Risk-averse Heteroscedastic Bayesian Optimization

Anastasiia Makarova
ETH Zürich
anmakaro@ethz.ch

Ilnura Usmanova
ETH Zürich
ilnurau@ethz.ch

Ilija Bogunovic
ETH Zürich
ilijab@ethz.ch

Andreas Krause
ETH Zürich
krausea@ethz.ch

Abstract

Many black-box optimization tasks arising in high-stakes applications require risk-averse decisions. The standard Bayesian optimization (BO) paradigm, however, optimizes the expected value only. We generalize BO to *trade mean and input-dependent variance* of the objective, both of which we assume to be unknown a priori. In particular, we propose a novel risk-averse heteroscedastic Bayesian optimization algorithm (RAHBO) that aims to identify a solution with high return and low noise variance, while learning the noise distribution on the fly. To this end, we model both expectation and variance as (unknown) RKHS functions, and propose a novel risk-aware acquisition function. We bound the regret for our approach and provide a robust rule to report the final decision point for applications where only a single solution must be identified. We demonstrate the effectiveness of RAHBO on synthetic benchmark functions and hyperparameter tuning tasks.

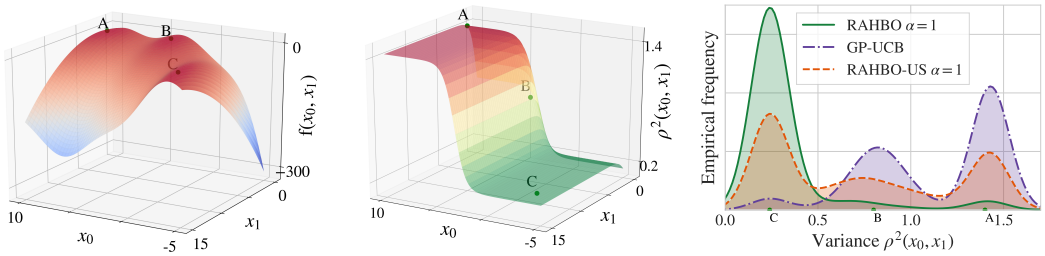
1 Introduction

Black-box optimization tasks arise frequently in high-stakes applications such as drug and material discovery [22, 17, 28], genetics [16, 27], robotics [5, 12, 25], hyperparameter tuning of complex learning systems [21, 13, 34], to name a few. In many of these applications, there is often a trade-off between achieving high utility and minimizing risk. Moreover, uncertain and costly evaluations are an inherent part of black-box optimization tasks, and modern learning methods need to handle these aspects when balancing between the previous two objectives.

Bayesian optimization (BO) is a powerful framework for optimizing such costly black-box functions from noisy zeroth-order evaluations. Classical BO approaches are typically *risk-neutral* as they seek to optimize the expected function value only. In practice, however, two different solutions might attain similar expected function values, but one might produce significantly noisier realizations. This is of major importance when it comes to actual deployment of the found solutions. For example, when selecting hyperparameters of a machine learning algorithm, we might prefer configurations that lead to slightly higher test errors but at the same time lead to smaller variance.

In this paper, we generalize BO to trade off mean and input-dependent noise variance when sequentially querying points and outputting final solutions. We introduce a practical setting where *both* the black-box objective and input-dependent noise variance are *unknown* a priori, and the learner needs to estimate them on the fly. We propose a novel optimistic risk-averse algorithm – RAHBO – that makes sequential decisions by simultaneously balancing between *exploration* (learning about uncertain actions), *exploitation* (choosing actions that lead to high gains) and *risk* (avoiding unreliable actions). We bound the cumulative regret of RAHBO as well as the number of samples required to output a single near-optimal risk-averse solution. In our experiments, we demonstrate the risk-averse performance of our algorithm and show that standard BO methods can severely fail in applications where reliability of the reported solutions is of utmost importance.

35th Conference on Neural Information Processing Systems (NeurIPS 2021).



(a) Unknown objective f (b) Unknown variance ρ^2 (c) Histogram of variance

Figure 1: When there is a choice between identical optima with different noise level, standard BO tends to query points corresponding to higher noise. (a) Unknown objective with 3 global maxima marked as (A, B, C); (b) Heteroscedastic noise variance over the same domain: the noise level at (A, B, C) varies according to the sigmoid function; (c) Empirical variance distribution at all points acquired during BO procedure (over 9 experiments with different seeds). The three bumps correspond to the three global optima with different noise variance. RAHBO dominates in choosing the risk-averse optimum, consequently yielding lower risk-averse regret in Figure 5a.

Related work. Bayesian optimization (BO) [26] refers to approaches for optimizing a noisy black-box objective that is typically expensive to evaluate. A great number of BO methods have been developed over the years, including a significant number of variants of popular algorithms such as GP-UCB [35], Expected Improvement [26], and Thompson Sampling [14]. While the focus of standard BO approaches is mainly on trading-off exploration vs. exploitation and optimizing for the expected performance, in this work, we additionally focus on the risk that is involved when working with noisy objectives, as illustrated in Figure 1.

The vast majority of previous BO works assume (sub-) Gaussian and *homoscedastic* noise (i.e., input independent and of some known fixed level). Both assumptions can be restrictive in practice. For example, as demonstrated in [15], the majority of hyperparameter tuning tasks exhibit heteroscedasticity. A few works relax the first assumption and consider, e.g., heavy-tailed noise models [31] and adversarially corrupted observations [8]. The second assumption is typically generalized via *heteroscedastic* Gaussian process (GP), allowing an explicit dependence of the noise distribution on the evaluation points [6, 10, 7, 20]. Similarly, in this work, we consider heteroscedastic GP models, but unlike the previous works, we specifically focus on the risk that is associated with querying and reporting noisy points.

Several works have recently considered robust and risk-averse aspects in BO. Their central focus is on designing robust strategies and protecting against the change/shift in uncontrollable covariates. They study various notions including worst-case robustness [9], distributional robustness [19, 30], robust mixed strategies [33] and other notions of risk-aversion [18, 11, 29], and while some of them report robust regret guarantees, their focus is primarily on the robustness in the homoscedastic GP setting. Instead, in our setting, we account for the risk that comes from the realization of random noise with *unknown* distribution. Rather than optimizing the expected performance, in our risk-averse setting, we prefer inputs with *lower variance*. To this end, we incorporate the learning of the noise distribution into the optimization procedure via a *mean-variance* objective. The closest to our setting is risk-aversion with respect to noise in multi-armed bandits [32]. Their approach, however, fails to exploit correlation in rewards among similar arms.

Contributions. We propose a novel *Risk-averse Heteroscedastic Bayesian optimization* (RAHBO) approach based on the optimistic principle that trades off the expectation and uncertainty of the mean-variance objective function. We model both the objective and variance as (unknown) functions belonging to RKHS space of functions, and propose a practical risk-aware algorithm in the heteroscedastic GP setting. In our theoretical analysis, we establish rigorous sublinear regret guarantees for our algorithm, and provide a robust reporting rule for applications where only a single solution is required. We demonstrate the effectiveness of RAHBO on synthetic benchmarks, as well as on hyperparameter tuning tasks for the Swiss free-electron laser and a machine learning model.

2 Problem setting

Let $\mathcal{X}$ be a given compact set of inputs ($\mathcal{X} \subset \mathbb{R}^d$ for some $d \in \mathbb{N}$). We consider a problem of sequentially interacting with a fixed and unknown objective $f : \mathcal{X} \to \mathbb{R}$. At every round of this procedure, the learner selects an action $x_t \in \mathcal{X}$, and obtains a noisy observation

$$y_t = f(x_t) + \xi(x_t), \tag{1}$$

where $\xi(x_t)$ is zero-mean noise independent across different time steps t. In this work, we consider sub-Gaussian heteroscedastic noise that depends on the query location.

Definition 1. *A zero-mean real-valued random variable ξ is ρ–sub-Gaussian, if there exists variance-proxy ρ^2 such that* $\forall \lambda \in \mathbb{R},\ \mathbb{E}[e^{\lambda \xi}] \leq e^{\frac{\lambda^2 \rho^2}{2}}$.

For a sub-Gaussian ξ, its variance $\mathbb{V}ar[\xi]$ lower bounds any valid variance-proxy ρ, i.e., $\mathbb{V}ar[\xi] \leq \rho^2$. In case $\mathbb{V}ar[\xi] = \rho^2$ holds, ξ is said to be *strictly ρ–sub-Gaussian*. Besides zero-mean Gaussian random variables, most standard symmetric bounded random variables (e.g., Bernoulli, beta, uniform, binomial) are strictly sub-Gaussian (see [2, Proposition 1.1]). Throughout the paper, we consider sub-Gaussian noise, and in Section 3.3, we specialize to the case of strictly sub-Gaussian noise.

Optimization objective. Unlike the previous works that mostly focus on sequential optimization of f in the homoscedastic noise case, in this work, we consider the trade-off between risk and return in the heteroscedastic case. While there exist a number of risk-averse objectives, we consider the simple and frequently used mean-variance objective (MV) [32]. Here, the objective value at $x \in \mathcal{X}$ is a trade-off between the (mean) return $f(x)$ and the risk expressed by its variance-proxy $\rho^2(x)$:

$$\mathrm{MV}(x) = f(x) - \alpha \rho^2(x), \tag{2}$$

where $\alpha \geq 0$ is a so-called *coefficient of absolute risk tolerance*. In this work, we assume α is fixed and known to the learner. In the case of $\alpha = 0$, maximizing $\mathrm{MV}(x)$ coincides with the standard BO objective.

Performance metrics. We aim to construct a sequence of input evaluations x_t that eventually maximizes the risk-averse objective $\mathrm{MV}(x)$. To assess this convergence, we consider two metrics. The first metric corresponds to the notion of cumulative regret similar to the one used in standard BO and bandits. Here, the learner's goal is to maximize its risk-averse cumulative reward over a time horizon T, or equivalently minimize its *risk-averse cumulative regret*:

$$R_T = \sum_{t=1}^{T} \Big[\mathrm{MV}(x^*) - \mathrm{MV}(x_t) \Big], \tag{3}$$

where $x^* \in \arg\max_{x \in \mathcal{X}} \mathrm{MV}(x)$. A sublinear growth of R_T with T implies vanishing average regret $R_T/T \to 0$ as $T \to \infty$. Intuitively, this implies the existence of some t such that $\mathrm{MV}(x_t)$ is arbitrarily close to the optimal value $\mathrm{MV}(x^*)$.

The second metric is used when the learner seeks to simultaneously minimize the number of expensive function evaluations T. Namely, for a given accuracy $\epsilon \geq 0$, we report a single "good" risk-averse point $\hat{x}_T \in \mathcal{X}$ after a total of T rounds, that satisfies:

$$\mathrm{MV}(\hat{x}_T) \geq \mathrm{MV}(x^*) - \epsilon. \tag{4}$$

Both metrics are important for choosing risk-averse solutions and which one is preferred depends on the application at hand. For example, risk-averse cumulative regret R_T might be of a greater interest in online recommendation systems, while reporting a single point with high MV value might be more suitable when tuning machine learning hyperparameters. We consider both performance metrics in our experiments.

Regularity assumptions. We consider standard smoothness assumptions [35, 9] when it comes to the unknown function $f : \mathcal{X} \to \mathbb{R}$. In particular, we assume that $f(\cdot)$ belongs to a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_\kappa$ (a space of smooth and real-valued functions defined on $\mathcal{X}$), i.e., $f \in \mathcal{H}_\kappa$, induced by a kernel function $\kappa(\cdot, \cdot)$. We also assume that $\kappa(x, x') \leq 1$ for every $x, x' \in \mathcal{X}$. Moreover, the RKHS norm of $f(\cdot)$ is assumed to be bounded $\|f\|_\kappa \leq \mathcal{B}_f$ for some fixed constant $\mathcal{B}_f > 0$. We assume that the noise $\xi(x)$ is $\rho(x)$–sub-Gaussian with variance-proxy $\rho^2(x)$ uniformly bounded $\rho(x) \in [\underline{\varrho}, \bar{\varrho}]$ for some constant values $\bar{\varrho} \geq \underline{\varrho} > 0$.

3 Algorithms

We first recall the Gaussian process (GP) based framework for sequential learning of RKHS functions from observations with heteroscedastic noise. Then, in Section 3.2, we consider a simple risk-averse Bayesian optimization problem with *known* variance-proxy, and later on in Section 3.3, we focus on our main problem setting in which the variance-proxy is *unknown*.

3.1 Bayesian optimization with heteroscedastic noise

Before addressing the risk-averse objective, we briefly recall the standard GP-UCB algorithm [35] in the setting of heteroscedastic sub-Gaussian noise. The regularity assumptions permit the construction of confidence bounds via GP model. Particularly, to decide which point to query at every round, GP-UCB makes use of the posterior GP mean and variance denoted by $\mu_t(\cdot)$ and $\sigma_t^2(\cdot)$, respectively. They are computed based on the previous measurements $y_{1:t} = [y_1, \dots, y_t]^\top$ and the given kernel $\kappa(\cdot,\cdot)$:

$$\mu_t(x) = \kappa_t(x)^T (K_t + \lambda \Sigma_t)^{-1} y_{1:t}, \tag{5}$$

$$\sigma_t^2(x) = \frac{1}{\lambda}\big(\kappa(x,x) - \kappa_t(x)^\top (K_t + \lambda\Sigma_t)^{-1}\kappa_t(x)\big), \tag{6}$$

where $\Sigma_t := \mathrm{diag}(\rho^2(x_1), \dots, \rho^2(x_t))$, $(K_t)_{i,j} = \kappa(x_i, x_j)$, $\kappa_t(x)^T = [\kappa(x_1,x), \dots, \kappa(x_t,x)]^T$, $\lambda > 0$ and prior modelling assumptions are $\xi(\cdot) \sim \mathcal{N}(0, \rho^2(\cdot))$ and $f \sim GP(0, \lambda^{-1}\kappa)$.

At time t, GP-UCB maximizes the upper confidence bound of $f(\cdot)$, i.e.,

$$x_t \in \arg\max_{x\in\mathcal{X}} \underbrace{\mu_{t-1}(x) + \beta_t \sigma_{t-1}(x)}_{=:\mathrm{ucb}_t^f(x)}. \tag{7}$$

If the noise $\xi_t(x_t)$ is heteroscedastic and $\rho(x_t)$-sub-Gaussian, the following confidence bounds hold:

Lemma 1 (Lemma 7 in [20])**.** *Let $f \in \mathcal{H}_\kappa$, and $\mu_t(\cdot)$ and $\sigma_t^2(\cdot)$ be defined as in Eqs.* (5) *and* (6) *with $\lambda > 0$. Assume that the observations $(x_t, y_t)_{t\geq 1}$ satisfy Eq.* (1)*. Then the following holds for all $t \geq 1$ and $x \in \mathcal{X}$ with probability at least $1-\delta$:*

$$|\mu_{t-1}(x) - f(x)| \leq \underbrace{\left(\sqrt{2\ln\left(\frac{\det(\lambda\Sigma_t + K_t)^{1/2}}{\delta\det(\lambda\Sigma_t)^{1/2}}\right)} + \sqrt{\lambda}\|f\|_\kappa\right)}_{:=\beta_t} \sigma_{t-1}(x). \tag{8}$$

Here, β_t stands for the parameter that balances between exploration vs. exploitation and ensures the validity of confidence bounds. The analogous concentration inequalities in case of homoscedastic noise were considered in [1, 14, 35].

Failure of GP-UCB in the risk-averse setting. GP-UCB is guaranteed to achieve sublinear cumulative regret with high probability in the risk-neutral (homoscedastic/heteroscedastic) BO setting [35, 14]. However, for the risk-averse setting in Eq. (2), the maximizers $x^* \in \arg\max_{x\in\mathcal{X}} \mathrm{MV}(x)$ and $x_f^* \in \arg\max_{x\in\mathcal{X}} f(x)$ might not coincide, and consequently, $\mathrm{MV}(x^*)$ can be significantly larger than $\mathrm{MV}(x_f^*)$. This is illustrated in Figure 1, where GP-UCB most frequently chooses optimum A of the highest risk.

3.2 Warm up: Known variance-proxy

We remedy the previous issue with GP-UCB by proposing a natural *Risk-averse Heteroscedastic* BO (RAHBO) in case of the known variance-proxy $\rho^2(\cdot)$. At each round t, RAHBO chooses the action:

$$x_t \in \arg\max_{x\in\mathcal{X}} \mu_{t-1}(x) + \beta_t\sigma_{t-1}(x) - \alpha\rho^2(x), \tag{9}$$

where β_t is from Lemma 1 and α is from Eq. (2). In the next section, we further relax the assumption of the variance-proxy and consider a more practical setting when $\rho^2(\cdot)$ is unknown to the learner. For the current setting, the performance of RAHBO is formally captured in the following proposition.

Proposition 1. *Consider any $f \in \mathcal{H}_\kappa$ with $\|f\|_\kappa \leq \mathcal{B}_f$ and sampling model from Eq.* (1) *with known variance-proxy $\rho^2(x)$. Let $\{\beta_t\}_{t=1}^T$ be set as in Lemma 1 with $\lambda = 1$. Then, with probability at least $1-\delta$,* RAHBO *attains cumulative risk-averse regret $R_T = \mathcal{O}\big(\beta_T\sqrt{T\gamma_T(\bar{\varrho}^2+1)}\big)$.*

Here, γ_T denotes the *maximum information gain* [35] at time T defined via mutual information $I(y_{1:T}, f_{1:T})$ between evaluations $y_{1:T}$ and $f_{1:T} = [f(x_1), \dots, f(x_T)]^\top$ at points $A \subset D$:

$$\gamma_T := \max_{A \subset \mathcal{X},\, |A|=T} I(y_{1:T}, f_{1:T}), \tag{10}$$

$$\text{where } I(y_{1:T}, f_{1:T}) = \frac{1}{2} \sum_{t=1}^{T} \ln \left(1 + \tfrac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}\right) \tag{11}$$

in case of heteroscedastic noise (see Appendix A.1.1). The upper bounds on γ_T are provided in [35] widely used kernels. These upper bounds typically scale sublinearly in T; for linear kernel $\gamma_T = \mathcal{O}(d \log T)$, and in case of squared exponential kernel $\gamma_T = \mathcal{O}(d(\log T)^{d+1})$. While these bounds are derived assuming the homoscedastic GP setting with some fixed constant noise variance, we show (in Appendix A.1.3) that the same rates (up to a multiplicative constant factor) apply in the heteroscedastic case.

3.3 RAHBO for unknown variance-proxy

In the case of unknown variance-proxy, the confidence bounds for the unknown $f(x)$ in Lemma 1 can not be readily used, and we construct new ones on the combined mean-variance objective. To learn about the unknown $\rho^2(x)$, we make some further assumptions.

Assumption 1. *The variance-proxy $\rho^2(x)$ belongs to an RKHS induced by some kernel κ^{var}, i.e., $\rho^2 \in \mathcal{H}_{\kappa^{var}}$, and its RKHS norm is bounded $\|\rho^2\|_{\kappa^{var}} \leq \mathcal{B}_{var}$ for some finite $\mathcal{B}_{var} > 0$. Moreover, the noise $\xi(x)$ in Eq.* (1) *is strictly $\rho(x)$–sub-Gaussian, i.e., $\mathbb{V}ar[\xi(x)] = \rho^2(x)$ for every $x \in \mathcal{X}$.*

As a consequence of our previous assumption, we can now focus on estimating the variance since $\mathbb{V}ar[\xi(\cdot)]$ and $\rho^2(\cdot)$ coincide. In particular, to estimate $\mathbb{V}ar[\xi(\cdot)]$ we consider a *repeated experiment setting*, where for each x_t we collect $k > 1$ evaluations $\{y_i(x_t)\}_{i=1}^k$, $y_i(x_t) = f(x_t) + \xi_i(x_t)$. Then, the sample mean and variance of $\xi(x_t)$ are given as:

$$\hat{m}_k(x_t) = \frac{1}{k} \sum_{i=1}^{k} y_i(x_t) \quad \text{and} \quad \hat{s}^2_k(x_t) = \frac{1}{k-1} \sum_{i=1}^{k} \big(y_i(x_t) - \hat{m}_k(x_t)\big)^2. \tag{12}$$

The key idea is that for strictly sub-Gaussian noise $\xi(x)$, $\hat{s}^2_{1:t} = [\hat{s}^2_k(x_1), \dots, \hat{s}^2_k(x_t)]^\top$ yields *unbiased, but noisy* evaluations of the unknown variance-proxy $\rho^2_{1:t} = [\rho^2(x_1), \dots, \rho^2(x_t)]^\top$, i.e.,

$$\hat{s}^2_k(x_t) = \rho^2(x_t) + \eta(x_t) \tag{13}$$

with zero-mean noise $\eta(x_t)$. In order to efficiently estimate $\rho^2(\cdot)$, we need an additional assumption.

Assumption 2. *The noise $\eta(x)$ in Eq.* (13) *is $\rho_\eta(x)$–sub-Gaussian with known $\rho^2_\eta(x)$ and the realizations $\{\eta(x_t)\}_{t \geq 1}$ are independent between t.*

We note that a similar assumption is made in [32] in the multi-armed bandit setting. The fact that $\rho^2_\eta(\cdot)$ is known is rather mild as Assumption 1 allows controlling its value. For example, in case of strictly sub-Gaussian $\eta(x)$ we show (in Appendix A.2) that $\mathbb{V}ar[\eta(\cdot)] = \rho^2_\eta(\cdot) \leq 2\rho^4(\cdot)/(k-1)$. Then, given that $\rho^2(\cdot) \leq \bar{\varrho}^2$, we can utilize the following (rather conservative) bound as a variance-proxy, i.e., $\rho^2_\eta(x) = 2\bar{\varrho}^4/(k-1)$.

RAHBO algorithm. We present our Risk-averse Heteroscedastic BO approach for unknown variance-proxy in Algorithm 1. Our method relies on building the following two GP models.

Firstly, we use sample variance evaluations $\hat{s}^2_{1:t}$ to construct a GP model for $\rho^2(\cdot)$. The corresponding $\mu^{var}_{t-1}(\cdot)$ and $\sigma^{var}_{t-1}(\cdot)$ are computed as in Eqs. (5) and (6) by using kernel κ^{var}, variance-proxy $\rho^2_\eta(\cdot)$ and noisy observations $\hat{s}^2_{1:t}$. Consequently, we build the upper and lower confidence bounds $\mathrm{ucb}^{var}_t(\cdot)$ and $\mathrm{lcb}^{var}_t(\cdot)$ of the variance-proxy $\rho^2(\cdot)$ and we set β^{var}_t according to Lemma 1:

$$\mathrm{ucb}^{var}_t(x) := \mu^{var}_{t-1}(x) + \beta^{var}_t \sigma^{var}_{t-1}(x), \tag{14}$$

$$\mathrm{lcb}^{var}_t(x) := \mu^{var}_{t-1}(x) - \beta^{var}_t \sigma^{var}_{t-1}(x). \tag{15}$$

Secondly, we use sample mean evaluations $\hat{m}_{1:t} = [\hat{m}_k(x_1), \dots, \hat{m}_k(x_t)]^\top$ to construct a GP model for $f(\cdot)$. The mean $\mu_t(\cdot)$ and variance $\sigma^2_t(\cdot)$ in Eqs. (5) and (6), however, rely on the unknown

Algorithm 1 Risk-averse Heteroscedastic Bayesian Optimization (RAHBO)

Require: Parameters $\alpha, \{\beta_t, \beta_t^{var}\}_{t\geq 1}, \lambda, k$, Prior $\mu_0^f = \mu_0^{var} = 0$, Kernel functions κ, κ^{var}

for $t = 1, 2, \ldots$ **do**
  Construct confidence bounds $\mathrm{ucb}_t^{var}(\cdot)$ and $\mathrm{lcb}_t^{var}(\cdot)$ as in Eqs. (14) and (15)
  Construct $\mathrm{ucb}_t^f(\cdot)$ as in Eq. (7)
  Select $x_t \in \arg\max_{x\in\mathcal{X}} \mathrm{ucb}_t^f(x) - \alpha\, \mathrm{lcb}_t^{var}(x)$
  Observe k samples: $y_i(x_t) = f(x_t) + \xi_i(x_t) \quad$ for every $\quad i \in [k]$
  Use samples $\{y_i(x_t)\}_{i=1}^k$ to compute sample mean $\hat{m}_k(x_t)$ and variance $\hat{s}_k^2(x_t)$ as in Eq. (12)
  Use $x_t, \hat{s}_k^2(x_t)$ to update posterior $\mu_t^{var}(\cdot)$ and $\sigma_t^{var}(\cdot)$ as in Eqs. (32) and (33)
  Use $\mathrm{ucb}_t^{var}(\cdot)$ to compute $\hat{\Sigma}_t$ as in Eq. (16)
  Use $x_t, \hat{m}_k(x_t)$ and $\hat{\Sigma}_t$ to update posterior $\mu_t(\cdot)$ and $\sigma_t(\cdot)$ as in Eqs. (5) and (6)
end for

variance-proxy $\rho^2(\cdot)$ in Σ_t, an we thus use its upper confidence bound $\mathrm{ucb}_t^{var}(\cdot)$ truncated with $\bar{\varrho}^2$:

$$\hat{\Sigma}_t := \tfrac{1}{k}\mathrm{diag}\big(\min\{\mathrm{ucb}_t^{var}(x_1), \bar{\varrho}^2\}, \ldots, \min\{\mathrm{ucb}_t^{var}(x_t), \bar{\varrho}^2\}\big), \tag{16}$$

where $\hat{\Sigma}_t$ is corrected by k since every evaluation in $\hat{m}_{1:t}$ is an average over k samples. This substitution of the unknown variance-proxy by its conservative estimate guarantees that the confidence bounds $\mathrm{ucb}_t^f(x) := \mu_{t-1}(x) + \beta_t\sigma_{t-1}(x)$ on f also hold with high probability (conditioning on the confidence bounds for $\rho(\cdot)$ holding true; see Appendix A.3 for more details).

Finally, we define the acquisition function as $\mathrm{ucb}_t^{\mathrm{MV}}(x) := \mathrm{ucb}_t^f(x) - \alpha\mathrm{lcb}_t^{var}(x)$, i.e., selecting $x_t \in \arg\max_{x\in\mathcal{X}} \mathrm{ucb}_t^{\mathrm{MV}}(x)$ at each round t.

The proposed algorithm leads to new maximum information gains $\hat{\gamma}_T = \max_A I(\hat{m}_{1:T}, f_{1:T})$ and $\Gamma_T = \max_A I(\hat{s}_{1:T}^2, \rho_{1:T}^2)$ for sample mean $\hat{m}_{1:T}$ and sample variance $\hat{s}_{1:T}^2$ evaluations. The corresponding mutual information in $\hat{\gamma}_T$ and Γ_T is computed according to Eq. (11) for heteroscedastic noise with variance-proxy $\bar{\varrho}^2/k$ and ρ_η^2, respectively (see Appendix A.4). The performance of RAHBO is captured in the following theorem.

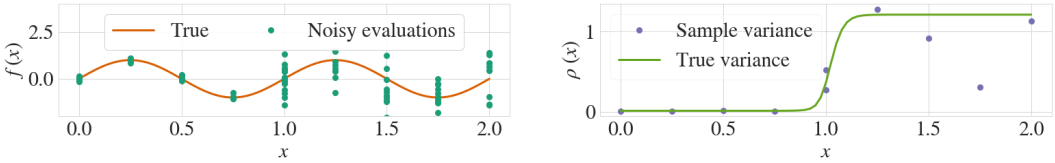
Theorem 1. *Consider any $f \in \mathcal{H}_\kappa$ with $\|f\|_\kappa \leq \mathcal{B}_f$ and sampling model in Eq.* (1) *with unknown variance-proxy $\rho^2(x)$ that satisfies Assumptions 1 and 2. Let $\{x_t\}_{t=1}^T$ denote the set of actions chosen by* RAHBO *(Algorithm 1) over T rounds. Set $\{\beta_t\}_{t=1}^T$ and $\{\beta_t^{var}\}_{t=1}^T$ according to Lemma 1 with $\lambda = 1$, $\mathcal{R}^2 = \max_{x\in\mathcal{X}} \rho_\eta^2(x_t)$ and $\rho(\cdot) \in [\underline{\varrho}, \bar{\varrho}]$. Then, the risk-averse cumulative regret R_T of* RAHBO *is bounded as follows:*

$$\Pr\left\{R_T \leq \beta_T k\sqrt{\frac{2T\hat{\gamma}_T}{\ln(1+k/\bar{\varrho}^2)}} + \alpha\beta_T^{var}k\sqrt{\frac{2T\Gamma_T}{\ln(1+\mathcal{R}^{-2})}}, \quad \forall T \geq 1\right\} \geq 1-\delta. \tag{17}$$

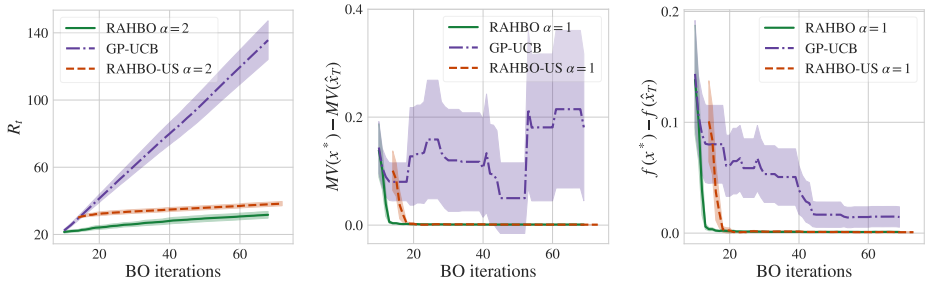
The risk-averse cumulative regret of RAHBO depends sublinearly on T for most of the popularly used kernels. This follows from the implicit sublinear dependence on T in β_T, β_T^{var} and $\hat{\gamma}_T, \Gamma_T$ (the bounds in case of heteroscedastic noise replicate the ones used in Proposition 1 as shown in Appendices A.1.2 and A.1.3). Finally, the result of Theorem 1 provides a non-trivial trade-off for number of repetitions k where larger k increases sample complexity but also leads to better estimation of the noise model. Furthermore, we obtain the bound for the number of rounds T required for identifying an ϵ-optimal point:

Corollary 1.1. *Consider the setup of Theorem 1. Let $A = \{x_t\}_{t=1}^T$ denote actions selected by* RAHBO *over T rounds. Then, with probability at least $1-\delta$, the reported point $\hat{x}_T := \arg\max_{x_t\in A} \mathrm{lcb}_t^{MV}(x_t)$, where $\mathrm{lcb}_t^{MV}(x_t) = \mathrm{lcb}_t^f(x) - \alpha\ \mathrm{ucb}_t^{var}(x)$, achieves ϵ-accuracy, i.e., $MV(x^*) - MV(\hat{x}_T) \leq \epsilon$, after $T \geq \frac{4\beta_T^2\hat{\gamma}_T/\ln(1+k/\bar{\varrho}^2) + 4\alpha(\beta_t^{var})^2\Gamma_T/\ln(1+\mathcal{R}^{-2})}{\epsilon^2}$ rounds.*

The previous result demonstrates the sample complexity rates when a single risk-averse reported solution is required. We note that both Theorem 1 and Corollary 1.1 provide guarantees for choosing risk-averse solutions, and depending on application at hand, we might consider either one of the proposed performance metrics. We demonstrate use-cases for both in the following section.



(a) Illustration of the sine function (left) and noise variance (right)



(b) Cumulative regret (c) Suboptimality w.r.t. MV (d) Suboptimality w.r.t. f

Figure 2: (a) Unknown true objective along with noisy evaluations with varying noise level (left) and unknown true noise variance and its evaluations (right). (b) Cumulative regret. (c) Simple MV regret for reporting rule $\hat{x}_T = \arg\max_{x_t} \mathrm{lcb}_T(x_t)$. (c) Simple regret $f(x^*) - f(\hat{x}_T)$ for the unknown function at the reported point $\hat{x}_T$ from (d). RAHBO not only leads to strong results in terms of MV but also in terms of the mean objective $f(x)$.

4 Experiments

In this section, we experimentally validate RAHBO on two synthetic examples and two real hyperparameter tuning tasks, and compare it with the baselines. We provide an open-source implementation of our method.[1]

Baselines. We compare against two baselines: As the first baseline, we use GP-UCB with heteroscedastic noise as a standard risk-neutral algorithm that optimizes the unknown $f(x)$. As the second one, we consider a risk-averse baseline that uniformly learns variance-proxy $\rho^2(x)$ *before* the optimization procedure, in contrast to RAHBO which learns the variance-proxy on the fly. We call it RAHBO-US, standing for RAHBO with uncertainty sampling. It consists of two stages: (i) uniformly learning $\rho^2(x)$ via uncertainty sampling, (ii) GP-UCB applied to the mean-variance objective, in which instead of the unknown $\rho^2(x)$ we use the mean of the learned model. Note that RAHBO-US is the closest to the contextual BO setting [18], where the context distribution is assumed to be known.

Experimental setup. At each iteration t, an algorithm queries a point x_t and observes sample mean and sample variance of k observations $\{y_i(x_t)\}_{i=1}^k$. We use a heteroscedastic GP for modelling $f(x)$ and a homoscedastic GP for $\rho^2(x)$. We set $\lambda = 1$ and $\beta_t = 2$, which is commonly used in practice to improve performance over the theoretical results. Before the BO procedure, we determine the GP hyperparameters maximizing the marginal likelihood. To this end, we use initial points that are same for all the baselines and are chosen via Sobol sequence that generates low discrepancy quasi-random samples. We repeat each experiment several times, generating new initial points for every repetition. We use two metrics: (a) risk-averse cumulative regret R_t computed for the acquired inputs; (b) simple regret $\mathrm{MV}(x^*) - \mathrm{MV}(\hat{x}_T)$ computed for inputs as reported via Corollary 1.1. For each metric, we report its mean $\pm$ two standard errors over the repetitions.

Example function We first illustrate the methods performance on a sine function depicted in Figure 2a. This function has two global optimizers. We induce a heteroscedastic zero-mean Gaussian noise on the measurements. We use a sigmoid function for the noise variance, as depicted in Figure 2a, that induces small noise on $[0, 1]$ and higher noise on $(1, 2]$. We initialize the algorithms by selecting 10 inputs x at random and keep these points the same for all the algorithms. We use $k = 10$ samples at each chosen x_t. The number of acquisition rounds is $T = 60$. We repeat the experiment 30 times for each method and show their average performances in Figure 2.

[1] `https://github.com/Avidereta/risk-averse-hetero-bo`

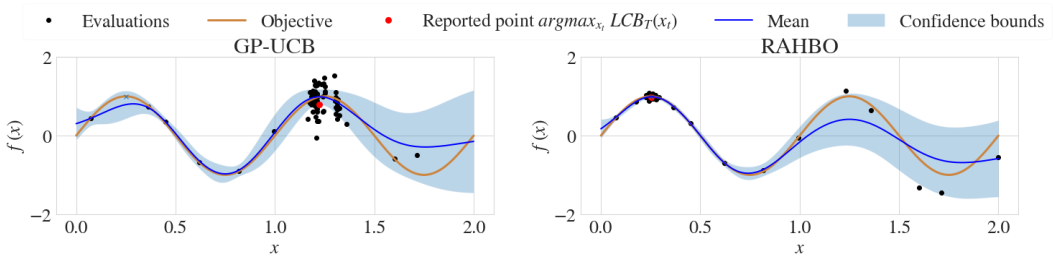


Figure 3: GP models fitted for GP-UCB (left) and RAHBO (right) for sine function. After initialization with the same sampled points, GP-UCB concentrates on the high-noise region whereas RAHBO prefers small variance. Additional plots are presented in Appendix A.6.

Branin benchmark Next, we evaluate the methods on the (negated) Branin benchmark function in Figure 1a, achieving its optimum value $f^* = -0.4$ at $(-\pi, 12.3), (\pi, 2.3), (9.4, 2.5)$. The heteroscedastic variance function illustrated in Figure 1b defines different noise variances for the three optima. We initialize all algorithms by selecting 10 inputs. We use $k = 10$ samples to estimate the noise variance. The number of acquisition rounds is $T = 150$. We repeat BO 25 times and show the results in Figures 1c and 5a. Figure 1c provides more intuition behind the observed regret: UCB exploits the noisiest maxima the most, while RAHBO prefers smaller variance.

Tuning Swiss free-electron laser In this experiment, we tune the parameters of Swiss X-ray free-electron laser (SwissFEL), an important scientific instrument that generates very short pulses of X-ray light and enables researchers to observe extremely fast processes. The main objective is to maximize the pulse energy measured by a gas detector, that is a time-consuming and repetitive task during the SwissFEL operation. Such (re-)tuning takes place while user experiments on SwissFEL are running, and thus the cumulative regret is the metric of high importance in this application.

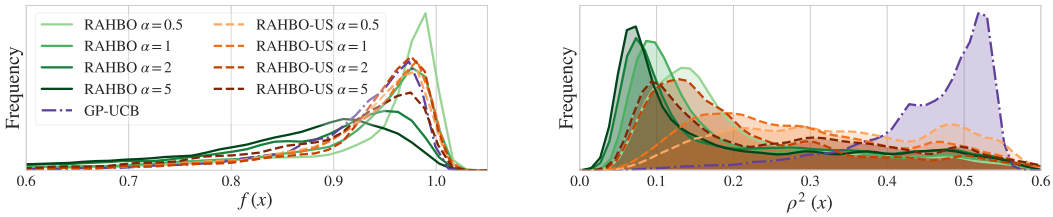
We use real SwissFEL measurements collected in [21] to train a neural network surrogate model, and use it to simulate the SwissFEL objective $f(x)$ for new parameter settings x. We similarly fit a model of the heteroscedastic variance by regressing the squared residuals via a GP model. Here, we focus on the calibration of the four most sensitive parameters.

We report our comparison in Figure 4 where we also assess the effect of varying the coefficient of absolute risk tolerance α. We use 30 points to initialize the baselines and then perform 200 acquisition rounds. We repeat each experiment 15 times. In Figure 4a we plot the empirical frequency of the true (unknown to the methods) values $f(x_t)$ and $\rho^2(x_t)$ at the inputs x_t acquired by the methods. The empirical frequency for $\rho^2(x)$ illustrates the tendency of risk-neutral GP-UCB to query points with higher noise, while risk-averse achieves substantially reduced variance and minimal reduction in mean performance. Sometimes, risk-neutral GP-UCB also fails to succeed in querying points with the highest f-value. That tendency results in lower cumulative regret for RAHBO in Figures 4c and 4d. We also compare the performance of the reporting rule from Corollary 1.1 in Figure 4b, where we plot error bars with standard deviation both for $f(\hat{x}_T)$ and $\rho^2(\hat{x}_T)$ at the reported point $\hat{x}_T$. As before, RAHBO drastically reduces the variance compared to GP-UCB, while having only slightly lower mean performance. Additional results are presented in Figure 10 in Appendix.

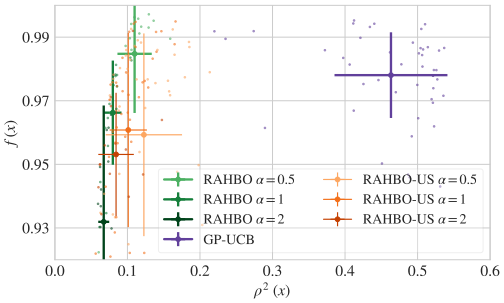
Random Forest tuning BO is widely used by cloud services for tuning machine learning hyperparameters and the resulting models might be then used in high-stakes applications such as credit scoring or fraud detection. In k-fold cross-validation, the average metric over the validation sets is optimized – a canonical example of the *repeated experiment setting* that we consider in the paper. High across-folds variance is a practical problem [24] where the mean-variance approach might be beneficial.

In our experiment, we tune hyperparameters of a random forest classifier (RF) on a dataset of fraudulent credit card transactions [23].[2]It consist of 285k transactions with 29 features (processed due to confidentiality issues) that are distributed over time, and only 0.2% are fraud examples (see Appendix for more details). The search space for the RF hyperparameters is also provided in the Appendix. We use the balanced accuracy score and 5 validation folds, i.e., $k = 5$, and each validation fold is shifted in time with respect to the training data. We seek not only for high performance *on average* but also for low variance across the validation folds that have different time shifts with respect to the training data.

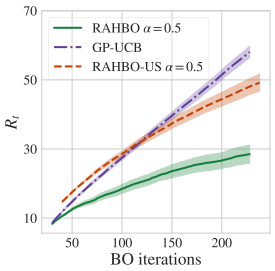
[2]`https://www.kaggle.com/mlg-ulb/creditcardfraud`



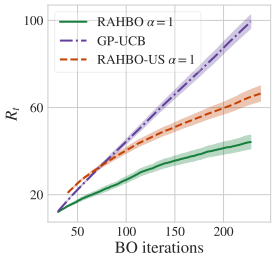
(a) Empirical distribution of true $f(x)$ (left) and $\rho^2(x)$ (right) for SwissFEL



(b) Mean-variance tradeoff (SwissFEL)



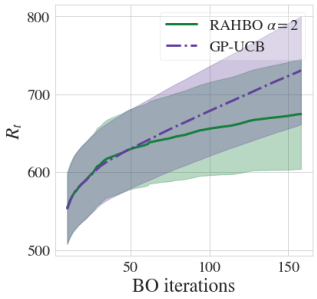
(c) Cum. regret ($\alpha = 0.5$)



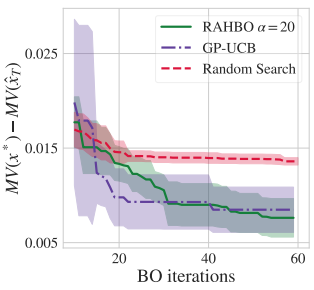
(d) Cum. regret ($\alpha = 1$)

Figure 4: **(a)** Distributions of $f(x)$ and $\rho^2(x)$ for *all points* queried by BO for the SwissFEL. Clearly, GP-UCB queries points with higher noise (but not necessarily high return f) in contrast to the risk-averse methods. Moreover, greater α leads to lower variance. A good trade-off is achieved at $\alpha = 0.5$. **(b)** Standard deviation error bars for $f(\hat{x}_T)$ and $\rho^2(\hat{x}_T)$ at the *reported* $\hat{x}_T = \arg\max_{x_t} \mathrm{lcb}_T(x_t)$ for SwissFEL computed over the repeated BO. Though points reported by GP-UCB have slightly higher $f(x)$ than most of the risk-averse methods, its noise variance is much higher. Again, $\alpha = 0.5$ yields a good mean-variance trade-off. **(c-d)** Cum. regret for $\alpha = 0.5$ and $\alpha = 1$ (see more in Appendix.)

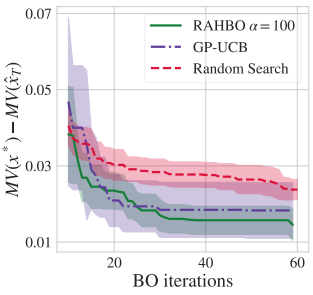
We initialize the algorithms by selecting 10 hyperparameter settings and keep these points the same for all algorithms. We use Matérn 5/2 kernels with Automatic Relevance Discovery (ARD) and normalize the input features to the unit cube. The number of acquisition rounds in one experiment is 50 and we repeat each experiment 15 times. We demonstrate our results in Figures 5b and 5c where we plot mean $\pm$ 2 standard errors. While both RAHBO and GP-UCB perform comparable in terms of the mean error, its standard deviation for RAHBO is smaller.



(a) Branin benchmark



(b) RF Tuning



(c) RF Tuning

Figure 5: **Branin**: (a) Cumulative regret. **Random Forest**: (b-c) Simple regret for the reported $\hat{x}_T = \arg\max_{x_t} MV(x_t)$ for (b) $\alpha = 20$ and (c) $\alpha = 100$. While both methods have comparable mean, RAHBO has consistently lower variance.

5 Conclusion

In this work, we generalize Bayesian optimization to the risk-averse setting and propose RAHBO algorithm aiming to find an input with both large expected return and small input-dependent noise variance. Both the mean objective and the variance are assumed to be unknown a priori and hence are estimated online. RAHBO is equipped with theoretical guarantees showing (under reasonable

assumptions) sublinear dependence on the number of evaluation rounds T both for cumulative risk-averse regret and ϵ-accurate mean-variance metric. The empirical evaluation of the algorithm on synthetic benchmarks and hyperparameter tuning tasks demonstrate promising examples of heteroscedastic use-cases benefiting from RAHBO.

Acknowledgements

This research has been gratefully supported by NCCR Automation grant 51NF40 180545, by ERC under the European Union's Horizon grant 815943, SNSF grant 200021_172781, and ETH Zürich Postdoctoral Fellowship 19-2 FEL-47. The authors thank Sebastian Curi, Mojmír Mutný and Johannes Kirschner as well as the anonymous reviewers of this paper for their helpful feedback.

References

[1] Y. Abbasi-Yadkori. *Online learning for linearly parametrized control problems*. PhD thesis, 2012.

[2] J. Arbel, O. Marchal, and H. D. Nguyen. On strict sub-Gaussianity, optimal proxy variance and symmetry for bounded random variables. *arXiv:1901.09188*, 2019.

[3] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy. BoTorch: a framework for efficient Monte-Carlo Bayesian optimization. In *Neural Information Processing Systems (NeurIPS)*, 2020.

[4] E. Benhamou. A few properties of sample variance. *arXiv:1809.03774*, 2018.

[5] F. Berkenkamp, A. Krause, and A. P. Schoellig. Bayesian optimization with safety constraints: Safe and automatic parameter tuning in robotics. *arXiv:1602.04450*, 2016.

[6] M. Binois, R. B. Gramacy, and M. Ludkovski. Practical heteroscedastic Gaussian process modeling for large simulation experiments. *Journal of Computational and Graphical Statistics*, 2018.

[7] M. Binois, J. Huang, R. B. Gramacy, and M. Ludkovski. Replication or exploration: Sequential design for stochastic simulation experiments. In *Technometrics*, 2019.

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

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

[10] I. Bogunovic, J. Scarlett, A. Krause, and V. Cevher. Truncated variance reduction: A unified approach to Bayesian optimization and level-set estimation. In *Neural Information Processing Systems (NeurIPS)*, 2016.

[11] S. Cakmak, R. Astudillo, P. I. Frazier, and E. Zhou. Bayesian optimization of risk measures. In *Neural Information Processing Systems (NeurIPS)*, 2020.

[12] R. Calandra, A. Seyfarth, J. Peters, and M. P. Deisenroth. Bayesian optimization for learning gaits under uncertainty. *Annals of Mathematics and Artificial Intelligence*, 2016.

[13] Y. Chen, A. Huang, Z. Wang, I. Antonoglou, J. Schrittwieser, D. Silver, and N. de Freitas. Bayesian optimization in Alphago. *arXiv:1812.06855*, 2018.

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

[15] A. I. Cowen-Rivers, W. Lyu, R. Tutunov, Z. Wang, A. Grosnit, R. R. Griffiths, H. Jianye, J. Wang, and H. B. Ammar. An empirical study of assumptions in Bayesian optimisation. *arXiv:2012.03826*, 2021.

[16] J. González, J. Longworth, D. C. James, and N. D. Lawrence. Bayesian optimization for synthetic gene design. *arXiv:1505.01627*, 2015.

[17] R.-R. Griffiths and J. M. Hernández-Lobato. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chemical Science*, 2020.

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

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

[20] J. Kirschner and A. Krause. Information directed sampling and bandits with heteroscedastic noise. In *Conference On Learning Theory (COLT)*, 2018.

[21] J. Kirschner, M. Mutny, N. Hiller, R. Ischebeck, and A. Krause. Adaptive and safe Bayesian optimization in high dimensions via one-dimensional subspaces. In *International Conference on Machine Learning (ICML)*, 2019.

[22] K. Korovina, S. Xu, K. Kandasamy, W. Neiswanger, B. Poczos, J. Schneider, and E. P. Xing. Chembo: Bayesian optimization of small organic molecules with synthesizable recommendations. *arXiv:1908.01425*, 2019.

[23] Y.-A. Le Borgne and G. Bontempi. *Machine Learning for Credit Card Fraud Detection - Practical Handbook*. 2021.

[24] A. Makarova, H. Shen, V. Perrone, A. Klein, J. B. Faddoul, A. Krause, M. Seeger, and C. Archambeau. Overfitting in Bayesian optimization: an empirical study and early-stopping solution. In *ICLR Workshop on Neural Architecture Search*, 2021.

[25] A. Marco, P. Hennig, J. Bohg, S. Schaal, and S. Trimpe. Automatic LQR tuning based on Gaussian process global optimization. In *International Conference on Robotics and Automation (ICRA)*, 2016.

[26] J. Močkus. On Bayesian methods for seeking the extremum. In *Optimization Techniques*, 1975.

[27] H. Moss, D. Leslie, D. Beck, J. González, and P. Rayson. BOSS: Bayesian optimization over string spaces. In *Neural Information Processing Systems (NeurIPS)*, 2020.

[28] D. M. Negoescu, P. I. Frazier, and W. B. Powell. The knowledge-gradient algorithm for sequencing experiments in drug discovery. *INFORMS J. on Computing*, 2011.

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

[30] T. Nguyen, S. Gupta, H. Ha, S. Rana, and S. Venkatesh. Distributionally robust Bayesian quadrature optimization. In *Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

[31] S. Ray Chowdhury and A. Gopalan. Bayesian optimization under heavy-tailed payoffs. In *Neural Information Processing Systems (NeurIPS)*, 2019.

[32] A. Sani, A. Lazaric, and R. Munos. Risk-aversion in multi-armed bandits. In *Neural Information Processing Systems (NeurIPS)*, 2012.

[33] P. G. Sessa, I. Bogunovic, M. Kamgarpour, and A. Krause. Mixed strategies for robust optimization of unknown objectives. In *Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

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

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

A Appendix

Risk-averse Heteroscedastic Bayesian Optimization

(Anastasiia Makarova, Ilnura Usmanova, Ilija Bogunovic, Andreas Krause)

A.1 Details on Proposition 1

We first provide the proof of Proposition 1 for cumulative risk-averse regret Eq. (3) with known variance-proxy $\rho^2(\cdot)$ (see Definition 1) (Appendix A.1.1). We further provide data-independent bounds for β_T (Appendix A.1.2) and maximum information gain γ_T (Appendix A.1.3) that together conclude the proof for sub-linear on T regret guarantees for most of the popularly used kernels.

A.1.1 Proof Proposition 1

Proposition 1. *Consider any $f \in \mathcal{H}_\kappa$ with $\|f\|_\kappa \leq \mathcal{B}_f$ and sampling model from Eq.* (1) *with known variance-proxy $\rho^2(x)$. Let $\{\beta_t\}_{t=1}^T$ be set as in Lemma 1 with $\lambda = 1$. Then, with probability at least $1-\delta$,* RAHBO *attains cumulative risk-averse regret $R_T = \mathcal{O}\big(\beta_T\sqrt{T\gamma_T(\bar{\varrho}^2+1)}\big)$.*

Proof. The main steps of the proof are as follows: In *Step 1*, we derive the upper and the lower confidence bounds, $\mathrm{ucb}_t^{\mathrm{MV}}(x_t)$ and $\mathrm{lcb}_t^{\mathrm{MV}}(x_t)$, on $\mathrm{MV}(x_t)$ at iteration t. In *Step 2*, we bound the instantaneous risk-averse regret $r(x_t) := \mathrm{MV}(x^*) - \mathrm{MV}(x_t)$. In *Step 3*, we derive mutual information $I(y_{1:T}, f_{1:T})$ in case of the heteroscedastic noise. In *Step 4*, we bound the sum of variances $\sum_{t=1}^T \sigma_{t-1}(x_t)$ via mutual information $I(y_{1:T}, f_{1:T})$. In *Step 5*, we bound the cumulative regret $R_T = \sum_{t=1}^T r(x_t)$ based on the previous steps.

Step 1: *On the confidence bounds for MV(x).*
In case of known variance-proxy $\rho^2(x)$, the confidence bounds for $\mathrm{MV}(x)$ at iteration t can be directly obtained based on the posterior $\mu_t(x)$ and $\sigma_t(x)$ for $f(x)$ defined in Eqs. (5) and (6). Particularly, for $\beta_t = \beta_t(\delta)$ defined in Eq. (8), $\Pr\big\{\mathrm{lcb}_t^{\mathrm{MV}}(x) \leq \mathrm{MV}(x) \leq \mathrm{ucb}_t^{\mathrm{MV}}(x)\ \forall x \in \mathcal{X}, \forall t \geq 0\big\} \geq 1-\delta$ with the confidence bounds:

$$\mathrm{lcb}_t^{\mathrm{MV}}(x) := \mu_{t-1}(x) - \beta_t\sigma_{t-1}(x) - \alpha\rho(x)^2, \tag{18}$$

$$\mathrm{ucb}_t^{\mathrm{MV}}(x) := \mu_{t-1}(x) + \beta_t\sigma_{t-1}(x) - \alpha\rho(x)^2. \tag{19}$$

Step 2: *On bounding the instantaneous risk-averse regret $r_t(x)$.* We have

$$\begin{aligned} r(x_t) &= \mathrm{MV}(x^*) - \mathrm{MV}(x_t) \\ &\leq \mathrm{ucb}_t^{\mathrm{MV}}(x^*) - \mathrm{lcb}_t^{\mathrm{MV}}(x_t) \\ &\leq \mathrm{ucb}_t^{\mathrm{MV}}(x_t) - \mathrm{lcb}_t^{\mathrm{MV}}(x_t) = 2\beta_t\sigma_{t-1}(x_t), \end{aligned}$$

where the first inequality is due to the definition of confidence bounds, the second is due to the acquisition strategy $x_t \in \arg\max_{x\in\mathcal{X}} \mathrm{ucb}_t^{\mathrm{MV}}(x)$; and the equality further expands $\mathrm{lcb}_t^{\mathrm{MV}}(x)$ and $\mathrm{ucb}_t^{\mathrm{MV}}(x)$. Thus, the cumulative regret can be bounded as follows:

$$R_T = \sum_{t=1}^T r(x_t) \leq \sum_{t=1}^T 2\beta_t\sigma_{t-1}(x_t) \leq 2\beta_T \sum_{t=1}^T \sigma_{t-1}(x_t), \tag{20}$$

where the last inequality holds since $\{\beta_t\}_{t=1}^T$ is a non-decreasing sequence.

Step 3: *On mutual information $I(y_{1:T}, f_{1:T})$ and maximum information gain γ_T.*
Mutual information $I(y_{1:T}, f_{1:T})$ between the vector of evaluations $y_{1:T} \in \mathbb{R}^T$ at points $A = \{x_t\}_{t=1}^T$ and $f_{1:T} = [f(x_1), \ldots, f(x_T)]^\top$ is defined by

$$I(y_{1:T}, f_{1:T}) = H(y_{1:T}) - H(y_{1:T}|f_{1:T}),$$

where $H(\cdot)$ denotes entropy. Under the modelling assumptions $f_{1:T} \sim \mathcal{N}(0, \lambda^{-1}K_T)$ and $\xi_{1:T} \sim \mathcal{N}(0, \Sigma_T)$ for the noise $\xi_{1:T} = [\xi(x_1), \ldots, \xi(x_T)]^\top$, the measurements are distributed as $y_{1:T} \sim \mathcal{N}(0, \lambda^{-1}K_T + \Sigma_T)$ and $y_t|y_{1:t-1} \sim \mathcal{N}(\mu_{t-1}(x_t), \rho^2(x_t) + \sigma_{t-1}^2(x_t))$, where $\sigma_{t-1}^2(\cdot)$ is defined as in

Eq. (6). Hence, the entropy of each new measurement y_t conditioned on the previous history $y_{1:t-1}$ is:

$$\begin{aligned}
H(y_t|y_{1:t-1}) &= \frac{1}{2}\ln\left(2\pi e\left(\rho^2(x_t)+\sigma^2_{t-1}(x_t)\right)\right)\\
&= \frac{1}{2}\ln\left(2\pi e\rho^2(x_t)\left(1+\frac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}\right)\right)\\
&= \frac{1}{2}\ln\left(2\pi e\rho^2(x_t)\right)+\frac{1}{2}\ln\left(1+\frac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}\right),\\
H(y_{1:T}) &= \sum_{t=1}^{T} H(y_t|y_{1:t-1}) = \frac{1}{2}\sum_{t=1}^{T}\ln\left(2\pi e\rho^2(x_t)\right)+\frac{1}{2}\sum_{t=1}^{T}\ln\left(1+\frac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}\right),\\
H(y_{1:T}|f_{1:T}) &= \sum_{t=1}^{T} H(y_t|f_t) = \frac{1}{2}\sum_{t=1}^{T}\ln(2\pi e\rho^2(x_t)).
\end{aligned}$$

Therefore, the information gain for $y_{1:T}$ is:

$$I(y_{1:T}, f_{1:T}) = H(y_{1:T}) - H(y_{1:T}|f_{1:T}) = \frac{1}{2}\sum_{t=1}^{T}\ln\left(1+\frac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}\right). \tag{21}$$

Then, by definition of maximum information gain:

$$\gamma_T := \max_{A\subset\mathcal{X},\,|A|=T} I(y_{1:T}, f_{1:T}) \geq \frac{1}{2}\sum_{t=1}^{T}\ln\left(1+\frac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}\right). \tag{22}$$

Step 4: *On bounding* $\sum_{t=1}^{T}\sigma_{t-1}(x_t)$.

$$\begin{aligned}
\sum_{t=1}^{T}\sigma_{t-1}(x_t) &= \sum_{t=1}^{T}\frac{\rho(x_t)}{\rho(x_t)}\sigma_{t-1}(x_t) \leq \sqrt{T\sum_{t=1}^{T}\rho^2(x_t)\frac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}}\\
&\leq \sqrt{T\sum_{t=1}^{T}\frac{1}{\ln(1+\rho^{-2}(x_t))}\ln\left(1+\frac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}\right)}\\
&\leq \sqrt{\frac{2T}{\ln(1+\bar{\varrho}^{-2})}\underbrace{\frac{1}{2}\sum_{t=1}^{T}\ln\left(1+\frac{\sigma^2_{t-1}(x_t)}{\rho^2(x_t)}\right)}_{\texttt{mutual information Eq. (21)}}},
\end{aligned} \tag{23}$$

where the first inequality follows from the Cauchy-Schwarz inequality. The second one is due to the fact that for any $s^2 \in [0, \rho^{-2}(x_t)]$ we can bound $s^2 \leq \frac{\rho^{-2}(x_t)}{\ln(1+\rho^{-2}(x_t))}\ln(1+s^2)$, that also holds for $s^2 := \rho^{-2}(x_t)\sigma^2_{t-1}(x_t)$ since $\rho^{-2}(x_t)\sigma^2_{t-1}(x_t) \leq \rho^{-2}(x_t)\lambda^{-1}\kappa(x_t,x_t) \leq \rho^{-2}(x_t)$ for $\lambda \geq 1$. The third inequality is due to $\rho(x) \in [\underline{\varrho}, \bar{\varrho}]$.

Step 5: *Bounding risk-averse cumulative regret* $R_T = \sum_{t=1}^{T} r(x_t)$.
Combining the previous three steps together: Eq. (20), Eq. (22), and Eq. (23) we finally obtain:

$$R_T \leq \sum_{t=1}^{T} 2\beta_t\sigma_{t-1}(x_t) \leq 2\beta_T\sum_{t=1}^{T}\sigma_{t-1}(x_t) \leq 2\beta_T\sqrt{\frac{2T}{\ln(1+\bar{\varrho}^{-2})}\gamma_T}$$

Also, note that for any $\alpha \geq 0$ the bound $\ln(1+\alpha) \geq \frac{\alpha}{1+\alpha}$ holds, thus $\frac{1}{\ln(1+\bar{\varrho}^{-2})} \leq \frac{1+\bar{\varrho}^{-2}}{\bar{\varrho}^{-2}} = \bar{\varrho}^2+1$. Therefore, the cumulative regret can be also bounded as $R_T = O(\beta_T\sqrt{T\gamma_T(\bar{\varrho}^2+1)})$. □

A.1.2 Bounds for β_T

We provide the bounds for the data-dependent β_T that appear in the regret bound (see Eq. (8)). Following our modelling assumptions $f_{1:T} \sim \mathcal{N}(0, \lambda^{-1}K_T)$ and $\xi_{1:T} \sim \mathcal{N}(0, \Sigma_T)$, the information gain $I(y_{1:T}, f_{1:T}) = H(y_{1:T}) - H(y_{1:T}|f_{1:T})$ is given as follows:

$$I(y_{1:T}, f_{1:T}) = \underbrace{\frac{1}{2}\ln\big(\det(2\pi e(\lambda^{-1}K_T + \Sigma_T))\big)}_{H(y_{1:T})} - \underbrace{\frac{1}{2}\ln\big(\det(2\pi e\Sigma_T)\big)}_{H(y_{1:T}|f_{1:T})} = \frac{1}{2}\ln\left(\frac{\det(K_T + \lambda\Sigma_T)}{\det(\lambda\Sigma_T)}\right). \tag{24}$$

By definition then $\gamma_T = \max_{A\subset\mathcal{X},\,|A|=T} I(y_{1:T}, f_{1:T}) \geq \frac{1}{2}\ln\left(\frac{\det(K_T+\lambda\Sigma_T)}{\det(\lambda\Sigma_T)}\right)$. On the other hand, β_T defined in Lemma 1 can be expanded in a data-independent manner as follows:

$$\begin{aligned}\beta_T &:= \sqrt{2\ln\left(\frac{\det(\lambda\Sigma_T + K_T)^{1/2}}{\delta\det(\lambda\Sigma_T)^{1/2}}\right)} + \sqrt{\lambda}\|f\|_\kappa \\ &= \sqrt{2\ln\frac{1}{\delta} + \ln\left(\frac{\det(\lambda\Sigma_T + K_T)}{\det(\lambda\Sigma_T)}\right)} + \sqrt{\lambda}\|f\|_\kappa \leq \sqrt{2\ln\frac{1}{\delta} + \gamma_T} + \sqrt{\lambda}B_f. \end{aligned} \tag{25}$$

A.1.3 Bounds for γ_T

Here, we show the relation between the information gains under heteroscedastic and homoscedastic noise. Note that for the latter the upper bounds are widely known, e.g., [35]. To distinguish between the maximum information gain for heteroscedastic noise with variance-proxy $\rho^2(x)$ and the maximum information gain for homoscedastic noise with fixed variance-proxy σ^2, we denote them as $\gamma_T^{\rho_x}$ and γ_T^{σ} respectively. Recall that $\varrho^2(\cdot) \in [\underline{\varrho}^2, \bar{\varrho}^2]$ for some constant values $\bar{\varrho}^2 \geq \underline{\varrho}^2 > 0$.

Below, we show that $\gamma_T^{\rho_x} \leq \gamma_T^{\sigma}\frac{\bar{\varrho}^2}{\underline{\varrho}^2}$ with σ^2 set to $\bar{\varrho}^2$, that only affects the constants but not the main scaling (in terms of T) of the known bound for the homoscedastic maximum information gain.

$$\gamma_T^{\rho_x} \overset{①}{=} \max_{A\subset\mathcal{X},|A|=T}\frac{1}{2}\sum_{t=1}^{T}\ln\left(1 + \frac{\sigma_{t-1}^2(x_t|\rho^2(x_t))}{\rho^2(x_t)}\right) \overset{②}{\leq} \max_{A\subset\mathcal{X},|A|=T}\frac{1}{2}\sum_{t=1}^{T}\ln\left(1 + \frac{\sigma_{t-1}^2(x_t|\bar{\varrho}^2)}{\underline{\varrho}^2}\right) \tag{26}$$

$$\overset{③}{=} \max_{A\subset\mathcal{X},|A|=T}\frac{1}{2}\sum_{t=1}^{T}\ln\left(1 + \frac{\bar{\varrho}^2}{\underline{\varrho}^2}\frac{\sigma_{t-1}^2(x_t|\bar{\varrho}^2)}{\bar{\varrho}^2}\right) \overset{④}{\leq} \max_{A\subset\mathcal{X},|A|=T}\frac{1}{2}\sum_{t=1}^{T}\frac{\bar{\varrho}^2}{\underline{\varrho}^2}\ln\left(1 + \frac{\sigma_{t-1}^2(x_t|\bar{\varrho}^2)}{\bar{\varrho}^2}\right) \tag{27}$$

$$\overset{⑤}{=} \max_{A\subset\mathcal{X},|A|=T}\frac{\bar{\varrho}^2}{\underline{\varrho}^2}\frac{1}{2}\sum_{t=1}^{T}\ln\left(1 + \frac{\sigma_{t-1}^2(x_t|\sigma^2)}{\sigma^2}\right) = \frac{\bar{\varrho}^2}{\underline{\varrho}^2}\gamma_T^{\sigma}, \tag{28}$$

where ① follows from Eq. (21). In ②, we lower bound the denominator $\rho^2(x_t)$ and upper bound the numerator $\sigma_{t-1}^2(x_t)$ (due to monotonicity w.r.t. noise variance, i.e., $\sigma_{t-1}^2(x_t|\Sigma_t) \leq \sigma_{t-1}^2(x_t|\bar{\varrho}^2\mathbf{I}_t)$). In ③, we multiply by $1 = \bar{\varrho}^2/\bar{\varrho}^2$. In ④ we use Bernoulli inequality since $\bar{\varrho}^2/\underline{\varrho}^2 \geq 1$. The obtained expression can be interpreted as a standard information gain for homoscedastic noise and, particularly, with the variance-proxy σ^2 set to $\bar{\varrho}^2$ due to ⑤. Finally, the upper bounds on γ_T^{σ} typically scale sublinearly in T for most of the popularly used kernels [35], e.g, for linear kernel $\gamma_T = \mathcal{O}(d\log T)$, and for squared exponential kernel $\gamma_T = \mathcal{O}(d(\log T)^{d+1})$.

A.2 Tighter bounds for the variance-proxy $\rho_\eta^2(x)$.

Assumption 2 states that noise $\eta(x)$ from Eq. (13) is $\rho_\eta^2(x)$-sub-Gaussian with variance-proxy $\rho_\eta^2(x)$ being known. In practice, $\rho_\eta^2(x)$ might be unknown. Here, we describe a way to estimate $\rho_\eta^2(x)$ under

the following two assumptions: the evaluation noise $\xi(x)$ is strictly sub-Gaussian (that is already reflected in the Assumption 1) and the noise $\eta(x)$ of variance evaluation is also strictly sub-Gaussian, that is, $\mathbb{V}ar[\eta(x)] = \rho_\eta^2(x)$ and $\mathbb{V}ar[\xi(x)] = \rho^2(x)$.

(i) Reformulation of the sample variance. We first rewrite the sample variance defined in Eq. (12) as the average over squared differences over all pairs $\{y_1(x), \dots, y_k(x)\}$:

$$\hat{s}_k^2(x) \overset{①}{=} \frac{1}{2k(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{k} (y_i(x) - y_j(x))^2 \overset{②}{=} \frac{1}{2k(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{k} (\xi_i(x) - \xi_j(x))^2, \tag{29}$$

where ② is due to $y_i(x) = f(x) + \xi_i(x)$, and ① is equivalent to the Eq. (12) due to the following:

$$\frac{1}{2k(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{k} (y_i - y_j)^2 = \frac{1}{2k(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{k} (y_i - \hat{m}_k + \hat{m}_k - y_j)^2 \tag{30}$$

$$= \frac{1}{2k(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{k} \left[\big(y_i - \hat{m}_k\big)^2 + \big(y_j - \hat{m}_k\big)^2 - 2\big(y_i - \hat{m}_k\big)\big(y_j - \hat{m}_k\big) \right]$$

$$= \frac{1}{2k(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{k} \left[\big(y_i - \hat{m}_k\big)^2 + \big(y_j - \hat{m}_k\big)^2 \right]$$

$$= \frac{1}{k-1} \sum_{i=1}^{k} (y_i - \hat{m}_k)^2. \tag{31}$$

(ii) Variance of the sample variance $\mathbb{V}ar[\hat{s}_k^2(x)]$. In Eq. (29), we show that sample variance can be written in terms of the noise $\xi(x)$. In [4] (see Eq. (37)), it is shown that for i.i.d observations $\{\xi_1(x), \dots, \xi_k(x)\}$, sampled from a distribution with the 2nd and 4th central moments being $\mathbb{V}ar[\xi(x)]$ and $\mu_4(x) = \mathbb{E}[\xi^4(x)]$, the variance of the sample variance can be computed as follows:

$$\mathbb{V}ar[\hat{s}_k^2(x)] = \mathbb{E}[(\hat{s}_k^2(x))^2] - \mathbb{E}[\hat{s}_k^2(x)]^2 = \frac{\mu_4(x)}{k} - \frac{(k-3)\mathbb{V}ar^2[\xi(x)]}{k(k-1)}.$$

Since $\xi(x)$ is strictly $\rho(x)$–sub-Gaussian, the latter can be further adapted as

$$\mathbb{V}ar[\hat{s}_k^2(x)] = \frac{\mu_4(x)}{k} - \frac{(k-3)\rho^4(x)}{k(k-1)}.$$

(iii) Due to $\eta(x)$ being strictly sub-Gaussian, i.e., $\rho_\eta^2(x) = \mathbb{V}ar[\eta(x)] = \mathbb{V}ar[\hat{s}_k^2(x)]$, the derivation above also holds for the variance-proxy $\rho_\eta^2(x)$:

$$\rho_\eta^2(x) = \frac{\mu_4(x)}{k} - \frac{(k-3)\rho^4(x)}{k(k-1)}.$$

(iv) Bound 4th moment $\mu_4(x)$. The 4th moment $\mu_4(x)$ can expressed in terms of the distribution kurtosis that is bounded under our assumptions. Particularly, *kurtosis* $\mathrm{Kurt}[\xi] := \frac{\mathbb{E}[(\xi - \mathbb{E}[\xi])^4]}{\mathbb{V}ar^2(\xi)}$ is measure that identifies the tails behaviour of the distribution of ξ; $\mathrm{Kurt}(\xi) = 3$ for normallly distribute ξ and $\mathrm{Kurt}(\xi) \leq 3$ for strictly sub-Gaussian random variable ξ (see [2]). This implies

$$\mu_4(x) = \mathrm{Kurt}(\xi(x))\rho^4(x) \leq 3\rho^4(x).$$

(v) Bound variance-proxy. There

$$\rho_\eta^2(x) \leq \frac{3(k-1)\rho^4(x) - (k-3)\rho^4(x)}{k(k-1)} = \frac{3k - 3 - k + 3}{k(k-1)}\rho^4(x) = \frac{2\rho^4(x)}{k-1}.$$

In case of the known bound $\bar{\varrho}^2 \geq \rho^2(x)$, we bound the unknown $\rho_\eta^2(x)$ as follows:

$$\rho_\eta^2(x) \leq \frac{2\bar{\varrho}^4}{k-1}.$$

A.3 Method details: GP-estimator of variance-proxy ρ^2

According to the Assumption 2, variance-proxy $\rho^2 \in \mathcal{H}_{\kappa^{var}}$ is smooth, and $\eta(x) = \hat{s}_k^2(x) - \rho^2(x)$ is $\rho_\eta(x)$-sub-Gaussian with known variance-proxy $\rho_\eta^2(x)$. In this case, confidence bounds for $\rho^2(x)$ follow the ones derived in Lemma 1 with β_t^{var} based on Σ_t^{var}. Particularly, we collect noise variance evaluations $\{x_t, \hat{s}_k(x_t)\}_{t=0}^T$. Then the estimates for $\mu_T^{var}(x)$ and $\sigma_T^{var}(x)$ for ρ^2 follow the corresponding estimates for $f(x)$. Particularly,

$$\mu_t^{var}(x) = \kappa_t^{var}(x)^T (K_t^{var} + \lambda \Sigma_t^{var})^{-1} \hat{s}_{1:t}, \tag{32}$$

$$\sigma_t^{var}(x)^2 = \frac{1}{\lambda}(\kappa^{var}(x,x) - \kappa_t^{var}(x)^\top (K_t^{var} + \lambda \Sigma_t^{var})^{-1} \kappa_t^{var}(x)), \tag{33}$$

where $\Sigma_t^{var} = \text{diag}[\rho_\eta^2(x_1), \ldots, \rho_\eta^2(x_t)]$, $\kappa_t^{var}(x) = [\kappa^{var}(x_1, x), \ldots, \kappa^{var}(x_t, x)]^T$ and $(K_t^{var})_{i,j} = \kappa^{var}(x_i, x_j)$. The confidence bounds are then:

$$\text{ucb}_t^{var}(x) = \mu_{t-1}^{var}(x) + \beta_t^{var} \sigma_{t-1}^{var}(x)$$
$$\text{lcb}_t^{var}(x) = \mu_{t-1}^{var}(x) - \beta_t^{var} \sigma_{t-1}^{var}(x),$$

with $\{\beta_t^{var}\}_{t=1}^T$ set according to Lemma 1.

A.4 Proof of Theorem 1

Theorem 1. *Consider any $f \in \mathcal{H}_\kappa$ with $\|f\|_\kappa \leq \mathcal{B}_f$ and sampling model in Eq.* (1) *with unknown variance-proxy $\rho^2(x)$ that satisfies Assumptions 1 and 2. Let $\{x_t\}_{t=1}^T$ denote the set of actions chosen by* RAHBO *(Algorithm 1) over T rounds. Set $\{\beta_t\}_{t=1}^T$ and $\{\beta_t^{var}\}_{t=1}^T$ according to Lemma 1 with $\lambda = 1$, $\mathcal{R}^2 = \max_{x \in \mathcal{X}} \rho_\eta^2(x_t)$ and $\rho(\cdot) \in [\underline{\varrho}, \bar{\varrho}]$. Then, the risk-averse cumulative regret R_T of* RAHBO *is bounded as follows:*

$$\Pr\left\{ R_T \leq \beta_T k \sqrt{\frac{2T\hat{\gamma}_T}{\ln(1 + k/\bar{\varrho}^2)}} + \alpha \beta_T^{var} k \sqrt{\frac{2T\Gamma_T}{\ln(1 + \mathcal{R}^{-2})}}, \quad \forall T \geq 1 \right\} \geq 1 - \delta. \tag{34}$$

Proof. The main steps of our proof are as follows: In *Step 1*, we derive the upper and the lower confidence bounds, $\text{ucb}_t^{\text{MV}}(x_t)$ and $\text{lcb}_t^{\text{MV}}(x_t)$, on $\text{MV}(x_t)$ at iteration t. In *Step 2*, we bound the instantaneous risk-averse regret $r(x_t) := \text{MV}(x^*) - \text{MV}(x_t)$. In *Step 3*, we derive mutual information both for function and variance-proxy evaluations. In *Step 4*, we bound the sum of variances via mutual information. In *Step 5*, we bound the cumulative regret $R_T = \sum_{t=1}^T r(x_t)$ based on the previous steps.

Step 1: *On confidence bounds for MV(x).*

(i) On confidence bounds for $\rho^2(x)$. According to Eq. (33), with probability $1 - \delta$ the following confidence bounds hold with $\{\beta_t^{var}\}_{t=1}^T$ set according to Lemma 1:

$$\text{ucb}_t^{var}(x) = \mu_{t-1}^{var}(x) + \beta_t^{var} \sigma_{t-1}^{var}(x),$$
$$\text{lcb}_t^{var}(x) = \mu_{t-1}^{var}(x) - \beta_t^{var} \sigma_{t-1}^{var}(x).$$

(ii) On confidence bounds for $f(x)$. Here we adapt confidence bounds introduced in Eq. (18)-(19) since Eq. (5) relies on the unknown variance-proxy $\rho^2(x)$ incorporated into Σ_T. Conditioning on the event that $\rho^2(x)$ is upper bounded by $\text{ucb}_t^{var}(x) \geq \rho(x)^2$ defined in (i), the confidence bounds for f with probability $1 - \delta$ are:

$$\text{ucb}_t^f(x) = \mu_{t-1}(x|\hat{\Sigma}_t) + \beta_t \sigma_{t-1}(x|\hat{\Sigma}_t), \tag{35}$$

$$\text{lcb}_t^f(x) = \mu_{t-1}(x|\hat{\Sigma}_t) - \beta_t \sigma_{t-1}(x|\hat{\Sigma}_t), \ \forall x, t. \tag{36}$$

(iii) On confidence bounds for MV(x). Finally, combining (i) and (ii) and using the union bound, with probability $1 - 2\delta$, we get $\text{lcb}_t^{\text{MV}}(x) \leq \text{MV}(x) \leq \text{ucb}_t^{\text{MV}}(x)$ with

$$\text{ucb}_t^{\text{MV}}(x) = \text{ucb}_t^f(x) - \alpha \text{lcb}_t^{var}(x), \tag{37}$$

$$\text{lcb}_t^{\text{MV}}(x) = \text{lcb}_t^f(x) - \alpha \text{ucb}_t^{var}(x), \ \forall x, t. \tag{38}$$

Step 2: *On bounding the instantaneous regret.*
First, we bound instantaneous regret of a single measurement at point x_t, but with unknown variance-proxy $\rho^2(x)$ as follows:

$$\begin{aligned} r_t := \text{MV}(x^*) - \text{MV}(x_t) &\leq \text{ucb}_t^{\text{MV}}(x^*) - \text{lcb}_t^{\text{MV}}(x_t) \\ &\leq \text{ucb}_t^{\text{MV}}(x_t) - \text{lcb}_t^{\text{MV}}(x_t) \\ &= \text{ucb}_t^{f}(x_t) - \text{lcb}_t^{f}(x_t) + \alpha(\text{ucb}_t^{var}(x_t) - \text{lcb}_t^{var}(x_t)) \\ &= 2\beta_t\sigma_{t-1}(x_t|\hat{\Sigma}_t) + 2\alpha\beta_t^{var}\sigma_{t-1}^{var}(x_t). \end{aligned} \tag{39}$$

The second inequality is due to the acquisition algorithm. The last equality is due to the fact that $\text{ucb}_t^f(x) - \text{lcb}_t^f(x) = 2\beta_t\sigma_{t-1}(x_t)$ by definition, as well as $\text{ucb}_t^{var}(x) - \text{lcb}_t^{var}(x) = 2\beta_t^{var}\sigma_{t-1}^{var}(x_t)$.

Note that at each iteration t we take k measurements, hence the total number of measurements is Tk. Thus, we can bound the cumulative regret by

$$\begin{aligned} R_T = \sum_{t=1}^{T} kr(x_t) &\leq k\sum_{t=1}^{T} 2\beta_t\sigma_{t-1}(x_t|\hat{\Sigma}_t) + k\sum_{t=1}^{T} 2\alpha\beta_t^{var}\sigma_{t-1}^{var}(x_t) \\ &\leq 2k\beta_T\sum_{t=1}^{T}\sigma_{t-1}(x_t|\hat{\Sigma}_t) + 2k\alpha\beta_T^{var}\sum_{t=1}^{T}\sigma_{t-1}^{var}(x_t). \end{aligned} \tag{40}$$

Step 3*: On bounding maximum information gain.*
We follow the notion of information gain $I(\hat{m}_{1:T}, f_{1:T})$ computed assuming that $\hat{m}_{1:T} = [\hat{m}_k(x_1), \ldots, \hat{m}_k(x_T)]^T$ with $\hat{m}_k(x_t) = \frac{1}{k}\sum_{i=1}^{k} y_i(x_t)$ (Eq. (12)). Under the modelling assumptions $f_{1:T} \sim \mathcal{N}(0, \lambda^{-1}K_T)$, and $\hat{m}_{1:T} \sim \mathcal{N}(f_{1:T}, \text{diag}(\bar{\varrho}^2/k))$ with variance-proxy $\bar{\varrho}^2/k$, the information gain is:

$$I(\hat{m}_{1:T}, f_{1:T}) := \sum_{t=1}^{T}\frac{1}{2}\ln\left(1 + \frac{\sigma_{t-1}^2(x_t|\text{diag}(\bar{\varrho}^2/k))}{\bar{\varrho}^2/k}\right).$$

We define the corresponding maximum information gain $\hat{\gamma}_T = \max_{A\subset\mathcal{X},|A|=T} I(\hat{m}_{1:T}, f_{1:T})$

$$\hat{\gamma}_T := \max_{A\subset\mathcal{X},|A|=T}\sum_{t=1}^{T}\frac{1}{2}\ln\left(1 + \frac{\sigma_{t-1}^2(x_t|\text{diag}(\bar{\varrho}^2/k))}{\bar{\varrho}^2/k}\right). \tag{41}$$

Analogously, for $\rho(x)$ with the posterior $\mathcal{N}(\mu_t^{var}(x), (\sigma_t^{var}(x))^2)$, the information gain is defined as:

$$I(\hat{s}_{1:T}^2, \rho_{1:T}^2) := \frac{1}{2}\sum_{t=1}^{T}\ln\left(1 + \frac{(\sigma_{t-1}^{var})^2(x)}{\rho_\eta^2(x_t)}\right).$$

Then, the corresponding maximum information gain Γ_T is as follows:

$$\Gamma_T := \max_{A\subset\mathcal{X},|A|=T} I(\hat{s}_{1:T}^2, \rho_{1:T}^2) = \max_{A\subset\mathcal{X},|A|=T}\frac{1}{2}\sum_{t=1}^{T}\ln\left(1 + \frac{(\sigma_{t-1}^{var})^2(x)}{\rho_\eta^2(x_t)}\right), \tag{42}$$

where A is again a set of size T with points $\{x_1, \ldots, x_T\}$.

Step 4: *On bounding* $\sum_{t=1}^T \sigma_{t-1}(x_t|\hat{\Sigma}_t)$ *and* $\sum_{t=1}^T \sigma_{t-1}^{var}(x_t)$
We repeat the corresponding derivation for known $\rho^2(x)$, recalling that $\rho^2(x) \le \bar{\varrho}^2, \forall x \in \mathcal{X}$:

$$\begin{aligned}\sum_{t=1}^T \sigma_t(x_t|\hat{\Sigma}_t) &= \sum_{t=1}^T \frac{\bar{\varrho}}{\bar{\varrho}}\sigma_{t-1}(x_t|\hat{\Sigma}_t) \le \sqrt{T\sum_{t=1}^T \frac{\bar{\varrho}^2}{k}\frac{\sigma_{t-1}^2\big(x_t|\text{diag}(\bar{\varrho}^2/k)\big)}{\bar{\varrho}^2/k}} \\ &\le \sqrt{T\frac{\bar{\varrho}^2}{k}\frac{k/\bar{\varrho}^2}{\ln(1+k/\bar{\varrho}^2)}\sum_{t=1}^T \ln\left(1+\frac{\sigma_{t-1}^2\big(x_t|\text{diag}(\bar{\varrho}^2/k)\big)}{\bar{\varrho}^2/k}\right)} \\ &\le \sqrt{\frac{2T}{\ln(1+k/\bar{\varrho}^2)}\underbrace{\sum_{t=1}^T \frac{1}{2}\ln\left(1+\frac{\sigma_{t-1}^2\big(x_t|\text{diag}(\bar{\varrho}^2/k)\big)}{\bar{\varrho}^2/k}\right)}_{\text{mutual information}}}. \end{aligned} \tag{43}$$

Here, the first inequality follows from Cauchy-Schwarz inequality and the fact that $\sigma_t(x_t|\hat{\Sigma}_t) \le \sigma_t(x_t|\text{diag}(\bar{\varrho}^2/k))$. The latter holds by the definition of $\hat{\Sigma}_t$, particularly:

$$\begin{aligned}\sigma_t^2(x_t|\hat{\Sigma}_t) &= \frac{1}{\lambda}(\kappa(x,x) - \kappa_t(x)^\top(K_t+\lambda\hat{\Sigma}_t)^{-1}\kappa_t(x)), \\ \sigma_t^2\big(x_t|\text{diag}(\bar{\varrho}^2/k)\big) &= \frac{1}{\lambda}(\kappa(x,x) - \kappa_t(x)^\top\big(K_t+\lambda\text{diag}(\bar{\varrho}^2/k)\big)^{-1}\kappa_t(x)), \\ \hat{\Sigma}_t &= \tfrac{1}{k}\text{diag}\big(\min\{\text{ucb}_t^{var}(x_1),\bar{\varrho}^2\},\dots,\min\{\text{ucb}_t^{var}(x_t),\bar{\varrho}^2\}\big),\end{aligned}$$

then $\hat{\Sigma}_t \preceq \text{diag}(\bar{\varrho}^2/k)$, and $-(K_t+\lambda\hat{\Sigma}_t)^{-1} \preceq -(K_t+\lambda\text{diag}(\bar{\varrho}^2/k))^{-1}$. That implies

$$\begin{aligned}\sigma_t^2(x_t|\hat{\Sigma}_t) - \sigma_t^2\big(x_t|\text{diag}(\bar{\varrho}^2/k)\big) &= -\kappa_t(x)^\top\big(K_t+\lambda\hat{\Sigma}_t)^{-1}\kappa_t(x)\big) \\ &\quad + \kappa_t(x)^\top(K_t+\lambda\text{diag}(\bar{\varrho}^2/k))^{-1}\kappa_t(x)\big) \\ &\le 0.\end{aligned}$$

The second inequality in Eq. (43) is due to the fact that for any $s^2 \in [0, k/\bar{\varrho}^2(x_t)]$ we can bound $s^2 \le \frac{k/\bar{\varrho}^2(x_t)}{\ln(1+k/\bar{\varrho}^2(x_t))}\ln(1+s^2)$, that also holds for $s^2 := \frac{\sigma_{t-1}^2\left(x_t|\text{diag}(\bar{\varrho}^2/k)\right)}{\bar{\varrho}^2/k}$ since for any $\lambda \ge 1$

$$\frac{\sigma_{t-1}^2\big(x_t|\text{diag}(\bar{\varrho}^2/k)\big)}{\bar{\varrho}^2/k} \le \frac{\lambda^{-1}\kappa(x_t,x_t)}{\bar{\varrho}^2/k} \le k/\bar{\varrho}^2.$$

Similarly, we bound

$$\begin{aligned}\sum_{t=1}^T \sigma_{t-1}^{var}(x_t) &= \sum_{t=1}^T \frac{\rho_\eta(x_t)}{\rho_\eta(x_t)}\sigma_{t-1}^{var}(x_t) \le \sqrt{T\sum_{t=1}^T \rho_\eta^2(x_t)\frac{(\sigma_{t-1}^{var})^2(x_t)}{\rho_\eta^2(x_t)}} \\ &\le \sqrt{\frac{2T}{\ln(1+\mathcal{R}^{-2})}\underbrace{\sum_{t=1}^T \frac{1}{2}\ln\left(1+\frac{(\sigma_{t-1}^{var})^2(x_t)}{\rho_\eta^2(x_t)}\right)}_{\text{mututal information}}},\end{aligned} \tag{44}$$

in the above we define $\mathcal{R}^2 := \max_{x\in A}\rho_\eta^2(x), A = \{x_1,\dots,x_T\}$.

Step 5: *On bounding cumulative regret* $R_T = \sum_{t=1}^T kr(x_t)$
Combining the above three steps together, we obtain with probability $1-2\delta$

$$R_T \le \beta_T k\sqrt{\frac{2T\hat{\gamma}_T}{\ln(1+k/\bar{\varrho}^2)}} + \alpha\beta_T^{var}k\sqrt{\frac{2T\Gamma_T}{\ln(1+\mathcal{R}^{-2})}}. \tag{45}$$

A.5 Proof of Corollary 1.1

Corollary 1.1 *Consider the setup of Theorem 1. Let $A = \{x_t\}_{t=1}^T$ denote actions selected by* RAHBO *over T rounds. Then, with probability at least $1-\delta$, the reported point $\hat{x}_T := \arg\max_{x_t \in A} \mathrm{lcb}_t^{MV}(x_t)$, where $\mathrm{lcb}_t^{MV}(x_t) = \mathrm{lcb}_t^f(x) - \alpha\ \mathrm{ucb}_t^{var}(x)$, achieves ϵ-accuracy, i.e., MV(x^*) − MV$(\hat{x}_T) \leq \epsilon$, after $T \geq \frac{4\beta_T^2\hat{\gamma}_T/\ln(1+k/\bar{\varrho}^2)+4\alpha(\beta_t^{var})^2\Gamma_T/\ln(1+\mathcal{R}^{-2})}{\epsilon^2}$ rounds.*

Proof. We select the maximizer of $\mathrm{lcb}_t^{\mathrm{MV}}(x_t)$ over the past points x_t:

$$\hat{x}_T := x_{t^*}, \text{ where } t^* := \arg\max_t \{\mathrm{lcb}_t^{\mathrm{MV}}(x_t)\} = \arg\min_t \{\mathrm{MV}(x^*) - \mathrm{lcb}_t^{\mathrm{MV}}(x_t)\},$$

since adding a constant does not change the solution. We denote $\hat{r}(x_t) := \mathrm{MV}(x^*) - \mathrm{lcb}_t^{\mathrm{MV}}(x_t)$. Then we obtain the following bound

$$\begin{aligned}
\mathrm{MV}(x^*) - \mathrm{MV}(x_{t^*}) \leq \mathrm{MV}(x^*) - \mathrm{lcb}_{t^*}^{\mathrm{MV}}(x_{t^*}) &= \frac{1}{T}\sum_{t=1}^T \hat{r}(x_{t^*}) \\
&\leq \frac{1}{T}\sum_{t=1}^T \hat{r}(x_t) = \frac{1}{T}\sum_{t=1}^T \left(\mathrm{MV}(x^*) - \mathrm{lcb}_t^{\mathrm{MV}}(x_t)\right) \\
&\leq \frac{1}{T}\sum_{t=1}^T \left(\mathrm{ucb}_t^{\mathrm{MV}}(x^*) - \mathrm{lcb}_t^{\mathrm{MV}}(x_t)\right) \\
&\leq \frac{1}{T}\sum_{t=1}^T \left(\mathrm{ucb}_t^{\mathrm{MV}}(x_t) - \mathrm{lcb}_t^{\mathrm{MV}}(x_t)\right). \qquad (46)
\end{aligned}$$

In the above, the first inequality holds with high probability by definition $\mathrm{lcb}_{t^*}^{\mathrm{MV}}(x_{t^*}) \leq \mathrm{MV}(x_{t^*})$, the second inequality is due to $t^* := \arg\min_t \hat{r}(x_t)$ and therefore $\hat{r}(x_{t^*}) \leq \hat{r}(x_t)\ \forall t = 1, \ldots, T$. The third inequality holds since $\mathrm{ucb}_t^{\mathrm{MV}}(x) \geq \mathrm{MV}(x)$ with high probability, and the fourth is due to $\mathrm{ucb}_t^{\mathrm{MV}}(x_t) \geq \mathrm{ucb}_t^{\mathrm{MV}}(x)$ for every x, since x_t is selected via Algorithm 1.

Recalling Eq. (45), note that the following bounds hold:

$$R_T = \sum_{t=1}^T kr(x_t) \leq \sum_{t=1}^T k(\mathrm{ucb}_t^{\mathrm{MV}}(x_t) - \mathrm{lcb}_t^{\mathrm{MV}}(x_t)) \leq \beta_T k\sqrt{\frac{2T\hat{\gamma}_T}{\ln(1+k/\bar{\varrho}^2)}} + \alpha\beta_T^{var} k\sqrt{\frac{2T\Gamma_T}{\ln(1+\mathcal{R}^{-2})}}. \qquad (47)$$

Combining the above Eq. (47) with Eq. (46) we can get the following upper bound

$$\begin{aligned}
\mathrm{MV}(x^*) - \mathrm{MV}(x_{t^*}) &\leq \frac{\beta_T k\sqrt{2T\hat{\gamma}_T/\ln(1+k/\bar{\varrho}^2)} + \alpha\beta_T^{var} k\sqrt{2T\Gamma_T/\ln(1+\mathcal{R}^{-2})}}{kT} \\
&\leq \frac{\sqrt{4\left(k\beta_T^2\hat{\gamma}_T/\ln(1+k/\bar{\varrho}^2) + \alpha k(\beta_t^{var})^2\Gamma_T/\ln(1+\mathcal{R}^{-2})\right)}}{\sqrt{kT}}.
\end{aligned}$$

Therefore, for Tk samples with $Tk \geq \frac{4(k\beta_T^2\hat{\gamma}_T/\ln(1+k/\bar{\varrho}^2)+\alpha k(\beta_t^{var})^2\Gamma_T/\ln(1+\mathcal{R}^{-2}))}{\epsilon^2}$ we finally obtain

$$\mathrm{MV}(x^*) - \mathrm{MV}(x_{t^*}) \leq \epsilon.$$

□

A.6 Experimental settings and extended results

Implementation and resources We implemented all our experiments using Python and BoTorch [3].[3] We ran our experiments on an Intel(R) Xeon(R) CPU E5-2699 v3 @ 2.30GHz machine.

[3] `https://botorch.org/`

A.6.1 Example function

We provide additional visualizations for the example sine function in Fig. 6. These examples demonstrate that exploration-exploitation trade-off (as in GP-UCB) might not be enough to prefer points with lower noise and GP-UCB might tend to acquire points with higher variance. In contrast, RAHBO, initialized with the same point, prefers points with lower risk inherited in noise.

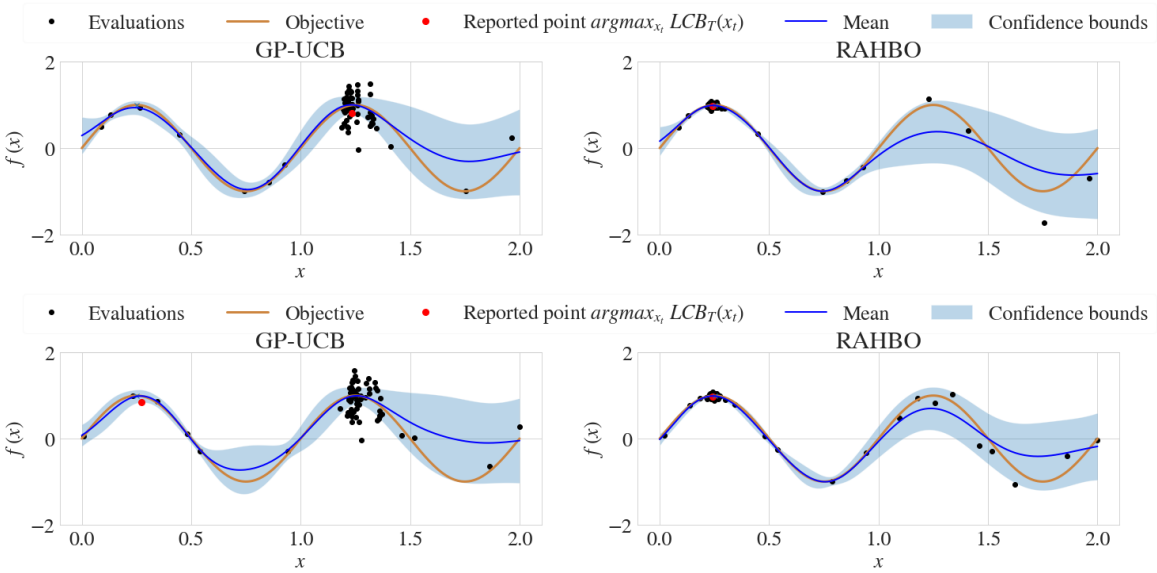
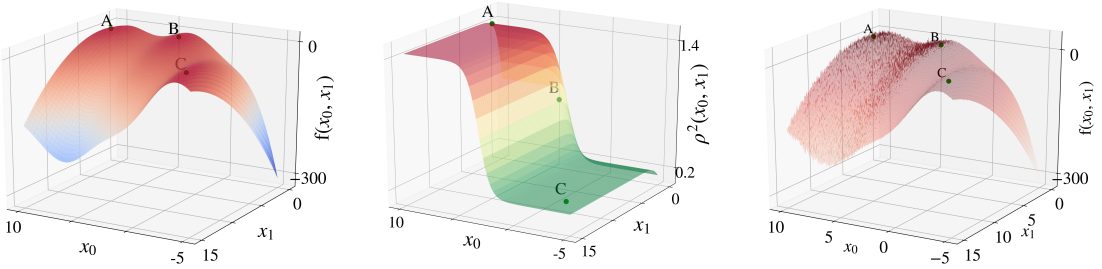


Figure 6: Additional examples for Section 4 (each row corresponds to one initialization). GP models fitted for GP-UCB (left column) and RAHBO (right column) for sine function. After initialization with the same sampled points, GP-UCB concentrates on the high-noise region whereas RAHBO prefers small variance.

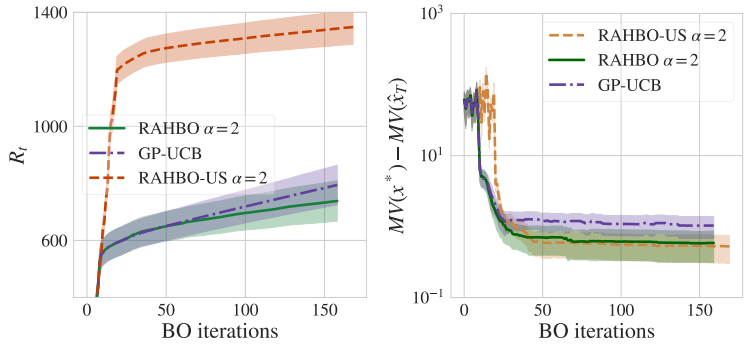
A.6.2 Branin

We provide additional visualizations, experimental details and results. Firstly, we plot the noise-perturbed objective function in Fig. 7 in addition to the visualization in Fig. 1c. In Fig. 8, we plot cumulative regret and simple mean-variance regrets that extends the results in Fig. 5a with RAHBO-US. The general setting is the same as described for Fig. 5a: we use 10 initial samples, repeat each evaluation $k = 10$ times, and RAHBO-US additionally uses 10 samples for learning the variance function with uncertainty sampling. During the optimization, RAHBO-US updates the GP model for variance function after every acquired point.



(a) Unknown objective (b) Unknown variance (c) Noise-perturbed evaluations

Figure 7: Visualization of noise-perturbed function landscape:(a) Unknown objective with 3 global maxima marked as (A, B, C). (b) Heteroscedastic noise variance over the same domain: the noise level at (A, B, C) varies according to the sigmoid function. (c) Noise-perturbed evaluations: A is located in the noisiest region.



(a) (b)

Figure 8: **Branin:** (a) Cumulative regret. (b) Suboptimality w.r.t. MV

A.7 Random Forest tuning

Experiment motivation: Consider the motivating example first: the optimized RF model will be exploited under the data drift over time, e.g., detecting fraud during a week. We are interested not only in high performance *on average* but also in low variance across the results. Particularly, the first can be a realization of the decent result in the first days and unacceptable result in the last days, and the latter ensures lower dispersion over the days while keeping a reasonable mean. In this case, when training an over-parametrized model that is prone to overfitting (to the training data), e.g., Random Forest (RF) with deep trees, high variance in validation error might be observed. In contrast, a model that is less prone to overfitting can result into a similar validation error with lower variance.

RF specifications: We use scikit-learn implementation of RF. The RF search spaces for BO are listed in Table 1 and other parameters are the default provided by scikit-learn. [4] During BO, we transform the parameter space to the unit-cube space.

Dataset: We tune RF on a dataset of fraudulent credit card transactions [23] originally announced for Kaggle competition.[5] It is a highly imbalanced dataset that consists of 285k transactions and only 0.2% are fraud examples. The transactions occurred in two days and each has a time feature that contains the seconds elapsed between each transaction and the first transaction in the dataset. We use the time feature to split the data into train and validation sets such that validation transactions happen later than the training ones. The distribution of the fraud and non-fraud transactions in time is presented in Fig. 9.

In BO, we collect evaluation in the following way: we fix the training data to be the first half of the transactions, and the rest we split into 5 validation folds that are consecutive in time. The RF model is then trained on the fixed training set, and evaluated on the validations sets. We use a balanced accuracy score that takes imbalance in the data into account.

task	hyperparameter	search space
RandomForest	n_estimators	[1, 100]
	max_features	[5, 28]
	max_depth	[1, 15]

Table 1: Search space description for RF.

[4] `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html`

[5] `https://www.kaggle.com/mlg-ulb/creditcardfraud`

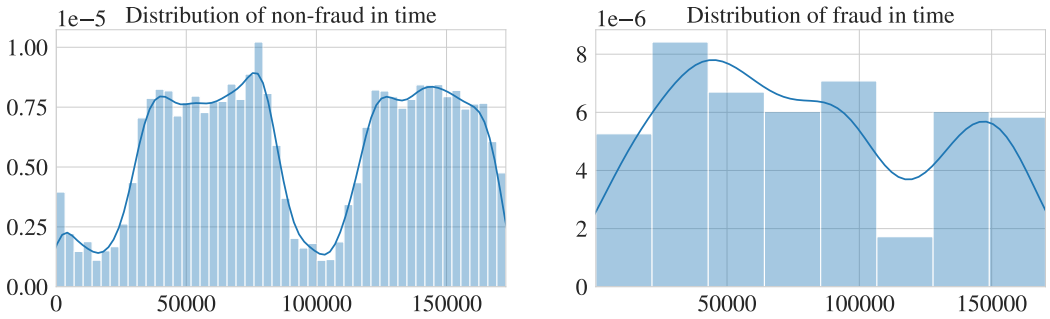
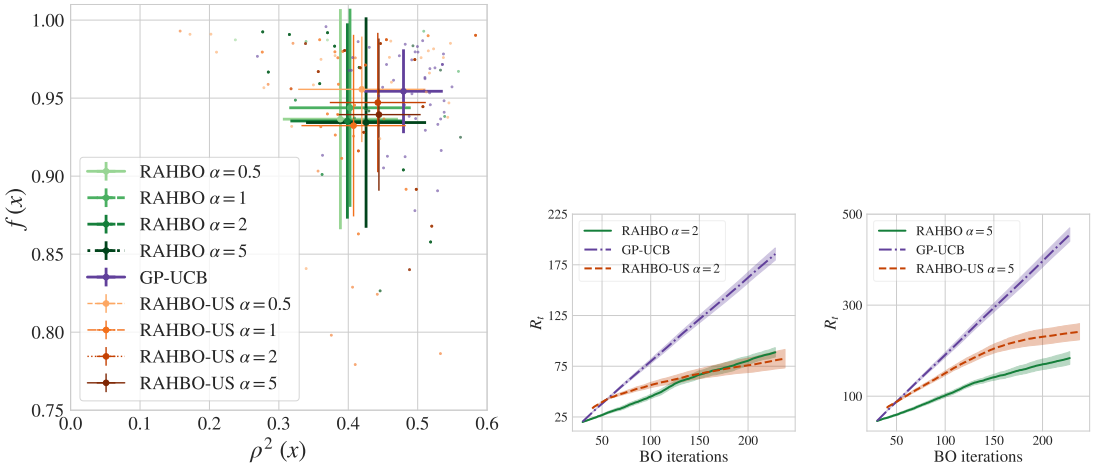


Figure 9: Distribution of non-fraud (left) and fraud (right) transactions in the dataset

A.7.1 Tuning Swiss free-electron laser (SwissFEL)



(a) Mean-variance tradeoff (SwissFEL) (b) Cum. regret ($\alpha = 0.5$) (c) Cum. regret ($\alpha = 1$)

Figure 10: (a) We plot standard deviation error bars for $f(x)$ and $\rho^2(x)$ at the *reported point* by the best observed value $x^{(T)} = \arg\max_{x_t} y_t(x_t)$ after BO completion for SwissFEL. The mean and std of the error bars are taken over the repeated BO experiments. The results demonstrate, that reporting based on the best observed value inherits high noise and as the result all methods perform similarly. Intuitively, when noise variance is high, it is possible to observe higher values. That however also inherits observing much lower value at this point, this leading to very non-robust solutions. (b-c) Cumulative regret for $\alpha = 2$ and $\alpha = 5$.