Bayesian optimization with partially specified queries

Shogo Hayashi¹ · Junya Honda^{2,3} · Hisashi Kashima²

Received: 18 May 2021 / Revised: 18 August 2021 / Accepted: 22 September 2021 /

Published online: 1 January 2022

© The Author(s) 2021

Abstract

Bayesian optimization (BO) is an approach to optimizing an expensive-to-evaluate black-box function and sequentially determines the values of input variables to evaluate the function. However, it is expensive and in some cases becomes difficult to specify values for all input variables, for example, in outsourcing scenarios where production of input queries with many input variables involves significant cost. In this paper, we propose a novel Gaussian process bandit problem, BO with partially specified queries (BOPSQ). In BOPSQ, unlike the standard BO setting, a learner specifies only the values of some input variables, and the values of the unspecified input variables are randomly determined according to a known or unknown distribution. We propose two algorithms based on posterior sampling for cases of known and unknown input distributions. We further derive their regret bounds that are sublinear for popular kernels. We demonstrate the effectiveness of the proposed algorithms using test functions and real-world datasets.

Keywords Bayesian optimization · multi-armed bandit · Thompson sampling

1 Introduction

Bayesian optimization (BO) (Shahriari et al. 2016b) is a promising approach for black-box optimization of an expensive-to-evaluate unknown function. Its applications include hyper-parameter tuning of machine learning models with the highest predictive accuracy (Snoek et al. 2012) and searching of compound structures with desirable properties (Korovina et al. 2020). BO is conducted in an iterative manner. In each iteration, a BO method determines the values of input variables in an input domain $\mathcal{X} \subset \mathbb{R}^d$, evaluates a black-box function $f : \mathcal{X} \rightarrow \mathbb{R}$ at that input query, and observes the corresponding output $y \in \mathbb{R}$. The

Editors: Yu-Feng Li, Mehmet Gönen, Kee-Eung Kim.

✉ Shogo Hayashi
shogo-hayashi@nec.com

¹ NEC Corporation, Minato-ku, Tokyo, Japan

² Kyoto University, Kyoto, Japan

³ RIKEN AIP, Chuo-ku, Tokyo, Japan

final goal is to find the input that maximizes the function with as few function evaluations as possible.

Although a learner can fully specify all values of input variables in the standard BO setting, this procedure is expensive or difficult to perform for some cases. One example is crowdsourcing, by which human-intelligent tasks are outsourced to an undetermined number of people on the Internet. Most existing crowdsourcing platforms allow us to ask workers to perform the task by specifying conditions such as age, country, and experience to obtain high-quality deliverables (Li et al. 2014). However, if the conditions are too strict, there is a risk that there will be no workers available to participate in the task. Therefore, we specify some conditions to increase the probability that one of the workers who meet the conditions will participate in the task. Upon completion of the task by a worker, we observe the quality of the deliverable and the unspecified conditions of the worker.

Another example is the design of airfoils with low self-noise in an outsourcing scenario. We ask a manufacturer to produce an airfoil by specifying values of its various variables, including angle of attack and chord length. However, specifying values of many variables in outsourcing incurs high cost, and our budget is often limited. Therefore, we are restricted to specify only values of certain variables. After obtaining the ordered airfoil, the values of the unspecified variables are revealed, and self-noise of the airfoil is measured via an acoustic test.

In this paper, we propose a novel Gaussian process (GP) bandit framework, which we call Bayesian optimization with partially specified queries (BOPSQ). In BOPSQ, unlike the standard BO setting, a learner selects a *subset* of input variables and specifies their values. We call such an input query a *partially specified query*. Next, the learner observes the values of the unspecified input variables determined according to a known or unknown distribution. Then, the learner evaluates the black-box function at the input values to obtain the corresponding output. Hence, the learner is required to consider the extent to which the input variables contribute to the function's output as well as which values the unspecified input variables will take. Williams et al. (2000) and Lattimore et al. (2016) also handled partially specified queries. However, in their works, the input variables whose values are specified are fixed or infinite input spaces are not considered for the GP bandit.

We propose two algorithms for known and unknown input distributions. In the case of the known input distribution, the uncertainty lies only in the function estimation, and we can naturally extend a posterior sampling approach to this setting. In the case of the unknown input distribution, the learner further needs to take the uncertainty in the distribution estimation into account. We adopt a multi-armed bandit approach to appropriately explore the input distribution.

The contributions of the present study are summarized as follows:

1. We propose a novel GP bandit framework with partially specified queries, which is a natural generalization of the standard BO problem and BO with environmental variables (Williams et al. 2000).
2. We propose two algorithms for known and unknown input distribution cases with their sublinear regret bounds for popular kernels.
3. We demonstrate the empirical effectiveness of the proposed algorithms using a set of test functions and real-world datasets.

The rest of the paper is organized as follows. We briefly describe the standard BO setting and approaches in Sect. 2. Section 3 formalizes the problem setting of BOPSQ as an extension of

the standard BO setting. In Sect. 4, we propose the two algorithms for known and unknown input distribution cases with theoretical guarantees. We validate the effectiveness of the proposed algorithms in Sect. 5. Section 6 summarizes the related works. We provide the concluding remarks in Sect. 7.

2 Review of Bayesian optimization

BO (Srinivas et al. 2012; Shahriari et al. 2016b) is an approach for finding the maximizer of an expensive-to-evaluate black-box function interactively. Let $\mathcal{X} \subset \mathbb{R}^d$ be a d -dimensional compact set and $f : \mathcal{X} \rightarrow \mathbb{R}$ be an unknown real-valued function. In each iteration, a learner determines the values of input variables $x \in \mathcal{X}$. The learner evaluates the function at x and obtains an output $y = f(x) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The goal is to find the maximizer

$$x^* = \operatorname{argmax}_{x \in \mathcal{X}} f(x)$$

with as few evaluations as possible.

BO methods typically assume a GP (Rasmussen and Williams 2006) as a prior of the black-box function: $f \sim \mathcal{GP}(\mu_0, k)$, where $\mu_0 : \mathcal{X} \rightarrow \mathbb{R}$ is a mean function and $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a covariance function. Without loss of generality, we assume a zero mean $\mu_0 = \mathbf{0}$. Suppose that a set of t input-output observations $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^t$ is given. Then, we have a posterior mean $\mu_t(x)$, posterior variance $\sigma_t^2(x)$, and posterior covariance $k_t(x, y)$ as follows:

$$\begin{aligned}\mu_t(x) &= k_x^\top (K + \sigma^2 I)^{-1} \mathbf{y}_t, & \sigma_t^2(x) &= k_t(x, x), \\ k_t(x, y) &= k(x, y) - k_x^\top (K + \sigma^2 I)^{-1} k_y,\end{aligned}$$

where k_x is a t -dimensional vector whose i -th entry is $k(x_i, x)$, K is a $t \times t$ matrix whose (i, j) -th entry is $k(x_i, x_j)$, I is an identity matrix, and $\mathbf{y}_t$ is a t -dimensional vector whose i -th entry is y_i .

BO methods often use an acquisition function $a_{t+1} : \mathcal{X} \rightarrow \mathbb{R}$ that quantifies how much an input should be evaluated next. The next input query is determined as

$$x_{t+1} = \operatorname{argmax}_{x \in \mathcal{X}} a_{t+1}(x).$$

Popular choices for the acquisition function are the Gaussian process upper confidence bound (GP-UCB) (Srinivas et al. 2012), the probability of improvement (Kushner 1964), and the expected improvement (Mockus et al. 1978; Jones et al. 1998). Thompson sampling (Thompson 1933; Russo and Roy 2014) based on a posterior sample $g_{t+1} \sim \mathcal{GP}(\mu_t, k_t)$ is also widely used, and its acquisition function is defined as

$$a_{t+1}(x) = g_{t+1}(x).$$

3 Problem setting

We propose a novel GP bandit problem that utilizes partially specified queries, BOPSQ. We first define some terms and notations. Then, we introduce the framework and define regret for BOPSQ.

3.1 Notation

Let $\mathcal{X} = \mathcal{X}^{(1)} \times \dots \times \mathcal{X}^{(d)} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a compact domain and $f : \mathcal{X} \rightarrow \mathbb{R}$ be an unknown function of interest. A *control variable set*, $I \subset [d]$, represents the indices of input variables which a learner can specify the values of. An *uncontrol variable set*, $\bar{I} = [d] \setminus I$, represents the remaining indices of input variables which the learner can not specify the values of. Let $\mathcal{I} \subset 2^{[d]}$ be a family of control variable sets. We call the values of control (uncontrol) variable set *control (uncontrol) variables*, denoted by $x^I \in \mathcal{X}^I$ ($x^{\bar{I}} \in \mathcal{X}^{\bar{I}}$). Here, $x^I = \{x^{(i)} \in \mathcal{X}^{(i)} \mid i \in I\}$ and $\mathcal{X}^I = \mathcal{X}^{i_1} \times \dots \times \mathcal{X}^{i_k}$ for $I = \{i_1, \dots, i_k\} \subset [d]$. Let $X = (X^{(1)}, \dots, X^{(d)})^\top \sim P(X)$ be a random vector according to an input distribution P over $\mathcal{X}$ and Y be a random variable over $\mathbb{R}$ corresponding to the output variable. With a slight abuse of notation, we write $f(x^I, x^{\bar{I}})$ for $f(x)$.

3.2 Framework

The optimization procedure consists of T iterations. A learner is given a non-empty family of control variable sets $\mathcal{I} \subset 2^{[d]}$, which we assume is static over T iterations for the sake of simplicity in this paper. In each iteration, the learner selects a control variable set $I \in \mathcal{I}$ and specifies their values $x^I \in \mathcal{X}^I$. Next, the learner observes uncontrol variables $X^{\bar{I}}$ drawn from a conditional distribution $P(X^{\bar{I}} \mid X^I = x^I)$, which is either known or unknown. By evaluating a black-box function f at the input variables $(x^I, X^{\bar{I}})$, the learner observes the corresponding output $Y = f(x^I, X^{\bar{I}}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

In the standard BO setting, the goal is to find the optimal input variables x^* . However, in BOPSQ, the learner may not always select x^* and obtain $f(x^*)$ because the input query is random. Therefore, for the objective function, we consider the conditional expectation of $f(x^I, X^{\bar{I}})$ over $X^{\bar{I}}$ given $X^I = x^I$. We define the goal as finding the optimal control variable set $I^* \in \mathcal{I}$ and control variables $x^{I^*} \in \mathcal{X}^{I^*}$ in as few evaluations as possible, defined as

$$(I^*, x^{I^*}) = \operatorname{argmax}_{I \in \mathcal{I}, x^I \in \mathcal{X}^I} \mathbb{E}[f(X^I, X^{\bar{I}}) \mid f, X^I = x^I]. \quad (1)$$

Hereafter, we write x^{I^*} instead of (I^*, x^{I^*}) for simplicity.

BOPSQ in the unknown input distribution case can be seen as a mixture of BO and multi-armed bandits. By selecting an action x^I from an infinite action space $\mathcal{X}^I$, a learner observes information $f(x^I, X^{\bar{I}})$, which corresponds to the standard BO. By selecting an action I (and x^I) from a finite action space $\mathcal{I}$, a learner observes information $X^{\bar{I}} \sim P(X^{\bar{I}} \mid X^I = x^I)$, which corresponds to a multi-armed bandit problem.

Some specific forms set as the family of control variable sets $\mathcal{I}$ recover existing BO settings. The simple setting in which the learner can select all the input variables as a control variable set, $\mathcal{I} = \{[d]\}$, corresponds to the standard BO setting. When the family of control variable sets is always fixed to one particular subset, $\mathcal{I} = \{I\}$, $I \subset [d]$, $0 < |I| < d$, we consider this setting as BO with environmental variables (Williams et al. 2000). In this sense, BOPSQ is a generalization of these BO settings. The major difference appears when the learner is required to select a control variable set for $|\mathcal{I}| > 1$. Moreover, when the input distribution $P(X)$ is unknown, the learner is required to consider the uncertainty of the input distribution estimation as well as the function estimation.

In practice, the sizes of all control variable sets may be same, i.e., $\mathcal{I} = \{I \mid |I| = k\}$ for $0 < k < d$. This is because control variable sets that are a subset of other control variable sets are useless. Let $I \subset J \in \mathcal{I}$, that is, I be a subset of J . It is clear that

the maximum of the objective function for I in Eq. (1) is not larger than that for J , i.e., $\max_{x^I \in \mathcal{X}^I} \mathbb{E}[f(X^I, X^J) \mid f, X^I = x^I] \leq \max_{x^J \in \mathcal{X}^J} \mathbb{E}[f(X^I, X^J) \mid f, X^J = x^J]$. Further, we typically want to specify the values of as many input variables as possible in black-box optimization. Therefore, in the experimental part, we set the sizes of all control variable sets to be same.

3.3 Regret

The final goal is to find the optimal control variables x^{I^*} defined in Eq. (1). We define instantaneous regret $\text{IR}(t)$ in iteration t , which evaluates the gap between the optimal solution and the choice x^{I_t} , described as

$$\text{IR}(t) = \mathbb{E}[f(X^{I^*}, X^{\bar{I}^*}) \mid f, X^{I^*} = x^{I^*}] - \mathbb{E}[f(X^{I_t}, X^{\bar{I}_t}) \mid f, X^{I_t} = x^{I_t}].$$

The resultant performance measure, Bayes cumulative regret $\text{BayesRegret}(T)$, is the expectation and summation over T iterations of the instantaneous regret, defined as

$$\text{BayesRegret}(T) = \sum_{t=1}^T \mathbb{E}[\text{IR}(t)].$$

Our target is designing algorithms that achieve sublinear regret, that is, $\lim_{T \rightarrow \infty} \text{BayesRegret}(T)/T \rightarrow 0$. In some situations, one may prefer simple regret, the best instantaneous regret so far $\min_{t=1, \dots, T} \mathbb{E}[\text{IR}(t)]$, to cumulative regret. However, sublinear cumulative regret guarantees that simple regret asymptotically converges to 0 since $\min_{t=1, \dots, T} \mathbb{E}[\text{IR}(t)] \leq \text{BayesRegret}(T)/T$.

4 Algorithms

We consider two cases where the input distribution is known and unknown, propose two algorithms based on posterior sampling for these cases, and also provide their regret bounds that are sublinear for popular kernels.

4.1 TSPSQ-KNOWN for known input distribution

We first focus on the case where the joint input distribution $P(X)$ or the set of conditional distributions $\{P(X^{\bar{I}} \mid X^I)\}_{I \in \mathcal{I}}$ is known. For example, the case of the the airfoil design provided by a manufacturer corresponds to the situation in which the manufacturer publishes random options for the unspecified features. We propose an algorithm based on Thompson sampling, which we call Thompson sampling with partially specified queries for the known input distribution case (TSPSQ-KNOWN).

4.1.1 Acquisition function

In iteration t , TSPSQ-KNOWN determines the next control variables as

$$x^{I_t} = \operatorname{argmax}_{I \in \mathcal{I}, x^I \in \mathcal{X}^I} \mathbb{E}[g_t(X^I, X^{\bar{I}}) \mid g_t, X^I = x^I], \quad (2)$$

where g_t is a sample from the GP posterior $g_t \sim \mathcal{GP}(\mu_{t-1}, k_{t-1})$. The pseudo-code is given in Algorithm 1.

Equation (2) approaches the ground-truth objective in Eq. (1) as the posterior sample g_t approximates the ground-truth function f better with more observations. TSPSQ-KNOWN is a natural extension of Thompson sampling; if we set $\mathcal{I} = \{[d]\}$ as the standard BO setting, the acquisition function in Eq. (2) is reduced to $g_t(x)$.

4.1.2 Regret bound

We first introduce the notion of maximum information gain, which is the basis of typical regret analysis for BO. The maximum information gain in iteration T , denoted by γ_T , is the maximum possible information gain achievable by any algorithm for f via queries $A = \{x_1, x_2, \dots, x_T\} \subset \mathcal{X}$ and the corresponding outputs $y_A = \{y_1, y_2, \dots, y_T\}$, defined as

$$\gamma_T = \max_{A \subset \mathcal{X}: |A|=T} I(f; y_A). \quad (3)$$

Here, $I(f; y_A)$ denotes the mutual information between f and y_A . Srinivas et al. (2012) provided the bounds of γ_T for various types of kernels: $\gamma_T = O(d \log T)$ for the linear kernel $k(x, y) = x^\top y$, $\gamma_T = O((\log T)^{d+1})$ for the Gaussian kernel $k(x, y) = \exp(0.5\|x - y\|^2/l^2)$, and $\gamma_T = O(T^{d(d+1)/(2\nu+d(d+1))} \log T)$ for the Matérn kernel $k(x, y) = (2^{1-\nu}/\Gamma(\nu))r^\nu B_\nu(r)$, where $r = (\sqrt{2\nu/l}\|x - y\|)$, B_ν is the modified Bessel function, and $l, \nu > 0$.

We then derive the regret bound of TSPSQ-KNOWN for the case where the ground-truth input distribution is known.

Theorem 1 Let $\mathcal{X} \subset [0, 1]^d$, $d \in \mathbb{N}$, be a compact domain. Assume $k(x, x') \leq 1$ and that there exist constants $a', b' > 0$ for the partial derivatives of sample paths f such that

$$l > 0, \forall j \in [d], \mathbb{P}\left(\sup_{x \in \mathcal{X}} \left| \frac{\partial f(x)}{\partial x^{(j)}} \right| > l\right) \leq a' e^{-(l/b')^2}. \quad (4)$$

Then, by running TSPSQ-KNOWN for T iterations, it holds that

$$\text{BayesRegret}(T) = O(\sqrt{dT\gamma_T \log T}).$$

The analysis is mainly based on the techniques by Russo and Roy (2014), Srinivas et al. (2012), where the cumulative regret is decomposed into terms with respect to discretization errors of $\mathcal{X}$, the difference between f and the upper confidence bound $\mu_{t-1}(x) + \beta_t^{1/2} \sigma_{t-1}(x)$, and the difference between x^{I^*} and x^I . A detailed proof of Theorem 1 can be found in Appendix A.

Note that the growth rate of the regret bound in T matches the existing results of Thompson sampling and GP-UCB in the standard BO setting shown by Russo and Roy (2014) and Srinivas et al. (2012), respectively, though their definitions of regret are different. This is because the ground-truth input distribution is given, and the uncertainty in estimation is therefore only in f , as in the standard BO setting. The regret bound is sublinear for the popular kernels, such as the linear kernel, Gaussian kernel, and Matérn kernel.

Algorithm 1 TSPSQ-KNOWN (TSPSQ-UNKNOWN)

Input: A GP prior $\mathcal{GP}(\mu_0, k_0)$, a query budget T , and a family of control variable sets $\mathcal{I}$ (and a set of conditional distributions $\{P(X^{\bar{I}} | X^I)\}_{I \in \mathcal{I}}$ for TSPSQ-KNOWN)
Initialize $\mathcal{D}_0 \leftarrow \emptyset$
for $t = 1, 2, \dots, T$ **do**
 Sample a posterior path $g_t \sim \mathcal{GP}(\mu_{t-1}, k_{t-1})$
 Determine the next control variables x^{I_t} using Equation (2) for TSPSQ-KNOWN (or Equation (6) for TSPSQ-UNKNOWN)
 Observe uncontrol variables $X^{\bar{I}_t}$
 Evaluate f at $(x^{I_t}, X^{\bar{I}_t})$, and observe the corresponding output y_t
 Update $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{(x^{I_t}, X^{\bar{I}_t}), y_t\}$ and $\mathcal{GP}$ (and $\mathcal{S}_t^{(i)} \leftarrow \mathcal{S}_{t-1}^{(i)} \cup \{X^{(i)}\}$ for $i \in \bar{I}_t$ for TSPSQ-UNKNOWN)
end for

4.2 TSPSQ-UNKNOWN for unknown input distribution

Next, we consider the case where the joint input distribution $P(X)$, or the set of conditional distributions $\{P(X^{\bar{I}} | X^I)\}_{I \in \mathcal{I}}$, is unknown. We propose Thompson sampling with partially specified queries for the unknown input distribution (TSPSQ-UNKNOWN) based on a multi-armed bandit approach.

In the unknown input distribution case, a set of conditional distributions of uncontrol variables needs to be estimated. For example, when the size of uncontrol variables is $d/2$, there could be $|\mathcal{I}| = \binom{d}{d/2} = \Omega(2^{d/2})$ combinations of uncontrol variables. However, samples from a conditional distribution $P(X^{\bar{I}} | X^I = x^I)$ may be uninformative for estimating other conditional distributions $P(X^{\bar{J}} | X^J)$ for $I \neq J$ without any assumption (a further discussion can be found in Appendix C). Then, the estimation of different conditional distributions may be separately addressed, which requires an exponential number of samples with respect to d .

Therefore, to make the estimation statistically tractable, we assume only that input variables are independent of each other. That is, the cumulative distribution function $F(X \leq x)$ is decomposed as $F(X \leq x) = \prod_{i=1}^d F^{(i)}(X^{(i)} \leq x^{(i)})$. Then, the number of distributions to be estimated does not increase exponentially with d . Note that we do not make further assumptions such as continuity or a certain form of the distribution.

4.2.1 Acquisition function

Let $\bar{g}_t$ be a function defined as

$$\bar{g}_t(x^I, x^{\bar{I}}) = g_t(x^I, x^{\bar{I}}) + \sum_{i \in \bar{I}} \frac{\alpha_t}{\sqrt{|\mathcal{S}_{t-1}^{(i)}|}}, \quad (5)$$

where $g_t \sim \mathcal{GP}(\mu_{t-1}, k_{t-1})$, $\mathcal{S}_{t-1}^{(i)}$ is a set of observations of the uncontrol variables until iteration $t - 1$ for $i \in [d]$, defined as $\mathcal{S}_{t-1}^{(i)} = \{X_s^{(i)} | i \in \bar{I}_s, s \in [t - 1]\}$, and α_t is a monotonically increasing function in t . TSPSQ-UNKNOWN determines the next control variables as

$$x^{I_t} = \operatorname{argmax}_{x^I \in \mathcal{X}^I, I \in \mathcal{I}} \mathbb{E}_{X^{\bar{I}} \sim \hat{F}^{\bar{I}}} [\bar{g}_t(x^I, X^{\bar{I}})], \quad (6)$$

where $\hat{F}^I$ denotes the empirical distribution function for I , defined as $\hat{F}^I(X^I \leq x^I) = \prod_{i \in I} \hat{F}^{(i)}(X^{(i)} \leq x^{(i)})$, $\hat{F}^{(i)}(X^{(i)} \leq x^{(i)}) = 1/|\mathcal{S}_{i-1}^{(i)}| \sum_{\hat{x}^{(i)} \in \mathcal{S}_{i-1}^{(i)}} \mathbb{1}[\hat{x}^{(i)} \leq x^{(i)}]$. The pseudo-code is given in Algorithm 1.

The acquisition function of TSPSQ-UNKNOWN in Eq. (6) works similarly to that of TSPSQ-KNOWN in Eq. (2) for the known input distribution case. TSPSQ-UNKNOWN computes expectation using the empirical distribution. However, the empirical distribution may not be accurate when the number of samples is small. Based on the observation discussed in Sect. 3.2, we employ a UCB approach for multi-armed bandits (Auer et al. 2002). The second term on the right hand side of Eq. (5) encourages increasing the number of samples from the input distributions with few samples. A hyperparameter α_i controls the exploration-exploitation trade-off in the distribution estimation: large α_i improves the distribution estimation, while small α_i attempts to achieve small regret.

TSPSQ-UNKNOWN avoids a regret bound exponentially dependent on the number of input variables d for the independence assumption, as shown later, because it does not need to consider the combinations of d input variables in the distribution estimation. However, the computational cost of TSPSQ-UNKNOWN in each iteration still exponentially depends on d due to the combinations of $\mathcal{S}^{(i)}$ for $i \in \bar{I}$. As a simple case, when $|\bar{I}| = d/2$ and $|\mathcal{S}^{(i)}| = n$ for $i \in \bar{I}$, the computation of the empirical expectation in Eq. (6) requires $n^{d/2}$ evaluations of $\bar{g}$ for each $x^I \in \mathcal{X}^I$ and $I \in \mathcal{I}$. This computational cost could be soon intractable as n increases through iterations for large $\mathcal{I}$ and d . In general, an exact computation with small cost may not be possible; however, the cost can be reduced when some assumptions hold. For example, Kandasamy et al. (2015) and Mutny and Krause (2018) assumed the additivity of f with respect to each $x^{(i)}$ for $i \in [d]$, i.e., $f(x) = \sum_{i \in [d]} f_i(x^{(i)})$. If that assumption holds, the computational cost is reduced to $O(dn)$.

4.2.2 Regret bound

We then derive a regret bound for the unknown input distribution case.

Theorem 2 *Let $\mathcal{X} \subset [0, 1]^d$, $d \in \mathbb{N}$, be a compact domain. Assume $k(x, x') \leq 1$ and that there exist constants $a, b, a', b' > 0$ for the sample paths f and their partial derivatives such that Assumption (4) holds and*

$$l > 0, \mathbb{P}\left(\sup_{x \in \mathcal{X}} |f(x)| > l\right) \leq ae^{-(l/b)^2}. \quad (7)$$

Suppose an unknown cumulative distribution function of the input variables is factorized into $F(X \leq x) = \prod_{i=1}^d F^{(i)}(X^{(i)} \leq x^{(i)})$. Define $\alpha_i = 2b' \log t$. Then, by running TSPSQ-UNKNOWN for T iterations, it holds that

$$\text{BayesRegret}(T) = O\left(\sqrt{dT \log T} \left(\sqrt{\gamma_T} + \sqrt{m \log T}\right)\right),$$

where γ_T is the maximum information gain defined in Eq. (3) and $m = \max_{I \in \mathcal{I}} |\bar{I}|$ is the maximum number of dimensions of uncontrol variables.

The main difference of the proof from that for Theorem 1 is the evaluation of the distribution estimation. We employ techniques in multi-armed bandit problems and

the Dvoretzky–Kiefer–Wolfowitz inequality (Dvoretzky et al. 1956; Massart 1990) to derive the upper bound. The whole proof can be found in Appendix B.

The growth rate of the regret bound in T in Theorem 2 matches the existing results for the standard BO setting (Russo and Roy 2014; Srinivas et al. 2012), as with TSPSQ-KNOWN. This is because, in most cases, the first term corresponding to the function estimation is dominant compared to the second term corresponding to the input distribution estimation thanks to the independence assumption. Recall that $m \leq d$. Since d is typically moderate in BO, e.g., $d < 10$, $\sqrt{m}$ is not large. Therefore, for the Gaussian kernel for example, where $\gamma_T = O((\log T)^{(d+1)})$, the first term $\sqrt{dT(\log T)^{(d+2)}}$ can be major compared to the second term $\sqrt{mdT \log^2 T}$ in the total regret bound. This argument also can be applied to the Matérn kernel. However, for the linear kernel, where $\gamma_T = O(d \log T)$, the first and second terms are equal in growth rate for $m = O(d)$.

The bound linearly depends on d for $m = O(d)$, and no exponential term appears in d . As m decreases, the bound also decreases. This is because the learner can specify the values of the many input variables as desired and the error with respect to the distribution estimation becomes small.

5 Experiments

We demonstrate the effectiveness of the proposed algorithms for the known and unknown input distributions using a set of test functions and real-world datasets.

5.1 Comparing methods

We compare the proposed algorithms to four baselines: DROPOUTBO, WRAPPERBOs, RANDOMBO, and RANDOM. DROPOUTBO (Li et al. 2017), which was originally proposed for the high-dimensional BO, sequentially determines a partially specified query by performing optimization over a low-dimensional input domain. In each iteration, DROPOUTBO randomly selects $I_t \in \mathcal{I}$ and then performs the standard BO over $\mathcal{X}^I$ to determine the next control variables x^{I_t} . WRAPPERBOs performs the standard BO over $\mathcal{X}^I$ separately for $I \in \mathcal{I}$ and selects x^{I_t} based on a wrapper method (Guyon and Elisseeff 2003) as feature selection. For each $I \in \mathcal{I}$, it applies GP regression to $\{(x_i^I, y_i)\}_{i=1}^t$, observations of input variables of I , where $x_i = (x^{I_t}, X^{I_t})$. Then, it selects x^{I_t} using $|\mathcal{I}|$ acquisition functions $\{a^I\}_{I \in \mathcal{I}}$ as $x^{I_t} = \arg\max_{I \in \mathcal{I}, x^I \in \mathcal{X}^I} a^I(x^I)$. RANDOMBO directly uses the standard BO method with a randomly determined control variable set. First, it determines the next d -dimensional input as in the standard BO, $x_t = \arg\max_{x \in \mathcal{X}} a_t(x)$. Then, I_t is randomly chosen from a uniform distribution over $\mathcal{I}$. Subsequently, it queries for $x_{t_i}^{I_t}$. RANDOM simply determines the next control variables according to a uniform distribution.

We use the Gaussian kernel function for GP models and employ Thompson sampling as an acquisition function for all methods. In the preliminary experiment in Sect. 5.3, we set the hyperparameters of the kernel function to those obtained by the type II maximum likelihood estimation using thousands of data points. In the other experiments in Sects. 5.4 and 5.5, the hyperparameters are learned every 10 iterations by using the type II likelihood estimation.

Fig. 1 Branin-Hoo function (left) and its input density function (right)

5.2 Approximation of sample paths

All methods except RANDOM use a sample path $g_t \sim \mathcal{GP}(\mu_{t-1}, k_{t-1})$. By approximating the GP with a finite-dimensional Bayesian linear model, we can efficiently optimize g_t (Hernández-Lobato et al. 2014). Bochner's theorem (Bochner 1959) guarantees that for any continuous shift-invariant kernel $k(x - y)$, its Fourier transform is a non-negative measure $\hat{k}(\omega)/\alpha$, where $\alpha = \int_{\omega} \hat{k}(\omega) d\omega$ is a normalization constant. Random Fourier features (Rahimi and Recht 2007) approximate the kernel function as

$$k(x - y) = \alpha \mathbb{E}[e^{-i\omega^\top(x-y)}] \approx \Phi(x)^\top \Phi(y),$$

where $\Phi(x) = \sqrt{2\alpha/M} (\sin(\omega_1^\top x), \cos(\omega_1^\top x), \dots, \sin(\omega_{M/2}^\top x), \cos(\omega_{M/2}^\top x))^\top$ and $\{\omega_i\}_{i=1}^{M/2} \stackrel{\text{i.i.d.}}{\sim} \hat{k}(\omega)$. Then, the GP is approximated with an M -dimensional Bayesian linear model. The sample path is approximated as

$$g_t(x) \approx \Phi(x)^\top \theta_t, \quad \theta_t \sim \mathcal{N}(\tilde{\mu}_t, \tilde{\Sigma}_t),$$

where $\tilde{\mu}_t = A^{-1} \Phi(X_{t-1})^\top Y_{t-1}$, $\tilde{\Sigma}_t = \sigma^2 A^{-1}$, $A = \Phi(X_{t-1})^\top \Phi(X_{t-1}) + \sigma^2 I$, $\Phi(X_{t-1}) = (\Phi(x_1), \dots, \Phi(x_{t-1}))^\top \in \mathbb{R}^{(t-1) \times M}$, and $Y_{t-1} = (y_1, \dots, y_{t-1})^\top$. The random Fourier features enable us to deal with a large number of samples. In the experiments, we set $M = 1000$.

5.3 Preliminary experiments

Here we study the effectiveness of our algorithms for known and unknown input distributions under simple settings. We use the Branin-Hoo function (Picheny et al. 2013) (Fig. 1(left)) over a two-dimensional input domain as a black-box target function without observation noise. We model a joint distribution of input variables with a Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ truncated over $[0, 1]^2$, where we set $\mu = (0.5, 0.5)^\top$ and $\Sigma = ((0.01, 0)^\top, (0, 0.05)^\top)$ (Fig. 1 (right)). We set a family of control variable sets as $\mathcal{I} = \{\{1\}, \{2\}\}$. The Branin-Hoo function has the three minimum points (dots in Fig. 1 (left)); however, the optimal control variable differs from them (line in Fig. 1 (left)). In Theorem 2 for TSPSQ-UNKNOWN, $b' > 0$ to define $\alpha_t = 2b' \log t$ may not be known in

Fig. 2 Cumulative regret for the Branin-Hoo function with the known input distribution

Fig. 3 Cumulative regret at 100th iteration (left) and the cumulative regret over the iterations (right) of TSPSQ-UNKNOWN with different hyperparameter values c for the Branin-Hoo function with the unknown input distribution

advance. Hence, we change the value of a hyperparameter c for $\alpha_t = c \log t$ and study the difference in performance. We set $T = 100$ and repeat each experiment 30 times.

The results for the known input distribution case are shown in Fig. 2, where the lines and shaded areas are the means and standard deviations for 30 trials, respectively. TSPSQ-KNOWN achieved small regret in the early iterations and outperformed the baselines by finding the nearly optimal solution. Figure 3 (left) shows the cumulative regret at the 100th iteration of TSPSQ-UNKNOWN, with different values of the hyperparameter c for the unknown input distribution case. We observed that the smaller c is better than the larger c in the mean for 30 trials. In particular, TSPSQ-UNKNOWN with $c = 0.12$ worked the best and found the nearly optimal solution, as shown in Fig. 3 (right), by controlling the balance between exploration and exploitation in the distribution estimation. The result of TSPSQ-UNKNOWN with $c = 0.12$ is competitive with that of TSPSQ-KNOWN for the known input distribution. The larger c values suffered from the higher regret because they heavily focused on exploration and obtained low rewards; however, the result is still better than the baselines. The variance of small c was large because it heavily focused on exploitation and sometimes fell into the local optimum.

Fig. 4 Cumulative regret for the cosine mixture function

Fig. 5 Cumulative regret for the Rosenbrock function

5.4 Experiments using test functions

Next, we conduct numerical simulations using two test functions: a cosine mixture function (Anderson et al. 2000) over a two-dimensional domain and the Rosenbrock function (Picheny et al. 2013) over a four-dimensional domain. We model a joint distribution of input variables with a Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, where we set $\mu = (0.7, 0.7)^\top$, $\Sigma = ((0.01, 0)^\top, (0, 0.05)^\top)$ for the cosine mixture function and $\mu_i = 0.7$, $\Sigma_{i,i} = 0.01$, $\Sigma_{i,j} = 0$, for $i, j \in [4]$, $i \neq j$ for the Rosenbrock function. We set a family of control variable sets $\mathcal{I}$ as a set of combinations of 1 of 2 input variables for the cosine mixture function and a set of combinations of 2 of 4 input variables for the Rosenbrock function. We set the observation noise variance $\sigma^2 = 10^{-3}$ for both functions, following Hernández-Lobato et al. (2014). According to the results in Sect. 5.3, we set $\alpha_t = 0.12 \log t$ for TSPSQ-UNKNOWN. We set $T = 100$ and repeat each experiment iterations 30 times.

Figures 4 and 5 show the cumulative regret for the known and unknown input distribution cases for the cosine mixture function and the Rosenbrock function, respectively. Note that the performance of the baselines does not change for both cases. The proposed algorithms, TSPSQ-KNOWN and TSPSQ-UNKNOWN, outperformed the baselines and achieved small regret. They finally yielded the nearly optimal solutions, while the baselines increased their cumulative regrets linearly. TSPSQ-UNKNOWN is again competitive with TSPSQ-KNOWN although it is not given the ground-truth input distribution. This result might be because the error in the distribution estimation is much smaller than the error in the function estimation, as discussed in Sect. 4.2.2.

Fig. 6 Cumulative regret for the airfoil self-noise function

5.5 Experiments using real-world datasets

Finally, we conduct experiments using two real-world datasets: the airfoil self-noise dataset¹ and the word similarity dataset (Snow et al. 2008). The airfoil self-noise dataset is a set of airfoil self-noise as an output variable and five input variables including angle of attack and chord length. The goal is to design airfoils with low self-noise via the five input variables in an outsourcing scenario as described in Sect. 1. The word similarity dataset is a set of answers given by workers for word pair similarity tasks in crowdsourcing. In a task, a word pair is given to a worker, and the worker assesses their similarity with an integer from 0 to 10, where a large number indicates stronger similarity. We randomly select 10 of 30 answers as the input variables. For the output variable, we calculate the negative mean of the absolute difference between the remaining 20 answers and the true answers so that the good workers have the high output value. The goal is to find good workers whose answers are accurate via the 10 input variables as described in Sect. 1. We set a family of control variable sets $\mathcal{I}$ as a set of combinations of 2 of 5 input variables for the airfoil self-noise dataset and a set of combinations of 5 of 10 input variables for the word similarity dataset. To reduce the computational cost, we randomly reduce the size of the family to 10 for the word similarity dataset. The discussion on large $|\mathcal{I}|$ can be found in Sect. 4.2.1. For these datasets, the ground-truth functions and the ground-truth input distributions are unknown and only available through noisy observations. Output data corresponding to the input points we query for is not always available. Therefore, we substitute a posterior mean function of a GP obtained by learning each dataset as a ground-truth function². The observation noise variance is also set to the learned parameter. For a proxy for the ground-truth input distributions, we apply kernel density estimation with the Gaussian kernel to the datasets, whose bandwidth is obtained using the median heuristic (Garreau et al. 2017). We set $\alpha_t = 0.12 \log t$ for TSPSQ-UNKNOWN based on the results in Sect. 5.3, $T = 100$, and repeat each experiment 30 times.

Figures 6 and 7 illustrate the cumulative regret for the known and unknown input distribution cases for the airfoil self-noise function and the word similarity function, respectively. Both TSPSQ-KNOWN and TSPSQ-UNKNOWN performed better than the baselines. In the early iterations, there was no significant difference in regret for all methods, but the

¹ <https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>

² This setting is common in real-world data experiments in the BO literature (Hernández-Lobato et al. 2014).

Fig. 7 Cumulative regret for the word similarity function

proposed algorithms were able to keep their regrets small after 20 iterations, as shown in Fig. 6, and after 30 iterations, as shown in Fig. 7. TSPSQ-UNKNOWN is competitive with TSPSQ-KNOWN in these experiments, possibly due to the more difficult function estimation than the distribution estimation, as discussed in Sect. 4.2.2.

6 Related work

There is a significant body of literature on BO, where a learner has control of only some of the input values but not the rest. Prior works are classified into two settings based on whether the remaining input values are revealed to the learner to determine the controllable input values.

First, in contextual GP bandits (Krause and Ong 2011), a black-box function takes control variables and context information as input. In each iteration, the learner receives context information and then determines the values of control variables. In BOPSQ, a learner has no access to the context information.

Second, several works tackled situations where a learner determines the controllable input values without knowing the remaining input values, as well as BOPSQ. In BO with environmental variables (Williams et al. 2000), the objective is to maximize the expected cumulative reward on the remaining input values determined according to a distribution. This setting can be seen as a special case of BOPSQ, as shown in Sect. 3. In a game theory setting (Sessa et al. 2019; Dai et al. 2020), each agent determines its action sequence to maximize each cumulative reward against the other agents. Bogunovic et al. (2018) dealt with an adversarial setting, where an adversary determines the remaining input values to minimize the learner's reward. Kirschner et al. (2020) and Nguyen et al. (2020) extended this adversarial setting into a setting where an adversary selects a distribution of the remaining input values. These works are closely related to BOPSQ; however, a learner can not select input variables to determine their values. BOPSQ involves the feature selection procedure.

Causal bandit (Lattimore et al. 2016) for causal intervention is also closely related to BOPSQ, which involves a feature selection procedure. In the problem setting, a causal model is provided to a learner as a directed acyclic graph over random variables of input and output variables. In each iteration, the learner specifies values of any

subset of input variables and then observes the remaining subset of input variables and the output variable. The objective is to maximize the expectation of the output variable conditioned on specified values. The primary difference is that BOPSQ deals with infinite input spaces as a GP bandit, while causal bandit deals with finite input spaces as a multi-armed bandit.

Noisy inputs have been tackled by several works (Oliveira et al. 2019; Luong et al. 2020). In BO with noisy inputs (Oliveira et al. 2019), the input values are perturbed, and a black-box function takes the unobservable perturbed values as input. In BO with missing inputs (Luong et al. 2020), some of the input values are replaced with unobservable random values. The input values in BOPSQ are also random; however, all input variables are controllable for a learner in those works, although input values are noisy. Further, feature selection is not addressed in those works.

The feature selection problem (Guyon and Elisseeff 2003) is related to BOPSQ in a wide sense, which is the task of selecting a subset of features that are most relevant to an output. Chen et al. (2012) proposed a method that jointly optimizes the objective and selects features for high-dimensional BO where it is assumed that only a subset of input variables are associated with a black-box function. Li et al. (2017) also dealt with high-dimensional BO by performing optimization only with respect to randomly chosen features. Random values or the values of the input variables with the best function value so far are used as the values of the unchosen features. However, these works operate in the standard BO setting: a learner determines all values of input variables. Further, the methods do not deal with the random input variables.

7 Conclusion

We proposed BOPSQ, a novel problem setting that utilizes partially specified queries for the situations in which the specification of all input variables is expensive or difficult. BOPSQ involves a features selection procedure, and values of unspecified input variables are randomly determined. In particular, BOPSQ with unknown input distribution corresponds to a mixture of BO and multi-armed bandits. We proposed the algorithms based on posterior sampling, TSPSQ-KNOWN and TSPSQ-UNKNOWN for the known and unknown input distribution cases. We derived their regret upper bounds that are sublinear for the popular kernels. In the experiments, we demonstrated the effectiveness of the proposed algorithms using the test functions and the real-world datasets.

In BOPSQ, uncontrol variables are sampled from a conditional distribution described by $X^I \sim P(X^I | X^I = x^I)$. For this sampling, we may consider another scenario: a causal intervention in a directed acyclic graph over random variables (Lattimore et al. 2016), described by $X^I \sim P(X^I | \text{do}(X^I = x^I))$. If the input distribution is neither known nor independent, the intervention scenario would differ from the scenario in this paper.

To make BOPSQ applicable to many real-world problems, it is important to extend BOPSQ to the setting in which selection costs of control variable sets are different (Tran-Thanh et al. 2012). In the problem setting of BOPSQ, in principle, we assume that the cost is uniform; however, specifying the values of more control variables

would be more costly in many real-world problems. Handling unbounded domains (Shahriari et al. 2016a) is another important extension of BOPSQ. We may not know the unbounded domain, especially for the unknown input distribution case.

A Proof of Theorem 1

In this section, we present the proof of Theorem 1, where the input distribution is known.

We first introduce a relevant result. Srinivas et al. (2012) provided a bound of sum of posterior variances over T iterations using the maximum information gain γ_T as

$$\sum_{t=1}^T \sigma_{t-1}^2(x_t) \leq \frac{2}{\log(1 + \sigma^{-2})} \gamma_T. \quad (8)$$

The analysis basically follows the frameworks by Russo and Roy (2014), Srinivas et al. (2012). First, we discretize the domain $\mathcal{X}$ in iteration t into grid points $[\mathcal{X}]_t$ so that $|[\mathcal{X}]_t| = \tau_t^d$ and $\|x - [x]_t\|_1 \leq d/\tau_t$ for all $x \in \mathcal{X}$, where $[x]_t$ is the closest point in $[\mathcal{X}]_t$ and $\tau_t = t^2 da'b'\sqrt{\pi}$. Note that this discretization is deterministic.

Define an upper confidence bound (UCB) sequence $U_t(\cdot) = \mu_{t-1}(\cdot) + \beta_t^{1/2} \sigma_{t-1}(\cdot)$ in iteration t , where $\beta_t = 4(d+1)\log(t) + 2d\log(da'b'\sqrt{\pi})$. Using the UCB sequence, we decompose $\text{BayesRegret}(T)$ into

$$\begin{aligned} \text{BayesRegret}(T) &= \underbrace{\sum_{t=1}^T \mathbb{E} \left[f(X^{I^*}, X^{\bar{I}^*}) - f([X^{I^*}]_t, [X^{\bar{I}^*}]_t) \right]}_{A_1} \\ &\quad + \underbrace{\sum_{t=1}^T \mathbb{E} \left[f([X^{I^*}]_t, [X^{\bar{I}^*}]_t) - U_t([X^{I^*}]_t, [X^{\bar{I}^*}]_t) \right]}_{A_2} \\ &\quad + \underbrace{\sum_{t=1}^T \mathbb{E} \left[U_t([X^{I^*}]_t, [X^{\bar{I}^*}]_t) - U_t([X^I]_t, [X^{\bar{I}^*}]_t) \right]}_{A_3} \\ &\quad + \underbrace{\sum_{t=1}^T \mathbb{E} \left[U_t([X^I]_t, [X^{\bar{I}^*}]_t) - f([X^I]_t, [X^{\bar{I}^*}]_t) \right]}_{A_4} \\ &\quad + \underbrace{\sum_{t=1}^T \mathbb{E} \left[f([X^I]_t, [X^{\bar{I}^*}]_t) - f(X^I, X^{\bar{I}^*}) \right]}_{A_5}. \end{aligned} \quad (9)$$

We will bound the terms A_1 and A_5 . By Assumption 4, we have

$$\begin{aligned}
A_1 &\leq \sum_{t=1}^T \mathbb{E} \left[\frac{d}{\tau_t} \sup_{x \in \mathcal{X}, j \in [d]} \left| \frac{\partial f(x)}{\partial x^{(j)}} \right| \right] \\
&\leq d \sum_{t=1}^T \frac{1}{t^2 da' b' \sqrt{\pi}} \int_0^\infty \mathbb{P} \left(\sup_{x \in \mathcal{X}, j \in [d]} \left| \frac{\partial f(x)}{\partial x^{(j)}} \right| \geq l \right) dl \\
&\leq \frac{\sqrt{\pi^3}}{6a' b'} \int_0^\infty a' e^{-l^2/b'^2} dl \\
&= \frac{\pi^2}{12}.
\end{aligned} \tag{10}$$

By applying the same argument to A_5 as above, we obtain the same upper-bound.

Next, we will upper-bound A_2 . For a random variable $Z \sim \mathcal{N}(\mu, \eta^2)$ with $\mu \leq 0$, $\mathbb{E}[Z \mathbb{1}[Z > 0]] = \int_0^\infty z/(\eta\sqrt{2\pi}) \exp(-(z - \mu)^2/(2\eta^2)) dz \leq (\eta/\sqrt{2\pi}) \exp(-\mu^2/(2\eta^2))$. Since $f(x) - U_t(x)$ conditioned on $\mathcal{D}_{t-1}$ follows $\mathcal{N}(-\beta_t^{1/2} \sigma_{t-1}(x), \sigma_{t-1}^2(x))$, we have

$$\mathbb{E}[\mathbb{1}[f(x) > U_t(x)] \cdot (f(x) - U_t(x)) \mid \mathcal{D}_{t-1}] \leq \frac{\sigma_{t-1}(x)}{\sqrt{2\pi}} \exp\left(-\frac{\beta_t}{2}\right) \leq \frac{1}{\sqrt{2\pi} |\mathcal{X}_t| t^2}, \tag{11}$$

where we used $\sigma_t(x) \leq 1$ and $\beta_t = 4(d+1)\log(t) + 2d\log(da'b'\sqrt{\pi}) = 2\log(t^2 |\mathcal{X}_t|)$. Then,

$$\begin{aligned}
A_2 &= \sum_{t=1}^T \mathbb{E} \left[\mathbb{E}[f([X^{I^*}]_t, [X^{\bar{I}^*}]_t) - U_t([X^{I^*}]_t, [X^{\bar{I}^*}]_t) \mid \mathcal{D}_{t-1}] \right] \\
&\leq \sum_{t=1}^T \mathbb{E}[\mathbb{E}[\mathbb{1}[f([X^{I^*}]_t, [X^{\bar{I}^*}]_t) - U_t([X^{I^*}]_t, [X^{\bar{I}^*}]_t) \geq 0] \\
&\quad \cdot (f([X^{I^*}]_t, [X^{\bar{I}^*}]_t) - U_t([X^{I^*}]_t, [X^{\bar{I}^*}]_t)) \mid \mathcal{D}_{t-1}]] \\
&\leq \sum_{t=1}^T \mathbb{E}[\mathbb{E}[\sum_{x \in \mathcal{X}_t} \mathbb{1}[f(x) - U_t(x) \geq 0] \cdot (f(x) - U_t(x)) \mid \mathcal{D}_{t-1}]]] \\
&\leq \sum_{t=1}^T \sum_{x \in \mathcal{X}_t} \frac{1}{\sqrt{2\pi} |\mathcal{X}_t| t^2} \\
&\leq \frac{\sqrt{2\pi^3}}{12}.
\end{aligned} \tag{12}$$

Since X^{I^*} and X^{I_t} are identically distributed from the same distribution conditioned on $\mathcal{D}_{t-1}$ and U_t is deterministic, we have

$$A_3 = \sum_{t=1}^T \mathbb{E} \left[\mathbb{E} \left[U_t([X^{I^*}]_t, [X^{\bar{I}^*}]_t) - U_t([X^{I_t}]_t, [X^{\bar{I}_t}]_t) \mid \mathcal{D}_{t-1} \right] \right] = 0. \tag{13}$$

Finally, we will bound A_4 . Since f and $(X^l, X^{\bar{l}})$ conditioned on $\mathcal{D}_{t-1}$ are independent, we have

$$\begin{aligned}
 A_4 &= \sum_{t=1}^T \mathbb{E} \left[\mathbb{E} \left[U_t([X^l]_t, [X^{\bar{l}}]_t) - f([X^l]_t, [X^{\bar{l}}]_t) \mid \mathcal{D}_{t-1}, X^l, X^{\bar{l}} \right] \right] \\
 &= \sum_{t=1}^T \beta_t^{1/2} \mathbb{E} \left[\sigma_{t-1}([X^l]_t, [X^{\bar{l}}]_t) \right] \\
 &\leq \beta_T^{1/2} \mathbb{E} \left[\sum_{t=1}^T \sigma_{t-1}([X^l]_t, [X^{\bar{l}}]_t) \right] \\
 &\leq \beta_T^{1/2} \mathbb{E} \left[\left(T \sum_{t=1}^T \sigma_{t-1}^2([X^l]_t, [X^{\bar{l}}]_t) \right)^{1/2} \right] \\
 &\leq \beta_T^{1/2} \mathbb{E} \left[\left(T \frac{2\gamma_T}{\log(1 + \sigma^{-2})} \right)^{1/2} \right] \\
 &= \sqrt{\frac{2T\beta_T\gamma_T}{\log(1 + \sigma^{-2})}},
 \end{aligned} \tag{14}$$

where we used the monotonic increase of β_t in the first inequality, the Cauchy-Schwarz inequality in the second inequality, and Inequality (8) in the third inequality.

By combining Eq. (9) and Inequalities (10), (12), (13), and (14), we complete the proof. $\square$

B Proof of Theorem 2

In this section, we present the proof of Theorem 2, where the input distribution is unknown and assumed to be independent of each other variables.

We define an UCB sequence $U_t(\cdot) = \mu_{t-1}(\cdot) + \beta_t^{1/2} \sigma_{t-1}(\cdot)$ in iteration t , where $\beta_t = 4(d+1) \log(t) + 2d \log(da'b' \sqrt{\pi})$. We further define

$$\begin{aligned}
 \bar{g}_t(x^l, x^{\bar{l}}) &= g_t(x^l, x^{\bar{l}}) + 2 \sum_{i \in \bar{I}} \frac{\alpha_i}{\sqrt{|\mathcal{S}_{i-1}^{(i)}|}} \\
 X^{J_t} &= \operatorname{argmax}_{x^l \in \mathcal{X}^l, J \in \bar{I}} \mathbb{E}_{X^{\bar{j}} \sim F^{\bar{j}}} \left[g_t(x^l, X^{\bar{j}}) \right].
 \end{aligned}$$

Then, we decompose the regret at iteration t into

$$\begin{aligned}
& \mathbb{E}\left[f\left(X^{I^*}, X^{\bar{I}^*}\right)\right] - \mathbb{E}\left[f\left(X^I, X^{\bar{I}}\right)\right] \\
&= \underbrace{\mathbb{E}\left[f\left(X^{I^*}, X^{\bar{I}^*}\right)\right] - \mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[g_t\left(X^I, X^{\bar{I}}\right)\right]\right]}_{B_{1,t}} \\
&+ \underbrace{\mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[g_t\left(X^I, \bar{J}_t\right)\right]\right] - \mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim \hat{F}^{\bar{I}_t}}\left[\bar{g}_t\left(X^I, X^{\bar{I}}\right)\right]\right]}_{B_{2,t}} \\
&+ \underbrace{\mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim \hat{F}^{\bar{I}_t}}\left[\bar{g}_t\left(X^I, X^{\bar{I}}\right)\right]\right] - \mathbb{E}_{X^{\bar{I}_t} \sim \hat{F}^{\bar{I}_t}}\left[\bar{g}_t\left(X^I, X^{\bar{I}}\right)\right]}_{B_{3,t}} \\
&+ \underbrace{\mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim \hat{F}^{\bar{I}_t}}\left[\bar{g}_t\left(X^I, X^{\bar{I}}\right)\right]\right] - \mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[\bar{\bar{g}}_t\left(X^I, X^{\bar{I}}\right)\right]\right]}_{B_{4,t}} \\
&+ \underbrace{\mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[\bar{\bar{g}}_t\left(X^I, X^{\bar{I}}\right)\right]\right] - \mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[g_t\left(X^I, X^{\bar{I}}\right)\right]\right]}_{B_{5,t}} \\
&+ \underbrace{\mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[g_t\left(X^I, X^{\bar{I}}\right)\right]\right] - \mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[U_t\left(X^I, X^{\bar{I}}\right)\right]\right]}_{B_{6,t}} \\
&+ \underbrace{\mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[U_t\left(X^I, X^{\bar{I}}\right)\right]\right] - \mathbb{E}\left[f\left(X^I, X^{\bar{I}}\right)\right]}_{B_{7,t}}.
\end{aligned} \tag{15}$$

First, since f and g_t follow the same distribution given $\mathcal{D}_{t-1}$, we have

$$\begin{aligned}
\mathbb{E}\left[f\left(X^{I^*}, X^{\bar{I}^*}\right)\right] &= \mathbb{E}\left[\max_{x^I} \mathbb{E}_{X^{\bar{I}} \sim F^{\bar{I}}}\left[f\left(x^I, X^{\bar{I}}\right)\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\max_{x^I} \mathbb{E}_{X^{\bar{I}} \sim F^{\bar{I}}}\left[f\left(x^I, X^{\bar{I}}\right)\right] \middle| \mathcal{D}_{t-1}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\max_{x^I} \mathbb{E}_{X^{\bar{I}} \sim F^{\bar{I}}}\left[g_t\left(x^I, X^{\bar{I}}\right)\right] \middle| \mathcal{D}_{t-1}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[g_t\left(X^I, X^{\bar{I}}\right)\right] \middle| \mathcal{D}_{t-1}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}_{X^{\bar{I}_t} \sim F^{\bar{I}_t}}\left[g_t\left(X^I, X^{\bar{I}}\right)\right]\right],
\end{aligned}$$

for which we have

$$B_{1,t} = 0. \tag{16}$$

Next, we will upper-bound $B_{2,t}$ and $B_{4,t}$. We decompose $B_{2,t}$ into

$$\begin{aligned}
B_{2,t} &= \mathbb{E} \left[\mathbb{1} [\|g'_t\|_\infty \leq \zeta_t] \left(\mathbb{E}_{X^{j_t} \sim F^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] - \mathbb{E}_{X^{j_t} \sim \hat{F}^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] \right) \right] \\
&\quad - \mathbb{E} \left[\sum_{i \in \bar{j}_t} \frac{\alpha_t}{\sqrt{|\mathcal{S}_t^{(i)}|}} \right] \\
&\quad + \mathbb{E} \left[\mathbb{1} [\|g'_t\|_\infty > \zeta_t] \left(\mathbb{E}_{X^{j_t} \sim F^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] - \mathbb{E}_{X^{j_t} \sim \hat{F}^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] \right) \right] \\
&\leq \mathbb{E} \left[\mathbb{1} [\|g'_t\|_\infty \leq \zeta_t] \left(\mathbb{E}_{X^{j_t} \sim F^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] - \mathbb{E}_{X^{j_t} \sim \hat{F}^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] \right) \right] \\
&\quad - \mathbb{E} \left[\sum_{i \in \bar{j}_t} \frac{\alpha_t}{\sqrt{|\mathcal{S}_t^{(i)}|}} \right] + 2\mathbb{E} [\mathbb{1} [\|g'_t\|_\infty > \zeta_t] \|g_t\|_\infty] \\
&\leq \mathbb{E} \left[\mathbb{1} [\|g'_t\|_\infty \leq \zeta_t] \left(\mathbb{E}_{X^{j_t} \sim F^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] - \mathbb{E}_{X^{j_t} \sim \hat{F}^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] \right) \right] \quad (17) \\
&\quad - \mathbb{E} \left[\sum_{i \in \bar{j}_t} \frac{\alpha_t}{\sqrt{|\mathcal{S}_t^{(i)}|}} \right] \\
&\quad + 2\eta_t \mathbb{E} [\mathbb{P} [\|g'_t\|_\infty > \zeta_t | \mathcal{D}_{t-1}]] + 2\mathbb{E} [\mathbb{E} [\mathbb{1} [\|g_t\|_\infty > \eta_t] \|g_t\|_\infty | \mathcal{D}_{t-1}]] \\
&= \underbrace{\mathbb{E} \left[\mathbb{1} [\|g'_t\|_\infty \leq \zeta_t] \left(\mathbb{E}_{X^{j_t} \sim F^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] - \mathbb{E}_{X^{j_t} \sim \hat{F}^{j_t}} [g_t(X^{j_t}, X^{\bar{j}_t})] \right) \right]}_{B_{2-1,t}} \\
&\quad - \underbrace{\mathbb{E} \left[\sum_{i \in \bar{j}_t} \frac{\alpha_t}{\sqrt{|\mathcal{S}_t^{(i)}|}} \right]}_{B_{2-2,t}} + \underbrace{2\eta_t \mathbb{P} [\|f'\|_\infty > \zeta_t]}_{B_{2-3,t}} + \underbrace{2\mathbb{E} [\mathbb{1} [\|f\|_\infty > \eta_t] \|f\|_\infty]}_{B_{2-4,t}},
\end{aligned}$$

where $\|g'_t\|_\infty := \sup_{x \in \mathcal{X}, j \in [d]} |\partial g_t(x)/\partial x^{(j)}|$, $\zeta_t := b' \sqrt{2 \log t}$, $\eta_t := b \sqrt{2 \log(t+1)}$, and $b', b > 0$ are the constants in Assumptions (7) and (4). Here, the last equality is because f and g'_t are independently distributed from the same distribution given $\mathcal{D}_{t-1}$.

Using the independence assumption of the input random variables and integration by parts, we have

$$\begin{aligned}
& \mathbb{E}_{X^{\bar{J}_t} \sim F^{\bar{J}_t}} \left[g_t \left(X^{\bar{J}_t}, X^{\bar{J}_t} \right) \right] - \mathbb{E}_{X^{\bar{J}_t} \sim \hat{F}^{\bar{J}_t}} \left[g_t \left(X^{\bar{J}_t}, X^{\bar{J}_t} \right) \right] \\
&= \sum_{i=1}^{|\bar{J}_t|} \mathbb{E}_{X^{\bar{J}_t \setminus \bar{J}_{t,i}}} \left[\int_0^1 g_t \left(x^{\bar{J}_{t,i}}, X^{\bar{J}_t \setminus \bar{J}_{t,i}} \right) \left(dF^{\bar{J}_{t,i}}(x^{\bar{J}_{t,i}}) - d\hat{F}^{\bar{J}_{t,i}}(x^{\bar{J}_{t,i}}) \right) \right] \\
&= \sum_{i=1}^{|\bar{J}_t|} \mathbb{E}_{X^{\bar{J}_t \setminus \bar{J}_{t,i}}} \left[- \int_0^1 \frac{\partial g_t}{\partial x^{\bar{J}_{t,i}}} \left(x^{\bar{J}_{t,i}}, X^{\bar{J}_t \setminus \bar{J}_{t,i}} \right) \left(F^{\bar{J}_{t,i}}(x^{\bar{J}_{t,i}}) - \hat{F}^{\bar{J}_{t,i}}(x^{\bar{J}_{t,i}}) \right) dx^{\bar{J}_{t,i}} \right] \\
&\leq \|g'_t\|_\infty \sum_{i \in \bar{J}_t} e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)},
\end{aligned} \tag{18}$$

where

$$e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} := \sup_{x^{(i)} \in \mathcal{X}^{(i)}} \left| F^{(i)}(x^{(i)}) - \hat{F}_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} \right|$$

and $\mathbb{E}_{X^{\bar{J}_t \setminus \bar{J}_{t,i}}}$ denotes an expectation operator for abbreviation defined as

$$\mathbb{E}_{X^{\bar{J}_t \setminus \bar{J}_{t,i}}} := \mathbb{E}_{X^{\bar{J}_{t,1}} \sim F^{\bar{J}_{t,1}}} \cdots \mathbb{E}_{X^{\bar{J}_{t,i-1}} \sim F^{\bar{J}_{t,i-1}}} \mathbb{E}_{X^{\bar{J}_{t,i+1}} \sim \hat{F}^{\bar{J}_{t,i+1}}} \cdots \mathbb{E}_{X^{\bar{J}_t \setminus \bar{J}_t} \sim \hat{F}^{\bar{J}_t \setminus \bar{J}_t}}.$$

Using Inequality (18), we have

$$\begin{aligned}
B_{2-1,t} + B_{2-2,t} &\leq \mathbb{E} \left[\mathbb{1} [\|g'_t\|_\infty \leq \zeta_t] \|g'_t\|_\infty \sum_{i \in \bar{J}_t} e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} \right] - \mathbb{E} \left[\sum_{i \in \bar{J}_t} \frac{\alpha_t}{\sqrt{|\mathcal{S}_{t-1}^{(i)}|}} \right] \\
&\leq \zeta_t \sum_{i \in \bar{J}_t} \mathbb{E} \left[e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} - \sqrt{\frac{2 \log t}{|\mathcal{S}_{t-1}^{(i)}|}} \right]
\end{aligned} \tag{19}$$

Since samples $x^i \in \mathcal{S}_{t-1}^i$ for $i \in \bar{J}_t$ are independently distributed given $|\mathcal{S}_{t-1}^i|$, by applying the Dvoretzky–Kiefer–Wolfowitz inequality (Dvoretzky et al. 1956; Massart 1990) conditioned on $n = |\mathcal{S}_{t-1}^i|$, for $\epsilon > 0$, we have

$$P_n^{(i)}(\epsilon) := \mathbb{P} \left(e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} > \epsilon \mid |\mathcal{S}_{t-1}^i| = n \right) \leq 2 \exp(-2n\epsilon^2).$$

Define $\epsilon_{t,n} := \epsilon + \sqrt{(2 \log t)/n}$ for $\epsilon > 0$, and we have

$$\begin{aligned}
& \mathbb{E} \left[e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} - \sqrt{\frac{2 \log t}{|\mathcal{S}_{t-1}^{(i)}|}} \right] \\
&= \mathbb{E} \left[\mathbb{1} \left[e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} - \sqrt{\frac{2 \log t}{|\mathcal{S}_t^{(i)}|}} \leq \epsilon \right] \left(e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} - \sqrt{\frac{2 \log t}{|\mathcal{S}_t^{(i)}|}} \right) \right] \\
&\quad + \mathbb{E} \left[\mathbb{1} \left[e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} - \sqrt{\frac{2 \log t}{|\mathcal{S}_t^{(i)}|}} > \epsilon \right] \left(e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} - \sqrt{\frac{2 \log t}{|\mathcal{S}_t^{(i)}|}} \right) \right] \\
&\leq \epsilon + \sum_{n=1}^T \mathbb{E} \left[\mathbb{1} \left[e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} - \sqrt{\frac{2 \log t}{n}} > \epsilon, |\mathcal{S}_t^{(i)}| = n \right] \left(e_{|\mathcal{S}_{t-1}^{(i)}|}^{(i)} - \sqrt{\frac{2 \log t}{n}} \right) \right] \\
&\leq \epsilon + \sum_{n=1}^T \int_{\epsilon_{t,n}}^{\infty} \left(x - \sqrt{\frac{2 \log t}{n}} \right) d(-P_n^{(i)}(x)) \\
&\leq \epsilon + \sum_{n=1}^T \left(- \left[\left(x - \sqrt{\frac{2 \log t}{n}} \right) P_n^{(i)}(x) \right]_{\epsilon_{t,n}}^{\infty} + \int_{\epsilon_{t,n}}^{\infty} P_n^{(i)}(x) dx \right) \tag{20} \\
&\leq \epsilon + \sum_{n=1}^T \left(2\epsilon e^{-2n\epsilon_{t,n}^2} + \int_{\epsilon_{t,n}}^{\infty} \frac{1}{nx} \cdot 2nxe^{-2nx^2} dx \right) \\
&= \epsilon + \sum_{n=1}^T \left(2\epsilon e^{-n\epsilon_{t,n}^2} + \left[-\frac{1}{nx} \cdot e^{-nx^2} \right]_{\epsilon_{t,n}}^{\infty} - \int_{\epsilon_{t,n}}^{\infty} \frac{1}{nx^2} \cdot e^{-nx^2} dx \right) \\
&\leq \epsilon + 2\epsilon \sum_{n=1}^T e^{-n\epsilon_{t,n}^2} + \sum_{n=1}^T \frac{1}{n\epsilon_{t,n}} e^{-n\epsilon_{t,n}^2} \\
&\leq \epsilon + 2\epsilon \sum_{n=1}^T e^{-n\epsilon^2 - 2 \log t} + \sum_{n=1}^T \frac{1}{n\epsilon} e^{-n\epsilon^2 - 2 \log t} \\
&= \epsilon + \frac{2\epsilon}{t^2} \sum_{n=1}^{\infty} e^{-n\epsilon^2} + \frac{1}{\epsilon t^2} \sum_{n=1}^T \frac{1}{n} e^{-n\epsilon^2} \\
&\leq \epsilon + \frac{2\epsilon}{t^2} \cdot \frac{1}{e^{\epsilon^2} - 1} - \frac{1}{\epsilon t^2} \log(1 - e^{-\epsilon^2})
\end{aligned}$$

$$\leq \epsilon + \frac{2\epsilon}{t^2} \cdot \frac{1}{\epsilon^2} + \frac{1}{\epsilon t^2} \left(\epsilon^2 + \log \left(\frac{1}{e^{\epsilon^2} - 1} \right) \right) \tag{21}$$

$$\leq \epsilon + \frac{2}{\epsilon t^2} + \frac{\epsilon}{t^2} + \frac{1}{\epsilon t^2} \log \frac{1}{\epsilon^2} \tag{22}$$

$$\leq \epsilon + \frac{2}{\epsilon t^2} + \frac{\epsilon}{t^2} + \frac{1}{\epsilon t^2} \log \left(\frac{1}{\epsilon^2} \right), \tag{23}$$

where we used $\sum_{n=1}^{\infty} x^n/n = -\log(1-x)$ in Equality (20) and $e^x \geq x+1$ in Inequalities (21) and (22). By taking the summation over T and combining Inequalities (19) and (23), we have

$$\begin{aligned} \sum_{t=1}^T (B_{2-1,t} + B_{2-2,t}) &\leq \sum_{t=1}^T \zeta_t \sum_{i \in \bar{J}_t} \left(\epsilon + \frac{2}{\epsilon t^2} + \frac{\epsilon}{t^2} + \frac{1}{\epsilon t^2} \log \left(\frac{1}{\epsilon^2} \right) \right) \\ &\leq b' m \sqrt{2 \log T} \left(\epsilon T + \frac{\pi^2}{3\epsilon} + \frac{\pi^2}{6} \epsilon + \frac{\pi^2}{6\epsilon} \log \left(\frac{1}{\epsilon^2} \right) \right), \end{aligned} \quad (24)$$

where $m = \max_{I \in \mathcal{I}} |\bar{I}|$.

$B_{2-3,t}$ after taking summation over T is upper-bounded because of Assumption (4).

$$\begin{aligned} \sum_{t=1}^T B_{2-3,t} &\leq \sum_{t=1}^T 2\eta_t a' e^{-(\zeta_t/b')^2} \\ &\leq 2a'b \sqrt{2 \log(T+1)} \sum_{t=1}^T \frac{1}{t^2} \\ &\leq \frac{\pi^2 a'b \sqrt{2 \log(T+1)}}{3}, \end{aligned} \quad (25)$$

where we used monotonic increase of B in the second inequality.

The term $B_{2-4,t}$ is upper-bounded as

$$\begin{aligned} B_{2-4,t} &= 2\mathbb{E} \left[\int_0^{\infty} \mathbb{1}[\mathbb{1}[\|f\|_{\infty} > \eta_t] \cdot \|f\|_{\infty} > z] dz \right] \\ &= 2\mathbb{E} \left[\int_0^{\infty} \mathbb{1}[\|f\|_{\infty} > \eta_t] \cdot \mathbb{1}[\|f\|_{\infty} > z] dz \right] \\ &= 2\mathbb{E} \left[\int_0^{\eta_t} \mathbb{1}[\|f\|_{\infty} > \eta_t] dz + \int_{\eta_t}^{\infty} \mathbb{1}[\|f\|_{\infty} > z] dz \right] \\ &= 2\eta_t \mathbb{P}[\|f\|_{\infty} > \eta_t] + 2 \int_{\eta_t}^{\infty} \mathbb{P}(\|f\|_{\infty} > z) dz \\ &\leq 2\eta_t a e^{-(\eta_t/b)^2} + 2 \int_{\eta_t}^{\infty} a e^{-(z/b)^2} dz \\ &\leq 2\eta_t a e^{-(\eta_t/b)^2} + 2a \int_{\eta_t}^{\infty} \frac{z}{\eta_t} e^{-(z/b)^2} dz \\ &= 2\eta_t a e^{-(\eta_t/b)^2} + \frac{ab^2}{\eta_t} e^{-(\eta_t/b)^2} \\ &= \left(2\eta_t + \frac{b^2}{\eta_t} \right) a e^{-(\eta_t/b)^2}, \end{aligned} \quad (26)$$

where we used Assumptions (7) and (4) in Inequality (26). Therefore, by taking summation over T , we have

$$\begin{aligned}
\sum_{t=1}^T B_{2-4,t} &\leq \sum_{t=1}^T \left(2\eta_t + \frac{b^2}{\eta_t} \right) a e^{-(\eta_t/b)^2} \\
&\leq \left(2\sqrt{2\log(T+1)} + \frac{1}{\sqrt{2\log 2}} \right) ab \sum_{t=1}^T \frac{1}{t^2 + 1} \\
&= \frac{\pi^2 ab \sqrt{2\log(T+1)}}{3} + \frac{\pi^2 ab}{6\sqrt{2\log 2}},
\end{aligned} \tag{28}$$

where we used Inequality (27) in the first inequality and the monotonic increase of η_t in the second inequality.

By combining Inequalities (17), (24), (25) and (28), we obtain

$$\begin{aligned}
\sum_{t=1}^T B_{2,t} &\leq \frac{\pi^2 ab}{6\sqrt{2\log 2}} + \frac{\pi^2 b(a+a')\sqrt{2\log(T+1)}}{3} \\
&\quad + b'm\sqrt{2\log T} \left(\epsilon T + \frac{\pi^2}{3\epsilon} + \frac{\pi^2}{6}\epsilon + \frac{\pi^2}{6\epsilon} \log\left(\frac{1}{\epsilon^2}\right) \right)
\end{aligned} \tag{29}$$

Applying the same argument to $\sum_{t=1}^T B_{4,t}$, we obtain the same bound.

Due to the definition of the acquisition function given $\mathcal{D}_{t-1}$, we have

$$B_{3,t} \leq 0. \tag{30}$$

Next, we upper-bound $B_{5,t}$. We have

$$\begin{aligned}
\sum_{t=1}^T B_{5,t} &\leq 2\alpha_T \mathbb{E} \left[\sum_{t=1}^T \sum_{i \in \bar{I}_t} \frac{1}{\sqrt{|\mathcal{S}_{t-1}^{(i)}|}} \right] \\
&= 2\alpha_T \mathbb{E} \left[\sum_{n=1}^T \sum_{i=1}^d \sum_{t=1}^T \frac{\mathbb{1}[i \in \bar{I}_t, |\mathcal{S}_{t-1}^{(i)}| = n]}{\sqrt{n}} \right] \\
&= 2\alpha_T \mathbb{E} \left[\sum_{n=1}^T \sum_{i=1}^d \frac{\mathbb{1}\left[\bigcup_{t=1}^T \{i \in \bar{I}_t, |\mathcal{S}_{t-1}^{(i)}| = n\}\right]}{\sqrt{n}} \right]
\end{aligned} \tag{31}$$

$$\leq 2\alpha_T \mathbb{E} \left[\sum_{n=1}^{\lceil mT/d \rceil} \sum_{i=1}^d \frac{1}{\sqrt{n}} \right] \tag{32}$$

$$\begin{aligned}
&\leq 2d\alpha_T \int_0^{mT/d+1} \frac{1}{\sqrt{x}} dx \\
&\leq 8b' \sqrt{d(mT+d) \log T},
\end{aligned} \tag{33}$$

where we used monotonic increase of α_t in the first inequality. Equality (31) is because events $\{i \in \bar{I}_t, |\mathcal{S}_{t-1}^{(i)}| = n\}$ for $n = 1, \dots, T$ happen once at most over $t = 1, \dots, T$. Inequality (32) is because $\sum_{n=1}^T \sum_{i=1}^d \mathbb{1} \left[\bigcup_{t=1}^T \left\{ i \in \bar{I}_t, |\mathcal{S}_{t-1}^{(i)}| = n \right\} \right] \leq mT$ and $1/\sqrt{n}$ is monotonically decreasing in n .

To upper-bound the term $B_{6,t}$, we decomposed it into three terms as

$$\begin{aligned}
B_{6,t} = & \underbrace{\mathbb{E} \left[\mathbb{E}_{X^l_t \sim F^l_t} \left[g_t(X^l_t, X^{\bar{l}}_t) - g_t([X^l_t]_t, [X^{\bar{l}}_t]_t) \right] \middle| \mathcal{D}_{t-1} \right]}_{B_{6-1,t}} \\
& + \underbrace{\mathbb{E} \left[\mathbb{E}_{X^l_t \sim F^l_t} \left[g_t([X^l_t]_t, [X^{\bar{l}}_t]_t) - U_t([X^l_t]_t, [X^{\bar{l}}_t]_t) \right] \middle| \mathcal{D}_{t-1} \right]}_{B_{6-2,t}} \\
& + \underbrace{\mathbb{E} \left[\mathbb{E}_{X^l_t \sim F^l_t} \left[(U_t([X^l_t]_t, [X^{\bar{l}}_t]_t) - U_t(X^l_t, X^{\bar{l}}_t)) \right] \middle| \mathcal{D}_{t-1} \right]}_{B_{6-3,t}}.
\end{aligned} \tag{34}$$

Because f and g'_t are independently distributed from the same distribution given $\mathcal{D}_{t-1}$, we upper-bound the term $B_{6-1,t}$ after summing over T as

$$\begin{aligned}
\sum_{t=1}^T B_{6-1,t} &\leq \sum_{t=1}^T \frac{d}{\tau_t} \mathbb{E} \left[\mathbb{E} \left[\sup_{x \in \mathcal{X}, j \in [d]} \left| \frac{\partial g_t(x)}{\partial x^{(j)}} \right| \middle| \mathcal{D}_{t-1} \right] \right] \\
&= \sum_{t=1}^T \frac{d}{\tau_t} \mathbb{E} \left[\mathbb{E} \left[\sup_{x \in \mathcal{X}, j \in [d]} \left| \frac{\partial f(x)}{\partial x^{(j)}} \right| \middle| \mathcal{D}_{t-1} \right] \right] \\
&= \sum_{t=1}^T \frac{d}{\tau_t} \mathbb{E} \left[\sup_{x \in \mathcal{X}, j \in [d]} \left| \frac{\partial f(x)}{\partial x^{(j)}} \right| \right] \\
&\leq \frac{\pi^2}{12},
\end{aligned} \tag{35}$$

where we applied the same argument as Inequality (10) in the last inequality.

Since $g_t(x) - U_t(x)$ conditioned on $\mathcal{D}_{t-1}$ follows $\mathcal{N}(-\beta_t^{1/2} \sigma_{t-1}(x), \sigma_{t-1}^2(x))$, by applying the same argument as Inequality (12), we have

$$\begin{aligned}
\sum_{t=1}^T B_{6-2,t} &\leq \sum_{t=1}^T \mathbb{E}[\mathbb{E}[\mathbb{1}[g_t([X^L]_t, [X^{\bar{L}}]_t) - U_t([X^L]_t, [X^{\bar{L}}]_t) \geq 0] \\
&\quad \cdot (g_t([X^L]_t, [X^{\bar{L}}]_t) - U_t([X^L]_t, [X^{\bar{L}}]_t)) \mid \mathcal{D}_{t-1}]] \\
&= \sum_{t=1}^T \mathbb{E}[\mathbb{E}[\sum_{x \in [\mathcal{X}]_t} \mathbb{1}[g_t(x) - U_t(x) \geq 0] \cdot (g_t(x) - U_t(x)) \mid \mathcal{D}_{t-1}]] \\
&\leq \sum_{t=1}^T \mathbb{E}[\mathbb{E}[\sum_{x \in [\mathcal{X}]_t} \mathbb{1}[f(x) - U_t(x) \geq 0] \cdot (f(x) - U_t(x)) \mid \mathcal{D}_{t-1}]] \\
&\leq \sum_{t=1}^T \sum_{x \in [\mathcal{X}]_t} \frac{1}{\sqrt{2\pi} |[\mathcal{X}]_t| t^2} \\
&\leq \frac{\sqrt{2\pi^3}}{12}.
\end{aligned} \tag{36}$$

Since f and $(X^L, X^{\bar{L}})$ are independent given $\mathcal{D}_{t-1}$, we upper-bound $B_{6-3,t}$ as

$$\begin{aligned}
\sum_{t=1}^T B_{6-3,t} &\leq \sum_{t=1}^T \mathbb{E} \left[\mathbb{E} \left[\left| f([X^L]_t, [X^{\bar{L}}]_t) - f(X^L, X^{\bar{L}}) \right| \mid X^L, X^{\bar{L}}, \mathcal{D}_{t-1} \right] \right] \\
&\quad + \sum_{t=1}^T \mathbb{E} \left[\beta_t^{1/2} (\sigma_{t-1}([X^L]_t, [X^{\bar{L}}]_t) - \sigma_{t-1}(X^L, X^{\bar{L}})) \right] \\
&\leq \sum_{t=1}^T \mathbb{E} \left[\frac{d}{\tau_t} \sup_{x \in \mathcal{X}, j \in [d]} \left| \frac{\partial f(x)}{\partial x^{(j)}} \right| \right] + \sum_{t=1}^T \beta_t^{1/2} \mathbb{E} \left[\sigma_{t-1}([X^L]_t, [X^{\bar{L}}]_t) \right] \\
&\leq \frac{\pi^2}{12} + \sqrt{\frac{2T\beta_T\gamma_T}{\log(1 + \sigma^{-2})}},
\end{aligned} \tag{37}$$

where we applied the same argument as Inequalities (10) and (14) in the last inequality.

By combining Eq. (34) and Inequalities (35), (36), and (37), we have

$$\sum_{t=1}^T B_{6,t} \leq \frac{\pi^2}{6} + \frac{\sqrt{2\pi^3}}{12} + \sqrt{\frac{2T\beta_T\gamma_T}{\log(1 + \sigma^{-2})}} \tag{38}$$

Since f and $(X^L, X^{\bar{L}})$ given $\mathcal{D}_{t-1}$ are independent, by applying the same argument as (14) to $B_{7,t}$ over T , we have

$$\sum_{t=1}^T B_{7,t} \leq \sqrt{\frac{2T\beta_T\gamma_T}{\log(1 + \sigma^{-2})}}. \tag{39}$$

Finally, we combine Eq. (15) and Inequalities (16), (29), (30), (33), (38), and (39), and we have

$$\begin{aligned}
\text{BayesRegret}(T) \leq & \frac{\pi^2}{6} + \frac{\sqrt{2\pi^3}}{12} + \frac{\pi^2 ab}{3\sqrt{2\log 2}} \\
& + \frac{2\pi^2(a+a')'b\sqrt{2\log(T+1)}}{3} + 8b'\sqrt{d(mT+d)}\log T \\
& + 2b'm\sqrt{2\log T} \left(\epsilon T + \frac{\pi^2}{3\epsilon} + \frac{\pi^2}{6}\epsilon + \frac{\pi^2}{6\epsilon} \log\left(\frac{1}{\epsilon^2}\right) \right) + 2\sqrt{\frac{2T\beta_T\gamma_T}{\log(1+\sigma^{-2})}}
\end{aligned} \tag{40}$$

By setting $\epsilon = \sqrt{\log T/T}$, we have

$$\begin{aligned}
\text{BayesRegret}(T) \leq & \frac{\pi^2}{6} + \frac{\sqrt{2\pi^3}}{12} + \frac{\pi^2 ab}{3\sqrt{2\log 2}} + \frac{\sqrt{2\pi^2 b'm} \log T}{3\sqrt{T}} \\
& + \frac{2\pi^2(a+a')'b\sqrt{2\log(T+1)}}{3} + \frac{2\pi^2 b'm\sqrt{2T}}{3} + \frac{\pi^2 b'm\sqrt{2T}}{3} \log\left(\frac{T}{\log T}\right) \\
& + 2b'm\sqrt{2T\log T} + 8b'\sqrt{d(mT+d)}\log T + 2\sqrt{\frac{2dT\beta_T\gamma_T}{\log(1+\sigma^{-2})}} \\
& = O\left(\sqrt{dT\log T}(\sqrt{\gamma_T} + \sqrt{m\log T})\right),
\end{aligned} \tag{41}$$

which completes the proof. $\square$

C Discussion on estimating unknown dependent input distributions

We consider a simple toy example that illustrates the essential hardness of the problem when the input distribution is not necessarily independent.

For simplicity we consider the binary input space, $\mathcal{X} = \{0, 1\}^6$ with $\mathcal{I} = \{I \mid I \subset [d], |I| = 3\}$, but the discussion below is easily extended to the continuous case with general input dimension d and the number of control variables $|I|$. Let us consider the case that the learner already knows the complete information of the objective function $f(x)$, which is expressed as

$$f(x) = \begin{cases} 1, & \text{if } |\{i \mid x^{(i)} = 0\}| \geq 4, \\ 0, & \text{otherwise.} \end{cases}$$

Now, consider the following two distributions P and Q . Under P , $X^{(i)}$ is i.i.d. with $P(X^{(i)} = 0) = \epsilon$ for sufficiently small $\epsilon > 0$. Under Q , the sequences in $\mathcal{S}$ given below appear more frequently, satisfying

$$Q(x) = \begin{cases} P(x)/\epsilon, & \text{if } x \in \mathcal{S}, \\ P(x)/A, & \text{otherwise,} \end{cases}$$

where

$$S = \{000011, 100011, 010011, 001011\},$$

$$A = \frac{1 - \sum_{x \in S} P(x)}{1 - \sum_{x \in S} Q(x)} = \frac{1 - \epsilon^3(1 - \epsilon)^2(3 - 2\epsilon^2)}{1 - \epsilon^2(1 - \epsilon)^2(3 - 2\epsilon^2)} \approx 1.$$

We can see that, under Q , it is uniquely optimal to choose $I = \{1, 2, 3\}$ with $x^I = 000$. This is because, in this choice $Q(X^I = 011|X^I = 000) \approx 1/2$ whereas X^I almost always becomes 111 (resulting in $f(x) = 0$) in any other choice. Nevertheless, it requires an enormous number of samples to distinguish P and Q unless choosing $I = \{1, 2, 3\}$ with $x = 000$, because X^I almost always takes value 111 in all the other cases.

By the same argument, for arbitrary I such that $|I| = 3$, we can consider a distribution Q' such that choosing I with $x^I = 000$ is optimal but Q' cannot be distinguished from P unless trying this choice. Therefore, it is necessary to try combinatorially many candidates of I to find the input such that $f(x)$ becomes 1 with a nonnegligible probability. From this observation, we can conclude that dependent input distributions essentially make the task exponentially hard.

Acknowledgements This research was supported by JSPS KAKENHI Grant Numbers 20H04244 and 21K11747.

Author Contributions Conceptualization: Shogo Hayashi; Methodology: Shogo Hayashi, Junya Honda; Formal analysis and investigation: Shogo Hayashi, Junya Honda, Hisashi Kashima; Writing - original draft preparation: Shogo Hayashi; Writing - review and editing: Junya Honda, Hisashi Kashima; Funding acquisition: Junya Honda, Hisashi Kashima; Resources: Hisashi Kashima; Supervision: Junya Honda, Hisashi Kashima.

Availability of data and material (data transparency) The real-world datasets used in the paper are available via – The airfoil self-noise (<https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>) – The word similarity dataset (<https://sites.google.com/site/nlpannotations/>) (Snow et al. 2008)

Declarations

Conflict of interest Not applicable.

Code availability (software application or custom code) We are preparing to publish the code online.

Additional declarations for articles in life science journals that report the results of studies involving humans and/or animals Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson, B. S., Moore, A. W., & Cohn, D. (2000). A nonparametric approach to noisy and costly optimization. In: *Proceedings of the 17th international conference on machine learning*, pp. 17–24.
- Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2–3), 235–256.
- Bochner, S. (1959). *Lectures on Fourier integrals*. Princeton University Press.
- Bogunovic, I., Scarlett, J., Jegelka, S., & Cevher, V. (2018). Adversarially robust optimization with Gaussian processes. *Advances in Neural Information Processing Systems*, 31, 5765–5775.
- Chen, B., Castro, R. M., & Krause, A. (2012). Joint optimization and variable selection of high-dimensional Gaussian processes. In: *Proceedings of the 29th international conference on machine learning*.
- Dai, Z., Chen, Y., Low, B. K. H., Jaillet, P., & Ho, T. (2020). R2-B2: Recursive reasoning-based Bayesian optimization for no-regret learning in games. In: *Proceedings of the 37th international conference on machine learning*, pp. 2291–2301.
- Dvoretzky, A., Kiefer, J., & Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3), 642–669.
- Garreau, D., Jitkrittum, W., & Kanagawa, M. (2017). Large sample analysis of the median heuristic. arXiv preprint [arXiv:1707.07269](https://arxiv.org/abs/1707.07269)
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182.
- Hernández-Lobato, J. M., Hoffman, M. W., & Ghahramani, Z. (2014). Predictive entropy search for efficient global optimization of black-box functions. *Advances in Neural Information Processing Systems*, 27, 918–926.
- Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4), 455–492.
- Kandasamy, K., Schneider, J. G., & Póczos, B. (2015). High dimensional Bayesian optimisation and bandits via additive models. In: *Proceedings of the 32nd international conference on machine learning*, pp. 295–304.
- Kirschner, J., Bogunovic, I., Jegelka, S., & Krause, A. (2020). Distributionally robust Bayesian optimization. In: *The 23rd international conference on artificial intelligence and statistics*, pp. 2174–2184.
- Korovina, K., Xu, S., Kandasamy, K., Neiswanger, W., Póczos, B., Schneider, J., & Xing, E. P. (2020). Chemo: Bayesian optimization of small organic molecules with synthesizable recommendations. In: *The 23rd international conference on artificial intelligence and statistics*, pp. 3393–3403.
- Krause, A., & Ong, C. S. (2011). Contextual Gaussian process bandit optimization. *Advances in Neural Information Processing Systems*, 24, 2447–2455.
- Kushner, H. J. (1964). A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, 86, 97–106.
- Lattimore, F., Lattimore, T., & Reid, M. D. (2016). Causal bandits: Learning good interventions via causal inference. *Advances in Neural Information Processing Systems*, 29, 1181–1189.
- Li, C., Gupta, S., Rana, S., Nguyen, V., Venkatesh, S., & Shilton, A. (2017). High dimensional Bayesian optimization using dropout. In: *Proceedings of the 26th international joint conference on artificial intelligence*, pp. 2096–2102.
- Li, H., Zhao, B., & Fuxman, A. (2014). The wisdom of minority: discovering and targeting the right group of workers for crowdsourcing. In: *23rd international world wide web conference*, pp. 165–176.
- Luong, P., Nguyen, D., Gupta, S., Rana, S., & Venkatesh, S. (2020). Bayesian optimization with missing inputs. In: *The European conference on machine learning and principles and practice of knowledge discovery in databases*, pp. 691–706.
- Massart, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *The Annals of Probability*, 18(3), 1269–1283.
- Mockus, J., Tiesis, V., & Zilinskas, A. (1978). The application of Bayesian methods for seeking the extremum. *Towards global optimization*, 2(117–129), 2.
- Mutny, M., & Krause, A. (2018). Efficient high dimensional Bayesian optimization with additivity and quadrature Fourier features. *Advances in Neural Information Processing Systems*, 31, 9019–9030.
- Nguyen, T. T., Gupta, S., Ha, H., Rana, S., & Venkatesh, S. (2020). Distributionally robust Bayesian quadrature optimization. In: *The 23rd international conference on artificial intelligence and statistics*, pp. 1921–1931.
- Oliveira, R., Ott, L., & Ramos, F. (2019). Bayesian optimisation under uncertain inputs. In: *The 22nd international conference on artificial intelligence and statistics*, pp. 1177–1184.

- Picheny, V., Wagner, T., & Ginsbourger, D. (2013). A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3), 607–626.
- Rahimi, A., & Recht, B. (2007). Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20, 1177–1184.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press.
- Russo, D., & Roy, B. V. (2014). Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4), 1221–1243.
- Sessa, P. G., Bogunovic, I., Kamgarpour, M., & Krause, A. (2019). No-regret learning in unknown games with correlated payoffs. *Advances in Neural Information Processing Systems*, 32, 13602–13611.
- Shahriari, B., Bouchard-Cote, A., & Freitas, N. (2016a). Unbounded Bayesian optimization via regularization. In: *Proceedings of the 19th international conference on artificial intelligence and statistics*, pp. 1168–1176.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & de Freitas, N. (2016b). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1), 148–175.
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25, 2960–2968.
- Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008). Cheap and fast - but is it good? evaluating non-expert annotations for natural language tasks. In: *Proceedings of the 2008 conference on empirical methods in natural language processing*, pp. 254–263.
- Srinivas, N., Krause, A., Kakade, S. M., & Seeger, M. W. (2012). Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5), 3250–3265.
- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4), 285–294.
- Tran-Thanh, L., Chapman, A. C., Rogers, A., & Jennings, N. R. (2012). Knapsack based optimal policies for budget-limited multi-armed bandits. In: *Proceedings of the 26th AAAI Conference on Artificial Intelligence*.
- Williams, B. J., Santner, T. J., & Notz, W. I. (2000). Sequential design of computer experiments to minimize integrated response functions. *Statistica Sinica*, 10(4), 1133–1152.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.