Electronic Journal of Statistics
Vol. 12 (2018) 3829–3874
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1497

Gaussian process bandits with adaptive discretization

Shubhanshu Shekhar[1] **and Tara Javidi**[1]

[1] *Department of Electrical and Computer Engineering*
University of California, San Diego.
e-mail: shshekha@eng.ucsd.edu; tjavidi@eng.ucsd.edu

Abstract: In this paper, the problem of maximizing a black-box function $f : \mathcal{X} \to \mathbb{R}$ is studied in the Bayesian framework with a Gaussian Process prior. In particular, a new algorithm for this problem is proposed, and high probability bounds on its simple and cumulative regret are established. The query point selection rule in most existing methods involves an exhaustive search over an increasingly fine sequence of uniform discretizations of $\mathcal{X}$. The proposed algorithm, in contrast, adaptively refines $\mathcal{X}$ which leads to a lower computational complexity, particularly when $\mathcal{X}$ is a subset of a high dimensional Euclidean space. In addition to the computational gains, sufficient conditions are identified under which the regret bounds of the new algorithm improve upon the known results. Finally, an extension of the algorithm to the case of contextual bandits is proposed, and high probability bounds on the contextual regret are presented.

Keywords and phrases: Gaussian processes, bandits, Bayesian optimization.

Received January 2018.

Contents

1. Introduction

We consider the problem of maximizing a function $f : \mathcal{X} \to \mathbb{R}$ from its noisy observations of the form

$$y_t = f(x_t) + \eta_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

where η_t is the observation noise at time t. We work in the Bayesian setting, assuming that the function f is a sample from a zero mean Gaussian Process (GP) indexed by the space $\mathcal{X}$, and η_t for $t \geq 1$ are i.i.d. $N(0, \sigma^2)$ Gaussian random variables. We further assume that the function f is expensive to evaluate, and we are allocated a budget of n function evaluations.

This problem can be thought of as an extension of the multi-armed bandit (MAB) problem to the case of infinite (possibly uncountable) arms indexed by the set $\mathcal{X}$. The goal is to design a strategy of sequentially selecting query points $x_t \in \mathcal{X}$ based on the past observations $\{(x_i, y_i); 1 \leq i \leq t-1\}$ and the prior on f. As in the case of MAB with finitely many arms, the performance of any query point selection strategy is usually measured by the cumulative regret $\mathcal{R}_n$:

$$\mathcal{R}_n = \sum_{t=1}^{n} f(x^*) - f(x_t). \tag{1.2}$$

An alternative measure of performance is the *simple regret* $\mathcal{S}_n$ which is used in the the pure exploration problem:

$$\mathcal{S}_n = f(x^*) - f(x(n)), \tag{1.3}$$

where $x(n)$ is the point recommended by the algorithm after n noisy function evaluations.

We note here that cumulative regret is a more stringent measure of performance than the simple regret, as it forces the agent to address the *exploration-exploitation* dilemma.

1.1. Prior work

Optimizing a black-box function from its noisy observations is an active area of research with a large body of literature. Here, we review existing methods which take a Bayesian approach with GP prior to this problem, and have provable guarantees on their performance.

Srinivas et al. (2012) formulated the task of black-box function optimization as a MAB problem and proposed the GP-UCB algorithm which is a modification of the Upper Confidence Bound (UCB) strategy widely used in bandit literature. The algorithm constructs high probability UCBs on the function values using the GP posterior and selects the evaluation points by maximizing the UCB over $\mathcal{X}$. For finite search spaces $\mathcal{X}$ they showed that the GP-UCB algorithm admits a high probability upper bound on the cumulative regret of the form:

$$\mathcal{R}_n \leq \mathcal{O}(\sqrt{n \log(n) \gamma_n}), \tag{1.4}$$

where γ_n is the *maximum information gain* with n evaluations. We will refer to cumulative regret bounds of this form as *information-type* regret bounds in this paper. In addition, to make the dependence on n explicit, Srinivas et al. (2012) further derived bounds on the term γ_n for some commonly used kernels. Finally, they presented an extension of the GP-UCB algorithm to the case of continuous $\mathcal{X}$ by applying it on a sequence of increasingly fine uniform discretizations of $\mathcal{X}$.

Follow up works to Srinivas et al. (2012) have extended the GP-UCB algorithm in several ways. Contal and Vayatis (2016) proposed a method of constructing a sequence of uniform discretizations with tight control over the approximation error, which allowed the extension of the GP-UCB algorithm to arbitrary compact metric spaces $\mathcal{X}$. Desautels et al. (2014) and Contal et al. (2013) considered the GP bandits problem with the additional assumption that the evaluations can be performed in parallel. Desautels et al. (2014) proposed the GP-BUCB algorithm which selects the points in a batch sequentially by maximizing a variant of the UCB, which is computed by keeping the mean function fixed and only updating the posterior variance. Contal et al. (2013) proposed the GP-UCB-PE which uses the UCB function for selecting the first point of a batch, and then proceeds in a greedy manner selecting the remaining points by maximizing the posterior variance. Krause and Ong (2011) proposed and analyzed the CGP-UCB algorithm for the contextual GP bandits problem, where the mean reward function corresponding to context-action pairs is modeled as a sample from a GP on the context-action product space. Kandasamy et al. (2016) considered a multi-fidelity version of the GP bandits problem in which

they assumed the availability of a sequence of approximations of the true function f with increasing accuracies which were cheaper to evaluate. They proposed an extension of GP-UCB called the MF-GP-UCB and derived information-type bounds on its cumulative regret.

Wang et al. (2016) proposed the GP-EST algorithm which looks at the optimization problem through the lens of estimation. In particular, the algorithm constructs an estimate of the maximum function value $f(x^*)$, and then selects a point for evaluation which has the largest probability of attaining this value. Russo and Van Roy (2014) analyzed the performance of the Thompson Sampling algorithm to a large class of problems, including the GP bandits problem. Thompson Sampling is a randomized strategy in which query points are sampled according to the posterior distribution on x^*. Since computing the posterior on x^* may be complicated, in practice, the query points are selected in the following two step procedure: first, a sample $\tilde{f}_t$ of the unknown function f is generated, and then the query point x_t is chosen by maximizing $\tilde{f}_t$ over $\mathcal{X}$. For the case of continuous $\mathcal{X}$, the function samples are generated over uniform discretizations $\mathcal{X}_t$ of $\mathcal{X}$. By deriving a relation between the expected regret of Thompson Sampling and UCB strategies, Russo and Van Roy (2014) obtained information-type bounds on the expected cumulative regret of the Thompson Sampling algorithm for GP bandits.

As observed in (Bubeck et al., 2011a), bounding the cumulative regret automatically gives us a bound on the expected simple regret by employing a randomized point recommendation strategy. Additionally, for the pure exploration setting, several algorithms specifically geared towards minimizing $\mathcal{S}_n$, such as *Expected Improvement (GP-EI)*, *Probability of Improvement(GP-PI)*, *Entropy Search* and *Bayesian Multi-Scale Optimistic Optimization (BaMSOO)* have been proposed (see (Shahriari et al., 2016) for a recent survey). Bogunovic et al. (2016b) considered the BO and Level Set Estimation problems in a unified manner and proposed the *Truncated Variance Reduction (TRUVAR)* algorithm which selects evaluation points greedily to obtain the largest reduction in the sum of truncated variances of the potential maximizers. The performance of all these algorithms have been empirically studied over various synthetic as well as real-world datasets. Furthermore, theoretical guarantees are also known for GP-EI (Bull, 2011) (in non-Bayesian setting) and BaMSOO(Wang et al., 2014) with noiseless observations, and for TRUVAR (Bogunovic et al., 2016b) with noisy observations and non-uniform cost of evaluations.

All the algorithms above, with the exception of BaMSOO, require solving an *auxiliary optimization* problem in each round t for selecting the query point x_t. The objective function of this auxiliary optimization problem is usually non-convex and multi-modal and hence requires an exhaustive search over an increasingly fine sequence of uniform discretizations to guarantee that a close approximation of the true optimum is found (Srinivas et al., 2012; Contal and Vayatis, 2016). The size of these uniform discretizations increases exponentially with the dimension of $\mathcal{X}$. This is because these discretizations are chosen off-line and do not depend on the function evaluations made up to round t. In contrast, BaMSOO adaptively constructs discretizations by locally refining the regions of

$\mathcal{X}$ in which f is more likely to take higher values based on the observations. As a result, the size of the discretizations under BaMSOO are independent of the dimension of $\mathcal{X}$ which leads to significantly lower computational costs when $\mathcal{X}$ is high dimensional. Our work is strongly motivated by this aspect of BaMSOO to provide the first algorithm for GP bandits with noisy observations whose computational complexity has a linear dependence on the dimension of the search space $\mathcal{X}$.

1.2. Our contributions

In this paper, we address two issues with existing approaches to the GP bandits problem:

1. As discussed above, all the existing GP bandit algorithms which minimize the cumulative regret require solving an auxiliary optimization problem over the entire search space for selecting a query point which may be computationally infeasible, and thus practical implementations resort to various approximation techniques which do not come with theoretical guarantees.
2. Furthermore, by constructing specific Gaussian Processes in Section 1.3, we show that the information-type regret bounds can be too pessimistic; thus motivating the need for designing algorithms that admit alternative analysis techniques.

To tackle these two problems, we design algorithms for GP bandits which utilize ideas from existing works in the Lipschitz function optimization literature, such as (Bubeck et al., 2011b; Munos, 2011; Munos et al., 2014; Kleinberg et al., 2013). More specifically, our main contributions are as follows:

- We first present an algorithm for GP bandits which employs a tree of partitions of the search space $\mathcal{X}$ to adaptively refine it based on observations. We show that because of the adaptive discretization, when $\mathcal{X} \subset \mathbb{R}^D$ and D is large, our algorithm has significantly less computational complexity than algorithms requiring auxiliary optimization.
- We obtain high probability bounds on the cumulative regret of our algorithm which are always as good as, and in some cases strictly better than, the existing regret bounds. In particular, we obtain the first explicit sublinear regret bounds for the GP with exponential kernel (Ornstein-Uhlenbeck process) and also identify sufficient conditions under which our bounds improve upon the current ones for Matérn family of kernels.
- We also derive high probability bounds on the simple regret for our algorithm. To the best of our knowledge, BaMSOO (Wang et al., 2014) is the only adaptive[1] algorithm for the black-box optimization problem in the Bayesian setting, for which theoretical guarantees on simple regret

[1] we use the term *adaptive* to refer to algorithms which adaptively discretize the search space $\mathcal{X}$ based on earlier observations

are known. Our algorithm matches BaMSOO's performance with the additional advantages that it requires fewer assumptions on the covariance functions and can work with noisy observations.

- We also study two extensions of our algorithm. First, we present a Bayesian Zooming algorithm based on (Kleinberg et al., 2013; Slivkins, 2014) and obtain theoretical guarantees on its regret performance. This algorithm assumes a *covering oracle* access to the metric space $\mathcal{X}$ instead of requiring a hierarchical tree of partitions of $\mathcal{X}$. We then extend our algorithm for GP bandits to the contextual GP bandits and obtain bounds on the contextual regret.
- Finally, our algorithms and the theoretical bounds rely on a set of technical results about Gaussian Process which may be of independent interest. We provide these results and discuss their implications in Section 6.

1.3. Toy examples

As mentioned earlier, our cumulative regret bounds for Matérn kernels improve upon the known information type bounds for GP bandits. In this section, we attempt to provide some intuition for this result. In particular, we construct two toy examples which serve to highlight a potential drawback of the information type regret bounds for GP bandit problems shown in (1.4).

The information-type regret bounds (1.4) depend on the *maximum information gain* γ_n which is defined as:

$$\gamma_n = \sup_{x[1:n]\in\mathcal{X}^n} I(f; y_{x[1:n]}), \tag{1.5}$$

Here $I(f; y_{x[1:n]})$ is the mutual information between the unknown function f and vector of observations $y_{x[1:n]}$ corresponding to the n query points $x[1:n]$. This term depends on the covariance function[2] of the Gaussian Process (GP), and upper bounds on γ_n for many commonly used GPs are given in (Srinivas et al., 2012). We note that since our aim is to gather information about a maximizer x^* of f, and not necessarily about the behavior of f over the entire space $\mathcal{X}$, information-type regret bounds can be quite loose. We present two examples which have been specifically constructed to illustrate the scenarios where the regret bounds implied by (1.4) are very pessimistic. Both examples utilize the fact that the maximum information gain (γ_n) can be large if the Gaussian Process has many independent components, even when the maximizer may be simple to learn.

For our first example, we construct a GP whose samples have simple structure around the maximum despite the highly complex structure away from the maximizer. More specifically, we begin by dividing the interval $[0, 1]$ into three equal subintervals. Over the second and third intervals, the GP sample varies smoothly as scaled and shifted versions of a smooth function $\varphi(\cdot)$, modulated by

[2] we will use the terms *covariance functions* and *kernels* interchangeably

a Standard Normal random variable X_1. The first subinterval is further divided into three parts, and this process continues infinitely.

Example 1. *Suppose* $\mathcal{X} = [0,1]$ *and let us define a GP* $= \{f(x)|x \in \mathcal{X}\}$ *as follows:*

$$f(x) = \sum_{i=1}^{\infty} a_i X_i \left(\varphi\left(\frac{x}{b_i} - 1\right) - \varphi\left(\frac{x}{b_i} - 2\right)\right), \tag{1.6}$$

where $(a_i)_{i\geq 1}$ *is a non-increasing positive sequence,* $b_i = 3^{-i}$ *for* $i \geq 1$*, and* $(X_i)_{i=1}^{\infty}$ *is a sequence of independent Standard Normal random variables. The function* φ *is defined as* $\varphi(x) = \big(1 - 4(x-1/2)^2\big)\mathbf{1}_{[0,1]}(x)$.

For this GP, we can claim the following (details in Appendix A.1):

- *For the choice of* a_i *described in Appendix A.1, we have* $\gamma_n = \Omega\big(\frac{n\sigma^2}{\log(n)}\big)$*, which means that the information-type bound (1.4) on the cumulative regret is linear in* n.
- *On the other hand, if* $a_1 >> a_i$ *for* $i \geq 2$*, then the true maximizer* $x^* \in \{1/2, 5/6\}$ *with high probability, and it can be identified with just one function evaluation implying a constant cumulative regret,* $\mathcal{R}_n = \mathcal{O}(1)$.

For our second example, we construct a GP in which the search space is partitioned at different scales, and statistically equivalent components are assigned to the sets of a given partition. This process is repeated with increasingly finer partitions, and we show that for certain choice of parameters, each observation of the GP sample results in diminishing the region of uncertainty associated with x^* by a constant factor. However, the information-type bound again is dominated by the information obtained from the large number of independent components of the GP and gives a linear upper bound on the cumulative regret.

Example 2. *We again take* $\mathcal{X} = [0,1]$ *and let* φ_1 *denote the following function*

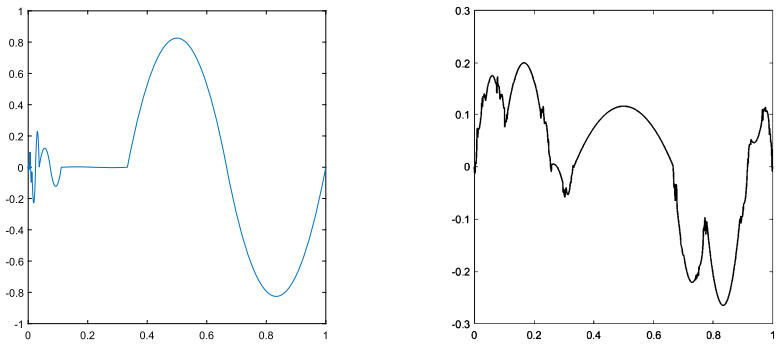
$$\varphi_1(x) = \begin{cases} \varphi(3x) & \text{if } x \in [0, 1/3) \\ \varphi(3x-1) & \text{if } x \in [1/3, 2/3) \\ -\varphi(3x-2) & \text{if } x \in [2/3, 1] \end{cases}$$

where φ *is the function used in Example 1. Let us now define a GP* $= \{f(x)|x \in \mathcal{X}\}$ *recursively as follows:*

$$\begin{aligned} f_i(x) &= a_i X_i \varphi_1(x) + f_{i+1}(3x) + f_{i+1}(3(x-2/3)) \quad \text{for } i \geq 2 \\ f(x) &= a_1 X_1 \varphi_1(x) + f_2(3x) + f_2(3(x-2/3)). \end{aligned} \tag{1.7}$$

As before $(a_i)_{i\geq 1}$ *is a decreasing sequence of positive real numbers, and* $(X_i)_{i\geq 1}$ *are i.i.d. Standard Normal random variables. For this example, we can claim the following:*

- *If the noise variance* σ^2 *is small enough, we have* $\gamma_n = \Omega\big(n\big)$ *which implies a linear in* n *information-type bound on cumulative regret.*



(a) Example 1 (b) Example 2

FIG 1. *An instance of the sample paths of the two toy Gaussian Processes.*

- *With the choice of parameters* $(a_i)_{i\geq 1}$ *described in Appendix A.2, we can select the evaluation points in such a way that with high probability after every observation, the size of the region containing* x^* *shrinks by a factor of* 3*, which in turn implies that the cumulative regret satisfies* $\mathcal{R}_n = \mathcal{O}(\log n)$.

Both our examples have been specifically crafted to highlight scenarios in which the information type upper bounds given in (1.4) may not reflect the actual performance of the algorithms due to its dependence on the term γ_n. In Section 4.2 we further strengthen this observation by showing that the information-type regret bounds are loose for a practically relevant class of Gaussian Processes.

The rest of the paper is organized as follows: In Section 2 we introduce the required definitions and present some background for the problem. We then describe our algorithm for GP bandits and analyze its regret in Section 3. We discuss the behavior of our algorithm in some specific problem instances in Section 4. In Section 5 we study two extensions of our approach and analyze their performance. Finally, Section 6 contains some technical results which were used in designing our algorithms.

2. Preliminaries

In this section we recall some definitions required for stating the results, and fix the notations used.

Definition 1. *A Gaussian Process is a collection of random variables* $\{f(x); x \in \mathcal{X}\}$ *which satisfy the property that* $(f(x_1), f(x_2), \ldots, f(x_m))$ *is a jointly Gaussian random variable for all* $\{x_1, x_2, \ldots, x_m\} \subset \mathcal{X}$ *and* $m \in \mathbb{N}$. *A Gaussian Process is completely specified by its mean function* $\mu(x) = \mathbb{E}[f(x)]$ *and its covariance function* $K(x_1, x_2) = \mathbb{E}[(f(x_1) - \mu(x_1))(f(x_2) - \mu(x_2))]$.

For a comprehensive discussion about Gaussian Processes and their applications in machine learning, see (Rasmussen and Williams, 2006).

Remark 1. Any zero mean Gaussian Process with covariance function K induces a metric[3] d on its index set $\mathcal{X}$, defined as

$$d(x_1, x_2) = \mathbb{E}[(f(x_1) - f(x_2))^2]^{1/2} \tag{2.1}$$

$$= \big(K(x_1, x_1) + K(x_2, x_2) - 2K(x_1, x_2)\big)^{1/2}. \tag{2.2}$$

which gives us the following useful tail bound for any $x_1, x_2 \in \mathcal{X}$ and $a \geq 0$:

$$Pr(|f(x_1) - f(x_2)| \geq a) \leq 2 \exp\left(- \frac{a^2}{2d(x_1, x_2)} \right). \tag{2.3}$$

As mentioned earlier, in this paper we assume that the black-box function f is a sample from a zero mean Gaussian process with known covariance function $K(\cdot, \cdot)$. Suppose $\mathcal{D}_t = \{(x_i, y_i); 1 \leq i \leq t\}$ represents the data collected by agent after t function evaluations, and let $\boldsymbol{x}_{[\mathbf{1:t}]}$ and $\boldsymbol{y}_{\boldsymbol{x}[\mathbf{1:t}]}$ denote the vector of evaluation points and the observed values respectively. Then the posterior distribution distribution of $f(x)$ for any $x \in \mathcal{X}$ is a univariate Gaussian with mean $\mu_t(x)$ and variance $\sigma_t^2(x)$ defined as:

$$\mu_t(x) = K(x, \boldsymbol{x}_{[\mathbf{1:t}]}) \big(K(\boldsymbol{x}_{[\mathbf{1:t}]}, \boldsymbol{x}_{[\mathbf{1:t}]}) + \sigma^2 \mathcal{I}_t\big)^{-1} \boldsymbol{y}_{\boldsymbol{x}_{[1:t]}}$$

$$\sigma_t^2(x) = K(x, x) + K(x, \boldsymbol{x}_{[\mathbf{1:t}]}) \big(K(\boldsymbol{x}_{[\mathbf{1:t}]}, \boldsymbol{x}_{[\mathbf{1:t}]}) + \sigma^2 \mathcal{I}_t\big)^{-1} K(\boldsymbol{x}_{[\mathbf{1:t}]}, x).$$

In the above display, $K(x, \boldsymbol{x}_{[\mathbf{1:t}]})$ is a $t \times 1$ row vector with i^{th} entry $K(x, x_i)$, $K(\boldsymbol{x}_{[\mathbf{1:t}]}, x)$ is the transpose of $K(x, \boldsymbol{x}_{[\mathbf{1:t}]})$, $K(\boldsymbol{x}_{[\mathbf{1:t}]}, \boldsymbol{x}_{[\mathbf{1:t}]})$ is a $t \times t$ matrix with $(i, j)^{th}$ entry $K(x_i, x_j)$ and $\mathcal{I}_t$ is the $t \times t$ identity matrix.

Next, we introduce some properties of any metric space (X, l) which will be used later on.

Definition 2. *Suppose $\mathcal{X}$ is a non-empty set and l is a metric on $\mathcal{X}$. Then we have the following:*

- *A subset $\mathcal{X}_1$ of $\mathcal{X}$ is called an r-covering set of $\mathcal{X}$ if for any $x \in \mathcal{X}$, we have $l(x, \mathcal{X}_1) \leq r$ where $l(x, \mathcal{X}_1) \coloneqq \inf\{l(x, y) : y \in \mathcal{X}_1\}$. The cardinality of the smallest such $\mathcal{X}_1$ is called the r-covering number of $\mathcal{X}$ with respect to l, denoted by $N(\mathcal{X}, r, l)$.*
- *The metric dimension of a space $\mathcal{X}$ with associated metric l, denoted by D_1, is then defined as follows:*

$$D_1 \coloneqq \inf\{a \geq 0 \mid \exists C \geq 0, \forall r > 0 : \log(N(\mathcal{X}, r, l)) \leq C - a \log(r)\}$$

Remark 2. If a metric space $(\mathcal{X}, l)$ has a finite metric dimension D_1, then the above definition implies that for any $D > D_1$, we can find a constant $C < \infty$ depending on D, such that for all $r > 0$, we have $N(\mathcal{X}, r, l) \leq C r^{-D}$. Furthermore, for the special case when $\mathcal{X}$ is a bounded subset of $\mathbb{R}^D$ with l being the

[3] d is a pseudo-metric for general K, and is a metric only for non-degenerate K

Euclidean metric, the metric dimension D_1 is equal to D (van Handel, 2014, page 125) and for any $r > 0$, we have $N(\mathcal{X}, r, l) \leq Cr^{-D}$ where the constant C depends on the dimension D.

The metric dimension D_1 gives us a notion of dimensionality intrinsic to the metric space $(\mathcal{X}, l)$. We now present a function specific measure of dimensionality of $(\mathcal{X}, l)$.

Definition 3. *Suppose $\mathcal{X}$ is a non-empty set, l is a metric on $\mathcal{X}$ and f is a function from $\mathcal{X}$ to $\mathbb{R}$. Then*

- *A subset $\mathcal{X}_2$ of $\mathcal{X}$ is called an r-separated set of $\mathcal{X}$ if for any $x_1, x_2 \in \mathcal{X}_2$ we have $l(x_1, x_2) \geq r$. The cardinality of the largest such set $\mathcal{X}_2$ is called the r-packing number of $\mathcal{X}$ with respect to l, and is denoted by $M(\mathcal{X}, r, l)$.*
- *For any $r > 0$ and $\zeta : \mathbb{R}^+ \to \mathbb{R}^+$, consider the $\zeta(r)$ near-optimal set $\mathcal{X}_{\zeta(r)} := \{x \in \mathcal{X} : f(x) \geq f(x^*) - \zeta(r)\}$ and its r-packing number $M(\mathcal{X}_{\zeta(r)}, r, l)$. Then we define the (Δ_0, ζ)-near-optimality dimension ($D^f(\Delta_0, \zeta)$) associated with $(\mathcal{X}, l)$ and the function f as*

$$D^f(\Delta_0, \zeta) := \inf\{a \geq 0 \mid \exists C > 0, \forall 0 < r \leq \Delta_0 : \log(M(\mathcal{X}_{\zeta(r)}, r, l)) \leq C - a\log(r)\} \tag{2.4}$$

Our definition of the near-optimality dimension is based on similar definitions used in existing works in literature such as (Bubeck et al., 2011b; Munos, 2011; Valko et al., 2013).

Remark 3. We note that for any $(\mathcal{X}, l)$ with finite metric dimension D_1, by using volume arguments (van Handel, 2014, Lemma 5.13) we can show that $D^f(\Delta_0, \zeta) \leq D_1$. An example (Bubeck et al., 2011b, Example 3) where this inequality is strict is the following: consider $\mathcal{X} = [-1, 1]$, $l(x_1, x_2) = |x_1 - x_2|$ and $f(x) = 1 - |x|^a$, and $\zeta(r) = r^b$ for some $0 < b \leq a$. Then $D^f(1, \zeta) = 1 - b/a \leq 1 = D$, and in particular $D^f(1, \zeta) = 0$ for $a = b$.

Definition 4. *We will call a compact metric space $(\mathcal{X}, l)$ well-behaved if there exists a sequence of subsets $(\mathcal{X}_h)_{h \geq 0}$ of $\mathcal{X}$ satisfying the following properties:*

P1 *Each subset $\mathcal{X}_h$ has N^h elements for some $N > 1$, i.e. $\mathcal{X}_h = \{x_{h,i}, 1 \leq i \leq N^h\}$, and to each element $x_{h,i} \in \mathcal{X}_h$ is associated a* cell *$\mathcal{X}_{h,i} = \{x \in \mathcal{X} : l(x, x_{h,i}) \leq l(x, x_{h,j})$ for all $j \neq i\}$.*

P2 *For all $h \geq 0$ and $0 \leq i \leq N^h$, we have*

$$\mathcal{X}_{h,i} = \cup_{j=N(i-1)+1}^{Ni} \mathcal{X}_{h+1,j}. \tag{2.5}$$

The nodes $x_{h+1,j}$ for $N(i-1)+1 \leq j \leq Ni$ are called the children *of $x_{h,i}$, which in turn is referred to as the* parent.

P3 *We assume that the cells have geometrically decaying radii, i.e., there exists $0 < \rho < 1$ and $0 < v_2 \leq 1 \leq v_1$ such that we have*

$$B(x_{h,i}, v_2\rho^h, l) \subset \mathcal{X}_{h,i} \subset B(x_{h,i}, v_1\rho^h, l). \tag{2.6}$$

From $P1$ we can see that the cells $\{\mathcal{X}_{h,i}; 1 \leq i \leq N^h\}$ partition the space $\mathcal{X}$ for every $h \geq 0$, while $P2$ implies that we get an increasingly fine sequence of partitions with increasing h. Finally $P3$ imposes the condition that for any h, the points $x_{h,i}$ are evenly spread out in the space $\mathcal{X}$. The subsets $(\mathcal{X}_h)_{h\geq 0}$ satisfying these properties are said to form a tree of partitions (Munos et al., 2014; Bubeck et al., 2011b).

Remark 4. We note that if $\mathcal{X} = [a, b]^D \subset \mathbb{R}^D$, and l is any metric on $\mathcal{X}$, then $\mathcal{X}$ is well-behaved according to the above definition. The cells $\mathcal{X}_{h,i}$ in this case are D dimensional hyper-rectangles such that $\mathcal{X}_{h+1,j}$ for $1 \leq j \leq N^h$ can be constructed from $\mathcal{X}_{h,i}$ by dividing it along its longest edge into N equal parts.

3. Algorithm for GP bandits

In Section 3.1, we describe the general template followed by all the algorithms studied in this paper. We then introduce our tree based algorithm for GP bandits in Section 3.2 and obtain high probability bounds on its regret in Section 3.3.

3.1. General approach

At any time t, we maintain a discretization (i.e., a finite subset) of $\mathcal{X}$, denoted by $\mathcal{X}_t$. To each $x \in \mathcal{X}_t$, we have an associated *confidence region* denoted by $Reg_t(x)$, which is a non-empty set containing x such that for all t, we have $\mathcal{X} \subset \cup_{x\in\mathcal{X}_t} Reg_t(x)$. Furthermore, we also associate an *index* $Ind_t(x)$ with each $x \in \mathcal{X}_t$, which is a high probability upper bound on the maximum value of the function f in $Reg_t(x)$. The index $Ind_t(x)$ depends on three quantities: (a) the actual function value at x, (b) the amount of uncertainty in the function value at x, and (c) the amount of *variation* in the function value in $Reg_t(x)$, defined as $\sup_{z_1,z_2\in Reg_t(x)} |f(z_1) - f(z_2)|$. We proceed as follows:

- In each round, we select a candidate point x_t *optimistically* by maximizing $Ind_t(x)$ over $\mathcal{X}_t$.
- If the uncertainty in the function value at x_t is smaller than an upper bound on the variation of f in the confidence region, it means that we must *refine* our discretization in the confidence region associated with x_t.
- If, on the other hand, the uncertainty in the function value at x_t is larger than the variation of f in the associated confidence region, our algorithm evaluates the function at this point to reduce this uncertainty.

The second and third steps ensure that for a given $x \in \mathcal{X}_t$, the uncertainty in the value of $f(x)$ and the variation of f in $Reg_t(x)$ are comparable to each other. Combined with the optimistic point selection rule, these two steps favor a coarse discretization of the regions of $\mathcal{X}$ with small values of $f(x)$, and finer exploration in the near-optimal regions.

Both the algorithms for GP bandits studied in this paper follow the general outline described above. However, they differ from each other in the manner

TABLE 1
Table of symbols used in the paper

Symbol	Description	Introduced in
f	black-box function	Section 1
n	function evaluation budget	—"—
η_t, σ^2	observation noise distributed as $N(0, \sigma^2)$	—"—
$GP(0, K)$	prior on f with covariance function K	—"—
$\mathcal{K}$	class of covariance functions considered	Section 3.3.1
δ_K, α, C_K, g	parameters associated with $K \in \mathcal{K}$	—"—
$\mu_{t-1}, \sigma^2_{t-1}$	posterior mean and variance functions	
$\mathcal{R}_n$	Cumulative regret	Section 1,(1.2)
$\mathcal{S}_n$	Simple regret	Section 1,(1.3)
γ_n	Maximum Information gain	Section 1.3, (1.5)
$\mathcal{R}^c_n$	Contextual regret	Section 5.2, (5.7)
$\Delta(x)$	$f(x^*) - f(x)$	
$\Delta^c(x_\tau)$	$\sup_{x^a \in \mathcal{X}_a} f(x^c_\tau, x^a) - f(x^c_\tau, x^a_\tau)$	Section 5.2
$(\mathcal{X}, l)$	Compact search space $\mathcal{X}$ with metric l	
$B(x, r, l)$	l-ball with center x, and radius r	(2.6)
d	metric induced by $GP(0, K)$ on $\mathcal{X}$	Section 2, (2.1)
D_1	Metric dimension	—"—,Definition 2
$N(\mathcal{X}, r, l)$	Covering number	—"—, —"—
$M(\mathcal{X}, r, l)$	Packing number	—"—, Definition 3
$D^f(\Delta_0, \zeta)$	(Δ_0, ζ)-near-optimality dimension	—"—, —"—
$\tilde{D}, \tilde{D}_Z$	instances of $D^f(\Delta_0, \zeta)$	Remark 10, Remark 15
N, v_1, v_2, ρ	Parameters of the tree of partitions	Section 2, Definition 4
Parameters of Algorithm 1 and Algorithm 3		
$\mathcal{L}_t$	the set of leaf nodes	Section 3.2
$I_t(x_{h,i})$	Index used for point selection in Algorithm 1	—"—, (3.1)
$\bar{U}_t(x_{h,i})$	upper bound on $f(x_{h,i})$	—"—, (3.2)
$p(x_{h,i})$	parent node of $x_{h,i}$	
β_n	multiplicative factor for confidence intervals	Section 3.3.2, Claim 1
$h_{\max}$	maximum depth of the tree	—"—, (3.4)
$(V_h)_{h \geq 0}$	upper bound on maximum variation of f in a cell at level h	—"—, Claim 2
$\mathcal{L}^{rel}_t$	relevant leaf nodes	Section 5.2.1
$\bar{x}^{(t)}_{h,i}$		—"—
I^c_t	Index used for action selection in Algorithm 3	—"—, (5.9)
$\mathcal{X}_c$, $\mathcal{X}_a$	Context space and Action space	Section 5.2
Parameters of Algorithm 2		
A_t	Set of active points	Section 5.1
$r(x)$	radius associated with a point $x \in A_t$	—"—
r_k	$diam(\mathcal{X})2^{-k}$	—"—
$W(r_k)$	upper bound on variation of f in $B(x, r_k, l)$ for any $x \in \mathcal{X}$	Claim 7

in which the discretization $\mathcal{X}_t$, the confidence region $Reg_t(x)$ and the index $Ind_t(x)$ are constructed. In Section 3.2 we present an algorithm for GP bandits which uses a hierarchical partitioning scheme for locally refining the search space similar to (Munos et al., 2014; Bubeck et al., 2011b; Wang et al., 2014). Alternatively, the *covering oracle* based approach used by Slivkins (2014); Kleinberg

et al. (2013) can also be employed for refining the discretization, and we describe such an algorithm in Section 5.1. We also apply this approach to design an adaptive algorithm for the Contextual GP bandits problem in Section 5.2.

3.2. Tree based algorithm

We now describe our tree based algorithm for GP bandits and derive high probability bounds on its regret. Our algorithm is motivated by several tree based methods that have been proposed for function optimization under Lipschitz-like assumptions, such as (Bubeck et al., 2011b; Munos, 2011; Munos et al., 2014). Assuming that the metric space $(\mathcal{X}, l)$ is *well behaved*, i.e., we have a sequence of subsets $(\mathcal{X}_h)_{h\geq 0}$ whose associated cells form a tree of partitions of $\mathcal{X}$, we proceed as follows:

- In every round t, the algorithm maintains an active set of leaf nodes denoted by $\mathcal{L}_t$, such that the cells of the nodes in $\mathcal{L}_t$ partition $\mathcal{X}$. This active set is initialized to $\mathcal{L}_0 = \{x_{0,1}\}$ with the associated cell $\mathcal{X}_{0,1} = \mathcal{X}$.
- The algorithm selects a node from $\mathcal{L}_t$ by maximizing an index I_t over its elements. The term $x_{h,i}$ represents the i^{th} node in the tree at depth h, with $1 \leq i \leq N^h$. The index $I_t(x_{h,i})$ is an upper confidence bound (UCB) on the maximum function value in cell $\mathcal{X}_{h,i}$ and is defined as

$$I_t(x_{h,i}) = \bar{U}_t(x_{h,i}) + V_h. \tag{3.1}$$

  The term $\bar{U}_t(x_{h,i})$ in the above equation is a high probability upper bound on the function value at $x_{h,i}$ and is defined as

$$\begin{aligned}\bar{U}_t(x_{h,i}) = \min \big(&\mu_{t-1}(x_{h,i}) + \beta_n \sigma_{t-1}(x_{h,i}), \mu_{t-1}(p(x_{h,i})) \\ &+ \beta_n \sigma_{t-1}(p(x_{h,i})) + V_{h-1}\big)\end{aligned} \tag{3.2}$$

  where $p(x_{h,i})$ is the parent node of $x_{h,i}$. For any $h \geq 0$, the term V_h is an upper bound (with high probability) on the maximum function variation in any cell $\mathcal{X}_{h,i}$ at level h, i.e., $\sup_{z_1,z_2\in\mathcal{X}_{h,i}} |f(z_1) - f(z_2)| \leq V_h$. Thus, we see that $\bar{U}_t(x_{h,i})$ computes an upper bound on the value of $f(x_{h,i})$ in two ways and takes their minimum, while adding V_h to it gives us an upper bound on the maximum function value in the cell $\mathcal{X}_{h,i}$.
- Having chosen the point (x_{h_t,i_t}) according to the selection rule (Line 2 of Algorithm 1) we take one of the following two actions:
  - *Refine:* If $\beta_n \sigma_{t-1}(x_{h_t,i_t}) \leq V_h$, then the node x_{h_t,i_t} is expanded, i.e., the N children nodes $\{x_{h_t+1,j} : N(i_t - 1) + 1 \leq j \leq N i_t\}$ of the node x_{h_t,i_t} are added to the set of leaves, and x_{h_t,i_t} is removed from it. (Lines 4-5 of Algorithm 1)
  - *Evaluate:* Otherwise, then the function is evaluated at the point x_{h_t,i_t}, i.e., we observe the noisy function value $y_t = f(x_{h_t,i_t}) + \eta_t$ and update the posterior distribution of f. (Lines 7-9 of Algorithm 1)

We note that this algorithm follows the general template of Section 3.1, with $\mathcal{L}_t$ representing the discretization $\mathcal{X}_t$, $I_t(\cdot)$ playing the role of $Ind_t(\cdot)$ and the cell $\mathcal{X}_{h,i}$ associated with a point $x_{h,i}$ playing the role of $Reg_t(x_{h,i})$.

The steps of the algorithm are shown as a pseudo-code in Algorithm 1. The exact expression of the terms β_n, V_h and h_{max} will be described in Section 3.3. The algorithm maintains two counters, t which counts the total number of function evaluations and refinements, and n_e which keeps track of the number of function evaluations. The algorithm stops after n function evaluations, and recommends a point from one of the deepest expanded cells (for minimizing $\mathcal{S}_n$). The second condition on Line 3 of Algorithm 1 is added to prevent the (unlikely) scenario in which the algorithm keeps refining indefinitely without evaluating the function.

Algorithm 1: Tree based Algorithm for GP bandits

Input : $n > 0$, $(\mathcal{X}_h)_{h\geq 0}$, β_n, $(V_h)_{h\geq 0}$, $h_{\max}$, N

Initialize: $\mathcal{L}_0 = \{x_{0,1}\}$, $t = 1$, $n_e = 0$

1 **while** $n_e \leq n$ **do**
2 choose $x_{h_t,i_t} = \arg\max_{x_i \in \mathcal{L}_t} I_t(x_{h,i})$
3 **if** $\beta_n \sigma_{t-1}(x_{h_t,i_t}) \leq V_h$ *AND* $h_t \leq h_{\max}$ **then**
4 $\mathcal{L}_{t+1} = \mathcal{L}_t \setminus \{x_{h_t,i_t}\}$
5 $\mathcal{L}_{t+1} = \mathcal{L}_{t+1} \cup \{x_{h_t+1,j} | N(i_t - 1) + 1 \leq j \leq Ni_t\}$
6 **else**
7 $y_t = f(x_{h_t,i_t}) + \eta_t$
8 update posterior $\mu_t(x)$ and $\sigma_t(x)$
9 $n_e \leftarrow n_e + 1$
10 **end**
11 $t \leftarrow t + 1$
12 **end**

Output : $x(n)$: the deepest expanded node

Remark 5. The parameter β_n of Algorithm 1 requires the knowledge of the horizon or the budget n. However, we can use the well known *doubling trick*(Cesa-Bianchi and Lugosi, 2006, Section 2.3) to make our algorithm *anytime* without any change in the theoretical regret guarantees. The trick is to work in phases of exponentially increasing lengths, and applying the algorithm with known horizon (equal to the duration of the phase) in each phase.

3.3. Analysis of Algorithm 1

In this section, we first specify the assumptions on the covariance functions required for the theoretical analysis and then furnish the missing details of our tree based algorithm for GP bandits. Finally, we derive high probability bounds on the cumulative and simple regret for our algorithm.

3.3.1. Assumptions on the covariance functions

To analyze our algorithm, we will restrict our attention to a class of covariance functions, denoted by $\mathcal{K}$, such that for any $K \in \mathcal{K}$, we have:

A1 For any $x, y \in \mathcal{X}$, we have $d(x,y) \leq g(l(x,y))$ for some non-decreasing continuous function $g : \mathbb{R}^+ \to \mathbb{R}^+$, such that $g(0) = 0$. Recall that l is assumed to be any metric on the space $\mathcal{X}$, and d is the natural metric induced on $\mathcal{X}$ by the zero mean GP with covariance function K.

A2 Moreover, we require that there exists a $\delta_K > 0$ such that for all $r \leq \delta_K$, we have for constants $C_K > 0$ and $0 < \alpha \leq 1$ satisfying

$$g(r) \leq C_K r^{\alpha}. \tag{3.3}$$

Assumption $A2$ informally requires that at least for small distances, points which are close in the metric l are also close in d. These assumptions are satisfied by all the commonly used kernels such as squared exponential (SE), and the Matérn family of kernels. It also includes other kernels such as $K(r) = \max(0, 1-r)$ and the rational quadratic kernel $K(r) = (1 + c_1 r^2)^{-c_2}$ for some $c_1, c_2 > 0$.

Remark 6. We note that $\mathcal{K}$ is closed under finite addition and multiplication operations. This is an important property as in many practical applications, often more than one kernels are combined through addition or multiplication to provide more accurate models (Duvenaud, 2014, Chapter-2),(Rasmussen and Williams, 2006).

Remark 7. Assumption $A2$ implies that if $(\mathcal{X}, l)$ has a finite metric dimension D_1, then the metric space $(\mathcal{X}, d)$ (where d is defined in (2.1)) has a finite metric dimension $D_1' = D_1/\alpha$. This fact is used in Proposition 1 in Section 6.

3.3.2. Details of the algorithm

To complete the description of Algorithm 1, we need to specify the choice of the parameters $h_{\max}$, β_n, and $(V_h)_{h \geq 0}$.

First we observe that for all t, we have $|\mathcal{L}_t| \leq M(\mathcal{X}, 2v_2\rho^{h_{\max}}, l)$. This follows from the assumption $P3$ in Definition 4. As will be evident in the proof of Theorem 1, an appropriate choice of the parameter $h_{\max}$ is:

$$h_{\max} = \frac{\log n}{2\alpha \log(1/\rho)}\big(1 + 1/\alpha\big). \tag{3.4}$$

We begin by describing the form of the constant β_n.

Claim 1. *With $\beta_n = \mathcal{O}(\sqrt{\log(n) + u})$, the following event occurs with probability at least $1 - e^{-u}$ for any $u > 0$:*

$$\Omega_{u5} = \{\forall 1 \leq t \leq t_n, \forall x \in \mathcal{L}_t : |f(x) - \mu_{t-1}(x)| \leq \beta_n \sigma_{t-1}(x)\}, \tag{3.5}$$

where t_n is the (random) number of rounds required by the algorithm to complete n function evaluations.

Proof. The largest value that the random variable t_n can take is $h_{\max}n$, and for any $t \leq t_n$ we have $|\mathcal{L}_t| \leq M(\mathcal{X}, 2v_2\rho^{h_{max}}, l)$. Following the discussion in Remark 2 with $D = 2D_1$, there exists a constant $C < \infty$ such that $M(\mathcal{X}, 2v_2\rho^{h_{max}}, l) \leq N(\mathcal{X}, v_2\rho^{h_{max}}, l) \leq C(v_2\rho^{h_{max}})^{-2D_1}$.

Based on these two observations, we can claim the following:

$$\begin{aligned}
1 - Pr(\Omega_{u5}) =& Pr(\exists t \leq t_n, \exists x_{h,i} \in \mathcal{L}_t : |\mu_{t-1}(x_{h,i}) - f(x_{h,i})| > \beta_n \sigma_{t-1}(x_{h,i})) \\
\leq& \sum_{t=1}^{t_n} \sum_{x_{h,i} \in \mathcal{L}_t} Pr(|\mu_{t-1}(x_{h,i}) - f(x_{h,i})| > \beta_n \sigma_{t-1}(x_{h,i})) \\
\leq& \sum_{t=1}^{t_n} \sum_{x_{h,i} \in \mathcal{L}_t} 2e^{-\beta_n^2/2} \\
\leq& 2(h_{\max}n)(C\rho^{-2D_1 h_{\max}})e^{-\beta_n^2/2}
\end{aligned}$$

Finally, we get the required bound by selecting $\beta_n^2 = \mathcal{O}\big(u + 2\log(h_{\max}n) + 2D_1 h_{\max}\log(1/\rho)\big)$. □

Remark 8. The calculation of β_n above is based on the worst case assumption that $|\mathcal{L}_t| = M(\mathcal{X}, 2v_2\rho^{h_{\max}}, l)$. In the case of $\mathcal{X} \subset \mathbb{R}^D$ and for odd values of N, we can use a tighter bound $|\mathcal{L}_t| \leq nNh_{\max}$ which gives us $\beta_n = \mathcal{O}(\sqrt{u + \log(nh_{\max})})$ which allows us to consider larger values of h_{max}.

Next, we obtain the expressions for the parameters $(V_h)_{h\geq 0}$ as an immediate consequence of Corollary 1 in Section 6.

Claim 2. *Suppose the metric space $(\mathcal{X}, l)$ is well-behaved in the sense of Definition 4 with subsets $(\mathcal{X}_h)_{h\geq 0}$ and associated parameters v_1, v_2 and ρ. Let us define the event Ω_{u6} as*

$$\Omega_{u6} = \{\forall h \geq 0; \forall 1 \leq i \leq N^h : \sup_{x \in B(x_{h,i}, v_1\rho^h, l)} |f(x) - f(x_{h,i})| \leq V_h\}. \tag{3.6}$$

Then for the choice of V_h

$$V_h = 4g(v_1\rho^h)\Big(\sqrt{2u + C_4 + h\log N + 4D_1\log(1/g(v_1\rho^h))} + C_3\Big),$$

we have $Pr(\Omega_{u6}) \geq 1 - e^{-u}$ for any $u > 0$. Here C_3 and C_4 are the positive constants defined in Corollary 1 in Section 6.

Remark 9. We note that that for $h \geq h_0 \coloneqq \log(v_1/\delta_K)/\log(1/\rho)$, the ratio V_h/V_{h+1} can be upper bounded by $\rho^{-\alpha}$. Thus for all values of $h \geq 0$, we have the condition $V_h \leq N_1 V_{h+1}$, where the term N_1 is defined as

$$N_1 \coloneqq \max\{\max_{0\leq h\leq h_0-1} \frac{V_h}{V_{h+1}},\ \rho^{-\alpha}\}. \tag{3.7}$$

3.3.3. Regret bounds

Before presenting the regret bounds, we first characterize the sub-optimality as well as the number of times points are evaluated by Algorithm 1.

Lemma 1. *Under the events* Ω_{u5} *(3.5), and* Ω_{u6} *(3.6), the following statements are true:*

- *If at time* t *a point* x_{h_t,i_t} *is evaluated by the algorithm, then the sub-optimality of the selected point (denoted by* $\Delta(x_{h_t,i_t})$*) can be upper bounded using* V_{h_t} *as follows:*

$$\Delta(x_{h_t,i_t}) \coloneqq f(x^*) - f(x_{h_t,i_t}) \leq (4N_1 + 1)V_{h_t}. \tag{3.8}$$

- *Furthermore, if the evaluated point* x_{h_t,i_t} *satisfies the condition that* $h_t < h_{\max}$*, then we have another bound on* $\Delta(x_{h_t,i_t})$ *in terms of the posterior standard deviation:*

$$\Delta(x_{h_t,i_t}) \leq 3\beta_n \sigma_{t-1}(x_{h_t,i_t}). \tag{3.9}$$

- *A point* $x_{h,i}$*, with* $h < h_{\max}$*, may be evaluated no more than* q_h *times before it is expanded, where*

$$q_h = \frac{\sigma^2 \beta_n^2}{V_h^2}.$$

  Furthermore for h *large enough so that* $v_1\rho^h \leq \delta_K$*, we have*

$$q_h \leq \frac{\sigma^2 \beta_n^2}{g(v_1\rho^h)^2 C_3},$$

  using the assumptions on the covariance function K*.*

Proof. We recall that under the event Ω_{u5} we have $|f(x_{h,i}) - \mu_{t-1}(x_{h,i})| \leq \beta_n \sigma_{t-1}(x_{h,i})$ for all $x_{h,i} \in \mathcal{L}_t$ and for all $t \geq 1$. Furthermore, from the definition of event Ω_{u6}, we have the following for all $h \geq 0$ and $1 \leq i \leq N^h$:

$$\sup_{x_1,x_2 \in \mathcal{X}_{h,i}} |f(x_1) - f(x_2)| \leq V_h.$$

Using these two facts we can prove the first part of this lemma in the following way:

- Suppose at time t, the true maximizer x^* lies in the cell $\mathcal{X}_{h_t^*,i_t*}$ associated with the point $x_{h_t^*,i_t^*}$, and the algorithm selects and evaluates the point x_{h_t,i_t}. Then we have the following sequence of inequalities:

$$\begin{aligned} f(x^*) \leq I_t(x_{h_t^*,i_t^*}) \leq I_t(x_{h_t,i_t}) &= \bar{U}_t(x_{h_t,i_t}) + V_{h_t} \\ &\overset{(a)}{\leq} \mu_{t-1}(p(x_{h_t,i_t})) + \beta_n \sigma_{t-1}(p(x_{h_t,i_t})) + V_{h_t-1} + V_{h_t} \end{aligned}$$

$$\begin{aligned}
&\overset{(b)}{\leq} f(p(x_{h_t,i_t})) + 2\beta_n\sigma_{t-1}(p(x_{h_t,i_t})) + V_{h_t-1} + V_{h_t} \\
&\overset{(c)}{\leq} \left(f(p(x_{h_t,i_t})) + V_{h_t-1}\right) + 2V_{h_t-1} + V_{h_t} \\
&\overset{(d)}{\leq} f(x_{h_t,i_t}) + 4V_{h_t-1} + V_{h_t} \\
&\overset{(e)}{\leq} f(x_{h_t,i_t}) + (4N_1+1)V_{h_t}
\end{aligned}$$

The inequality (a) follows from the definition of $\bar{U}_t(x_{h_t,i_t})$, while (b) uses the fact that $f(p(x_{h_t,i_t})) \geq \mu_{t-1}(p(x_{h_t,i_t})) - \beta_n\sigma_{t-1}(p(x_{h_t,i_t})$ under event Ω_{u5}. For (c), we use the fact that $p(x_{h_t,i_t})$ must have been expanded which means $\beta_n\sigma_{t-1}(p(x_{h_t,i_t}))$ must be smaller than V_{h_t-1}. For inequality (d) we observe that x_{h_t,i_t} must lie in the cell associated with $p(x_{h_t,i_t})$ and then use the definition of V_{h_t-1}, while (e) follows from the property that $V_h \leq N_1 V_{h+1}$ (see Remark 9).

- For obtaining the bound in (3.9), we again use the definition of $\bar{U}_t(x_{h_t,i_t})$ to now upper bound it by the other term in its definition to get:

$$\begin{aligned}
f(x^*) \leq I_t(x_{h_t^*,i_t^*}) &\leq I_t(x_{h_t,i_t}) = \bar{U}_t(x_{h_t,i_t}) + V_{h_t} \\
&\leq \mu_{t-1}(x_{h_t,i_t}) + \beta_n\sigma_{t-1}(x_{h_t,i_t}) + V_{h_t} \\
&\leq f(x_{h_t,i_t}) + 2\beta_n\sigma_{t-1}(x_{h_t,i_t}) + V_{h_t} \\
&\overset{(f)}{\leq} f(x_{h_t,i_t}) + 3\beta_n\sigma_{t-1}(x_{h_t,i_t})
\end{aligned}$$

The inequality (f) above uses the fact that since the function is evaluated at time t, we must have $\beta_n\sigma_{t-1}(x_{h_t,i_t}) \geq V_{h_t}$.

- A point $x_{h,i}$ must be evaluated by the algorithm sufficiently many times to reduce the uncertainty in the function value at $x_{h,i}$ from below V_{h-1} to below V_h. We provide a loose upper bound on this quantity, by providing an upper bound on the number of function evaluations sufficient to reduce the uncertainty in the value of $f(x_{h,i})$ to below V_h. Using the first part of Proposition 3, we define q_h as follows to get the required result.

$$q_h = \min\{m : \beta_n \frac{\sigma}{\sqrt{m}} \leq V_h\}. \qquad \square$$

Remark 10. From Lemma 1, we can see that the algorithm only selects points lying in $\mathcal{X}_{(4N_1+1)V_h} = \{x \in \mathcal{X} : f(x) \geq f(x^*) - (4N_1+1)V_h\}$ for $h \geq 0\}$. Now, for $r \leq \delta_K$, let us define $\zeta_K(r) = V_{h_r}$ where $h_r = \min\{h \geq 0 : v_1\rho^h \geq r\}$ and $\tilde{D} = D^f(\delta_K, \zeta_K)$ where the term $D^f(\cdot,\cdot)$ was introduced in Definition 3. We will use this term $\tilde{D}$ for presenting our regret bounds, and will refer to it as the near optimality dimension of $\mathcal{X}$ associated with the function f.

The term $\tilde{D}$ defined above is, in general, a complicated random variable bounded above by D_1. However, in some cases we can show that it is strictly smaller than the metric dimension D_1. Our next result shows that this indeed is true for a large class of Gaussian Processes.

Claim 3. *For $\mathcal{X} \subset \mathbb{R}^D$, and Matërn kernels with $\nu > 2$, we have with high probability*

$$\tilde{D} \leq \frac{3D}{4}$$

where $\tilde{D}$ is defined in Remark 10.

Proof. We show the result for any covariance function $K \in \mathcal{K}$ which also satisfies both the assumptions in (Wang et al., 2014, Section 5). This, then implies (from the discussion following Assumption 2 in Wang et al. (2014)) that the samples of such GPs have quadratic behavior near the maximum. More specifically, there exist constants $c_1, c_2, \varrho > 0$ and $\alpha_1 \in \{1, 2\}$ such that the following holds with high probability for all $x \in B(x^*, \varrho, l)$:

$$c_1 \|x - x^*\|^{\alpha_1} \leq f(x^*) - f(x) \leq c_2 \|x - x^*\|^2.$$

Now, with $\zeta(r) \coloneqq c_3 r^{\alpha} \sqrt{\log(1/r) + c_4}$ for constants $c_3, c_4 > 0$, and replacing δ_K with $\min\{\delta_K, \varrho\}$ in the definition of $\tilde{D}$ in Remark 10, we get that

$$\mathcal{X}_r = \{x : f(x) \geq f(x^*) - \zeta(r)\} \subset B\left(x^*, \left(\frac{c_3}{c_1} r^{\alpha} \sqrt{\log(1/r) + c_4}\right)^{1/\alpha_1}\right)$$

Now, setting $a = \frac{\alpha}{2\alpha_1}$, and using the fact that $c_a \coloneqq \sup_{0<r\leq\varrho} r^a (c_4 + \log(1/r))^{1/(2\alpha_1)} < \infty$, we can conclude by using volume arguments (van Handel, 2014, Lemma 5.13) that $\tilde{D} \leq D_1(1 - \frac{\alpha}{2\alpha_1})$. More specifically for $\mathcal{X} \subset \mathbb{R}^D$, and considering Matërn kernels with $\nu > 2$, we have $\alpha = 1$ and $\alpha_1 \in \{1, 2\}$. This implies that with high probability for these kernels we must have $\tilde{D} \leq \frac{3D}{4}$. □

We can now state the main result of this section which gives us high probability bounds on the cumulative as well as simple regret of Algorithm 1.

Theorem 1. *Suppose the unknown function f is a sample from a $GP(0, K)$, with $K \in \mathcal{K}$ and $\mathcal{X}$ is a well behaved metric space (in the sense of Definition 4) with finite metric dimension (see Definition 3) D_1.*

For any $u > 0$, and for any $D' > \tilde{D}$ (defined in Remark 10), the following bounds are true with probability at least $1 - 2e^{-u}$ for Algorithm 1:

- *The cumulative regret incurred by Algorithm 1 satisfies*

$$\mathcal{R}_n = \tilde{\mathcal{O}}(n^{1 - \frac{\alpha}{2\alpha + D'}}). \tag{3.10}$$

- *Furthermore, if we make the assumption that $K(x, x) \leq 1$ for all $x \in \mathcal{X}$, we have an information type bound on the cumulative regret:*

$$\mathcal{R}_n = \mathcal{O}(\sqrt{n \gamma_n \log(n)}). \tag{3.11}$$

- *Finally, we also have an upper bound on the simple regret:*

$$\mathcal{S}_n = \tilde{\mathcal{O}}(n^{-\frac{\alpha}{2\alpha + D'}}). \tag{3.12}$$

The proof of this result is given in Appendix C.

Remark 11. The bounds in (3.10) and (3.12) which depend on near-optimality dimension will be referred to as *dimension-type* regret bounds in accordance with the terminology used by Slivkins (2014). The hidden multiplicative constant terms in (3.10) and (3.12) depend on D' as well as all the other parameters associated with $(\mathcal{X}, l)$ and the covariance function. We note that since the cumulative regret of the algorithm can be bounded in two ways, by taking the minimum of the bounds in (3.10) and (3.11), we can get a uniformly better upper bound on the cumulative regret for our algorithms for all GPs with admissible covariance functions with $K(x, x) \leq 1$.

4. Discussion

The analysis of Algorithm 1 presented in the previous section is valid for arbitrary well-behaved search space $\mathcal{X}$, any covariance function $K \in \mathcal{K}$ and in the presence of observation noise. In this section, we discuss the performance of our algorithm under some specific problem instances. In particular, we first show that our adaptive approach leads to computational requirements which do not explode with the dimension D when $\mathcal{X} \subset \mathbb{R}^D$, unlike the existing algorithms for GP bandits. We then validate the intuition provided by our toy examples in Section 1.3 by showing that the information-type bounds are indeed loose for an important family of Gaussian Processes. Finally, we specialize our results to the noiseless case, and show that our algorithm compares favorably with BaMSOO in the pure exploration problem.

4.1. Computational benefits of adaptivity

As an upshot of the adaptive discretization of the search space, the computational complexity of Algorithm 1 does not grow exponentially with the dimension of the search space, as shown in the following result.

Claim 4. *If $\mathcal{X} = [a, b]^D$ for $a, b \in \mathbb{R}$ and $D < \infty$, and N is an odd positive integer, then the computational complexity of running Algorithm 1 with a budget of n evaluations is $\mathcal{O}(Nh_{\max}n^4 + NDnh_{max}) = \mathcal{O}(Nn^4 \log n + NDn \log n)$.*

Proof. Recall that the search space considered here is well-behaved in the sense of Definition 4, and has a finite metric dimension $D_1 = D$. Furthermore, since N is odd, we observe that the sequence of partitions $(\mathcal{X}_h)_{h \geq 0}$ are nested. More specifically, if the cell associated with a node $x_{h,i}$ is refined to add the nodes $\{x_{h+1,j}; (N-1)i+1 \leq j \leq Ni\}$ to the leaf set, then we have $x_{h+1,(N-1)i+(N+1)/2} = x_{h,i}$.

Let t_n denote the number of rounds required for n function evaluations by the algorithm, and let $(\tau_j)_{j=1}^n$ denote the round numbers in which function evaluations are performed. Now, if we define $\tau_0 = 1$, then we claim that the following:

- The posterior distribution is recomputed in rounds $(\tau_j + 1)_{j=0}^n$ based on the observations. The computational task of updating the posterior based

on j observations in the round τ_j+1 can be performed in $\mathcal{O}(j^2)$ operations by using the Cholesky Decomposition. Thus the total cost for posterior computation is $\mathcal{O}(n^3)$.

- For all t such that $\tau_j + 1 < t \leq \tau_{j+1}$, the index I_t at a given point can be computed in $\mathcal{O}(j^2)$ operations. Since every refinement step adds $N-1$ new points to the leaf set and $\tau_{j+1} - \tau_j \leq h_{\max}$ for all $j \geq 0$, the total cost of computing the index in this time interval is $\mathcal{O}((N-1)h_{\max}j^2)$. For $t \in \{\tau_j + 1; 0 \leq j \leq n\}$, the index must be recomputed for the entire leaf set $\mathcal{L}_t$ whose cardinality is upper bounded by $(N-1)h_{\max}j$, and thus the computational cost is $\mathcal{O}(j^2(N-1)h_{\max}j) = \mathcal{O}(h_{\max}j^3)$. Thus the total cost of computing the index I_t for all $t \leq t_n$ is $\mathcal{O}((N-1)h_{\max}n^4)$.
- For selecting the candidate points x_{h_t,i_t} for $t \in \{\tau_j + 1, 0 \leq j < n\}$, we need to perform an exhaustive search over the entire leaf set $\mathcal{L}_t$ which is a $\mathcal{O}((N-1)h_{\max}j)$ operation. At all other times, we only need to search over the $(N-1)$ new descendants of the previous candidate point. Thus the total cost of selecting x_{h_t,i_t} for $1 \leq t \leq t_n$ is $\mathcal{O}((N-1)h_{\max}n^2+(N-1)h_{\max}n) = \mathcal{O}((N-1)h_{\max}n^2)$.
- As mentioned earlier, the refinement of a cell $\mathcal{X}_{h,i}$ when $\mathcal{X} \subset \mathbb{R}^D$ is performed by dividing it equally in N parts along its longest side. This operation involves specifying the centers and the D side lengths each of the N new cells, and is thus a $\mathcal{O}(DN)$ operation. So the total cost of refining the search space is $\mathcal{O}(h_{\max}nDN)$.

Thus the overall computational cost of running the algorithm with a budget of n function evaluations for fixed D and N is $\mathcal{O}(h_{\max}n^4)$, which is equal to $\mathcal{O}(n^4 \log n)$ using the constraint on $h_{\max}$ given in (3.4). □

As shown above, the computational complexity of our algorithm scales linearly with the dimension of the search space. This is in contrast to the existing algorithms for GP bandits with guarantees on cumulative regret, which perform a global maximization of an *acquisition function* $(\psi_t(\cdot))$ for selecting a query point:

$$x_t \in \arg\max_{x \in \mathcal{X}} \psi_t(x).$$

The computational cost of performing this operation exactly can be exponential in D. For example in the GP-UCB algorithm the acquisition function is the upper confidence bound at each point $x \in \mathcal{X}$. Over a search space $\mathcal{X} \subset \mathbb{R}^D$, for the theoretical results to be valid, the GP-UCB algorithm must select a query point at time t by calculating and then maximizing the UCB over a uniform grid of size $\mathcal{O}(t^{2D})$ (Srinivas et al., 2012). Thus the overall computational cost of running this algorithm for n rounds is $\mathcal{O}\big(\sum_{t=1}^{n} t^{2D+2}\big) = \mathcal{O}\big(n^{2D+3}\big)$. In the case of the algorithm for GP bandits studied in (Contal and Vayatis, 2016), the query point at time t is chosen by maximizing a UCB over a $\frac{1}{\sqrt{t}}$ cover of $\mathcal{X}$. The size of such a cover is $\mathcal{O}(t^{D/2})$ and the cost of computing the UCB for these points is $\mathcal{O}(t^{D/2+2})$ which leads to an overall computational complexity of $\mathcal{O}(n^{D/2+3})$. Similarly, the Thompson Sampling approach of Russo and

Van Roy (2014) as well as the GP-EST algorithm of Wang et al. (2016) also involve maximizing the acquisition function over uniform grids which results in an exponential dependence of the computational complexity on D.

4.2. Improved bounds for Matérn kernels

Matérn kernels are a widely used class of kernels parameterized by a smoothness parameter ν. For half integer values of $\nu = m + 1/2$, the Matérn kernels have the form:

$$K(r) = K(0)(1 + p_m(r))e^{-c_1\sqrt{\nu}r},$$

where $p_m = \sum_{j=1}^{m} a_i r^i$ for some $a_i > 0$ for all $1 \leq i \leq m$. Thus we can write for any $x, y \in \mathcal{X}$ such that $l(x, y) = r$:

$$\begin{aligned} d(x, y) &= [2K(0)(1 - (1 + p_m(r))e^{-c_1\sqrt{\nu}r})]^{1/2} \\ &\leq C_K(\nu)r^{\alpha}. \end{aligned}$$

It is easy to check that for $\nu = 1/2$, we have $\alpha = 1/2$, and for all other half-integer values of ν, we have $\alpha = 1$. So, for Matérn kernels, our algorithm has a dimension-type upper bound on regret of the form $\tilde{\mathcal{O}}(n^{(D'+\alpha)/(D'+2\alpha)})$ for any $D' > \tilde{D}$ and for all $\nu = m + 1/2$ with $m \geq 0$ and $\alpha \in \{1/2, 1\}$. This improves upon the existing upper bounds on Matérn kernels in the following two ways (since the existing bounds are true only when $\mathcal{X} \subset \mathbb{R}^D$, we will restrict our comparison to this case, and so we have $D_1 = D$ here):

- The existing regret bounds are only valid for the case of $\nu > 1$ (Srinivas et al., 2012; Contal and Vayatis, 2016), whereas the dimension-type regret bounds of our algorithm is valid for all $\nu \geq 1/2$. In particular, for the exponential kernel ($\nu = 1/2$, also referred to as the Ornstein-Uhlenbeck process), Srinivas et al. (2012) conjectured that it may not be possible to derive a regret bound of the form shown in (1.4). This conjecture was refuted by Contal and Vayatis (2016), but the authors did not provide an explicit characterization of $\mathcal{R}_n$ as no suitable bounds for γ_n for this kernel are known. Our result provides an upper bound on the cumulative regret for the exponential kernel of the form $\mathcal{R}_n \leq \tilde{\mathcal{O}}(n^{(2D'+1)/(2D'+2)})$ for any $D' > \tilde{D}$, which is, to the best of our knowledge, the first explicit sublinear bound on the cumulative regret for the GP bandits problem with exponential kernel.
- The existing regret bounds for Matérn kernels have the form $\tilde{\mathcal{O}}\big(n^{\frac{D(D+1)+\nu}{D(D+1)+2\nu}}\big)$ (Srinivas et al., 2012; Contal and Vayatis, 2016) for $\nu > 1$. As compared to this, the bounds obtained by our algorithm, after substituting $\alpha = 1$ for Matérn kernels with $\nu > 1$ depend upon $\tilde{D}$, which itself is a random variable dependent on the sample function f of the Gaussian Process and can take values anywhere from 0 to D. Assuming the worst case value of $\tilde{D} = D$, we observe that for $D \geq \nu - 1$, we have

$$\frac{D+1}{D+2} \leq \frac{D(D+1)+\nu}{D(D+1)+2\nu}.$$

Thus $D \geq \nu - 1$ is a sufficient condition for our upper bounds to be tighter than the best known bounds for Matérn kernels. The two most commonly used Matérn kernels in Machine learning correspond to $\nu = 3/2$ and $\nu = 5/2$ (Rasmussen and Williams, 2006, Chapter 4), for which the sufficient condition reduces to $D \geq 1$ and $D \geq 2$ respectively.

4.3. Regret under noiseless observations

In this section, we consider the special case where there is no observation noise, and specialize the regret bounds of our algorithm to this setting. In particular we have the following bounds:

Claim 5. *If in addition to the assumptions of Theorem 1, we further assume that the observations are noiseless, i.e., $\sigma = 0$, we get with high probability for any $D' > \tilde{D}$*

$$\mathcal{R}_n = \tilde{\mathcal{O}}(n^{1-\frac{\alpha}{D'}}), \tag{4.1}$$

$$\mathcal{S}_n = \tilde{\mathcal{O}}(n^{-\alpha/D'}), \tag{4.2}$$

if $\tilde{D} > 0$, and

$$\mathcal{R}_n = \tilde{\mathcal{O}}(1), \tag{4.3}$$

$$\mathcal{S}_n = \tilde{\mathcal{O}}(e^{-c_1 \log(1/\rho) n}), \tag{4.4}$$

if $\tilde{D} = 0$ and $h_{\max} = \Omega(n)$, for some constant $c_1 > 0$.

Remark 12. We note that unlike Theorem 1, we do not present information-type bounds on the cumulative regret in Claim 5. This is because the information-type bounds given by (1.4) are not directly applicable in the noiseless setting as the term γ_n becomes undefined for $\sigma = 0$.

As mentioned earlier, our work is motivated by BaMSOO, an adaptive algorithm for the Bayesian optimization problem which works only with noiseless observations (Wang et al., 2014). BaMSOO builds upon the *Simultaneous Optimistic Optimization*(SOO) algorithm of Munos (2011) by making the further assumption that the unknown function is a sample from a GP, and then utilizes the posterior confidence intervals in selection of points. Wang et al. (2014) obtained an upper bound on the simple regret of the order $\tilde{\mathcal{O}}(n^{-c/D})$ for some $c > 0$ which is similar to our simple regret bound in Claim 5. However, our approach extracts more information about the function from the GP prior and has some advantages over BaMSOO in the pure exploration setting. In particular, the derivation of regret bounds for BaMSOO required the assumption (Wang et al., 2014, Assumption 2) that the unknown function is approximately quadratic in the region around the maximum x^*, which for example is ensured if the covariance function has continuous partial derivative of order 6. Our result does not require this quadratic behavior, and is valid for kernels not satisfying the smoothness requirements, such as the exponential kernel $K(r) = ce^{-c_1 r}$, and

the Slepian kernel $K(r) = (1 - r)^+$ (Slepian, 1961). Furthermore, if for some instances of the function f, the random variable $\tilde{D}$ equals zero, then we obtain an exponentially decaying simple regret bound for Algorithm 1. This is unlike the simple regret bounds for BaMSOO which decay polynomially in n for all admissible kernels.

5. Extensions

In this section, we first present an algorithm for GP bandits which uses an alternative approach to locally refining the search space as compared to Algorithm 1. While Algorithm 1 requires a tree of partitions to adaptively discretize the space $\mathcal{X}$, the algorithm presented in Section 5.1 instead utilizes a *covering oracle* to explore the search space.

Next, in Section 5.2 we apply our general approach to design an adaptive algorithm for the problem of contextual GP bandits, an extension of the usual GP bandits problem first studied in (Krause and Ong, 2011).

5.1. Bayesian zooming algorithm

We now present a Bayesian version of the zooming algorithm for Lipschitz optimization introduced by Kleinberg et al. (2013) and analyze its regret. In particular, instead of assuming that the metric space $(\mathcal{X}, l)$ is well-behaved in the sense of Definition 4, this algorithm requires access to the space $(\mathcal{X}, l)$ through a *covering oracle* (see Remark 14 for definition) to locally refine the discretization.

The algorithm proceeds by constructing an increasing sequence of *active* subsets of $\mathcal{X}$ denoted by $(A_t)_{t\geq 1}$. As with Algorithm 1, we can compute upper and lower confidence intervals for the function values at points in A_t for all $t \geq 1$ such that

$$f(x) \in [\mu_{t-1}(x) - \beta_n \sigma_{t-1}(x), \mu_{t-1} + \beta_n \sigma_{t-1}(x)]$$

with high probability for a suitable factor β_n.

Also, to each point x that has been evaluated at least once, we assign a radius denoted by $r(x)$. The radius $r(x)$ can take values in a set $S = \{r_k = diam(\mathcal{X})2^{-k}; k \in \mathbb{N}\}$, where

$$diam(\mathcal{X}) \coloneqq \sup_{x_1, x_2 \in \mathcal{X}} l(x_1, x_2)$$

is the diameter of the metric space $(\mathcal{X}, l)$ and is assumed to be finite. For implementing the algorithm, we further require bounds $(W(r_k))_{k\in\mathbb{N}}$ such that for all $x \in \mathcal{X}$ and for all $k \in \mathbb{N}$, $W(r_k)$ is a bound on the variation of the GP sample in the ball $B(x, r_k, l)$ with high probability. We obtain these $W(r_k)$ using Proposition 2. We also require a parameter $r_{\min}$ as input, which plays a role similar to $h_{\max}$ in Algorithm 1. The details behind the choice of these parameters are provided in Appendix D.

Corresponding to each point that has been evaluated at least once, we have an associated confidence region $B(x, r(x), l)$, and furthermore, we also have an

upper bound on the maximum value of the function in that region (w.h.p.) given by the index:

$$J_t(x) = \mu_{t-1}(x) + \beta_n \sigma_{t-1}(x) + W(r(x)). \tag{5.1}$$

In each round t, a candidate point is selected in an optimistic manner from the set A_t, i.e.,

$$x_t \in \arg\max_{x \in A_t} \quad J_t(x). \tag{5.2}$$

The index $J_t(x)$ can take a large values if:

- the point x has been evaluated very few times, in which case the uncertainty at x (i.e., $\beta_n \sigma_{t-1}(x)$) as well as the bound on the variation of f in the confidence region ($W(r(x))$) are large.
- or if the point x has been observed many times, and the true function value $f(x)$ is large.

In this way the selection rule strikes a balance between exploration of poorly understood regions, and exploitation of well explored regions with high function values.

Having chosen a candidate point x_t at time t, the algorithm takes one of two actions:

- *Refine:* If the uncertainty in the function value a point x_t is smaller than the bound on the variation of the function in the confidence region associated with point x_t, then the algorithm locally refines the search space, that is, it shrinks the radius of the confidence region associated with x_t by a factor of 2.
- *Evaluate:* Otherwise, if the uncertainty in the function value is larger than the variation in the confidence region, the function is evaluated at the candidate point x_t.

In order to ensure that the entire search space is taken into consideration, the algorithm maintains at all times the following invariant:

$$\mathcal{X} \subset \cup_{x \in A_t} B(x, r(x), l). \tag{5.3}$$

If this invariant is violated, a point from the *uncovered region* (i.e., $\mathcal{X} \setminus \cup_{x \in A_t} B(x, r(x), l)$ is added to the active set of points with an associated radius $r_0 = diam(\mathcal{X})$.

All the steps described above are formally stated as a pseudo-code in Algorithm 2.

Remark 13. A key difference between Algorithm 2 and the zooming algorithm for Lipschitz functions is that our algorithm only evaluates a point if the confidence radius associated with it is small enough (Lines 3-4 of Algorithm 2). This is unlike the zooming algorithm in (Kleinberg et al., 2013), in which a point is evaluated every round. This modification is necessary to obtain the information type bounds on the cumulative regret for our algorithm.

Algorithm 2: Zooming Algorithm for GP bandits

Input : $n > 0$, $(r_k)_{k\geq 0}$, $(W(r_k))_{k\geq 0}$, $r_{\min}$
Initialize: $t = 1$, $n_e = 0$, $A_t = \{\}$

1 **while** $n_e \leq n$ **do**
2 choose $x_t = \arg\max_{x_i \in A_t} \mu_{t-1}(x_i) + \beta_n \sigma_{t-1}(x_i) + W(r(x_i))$
3 **if** $(\beta_n \sigma_{t-1}(x_t) \leq W(r(x_t)))$ *AND* $(r(x_t) \geq r_{\min})$ **then**
4 $r(x_t) \leftarrow r(x_t)/2$
5
6 **else**
7 evaluate $y_t = f(x_t) + \eta_t$
8 update posterior $\mu_t(x)$ and $\sigma_t(x)$
9 update $n_e \leftarrow n_e + 1$
10 **end**
11 **if** $\mathcal{X} \not\subset \cup_{x_i \in A_t} B(x_i, r(x_i), l)$ **then**
12 Add a point $x \in \mathcal{X} \setminus \cup_{x_i \in A_t} B(x_i, r(x_i), l)$ to A_t, with $r(x) = r_0 = diam(\mathcal{X})$.
13 **end**
14 $t \leftarrow t + 1$
15 **end**

Output : $x(n)$: point with the smallest radius

Remark 14. For maintaining the invariant described in (5.3) and in Lines 10-12 of Algorithm 2, we assume the existence of a *covering oracle* (Kleinberg et al., 2013, Section 1.5), which takes in as inputs a finite set of balls and outputs whether these balls cover the entire space $\mathcal{X}$ or not. In the latter case, the covering oracle also returns an arbitrary point from the uncovered region of $\mathcal{X}$. In our case, if at the beginning of round t the entire space is covered by the balls (this is true at $t = 2$), and suppose a point x is selected by the algorithm and its confidence radius is shrunk from $r(x)$ to $r(x)/2$. Then at the beginning of the next round, we only need to check whether the annular region $B(x, r(x), l) \setminus B(x, r(x)/2, l)$ is fully covered by the other balls or not.

Our next result shows that we can obtain the same regret performance for Algorithm 2 as we did for the tree-based algorithm.

Theorem 2. *Suppose the unknown function f is a sample from a $GP(0, K)$, with $K \in \mathcal{K}$. $(\mathcal{X}, l)$ is assumed to be a compact metric space with finite metric dimension D_1 (see Definition 3). Moreover, we assume that we can access the metric space $(\mathcal{X}, l)$ through a* covering oracle.

Then, for any $u > 0$ and for $D' > \tilde{D}_Z$ (defined in Remark 15), the following bounds are true with probability at least $1 - 2e^{-u}$ for Algorithm 2:

- *We have the following dimension-type bound on the cumulative regret.*

$$\mathcal{R}_n = \tilde{\mathcal{O}}(n^{1-\frac{\alpha}{2\alpha+D'}}). \tag{5.4}$$

- *Under the extra assumption that $K(x, x) \leq 1$ for all $x \in \mathcal{X}$, we also have an information type bound on the cumulative regret:*

$$\mathcal{R}_n = \mathcal{O}(\sqrt{n\gamma_n \log(n)}). \tag{5.5}$$

- *Finally, we also have an upper bound on the simple regret:*

$$\mathcal{S}_n = \tilde{\mathcal{O}}(n^{-\frac{\alpha}{2\alpha+D'}}). \tag{5.6}$$

The details of the choice of the parameters of Algorithm 2 as well as an outline of the proof of Theorem 2 is provided in Appendix D.

Remark 15. The near-optimality dimension $\tilde{D}_Z$ used in the statement of Theorem 2 can be defined similar to the definition of $\tilde{D}$ introduced in Remark 10. More specifically, by Lemma 2 in Appendix D we know that Algorithm 2 only selects evaluation points from sets of the form $\mathcal{X}_{5W(r_k)} = \{x \in \mathcal{X} : f(x^*) - f(x) \leq 5W(r_k)\}$ for $k \geq 0$. So we can proceed as in Remark 10 to define $\tilde{D}_Z = D^f(\delta_K, \zeta_K)$, with $\zeta_K(z) \coloneqq 5W(r_{k_z})$ and $k_z \coloneqq \min\{k \geq 0 : r_k \leq z\}$.

5.2. Extension to contextual GP bandits

The contextual bandit problem is a generalization of the multi-armed bandit (MAB) problem in which at the beginning of each round, the agent receives a context, and the task is to select an action (or arm) which is optimal for the context received. This problem was studied in the non-Bayesian framework by Rigollet and Zeevi (2010) for the case of two arms under certain smoothness and margin assumptions on the reward function. It was further extended to the case of finitely many arms by Perchet and Rigollet (2013), and to the continuum armed setting under Lipschitz assumptions by Slivkins (2014). Krause and Ong (2011) considered this problem in the Bayesian framework with GP prior and proposed the CGP-UCB algorithm which is a variant of the GP-UCB algorithm. They obtained information-type regret bounds on the contextual regret for CGP-UCB and additionally, provided bounds on the maximum information gain (γ_n) for composite kernels over the product space.

For this problem, the set $\mathcal{X}$ is a product of two sets, $\mathcal{X}_c$ the context set and $\mathcal{X}_a$ the action set, and $f : \mathcal{X} \to \mathbb{R}$ is the mean reward observed for a context-action pair. As before, we will assume that the unknown function f is a sample from a Gaussian process $GP(0, K)$ now indexed by the product set $\mathcal{X} = \mathcal{X}_c \times \mathcal{X}_a$. In each round τ, the agent receives a *context* $x_\tau^c \in \mathcal{X}_c$, and must select an action $x_\tau^a \in \mathcal{X}_a$ corresponding to that context and observe the reward $y_\tau = f(x_\tau) + \eta_\tau$ where $x_\tau = (x_\tau^c, x_\tau^a) \in \mathcal{X}$. The goal of the agent is to design a strategy of selecting actions to minimize the *contextual cumulative regret* defined as

$$\mathcal{R}_n^c \coloneqq \sum_{\tau=1}^{n} \Delta^c(x_\tau), \tag{5.7}$$

where we have

$$\Delta^c(x_\tau) \coloneqq \sup_{x^a \in \mathcal{X}_a} f(x_\tau^c, x^a) - f(x_\tau^c, x_\tau^a). \tag{5.8}$$

5.2.1. Tree based algorithm for contextual GP bandits

We again make the assumption that the space $\mathcal{X}$ admits a tree of partitions satisfying the properties described in Definition 4. To simplify the description of the algorithm, we will assume that the metric space admits a binary tree of partitions (i.e., $N = 2$). We show that with a small modification to the point selection rule and the cell expansion strategy, we can easily adapt Algorithm 1 to the problem of contextual GP bandits.

We need to introduce a couple of definitions in order to describe the algorithm. We call a cell $\mathcal{X}_{h,i}$ *active* with respect to a context $x^c \in \mathcal{X}_c$, if there exists an action $x^a \in \mathcal{X}_a$ such that $(x^c, x^a) \in \mathcal{X}_{h,i}$. Now, given a context x^c, for every active cell $\mathcal{X}_{h,i}$ (corresponding to a point $x_{h,i} \in \mathcal{L}_t$) we find a point of the from $(x^c, x^a_{h,i}) \in \mathcal{X}_{h,i}$, and we will refer the collection of these points as a leaf set relevant to the context x^c denoted by $\mathcal{L}_t^{rel}$.

Suppose a cell $\mathcal{X}_{h,i}$ with $0 < h < h_{\max}$ is expanded by the algorithm at time t_0. Then for all $t \geq t_0$, we use $\bar{x}^{(t)}_{h,i}$ to denote the candidate point in the cell $\mathcal{X}_{h,i}$ which was chosen by the algorithm at time t_0. This point has the property that $\beta_n \sigma_{t-1}(\bar{x}^{(t)}_{h,i}) \leq V_h + \beta_n g(v_1 \rho^h)$ for all $t \geq t_0$. Clearly, this property is true at time $t = t_0$ (by Line 6 of Algorithm 3). Furthermore, since the posterior variance at a point cannot increase as more observations are made, the inequality holds for all $t > t_0$ as well.

For all points in $\mathcal{L}_t^{rel}$, we define as index as follows:

$$I_t^c(x_{h,i}) = \min\{\mu_{t-1}(x_{h,i}) + \beta_n \sigma_{t-1}(x_{h,i}), \mu_{t-1}(\bar{x}^{(t)}_{h-1,\lfloor i/2 \rfloor}) + \beta_n \sigma_{t-1}(\bar{x}^{(t)}_{h-1,\lfloor i/2 \rfloor}) + V_{h-1}\} + V_h \tag{5.9}$$

The rest of the algorithm proceeds in a manner similar to Algorithm 1. We select a candidate point by maximizing the index I_t^c over the relevant leaf set $\mathcal{L}_t^{rel}$. Having selected the candidate point, we either evaluate the function or refine the discretization depending on the uncertainty in the function value at the chosen point.

The steps of the algorithm are shown as a pseudo-code in Algorithm 3. The values of the parameters $h_{\max}$, β_n and $(V_h)_{h \geq 0}$ used here are the same as those used in the algorithms for GP bandits, with the modification that n now represents the total number of context arrivals and $\mathcal{X} = \mathcal{X}_c \times \mathcal{X}_a$.

5.2.2. Bounds on contextual regret

For the algorithm for contextual GP bandits described above, we now present high probability bounds on the contextual regret $\mathcal{R}_n^c$:

Theorem 3. *Suppose Algorithm 3 is applied to a contextual GP bandits problem, where the reward function f is a sample from a zero mean GP with covariance function $K \in \mathcal{K}$ and furthermore K is assumed to be isotropic*[4] *The*

[4] i.e., covariance between two points x_1 and x_2 satisfies $K(x_1, x_2) = K(l(x_1, x_2))$

Algorithm 3: Tree based Algorithm for Contextual GP bandits

Input : $n > 0$, $(\mathcal{X}_h)_{h\geq 0}$, $(V_h)_{h\geq 0}$, $h_{\max}$

Initialize: $\mathcal{L}_0 = \{x_{0,1}\}$, $t = 0$, $\tau = 1$, $flag$=TRUE

while $\tau \leq n$ **do**
  Observe a context x_τ^c
  while *flag* **do**
    Obtain $\mathcal{L}_t^{rel}$
    choose $x_{h_t,i_t} = (x_\tau^c, x_t^a) \in \arg\max_{x_i \in \mathcal{L}_t^{rel}} I_t^c(x_{h,i})$
    if $\beta_n \sigma_{t-1}(x_{h_t,i_t}) \leq V_{h_t} + \beta_n g(v_1 \rho^{h_t})$ *AND* $h_t < h_{\max}$ **then**
      $\mathcal{L}_{t+1} = \mathcal{L}_t \setminus \{x_{h_t,i_t}\}$
      $\mathcal{L}_{t+1} = \mathcal{L}_{t+1} \cup \{x_{h_t+1,2i_t-1}, x_{h_t+1,2i_t}\}$
    else
      play the action x_t^a
      observe the reward $y_t = f(x_{h_t,i_t}) + \eta_t$
      update posterior $\mu_t(x)$ and $\sigma_t(x)$
      $flag$ = FALSE
    end
    $t \leftarrow t + 1$
  end
  $\tau \leftarrow \tau + 1$
  $flag$ = TRUE
end

product space $\mathcal{X} = \mathcal{X}_c \times \mathcal{X}_a$ is assumed to be well-behaved *(Definition 4) with finite metric dimension D_1. Then after observing n contexts, for any $u > 0$ and $D' > \tilde{D}$ we have with probability at least $1 - 2e^{-u}$*

$$\mathcal{R}_n^c = \tilde{\mathcal{O}}(n^{1-\alpha/(\tilde{D}+2\alpha)}). \tag{5.10}$$

In addition if we further assume that $K(x, x) \leq 1$ for all $x \in \mathcal{X}$, then we can also have an information type bound on the contextual regret:

$$\mathcal{R}_n^c = \mathcal{O}(\sqrt{n\gamma_n \log(n)}). \tag{5.11}$$

The proof of the above result essentially follows the same arguments used in the proof of Theorem 1, and we omit the details here. For deriving the dimension type contextual regret bound, we will require an intermediate lemma analogous to Lemma 1. The derivation of this result differs from Lemma 1 in the following two ways:

- Unlike Algorithm 1, a single point cannot be evaluated repeatedly in the contextual case as the contexts are not chosen by the algorithm. Thus to get a bound on the term q_h here, we need to upper bound the posterior variance at a point given a certain number of function evaluations at points in a ball $B(x, r, l)$. For this we use the result in the second part of Proposition 3.
- The definition of $\bar{x}_{h,i}^{(t)}$ introduced earlier is crucial in obtaining a bound on the sub-optimality of the chosen action analogous to that in (3.8).

Suppose the algorithm selects an action x_t^c which is at level h_t of the tree, in response to a context x_τ^c and let $x_\tau^* := (x_\tau^c, \arg\sup_{x^a \in \mathcal{X}_a} f(x_\tau^c, x^a))$ and $x_{h_t,i_t} = (x_\tau^c, x_t^a)$ (note that τ is the index of the context (i.e.,$1 \leq \tau \leq n$) and t is the index of the round (i.e.,$1 \leq t \leq nh_{\max}$) in Algorithm 3). We then proceed as follows:

$$\begin{aligned} f(x_\tau^*) &\leq \mu_{t-1}(\bar{x}_{h_t-1,\lceil i_t/2 \rceil}) + \beta_n \sigma_{t-1}(\bar{x}_{h_t-1,\lceil i_t/2 \rceil}) + V_{h_t-1} + V_{h_t} \\ &\leq f(\bar{x}^{(t)}_{h_t-1,\lceil i_t/2 \rceil}) + 2\beta_n \sigma_{t-1}(\bar{x}^{(t)}_{h_t-1,\lceil i_t/2 \rceil}) + V_{h_t-1} + V_{h_t} \\ &\leq f(x_{h_t,i_t}) + V_{h_t-1} + 2\beta_n \sigma_{t-1}(\bar{x}^{(t)}_{h_t-1,\lceil i_t/2 \rceil}) + V_{h_t-1} + V_{h_t} \\ &\overset{(a)}{\leq} f(x_{h_t,i_t}) + 2(V_{h_t-1} + \beta_n g(v_1 \rho^{h_t-1})) + 2V_{h_t-1} + V_{h_t} \\ &\overset{(b)}{\leq} f(x_{h_t,i_t}) + 5V_{h_t-1} + 2\beta_n g(v_1 \rho^{h_t-1}) \\ \Rightarrow \Delta^c(x_\tau^c, x_t^a) &\leq 5V_{h_t-1} + 2\beta_n g(v_1 \rho^{h_t-1}) \end{aligned}$$

The inequality (a) above uses the definition of $\bar{x}^{(t)}_{h_t-1,\lceil i_t/2 \rceil}$ and (b) uses the fact that $V_{h_t} \leq V_{h_t-1}$.

With these results available, the remainder of the proof of Theorem 3 mirrors the proof of Theorem 1.

Remark 16. Compared to the CGP-UCB algorithm of Krause and Ong (2011), Algorithm 3 again has two benefits. First, if $\mathcal{X} \subset \mathbb{R}^D$, then the computational cost of running the algorithm does not depend on the dimension of the space, unlike the CGP-UCB whose practical implementation cost increases exponentially with D. Second, as with Algorithm 1, our theoretical regret bounds are tighter for Matérn kernels when we have $D \geq \nu - 1$.

Remark 17. Krause and Ong (2011) considered composite covariance functions formed either by taking products $K(x^c, c^a) = K_c(x^c) \times K_a(x^a)$ or by taking sums $K(x^c, x^a) = K_c(x^c) + K_a(x^a)$ of different covariance functions over the context space and the action space. Since our class of covariance functions $\mathcal{K}$ is closed under such operations, if K_c and K_a lie in $\mathcal{K}$ then their combination will also be in $\mathcal{K}$, and thus our dimension-type bounds on the contextual regret is valid for such composite covariance functions. In addition, for the information type bound we can use (Krause and Ong, 2011, Theorem 2 and Theorem 3) to get the required explicit upper bound on γ_n.

6. Technical results

In this section, we present some analytical results about the Gaussian Processes satisfying the assumptions described in Section 3.3.1, which were used in the design of our algorithms.

We begin by deriving a high probability bound on the maximum variation of the sample functions of a Gaussian Process within a d-ball of radius b around some fixed point x.

Proposition 1. *Suppose* $\{f(x); x \in \mathcal{X}\}$ *is a separable zero mean Gaussian Process* $GP(0, K)$*, and let* d *denote the usual metric on* $\mathcal{X}$ *induced by the GP. Let* $B(x_0, b, d) \subset \mathcal{X}$ *be a* d*-ball of radius* $b > 0$*. Then we have for any* $u > 0$*:*

$$Pr\Big(\sup_{x \in B(x_0,b,d)} |f(x) - f(x_0)| > w_b\Big) \leq e^{-u}, \tag{6.1}$$

with $w_b \leq 4b\big(\sqrt{C_2 + 2u + 4D_1' \log(1/b)} + C_3\big)$*. Here* C_2 *and* C_3 *are positive constants and* D_1' *is the metric dimension of* $B(x_0, b, d)$ *with respect to* d*.*

The details of the proof of this statement is given in Appendix B.1. The proof uses the classical chaining technique for bounding the suprema of Gaussian Processes, and follows the same line of arguments used in some existing results in literature such as (Contal, 2016, Theorem 3.3) and (van Handel, 2014, Theorem 5.24).

The previous result gives us a bound on the variation of the samples of a given Gaussian process within a given d-ball of radius b. Using this and the union bound, we can easily extend this to a sequence of discretizations of $\mathcal{X}$:

Corollary 1. *Suppose* $\{f(x); x \in \mathcal{X}\}$ *is a zero mean Gaussian Process which induces the metric* d *on* $\mathcal{X}$*. Let* $(\mathcal{X}_k)_{k \geq 0}$ *be a sequence of finite subsets of* $\mathcal{X}$*, and to every point in* $\mathcal{X}_k$ *we associate a radius* b_k *with respect to the metric* d*. Then we have* $Pr(\Omega_u) \geq 1 - e^{-u}$*, where the event* Ω_u *is defined as*

$$\Omega_u = \{\forall n \geq 0, \forall x \in \mathcal{X}_k : \sup_{y \in B(x,b_k,d)} |f(y) - f(x)| \leq w_k\}, \tag{6.2}$$

with the value of w_k *given by:*

$$w_k \leq 4b_k\big(\sqrt{C_4 + 2u + 2\log(|\mathcal{X}_k|/(b_k^{2D_1'}))} + C_3\big), \tag{6.3}$$

where $C_4 = C_2 + 2\log(n^2\pi^2/6)$*, and* D_1' *is the metric dimension of* $(\mathcal{X}, d)$*.*

Proof. The result is obtained by replacing $u_k \leftarrow u_k + \log(n^2\pi^2/6) + \log(|\mathcal{X}_k|)$ in the proof of Proposition 1 and then taking two union bounds, one over points in $\mathcal{X}_k$ for a fixed n and the other over all values of $n \in \mathbb{N}$. □

Specializing this result to the class of Gaussian Processes with covariance functions $K \in \mathcal{K}$, we can obtain bounds on the variation of the GP samples in l-balls.

Corollary 2. *Suppose* $\{f(x); x \in \mathcal{X}\}$ *is a Gaussian Process with its covariance function* $K \in \mathcal{K}$*, and let* l *be a metric defined on* $\mathcal{X}$*. Then for* $(\mathcal{X}_k)_{k \geq 0}$ *subsets of* $\mathcal{X}$*, and* $(r_k)_{k \geq 0}$ *the associated radius values, we have for any* $u > 0$*:*

$$P(\Omega_{u1}) \geq 1 - e^{-u},$$

where the event Ω_{u1} *is defined as*

$$\Omega_{u1} = \{\forall n \geq 0, \forall x \in \mathcal{X}_k : \sup_{y \in B(x,r_k,l)} |f(y) - f(x)| \leq w(r_k)\}, \tag{6.4}$$

with the value of $w(r_k)$ *given by:*

$$w(r_k) \leq 4g(r_k)\big(\sqrt{C_4 + 2u + 2\log(|\mathcal{X}_k|/(g(r_k)^{2D_1})} + C_3\big). \tag{6.5}$$

This result gives us control over the variation of the Gaussian process samples in balls centered at points in $(\mathcal{X}_k)_{k\geq 0}$.

Now suppose we want to obtain high probability bounds on the variation of the GP samples in l-balls of radius $(r_k)_{k\geq 0}$ for all points $x \in \mathcal{X}$ and not just those in $(\mathcal{X}_k)_{k\geq 0}$. Our next result shows that we can obtain this by a small modification of the previous result.

Proposition 2. *For a given sequence* $(r_k)_{k\geq 0}$*, we have for any* $u>0$*,* $Pr(\Omega_{u2}) \geq 1-e^{-u}$*, where the event* Ω_{u2} *is defined as*

$$\Omega_{u2} = \{\forall k \geq 1, \forall x \in \mathcal{X} : \sup_{y \in B(x,r_k,l)} |f(x) - f(y)| \leq \tilde{w}_k\}, \tag{6.6}$$

and $\tilde{w}_k = 2w(R_k)$*, where* $w(R_k)$ *is as defined in 6.5 by selecting* $\mathcal{X}_k$ *to be an* ϵ_k *cover (for any* $\epsilon_k > 0$*) of* $\mathcal{X}$*, and choosing* R_k *satisfying* $R_k \geq r_k + \epsilon_k$*.*

This result is crucial in the design of Algorithm 2 as the covering oracle can return an arbitrary point in the uncovered region of the search space $\mathcal{X}$, and thus we need to bound the variation of f in ball centered at *any* point $x \in \mathcal{X}$ with radius r_k for $k \geq 0$.

Remark 18. The result follows by application of Corollary-2 for the given choice of R_k and ϵ_k. However, the idea behind this result can be better understood through Figure.2. Let us consider a fixed radius r_k. We want a bound $\tilde{w}_k$ such that for all $x \in \mathcal{X}$ we know that with high probability the variation of a Gaussian process sample within the ball $B(x, r_k)$ is no more than $\tilde{w}_k$. Since the set $\mathcal{X}$ in general can be uncountable, we cannot directly use union bound to get this result. However, we can get a bound in the following way: For some $\epsilon_k > 0$, consider an ϵ_k covering of $\mathcal{X}$, denoted by $\mathcal{X}_{\epsilon_k}$. Now for every point $z \in \mathcal{X}_{\epsilon_k}$, we associate a ball $B(x, R_k, l)$ with $R_k \geq r_k + \epsilon_k$ and compute the corresponding variation $(w(R_k))$ within this ball for all $x \in \mathcal{X}_{\epsilon_k}$ by using Corollary 2. By definition of $\mathcal{X}_{\epsilon_k}$, for all $x \in \mathcal{X}$ there exists a z_x within ϵ_k distance of x, and by the choice of radius R_k, we know that $B(x, r_k, l) \subset B(z_x, R_k, l)$. Now, by the triangle inequality, we have for all $y \in B(x, r_k, l)$, $|f(x)-f(y)| \leq |f(x)-f(z_x)|+|f(y)-f(z_x)|$, which gives us the required bound $\tilde{w}_k \leq 2w(R_k)$.

Finally, we present a result about the posterior variance at a point x at which we have multiple noisy observations.

Proposition 3. *Suppose the unknown function* f *is a sample from* $GP(0, K)$ *with* $K \in \mathcal{K}$*.*

- *If a point* x *has been evaluated* $n_t(x)$ *times before time* t *according to the observation model* $y(x) = f(x) + \eta$ *where the noise term* η *is distributed according to* $N(0, \sigma^2)$*. Then we have*

$$\sigma_t(x) \leq \frac{\sigma}{\sqrt{n_t(x)}}, \tag{6.7}$$

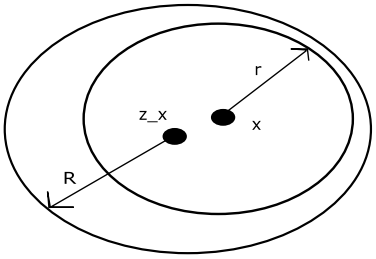


FIG 2. *If $\mathcal{X}_\epsilon$ is an ϵ cover of $\mathcal{X}$, then for any $x \in \mathcal{X}$ there exists a $z_x \in \mathcal{X}_\epsilon$ within ϵ distance of x. A ball of radius $R \geq r + \epsilon$ will contain the ball $B(x, r)$ and so twice the variation of f in $B(z_x, R)$ (denoted by $w(R)$) is an upper bound on the variation of f in $B(x, r)$*

where $\sigma_t(x)$ is the posterior variance at the point x after t observations.

- *Suppose, we make the further assumption that the covariance function K is isotropic, i.e., $K(x, y) = K(r)$ where $r = l(x, y)$. Now, if $n_t(x, r)$ denotes the number of times a point from the ball $B(x, r, l)$ has been evaluated up to time t, then we have*

$$\sigma_t(x) \leq \frac{\sigma}{\sqrt{n_t(x, r)}} + d(r) \leq \frac{\sigma}{\sqrt{n_t(x, r)}} + g(r). \tag{6.8}$$

This result allows us to estimate the number of evaluations required to bring the uncertainty about the function value at a point below a certain threshold. The first part of the above result is used in the analysis of the two algorithms proposed for GP bandits (Algorithm 1 and Algorithm 2), while the second part is used in the analysis of Algorithm 3 for Contextual GP bandits.

7. Conclusion

In this paper, we considered the problem of optimizing an unknown function under noisy bandit feedback, and presented an algorithm which adaptively discretizes the search space using a hierarchical tree of partitions. We then obtained high probability bounds on the cumulative and simple regret for our algorithm. Because of adaptive refinement of the search space, our algorithms can be computationally much cheaper than the existing approaches using uniform discretizations. Furthermore, we also identified sufficient conditions under which the regret bounds of our algorithms improve upon the existing theoretical results.

Finally, we note that the tools described in Section 6, along with some stronger bounds on suprema of GPs such as those presented in (Contal and Vayatis, 2016; Van Handel, 2015) may be useful for designing adaptive algorithms for some other settings, such as time varying GP bandits problem (Bogunovic et al., 2016a).

Appendix A: Details of toy examples in Section 1.3

A.1. Example 1

First we note that the covariance function of the Gaussian Process is uniformly upper bounded by a_1^2 which implies that the information type regret bound is valid for it (Srinivas et al., 2012). Before obtaining the lower bound on γ_n, let us select the parameters $(a_i)_{i\geq 1}$ in the following way for a fixed $\delta > 0$:

$$a_1 = \frac{1}{(\Phi)^{-1}((1+\delta)/2)}$$
$$a_i = \frac{1}{2\sqrt{2\log(\frac{\pi^2 i^2}{6\delta})}},$$

where $\Phi(.)$ is the cdf of Standard Normal random variable.

Now, we have the following:

$$\gamma_n \overset{(a)}{\geq} \sum_{i=1}^{n} \log\left(1+\frac{a_i^2}{\sigma^2}\right) \overset{(b)}{\geq} n\log\left(1+\frac{a_n^2}{\sigma^2}\right)$$
$$\overset{(c)}{\geq} n\frac{a_n^2}{a_n^2+\sigma^2} = \frac{n}{1+8\sigma^2\log\left(\frac{\pi^2 n^2}{3\delta}\right)}.$$

The inequality (a) is obtained by the following observations: by definition γ_n is greater than or equal to the mutual information between f and the observations at any set of n points. In particular, the inequality holds if the n observation points are $x_i = 1/3^i + 1/(2*3^i)$ for $i = 1, 2, \ldots, n$. Finally, we use (Srinivas et al., 2012, Lemma 5.3) to write the mutual information between f and the observations at these points in terms of the predictive variances, which in this case are just $(a_i^2)_{i=1}^n$ because of the independence assumption. The inequality (b) uses the fact that $(a_i)_{i\geq 0}$ is a decreasing sequence, and (c) follows from the inequality $\log(1+x) \geq \frac{x}{1+x}$ for $x \geq 0$. From the above, we get the following bound:

$$\gamma_n \geq n\max\left(\frac{1}{2}, \frac{1}{16\sigma^2\log\left(\frac{\pi^2 n^2}{3\delta}\right)}\right).$$

This implies that for all $\sigma^2 > 0$, the information type regret bound for this Gaussian Process increases linearly with n.

Now we show that for the given choice of parameters for this Gaussian Process, the global maximizer of the sample function f can be found from just one evaluation with high probability. Let us define the following events: $E_1 = \{|a_1 X_1| \geq 1\}$, $E_2 = \{\forall i \geq 2 : |a_i X_i| \leq 1/2\}$ and $E_3 = \{|\eta_1| \leq 1/2\}$ where η_1 is the observation noise at time $t = 1$. Then, we have $Pr(E_1 \cap E_2 \cap E_3) \geq 1-3\delta$, and it is easy to see that the global maximum of the function f under the event $E_1 \cap E_2 \cap E_3$ will lie either at $x = 1/2$ or $x = 5/6$. Since, by construction we have $f(1/2) = -f(5/6)$, a single evaluation of the function at either of these two points is sufficient to find the global maximum, and hence the regret $\mathcal{R}_n \leq \mathcal{O}(1)$.

A.2. Example 2

We observe that the covariance function of the Gaussian Process is upper bounded by $\sum_{i=1}^{\infty} a_i^2$ which for our choice of parameters a_i will be finite. If we make the extra assumption that the noise variance is smaller than a_n^2, we get that $\gamma_n \geq n \log(2)$ which implies that the information type regret bound increases linearly with n.

Now, for a fixed $\delta > 0$, let us define the event

$$E_4 = \{|X_i| \leq \sqrt{2 \log\big(\frac{\pi^2 i^2}{3\delta}\big)} \quad \text{for all } i \geq 1\}.$$

By using the tail bounds for Gaussian random variables and the union bound, we get that $Pr(E_4) \geq 1 - \delta$. We now set the parameters as follows:

$$a_i = \frac{1}{i^2 \sqrt{2 \log\big(\frac{\pi^2 i^2}{3\delta}\big)}} \quad \text{for all } i \geq 1$$
$$\sigma = a_n / \sqrt{2}.$$

Next, suppose $(\eta_t)_{t \geq 1}$ denote the *i.i.d.* $N(0, \sigma^2)$ noise random variables. We define the following event which also occurs with probability at least $1 - \delta$

$$E_5 = \{|\eta_t| \leq 1/(\sqrt{2} t^2) \quad \text{for all } t \geq 1\}.$$

Now, we need to show that there exists a strategy which will ensure with high probability that the cumulative regret is upper bounded by $\mathcal{O}(\log(n))$. Assuming that the events E_4 and E_5 hold (which happens with probability at least $1 - 2\delta$), we proceed as follows:

- We first note that we can construct a ternary tree of intervals ($\{\mathcal{I}_{j,k} : j \geq 0, \text{ and } 1 \leq k \leq 3^j\}$) which form an increasing sequence of partition of the input space $\mathcal{X} = [0, 1]$. The root of the tree is the entire unit interval $\mathcal{I}_{0,1} = [0, 1]$ while the nodes at level 1 are obtained by partitioning $\mathcal{I}_{0,1}$ into three equal intervals $\mathcal{I}_{1,1} = [0, 1/3)$, $\mathcal{I}_{1,2} = [1/3, 2/3)$ and $\mathcal{I}_{1,3} = [2/3, 1]$. This process is repeated indefinitely to get an infinite ternary tree.
- Because of the definition of the Gaussian Process, the function value in the interval $\mathcal{I}_{1,1}$ is $a_1 X_1 \varphi(3x) + f_2(3x)$ and in the interval $\mathcal{I}_{1,3}$ is $-a_1 X_1 \varphi(3x - 2) + f_2(3(x - 2/3))$, we note that x^* must lie either in $\mathcal{I}_{1,1}$ or $\mathcal{I}_{1,3}$. To decide which one, we need to know the sign of X_1 for which we observe the function at the mid point of the interval $\mathcal{I}_{1,2}$. If the observed value is positive, we can conclude that x^* must lie in $\mathcal{I}_{1,1}$. Otherwise, x^* lies in $\mathcal{I}_{1,3}$. Thus our region of uncertainty shrinks from $\mathcal{I}_{0,1}$ to $\mathcal{I}_{1,1}$ or $\mathcal{I}_{1,3}$.
- For $t > 1$, we proceed similarly by evaluating the function at a point x_t in the middle sub-interval of the current region of uncertainty. Based on the observed value, we can infer the sign of $a_t X_t$ which allows us the pick the next subinterval. Thus at any time t, the suboptimality of the evaluated

point is upper bounded by

$$f(x^*) - f(x_t) \leq \sum_{i \geq t} |a_i X_i| + |\eta_i| \leq \sum_{i \geq t} \frac{2}{i^2} \leq \frac{2}{t},$$

where the second inequality follows from the definition of event E_4 and the choice of $(a_i)_{i \geq 1}$.

- Finally, summing up all such terms gives us the required bound on the cumulative regret

$$\mathcal{R}_n \leq \sum_{t=1}^{n} \frac{2}{t} \leq 2\log(n).$$

Appendix B: Deferred proofs from Section 6

B.1. Proof of Proposition 1

Proof. Let $T = B(x_0, r, d)$ and let us assume we have a sequence of increasingly fine discretizations $(T_n)_{n \geq 0}$ of T with $T_0 = \{x_0\}$, and let $\pi_n : T \to T_n$ represent the projection operator onto T_n, i.e., $\pi_n(x) = \arg\min_{y \in T_n} d(x, y)$. Then we have the following:

$$|f(x) - f(x_0)| = |\sum_{n \geq 1} f(\pi_n(x)) - f(\pi_{n-1}(x))| \leq \sum_{n \geq 1} |f(\pi_n(x)) - f(\pi_{n-1}(x))|.$$

Now we use the concentration property of Gaussian Process (2.3) and union bounds, to get:

$$\begin{aligned}
&Pr\big(|f(\pi_n(x)) - f(\pi_{n-1}(x))| > \sqrt{u_n} d(\pi_n(x), \pi_{n-1}(x))\big) \\
&\quad \leq 2\exp(-u_n/2) \\
\Rightarrow\ &Pr\big(\exists x \in T : |f(\pi_n(x)) - f(\pi_{n-1}(x))| > \sqrt{u_n} d(\pi_n(x), \pi_{n-1}(x))\big) \\
&\quad \leq 2|T_n||T_{n-1}|\exp(-u_n/2) \\
\Rightarrow\ &Pr\big(\exists n \in \mathbb{N}, \exists x \in T : |f(\pi_n(x)) - f(\pi_{n-1}(x))| > \sqrt{u_n} d(\pi_n(x), \pi_{n-1}(x))\big) \\
&\quad \leq \sum_{n \geq 1} 2|T_n||T_{n-1}|\exp(-u_n/2) \\
&\quad := P_e.
\end{aligned}$$

Let us define the event $E_1 = \{\exists n \in \mathbb{N}, \exists x \in T : |f(\pi_n(x)) - f(\pi_{n-1}(x))| > \sqrt{u_n} d(\pi_n(x), \pi_{n-1}(x))\}$. Then under the event E_1^c, we know that for all x and n, we have $|f(\pi_n(x)) - f(\pi_{n-1}(x))| \leq \sqrt{u_n} d(\pi_n(x), \pi_{n-1}(x))$, which means that

$$\begin{aligned}
\sup_{x \in T} |f(x) - f(x_0)| &\leq \sup_{x \in T} \sum_{n \geq 1} |f(\pi_n(x)) - f(\pi_{n-1}(x))| \\
&\leq \sup_{x \in T} \sum_{n \geq 1} \sqrt{u_n} d(\pi_n(x), \pi_{n-1}(x)).
\end{aligned}$$

Now, let us choose T_n to be the $\epsilon_n = r2^{-n}$ covering of T with respect to the metric d. Since T has a finite metric dimension D_1', there exists a constant $C_1 < \infty$ depending on $2D_1'$ such that for all $z \leq diam(\mathcal{X})$, we have $N(\mathcal{X}, z, l) \leq C_1 z^{-2D_1'}$. Thus we have $|T_n| \leq C_1 2^{2nD_1'}/r^{2D_1'}$. Now in order to keep P_e below e^{-u} for some $u > 0$, we set $u_n = 2(u + v_n)$ with v_n to be defined later. This gives us

$$P_e \leq \sum_{n\geq 1} 2(C_1^2 2^{2(2n-1)D_1'}/r^{4D_1'})e^{-u_n/2} \leq \sum_{n\geq 1} 2C_1^2 \frac{2^{4nD_1'}}{r^{4D_1'}} e^{-u_n/2}.$$

Now by choosing

$$v_n = \log\left(2C_1^2 \frac{2^{4nD_1'}}{r^{4D_1'}}\right) + \log(n^2\pi^2/6),$$

we get the required bound on P_e. Now it remains to get the upper bound on w_r for this choice of u_n. We use the fact that $d(\pi_n(x), \pi_{n-1}(x)) \leq d(\pi_n(x), x) + d(\pi_{n-1}(x), x) \leq 2r2^{-(n-1)}$ to get

$$w_r \leq 2r \sum_{n\geq 1} 2^{-(n-1)} \times \sqrt{2u + 4D_1' \log(1/r) + 2\log(n^2) + 4nD_1' \log(2) + 2\log(2C_1^2\pi^2/6)}.$$

Finally, replacing $2\log(2C_1^2\pi^2/6)$ with C_2, and writing $\sum_{n\geq 1} 2^{-(n-1)}\sqrt{\log n} = \alpha_1$ and $\sum_{n\geq 1} 2^{-(n-1)}\sqrt{n} = \alpha_2$, we get

$$w_r \leq 4r\big(\sqrt{C_2 + 2u + 4D_1' \log(1/r)} + C_3\big), \tag{B.1}$$

where $C_3 = \alpha_1 + \alpha_2\sqrt{2D_1' \log 2}$. □

B.2. Proof of Proposition 3

Proof. Let $\bar{y}_{1:t-1}$ denote all the observations before time t, and $\bar{y}_x$ be the vector of observations at x. Also, let $\bar{y}_{x^c}$ be the vector of observations at points other than x. Then by the non-negativity of mutual information we have

$$\begin{aligned}
&I(f(x); \bar{y}_{x^c} | \bar{y}_x) \geq 0 \\
\Rightarrow &h(f(x)|\bar{y}_x) - h(f(x)|\bar{y}_x, \bar{y}_{x^c}) \geq 0 \\
\Rightarrow &\log\left(\frac{1}{\sqrt{\frac{n_t(x)}{\sigma^2} + \frac{1}{K(x,x)}}}\right) - \log(\sigma_t(x)) \overset{(a)}{\geq} 0 \\
\Rightarrow &\frac{\sigma}{\sqrt{n_t(x)}} \geq \sigma_t(x),
\end{aligned}$$

where $h(X)$ is the differential entropy of X and $I(X;Y)$ denotes the mutual information between random variables X and Y. For inequality (a), we used the formula for the differential entropy of a Gaussian random variable.

For the second part, let us define $S_x = \{x_1, x_2, \ldots, x_{n_t(x,r)}\}$ as the set of points in $B(x,r,l)$ which have been evaluated up to time t. Further introducing the vector $K_x = [K(x,x_1), K(x,x_2), \ldots, K(x,x_{n_t(x,r)}]^{tr}$ where tr denote the transpose operation, and the matrix $K_{xx} = [K(x_i,x_j)]_{(x_i,x_j)\in S_x \times S_x}$, we have by the formula for the posterior variance at x (Rasmussen and Williams, 2006, (2.26)):

$$\sigma_t^2(x) \leq K(0) - K_x^T(K_{xx} + \sigma^2 I)^{-1} K_x.$$

Now, based on the assumption that K is isotropic, we can make the following two observations,

$$(K_{xx} + \sigma^2 I) \preccurlyeq (K(0)\mathbf{1}\mathbf{1}^T + \sigma^2 I)$$
$$K(r)\mathbf{1} \preccurlyeq K_x,$$

which gives us

$$-K_x^T(K_{xx} + \sigma^2 I)^{-1} K_x \leq K(r)^2 \mathbf{1}^T (K(0)\mathbf{1}\mathbf{1}^T + \sigma^2 I)^{-1}\mathbf{1}. \tag{B.2}$$

Now, using the Woodbury matrix inversion identity, and some simplification, we get:

$$\begin{aligned}
\sigma_t^2(x) &\leq \frac{K(0)\sigma^2 + n_t(x,r)(K(0)^2 - K(r)^2}{\sigma^2 + n_t(x,r)K(0)} \\
\Rightarrow \sigma_t^2(x) &\leq \frac{\sigma^2}{n_t(x,r)} + 2(K(0) - K(r)) \\
\Rightarrow \sigma_t(x) &\overset{(a)}{\leq} \frac{\sigma}{\sqrt{n_t(x,r)}} + d(r) \\
\Rightarrow \sigma_t(x) &\overset{(b)}{\leq} \frac{\sigma}{\sqrt{n_t(x,r)}} + g(r),
\end{aligned}$$

where (a) uses the inequality $\sqrt{z_1 + z_2} \leq \sqrt{z_1} + \sqrt{z_2}$ for $z_1, z_2 \geq 0$ and (b) follows from the fact that $K \in \mathcal{K}$. □

Appendix C: Proof of Theorem 1

For the entirety of this proof, we will assume that the events Ω_{u5} and Ω_{u6} hold, which is true with probability at least $1 - 2e^{-u}$. Let $(\tau_j)_{j\geq 1}^n$ denote the rounds in which the function evaluations were performed, and let $Q_n = \{x_{h_{\tau_j}, i_{\tau_j}} | 1 \leq j \leq n\}$ denote the multiset of points evaluated by the algorithm.

C.1. Information-type bound on $\mathcal{R}_n$

To obtain the information-type cumulative regret bound, we divide the set Q_n into Q_{n1} and Q_{n2}, where

$$Q_{n1} = \{x_{h,i} \in Q_n | h < h_{\max}\}$$

and $Q_{n2} = Q_n \setminus Q_{n1}$. For bounding the regret due to the elements of Q_{n2}, we use the fact that all nodes in Q_{n_2} are at most $(4N_1 + 1)V_{h_{max}}$ suboptimal and then trivially bound $|Q_{n2}|$ with n. For the nodes in Q_{n1}, we can bound the sub-optimality of the points in terms of their predictive variance which allows us to obtain information type bounds for them. The term h_{max} is then chosen to balance the contribution of these two terms to the overall cumulative regret.

From Lemma 1, we know that for all $x_{h,i} \in Q_{n2}$, we have $\Delta(x_{h,i}) \leq (4N_1 + 1)V_h$, and assuming n is large enough so that $h_{\max} \geq h_0 := \frac{\log(v_1/\delta_K)}{\log(1/\rho)}$ (which is true for $n \geq \left(\frac{v_1}{\delta_K}\right)^{2\alpha}$), we can upper bound the contribution of the terms in Q_{n2} to the cumulative regret (denoted by $\mathcal{R}_{n2}$) as follows:

$$\begin{aligned}
\mathcal{R}_{n2} &:= \sum_{x_{h,i} \in Q_{n2}} f(x^*) - f(x_{h,i}) \\
&\leq (4N_1 + 1)V_{h_{\max}} |Q_{n2}| \\
&\leq (4N_1 + 1)V_{h_{\max}} n \\
&\leq \mathcal{O}(\rho^{h_{\max}\alpha} \sqrt{h_{\max}} n),
\end{aligned}$$

where the last inequality relies on the assumption that $h_{\max} \geq h_0$ and the properties of the covariance functions in the class $\mathcal{K}$. Now, using the fact that $h_{\max} \geq \frac{(1/2)\log(n)}{\alpha \log(1/\rho)}$ we get that $\mathcal{R}_{n2} \leq \mathcal{O}(\sqrt{n \log(n)})$.

Now, for the terms x_{h_τ, i_τ} in Q_{n1} we observe from Lemma 1 that $f(x^*) - f(x_{h_\tau,i_\tau}) \leq 3\beta_n \sigma_{\tau-1}(x_{h_\tau,i_\tau})$. If $|Q_{n1}| = n_1$, then by using (Srinivas et al., 2012, Lemma 5.3 and Lemma 5.4), and the assumption that $K(x, x) \leq 1$ for all $x \in \mathcal{X}$, we get

$$\mathcal{R}_{n1} \leq \mathcal{O}(\sqrt{n_1 \gamma_{n_1} \log(n_1)}) \leq \mathcal{O}(\sqrt{n \gamma_n \log(n)}).$$

On adding the two terms, we get the required information type bound $\mathcal{R}_n \leq \mathcal{O}(\sqrt{n\gamma_n \log(n)})$.

C.2. Dimension-type regret bounds

To obtain the dimension-type bounds, we upper bound the number of times the algorithm selects points at some level h of the tree, and the sub-optimality of such points using Lemma 1. Furthermore, by using the near-optimality dimension, we can upper bound the number of unique points evaluated by the algorithm at level h of the tree. Combining these results with an appropriate choice of h_{max}, we obtain the required bounds on cumulative and simple regret of the algorithm.

We first describe the steps for bounding the cumulative regret. Recall that the algorithm only selects points for evaluation from the sets of the form $\mathcal{X}_{(4N_1+1)V_h} = \{x_{h,i} \in \mathcal{X} : f(x^*) - f(x_{h,i}) \leq (4N_1+1)V_h\}$, and furthermore, by the assumption on the metric space that any two points in $\mathcal{X}_h$ are separated by at least $2v_2\rho^h$. These two facts imply that $|\mathcal{X}_{(4N_1+1)V_h} \cap \mathcal{X}_h| \leq M(\mathcal{X}_{(4N_1+1)V_h}, 2v_2\rho^h, l) \leq N(\mathcal{X}_{(4N_1+1)V_h}, v_2\rho^h, l)$.

We first consider the contribution of the terms $x_{h,i}$ for which $h < h_0$:

$$\begin{aligned}
\mathcal{R}_1 &= \sum_{x_{h,i} \in Q_n : h < h_0} f(x^*) - f(x_{h,i}) \\
&\leq \sum_{h=0}^{h_0-1} |\mathcal{X}_{(4N_1+1)V_h} \cap \mathcal{X}_h| q_h \\
&\overset{(a)}{\leq} \mathcal{O}\Big(\sum_{h=0}^{h_0-1} \rho^{-2hD_1} q_h \Big) \\
&\overset{(b)}{\leq} \mathcal{O}\Big(\sum_{h=0}^{h_0-1} \rho^{-2hD_1} \frac{\sigma^2 \beta_n^2}{g(v_2\rho^{h_0})} \Big) \\
&= \mathcal{O}(\log(n)).
\end{aligned}$$

In the above display, the inequality (a) relies on the fact that $\mathcal{X}$ has a finite metric dimension D_1 and thus there exists a constant $C < \infty$ such that for all $r > 0$, we have $N(\mathcal{X}, r, l) \leq Cr^{-2D_1}$. The inequality (b) uses Lemma 1 and the fact that g is a non-decreasing function.

Now, we fix an H such that $h_0 \leq H \leq h_{\max}$. Then for $D' > \tilde{D}$, we have the following:

$$\begin{aligned}
\mathcal{R}_2 &= \sum_{x_{h,i} \in Q_n : h_0 \leq h \leq H} f(x^*) - f(x_{h,i}) \\
&\leq \sum_{h=h_0}^{H} |\mathcal{X}_{(4N_1+1)V_h} \cap \mathcal{X}_h| (4N_1+1) V_h q_h \\
&\leq \sum_{h=h_0}^{H} M(\mathcal{X}_{(4N_1+1)V_h}, 2v_2\rho^h, l) V_h q_h \\
&\overset{(c)}{\leq} \mathcal{O}\Big(\sum_{h=h_0}^{H} \beta_n^2 \rho^{-h(D'+\alpha)} \sqrt{h} \Big) \\
&= \mathcal{O}\Big(\rho^{H(\alpha+D')} \sqrt{H} \log(n) \Big).
\end{aligned}$$

In the inequality (c) above, we use the fact that for $h \geq h_0$, and for any $D' > \tilde{D}$, there exists a $C < \infty$ such that for all $r \leq v_2$, we have $M(\mathcal{X}_{(4N_1+1)V_h}, r, l) \leq Cr^{-D'}$. Furthermore, we upper bound V_h with $\mathcal{O}(\rho^{h\alpha}\sqrt{h})$ and q_h with $\mathcal{O}(\frac{\beta_n^2}{\rho^{2h\alpha}})$ by the assumptions on the covariance function.

Finally, the contribution of the remaining points in Q_n can be trivially upper bounded as:

$$\begin{aligned}\mathcal{R}_3 &= \sum_{x_{h,i}\in Q_n : h\geq H} f(x^*) - f(x_{h,i}) \leq nV_H \\ &\leq \mathcal{O}(n\rho^{H\alpha}\sqrt{H}).\end{aligned}$$

Now, if we select $H = \frac{\log(n)}{\log(1/\rho)\alpha(D+2\alpha)} < h_{\max}$, we get

$$\mathcal{R}_n \leq \mathcal{O}\left(\log(n)^{3/2} n^{1-\frac{1}{D'+2\alpha}}\right)$$

as required.

To obtain the bound on the simple regret, we introduce the terms $Q_h = Q_n \cap \mathcal{X}_h$ for $h \geq 0$. for any $H > 0$, we have the following:

$$\begin{aligned}\sum_{h=0}^{H} |Q_h| &= \sum_{h=0}^{h_0-1} |Q_h| + \sum_{h=h_0}^{H} |Q_h| \leq n \\ &\leq \mathcal{O}(1) + \mathcal{O}\left(\sum_{h=h_0}^{H} \rho^{-h(D'+2\alpha)}\beta_n^2\right) \\ &\leq \mathcal{O}(\rho^{-H(D'+2\alpha)}\beta_n^2).\end{aligned}$$

Now, if we find the largest H (denoted by $\bar{H}$) such that the upper bound on $\sum_{h=0}^{H} |Q_h|$ given above is smaller than n, then $\bar{H}$ will be a lower bound on the maximum depth explored by the algorithm. From the definition of $\bar{H}$, we can show that there exists some constant $C' > 0$, such that

$$\left(\frac{C'\log(n)}{n}\right)^{1/(D'+2\alpha)} \leq \rho^{\bar{H}} \leq \frac{1}{\rho}\left(\frac{C'\log(n)}{n}\right)^{1/(D'+2\alpha)}.$$

Assuming that n is large enough so that $\bar{H} \geq h_0$ and that $h_{\max} \geq \bar{H}$ (which is true if $h_{\max} \geq \frac{\log(n)}{2\alpha\log(1/\rho)}$), we can now upper bound the simple regret as follows:

$$\begin{aligned}\mathcal{S}_n &\leq (4N_1+1)V_{\bar{H}} \leq \mathcal{O}(\rho^{\bar{H}\alpha}\sqrt{\bar{H}}) \\ &\leq \mathcal{O}\big(n^{-\frac{\alpha}{D'+2\alpha}}(\log n)^{\frac{\alpha}{D'+2\alpha}+\frac{1}{2}}\big) \\ &\leq \tilde{\mathcal{O}}\big(n^{-\frac{\alpha}{D'+2\alpha}}\big).\end{aligned}$$

Remark 19. The condition on h_{max} given in (3.4) in Section 3 follows by summing up the two lower bounds on h_{max} required by the two parts of the proof of Theorem 1.

Appendix D: Deferred proofs from Section 5.1

D.1. Details of the algorithm

To complete the description of the algorithm, we need to calculate the terms β_n and the term $(W(r_k))_{k\geq 0}$ for radius $r_k = diam(\mathcal{X})2^{-k}$.

We begin with the following simple claim which gives us the appropriate choice of β_n.

Claim 6. *For the choice of* $\beta_n = \mathcal{O}(\sqrt{(D_1/\alpha+1)\log(n)+u})$*, we have for any* $u > 0$*:*

$$Pr(\Omega_{u3}) \geq 1 - e^{-u}$$

where the event Ω_{u3} *is defined as*

$$\Omega_{u3} = \{\forall t \leq t_n, \forall x \in A_t : |f(x) - \mu_{t-1}(x)| \leq \beta_n \sigma_{t-1}(x)\} \tag{D.1}$$

with t_n *is the (random) number of rounds of the algorithm required for* n *function evaluations.*

Proof. Let $M_n = M(\mathcal{X}, n^{-1/(2\alpha)}, l)$ be the $n^{-1/(2\alpha)}$ packing number of $\mathcal{X}$ with respect to the metric l. Then by the design of the algorithm, at any time t we have $|A_t| \leq M_n$, and also $t_n \leq M_n + n$ almost surely. So, we get by two union bounds:

$$\begin{aligned} Pr(\Omega_{u3}^c) &\leq \sum_{t=1}^{t_n} \sum_{x\in A_t} 2e^{-\beta_n^2/2} \leq 2n^2 2e^{-\beta_n^2/2} \\ &\leq 2M_n(M_n+n)e^{-\beta_n^2/2}. \end{aligned}$$

Now, by using the fact that $\mathcal{X}$ has a finite metric dimension D_1 we have $M_n \leq Cn^{2D_1/(2\alpha)}$ for some constant $C > 0$. This implies that $M_n(M_n+n) \leq C^2 n^{1+2D_1/\alpha}$ for $n \geq 1$.

Thus for any $u > 0$, the choice of $\beta_n = \sqrt{2(u + 2\log(C) + (2D_1/\alpha+1)\log(n))}$ ensures that $Pr(\Omega_{u3}) \geq 1 - e^{-u}$. □

Now, we obtain the terms $W(r_k)$ which denotes a high probability upper bound on the maximum variation in the GP sample within any ball of radius r_k in $\mathcal{X}$.

Claim 7. *Consider the choice of radius values* $r_k = 2^{-k}diam(\mathcal{X})$ *for* $k \geq 0$*. Then we have for any* $u > 0$*,* $Pr(\Omega_{u4}) \geq 1 - e^{-u}$*, where the event* Ω_{u4} *is defined as*

$$\Omega_{u4} = \{\forall k \geq 0, \forall x \in \mathcal{X} : \sup_{y\in B(x,r_k,l)} |f(x) - f(y)| \leq W(r_k)\}. \tag{D.2}$$

The term $W(r_k)$ *is given by*

$$\begin{aligned} W(r_k) &= 2w(2r_k) \\ &\leq 8g(r_k)\big(\sqrt{C_4 + 2u + 2\log(|\mathcal{X}_k|) + 4D_1\log(2^k/diam(\mathcal{X}))} + C_3\big), \end{aligned}$$

with $\mathcal{X}_k$ *being the* r_k *cover of* $\mathcal{X}$*.*

Proof. The result follows immediately by applying Proposition 2 with $\epsilon_k = r_k$ and $R_k = 2r_k$. □

Without loss of generality, we can assume that the diameter of the search space $\mathcal{X}$ is 1. Then, in the expression for $W(r_k)$ above, we can upper bound the term $|\mathcal{X}_k|$ for all k by $C2^{2kD_1}$ due to the assumption of finite metric dimension of $\mathcal{X}$. Thus for all $k \geq 0$ we have $W(r_k) \leq \mathcal{O}(g(r_k)\sqrt{u + 4D_1 k \log(C^2/diam(\mathcal{X}))})$ and in particular for $k \geq \log_2(1/\delta_K)$ we have

$$W(r_k) \leq \mathcal{O}(2^{-k\alpha}\sqrt{u + 4D_1 k \log(C^2/diam(\mathcal{X}))}).$$

Having described the algorithm parameters, we now present an outline of the derivation of the regret bounds for the Bayesian Zooming algorithm. We characterize the properties of the points selected by the algorithm in the following lemma. The proof of the regret bounds can be completed in an analogous manner to the proof of Theorem 1

Lemma 2. *Under the events Ω_{u3} and Ω_{u4}, the following statements are true:*

- *Any point x at which the function is evaluated by the algorithm satisfies*

$$f(x^*) - f(x) \leq 5W(r(x)). \tag{D.3}$$

- *If in round t, the function value is evaluated at a point x with $r(x) > r_{min}$, then we have*

$$f(x^*) - f(x) \leq 3\beta_n \sigma_{t-1}(x). \tag{D.4}$$

- *Any two points x_1 and x_2 which have been evaluated at least k times each must satisfy $l(x_1, x_2) > r_k$.*
- *A point x with radius r_k will be evaluated no more than q_{r_k} times before its radius is shrunk, where q_{r_k} is defined as*

$$q_{r_k} = \frac{\sigma^2 \beta_n^2}{2W(r_k)^2}. \tag{D.5}$$

Proof. The results stated above follow directly from the point selection and refinement strategy used in the algorithm:

- Suppose x_t^* denotes the point which contains the maximizer x^* in its confidence region. Then we have the following:

$$\begin{aligned} f(x^*) &\leq J_t(x_t^*) \leq J_t(x) \\ &\leq f(x) + 2\beta_n \sigma_{t-1}(x) + W(r(x)) \\ &\overset{(a)}{\leq} f(x) + 2W(2r(x)) + W(r(x)) \\ &\leq f(x) + 5W(r(x)), \end{aligned}$$

  where (a) follows from the condition required for shrinking the radius associated with x from $2r(x)$ to $r(x)$, assuming $r(x) < diam(\mathcal{X})$. If $r(x) = diam(\mathcal{X})$ then the inequality is trivially true by the definition of $W(r(x))$.

- If $r(x)$ is strictly greater than r_{min} and the point x is evaluated by the algorithm, then we must have $\beta_n \sigma_{t-1}(x) \geq W(r(x))$ which gives us the required inequality.
- Since the covering oracle only adds points from the uncovered region, the distance between two points with associated radius r_k must be greater than r_k.
- Finally, the maximum number of times a point is evaluated by the algorithm before shrinking the radius is upper bounded by using the result in the first part of Proposition 3 to get the required expression of q_{r_k}. □

Having obtained the above results, we can retrace the steps in the proof of Theorem 1 to obtain similar regret bounds for the Bayesian Zooming algorithm.

Acknowledgements

Tara Javidi would like to thank Galen Reeves for introducing her to the problem studied in this paper. Shubhanshu Shekhar would like to thank Emile Contal for helpful discussions. The authors would also like to thank Jonathan Scarlett, the two anonymous referees, the Associate Editor and the Editor for helpful feedback on the manuscript.

References

Bogunovic, I., Scarlett, J., and Cevher, V. (2016a). Time-varying gaussian process bandit optimization. In *Artificial Intelligence and Statistics*, pages 314–323.

Bogunovic, I., Scarlett, J., Krause, A., and Cevher, V. (2016b). Truncated variance reduction: A unified approach to bayesian optimization and level-set estimation. In *Advances in Neural Information Processing Systems*, pages 1507–1515.

Bubeck, S., Munos, R., and Stoltz, G. (2011a). Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852. MR2814373

Bubeck, S., Munos, R., Stoltz, G., and Szepesvári, C. (2011b). X-armed bandits. *Journal of Machine Learning Research*, 12(May):1655–1695. MR2813150

Bull, A. D. (2011). Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12(Oct):2879–2904. MR2854351

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge university press. MR2409394

Contal, E. (2016). *Statistical learning approaches for global optimization*. PhD thesis, Université Paris-Saclay.

Contal, E., Buffoni, D., Robicquet, A., and Vayatis, N. (2013). Parallel gaussian process optimization with upper confidence bound and pure exploration. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 225–240. Springer.

Contal, E. and Vayatis, N. (2016). Stochastic process bandits: Upper confidence bounds algorithms via generic chaining. *arXiv preprint arXiv:1602.04976*.

Desautels, T., Krause, A., and Burdick, J. W. (2014). Parallelizing exploration-exploitation tradeoffs in gaussian process bandit optimization. *The Journal of Machine Learning Research*, 15(1):3873–3923. MR3317214

Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge.

Kandasamy, K., Dasarathy, G., Oliva, J. B., Schneider, J., and Poczos, B. (2016). Multi-fidelity gaussian process bandit optimisation. *arXiv preprint arXiv:1603.06288*.

Kleinberg, R., Slivkins, A., and Upfal, E. (2013). Bandits and experts in metric spaces. *arXiv preprint arXiv:1312.1277*.

Krause, A. and Ong, C. S. (2011). Contextual gaussian process bandit optimization. In *Advances in Neural Information Processing Systems*, pages 2447–2455.

Munos, R. (2011). Optimistic optimization of a deterministic function without the knowledge of its smoothness. In *NIPS*, pages 783–791.

Munos, R. et al. (2014). From bandits to monte-carlo tree search: The optimistic principle applied to optimization and planning. *Foundations and Trends® in Machine Learning*, 7(1):1–129.

Perchet, V. and Rigollet, P. (2013). The multi-armed bandit problem with covariates. *The Annals of Statistics*, pages 693–721. MR3099118

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*, volume 1. MIT press Cambridge. MR2514435

Rigollet, P. and Zeevi, A. (2010). Nonparametric bandits with covariates. *arXiv preprint arXiv:1003.1630*.

Russo, D. and Van Roy, B. (2014). Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243. MR3279764

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and de Freitas, N. (2016). Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.

Slepian, D. (1961). First passage time for a particular gaussian process. *The Annals of Mathematical Statistics*, 32(2):610–612. MR0125619

Slivkins, A. (2014). Contextual bandits with similarity information. *Journal of Machine Learning Research*, 15:2533–2568. MR3247150

Srinivas, N., Krause, A., Kakade, S. M., and 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. MR2952544

Valko, M., Carpentier, A., and Munos, R. (2013). Stochastic simultaneous optimistic optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 19–27.

van Handel, R. (2014). Probability in high dimension. Technical report, DTIC Document.

Van Handel, R. (2015). Chaining, interpolation, and convexity. *arXiv preprint arXiv:1508.05906*. MR3852183

Wang, Z., Shakibi, B., Jin, L., and Freitas, N. (2014). Bayesian multi-scale optimistic optimization. In *Artificial Intelligence and Statistics*, pages 1005–1014.

Wang, Z., Zhou, B., and Jegelka, S. (2016). Optimization as estimation with gaussian processes in bandit settings. In *Artificial Intelligence and Statistics*, pages 1022–1031.