ are defined as follows:

$$\Delta_1(T) = \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{t\in[T]} f(\boldsymbol{x}^*, W_t) - \max_{t\in[T]} f(\hat{\boldsymbol{x}}^*, W_t) \right],$$

$$\Delta_2(T) = \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{t\in[T]} f(\hat{\boldsymbol{x}}^*, W_t) - \max_{t\in[T]\setminus[\tilde{T}]} f(\hat{\boldsymbol{x}}^*, W_t) \right].$$

The proof of Eq. (1) is in Lemma A.1 in Appendix A. In Eq. (1), $\tilde{\mathbb{E}}_{\boldsymbol{W}}[\cdot]$ is the expectation operator taken for the independent random variables $\boldsymbol{W} := (W_1, \dots W_T)^\top$, where $W_1, \dots, W_T \sim_{i.i.d.} p(\boldsymbol{w})$. Namely, $\tilde{\mathbb{E}}_{\boldsymbol{W}}[g(\boldsymbol{W})] := \sum_{(\boldsymbol{w}^{(1)},\dots,\boldsymbol{w}^{(T)})\in\mathcal{W}^T} g(\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(T)}) \prod_{t=1}^T p(\boldsymbol{w}^{(t)})$ for any (measurable) function g. As shown in Lemma A.2 in Appendix A, the second term converges to zero when $\tilde{T}$ is chosen as $\Theta(T)$. Thus, intuitively, $\boldsymbol{x}_t$ and $\hat{\boldsymbol{x}}^*$ should be designed to maximize $\tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{t\in[T]} f(\hat{\boldsymbol{x}}^*, W_t)\right]$ such that the remaining first term becomes small. From this insight, we adopt the GP upper confidence bound (GP-UCB)-based strategy against $\tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{t\in[T]} f(\boldsymbol{x}, W_t)\right]$ to choose $\boldsymbol{x}_t$ in the exploration period $t \le \tilde{T}$:

$$\boldsymbol{x}_t \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{t\in[T]} \mathrm{ucb}_t(\boldsymbol{x}, W_t)\right], \tag{2}$$

where ucb_t is defined as $\mathrm{ucb}_t(\boldsymbol{x}, \boldsymbol{w}) = \mu_{t-1}(\boldsymbol{x}, \boldsymbol{w}) + \beta^{1/2}(t)\sigma_{t-1}(\boldsymbol{x}, \boldsymbol{w})$ with the parameter $\beta(t)$, which specifies the width of confidence bounds. In addition, $\hat{\boldsymbol{x}}^*$ is defined based on the lower confidence bound of $\tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{t\in[T]} f(\boldsymbol{x}, W_t)\right]$, which is common strategy for GP-UCB-based algorithms (Bogunovic et al., 2018; Kirschner et al., 2020; Iwazaki et al., 2021b):

$$\hat{\boldsymbol{x}}^* = \boldsymbol{x}_{\tilde{t}} \text{ where } \tilde{t} \in \arg\max_{t\in[\tilde{T}]} \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{j\in[T]} \mathrm{lcb}_t(\boldsymbol{x}_t, W_j)\right],$$

where $\mathrm{lcb}_t(\boldsymbol{x}, \boldsymbol{w}) = \mu_{t-1}(\boldsymbol{x}, \boldsymbol{w}) - \beta^{1/2}(t)\sigma_{t-1}(\boldsymbol{x}, \boldsymbol{w})$.

The computations of $\boldsymbol{x}_t$ and $\hat{\boldsymbol{x}}^*$ require the expectation of the maximum of T independent random variables, which is analytically solved because of the finiteness assumption of $\mathcal{W}$ (see Appendix D.1 for details).

Theoretical Analysis The following theorem gives the upper bound of $\Delta(T)$ for the environmental model setting. Full proof is given in Appendix A.

Theorem 2.1. *Fix $f \in \mathcal{H}(k)$ with $\|f\|_{\mathcal{H}(k)} \le B$. When running Algorithm 1 with $\beta^{1/2}(t) = B + \sqrt{2(\gamma(t) + \ln(2T))}$ and $\alpha \in (0, 1]$, the following upper bound of the extreme regret $\Delta(T)$ holds:*

$$\Delta(T) \le 2B(1-\underline{p})^{(1-\alpha)T} + \frac{2B}{T} + 2C\sqrt{\frac{\beta(\tilde{T})(Q-1)\ln T}{\tilde{T}\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8\ln(12T)\right\}},$$

where $\tilde{T} = \lceil \alpha(T-1) \rceil$, $\underline{p} = \min_{\boldsymbol{w}\in\mathcal{W}} p(\boldsymbol{w})$, $Q = 2\underline{p}^{-1}$, and $C > 0$ is an absolute constant.

The first term $2B(1-\underline{p})^{(1-\alpha)T}$ comes from the second term of Eq. (1), and the rest comes from the first term of Eq. (1). When we fix α, $p(\cdot)$, and σ^{-2}, the first term $2B(1-\underline{p})^{(1-\alpha)T}$ decays exponentially as T increases and is ignorable compared with the third term. The second term is also ignorable compared with the third term. Thus, the overall order of the regret is dominated by the third term. Then, the order of the regret bound is $\mathcal{O}(\gamma(T)(\ln T)/\sqrt{T})$. In our bound, there is the additional logarithmic dependence which does not appear in the simple regret bound of the standard GP-UCB algorithm (Chowdhury and Gopalan, 2017); however, the no-regret guarantees hold for commonly used kernels such as the squared exponential (SE) kernel: $k_{\mathrm{SE}}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) := \exp(-\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2^2/(2\ell^2))$. Actually, since $\gamma(T) = \mathcal{O}((\ln T)^{d_1+d_2})$ holds in the SE kernel, $\Delta(T)$ becomes $\mathcal{O}((\ln T)^{d_1+d_2+1}/\sqrt{T}) \to 0$ (as $T \to \infty$).

Proof Sketch As previously mentioned, the second term of Eq. (1) converges to zero (Lemma A.2). The remaining interest is the first term $\mathbb{E}[\Delta_1(T)]$ of Eq. (1) here. To bound $\Delta_1(T)$, we first derive the following high probability upper bound by resorting to similar arguments of the proof of Srinivas et al. (2010):

$$\Delta_1(T) \le \frac{2\beta^{1/2}(\tilde{T})}{\tilde{T}} \sum_{t=1}^{\tilde{T}} \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{j\in[T]} \sigma_{t-1}(\boldsymbol{x}_t, W_j)\right]. \tag{3}$$

To use the well-known result of Srinivas et al. (2010), which describes the upper bound of $\sum_{t=1}^{\tilde{T}} \sigma_{t-1}(\boldsymbol{x}_t, \boldsymbol{w}_t)$ by using the MIG (Lemma A.3 in Appendix A), we next leverage the following Lemma 2.2.

Lemma 2.2 (Exercise 2.5.10 in Vershynin (2018))**.** *Let $X_1, \dots, X_n$ be independent and identically distributed random variables with $n \geq 2$. Suppose $\|X_1\|_{\psi_2} < \infty$, where $\|X_1\|_{\psi_2}$ is defined as $\|X_1\|_{\psi_2} = \inf\{a > 0 \mid \mathbb{E}[\exp(X_1^2/a^2)] \leq 2\}$.*

Then, there exist an absolute constant $C > 0$, and the inequality: $\mathbb{E}[\max_{i\in[n]} |X_i|] \leq C\|X_1\|_{\psi_2}\sqrt{\ln n}$ holds.

The quantity $\|X\|_{\psi_2}$ is called the sub-Gaussian norm of random variable X (see, e.g., Chapter 2.5 in Vershynin (2018)). Roughly speaking, our idea is to analyze the upper bound of $\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} \sigma_{t-1}(\boldsymbol{x}_j, W_t)\right]$ via the sub-Gaussian norm $\|\sigma_{t-1}(\boldsymbol{x}_j, W_t)\|_{\psi_2}$, instead of directly analyzing it. By applying Lemma 2.2, $\sum_{t=1}^{\tilde{T}} \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} \sigma_{t-1}(\boldsymbol{x}_t, W_t)\right] \leq \sqrt{\ln T}\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_j, W_t)\|_{\psi_2}$ holds. Finally, we bound $\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_j, W_t)\|_{\psi_2}$ from above by using the inequality about the MIG (Lemma A.5 in Appendix A) and obtain the upper bound of Eq. (3).

Discussions The drawback of our analysis is that an undesirable dependence on $|\mathcal{W}|$ implicitly appeared in our regret upper bound. From the definition of $\underline{p}$, $\underline{p} \leq 1/|\mathcal{W}|$ holds; thus $Q = \mathcal{O}(|\mathcal{W}|)$, which leads to our regret bound of $\mathcal{O}(\sqrt{|\mathcal{W}|\gamma_T \ln T/T})$. This indicates that our algorithm does not guarantee to work on the regime of $\sqrt{T} \ll |\mathcal{W}|$. However, note that the same dependence of $\overline{p}^{-1}$ also appears in the existing analysis of the uncontrollable setting (Inatsu et al., 2022). We leave the additional analysis to study whether the $\mathcal{O}(\sqrt{|\mathcal{W}|})$ dependence is avoidable as future work.

3 HETEROSCEDASTIC MODEL SETTING

Problem setup Let $f : \mathcal{X} \to \mathbb{R}$ be an unknown function whose input domain $\mathcal{X} \subset \mathbb{R}^d$ is a compact and convex set. At each step t, the learner chooses $\boldsymbol{x}_t \in \mathcal{X}$ and obtain a reward $y_t := f(\boldsymbol{x}_t) + \eta_t(\boldsymbol{x}_t)$. The random variable $\eta_t(\boldsymbol{x})$ is an additional stochastic term whose variance depends on the input. We assume that the random variables $(\eta_t(\boldsymbol{x}))_{t\in\mathbb{N}_+,\boldsymbol{x}\in\mathcal{X}}$ are independent. Furthermore, $\eta_t(\boldsymbol{x})$ follows a zero-mean Gaussian distribution whose variance is given as $\rho^2(\boldsymbol{x})$, where $\rho : \mathcal{X} \to [0, \infty)$ is an unknown function. It should be noted that the minimization of extreme regret without specifying the class of reward distributions is infeasible (Streeter and Smith, 2006). Thus, we assume Gaussian rewards commonly used in the output model of BO (e.g., Sec. 5 in Frazier (2018)).

The learner's goal is to obtain a reward that is as high as possible within a known total step size T. The extreme regret $\Delta(T)$ in this setting is defined as follows:

$$\Delta(T) = \mathbb{E}\left[\max_{t\in[T]} \{f(\boldsymbol{x}^*) + \eta_t(\boldsymbol{x}^*)\}\right] - \mathbb{E}\left[\max_{t\in[T]} y_t\right],$$

where $\boldsymbol{x}^* \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}}\mathbb{E}[\max_{t\in[T]} \{f(\boldsymbol{x}) + \eta_t(\boldsymbol{x})\}]$.

Regularity Assumptions As with the regularity assumptions in Sec. 2, we assume that f is an element of RKHS with $\|f\|_{\mathcal{H}(k_f)} \leq B_f < \infty$, where $k_f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a known positive definite kernel such that $k_f(\boldsymbol{x}, \boldsymbol{x}) \leq 1$ holds for all $\boldsymbol{x} \in \mathcal{X}$. To efficiently estimate ρ, we assume that the function ρ lies on some RKHS $\mathcal{H}(k_\rho)$ with $\|\rho\|_{\mathcal{H}(k_\rho)} \leq B_\rho < \infty$. Here, $k_\rho : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a known positive definite kernel such that $k_\rho(\boldsymbol{x}, \boldsymbol{x}) \leq 1$ holds for all $\boldsymbol{x} \in \mathcal{X}$, which can be different from k_f. Moreover, we assume that the range of ρ is bounded with known constants: $\underline{\rho}$, $\overline{\rho}$. Namely, $\rho(\boldsymbol{x}) \in [\underline{\rho}, \overline{\rho}]$ holds for all $\boldsymbol{x} \in \mathcal{X}$. It should be noted that similar boundness assumptions were also used in previous heteroscedastic BO literature (Makarova et al., 2021).

Estimation of $\rho(\cdot)$ As used in Makarova et al. (2021), we adopt the *repeated experiment strategy* to make the confidence bound of $\rho(\cdot)$ during the exploration period of our ETC-based algorithm. Namely, the learner selects an input at once and then repeatedly queries the same input in the subsequent m steps. We call this one block of the same m query points a *batch*. At the end of each i-th batch, Makarova et al. (2021) proposed to leverage the unbiased estimator $\hat{m}^{(i)} := \sum_{l=1}^{m}(y_l^{(i)} - \hat{y}^{(i)})^2/(m-1)$ of the variance $\rho^2(\boldsymbol{x})$, where $y_l^{(i)}$ is the l-th reward obtained within batch i, and $\hat{y}^{(i)} := \sum_{l=1}^{m} y_l^{(i)}/m$ is the sample mean of rewards within batch i. Hereafter, we let $\boldsymbol{x}^{(i)}$ be the query point chosen by the learner and $\tilde{\zeta}^{(i)}$ be the error $\tilde{\zeta}^{(i)} := \rho^2(\boldsymbol{x}^{(i)}) - \hat{m}^{(i)}$ of $\hat{m}^{(i)}$. If we assume that the error $\tilde{\zeta}^{(i)}$ is sub-Gaussian, the confidence bound of $\rho^2(\cdot)$ can be obtained through a GP model, which leverages $\hat{m}^{(i)}$ as the outputs of training data (Makarova et al., 2021). However, the sub-Gaussian assumption of $\tilde{\zeta}^{(i)}$ is not valid under the assumption that η_t is Gaussian (Lemma B.10 in Appendix B). To avoid this problem, we use the fact that the square root of $\hat{m}^{(i)}$ has sub-Gaussian property even though $\hat{m}^{(i)}$ and $\tilde{\zeta}^{(i)}$ themselves do not. Namely, we construct a GP-model of $\rho(\cdot)$ instead of $\rho^2(\cdot)$ by plugging the following unbiased estimator $\hat{s}^{(i)}$ of $\rho(\boldsymbol{x}^{(i)})$ into the training outputs:

$$\hat{s}^{(i)} = \left\{\sqrt{\frac{2}{m-1}}\frac{\Gamma(m/2)}{\Gamma((m-1)/2)}\right\}^{-1}\sqrt{\hat{m}^{(i)}},$$

where $\Gamma(\cdot)$ is a gamma function. The posterior mean $\mu_\rho^{(i)}(\boldsymbol{x})$ and variance $\sigma_\rho^{(i)2}(\boldsymbol{x})$ of the constructed GP-

model of $\rho(\boldsymbol{x})$ at the end of batch i are defined as

$$\mu_\rho^{(i)}(\boldsymbol{x}) = \boldsymbol{k}_\rho^{(i)}(\boldsymbol{x})^\top (\boldsymbol{K}_\rho^{(i)} + \lambda_\rho^2 \boldsymbol{I}_i)^{-1} \hat{\boldsymbol{s}}^{(i)},$$
$$\sigma_\rho^{(i)2}(\boldsymbol{x}) = k_\rho(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}_\rho^{(i)}(\boldsymbol{x})^\top (\boldsymbol{K}_\rho^{(i)} + \lambda_\rho^2 \boldsymbol{I}_i)^{-1} \boldsymbol{k}_\rho^{(i)}(\boldsymbol{x}),$$

where $\hat{\boldsymbol{s}}^{(i)} = (\hat{s}^{(1)}, \ldots, \hat{s}^{(i)})^\top$ and $\boldsymbol{k}_\rho^{(i)}(\boldsymbol{x})$ is the i- dimensional vector whose j-th elemnt is $k_\rho(\boldsymbol{x}, \boldsymbol{x}^{(j)})$. Furthermore, $\boldsymbol{K}_\rho^{(i)}$ is a $i \times i$ kernel matrix whose (j, l)-th element is $k_\rho(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(l)})$, and $\lambda_\rho > 0$ is a pre-specified noise variance parameter.

Estimation of $f(\cdot)$ To efficiently estimate $f(\cdot)$ under the unknown $\rho(\cdot)$ without breaking the theoretical guarantee, we leverage the UCB-based estimation of the noise level to construct a GP model of $f(\cdot)$ as in Makarova et al. (2021). Let $\hat{\boldsymbol{\Sigma}}^{(i)} \in \mathbb{R}^{i \times i}$ be a UCB-based noise matrix, which is defined as $\hat{\boldsymbol{\Sigma}}^{(i)} = \mathrm{diag}\left(\overline{\rho}(\boldsymbol{x}^{(1)})^2, \ldots, \overline{\rho}(\boldsymbol{x}^{(i)})^2\right)/m$, where $\overline{\rho}(\boldsymbol{x}) = \max\{\min\{\overline{\rho}, \mathrm{ucb}_\rho^{(i)}(\boldsymbol{x})\}, \underline{\rho}\}$ and $\mathrm{ucb}_\rho^{(i)}(\boldsymbol{x}) = \mu_\rho^{(i-1)}(\boldsymbol{x}) + \beta_\rho^{1/2}(i)\sigma_\rho^{(i-1)}(\boldsymbol{x})$ with parameter $\beta_\rho(i) > 0$. By using $\hat{\boldsymbol{\Sigma}}^{(i)}$, the posterior mean $\mu_f^{(i)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i)})$ and variance $\sigma_f^{(i)2}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i)})$ of f at the end of batch i are respectively defined as follows:

$$\mu_f^{(i)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i)}) = \boldsymbol{k}_f^{(i)}(\boldsymbol{x})^\top (\boldsymbol{K}_f^{(i)} + \hat{\boldsymbol{\Sigma}}^{(i)})^{-1} \hat{\boldsymbol{y}}^{(i)},$$
$$\sigma_f^{(i)2}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i)}) = k_f(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}_f^{(i)}(\boldsymbol{x})^\top (\boldsymbol{K}_f^{(i)} + \hat{\boldsymbol{\Sigma}}^{(i)})^{-1} \boldsymbol{k}_f^{(i)}(\boldsymbol{x}),$$

where $\hat{\boldsymbol{y}}^{(i)} = (\hat{y}^{(1)}, \ldots, \hat{y}^{(i)})^\top$. Furthermore, $\boldsymbol{k}_f^{(i)}(\boldsymbol{x})$ and $\boldsymbol{K}_f^{(i)}$ are defined by replacing $\boldsymbol{k}_\rho^{(i)}(\boldsymbol{x})$ and $\boldsymbol{K}_\rho^{(i)}$ of k_ρ with k_f.

Proposed Algorithm We propose a kernel-ETC algorithm in the heteroscedastic model setting, and Algorithm B.2 in Appendix B shows the pseudocode. In the exploration period, our algorithm is designed to find the maximizer of the quantity $f(\boldsymbol{x}) + \theta_T \rho(\boldsymbol{x})$ based on Lemma B.4 in Appendix B, which describes the decomposition of the regret similar to Eq. (1). Here, $\theta_T \coloneqq \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t \in [T]} Z_t\right]$ denotes the expected maximum of the T independent standard Gaussian random variables. The expectation operator $\tilde{\mathbb{E}}_{\boldsymbol{Z}}[\cdot]$ is defined as $\tilde{\mathbb{E}}_{\boldsymbol{Z}}[g(\boldsymbol{Z})] = \int g(z^{(1)}, \ldots, z^{(T)}) \prod_{t=1}^T \phi(z^{(t)}) \mathrm{d}z^{(1)} \ldots \mathrm{d}z^{(T)}$ for any measurable function $g : \mathbb{R}^T \to \mathbb{R}$. The function $\phi(\cdot)$ is the probability density function of $\mathcal{N}(0, 1)$.

To find the maximizer of $f(\boldsymbol{x}) + \theta_T \rho(\boldsymbol{x})$, we choose the query point $\boldsymbol{x}^{(i)}$ of batch i as $\boldsymbol{x}^{(i)} = \mathrm{argmax}_{\boldsymbol{x} \in \mathcal{X}}[\mathrm{ucb}_f^{(i)}(\boldsymbol{x}) + \theta_T \mathrm{ucb}_\rho^{(i)}(\boldsymbol{x})]$, where $\mathrm{ucb}_f^{(i)}(\boldsymbol{x}) = \mu^{(i-1)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)}) + \beta_f^{1/2}(i)\sigma_f^{(i-1)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)})$ with a pre-specified parameter $\beta_f(i) > 0$. We also define the query point $\hat{\boldsymbol{x}}^*$ of the exploitation period based on the LCB similar to Sec. 2. The definition of $\hat{\boldsymbol{x}}^*$ is $\hat{\boldsymbol{x}}^* = \boldsymbol{x}^{(\tilde{i})}$ with $\tilde{i} = \mathrm{argmax}_{i \in [M]}[\mathrm{lcb}_f^{(i)}(\boldsymbol{x}) + \theta_T \mathrm{lcb}_\rho^{(i)}(\boldsymbol{x})]$, where $M \coloneqq \lfloor \tilde{T}/m \rfloor$ and $\tilde{T} \coloneqq \lceil \alpha(T-1) \rceil$ are the number of batches and step size in the exploration period, respectively. Furthermore, α is the exploration ratio. Here, we set $\mathrm{lcb}_f^{(i)}(\boldsymbol{x}) = \mu_f^{(i-1)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)}) - \beta_f^{1/2}(i)\sigma_f^{(i-1)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)})$ and $\mathrm{lcb}_\rho^{(i)}(\boldsymbol{x}) = \mu_\rho^{(i-1)}(\boldsymbol{x}) - \beta_\rho^{1/2}(i)\sigma_\rho^{(i-1)}(\boldsymbol{x})$.

Theoretical Analysis The following Theorem 3.1 gives the regret upper bound of the kernel-ETC algorithm. Full proofs are described in Appendix B.

Theorem 3.1. *Fix any $\tau \in (0, 1)$, $m \geq 2$, and $\overline{\rho} \geq \underline{\rho} > 0$. Assume $f \in \mathcal{H}(k_f)$ with $\|f\|_{\mathcal{H}(k_f)} \leq B_f$, $\rho \in \mathcal{H}(k_\rho)$ with $\|\rho\|_{\mathcal{H}(k_\rho)} \leq B_\rho$, and $\forall \boldsymbol{x} \in \mathcal{X}$, $\rho(\boldsymbol{x}) \in [\underline{\rho}, \overline{\rho}]$. When running Algorithm B.2 with $\beta_f^{1/2}(i) = B_f + \sqrt{2(\gamma_f(i) + 1 + \ln 2T)}$, $\beta_\rho^{1/2}(i) = B_\rho + \sqrt{2(\gamma_\rho(i) + \ln 2T)}$, $\lambda_\rho = c\kappa(m)\overline{\rho}$, and $\alpha = T^\tau/T$, the following holds:*

$$\Delta(T) = \mathcal{O}\left(\sqrt{\frac{(\ln T + \gamma_f(T^\tau))\gamma_f(T^\tau)}{T^\tau}} + \frac{T^\tau \sqrt{\ln T}}{T} + \sqrt{\frac{(\ln T + \gamma_\rho(T^\tau))\gamma_\rho(T^\tau) \ln T}{T^\tau}} \right), \tag{4}$$

where $\kappa(m) = (m-1)^{1/4}\Gamma((m-1)/2)/\Gamma(m/2)$, and $c > 0$ is an absolute constant. Furthermore, $\gamma_f(i) \coloneqq 0.5 \max_{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(i)}} \ln\det(\boldsymbol{I}_i + m\overline{\rho}^{-2}\boldsymbol{K}_f^{(i)})$ and $\gamma_\rho(i) \coloneqq 0.5 \max_{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(i)}} \ln\det(\boldsymbol{I}_i + \lambda_\rho^{-2}\boldsymbol{K}_\rho^{(i)})$ are MIGs at the end of batch i, which are derived from the GP models of f and ρ, respectively.

The detailed version of Theorem 3.1 that describes the explicit upper bound without order notation is also given as Theorem B.1 in Appendix B. By substituting γ_f and γ_ρ with the known upper bound of MIGs, Eq. (4) becomes more explicit. For example, when both k_f and k_ρ are SE kernels, the regret upper bound Eq. (4) becomes $\Delta(T) = \mathcal{O}(\sqrt{(\ln T)^{2d+1}/T^\tau} + T^{\tau-1}\sqrt{\ln T})$ and $\Delta(T) \to 0$ (as $T \to \infty$).

One clear difference between Theorem 3.1 and Theorem 2.1 of Sec. 2 is that the exploration ratio α is chosen adaptively with respect to T in Theorem 3.1, whereas Theorem 2.1 assumes α is fixed. Actually, in our analysis, the regret upper bound for fixed α becomes $\Delta(T) = \mathcal{O}(\sqrt{\ln T})$, as shown in Theorem 3.1 that describes the explicit upper bound without order notation is also given as Theorem B.1 in Appendix B. Intuitively, this phenomenon arises from the fact that uncontrollable additional terms η_t have unbounded support in contrast to the setting of Sec. 2. Consequently, to keep no-regret guarantees, we must spend

more exploitation costs than constant fractions of the total step size.

4 MAXIMUM VARIANCE REDUCTION-BASED VARIANTS

Our ETC-based algorithm relies on the confidence bound whose width $\beta(t)$ increases as $\sqrt{\gamma(t)}$, and the resulting regret upper bound does not vanish in several important kernels such as those of the Matérn family. Recently, Vakili et al. (2021) showed that the width of the confidence bound with the fixed confidence level can be improved from $\mathcal{O}(\sqrt{\gamma(t)})$ to $\mathcal{O}(1)$ for *non-adaptive* strategy, whose query points selections are independent of observation noises. Here, we consider the maximum variance reduction (MVR)-based variants of our kernel-ETC algorithm, whose exploration strategy is non-adaptive since the query points only rely on the posterior variances of GP. The pseudo-codes of our MVR-based algorithms are in Appendix C. In the environmental model setting, the following Theorem 4.1 shows the regret upper bound of our MVR-based algorithm, which improves the regret $\mathcal{O}(\gamma(T)(\ln T)/\sqrt{T})$ of Theorem 2.1 to $\mathcal{O}(\sqrt{\gamma(T)}(\ln T)/\sqrt{T})$.

Theorem 4.1. *Fix $f \in \mathcal{H}(k)$ with $\|f\|_{\mathcal{H}(k)} \leq B$. Suppose that $\mathcal{X}$ is a finite set. When running Algorithm C.3 in Appendix C with $\alpha \in (0, 1]$, the following upper bound of the extreme regret $\Delta(T)$ holds:*

$$\Delta(T) \leq 2B(1-\underline{p})^{(1-\alpha)T} + \frac{2B}{T} + 2C\sqrt{\frac{\beta(T)(Q-1)\ln T}{\tilde{T}\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8\ln(12T)\right\}},$$

where $\tilde{T} = \lceil \alpha(T-1) \rceil$ $\underline{p} = \min_{\boldsymbol{w}\in\mathcal{W}} p(\boldsymbol{w})$, $Q = 2\underline{p}^{-1}$, and $\beta^{1/2}(T) = B + \sqrt{2\ln(4|\mathcal{X}||\mathcal{W}|T)}$. Furthermore, C is an absolute constant.

Theorem 4.1 assumes that the input space $\mathcal{X}$ is finite in order to simplify the proof. The extensions to continuous input spaces are easily made by resorting to the discretizing arguments of the input space (Vakili et al., 2021; Li and Scarlett, 2022). We also propose an MVR-based kernel-ETC algorithm in the heteroscedastic model setup. The details and its regret upper bound are in Appendix C.2.

We emphasize that previous works report that UCB-based algorithms work well in practice when the width of the confidence bound is tuned as a hyperparmeter (Srinivas et al., 2010). Furthermore, in our numerical experiments, we found that UCB-based kernel-ETC tends to outperform MVR-based ones; thus, we believe both of them are useful [1].

[1]Whitehouse et al. (2023), which proves the tight regret bound of GP-UCB strategy, appears online as a preprint. We leave extensions based on their results as future work.

5 NUMERICAL EXPERIMENTS

We show the performance of the kernel-ETC in numerical experiments. The details are shown in Appendix E, including the results with standard errors.

We compare our algorithm with `Random`, which explores the search space $\mathcal{X}$ uniformly at random. Furthermore, in the environmental model setting experiments, we adopt MVABO (Iwazaki et al., 2021b), which aims to optimize mean-variance objectives. Specifically, we employ MVABO with parameters set to maximize either the mean or the variance. We denote the versions of MVABO that focus on maximizing the mean or variance as `Mean-MVABO` or `Variance-MVABO`, respectively. In the heteroscedastic model setup, we adopt standard `GP-UCB` with noise parameter $\overline{\rho}^2$ and the risk-averse heteroscedastic BO (RAHBO) algorithm from (Makarova et al., 2021). Similar to MVABO, we consider maximization of the mean (variance) in RAHBO, which is denoted as `Mean-RAHBO` (`Variance-RAHBO`). It should be noted that the original MVABO and RAHBO focus on minimizing the variance; however, to adapt to our risk-seeking setting, we modify the original algorithms to maximize the variance. The details of the modified version of MVABO and RAHBO in our experiments are in Appendix E. Furthermore, we conduct either kernel-ETC or MVR-based kernel-ETC with $\alpha = 0.75$ and $\alpha = 0.95$ ($\tau = 0.75$ and $\tau = 0.95$) in environmental (heteroscedastic) model setting[2].

In all experiments, we use the SE kernel $k_{\mathrm{SE}}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \coloneqq \sigma_{\mathrm{ker}}^2 \exp(-\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2^2/(2\ell^2))$ with fixed lengthscale parameter ℓ and outputscale parameter σ_{ker}. The specified ℓ and σ_{ker} are described in Appendix E.

Synthetic Benchmark Functions We conduct experiments with synthetic functions whose desired inputs of existing methods are different from our risk-seeking formulation. In the environmental model setting, we use the function depicted in Fig. 1, and we also create a 1D-synthetic function in the heteroscedastic model setting. The precise definitions and illustrations of both functions are given in Appendix E.

We conduct experiments with 100 different random seeds and report average extreme regrets. The left-most figures in Fig. 2 and Fig. 3 show the results with the synthetic benchmark functions in the environmental and heteroscedastic model settings, respectively. We confirm the extreme regret decreases as T increases for kernel-ETC, but not for other methods. Specifi-

[2]Additional experiments about the performance sensitivity to the settings of α and τ are in Appendix F.

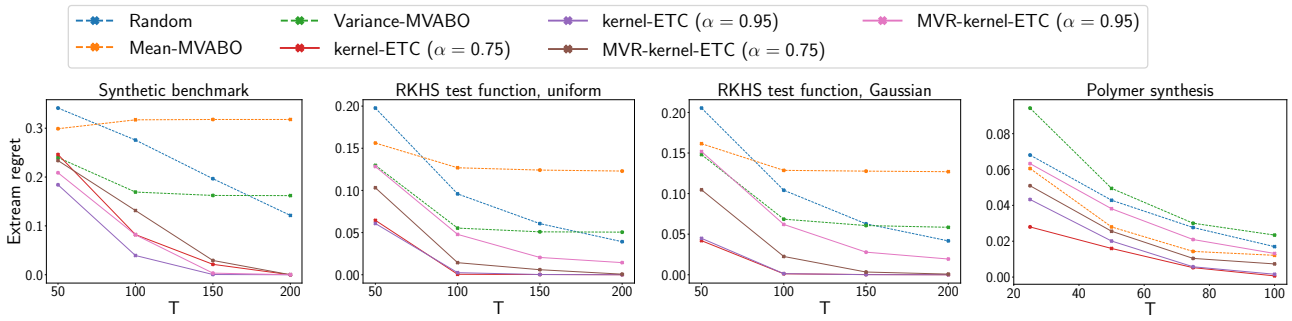

Figure 2: The average extreme regret of the numerical experiments in the environmental model settings.

cally, we find the regrets of existing methods in the heteroscedastic model setting tend to become larger as T increases. This is reasonable since the first term of the extreme regret becomes large at the order of $\mathcal{O}(\sqrt{\ln T})$; thus, the regret of existing methods diverges if their query points are concentrated on sub-optimal points.

2D RKHS Test Functions We further run experiments with 2D RKHS test functions, which are constructed as with the experiments in Chowdhury and Gopalan (2017). In the environmental model setting, the first (second) dimensions of test functions are used as controllable (uncontrollable) inputs. We set $\mathcal{X}$ and $\mathcal{W}$ as 50 and 10 uniformly-spaced grids of $[0, 1]$, respectively. Furthermore, we run experiments with two types of probability mass functions: p_{uniform} and p_{Gaussian}, where $p_{\mathrm{uniform}}(w) = 1/|\mathcal{W}|$ and $p_{\mathrm{Gaussian}}(w) = \phi(w)/\sum_{w \in \mathcal{W}} \phi(w)$. In heteroscedastic model setting, $\mathcal{X}$ is defined as $\mathcal{X} = \tilde{\mathcal{X}} \times \tilde{\mathcal{X}}$, where $\tilde{\mathcal{X}}$ is 25 uniformly spaced grids of $[0, 1]$.

We generate 20 test functions and then run 10 experiments with different initial points in each test function. Namely, we run 200 trials of experiments in total. The second and third figures from left in Fig. 2 show the results in the environmental model setting. The result of the heteroscedastic model setting is shown in the right figure in Fig. 3. We confirm that our methods reduce the regret with increasing T, in contrast to others.

Polymer Synthesis Simulation Function As a potential application of our environmental model setting, we conducted an experiment with a 2D-simulation function, which is created using real data of polymer synthesis (Belabed et al., 2012). In this experiment, the control parameter is the mixing ratio of two polymers. The uncontrollable parameter is the ingredient of the polymer subcomponent, whose uncertainty arises from the manufacturing process. The learner's goal is to obtain the polymer whose glass transition temperature is high. The rightmost figure in Fig. 2 shows the results with 100 different random seeds. We confirm kernel-ETC outperforms the existing methods.

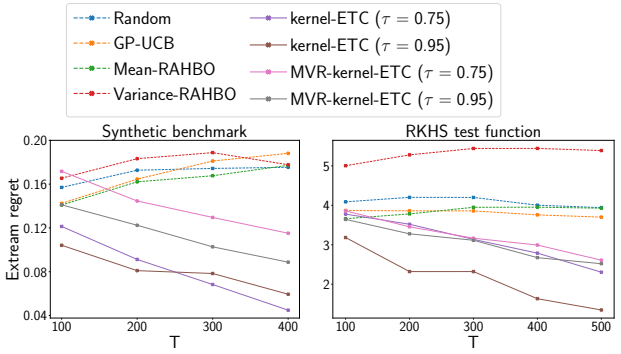

Figure 3: The average extreme regret of the numerical experiments in the heteroscedastic model settings.

Table 1: The average best validation loss with standard errors obtained from CNN tuning tasks with 10 different random seeds.

GP-UCB	Variance-RAHBO	kernel-ETC
1.457 ± 0.006	1.494 ± 0.011	**1.446** ± 0.005

Hyperparameter Tuning of Convolutional Neural Network As a demonstration of our heteroscedastic model setting, we conduct experiments on a hyperparameter tuning task of a convolutional neural network (CNN) with the CIFAR-10 dataset (Krizhevsky, 2009). We build a 3-layer CNN with 8 channels and define tuning parameters as epochs, batch size, and learning rate of stochastic gradient descent optimizer. Our goal is to obtain the model whose validation error is low with 100 trials of training. We omit `random`, `MVR-kernel-ETC`, and `kernel-ETC` with $\tau = 0.75$ since their performances tend to be worse than `kernel-ETC` with $\tau = 0.95$ in previous experiments. Moreover, we omit `Mean-RAHBO` whose objective function is the same as `GP-UCB`. Table 1 shows the results, in which we observe the superiority of kernel-ETC.

6 CONCLUSIONS

We formulate the risk-seeking BO problems under uncontrollable uncertainty factors and propose a novel kernel-ETC algorithm. Specifically, we consider the two types of settings based on the treatment of the uncontrollable factors. In both settings, we prove the regret upper bound of kernel-ETC and show that our kernel-ETC algorithm has no-regret guarantees.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP20H00601, JP23K19967, and JP23K16943, and RIKEN Center for Advanced Intelligence Project.

References

Abbasi-Yadkori, Y. (2013). *Online learning for linearly parametrized control problems*. PhD thesis, University of Alberta.

Achab, M., Clémençon, S., Garivier, A., Sabourin, A., and Vernade, C. (2017). Max k-armed bandit: On the extremehunter algorithm and beyond. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017*.

Baudry, D., Russac, Y., and Kaufmann, E. (2022). Efficient algorithms for extreme bandits. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*.

Belabed, C., Benabdelghani, Z., Granado, A., and Etxeberria, A. (2012). Miscibility and specific interactions in blends of poly (4-vinylphenol-co-methyl methacrylate)/poly (styrene-co-4-vinylpyridine). *Journal of applied polymer science*.

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

Carpentier, A. and Valko, M. (2014). Extreme bandits. *Proc. Neural Information Processing Systems (NeurIPS)*.

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

Cicirello, V. A. and Smith, S. F. (2005). The max k-armed bandit: A new model of exploration applied to search heuristic selection. In *Proc. Conference on Artificial Intelligence (AAAI)*.

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

Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Gardner, J. R., Kusner, M. J., Xu, Z. E., Weinberger, K. Q., and Cunningham, J. P. (2014). Bayesian optimization with inequality constraints. *Proc. International Conference on Machine Learning (ICML)*.

González, J., Longworth, J., James, D. C., and Lawrence, N. D. (2015). Bayesian optimization for synthetic gene design. *arXiv preprint arXiv:1505.01627*.

Inatsu, Y., Takeno, S., Karasuyama, M., and Takeuchi, I. (2022). Bayesian optimization for distributionally robust chance-constrained problem. In *Proc. International Conference on Machine Learning (ICML)*.

Iwazaki, S., Inatsu, Y., and Takeuchi, I. (2021a). Bayesian quadrature optimization for probability threshold robustness measure. *Neural Computation*.

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

Kandasamy, K., Dasarathy, G., Oliva, J., Schneider, J., and Poczos, B. (2019). Multi-fidelity Gaussian process bandit optimisation. *Journal of Artificial Intelligence Research*.

Kandasamy, K., Schneider, J., and Poczos, B. (2015). High dimensional Bayesian optimisation and bandits via additive models. In *Proc. International Conference on Machine Learning (ICML)*.

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

Kirschner, J. and Krause, A. (2018). Information directed sampling and bandits with heteroscedastic noise. In *Proc. Conference on Learning Theory (COLT)*.

Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. Master's thesis, University of Toronto.

Kwei, T. (1984). The effect of hydrogen bonding on the glass transition temperatures of polymer mixtures. *Journal of Polymer Science: Polymer Letters Edition*.

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

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

Makarova, A., Usmanova, I., Bogunovic, I., and Krause, A. (2021). Risk-averse heteroscedastic Bayesian optimization. In *Proc. Neural Information Processing Systems (NeurIPS)*.

Močkus, J. (1975). On Bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference*.

Nguyen, Q. P., Dai, Z., Low, B. K. H., and Jaillet, P. (2021a). Optimizing conditional value-at-risk of black-box functions. *Proc. Neural Information Processing Systems (NeurIPS)*.

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

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

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

Streeter, M. J. and Smith, S. F. (2006). A simple distribution-free approach to the max k-armed bandit problem. In *International Conference on Principles and Practice of Constraint Programming*.

Toscano-Palmerin, S. and Frazier, P. I. (2018). Bayesian optimization with expensive integrands. *arXiv preprint arXiv:1803.08661*.

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

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*. Cambridge university press.

Wang, Z. and Jegelka, S. (2017). Max-value entropy search for efficient Bayesian optimization. In *Proc. International Conference on Machine Learning (ICML)*.

Whitehouse, J., Wu, Z. S., and Ramdas, A. (2023). Improved self-normalized concentration in Hilbert spaces: Sublinear regret for GP-UCB. *Proc. Neural Information Processing Systems (NeurIPS)*.

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]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [No]
   (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] The detailed information on experiments is provided in Appendix E.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] The detailed results with a standard error are provided in Appendix E.
   (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 Material: Risk Seeking Bayesian Optimization under Uncertainty for Obtaining Extremum

A PROOFS OF SECTION 2

Lemma A.1. *Fix any natural number $\tilde{T} < T$ and algorithm. Suppose that $\boldsymbol{x}_t = \hat{\boldsymbol{x}}^*$ holds for any $t \in [T] \setminus [\tilde{T}]$, where $\hat{\boldsymbol{x}}^*$ is the random variable defined based on the history up to step $\tilde{T}$. Then, the following inequality holds:*

$$\Delta(T) \leq \mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\boldsymbol{x}^*, W_t)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\hat{\boldsymbol{x}}^*, W_t)\right]\right] + \mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\hat{\boldsymbol{x}}^*, W_t)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\hat{\boldsymbol{x}}^*, W_t)\right]\right].$$

Proof. Let $\mathcal{H}_{\tilde{T}} \coloneqq (\boldsymbol{x}_1, \boldsymbol{w}_1, y_1, \ldots, \boldsymbol{x}_{\tilde{T}}, \boldsymbol{w}_{\tilde{T}}, y_{\tilde{T}})$ be the history up to step $\tilde{T}$. Then,

$$\begin{aligned}
\Delta(T) &= \mathbb{E}\left[\max_{t\in[T]} f(\boldsymbol{x}^*, \boldsymbol{w}_t)\right] - \mathbb{E}\left[\max_{t\in[T]} f(\boldsymbol{x}_t, \boldsymbol{w}_t)\right] \\
&\leq \mathbb{E}\left[\max_{t\in[T]} f(\boldsymbol{x}^*, \boldsymbol{w}_t)\right] - \mathbb{E}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\boldsymbol{x}_t, \boldsymbol{w}_t)\right] \\
&= \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\boldsymbol{x}^*, W_t)\right] - \mathbb{E}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\hat{\boldsymbol{x}}^*, \boldsymbol{w}_t)\right] \\
&= \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\boldsymbol{x}^*, W_t)\right] - \mathbb{E}\left[\mathbb{E}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\hat{\boldsymbol{x}}^*, \boldsymbol{w}_t) \,\middle|\, \mathcal{H}_{\tilde{T}}\right]\right] \\
&= \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\boldsymbol{x}^*, W_t)\right] - \mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\hat{\boldsymbol{x}}^*, W_t)\right]\right] \\
&= \mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\boldsymbol{x}^*, W_t)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\hat{\boldsymbol{x}}^*, W_t)\right]\right] + \mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\hat{\boldsymbol{x}}^*, W_t)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\hat{\boldsymbol{x}}^*, W_t)\right]\right].
\end{aligned}$$

□

Lemma A.2. *Fix any algorithm and $f \in \mathcal{H}(k)$ with $\|f\|_{\mathcal{H}(k)} \leq B < \infty$. Let $\tilde{T}$ be a natural number that $\tilde{T} < \alpha T$ holds with some $\alpha \in (0, 1]$. Furthermore, suppose that $\boldsymbol{x}_t = \hat{\boldsymbol{x}}^*$ holds for any $t \in [T] \setminus [\tilde{T}]$, where $\hat{\boldsymbol{x}}^*$ is the random variable defined based on the history up to step $\tilde{T}$. Then, the following inequality holds:*

$$\mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\hat{\boldsymbol{x}}^*, W_t)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\hat{\boldsymbol{x}}^*, W_t)\right]\right] \leq 2B(1 - \underline{p})^{(1-\alpha)T},$$

where $\underline{p}$ is defined as $\underline{p} = \min_{\boldsymbol{w}\in\mathcal{W}} p(\boldsymbol{w})$.

Proof. Let us define $p_{\boldsymbol{x}}$ as $p_{\boldsymbol{x}} = \sum_{\boldsymbol{w}\in\mathcal{W}} \mathbb{1}\{f(\boldsymbol{x},\boldsymbol{w}) \neq \max_{\tilde{\boldsymbol{w}}\in\mathcal{W}} f(\boldsymbol{x},\tilde{\boldsymbol{w}})\}p(\boldsymbol{w})$ for any $\boldsymbol{x}\in\mathcal{X}$. Then, for any $\boldsymbol{x}\in\mathcal{X}$,

$$
\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\boldsymbol{x},W_t)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\boldsymbol{x},W_t)\right] \\
&= \left\{(1-p_{\boldsymbol{x}}^{T}) - (1-p_{\boldsymbol{x}}^{T-\tilde{T}})\right\}\max_{\boldsymbol{w}\in\mathcal{W}} f(\boldsymbol{x},\boldsymbol{w}) \\
&+ \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\mathbb{1}\left\{\max_{t\in[T]} f(\boldsymbol{x},W_t) \neq \max_{\boldsymbol{w}\in\mathcal{W}} f(\boldsymbol{x},\boldsymbol{w})\right\}\max_{t\in[T]} f(\boldsymbol{x},W_t)\right] \\
&- \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\mathbb{1}\left\{\max_{t\in[T]\setminus[\tilde{T}]} f(\boldsymbol{x},W_t) \neq \max_{\boldsymbol{w}\in\mathcal{W}} f(\boldsymbol{x},\boldsymbol{w})\right\}\max_{t\in[T]\setminus[\tilde{T}]} f(\boldsymbol{x},W_t)\right] \\
&\leq (p_{\boldsymbol{x}}^{T-\tilde{T}} - p_{\boldsymbol{x}}^{T})B + p_{\boldsymbol{x}}^{T}B + p_{\boldsymbol{x}}^{T-\tilde{T}}B \qquad \text{(A.5)} \\
&\leq 2Bp_{\boldsymbol{x}}^{T-\tilde{T}} \\
&\leq 2Bp_{\boldsymbol{x}}^{T-\alpha T} \\
&\leq 2B(1-\underline{p})^{(1-\alpha)T},
\end{aligned}
$$

where (A.5) follows from the inequality $\|f\|_\infty \leq \|f\|_{\mathcal{H}(k)} \leq B$. Therefore,

$$
\mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]} f(\hat{\boldsymbol{x}}^*,W_t)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{t\in[T]\setminus[\tilde{T}]} f(\hat{\boldsymbol{x}}^*,W_t)\right]\right] \leq 2B(1-\underline{p})^{(1-\alpha)T}.
$$

□

Lemma A.3. *The following inequality holds for any $T\in\mathbb{N}$:*

$$
\sum_{t=1}^{T}\sigma_{t-1}^2(\boldsymbol{x}_t,\boldsymbol{w}_t) \leq \frac{2\gamma(T)}{\ln(1+\sigma^{-2})}.
$$

Lemma A.3 follows from the proofs of Srinivas et al. (2010). See Lemma 5.3 and Lemma 5.4 in Srinivas et al. (2010) for more details.

Lemma A.4 (Lemma 3 in Kirschner and Krause (2018) or Lemma 7 in Kirschner et al. (2020))**.** *Let S_t be any non-negative stochastic process adapted to a filtration $\{\mathcal{F}_t\}$, and define m_t as $m_t = \mathbb{E}[S_t \mid \mathcal{F}_{t-1}]$. Suppose that there exists $K\geq 1$ which $S_t \leq K$ holds for any $t\in\mathbb{N}$. Then, for any $T\geq 1$, the following inequality holds with probability at least $1-\delta$:*

$$
\sum_{t=1}^{T} m_t \leq 2\sum_{t=1}^{T} S_t + 8K\ln\frac{6K}{\delta}.
$$

Lemma A.5. *For any $t\in[\tilde{T}]$, let us define $\|\sigma_{t-1}(\boldsymbol{x}_t,W_1)\|_{\psi_2}$ as*

$$
\|\sigma_{t-1}(\boldsymbol{x}_t,W_1)\|_{\psi_2} = \inf\left\{a>0 \;\middle|\; \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\exp\left(\frac{\sigma_{t-1}^2(\boldsymbol{x}_t,W_1)}{a^2}\right)\right] \leq 2\right\}. \qquad \text{(A.6)}
$$

Then, for any algorithm and $\delta\in(0,1)$, the following inequality holds with probability at least $1-\delta$:

$$
\sum_{t=1}^{\tilde{T}}\|\sigma_{t-1}(\boldsymbol{x}_t,W_1)\|_{\psi_2} \leq \sqrt{\frac{\tilde{T}(Q-1)}{\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8\ln\frac{6}{\delta}\right\}}, \qquad \text{(A.7)}
$$

where $Q = 2\underline{p}^{-1}$ with $\underline{p} = \min_{\boldsymbol{w}\in\mathcal{W}} p(\boldsymbol{w})$.

Proof. If there exists $\boldsymbol{w} \in \mathcal{W}$ that $\sigma_{t-1}^2(\boldsymbol{x}_t, \boldsymbol{w}) > 0$ holds, $\mathbb{E}\left[\exp\left(\sigma_{t-1}^2(\boldsymbol{x}_t, W_1)\|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^{-2}\right)\right] = 2$ holds. Then, for any $\boldsymbol{w} \in \mathcal{W}$,

$$\exp(\sigma_{t-1}^2(\boldsymbol{x}_t, \boldsymbol{w})\|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^{-2}) = p(\boldsymbol{w})^{-1}\left(2 - \sum_{\tilde{\boldsymbol{w}} \in \mathcal{W}\setminus\{\boldsymbol{w}\}} p(\tilde{\boldsymbol{w}})\exp(\sigma_{t-1}^2(\boldsymbol{x}_t, \tilde{\boldsymbol{w}})\|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^{-2})\right) \leq 2\underline{p}^{-1}$$
$$\Rightarrow \sigma_{t-1}^2(\boldsymbol{x}_t, \boldsymbol{w})\|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^{-2} \leq \ln Q.$$

Since the inequality $\exp(a) \leq 1 + a(\exp(\overline{a}) - 1)/\overline{a}$ holds for any $a \in [0, \overline{a}]$,

$$\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\exp\left(\sigma_{t-1}^2(\boldsymbol{x}_t, W_1)\|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^{-2}\right)\right] = 2 \\
&\Rightarrow \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[1 + \frac{(Q-1)\sigma_{t-1}^2(\boldsymbol{x}_t, W_1)\|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^{-2}}{\ln Q}\right] \geq 2 \\
&\Rightarrow \frac{Q-1}{\ln Q}\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\sigma_{t-1}^2(\boldsymbol{x}_t, W_1)\right] \geq \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^2. \qquad \text{(A.8)}
\end{aligned}$$

In addition, Eq. (A.8) also holds if $\sigma_{t-1}^2(\boldsymbol{x}_t, \boldsymbol{w}) = 0$ for all $\boldsymbol{w} \in \mathcal{W}$. Thus, by applying Lemma A.4 with $S_t = \sigma_{t-1}^2(\boldsymbol{x}_t, \boldsymbol{w}_t)$ and $B = 1$, the following inequality holds with probability at least $1 - \delta$:

$$\begin{aligned}
\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^2 &\leq \frac{Q-1}{\ln Q}\sum_{t=1}^{\tilde{T}} \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\sigma_{t-1}^2(\boldsymbol{x}_t, W_1)\right] \\
&= \frac{Q-1}{\ln Q}\sum_{t=1}^{\tilde{T}} \mathbb{E}\left[\sigma_{t-1}^2(\boldsymbol{x}_t, \boldsymbol{w}_t) \mid \mathcal{H}_{t-1}\right] \\
&\leq \frac{Q-1}{\ln Q}\left\{2\sum_{t=1}^{\tilde{T}} \sigma_{t-1}^2(\boldsymbol{x}_t, \boldsymbol{w}_t) + 8\ln\frac{6}{\delta}\right\} \\
&\leq \frac{Q-1}{\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8\ln\frac{6}{\delta}\right\}. \qquad \text{(A.9)}
\end{aligned}$$

Here, we set $\mathcal{H}_t$ as $\mathcal{H}_t = (\boldsymbol{x_1}, \boldsymbol{w}_1, y_1, \ldots, \boldsymbol{x}_t, \boldsymbol{w}_t, y_t)$. Moreover, Eq. (A.9) follows from Lemma A.3. By applying Schwarz inequality to $\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}$, we have $\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2} \leq \sqrt{\tilde{T}\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^2}$. We complete the proof by combining the inequality $\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2} \leq \sqrt{\tilde{T}\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}^2}$ with Eq. (A.9).

□

A.1 Proof of Theorem 2.1

We describe the full proof of Theorem 2.1 below.

Proof of Theorem 2.1. Let us define the event E as follows:

$$\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2} \leq \sqrt{\frac{\tilde{T}(Q-1)}{\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8\ln 12T\right\}} \qquad \text{(A.10)}$$
$$\text{and} \quad \forall t \in \mathbb{N}_+,\ \forall(\boldsymbol{x}, \boldsymbol{w}) \in \mathcal{X} \times \mathcal{W},\ |f(\boldsymbol{x}, \boldsymbol{w}) - \mu_{t-1}(\boldsymbol{x}, \boldsymbol{w})| \leq \beta^{1/2}(t)\sigma_{t-1}(\boldsymbol{x}, \boldsymbol{w}).$$

It should be noted that the event E is true with probability at least $1 - 1/T$, by applying the union bound to

Lemma 2.1 and Lemma A.5. Thus, under the event E, the following inequality holds:

$$\begin{aligned}
&\mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right]\right] \\
&= \mathbb{E}\left[\mathbb{1}\{E^c\}\left\{\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right]\right\}\right] \\
&+ \mathbb{E}\left[\mathbb{1}\{E\}\left\{\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right]\right\}\right] \\
&\leq \frac{2B}{T} + \mathbb{E}\left[\mathbb{1}\{E\}\left\{\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right]\right\}\right]. \qquad \text{(A.11)}
\end{aligned}$$

Furthermore, under the event E, the following inequality holds for any $t \in [\tilde{T}]$:

$$\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right] \\
&\leq \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} \mathrm{ucb}_t(\boldsymbol{x}_t, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} \mathrm{lcb}_t(\boldsymbol{x}_t, W_j)\right] \\
&\leq \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} \{\mathrm{ucb}_t(\boldsymbol{x}_t, W_j) - \mathrm{lcb}_t(\boldsymbol{x}_t, W_j)\}\right] \\
&\leq 2\beta(\tilde{T})^{1/2}\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} \sigma_{t-1}(\boldsymbol{x}_t, W_j)\right] \\
&\leq 2\beta(\tilde{T})^{1/2} C \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2}\sqrt{\ln T}, \qquad \text{(A.12)}
\end{aligned}$$

where $\|\cdot\|_{\psi_2}$ is defined in Eq. (A.6) and C is the absolute constant of Lemma 2.2. In Eq. (A.12), we use Lemma 2.2 with $X_i = \sigma_{t-1}(\boldsymbol{x}_t, W_i)$. By taking arithmetic mean in both sides of Eq. (A.12),

$$\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right] \\
&\leq 2\tilde{T}^{-1}\beta(\tilde{T})^{1/2} C\sqrt{\ln T}\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2} \\
&\leq 2C\sqrt{\frac{\beta(\tilde{T})(Q-1)\ln T}{\tilde{T}\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8\ln 12T\right\}}. \qquad \text{(A.13)}
\end{aligned}$$

In the last line, we use Lemma A.5. By combining Lemma A.1 with Eq. (A.11), (A.13), and Lemma A.2, we complete the proof. □

B DETAILS OF SECTION 3

B.1 Pseudo-code of kernel-ETC in Heteroscedastic Model Setting

Algorithm B.2 shows the pseudo-code of our kernel-ETC algorithm in the heteroscedastic model setting.

B.2 Proof of Theorem 3.1

Definition B.1 (Sub-exponential random variable and sub-exponential norm)**.** *Define sub-exponential norm* $\|X\|_{\psi_1}$ *of random variable* X *as*

$$\|X\|_{\psi_1} = \inf\left\{t > 0 \;\middle|\; \mathbb{E}\left[\exp\left(\frac{|X|}{t}\right)\right] \leq 2\right\}.$$

Furthermore, if $\|X\|_{\psi_1} < \infty$*,* X *is called sub-exponential random variable.*

Algorithm B.2 The kernel-ETC algorithm for heteroscedastic model setting.

Input: Kernel k_f, k_ρ, exploration ratio α, variance parameter $\lambda_\rho > 0$, number of repetition $m \geq 2$, lower and upper bound of $\rho(\cdot)$: $\underline{\rho}$, $\overline{\rho}$, width of confidence bounds $\{\beta_f(i)\}_{i\in\mathbb{N}_+}, \{\beta_\rho(i)\}_{i\in\mathbb{N}_+}$.

$\tilde{T} \leftarrow \lceil \alpha(T-1) \rceil$.
$t \leftarrow 1$.
$M \leftarrow \lfloor \tilde{T}/m \rfloor$.
Initialize GP prior of $f(\cdot)$ and $\rho(\cdot)$.
$\theta_T \leftarrow \tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t\in[T]} Z_t \right]$.
for $i = 1$ to M **do**
  $\boldsymbol{x}^{(i)} \leftarrow \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \left\{ \mathrm{ucb}_f^{(i)}(\boldsymbol{x}) + \theta_T \mathrm{ucb}_\rho^{(i)}(\boldsymbol{x}) \right\}$.
  for $j = 1, \ldots, m$ **do**
    $\boldsymbol{x}_t \leftarrow \boldsymbol{x}^{(i)}$.
    Observe $y_t = f(\boldsymbol{x}_t) + \eta_t(\boldsymbol{x}_t)$.
    $t \leftarrow t + 1$.
  end for
  Update GP posterior of $f(\cdot)$ and $\rho(\cdot)$.
end for
$\tilde{i} \leftarrow \mathrm{argmax}_{i\in[M]} \left\{ \mathrm{lcb}_f^{(M)}(\boldsymbol{x}) + \theta_T \mathrm{lcb}_\rho^{(M)}(\boldsymbol{x}) \right\}$.
$\hat{\boldsymbol{x}}^* \leftarrow \boldsymbol{x}^{(\tilde{i})}$.
for $t = mM + 1$ to T **do**
  $\boldsymbol{x}_t \leftarrow \hat{\boldsymbol{x}}^*$.
  Observe y_t.
end for

The following Lemma B.1 describes the relation between a sub-exponential norm and a variance proxy of sub-Gaussian random variable.

Lemma B.1. *If X^2 is a sub-exponential random variable, $X - \mathbb{E}[X]$ is a $c_1\sqrt{\|X^2\|_{\psi_1}}$-sub-Gaussian random variable, where $c_1 > 0$ is an absolute constant. That is, the following inequality holds for all $\lambda \in \mathbb{R}$:*

$$\mathbb{E}\left[\exp\left(\lambda\left(X - \mathbb{E}[X]\right)\right)\right] \leq \exp\left(\frac{\lambda^2 c_1^2 \|X^2\|_{\psi_1}}{2}\right).$$

Lemma B.1 directly follows by applying Lemma 2.7.6 and Eq. (2.16) of Vershynin (2018).

Lemma B.2. *Fix any natural number $m \geq 2$. If X is a chi-square random variable with $m-1$ degree of freedom, then $\|X\|_{\psi_1} \leq c_2\sqrt{m-1}$ holds with an absolute constant $c_2 > 0$.*

Proof. Let $Z_1, \ldots, Z_{m-1}$ be $m-1$ independent standard normal random variables. Since X equals $\sum_{i=1}^{m-1} Z_i^2$ in distribution, $\|X\|_{\psi_1} = \|\sum_{i=1}^{m-1} Z_i^2\|_{\psi_1}$ holds. Here, it is easy to see that a moment generating function of $Z_1^2 - \mathbb{E}[Z_1^2]$ satisfies the following:

$$\mathbb{E}[\exp(\lambda(Z_1^2 - \mathbb{E}[Z_1^2]))] = \frac{1}{\sqrt{1-2\lambda}}\exp(-\lambda) \leq \exp(2\lambda^2) \quad \text{for all } |\lambda| \leq \frac{1}{4}.$$

Therefore, due to the independence of $(Z_i)_{i\in[m-1]}$, the following inequality (B.14) holds for all $0 \leq \lambda \leq 1/4$:

$$\mathbb{E}\left[\exp\left(\lambda\left(\sum_{i=1}^{m-1} Z_i^2 - \mathbb{E}\left[\sum_{i=1}^{m-1} Z_i^2\right]\right)\right)\right] \leq \exp(2(m-1)\lambda^2). \tag{B.14}$$

Equation (B.14) implies

$$\mathbb{E}\left[\exp\left(\lambda\left(\sum_{i=1}^{m-1} Z_i^2 - \mathbb{E}\left[\sum_{i=1}^{m-1} Z_i^2\right]\right)\right)\right] \leq \exp\left((\max\{\sqrt{2(m-1)}, 4\})^2\lambda^2\right) \quad \text{for all } |\lambda| \leq \frac{1}{\max\{\sqrt{2(m-1)}, 4\}}. \tag{B.15}$$

Equation (B.15) implies that $\| \sum_{i=1}^{m-1} Z_i^2 - \mathbb{E}[\sum_{i=1}^{m-1} Z_i^2] \|_{\psi_1} \leq \tilde{c}_2 \max\{\sqrt{2(m-1)}, 4\}$ holds with some absolute constant $\tilde{c}_2 > 0$ (e.g., Proposition 2.7.1 in Vershynin (2018)). Finally, by applying the centering lemma (e.g., Exercise 2.7.10 in Vershynin (2018)), we can obtain the inequality: $\| \sum_{i=1}^{m-1} Z_i^2 \|_{\psi_1} \leq \hat{c}_2 \tilde{c}_2 \max\{\sqrt{2(m-1)}, 4\}$ with some absolute constant $\hat{c}_2 > 0$. Since $\max\{\sqrt{2(m-1)}, 4\} \leq 4\sqrt{m-1}$ for any $m \geq 2$, we complete the proof by setting c_2 as $c_2 = 4\hat{c}_2\tilde{c}_2$.

□

Lemma B.3. *Fix any $m \geq 2$ and $j \in \mathbb{N}_+$. Suppose $\rho(\cdot)$ satisfies $\|\rho\|_\infty \leq \overline{\rho}$ with some $\overline{\rho} > 0$. Let us define ζ_j as $\zeta_j = \rho(\boldsymbol{x}^{(j)}) - \hat{s}^{(j)}$, where $\boldsymbol{x}^{(j)}$ and $\hat{s}^{(j)}$ are defined in Sec. 3. Then, ζ_j is $c_3\kappa(m)\overline{\rho}$-sub-Gaussian random variable with $\kappa(m) = (m-1)^{1/4}\Gamma((m-1)/2)/\Gamma(m/2)$. Here, $c_3 > 0$ is an absolute constant.*

Proof. From the definition of $\hat{s}^{(j)}$,

$$\begin{aligned}\hat{s}^{(j)} &= \left\{ \sqrt{\frac{2}{m-1}} \frac{\Gamma(m/2)}{\Gamma((m-1)/2)} \right\}^{-1} \sqrt{\frac{1}{m-1} \sum_{l=1}^{m} \left(y_l^{(j)} - \hat{y}^{(j)} \right)^2} \\ &= \frac{\rho(\boldsymbol{x}^{(j)})}{\sqrt{2}} \frac{\Gamma((m-1)/2)}{\Gamma(m/2)} \sqrt{\frac{1}{\rho^2(\boldsymbol{x}^{(j)})} \sum_{l=1}^{m} \left(y_l^{(j)} - \hat{y}^{(j)} \right)^2}.\end{aligned}$$

Since $y_l^{(j)}$ independently follows $\mathcal{N}(f(\boldsymbol{x}^{(j)}), \rho^2(\boldsymbol{x}^{(j)}))$, $\sum_{l=1}^{m} (y_l^{(j)} - \hat{y}^{(j)})^2 / \rho^2(\boldsymbol{x}^{(j)})$ follows a chi-square distribution with $m-1$ degree of freedom. Thus, from Lemma B.2,

$$\|\hat{s}^{(j)2}\|_{\psi_1} \leq \frac{\overline{\rho}^2}{2} \frac{\Gamma((m-1)/2)^2}{\Gamma(m/2)^2} c_2 \sqrt{m-1}. \tag{B.16}$$

Finally, by using Lemma B.1, Eq. (B.16), and the fact that $\mathbb{E}[\hat{s}^{(j)}] = \rho(\boldsymbol{x}^{(j)})$ holds, we find ζ_j is $c_3\kappa(m)\overline{\rho}$-sub-Gaussian random variable with $c_3 = c_1\sqrt{c_2/2}$. □

Lemma B.4. *Fix any natural number $\tilde{T} < T$ and algorithm. Suppose that $\boldsymbol{x}_t = \hat{\boldsymbol{x}}^*$ holds for any $t \in [T] \setminus [\tilde{T}]$, where $\hat{\boldsymbol{x}}^*$ is the random variable defined based on the history up to step $\tilde{T}$. Then, if $\rho(\boldsymbol{x}) \leq \overline{\rho}$ for all $\boldsymbol{x} \in \mathcal{X}$, the following inequality holds:*

$$\Delta(T) \leq \mathbb{E}\left[\{ f(\boldsymbol{x}^*) + \theta_T \rho(\boldsymbol{x}^*) \} - \{ f(\hat{\boldsymbol{x}}^*) + \theta_T \rho(\hat{\boldsymbol{x}}^*) \} \right] + \overline{\rho} \left\{ \tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t \in [T]} Z_t \right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t \in [T] \setminus [\tilde{T}]} Z_t \right] \right\}.$$

Proof.

$$
\begin{aligned}
\Delta(T) &= \mathbb{E}\left[\max_{t\in[T]}\{f(\boldsymbol{x}^*)+\eta_t(\boldsymbol{x}^*)\}\right] - \mathbb{E}\left[\max_{t\in[T]}\{f(\boldsymbol{x}_t)+\eta_t(\boldsymbol{x}_t)\}\right] \\
&\leq \mathbb{E}\left[\max_{t\in[T]}\{f(\boldsymbol{x}^*)+\eta_t(\boldsymbol{x}^*)\}\right] - \mathbb{E}\left[\max_{t\in[T]\setminus[\tilde{T}]}\{f(\boldsymbol{x}_t)+\eta_t(\boldsymbol{x}_t)\}\right] \\
&= f(\boldsymbol{x}^*) + \mathbb{E}\left[\max_{t\in[T]}\eta_t(\boldsymbol{x}^*)\right] - \mathbb{E}\left[f(\hat{\boldsymbol{x}}^*) + \max_{t\in[T]\setminus[\tilde{T}]}\eta_t(\hat{\boldsymbol{x}}^*)\right] \\
&= f(\boldsymbol{x}^*) + \rho(\boldsymbol{x}^*)\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]} Z_t\right] - \mathbb{E}\left[f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*)\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]\setminus[\tilde{T}]} Z_t\right]\right] \\
&= f(\boldsymbol{x}^*) + \rho(\boldsymbol{x}^*)\theta_T - \mathbb{E}\left[f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*)\theta_T\right] \\
&\quad + \mathbb{E}\left[f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*)\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]} Z_t\right]\right] - \mathbb{E}\left[f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*)\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]\setminus[\tilde{T}]} Z_t\right]\right] \\
&\leq \mathbb{E}\left[\{f(\boldsymbol{x}^*)+\theta_T\rho(\boldsymbol{x}^*)\} - \{f(\hat{\boldsymbol{x}}^*)+\theta_T\rho(\hat{\boldsymbol{x}}^*)\}\right] + \mathbb{E}\left[\overline{\rho}\left\{\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]} Z_t\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]\setminus[\tilde{T}]} Z_t\right]\right\}\right] \\
&= \mathbb{E}\left[\{f(\boldsymbol{x}^*)+\theta_T\rho(\boldsymbol{x}^*)\} - \{f(\hat{\boldsymbol{x}}^*)+\theta_T\rho(\hat{\boldsymbol{x}}^*)\}\right] + \overline{\rho}\left\{\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]} Z_t\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]\setminus[\tilde{T}]} Z_t\right]\right\}.
\end{aligned}
$$

□

As used in a proof of Theorem 1 in Makarova et al. (2021), we use the following Lemma B.5 and Lemma B.6, which state confidence bounds of ρ and f, respectively.

Lemma B.5. *Fix any $\delta \in (0,1)$, $m \geq 2$, and $\rho \in \mathcal{H}(k_\rho)$. Suppose $\|\rho\|_{\mathcal{H}(k_\rho)} \leq B_\rho$ and $\forall \boldsymbol{x} \in \mathcal{X},\ \rho(\boldsymbol{x}) \in [\underline{\rho}, \overline{\rho}]$. Furthermore, let $\beta_\rho(i)$ and λ_ρ be $\beta_\rho^{1/2}(i) = B_\rho + \sqrt{2(\gamma_\rho(i)+\delta^{-1})}$ and $\lambda_\rho = c\kappa(m)\overline{\rho}$, respectively. Then, the following statement holds with probability at least $1-\delta$:*

$$
\forall i \in \mathbb{N}_+,\ \forall \boldsymbol{x} \in \mathcal{X},\ |\rho(\boldsymbol{x}) - \mu_\rho^{(i-1)}(\boldsymbol{x})| \leq \beta_\rho^{1/2}(i)\sigma_\rho^{(i-1)}(\boldsymbol{x}). \tag{B.17}
$$

Here, $\mu_\rho^{(i)}(\boldsymbol{x})$, $\sigma_\rho^{(i)}(\boldsymbol{x})$, and $\gamma_\rho(i)$ are defined in Sec. 3, and c is the absolute constant, which is the same as the constant c_3 in Lemma B.3.

Lemma B.6. *Fix any $\delta \in (0,1)$, $m \geq 2$, and $f \in \mathcal{H}(k_f)$ with $\|f\|_{\mathcal{H}(k_f)} \leq B_f$. Let $\beta_f(i)$ be $\beta_f^{1/2}(i) = B_f + \sqrt{2(\gamma_f(i)+1+\delta^{-1})}$. Then, under the condition that the event* (B.17) *holds, the following statement holds with probability at least $1-\delta$:*

$$
\forall i \in \mathbb{N}_+,\ \forall \boldsymbol{x} \in \mathcal{X},\ |f(\boldsymbol{x}) - \mu_f^{(i-1)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)})| \leq \beta_f^{1/2}(i)\sigma_f^{(i-1)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)}).
$$

Here, $\mu_f^{(i)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i)})$, $\sigma_f^{(i)}(\boldsymbol{x} \mid \hat{\boldsymbol{\Sigma}}^{(i)})$, and $\gamma_f(i)$ are defined in Sec. 3.

By noting that ζ_t is $c\kappa(m)\overline{\rho}$-sub-Gaussian random variable (Lemma B.3), Lemma B.5 follows from a direct application of Theorem 3.11 in Abbasi-Yadkori (2013). Furthermore, Lemma B.6 immediately follows by using Eq.(25) of Makarova et al. (2021) and Lemma 7 in Kirschner and Krause (2018), which is the heteroscedastic model version of Theorem 3.11 in Abbasi-Yadkori (2013).

The following Lemma B.7 describe a relation between the sums of posterior variances and the MIGs, which are provided in Appendix A.4 in Makarova et al. (2021).

Lemma B.7. *For any $m \geq 2$ and $M \in \mathbb{N}_+$, the following holds:*

$$
\sum_{i=1}^{M} \sigma_f^{(i-1)}(\boldsymbol{x}^{(i)} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)}) \leq \sqrt{\frac{2M\gamma_f(M)}{\ln(1+m\overline{\rho}^{-2})}} \quad \textit{and} \quad \sum_{i=1}^{M} \sigma_\rho^{(i-1)}(\boldsymbol{x}^{(i)}) \leq \sqrt{\frac{2M\gamma_\rho(M)}{\ln(1+\lambda_\rho^{-2})}}.
$$

Lemma B.8. *Fix $\alpha \in (0,1)$, $m \geq 2$, $f \in \mathcal{H}(k_f)$ with $\|f\|_{\mathcal{H}(k_f)} \leq B_f$, and $\rho \in \mathcal{H}(k_\rho)$ with $\|\rho\|_{\mathcal{H}(k_\rho)} \leq B_\rho$. Suppose $\forall \boldsymbol{x} \in \mathcal{X},\ \rho(\boldsymbol{x}) \in [\underline{\rho}, \overline{\rho}]$ holds for some $\overline{\rho}$ and $\underline{\rho}$ with $\overline{\rho} \geq \underline{\rho} > 0$. Then, when running Algorithm B.2 with $\beta_f^{1/2}(i) = \sqrt{2(\ln 2T + 1 + \gamma_f(i))} + B_f$, $\beta_\rho^{1/2}(i) = \sqrt{2(\ln 2T + \gamma_\rho(i))} + B_\rho$, and $\lambda_\rho = c\kappa(m)\overline{\rho}$, the following inequality holds:*

$$\begin{aligned}
&\mathbb{E}\left[\{f(\boldsymbol{x}^*) + \theta_T \rho(\boldsymbol{x}^*)\} - \{f(\hat{\boldsymbol{x}}^*) + \theta_T \rho(\hat{\boldsymbol{x}}^*)\}\right] \\
&\leq \frac{(2B + \overline{\rho}\sqrt{2\ln T})}{T} + \sqrt{\frac{8m\beta_f\left(\frac{\alpha T+1}{m}\right)}{\alpha(T-m-1)\ln(1+m\overline{\rho}^{-2})}\gamma_f\left(\frac{\alpha T+1}{m}\right)} \\
&\quad + \sqrt{\frac{16m\beta_\rho\left(\frac{\alpha T+1}{m}\right)\ln T}{\alpha(T-m-1)\ln(1+c^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}\gamma_\rho\left(\frac{\alpha T+1}{m}\right)},
\end{aligned}$$

where $\kappa(m) = (m-1)^{1/4}\Gamma((m-1)/2)/\Gamma(m/2)$, and $c > 0$ is an absolute constant, which is the same as the constant c_3 in Lemma B.3.

Proof. Let us assume the following event (B.18) holds:

$$\forall \boldsymbol{x} \in \mathcal{X},\ \forall i \in \mathbb{N}_+,\ f(\boldsymbol{x}) \in \left[\mathrm{lcb}_f^{(i)}(\boldsymbol{x}), \mathrm{ucb}_f^{(i)}(\boldsymbol{x})\right] \quad \text{and} \quad \rho(\boldsymbol{x}) \in \left[\mathrm{lcb}_\rho^{(i)}(\boldsymbol{x}), \mathrm{ucb}_\rho^{(i)}(\boldsymbol{x})\right] \tag{B.18}$$

Then, for any $i \in [M]$,

$$\begin{aligned}
&\{f(\boldsymbol{x}^*) + \theta_T \rho(\boldsymbol{x}^*)\} - \{f(\hat{\boldsymbol{x}}^*) + \theta_T \rho(\hat{\boldsymbol{x}}^*)\} \\
&\leq \mathrm{ucb}_f^{(i)}(\boldsymbol{x}^*) + \theta_T \mathrm{ucb}_\rho^{(i)}(\boldsymbol{x}^*) - \mathrm{lcb}_f^{(\tilde{i})}(\boldsymbol{x}^{(\tilde{i})}) - \theta_T \mathrm{lcb}_\rho^{(\tilde{i})}(\boldsymbol{x}^{(\tilde{i})}) \\
&\leq \mathrm{ucb}_f^{(i)}(\boldsymbol{x}^{(i)}) + \theta_T \mathrm{ucb}_\rho^{(i)}(\boldsymbol{x}^{(i)}) - \mathrm{lcb}_f^{(i)}(\boldsymbol{x}^{(i)}) - \theta_T \mathrm{lcb}_\rho^{(i)}(\boldsymbol{x}^{(i)}) \\
&= 2\beta_f^{1/2}(i)\sigma_f^{(i-1)}(\boldsymbol{x}^{(i)} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)}) + 2\theta_T \beta_\rho^{1/2}(i)\sigma_\rho^{(i-1)}(\boldsymbol{x}^{(i)}),
\end{aligned} \tag{B.19}$$

where $M = \lfloor \tilde{T}/m \rfloor$, $\tilde{T} = \lceil \alpha(T-1) \rceil$ and $\tilde{i} = \mathrm{argmax}_{i \in [M]} \{\mathrm{lcb}_f^{(M)}(\boldsymbol{x}) + \theta_T \mathrm{lcb}_\rho^{(M)}(\boldsymbol{x})\}$. By taking arithmetic mean in Eq. (B.19),

$$\begin{aligned}
&\{f(\boldsymbol{x}^*) + \theta_T \rho(\boldsymbol{x}^*)\} - \{f(\hat{\boldsymbol{x}}^*) + \theta_T \rho(\hat{\boldsymbol{x}}^*)\} \\
&\leq \frac{2}{M}\sum_{i=1}^{M} \beta_f^{1/2}(i)\sigma_f^{(i-1)}(\boldsymbol{x}^{(i)} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)}) + \frac{2\theta_T}{M}\sum_{i=1}^{M} \beta_\rho^{1/2}(i)\sigma_\rho^{(i-1)}(\boldsymbol{x}^{(i)}) \\
&\leq \frac{2\beta_f^{1/2}(M)}{M}\sum_{i=1}^{M} \sigma_f^{(i-1)}(\boldsymbol{x}^{(i)} \mid \hat{\boldsymbol{\Sigma}}^{(i-1)}) + \frac{2\theta_T \beta_\rho^{1/2}(M)}{M}\sum_{i=1}^{M} \sigma_\rho^{(i-1)}(\boldsymbol{x}^{(i)}) \qquad \text{(B.20)} \\
&\leq \frac{2\beta_f^{1/2}(M)}{M}\sqrt{\frac{2M}{\ln(1+m\overline{\rho}^{-2})}\gamma_f(M)} + \frac{2\theta_T \beta_\rho^{1/2}(M)}{M}\sqrt{\frac{2M}{\ln(1+c^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}\gamma_\rho(M)} \qquad \text{(B.21)} \\
&= \sqrt{\frac{8\beta_f(M)}{M\ln(1+m\overline{\rho}^{-2})}\gamma_f(M)} + \theta_T\sqrt{\frac{8\beta_\rho(M)}{M\ln(1+c^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}\gamma_\rho(M)}, \qquad \text{(B.22)}
\end{aligned}$$

where:

- Eq. (B.20) follows from monotonicity of $\beta_f(i)$ and $\beta_\rho(i)$.
- Eq. (B.21) follows from Lemma B.7.

Since the event (B.18) holds with probability at least $1 - 1/T$ by taking union bound in Lemma B.5 and

Lemma B.6, the following inequality holds:

$$
\begin{aligned}
&\mathbb{E}\left[\{f(\boldsymbol{x}^*)+\theta_T\rho(\boldsymbol{x}^*)\}-\{f(\hat{\boldsymbol{x}}^*)+\theta_T\rho(\hat{\boldsymbol{x}}^*)\}\right] \\
&=\mathbb{E}\left[\mathbb{1}\{\text{(B.18) is true}\}\left[\{f(\boldsymbol{x}^*)+\theta_T\rho(\boldsymbol{x}^*)\}-\{f(\hat{\boldsymbol{x}}^*)+\theta_T\rho(\hat{\boldsymbol{x}}^*)\}\right]\right] \\
&\quad+\mathbb{E}\left[\mathbb{1}\{\text{(B.18) is false}\}\left[\{f(\boldsymbol{x}^*)+\theta_T\rho(\boldsymbol{x}^*)\}-\{f(\hat{\boldsymbol{x}}^*)+\theta_T\rho(\hat{\boldsymbol{x}}^*)\}\right]\right] \\
&\leq \frac{(2B+\theta_T\overline{\rho})}{T}+\left(1-\frac{1}{T}\right)\left[\sqrt{\frac{8\beta_f(M)}{M\ln(1+m\overline{\rho}^{-2})}\gamma_f(M)}+\theta_T\sqrt{\frac{8\beta_\rho(M)}{M\ln(1+c^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}\gamma_\rho(M)}\right] \qquad \text{(B.23)} \\
&\leq \frac{(2B+\overline{\rho}\sqrt{2\ln T})}{T}+\sqrt{\frac{8\beta_f(M)}{M\ln(1+m\overline{\rho}^{-2})}\gamma_f(M)}+\sqrt{\frac{16\beta_\rho(M)\ln T}{M\ln(1+c^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}\gamma_\rho(M)} \qquad \text{(B.24)} \\
&\leq \frac{(2B+\overline{\rho}\sqrt{2\ln T})}{T}+\sqrt{\frac{8m\beta_f\left(\frac{\alpha T+1}{m}\right)}{\alpha(T-m-1)\ln(1+m\overline{\rho}^{-2})}\gamma_f\left(\frac{\alpha T+1}{m}\right)} \\
&\quad+\sqrt{\frac{16m\beta_\rho\left(\frac{\alpha T+1}{m}\right)\ln T}{\alpha(T-m-1)\ln(1+c^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}\gamma_\rho\left(\frac{\alpha T+1}{m}\right)}, \qquad \text{(B.25)}
\end{aligned}
$$

where:

- Eq. (B.23) follows from Eq. (B.22) and the fact that $\|f\|_\infty \leq \|f\|_{\mathcal{H}(k_f)} \leq B_f$ and $\|f\|_\infty \leq \|\rho\|_\infty \leq \overline{\rho}$ hold.
- Eq. (B.24) follows from a upper bound of the expectation for T standard normal random variables: $\theta_T \leq \sqrt{2\ln T}$.
- Eq. (B.25) follows from $\alpha(T-m-1)/m \leq M \leq (\alpha T+1)/m$.

The final inequality (B.25) is the desired inequality of the lemma; thus, the proof is completed. □

Lemma B.9. *Fix any natural number $\tilde{T} < T$. Then,*

$$
\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]} Z_t\right]-\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]\setminus[\tilde{T}]} Z_t\right] \leq \frac{\tilde{T}}{T}\left(\sqrt{2\ln T}+\frac{1}{2\pi}+\frac{1}{2\sqrt{2\pi}\ln T}\right)+\frac{1}{2^{T-\tilde{T}-1}}. \qquad \text{(B.26)}
$$

A proof strategy of Lemma B.9 follows discussions in Appendix A.4 of Baudry et al. (2022).

Proof of Lemma B.9.

$$
\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]} Z_t\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]\setminus[\tilde{T}]} Z_t\right] \\
&= \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[T]\setminus[\tilde{T}]} Z_t\right\}\left\{\max_{t\in[T]} Z_t - \max_{t\in[T]\setminus[\tilde{T}]} Z_t\right\}\right] + \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\left\{\max_{t\in[T]} Z_t - \max_{t\in[T]\setminus[\tilde{T}]} Z_t\right\}\right] \\
&= \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\left\{\max_{t\in[\tilde{T}]} Z_t - \max_{t\in[T]\setminus[\tilde{T}]} Z_t\right\}\right] \\
&= \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\max_{t\in[\tilde{T}]} Z_t\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\max_{t\in[T]\setminus[\tilde{T}]} Z_t\right] \\
&\leq \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\max_{t\in[\tilde{T}]} Z_t\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]\setminus[\tilde{T}]} Z_t \leq 0\right\}\max_{t\in[T]\setminus[\tilde{T}]} Z_t\right] \\
&\leq \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\max_{t\in[\tilde{T}]} Z_t\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]\setminus[\tilde{T}]} Z_t \leq 0\right\} Z_T\right] \\
&= \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\max_{t\in[\tilde{T}]} Z_t\right] - \left\{\prod_{t=\tilde{T}}^{T-1}\tilde{\mathbb{P}}\left(Z_t \leq 0\right)\right\}\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{Z_T \leq 0\right\} Z_T\right] \\
&= \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\max_{t\in[\tilde{T}]} Z_t\right] + \frac{1}{\sqrt{2\pi}2^{T-\tilde{T}-1}}, \qquad \text{(B.27)}
\end{aligned}
$$

where Eq. (B.27) follows from $\tilde{\mathbb{P}}\left(Z_t \leq 0\right) = 1/2$ and $\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{Z_T \leq 0\right\} Z_T\right] = -2/\sqrt{2\pi}$. Here, for any $q > 0$,

$$
\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\max_{t\in[\tilde{T}]} Z_t\right] \\
&\leq q\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t\right\}\right] + \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[T]} Z_t = \max_{t\in[\tilde{T}]} Z_t \ \text{ and } \ \max_{t\in[\tilde{T}]} Z_t > q\right\}\max_{t\in[\tilde{T}]} Z_t\right] \\
&\leq q\sum_{j=1}^{\tilde{T}}\tilde{\mathbb{P}}\left(\max_{t\in[T]} Z_t = Z_j\right) + \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[\tilde{T}]} Z_t > q\right\}\max_{t\in[\tilde{T}]} Z_t\right] \\
&\leq \frac{\tilde{T}}{T}q + \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[\tilde{T}]} Z_t > q\right\}\max_{t\in[\tilde{T}]} Z_t\right]. \qquad \text{(B.28)}
\end{aligned}
$$

Furthermore, by using the fact that $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)\mathrm{d}x$ holds for any non-negative random variable X,

$$
\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathbb{1}\left\{\max_{t\in[\tilde{T}]} Z_t > q\right\}\max_{t\in[\tilde{T}]} Z_t\right] \\
&= \int_0^\infty \tilde{\mathbb{P}}\left(\mathbb{1}\left\{\max_{t\in[\tilde{T}]} Z_t > q\right\}\max_{t\in[\tilde{T}]} Z_t > x\right)\mathrm{d}x \\
&= \int_0^q \tilde{\mathbb{P}}\left(\mathbb{1}\left\{\max_{t\in[\tilde{T}]} Z_t > q\right\}\max_{t\in[\tilde{T}]} Z_t > x\right)\mathrm{d}x + \int_q^\infty \tilde{\mathbb{P}}\left(\mathbb{1}\left\{\max_{t\in[\tilde{T}]} Z_t > q\right\}\max_{t\in[\tilde{T}]} Z_t > x\right)\mathrm{d}x. \qquad \text{(B.29)}
\end{aligned}
$$

By noting $\mathbb{1}\{\max_{t\in[\tilde{T}]} Z_t > q\}\max_{t\in[\tilde{T}]} Z_t > x \Leftrightarrow \max_{t\in[\tilde{T}]} Z_t > q$ holds for any $x \in [0, q]$, the following inequality

holds for the first term of Eq. (B.29):

$$\begin{aligned}
&\int_0^q \tilde{\mathbb{P}}\left(\mathbb{1}\left\{\max_{t\in[\tilde{T}]} Z_t > q\right\} \max_{t\in[\tilde{T}]} Z_t > x\right) \mathrm{d}x \\
&= q\tilde{\mathbb{P}}\left(\max_{t\in[\tilde{T}]} Z_t > q\right) \\
&= q\tilde{\mathbb{P}}\left(\bigcup_{j\in[\tilde{T}]} \{Z_j > q\}\right) \\
&\leq q\sum_{j=1}^{\tilde{T}} \tilde{\mathbb{P}}\left(Z_j > q\right) \\
&\leq \frac{\tilde{T}}{\sqrt{2\pi}} \exp\left(-\frac{q^2}{2}\right), \qquad \text{(B.30)}
\end{aligned}$$

where Eq. (B.30) follows from an inequality: $\tilde{\mathbb{P}}(Z_t \geq q) \leq \phi(q)/q$ of tail probability for a standard normal random variable. Moreover, by noting $\mathbb{1}\{\max_{t\in[\tilde{T}]} Z_t > q\} \max_{t\in[\tilde{T}]} Z_t > x \Leftrightarrow \max_{t\in[\tilde{T}]} Z_t > x$ holds for any $x \in [q, \infty)$, the following inequality holds for the second term of Eq. (B.29):

$$\begin{aligned}
&\int_q^\infty \tilde{\mathbb{P}}\left(\mathbb{1}\left\{\max_{t\in[\tilde{T}]} Z_t > q\right\} \max_{t\in[\tilde{T}]} Z_t > x\right) \mathrm{d}x \\
&= \int_q^\infty \tilde{\mathbb{P}}\left(\max_{t\in[\tilde{T}]} Z_t > x\right) \mathrm{d}x \\
&\leq \sum_{j=1}^{\tilde{T}} \int_q^\infty \tilde{\mathbb{P}}\left(Z_j > x\right) \mathrm{d}x \\
&\leq \tilde{T}\int_q^\infty \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\
&\leq \tilde{T}\int_q^\infty \frac{x}{q^2\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\
&\leq \frac{\tilde{T}}{q^2\sqrt{2\pi}} \left[-\exp\left(-\frac{x^2}{2}\right)\right]_q^\infty \\
&= \frac{\tilde{T}}{q^2\sqrt{2\pi}} \exp\left(-\frac{q^2}{2}\right). \qquad \text{(B.31)}
\end{aligned}$$

By choosing q as $q = \sqrt{2\ln T}$, we can obtain Eq. (B.26) by using Eq. (B.27), (B.28), (B.29), (B.30), and (B.31). □

The following Theorem B.1 is a detailed version of Theorem 3.1.

Theorem B.1. *Fix $\alpha \in (0, 1]$, $m \geq 2$, $f \in \mathcal{H}(k_f)$ with $\|f\|_{\mathcal{H}(k_f)} \leq B_f$, and $\rho \in \mathcal{H}(k_\rho)$ with $\|\rho\|_{\mathcal{H}(k_\rho)} \leq B_\rho$. Suppose $\forall \boldsymbol{x} \in \mathcal{X}$, $\rho(\boldsymbol{x}) \in [\underline{\rho}, \overline{\rho}]$ holds for some $\overline{\rho}$ and $\underline{\rho}$ with $\overline{\rho} \geq \underline{\rho} > 0$. Then, when running Algoritm B.2 with $\beta_f^{1/2}(i) = \sqrt{2(\ln 2T + 1 + \gamma_f(i))} + B_f$, $\beta_\rho^{1/2}(i) = \sqrt{2(\ln 2T + \gamma_\rho(i))} + B_\rho$, and $\lambda_\rho = c\kappa(m)\overline{\rho}$, the following upper bound of extreme regret holds:*

$$\begin{aligned}
\Delta(T) \leq{}& \frac{(2B + \overline{\rho}\sqrt{2\ln T})}{T} + \sqrt{\frac{8m\beta_f\left(\frac{\alpha T+1}{m}\right)}{\alpha(T-m-1)\ln(1+m\overline{\rho}^{-2})}\gamma_f\left(\frac{\alpha T+1}{m}\right)} \\
&+ \sqrt{\frac{16m\beta_\rho\left(\frac{\alpha T+1}{m}\right)\ln T}{\alpha(T-m-1)\ln(1+c^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}\gamma_\rho\left(\frac{\alpha T+1}{m}\right)} + \alpha\overline{\rho}\left(\sqrt{2\ln T} + \frac{1}{2\pi} + \frac{1}{2\sqrt{2\pi}\ln T}\right) + \frac{\overline{\rho}}{2^{(1-\alpha)T-2}}, \qquad \text{(B.32)}
\end{aligned}$$

where $\kappa(m) = (m-1)^{1/4}\Gamma((m-1)/2)/\Gamma(m/2)$*, and* $c > 0$ *is an absolute constant. Especially, by setting* $\alpha = T^\tau/T$ *with* $\tau \in (0,1)$*, then*

$$\Delta(T) = \mathcal{O}\left(\frac{B}{T} + \sqrt{\frac{m\beta_f(T^\tau/m)\gamma_f(T^\tau/m)}{(T^\tau - m)\ln(1+m\overline{\rho}^{-2})}} + \frac{\overline{\rho}T^\tau\sqrt{\ln T}}{T} + \sqrt{\frac{m\beta_\rho(T^\tau/m)\gamma_\rho(T^\tau/m)\ln T}{(T^\tau - m)\ln(1+mc^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}}\right).$$

Proof. By combining Lemma B.4, Lemma B.8, and Lemma B.9 with $\tilde{T} = \lceil \alpha(T-1) \rceil$, the following inequality holds:

$$\begin{aligned}\Delta(T) \leq{}& \frac{(2B + \overline{\rho}\sqrt{2\ln T})}{T} + \sqrt{\frac{8m\beta_f\left(\frac{\alpha T+1}{m}\right)}{\alpha(T-m-1)\ln(1+m\overline{\rho}^{-2})}\gamma_f\left(\frac{\alpha T+1}{m}\right)} + \frac{\overline{\rho}}{2^{T-\lceil\alpha(T-1)\rceil-1}} \\ &+ \sqrt{\frac{16m\beta_\rho\left(\frac{\alpha T+1}{m}\right)\ln T}{\alpha(T-m-1)\ln(1+c^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}\gamma_\rho\left(\frac{\alpha T+1}{m}\right)} + \frac{\overline{\rho}\lceil\alpha(T-1)\rceil}{T}\left(\sqrt{2\ln T} + \frac{1}{2\pi} + \frac{1}{2\sqrt{2\pi}\ln T}\right).\end{aligned} \tag{B.33}$$

Equation (B.32) is derived from Eq. (B.33) since $\lceil \alpha(T-1) \rceil \leq \alpha T + 1$ and $1/T \leq 1$ hold. □

B.3 Additional Lemma

Lemma B.10 (Unbiased sample variance $\hat{m}^{(i)}$ is not sub-Gaussian random variable)**.** *Let* $Y_1, \ldots, Y_m$ *be random variables which independently follows* $\mathcal{N}(a, b^2)$ *with any* $a \in \mathbb{R}, b > 0$*. Define* $\hat{M}$ *as*

$$\hat{M} = \frac{1}{m-1}\sum_{l=1}^{m}(Y_l - \overline{Y})^2,$$

where $\overline{Y} = \sum_{l=1}^m Y_l/m$*. Then,* $\hat{M} - b^2$ *is not sub-Gaussian random variable. Namely, the following statement holds:*

$$\forall \lambda_1 \geq 0,\ \exists \lambda_2 \in \mathbb{R},\ \mathbb{E}[\exp(\lambda_2(\hat{M} - b^2))] > \exp\left(\frac{\lambda_1^2\lambda_2^2}{2}\right) \tag{B.34}$$

Proof. For any $\lambda_2 \in \mathbb{R}$, we have

$$\mathbb{E}\left[\exp\left(\lambda_2(\hat{M} - b^2)\right)\right] = \exp(-(m-1))\mathbb{E}\left[\exp\left(\frac{\lambda_2 b^2}{m-1}\frac{1}{b^2}\sum_{l=1}^{m}(Y_l - \overline{Y})^2\right)\right].$$

Here, it should be noted that $\frac{1}{b^2}\sum_{l=1}^m(Y_l - \overline{Y})^2$ follows chi-square distribution. Since a moment-generating function of chi-square distribution becomes infinity when the input is greater than $1/2$, Eq. (B.34) holds by choosing λ_2 as $\lambda_2 > (m-1)/(2b^2)$. □

C DETAILS OF SECTION 4

C.1 Environmental Model Setting

Algorithm C.3 shows a pseudo-code of MVR-based kernel-ETC algorithm. The differences between UCB-based and MVR-based kernel-ETC in the environmental model setting are described as follows.

- The query strategy in the exploration period (Line 3 in Algorithm 1 and Algorithm C.3): Algorithm 1 chooses $\boldsymbol{x}_t$ based on UCB, which depends on the noise terms (ϵ_t), whereas Algorithm C.3 chooses $\boldsymbol{x}_t$ based on a posterior variance of GP, which are independent of (ϵ_t).
- The query point $\hat{\boldsymbol{x}}^*$ of exploitation phase (Lines 6-7 in Algorithm 1 and Line 6 in Algorithm C.3): The query point$\hat{\boldsymbol{x}}^*$ is defined based on LCB in the UCB-based exploration strategy of Algorithm 1, whereas the posterior means are used in MVR-based exploration strategy of Algorithm C.3. These definitions of $\hat{\boldsymbol{x}}^*$ based on LCBs and posterior means are the same as in the theoretical analysis of simple regret in UCB and MVR-based BO algorithms, respectively (Bogunovic et al., 2018; Vakili et al., 2021).

Algorithm C.3 The MVR-based kernel-ETC algorithm for environmental model setting.

Input: GP prior $\mathcal{GP}(0,\ k)$, exploration ratio $\alpha \in (0, 1]$.
$\tilde{T} \leftarrow \lceil \alpha(T-1) \rceil$.
for $t = 1$ to $\tilde{T}$ **do**
  $\boldsymbol{x}_t \leftarrow \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{j\in[T]} \sigma_{t-1}(\boldsymbol{x}, W_j)\right]$.
  Observe y_t and update GP posterior.
end for
$\hat{\boldsymbol{x}}^* \leftarrow \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \tilde{\mathbb{E}}_{\boldsymbol{W}} \left[\max_{j\in[T]} \mu_{\tilde{T}}(\boldsymbol{x}, W_j)\right]$.
for $t = \tilde{T}+1$ to T **do**
  $\boldsymbol{x}_t \leftarrow \hat{\boldsymbol{x}}^*$.
  Observe y_t.
end for

C.1.1 Proof of Theorem 4.1

We first describe the following Lemma C.1, which gives tight confidence bounds under a non-adaptive learner's strategy.

Lemma C.1 (Theorem 1 in Vakili et al. (2021))**.** *Fix $f \in \mathcal{H}(k)$ with $\|f\|_{\mathcal{H}(k)} \leq B$, $\delta \in (0, 1)$, $\tilde{T} \in \mathbb{N}_+$, and $\mathcal{X}$ with $|\mathcal{X}| < \infty$. Let us assume that the noise term ϵ_t independently follows $\mathcal{N}(0, \sigma^2)$. Furthermore, suppose that the learner's decisions $(\boldsymbol{x}_t)_{t\in\mathbb{N}_+}$ are independent of the noise terms $(\epsilon_t)_{t\in\mathbb{N}_+}$. Then, with probability at least $1-\delta$, the following inequality holds for any $(\boldsymbol{x}, \boldsymbol{w}) \in \mathcal{X} \times \mathcal{W}$:*

$$|f(\boldsymbol{x}, \boldsymbol{w}) - \mu_{\tilde{T}}(\boldsymbol{x}, \boldsymbol{w})| \leq \left(B + \sqrt{2 \ln \frac{2|\mathcal{X}||\mathcal{W}|}{\delta}}\right) \sigma_{\tilde{T}}(\boldsymbol{x}, \boldsymbol{w}).$$

Now, we show a proof of Theorem 4.1 below.

Proof of Theorem 4.1. Let us consider the following event (C.35):

$$\sum_{t=1}^{\tilde{T}} \|\sigma_{t-1}(\boldsymbol{x}_t, W_1)\|_{\psi_2} \leq \sqrt{\frac{\tilde{T}(Q-1)}{\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8 \ln 12T\right\}} \quad \text{(C.35)}$$
$$\text{and} \ \ \forall(\boldsymbol{x}, \boldsymbol{w}) \in \mathcal{X} \times \mathcal{W},\ |f(\boldsymbol{x}, \boldsymbol{w}) - \mu_{\tilde{T}}(\boldsymbol{x}, \boldsymbol{w})| \leq \beta^{1/2}(T)\sigma_{\tilde{T}}(\boldsymbol{x}, \boldsymbol{w}).$$

By noting that the event (C.35) holds with probability at least $1 - 1/T$ from Lemma A.5 and Lemma C.1, we have,

$$\mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right]\right]$$
$$\leq \frac{2B}{T} + \mathbb{E}\left[\mathbb{1}\{\text{(C.35) is true}\}\left\{\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right]\right\}\right]. \quad \text{(C.36)}$$

Furthermore, under the event (C.35), we have:

$$
\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right] \\
&\leq \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\left\{\mu_{\tilde{T}}(\boldsymbol{x}^*, W_j) + \beta^{1/2}(T)\sigma_{\tilde{T}}(\boldsymbol{x}^*, W_j)\right\}\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\left\{\mu_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j) - \beta^{1/2}(T)\sigma_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j)\right\}\right] \\
&\leq \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\left\{\mu_{\tilde{T}}(\boldsymbol{x}^*, W_j) + \beta^{1/2}(T)\sigma_{\tilde{T}}(\boldsymbol{x}^*, W_j)\right\}\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\mu_{\tilde{T}}(\boldsymbol{x}^*, W_j)\right] \\
&\quad + \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\mu_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\left\{\mu_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j) - \beta^{1/2}(T)\sigma_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j)\right\}\right] \\
&\leq \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\left\{\mu_{\tilde{T}}(\boldsymbol{x}^*, W_j) + \beta^{1/2}(T)\sigma_{\tilde{T}}(\boldsymbol{x}^*, W_j) - \mu_{\tilde{T}}(\boldsymbol{x}^*, W_j)\right\}\right] \\
&\quad + \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\left\{\mu_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j) - \mu_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j) + \beta^{1/2}(T)\sigma_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j)\right\}\right] \\
&= \beta^{1/2}(T)\left\{\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\sigma_{\tilde{T}}(\boldsymbol{x}^*, W_j)\right] + \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\sigma_{\tilde{T}}(\hat{\boldsymbol{x}}^*, W_j)\right]\right\} \\
&\leq 2\beta^{1/2}(T)\max_{\boldsymbol{x}\in\mathcal{X}}\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\sigma_{\tilde{T}}(\boldsymbol{x}, W_j)\right].
\end{aligned}
$$

From the monotonicity of the posterior variance and the definition of $\boldsymbol{x}_t$, the following inequality holds for any $t \in [\tilde{T}]$:

$$
\max_{\boldsymbol{x}\in\mathcal{X}}\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\sigma_{\tilde{T}}(\boldsymbol{x}, W_j)\right] \leq \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]}\sigma_{t-1}(\boldsymbol{x}_t, W_j)\right] \leq C\|\sigma_{t-1}(\boldsymbol{x}_t, W_j)\|_{\psi_2}\sqrt{\ln T},
$$

where we use Lemma 2.2 in the second inequality. By taking arithmetic mean over $t \in [\tilde{T}]$, we have,

$$
\begin{aligned}
\tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\boldsymbol{x}^*, W_j)\right] - \tilde{\mathbb{E}}_{\boldsymbol{W}}\left[\max_{j\in[T]} f(\hat{\boldsymbol{x}}^*, W_j)\right] &\leq \frac{2C\beta^{1/2}(T)\sqrt{\ln T}}{\tilde{T}}\sum_{t=1}^{\tilde{T}}\|\sigma_{t-1}(\boldsymbol{x}_t, W_j)\|_{\psi_2} \\
&\leq 2C\sqrt{\frac{\beta(T)(Q-1)\ln T}{\tilde{T}\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8\ln 12T\right\}}. \qquad \text{(C.37)}
\end{aligned}
$$

□

Finally, by using Lemma A.1 with Eq. (C.36), (C.37), and Lemma A.2, we have

$$
\Delta(T) \leq 2B(1-\underline{p})^{(1-\alpha)T} + \frac{2B}{T} + 2C\sqrt{\frac{\beta(T)(Q-1)\ln T}{\tilde{T}\ln Q}\left\{\frac{4\gamma(\tilde{T})}{\ln(1+\sigma^{-2})} + 8\ln(12T)\right\}}.
$$

C.2 Heteroscedastic Model Setting

Algorithm C.4 shows a pseudo-code of the MVR-based kernel-ETC algorithm in the heteroscedastic model setting. The differences between UCB-based and MVR-based kernel-ETC in the heteroscedastic model setting are as follows:

- The query strategy in the exploration period (Line 7 in Algoritm B.2 and Line 7 in Algorithm C.4).
- The query point $\hat{\boldsymbol{x}}^*$ in the exploitation period (Lines 15-16 in Algoritm B.2 and Line 15 in Algorithm C.4).
- The noise matrix used in the GP model of f: Algoritm B.2 plugs the UCB-based noise matrix $\hat{\boldsymbol{\Sigma}}^{(i)}$ into the GP model of f, whereas Algorithm C.4 uses the noise matrix $\lambda_f^2 \boldsymbol{I}_i$ with fixed variance parameter $\lambda_f > 0$. This modification is needed to guarantee that the query strategy of Algorithm C.4 is non-adaptive.

Algorithm C.4 The MVR-based kernel-ETC algorithm for heteroscedastic model setting.

Input: Kernel k_f, k_ρ, exploration ratio $\alpha \in (0,1]$, variance parameters λ_f, $\lambda_\rho > 0$, number of repetition $m \geq 2$, lower and upper bound of $\rho(\cdot)$: $\underline{\rho}$, $\overline{\rho}$, width of confidence bounds $\beta_f(T), \beta_\rho(T)$.

$\tilde{T} \leftarrow \lceil \alpha(T-1) \rceil$.
$t \leftarrow 1$.
$M \leftarrow \lfloor \tilde{T}/m \rfloor$.
Initialize GP prior of $f(\cdot)$ and $\rho(\cdot)$.
$\theta_T \leftarrow \tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t \in [T]} Z_t \right]$.
for $i = 1$ to M **do**
  $\boldsymbol{x}^{(i)} \leftarrow \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}} \left\{ \beta_f^{1/2}(T) \sigma_f^{(i-1)}(\boldsymbol{x} \mid \lambda_f^2 \boldsymbol{I}_{i-1}) + \beta_\rho^{1/2}(T) \theta_T \sigma_\rho^{(i-1)}(\boldsymbol{x}) \right\}$.
  for $j = 1, \ldots, m$ **do**
    $\boldsymbol{x}_t \leftarrow \boldsymbol{x}^{(i)}$.
    Observe $y_t = f(\boldsymbol{x}_t) + \eta_t(\boldsymbol{x}_t)$.
    $t \leftarrow t + 1$.
  end for
  Update GP posterior of $f(\cdot)$ and $\rho(\cdot)$.
end for
$\hat{\boldsymbol{x}}^* \leftarrow \operatorname{argmax}_{i \in [M]} \left\{ \mu_f^{(M)}(\boldsymbol{x}) + \theta_T \mu_\rho^{(M)}(\boldsymbol{x}) \right\}$.
for $t = mM + 1$ to T **do**
  $\boldsymbol{x}_t \leftarrow \hat{\boldsymbol{x}}^*$.
  Observe y_t.
end for

C.2.1 Regret Upper Bound of Algorithm C.4

We show the regret upper bound of Algorithm C.4 in the following Theorem C.1.

Theorem C.1. *Fix $\alpha \in (0,1]$, $m \geq 2$, $\mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| < \infty$, $f \in \mathcal{H}(k_f)$ with $\|f\|_{\mathcal{H}(k_f)} \leq B_f$, and $\rho \in \mathcal{H}(k_\rho)$ with $\|\rho\|_{\mathcal{H}(k_\rho)} \leq B_\rho$. Suppose $\forall \boldsymbol{x} \in \mathcal{X},\ \rho(\boldsymbol{x}) \in [\underline{\rho}, \overline{\rho}]$ holds for some $\overline{\rho}$ and $\underline{\rho}$ with $\overline{\rho} \geq \underline{\rho} > 0$. Then, when running Algorithm C.4 with $\beta_f^{1/2}(T) = \overline{\rho} m^{-1} \lambda_f^{-1} \sqrt{2 \ln(4|\mathcal{X}|T)} + B_f$ and $\beta_\rho^{1/2}(T) = c\kappa(m) \overline{\rho} \lambda_\rho^{-1} \sqrt{2 \ln(4|\mathcal{X}|T)} + B_\rho$, the following upper bound of extreme regret holds:*

$$\begin{aligned} \Delta(T) \leq & \frac{(2B + \overline{\rho}\sqrt{2 \ln T})}{T} + \sqrt{\frac{8m\beta_f(T)}{\alpha(T-m-1)\ln(1+\lambda_f^{-2})} \gamma_f \left(\frac{\alpha T + 1}{m} \right)} \\ & + \sqrt{\frac{16m\beta_\rho(T) \ln T}{\alpha(T-m-1)\ln(1+\lambda_\rho^{-2})} \gamma_\rho \left(\frac{\alpha T + 1}{m} \right)} + \alpha \overline{\rho} \left(\sqrt{2 \ln T} + \frac{1}{2\pi} + \frac{1}{2\sqrt{2\pi \ln T}} \right) + \frac{\overline{\rho}}{2^{(1-\alpha)T-2}}, \end{aligned} \tag{C.38}$$

where $\kappa(m) = (m-1)^{1/4}\Gamma((m-1)/2)/\Gamma(m/2)$, and $c > 0$ is an absolute constant. Especially, by setting $\alpha = T^\tau/T$ with $\tau \in (0,1)$, then

$$\Delta(T) = \mathcal{O}\left(\frac{B}{T} + \sqrt{\frac{m\beta_f(T^\tau/m)\gamma_f(T^\tau/m)}{(T^\tau - m)\ln(1 + m\overline{\rho}^{-2})}} + \frac{\overline{\rho} T^\tau \sqrt{\ln T}}{T} + \sqrt{\frac{m\beta_\rho(T^\tau/m)\gamma_\rho(T^\tau/m)\ln T}{(T^\tau - m)\ln(1 + mc^{-2}\kappa(m)^{-2}\overline{\rho}^{-2})}} \right).$$

To prove Theorem C.1, we first describe the following Lemma C.2 and Lemma C.3, which give tight confidence bounds of ρ and f under a non-adaptive learner's strategy, respectively.

Lemma C.2. *Fix any $\delta \in (0,1)$, $m \geq 2$, $M \in \mathbb{N}_+$, and $\rho \in \mathcal{H}(k_\rho)$. Suppose $\|\rho\|_{\mathcal{H}(k_\rho)} \leq B_\rho$ and $\forall \boldsymbol{x} \in \mathcal{X},\ \rho(\boldsymbol{x}) \in [\underline{\rho}, \overline{\rho}]$. Furthermore, assume the learner's decisions $(\boldsymbol{x}_t)_{t \in \mathbb{N}_+}$ are independent of $(\eta_t)_{t \in \mathbb{N}_+}$. Then, the following statement holds with probability at least $1 - \delta$:*

$$\forall \boldsymbol{x} \in \mathcal{X},\ |\rho(\boldsymbol{x}) - \mu_\rho^{(M)}(\boldsymbol{x})| \leq \left(B_\rho + \frac{c\kappa(m)\overline{\rho}}{\lambda_\rho} \sqrt{2 \ln \frac{2}{\delta}} \right) \sigma_\rho^{(M)}(\boldsymbol{x}), \tag{C.39}$$

where $\kappa(m) = (m-1)^{1/4}\Gamma((m-1)/2)/\Gamma(m/2)$, and $c > 0$ is an absolute constant.

Lemma C.3. *Fix any $\delta \in (0,1)$, $m \geq 2$, $M \in \mathbb{N}_+$, $f \in \mathcal{H}(k_f)$ with $\|f\|_{\mathcal{H}(k_f)} \leq B_f$, and $\rho(\cdot)$ with $\|\rho\|_\infty \leq \overline{\rho}$. Suppose that the learner's decisions $(\boldsymbol{x}_t)_{t\in\mathbb{N}_+}$ are independent of $(\eta_t)_{t\in\mathbb{N}_+}$. Then, the following statement holds with probability at least $1-\delta$:*

$$\forall \boldsymbol{x} \in \mathcal{X},\ |f(\boldsymbol{x}) - \mu_f^{(M)}(\boldsymbol{x} \mid \lambda_f^2 \boldsymbol{I}_M)| \leq \left(B_f + \frac{\overline{\rho}}{m\lambda_f} \sqrt{2 \ln \frac{2}{\delta}} \right) \sigma_f^{(M)}(\boldsymbol{x} \mid \lambda_f^2 \boldsymbol{I}_M).$$

It should be noted that the error terms $\hat{m}^{(i)} - f(\boldsymbol{x}^{(i)})$ and $\hat{s}^{(i)} - \rho(\boldsymbol{x}^{(i)})$, which are included into GP model of f and ρ, are $\overline{\rho}/m$ and $c\kappa(m)\overline{\rho}$-sub-Gaussian random variables, respectively (Lemma B.3). Thus, Lemma C.2 and Lemma C.3 are obtained by direct applications of Theorem 1 in Vakili et al. (2021).

Next, similar to Lemma B.7, the following Lemma C.4 gives the relation between the summations of posterior variances and the MIGs for our MVR-based algorithm.

Lemma C.4. *For any $m \geq 2$ and $M \in \mathbb{N}_+$, the following holds:*

$$\sum_{i=1}^{M} \sigma_f^{(i-1)}(\boldsymbol{x}^{(i)} \mid \lambda_f^2 \boldsymbol{I}_i) \leq \sqrt{\frac{2M\gamma_f(M)}{\ln(1+\lambda_f^{-2})}} \quad \text{and} \quad \sum_{i=1}^{M} \sigma_\rho^{(i-1)}(\boldsymbol{x}^{(i)}) \leq \sqrt{\frac{2M\gamma_\rho(M)}{\ln(1+\lambda_\rho^{-2})}}.$$

Lemma B.7 uses $\hat{\boldsymbol{\Sigma}}^{(i)}$ as the noise matrix for GP-model of f, whereas Lemma C.4 uses $\lambda_f^2 \boldsymbol{I}_i$, which is the same as standard homoscedastic GP-model with fixed noise variance parameter λ_f^2. Thus, Lemma C.4 is simply obtained from Lemma 5.3 in Srinivas et al. (2010).

We describe the proof of Theorem C.1 below.

Proof of Theorem C.1. From Lemma C.3 and Lemma C.2, the following event (C.40) holds with probability at least $1 - 1/T$:

$$\forall \boldsymbol{x} \in \mathcal{X},\ |f(\boldsymbol{x}) - \mu_f^{(M)}(\boldsymbol{x} \mid \lambda_f^2 \boldsymbol{I}_M)| \leq \beta_f^{1/2}(T)\sigma_f^{(M)}(\boldsymbol{x} \mid \lambda_f^2 \boldsymbol{I}_M) \ \text{ and } \ |\rho(\boldsymbol{x}) - \mu_\rho^{(M)}(\boldsymbol{x})| \leq \beta_\rho^{1/2}(T)\sigma_\rho^{(M)}(\boldsymbol{x}) \quad \text{(C.40)}$$

where $M = \lfloor \tilde{T}/m \rfloor$ with $\tilde{T} = \lceil \alpha(T-1) \rceil$. When the event (C.40) holds, we have,

$$\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t\in[T]} \{ f(\boldsymbol{x}^*) + \rho(\boldsymbol{x}^*) Z_t \} \right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t\in[T]} \{ f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*) Z_t \} \right] \\
&= f(\boldsymbol{x}^*) + \rho(\boldsymbol{x}^*) \tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t\in[T]} Z_t \right] - f(\hat{\boldsymbol{x}}^*) - \rho(\hat{\boldsymbol{x}}^*) \tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t\in[T]} Z_t \right] \\
&\leq \mu_f^{(M)}(\boldsymbol{x}^* \mid \lambda_f^2 \boldsymbol{I}_M) + \beta_f^{1/2}(T)\sigma_f^{(M)}(\boldsymbol{x}^* \mid \lambda_f^2 \boldsymbol{I}_M) - \mu_f^{(M)}(\hat{\boldsymbol{x}}^* \mid \lambda_f^2 \boldsymbol{I}_M) + \beta_f^{1/2}(T)\sigma^{(M)}(\hat{\boldsymbol{x}}^* \mid \lambda_f^2 \boldsymbol{I}_M) \\
&\quad + \left\{ \mu_\rho^{(M)}(\boldsymbol{x}^*) + \beta_\rho^{1/2}(T)\sigma_\rho^{(M)}(\boldsymbol{x}^*) \right\} \theta_T - \left\{ \mu_\rho^{(M)}(\hat{\boldsymbol{x}}^*) - \beta_\rho^{1/2}(T)\sigma_\rho^{(M)}(\hat{\boldsymbol{x}}^*) \right\} \theta_T \\
&\leq \beta_f^{1/2}(T) \left\{ \sigma_f^{(M)}(\boldsymbol{x}^* \mid \lambda_f^2 \boldsymbol{I}_M) + \sigma_f^{(M)}(\hat{\boldsymbol{x}}^* \mid \lambda_f^2 \boldsymbol{I}_M) \right\} + \beta_\rho^{1/2}(T) \left\{ \sigma_\rho^{(M)}(\boldsymbol{x}^*) + \sigma_\rho^{(M)}(\hat{\boldsymbol{x}}^*) \right\} \theta_T \\
&\leq 2\beta_f^{1/2}(T) \max_{\boldsymbol{x}\in\mathcal{X}} \sigma_f^{(M)}(\boldsymbol{x} \mid \lambda_f^2 \boldsymbol{I}_M) + 2\theta_T \beta_\rho^{1/2}(T) \max_{\boldsymbol{x}\in\mathcal{X}} \sigma_\rho^{(M)}(\boldsymbol{x}).
\end{aligned}$$

From the definition of $\boldsymbol{x}^{(i)}$ and monotonicity of posterior variances,

$$\begin{aligned}
&\beta_f^{1/2}(T) \max_{\boldsymbol{x}\in\mathcal{X}} \sigma_f^{(M)}(\boldsymbol{x} \mid \lambda_f^2 \boldsymbol{I}_M) + \theta_T \beta_\rho^{1/2}(T) \max_{\boldsymbol{x}\in\mathcal{X}} \sigma_\rho^{(M)}(\boldsymbol{x}) \\
&\leq \beta_f^{1/2}(T)\sigma_f^{(i-1)}(\boldsymbol{x}^{(i)} \mid \lambda_f^2 \boldsymbol{I}_{i-1}) + \theta_T \beta_\rho^{1/2}(T)\sigma_\rho^{(i-1)}(\boldsymbol{x}^{(i)})
\end{aligned}$$

for any $i \in [M]$. Thus, we have,

$$\begin{aligned}
&\tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t\in[T]} \{ f(\boldsymbol{x}^*) + \rho(\boldsymbol{x}^*) Z_t \} \right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}} \left[\max_{t\in[T]} \{ f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*) Z_t \} \right] \\
&\leq \frac{2\beta_f^{1/2}(T)}{M} \sum_{i=1}^{M} \sigma_f^{(i-1)}(\boldsymbol{x}^{(i)} \mid \lambda_f^2 \boldsymbol{I}_{i-1}) + \frac{2\theta_T \beta_\rho^{1/2}(T)}{M} \sum_{i=1}^{M} \sigma_\rho^{(i-1)}(\boldsymbol{x}^{(i)}) \\
&\leq \sqrt{\frac{8\beta_f(T)\gamma_f(M)}{M\ln(1+\lambda_f^{-2})}} + \sqrt{\frac{16\beta_\rho(T)\gamma_\rho(M)\ln T}{M\ln(1+\lambda_\rho^{-2})}}. \qquad \text{(C.41)}
\end{aligned}$$

Algorithm D.5 The computation of $\tilde{\mathbb{E}}_{\boldsymbol{W}}[\max_{j\in[T]} g(\boldsymbol{x}, W_j)]$.

Input: Input $\boldsymbol{x}$, function g, total step size T, probability mass function $p(\boldsymbol{w})$.
Output: $\tilde{\mathbb{E}}_{\boldsymbol{W}}[\max_{j\in[T]} g(\boldsymbol{x}, W_j)]$.
1: $\mathcal{Y} \leftarrow \{g(\boldsymbol{x}, \boldsymbol{w}) \mid \boldsymbol{w} \in \mathcal{W}\}$.
2: Calculate the probability mass function $\tilde{p} : \mathcal{Y} \to [0, 1]$ of $g(\boldsymbol{x}, W_1)$ as $\tilde{p}(y) = \sum_{\boldsymbol{w}\in\mathcal{W}; g(\boldsymbol{x},\boldsymbol{w})=y} p(\boldsymbol{w})$.
3: Sort and index the elements of $\mathcal{Y}$ as $y_1, \ldots, y_{|\mathcal{Y}|}$ by descending order.
4: $g_{\max} \leftarrow 0$, $p_{\text{total}} \leftarrow 0$.
5: **for** $i = 1$ to $|\mathcal{Y}| - 1$ **do**
6: $\quad p_{\max} \leftarrow 1 - \left[\sum_{l=i+1}^{|\mathcal{Y}|} \tilde{p}(y_l)\right]^T - p_{\text{total}}$.
7: $\quad g_{\max} \leftarrow g_{\max} + p_{\max} y_i$.
8: $\quad p_{\text{total}} \leftarrow p_{\text{total}} + p_{\max}$.
9: **end for**
10: $g_{\max} \leftarrow g_{\max} + (1 - p_{\text{total}}) y_{|\mathcal{Y}|}$.
11: **return** $g_{\max}$.

In Eq. (C.41), we use Lemma C.4 and the upper bound of the expectation for T standard normal random variables: $\theta_T \le \sqrt{2 \ln T}$. Furthermore, due to $\alpha(T - m - 1)/m \le M \le (\alpha T + 1)/m$,

$$\begin{aligned} &\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]}\{f(\boldsymbol{x}^*) + \rho(\boldsymbol{x}^*) Z_t\}\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]}\{f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*) Z_t\}\right] \\ &\le \sqrt{\frac{8m\beta_f(T)\gamma_f(\frac{\alpha T+1}{m})}{\alpha(T-m-1)\ln(1+\lambda_f^{-2})}} + \sqrt{\frac{16m\beta_\rho(T)\gamma_\rho(\frac{\alpha T+1}{m})\ln T}{\alpha(T-m-1)\ln(1+\lambda_\rho^{-2})}} \end{aligned} \tag{C.42}$$

By noting the event (C.40) holds with probability at least $1 - 1/T$, we have,

$$\begin{aligned} &\mathbb{E}\left[\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]}\{f(\boldsymbol{x}^*) + \rho(\boldsymbol{x}^*) Z_t\}\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]}\{f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*) Z_t\}\right]\right] \\ &\le \frac{2B_f + \overline{\rho}\sqrt{2\ln T}}{T} + \mathbb{E}\left[\mathbb{1}\{\text{(C.40) is true}\}\left\{\tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]}\{f(\boldsymbol{x}^*) + \rho(\boldsymbol{x}^*) Z_t\}\right] - \tilde{\mathbb{E}}_{\boldsymbol{Z}}\left[\max_{t\in[T]}\{f(\hat{\boldsymbol{x}}^*) + \rho(\hat{\boldsymbol{x}}^*) Z_t\}\right]\right\}\right] \end{aligned} \tag{C.43}$$

$$\le \frac{2B_f + \overline{\rho}\sqrt{2\ln T}}{T} + \sqrt{\frac{8m\beta_f(T)\gamma_f(\frac{\alpha T+1}{m})}{\alpha(T-m-1)\ln(1+\lambda_f^{-2})}} + \sqrt{\frac{16m\beta_\rho(T)\gamma_\rho(\frac{\alpha T+1}{m})\ln T}{\alpha(T-m-1)\ln(1+\lambda_\rho^{-2})}}. \tag{C.44}$$

where:

- Eq. (C.43) follows from the fact that $\|f\|_\infty \le \|f\|_{\mathcal{H}(k)} \le B_f$, $\|\rho\|_\infty \le \overline{\rho}$, and $\theta_T \le \sqrt{2\ln T}$.
- Eq. (C.44) follows from Eq. (C.42).

By combining Lemma B.4 with Eq. (C.44) and Lemma B.9, Eq. (C.38) is obtained. □

D COMPUTATIONAL DETAILS

D.1 Computation of the Maximum in Environmental Model Setting

Our kernel-ETC algorithm for environmental model setting requires us to compute the maximum of the expectation, whose form is $\tilde{\mathbb{E}}_{\boldsymbol{W}}[\max_{j\in[T]} g(\boldsymbol{x}, W_j)]$ for some function g. For example, $\tilde{\mathbb{E}}_{\boldsymbol{W}}[\max_{j\in[T]} \mathrm{ucb}_t(\boldsymbol{x}, W_j)]$ and $\tilde{\mathbb{E}}_{\boldsymbol{W}}[\max_{j\in[T]} \mathrm{lcb}_{\bar{T}}(\boldsymbol{x}, W_j)]$ are needed to compute Line 3 and Line 6 in Algorithm 1, respectively. We can obtain $\tilde{\mathbb{E}}_{\boldsymbol{W}}[\max_{j\in[T]} g(\boldsymbol{x}, W_j)]$ exactly by calculating the probability mass function of $\max_{j\in[T]} g(\boldsymbol{x}, W_j)$. We show the details in Algorithm D.5.

D.2 Computation of θ_T in Heteroscedastic Model Setting

To our knowledge, the way to compute $\theta_T \coloneqq \mathbb{E}_{\boldsymbol{Z}}[\max_{j\in[T]} Z_t]$ exactly for any $T \in \mathbb{N}_+$ is unknown. Thus, we need to resort to some approximation method to estimate θ_T. In our experiments, we adopt the Monte-Carlo estimate of θ_T with 1000 samples. It should be noted that the estimation of θ_T is only required at the beginning of our algorithm only once, so the computational time of each step is not affected.

E DETAILS OF EXPERIMENTS

E.1 Methods

We describe the details of the methods that are used in numerical experiments.

MVABO (Iwazaki et al., 2021a) : As described in Sec. 5, we use the modified version of the original algorithm. Original MVABO is formulated to maximize scalarized objective function $G(\boldsymbol{x}) \coloneqq b\mathbb{E}_{\boldsymbol{w}}[f(\boldsymbol{x},\boldsymbol{w})] - (1-b)\sqrt{\mathbb{V}_{\boldsymbol{w}}[f(\boldsymbol{x},\boldsymbol{w})]}$, where b is a parameter that controls the balance between the mean and variance. To adapt our risk-seeking setting, we consider the modified objective function $\tilde{G}(\boldsymbol{x}) \coloneqq b\mathbb{E}_{\boldsymbol{w}}[f(\boldsymbol{x},\boldsymbol{w})] + (1-b)\sqrt{\mathbb{V}_{\boldsymbol{w}}[f(\boldsymbol{x},\boldsymbol{w})]}$. Original MVABO chooses $\boldsymbol{x}_t$ as $\boldsymbol{x}_t \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} b\,\mathrm{ucb}^{(e)}(\boldsymbol{x}) - (1-b)\sqrt{\mathrm{ucb}^{(v)}(\boldsymbol{x})}$, where $\mathrm{ucb}^{(e)}(\boldsymbol{x})$ and $\mathrm{ucb}^{(v)}(\boldsymbol{x})$ are UCB of $\tilde{\mathbb{E}}_{\boldsymbol{W}}[f(\boldsymbol{x},W_1)]$ and $\tilde{\mathbb{V}}_{\boldsymbol{W}}[f(\boldsymbol{x},W_1)]$. See Lemma 3.1 in Iwazaki et al. (2021a) for the precise definitions of $\mathrm{ucb}^{(e)}(\boldsymbol{x})$ and $\mathrm{ucb}^{(v)}(\boldsymbol{x})$. We extend the query strategy of the original MVABO to the maximization strategy of the modified objective function $\tilde{G}$ by flipping the sign before the variance. Namely, we choose $\boldsymbol{x}_t$ as $\boldsymbol{x}_t \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} b\mathrm{ucb}^{(e)}(\boldsymbol{x}) + (1-b)\sqrt{\mathrm{ucb}^{(v)}}(\boldsymbol{x})$. We set $b = 1.0$ and $b = 0.0$ in `Mean-RAHBO` and `Variance-RAHBO`, respectively. Finally, we set the parameter β_t, which specifies the width of the confidence interval, as $\beta_t^{1/2} = 3$.

RAHBO (Makarova et al., 2021) : As with the MVABO, we modify the original query strategy $\boldsymbol{x}_t \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \mathrm{ucb}_t^f(\boldsymbol{x}) - b\mathrm{lcb}_t^{\mathrm{var}}(\boldsymbol{x})$ to $\boldsymbol{x}_t \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \mathrm{ucb}_t^f(\boldsymbol{x}) + b\mathrm{ucb}_t^{\mathrm{var}}(\boldsymbol{x})$. The function $\mathrm{ucb}_t^f(\boldsymbol{x})$, $\mathrm{ucb}_t^{\mathrm{var}}(\boldsymbol{x})$, and $\mathrm{lcb}_t^{\mathrm{var}}(\boldsymbol{x})$ are defined as the UCB of f, the UCB of ρ^2, and the LCB of ρ^2, respectively. See Makarova et al. (2021) for the precise definitions. The above modifications of the query strategy correspond to considering the modified objective function $\tilde{\mathrm{MV}}(\boldsymbol{x}) \coloneqq f(\boldsymbol{x}) + b\rho^2(\boldsymbol{x})$ of RAHBO, which is different from the original objective function $\mathrm{MV}(\boldsymbol{x}) = f(\boldsymbol{x}) - b\rho^2(\boldsymbol{x})$. In `Mean-RAHBO`, we set $b = 0$. In `Variance-RAHBO`, we only focus on the variance term and choose $\boldsymbol{x}_t$ as $\boldsymbol{x}_t \in \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{X}} \mathrm{ucb}_t^{\mathrm{var}}(\boldsymbol{x})$. We set β_t and β_t^{var}, as $\beta_t = 3$ and $\beta_t^{\mathrm{var}} = 3$, respectively. These parameters specify the width of the confidence intervals of f and ρ^2 and are required to calculate $\mathrm{ucb}_t^f(\boldsymbol{x})$, $\mathrm{ucb}_t^{\mathrm{var}}(\boldsymbol{x})$, and $\mathrm{lcb}_t^{\mathrm{var}}(\boldsymbol{x})$, respectively. Finally, we set the variance parameter of the GP model of ρ^2 as $2\overline{\rho}^4/(m-1)$ by following the discussions of Appendix A.2 in Makarova et al. (2021).

GP-UCB (Srinivas et al., 2010) : We adopt the standard GP-UCB in the heteroscedastic model settings. As described in Sec. 5, we set the noise parameter of GP as $\overline{\rho}^2$. Furthermore, we set the confidence width parameter $\beta^{1/2}(t)$ as $\beta^{1/2}(t) = 3$.

Kernel-ETC and MVR-based kernel-ETC : In environmental model settings, we set $\beta^{1/2}(t)$ as $\beta^{1/2}(t) = 3$. Similarly, in heteroscedastic model setting, we set $\beta_f^{1/2}(t)$ and $\beta_\rho^{1/2}(t)$ as $\beta_f^{1/2}(t) = 3$ and $\beta_\rho^{1/2}(t) = 3$, respectively. Moreover, we set the $\lambda_\rho = \kappa(m)\overline{\rho}/4$. Furthermore, in standard kernel-ETC for both settings, we define $\hat{\boldsymbol{x}}^*$ as the same as in the MVR-based algorithms. That is, we define $\hat{\boldsymbol{x}}^*$ as in Line 6 of Algorithm C.3 and Line 15 of Algorithm C.4 in environmental and heteroscedastic model settings, respectively. In practice, such modifications are also made in the estimated solutions of existing UCB-based algorithms to improve the performance (Nguyen et al., 2021b).

Finally, except for the hyperparameter tuning experiment of CNN, we set $\overline{\rho}$ and $\underline{\rho}$ as $\overline{\rho} = \max_{\boldsymbol{x}\in\mathcal{X}} \rho(\boldsymbol{x})$ and $\underline{\rho} = \min_{\boldsymbol{x}\in\mathcal{X}} \rho(\boldsymbol{x})$, respectively. In the CNN tuning experiments, we first select 100 sets of hyperparameters over $\mathcal{X}$ uniformly at random. Then, we compute the unbiased standard deviation $\hat{s}^{(i)}$ with $m = 3$ by training the CNN in such 100 sets of hyperparameters, and set $\overline{\rho}$ and $\underline{\rho}$ as the maximum and minimum of the computed $\hat{s}^{(i)}$, respectively.

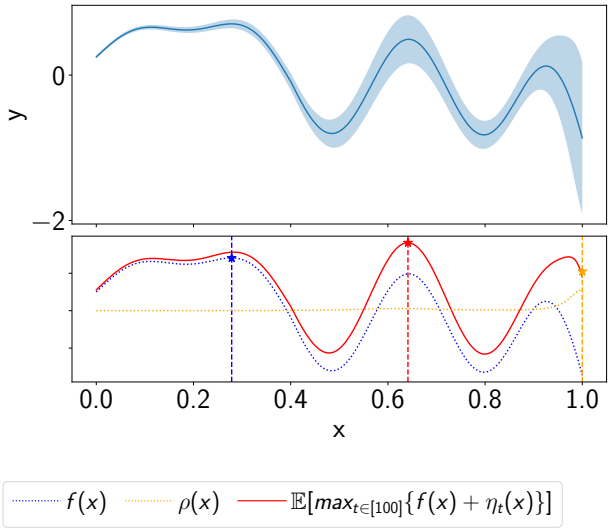

Figure E.4: The visualization of the synthetic function used in the heteroscedastic model setting. In the top figure, the blue line shows f, and the widths of the blue shaded areas are defined as $2\rho(\boldsymbol{x})$. From the bottom figure, we can see that the maximum of $f(\boldsymbol{x})$, $\rho(\boldsymbol{x})$, and $\mathbb{E}[\max_{t\in[100]}\{f(\boldsymbol{x})+\eta_t(\boldsymbol{x})\}]$ are different; thus, `Mean-RAHBO`, `Variance-RAHBO`, and `kernel-ETC` are expected to show different behaviors in this problem.

.

E.2 Details of Objective Functions

E.2.1 Synthetic Benchmark Functions

The synthetic benchmark function experiments in the environmental model setting are conducted with the following function $f_{\mathrm{env}} : [0,1]^2 \to \mathbb{R}$:

$$f_{\mathrm{env}}(x,w) = 0.75xw^{15x} + 0.5\max\{1-x, 0.5\} + 0.05\sin(10w+x) - \min\{x, 1-x\}\sin(9w) - 0.25.$$

We set the probability mass $p(\boldsymbol{w})$ as $p_{\mathrm{Gaussian}} \coloneqq \phi(w)/\sum_{w\in\mathcal{W}}\phi(w)$ in this experiment. The shapes of f are depicted in Fig. 1.

Furthermore, we create the synthetic mean function $f_{\mathrm{hetero}} : [0,1] \to \mathbb{R}$ and the variance function $\rho^2_{\mathrm{hetero}} : [0,1] \to \mathbb{R}_+$ for the heteroscedastic model setting respectively as follows:

$$f_{\mathrm{hetero}}(x) = 2.5\min\{x-0.4, 0.0\} + 0.5\sin(10x) + 2.25(1-x) + x\cos(20x) - 1.0,$$
$$\rho^2_{\mathrm{hetero}}(x) = \left(10^{-4} + \frac{0.4}{|10(0.62-x)|^2+2.5} + \frac{1}{|30(1-x)|^2+2.0}\right)^2.$$

Figure E.4 shows the visualization of f_{hetero} and ρ^2_{hetero}.

In synthetic function experiments, we set the kernel hyperparameters as $\ell = 0.2$ and $\sigma^2_{\mathrm{ker}} = 1.0$ for both settings.

E.2.2 2D-RKHS Test Functions

We construct the test function as with the experiments in Chowdhury and Gopalan (2017). We first make 50 pairs of training inputs and function values. Training inputs are chosen uniformly at random in two-dimensional unit hypercube $[0,1]^2$, and their corresponding function values are generated from GP. After that, we fit the GP model by using these 50 training input and function value pairs and then use the GP posterior mean as a test function. Moreover, in generating the test function of ρ in heteroscedastic variance model setting, we shift the function values by adding $\underline{\rho} - \rho_{\min}$ if $\rho_{\min} < \underline{\rho}$, where $\rho_{\min}$ is the minimum of the generated test function, and $\rho = 10^{-4}$. To generate the test functions, we adopt the SE kernel with $\sigma^2_{\mathrm{ker}} = 1$ and $\ell = 0.2$.

E.2.3 Simulation Function of Polymer Synthesis

The synthetic benchmark function is based on real data from polymer syntehsis. The objective function $f_{\mathrm{sim}} : \mathcal{X} \times \mathcal{W} \to \mathbb{R}$ has a controllable parameter $x \in \mathcal{X}$, an uncontrollable parameter $w \in \mathcal{W}$, and one output value. We define the controllable and uncontrollable parameter sets as $\mathcal{X} = \{(i-1)/19 \mid i \in [20]\}$ and $\mathcal{W} = \{(i-1)/9 \mid i \in [10]\}$, respectively. The probability mass $p(w)$ is assumed to be uniform, i.e., $p(w) = 1/10$ for all $w \in \mathcal{W}$. We use the following modeling for the function f_{sim}:

$$f_{\mathrm{sim}}(x, w) = \frac{T_g(x, w) - 400}{15}, \tag{E.45}$$

$$T_g(x, w) = T_{g,A}(45w + 5) \cdot (1 - x) + 410\, x + q(45w + 5) \cdot (1 - x)x, \tag{E.46}$$

$$T_{g,A}(z) = 374.374 + 0.815146\, z - 0.0215356\, z^2 + 0.000269113\, z^3, \tag{E.47}$$

$$q(z) = 4.94286 + 3.71676\, z - 0.0906406\, z^2 + 0.000778145\, z^3. \tag{E.48}$$

The problem is to maximize f_{sim} in the environmental model setting. In this experiment, we use $\sigma^2_{\mathrm{ker}} = 1$ and $\ell = 0.2$.

The form of f_{sim} is obtained from the following considerations. We consider a hypothetical experiment involving the blending of two ingredients A and B whose respective fractions are given by w_A and w_B. It is assumed that ingredient A is further made up of its subcomponents. One subcomponent is assumed to be important since it affects a certain target property of the final mixture, which we wish to optimize. We will call the fraction of this subcomponent within ingredient A to be w_s. We assume that the subcomponent fraction w_s has an uncertainty that arises from an uncontrollable parameter in the manufacturing process. Since the weights must sum to unity, i.e., $w_A + w_B = 1$, only one of them is truly an independent variable. Without loss of generality, we choose w_B as the independent variable. The goal of this experiment is to maximize the objective function f_{sim} that models a certain property of the final mixture. We use the data from Belabed et al. (2012) which reports on the glass transition temperature T_g in the blends of poly(4-vinylphenol-*co*-methyl methacrylate) (PSMA) and poly(styrene-*co*-4-vinylpyridine) (PS4VP). The T_g values are reported for various fractions of PSMA and PS4VP and for different values of the subcomponent of PS4VP, namely the fraction of 4VP, $w_{4\mathrm{VP}}$. In our numerical experiment, we make the identification of $w_B = w_{\mathrm{PSMA}}$ and $w_s = w_{4\mathrm{VP}}$. The objective function to maximize is T_g of the final blend. We follow the paper's use of the Kwei equation (Kwei, 1984) to model T_g,

$$T_g = \frac{w_A T_{g,A} + k w_B T_{g,B}}{w_A + k w_B} + q w_A w_B. \tag{E.49}$$

The formula takes into account the intermolecular interactions which make the T_g model less trivial compared to the classical Fox equation. The parameters $k = 1$ and $T_{g,B} = 410\,\mathrm{K}$ are fixed using data. The subcomponent fraction w_s affects both the parameters q and $T_{g,A}$. We perform a fit to the data using a third-degree polynomial function to obtain the expressions for q and $T_{g,A}$ in terms of w_s. The objective function is identified as $f_{\mathrm{sim}} \coloneqq (T_g - 400)/15$ by rescaling the output. After substituting $x \coloneqq w_B$ and $w \coloneqq (w_s - 5)/45$ for input normalization, we obtain Eqs. (E.45)–(E.48).

E.2.4 Hyperparameter Tuning of CNN

We build a 3-layer CNN, whose first and third layers are convolution layers with 8 channels, and the second layer is the 2×2 max-pooling layer. The input domain $\mathcal{X}$ is defined as $\mathcal{X} = \mathcal{X}_{\mathrm{lr}} \times \mathcal{X}_{\mathrm{bs}} \times \mathcal{X}_{\mathrm{epoch}}$, where $\mathcal{X}_{\mathrm{lr}}$, $\mathcal{X}_{\mathrm{bs}}$, and $\mathcal{X}_{\mathrm{epoch}}$ are the domain of learning rate, batch size, and epoch, respectively. We define $\mathcal{X}_{\mathrm{lr}}$, $\mathcal{X}_{\mathrm{bs}}$, and $\mathcal{X}_{\mathrm{epoch}}$ respectively as

$$\begin{aligned}
\mathcal{X}_{\mathrm{lr}} &= \{10^{-5+4.5\frac{x-1}{19}} \mid x \in [20]\}, \\
\mathcal{X}_{\mathrm{bs}} &= \{\lfloor 2^{7(x-1)/19} \rfloor \mid x \in [20]\}, \\
\mathcal{X}_{\mathrm{epoch}} &= [20].
\end{aligned}$$

When running the experiments, we transform the original validation error in log scale and multiply the log-scaled validation error by -1. Then, we define this negative log-scaled validation error as the reward y_t. To simplify the experiments, we use randomly chosen 5000 (2500) training (validation) data from the original 50000 CIFAR-10 data.

Finally, we set σ^2_{ker} and l of k_f as $\sigma^2_{\mathrm{ker}} = 1.0$ and $\ell = 0.5$, respectively. For k_ρ, we set $\sigma^2_{\mathrm{ker}} = 0.1$ and $\ell = 0.5$.

Table E.2: The average extreme regret of the experiments with synthetic benchmark function in the environmental model setting. The numbers in parentheses correspond to one standard error.

	T=50	T=100	T=150	T=200
Random	0.341 (0.024)	0.276 (0.021)	0.197 (0.018)	0.121 (0.013)
Mean-MVABO	0.299 (0.030)	0.317 (0.029)	0.318 (0.029)	0.318 (0.029)
Variance-MVABO	0.240 (0.028)	0.169 (0.022)	0.162 (0.021)	0.162 (0.021)
kernel-ETC ($\alpha = 0.75$)	0.246 (0.026)	**0.082** (0.017)	**0.021** (0.009)	**0.000** (0.000)
kernel-ETC ($\alpha = 0.95$)	**0.184** (0.025)	**0.039** (0.013)	**0.000** (0.000)	**0.000** (0.000)
MVR-kernel-ETC ($\alpha = 0.75$)	0.234 (0.026)	0.131 (0.019)	**0.029** (0.010)	**0.000** (0.000)
MVR-kernel-ETC ($\alpha = 0.95$)	0.209 (0.026)	**0.082** (0.017)	**0.003** (0.003)	**0.000** (0.000)

Table E.3: The average extreme regret of the RKHS test function experiments with p_{uniform} in the environmental model setting.

	T=50	T=100	T=150	T=200
Random	0.198 (0.043)	0.096 (0.027)	0.061 (0.018)	0.039 (0.012)
Mean-MVABO	0.156 (0.032)	0.127 (0.025)	0.124 (0.024)	0.123 (0.024)
Variance-MVABO	0.130 (0.036)	0.055 (0.021)	0.051 (0.020)	0.051 (0.020)
kernel-ETC ($\alpha = 0.75$)	0.065 (0.033)	**0.001** (0.001)	**0.000** (0.001)	**0.000** (0.000)
kernel-ETC ($\alpha = 0.95$)	**0.061** (0.030)	**0.002** (0.004)	**0.000** (0.001)	**0.000** (0.000)
MVR-kernel-ETC ($\alpha = 0.75$)	0.103 (0.034)	**0.014** (0.014)	**0.006** (0.008)	**0.001** (0.001)
MVR-kernel-ETC ($\alpha = 0.95$)	0.128 (0.038)	0.048 (0.023)	0.020 (0.014)	**0.014** (0.013)

E.3 Details of Experimental Results

We give the details of the experimental results, including the standard errors. As for the environmental model settings experiments, Tabs. E.2, E.3, E.4, and E.5 show the results of the synthetic benchmark function, RKHS test function with p_{uniform}, RKHS test function with p_{Gaussian}, and polymer synthesis simulation function, respectively. Furthermore, Tabs. E.6 and E.7 show the experiment results of the synthetic benchmark function and RKHS test functions in the heteroscedastic model setting, respectively.

F SENSITIVITY ANALYSIS

By using the synthetic function used in the experiments of Sec. 5, we analyze the performance sensitivity of kernel-ETC with respect to the parameters α and τ. We conduct experiments of kernel-ETC by setting α and τ as $\alpha, \tau = \{0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$. Other settings are the same as the experiments with synthetic benchmark functions in Sec. 5. Figures F.5 and F.6 show the results in the environmental and heteroscedastic model setting, respectively. In the environmental model setting, we can confirm that our algorithms work well in various parameter settings. On the other hand, in the heteroscedastic setting, we find that values of τ between 0.8 and 0.9 work well, whereas values of $\tau = 0.70, 0.75, 0.95$ have worse performance than $\tau = 0.8, 0.85, 0.90$ in both kernel-ETC and MVR-based variants.

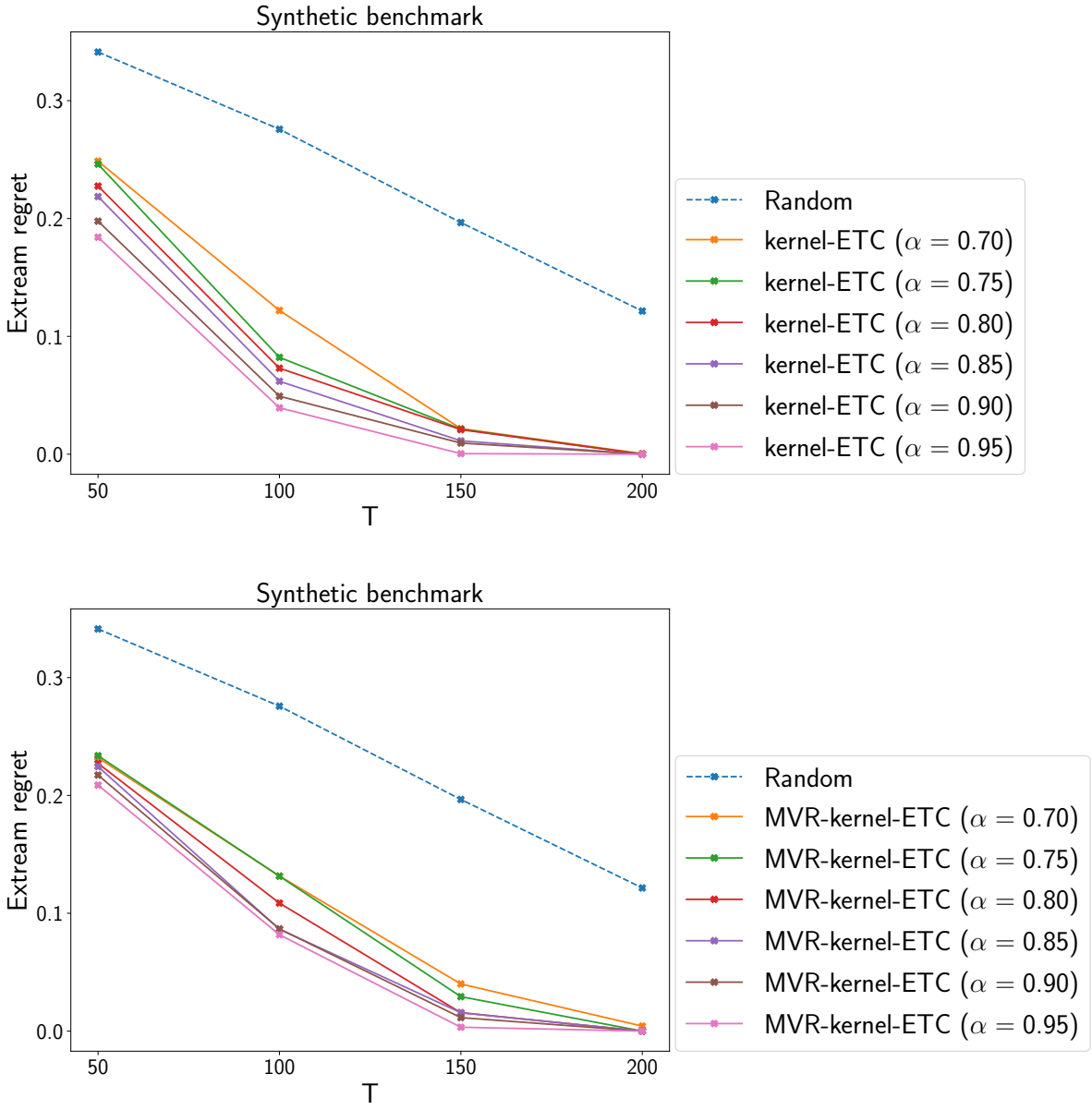

Figure F.5: The average extreme regrets with kernel-ETC (top) and MVR-based kernel-ETC (top) in the environmental model setting.

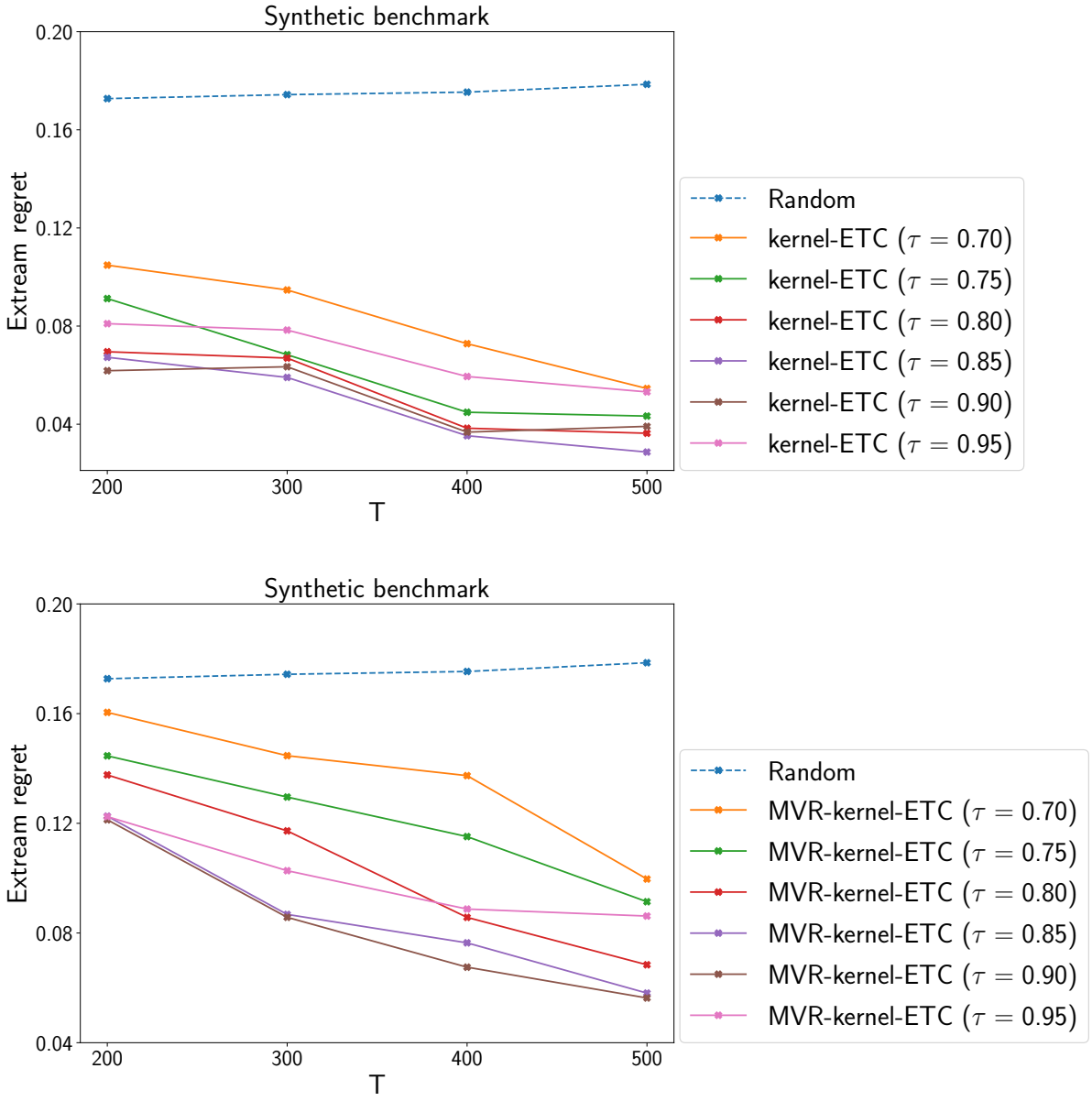

Figure F.6: The average extreme regrets with kernel-ETC (top) and MVR-based kernel-ETC (top) in the heteroscedastic model setting.

Table E.4: The average extreme regret of the RKHS test function experiments with p_{Gauss} in the environmental model setting.

	T=50	T=100	T=150	T=200
Random	0.205 (0.043)	0.104 (0.027)	0.063 (0.018)	0.042 (0.014)
Mean-MVABO	0.161 (0.036)	0.129 (0.028)	0.128 (0.028)	0.127 (0.028)
Variance-MVABO	0.148 (0.038)	0.068 (0.024)	0.061 (0.023)	0.058 (0.023)
kernel-ETC ($\alpha = 0.75$)	**0.042** (0.024)	**0.001** (0.001)	**0.000** (0.000)	**0.000** (0.000)
kernel-ETC ($\alpha = 0.95$)	**0.045** (0.023)	**0.001** (0.001)	**0.000** (0.000)	**0.000** (0.000)
MVR-kernel-ETC ($\alpha = 0.75$)	0.105 (0.035)	0.023 (0.020)	**0.003** (0.004)	**0.001** (0.001)
MVR-kernel-ETC ($\alpha = 0.95$)	0.152 (0.043)	0.062 (0.029)	0.028 (0.018)	**0.019** (0.016)

Table E.5: The average extreme regret of the experiments with polymer synthesis simulation function in the environmental model setting.

	T=25	T=50	T=75	T=100
Random	0.068 (0.008)	0.043 (0.005)	0.028 (0.004)	0.017 (0.003)
Mean-MVABO	0.061 (0.006)	0.028 (0.004)	0.014 (0.002)	0.012 (0.001)
Variance-MVABO	0.094 (0.008)	0.049 (0.005)	0.030 (0.004)	0.023 (0.003)
kernel-ETC ($\alpha = 0.75$)	**0.028** (0.005)	**0.016** (0.003)	**0.005** (0.001)	**0.001** (0.000)
kernel-ETC ($\alpha = 0.95$)	**0.043** (0.006)	**0.020** (0.003)	**0.006** (0.001)	**0.002** (0.001)
MVR-kernel-ETC ($\alpha = 0.75$)	0.051 (0.007)	0.026 (0.004)	0.010 (0.003)	**0.007** (0.002)
MVR-kernel-ETC ($\alpha = 0.95$)	0.063 (0.007)	0.038 (0.005)	0.021 (0.004)	0.013 (0.003)

Table E.6: The average extreme regret of the experiments with synthetic benchmark function in the heteroscedastic model setting.

	T=100	T=200	T=300	T=400
Random	0.157 (0.005)	0.173 (0.006)	0.174 (0.006)	0.175 (0.006)
GP-UCB	0.142 (0.003)	0.165 (0.003)	0.181 (0.003)	0.188 (0.003)
Mean-RAHBO	0.141 (0.005)	0.162 (0.005)	0.168 (0.005)	0.177 (0.005)
Variance-RAHBO	0.165 (0.005)	0.183 (0.005)	0.189 (0.006)	0.178 (0.009)
kernel-ETC ($\tau = 0.75$)	**0.121** (0.007)	**0.091** (0.009)	**0.068** (0.007)	**0.045** (0.008)
kernel-ETC ($\tau = 0.95$)	**0.104** (0.006)	**0.081** (0.007)	**0.078** (0.007)	**0.059** (0.007)
MVR-kernel-ETC ($\tau = 0.75$)	0.172 (0.005)	**0.145** (0.009)	**0.130** (0.009)	**0.115** (0.010)
MVR-kernel-ETC ($\tau = 0.95$)	0.141 (0.005)	**0.122** (0.008)	**0.103** (0.008)	**0.089** (0.008)

Table E.7: The average extreme regret of the experiments with RKHS test function in the heteroscedastic model setting.

	T=100	T=200	T=300	T=400	T=500
Random	4.088 (0.365)	4.203 (0.331)	4.201 (0.324)	4.003 (0.332)	3.942 (0.328)
GP-UCB	3.869 (0.370)	3.861 (0.318)	3.858 (0.292)	3.758 (0.310)	3.701 (0.331)
Mean-RAHBO	3.661 (0.353)	3.783 (0.351)	3.950 (0.349)	3.953 (0.340)	3.927 (0.370)
Variance-RAHBO	5.004 (0.415)	5.280 (0.382)	5.443 (0.424)	5.444 (0.456)	5.390 (0.441)
kernel-ETC ($\tau = 0.75$)	3.777 (0.416)	3.520 (0.367)	**3.136** (0.400)	**2.788** (0.458)	**2.305** (0.437)
kernel-ETC ($\tau = 0.95$)	3.183 (0.399)	**2.320** (0.310)	**2.320** (0.336)	**1.631** (0.357)	**1.346** (0.387)
MVR-kernel-ETC ($\tau = 0.75$)	3.858 (0.403)	3.456 (0.413)	**3.166** (0.393)	**2.993** (0.441)	**2.609** (0.484)
MVR-kernel-ETC ($\tau = 0.95$)	3.643 (0.372)	3.277 (0.353)	**3.114** (0.339)	**2.673** (0.410)	**2.520** (0.394)