Lenient Regret and Good-Action Identification in Gaussian Process Bandits

Xu Cai [1] **Selwyn Gomes** [1] **Jonathan Scarlett** [1 2]

Abstract

In this paper, we study the problem of Gaussian process (GP) bandits under relaxed optimization criteria stating that any function value above a certain threshold is "good enough". On the theoretical side, we study various *lenient regret* notions in which all near-optimal actions incur zero penalty, and provide upper bounds on the lenient regret for GP-UCB and an elimination algorithm, circumventing the usual $O(\sqrt{T})$ term (with time horizon T) resulting from zooming extremely close towards the function maximum. In addition, we complement these upper bounds with algorithm-independent lower bounds. On the practical side, we consider the problem of finding a single "good action" according to a known pre-specified threshold, and introduce several good-action identification algorithms that exploit knowledge of the threshold. We experimentally find that such algorithms can often find a good action faster than standard optimization-based approaches.

1. Introduction

Gaussian Process (GP) methods have recently gained popularity as a highly effective tool in finding the optimum $f(\mathbf{x}^*)$ of a black-box function f (Shahriari et al., 2016), with a particularly notable advantage being sample efficiency. Alongside the practical developments, the theory of GP bandits has also seen several interesting advances. The results can broadly be classified according to whether the mathematical model adopted is Bayesian (i.e., the function is assumed to be random and drawn from a GP) or non-Bayesian (i.e., the function is deterministic and assumed to have a bounded norm in a suitably-defined Reproducing Kernel Hilbert Space (RKHS)), and the same GP-based algorithms can often be applied in a unified manner in these two settings.

Perhaps the most prominent class of existing results concerns cumulative regret bounds that scale with the time horizon as $\sqrt{T}$ or higher, and simple regret bounds that show convergence to the optimum at a rate of $\frac{1}{\sqrt{T}}$ or slower (Bogunovic et al., 2016; Chowdhury & Gopalan, 2017; Contal et al., 2013; Janz et al., 2020; Srinivas et al., 2010). While algorithm-independent lower bounds show such behavior to be unavoidable (Scarlett, 2018; Scarlett et al., 2017), their proofs suggest these regret terms are predominantly dictated by the hardness of zooming increasingly close to the locally-quadratic maximum (Bayesian setting), or of finding a very small and narrow bump hidden in an otherwise flat function (RKHS setting). In practice, one may not be concerned with the distinction between being "very close" vs. "extremely close" to the maximum, or one may not mind missing the existence of a very small bump. In this sense, there is potentially a wide gap between standard theoretical guarantees and practical desiderata.

Motivated by these considerations, we investigate theory and algorithms for Gaussian process bandits under various notions that only seek to find "good enough" actions, where an action $\mathbf{x}$ is considered good if $f(\mathbf{x})$ is within a certain threshold $\Delta > 0$ of the optimum $f(\mathbf{x}^*)$. In particular, following a recent work in the multi-armed bandit literature (Merlis & Mannor, 2021) and focusing on the non-Bayesian RKHS setting, we study *lenient regret* notions that incur no penalty for good-actions. We show that this circumvents the $\sqrt{T}$ term appearing (and being unavoidable) in the standard cumulative regret, and that GP-UCB (Srinivas et al., 2010) and an elimination algorithm (Contal et al., 2013) can instead incur a significantly smaller lenient regret such as $\mathrm{poly}(\log T)$, or even just a constant value (i.e, $O(1)$) depending on Δ.

In addition, we consider the related problem of finding a single point whose function value exceeds some pre-specified threshold $\eta > 0$ (we may set $\eta = f(\mathbf{x}^*) - \Delta$), which we call the *good-action identification* problem. This problem may be of interest, for example, in the context of hyperparameter tuning, where narrowing down a near-optimal configuration may be prohibitively expensive, so one may instead resort to seeking a "sufficiently good" configuration.

[1]Department of Computer Science, National University of Singapore [2]Department of Mathematics & Institute of Data Science, National University of Singapore. Correspondence to: Xu Cai <caix@u.nus.edu>, Selwyn Gomes <selwyn@comp.nus.edu.sg>, Jonathan Scarlett <scarlett@comp.nus.edu.sg>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

We connect the good-action identification problem to the notion of lenient regret, and provide novel algorithms that are specifically targeted to this setting and exploit the knowledge of η. We empirically observe that these algorithms can improve on standard optimization-based approaches, using both synthetic and non-synthetic functions.

1.1. Related Work

Theoretical works on GP bandits have focused mainly on the cumulative regret (see (3) below), and in some cases the simple regret (see (4) below). Perhaps most related to our work are the analyses of GP-UCB in (Chowdhury & Gopalan, 2017; Srinivas et al., 2010), and of elimination-based algorithms in (Bogunovic et al., 2016; Contal et al., 2013), as well as the algorithm-independent lower bounds in (Cai & Scarlett, 2021; Scarlett et al., 2017).

The preceding works provide near-tight scaling laws for the squared-exponential (SE) kernel, while incurring larger gaps for the Matérn kernel; however, these gaps have been narrowed in a recent line of works (Janz et al., 2020; Shekhar & Javidi, 2020; Valko et al., 2013). Other theoretical studies include those for the noiseless setting (Bull, 2011; Grünewälder et al., 2010) and the Bayesian setting (Scarlett, 2018; Shekhar & Javidi, 2018), but these are less relevant to the present paper.

Our work is motivated by recent works in the multi-armed bandit (MAB) literature studying various notions of lenient regret (Merlis & Mannor, 2021) and good-arm identification (Kano et al., 2019; Katz-Samuels & Jamieson, 2020). Like with these works, we seek to show that such notions can be attained with significantly fewer samples; however, the associated algorithms, results, and analyses have minimal similarity with these works, due to the very different continuous action space along with smoothness assumptions.

Some works on GP bandits have sought to incorporate prior information such as monotonicity (Li et al., 2017) and knowledge of the function maximum (Nguyen & Osborne, 2020), but to our knowledge, none have considered notions relating to lenient regret and good-action identification.

Finally, the problem of identifying an action whose function value exceeds a given threshold is related to *level-set estimation* (LSE), which has been studied using GP methods (Bogunovic et al., 2016; Bryan et al., 2006; Gotovos et al., 2013; Shekhar & Javidi, 2019). However, the goal of LSE is to classify the *entire domain* into points falling above/below the threshold, whereas our focus is on finding just a single point above the threshold. Thus, applying LSE methods to our setting would amount to unnecessarily solving a harder problem as an intermediate step.

2. Problem Setup

We consider the problem of sequentially optimizing an unknown function f on a compact domain D, taking $D = [0,1]^d$ for concreteness. In each round indexed by $t = 1, \dots, T$, the algorithm selects $\mathbf{x}_t \in D$ and observes a noisy sample $y_t = f(\mathbf{x}_t) + z_t$, with $z_t \sim \mathcal{N}(0, \sigma^2)$.

We focus on the non-Bayesian RKHS setting (briefly turning to the Bayesian setting in Section 3.4), adopting the assumption that $f \in \mathcal{F}_k(B)$, where $\mathcal{F}_k(B)$ denotes the set of all functions whose RKHS norm $\|f\|_k$ is upper bounded by some constant $B > 0$. We consider arbitrary choices of the kernel $k(\mathbf{x}, \mathbf{x}')$ for the most part, but will sometimes pay particular attention to the squared exponential (SE) and Matérn kernels (Rasmussen, 2006), parametrized by the length-scale l (both cases) and the smoothness parameter ν (Matérn only). Throughout the paper, we assume normalization such that $k(\mathbf{x}, \mathbf{x}) \le 1$ for all $\mathbf{x} \in D$, with equality for the SE and Matérn kernels.

Despite considering the non-Bayesian RKHS setting, it is useful to consider a 'fictitious' Bayesian GP posterior: Given a sequence of inputs $(\mathbf{x}_1, \dots, \mathbf{x}_t)$ and their noisy observations $(y_1, \dots, y_t)$, the posterior distribution under a $\mathrm{GP}(\mathbf{0}, k)$ prior and $\mathcal{N}(0, \lambda)$ sampling noise[1] is also Gaussian, with mean and variance given by

$$\mu_t(\mathbf{x}) = \mathbf{k}_t(\mathbf{x})^T \big(\mathbf{K}_t + \lambda \mathbf{I}_t\big)^{-1} \mathbf{y}_t, \tag{1}$$

$$\sigma_t^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_t(\mathbf{x})^T \big(\mathbf{K}_t + \lambda \mathbf{I}_t\big)^{-1} \mathbf{k}_t(\mathbf{x}), \tag{2}$$

where $\mathbf{k}_t(\mathbf{x}) = \big[k(\mathbf{x}_i, \mathbf{x})\big]_{i=1}^t$, and $\mathbf{K}_t = \big[k(\mathbf{x}_t, \mathbf{x}_{t'})\big]_{t,t'}$ is the kernel matrix.

The most widely-adopted performance measure in the literature is the (standard) cumulative regret, defined as

$$R_T = \sum_{t=1}^T \big(f(\mathbf{x}^*) - f(\mathbf{x}_t)\big), \tag{3}$$

where $\mathbf{x}^*$ denotes any maximizer of f. Another popular notion is the *simple regret* $r^{(T)}$, in which the algorithm returns an additional point $\mathbf{x}^{(T)}$ (not necessarily a sampled one) after T rounds, and

$$r^{(T)} = f(\mathbf{x}^*) - f(\mathbf{x}^{(T)}). \tag{4}$$

2.1. Lenient Regret

In light of the motivation in the introduction, and following recent study of (Merlis & Mannor, 2021) for the multi-armed bandit setting, we consider notions of *lenient regret* in which no penalty is incurred when $f(\mathbf{x}_t)$ is within Δ of the

[1] Since this is a fictitious update model, the parameter λ may differ from the true noise variance σ^2.

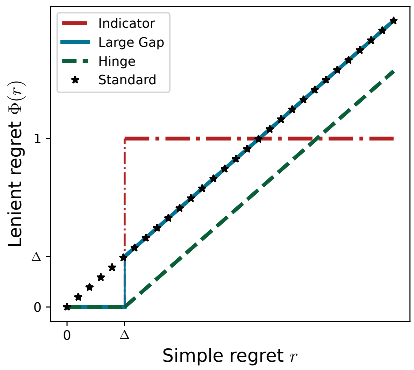


Figure 1. Illustration of three choices of Φ for the lenient regret, along with the choice that yields the standard regret.

optimum, for some small $\Delta > 0$. In view of this property, we henceforth refer to $\mathbf{x} \in D$ satisfying $f(\mathbf{x}) \geq f(\mathbf{x}^*) - \Delta$ as *good actions*, and to other $\mathbf{x}$ as *bad actions*.

In generic notation, we consider (cumulative) lenient regret notions of the form

$$\widetilde{R}_T = \sum_{t=1}^{T} \Phi(r_t), \quad r_t = f(\mathbf{x}^*) - f(\mathbf{x}_t) \tag{5}$$

for some function $\Phi(\cdot)$ such that $\Phi(r) = 0$ for all $r \leq \Delta$ (whereas $\Phi(r) = r$ would recover (3)).

We focus our attention on the following three choices of Φ suggested in (Merlis & Mannor, 2021):

- **Indicator:** $\Phi^{\text{ind}}(r) = \mathbb{1}\{r > \Delta\}$, implying that $\widetilde{R}_T^{\text{ind}}$ counts the number of bad actions.
- **Large Gap:** $\Phi^{\text{gap}}(r) = r \cdot \mathbb{1}\{r > \Delta\}$, implying that $\widetilde{R}_T^{\text{gap}}$ only accumulates the simple regret of bad actions.
- **Hinge:** $\Phi^{\text{hinge}}(r) = \max(r - \Delta, 0)$, implying that $\widetilde{R}_T^{\text{hinge}}$ accumulates the distances of bad actions' function values to the good-action threshold.

These functions are illustrated in Figure 1. Intuitively, one might expect the large-gap regret and hinge regret to behave similarly when Δ is small, whereas the indicator regret may be larger due to the rapid transition from zero to one; our theory will support this intuition.

2.2. Good-Action Identification

In addition to the above lenient regret notions that increase in a cumulative manner, it is also of interest to consider the case that the algorithm is only required to return a single point, and is considered successful if that point is a good action (i.e., its function value is within Δ of the optimum). If the time horizon T is fixed and the returned point is $\mathbf{x}^{(T)}$, then this is equivalent to attaining simple regret at most Δ (see (4)). Since several theoretical guarantees are already known for the simple regret (e.g., see (Bogunovic et al., 2016; Scarlett et al., 2017; Shekhar & Javidi, 2018)), we do not explore them further in this paper, though analogous guarantees can indeed be inferred via simple modifications to our lenient regret analysis.

Instead, in order to move further beyond what is already known, we consider the problem of *fixed-threshold good-arm identification*, where an action $\mathbf{x} \in D$ is considered *good* if $f(\mathbf{x}^*) \geq \eta$ for some pre-specified threshold $\eta > 0$, and *bad* otherwise. This coincides with our above notion of "good" and "bad" actions when $\eta = f(\mathbf{x}^*) - \Delta$.

On the other hand, in contrast to our studies of lenient regret, when η is pre-specified, it is natural to assume that it is *known to the algorithm*. Thus, in Section 4, we introduce algorithms for good-action identification that exploit the prior knowledge of η, and provide experimental evidence that this can be beneficial in Section 5.

2.3. GP-UCB Algorithm

In our study of the lenient regret, we focus on the widely-considered Gaussian process upper confidence bound (GP-UCB) algorithm (Srinivas et al., 2010), which selects the t-th point $\mathbf{x}_t$ to maximize the acquisition function

$$\alpha_t^{\text{UCB}}(\mathbf{x}) = \mu_{t-1}(\mathbf{x}) + \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}), \tag{6}$$

for some suitably-chosen exploration parameter β_t. We use the following well-known result (Abbasi-Yadkori, 2013) (see also (Chowdhury & Gopalan, 2017)) to select β_t. Here and subsequently, we make use of the *maximum information gain*, which is widely used in the GP bandit literature, and is defined as

$$\gamma_t = \max_{\mathbf{x}_1,\dots,\mathbf{x}_t} \frac{1}{2} \ln\det(\mathbf{I}_t + \lambda^{-1}\mathbf{K}_t) \tag{7}$$

with $\mathbf{K}_t$ defined following (2).

Lemma 1. (Abbasi-Yadkori, 2013) *For any $\lambda > 0$ and $f \in \mathcal{F}_k(D)$ with $\|f\|_k \leq B$, under the choice*[2]

$$\beta_t^{1/2} = B + \sigma\lambda^{-1/2}\sqrt{2(\gamma_{t-1} + \ln(1/\delta))}, \tag{8}$$

we have with probability at least $1-\delta$ that $\text{lcb}_t(\mathbf{x}) \leq f(\mathbf{x}) \leq \text{ucb}_t(\mathbf{x})$ for all t and $\mathbf{x} \in D$, where

$$\text{ucb}_t(\mathbf{x}) = \mu_{t-1}(\mathbf{x}) + \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}), \tag{9}$$

$$\text{lcb}_t(\mathbf{x}) = \mu_{t-1}(\mathbf{x}) - \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}). \tag{10}$$

and where $\mu_{t-1}(\cdot)$ and $\sigma_{t-1}(\cdot)$ are given in (1)–(2).

[2] We follow the convention of (Srinivas et al., 2010) and equate this expression with $\beta_t^{1/2}$, whereas some other works denote the right-hand side by β_t.

2.4. Elimination Algorithm

In addition to GP-UCB, we consider a simple algorithm that selects actions with the maximum uncertainty, while using the confidence bounds to eliminate suboptimal actions. While we are not aware of this exact algorithm being used before, it is of a very standard form, and can be viewed as a simplified variant of elimination algorithms such as GP-UCB-PE (Contal et al., 2013) and truncated variance reduction (Bogunovic et al., 2016).

The idea is to define a set of *potential maximizers*

$$M_t = \Big\{ \mathbf{x} \in M_{t-1} \,:\, \mathrm{ucb}_t(\mathbf{x}) \ge \max_{\mathbf{x}'} \mathrm{lcb}_t(\mathbf{x}') \Big\} \qquad (11)$$

and observe that when the UCB and LCB functions in (9)–(10) provide valid confidence bounds, M_t contains $\mathbf{x}^*$ while also eliminating suboptimal points.

With the above definitions in place, the algorithm initializes $M_0 = D$ and $t = 1$, and repeats the following:

(i) Select $\mathbf{x}_t = \arg\max_{\mathbf{x} \in M_{t-1}} \sigma_{t-1}(\mathbf{x})$;

(ii) Observe y_t and update the posterior (i.e., $\mu_t(\cdot)$ and $\sigma_t(\cdot)$) and set of potential maximizers (i.e., M_t in (11)), and increment t.

3. Lenient Regret Bounds

In this section, we provide our main theoretical results on the lenient regret of GP-UCB and the elimination algorithm. The proofs are deferred to Appendix B.

3.1. Lenient Regret of GP-UCB

Our first main result is as follows.

Theorem 1. (Lenient Regret of GP-UCB) *Define*

$$N_{\max} = \max \Big\{ N \,:\, N \le \frac{C_1 \gamma_N \beta_T}{\Delta^2} \Big\}, \qquad (12)$$

where $C_1 = \frac{8\lambda^{-1}}{\log(1+\lambda^{-1})}$. *For any* $f \in \mathcal{F}_k(B)$ *and any* $\delta \in (0,1)$, $\lambda > 0$, *and* $\Delta > 0$, *GP-UCB run with the choice of* β_t *in* (8) *satisfies the following lenient regret bounds with probability at least* $1-\delta$*:*

(i) $\widetilde{R}_T^{\mathrm{ind}} \le N_{\max}$*;*

(ii) $\widetilde{R}_T^{\mathrm{hinge}} \le \widetilde{R}_T^{\mathrm{gap}} \le \frac{C_1 \gamma_{N_{\max}} \beta_T}{\Delta}$.

Specialization to SE and Matérn kernels. To bound $N_{\max}$ under the widely-considered SE and Matérn kernels, we use the following known bounds on γ_t:

- For the SE kernel, we have $\gamma_t = O^*((\log t)^d)$ (Srinivas et al., 2010);
- For the Matérn-ν kernel, we have $\gamma_t = O^*(t^{\frac{d}{2\nu+d}})$ (Vakili et al., 2021).

Here and subsequently, $O^*(\cdot)$ hides dimension-independent logarithmic factors, and will also hide $\log\log T$ factors in expressions for which $\log T$ factors are present. In addition, we treat B, σ, λ, d, l, and ν as being constant as T increases.

We have from (8) that $\beta_t = \Theta(\gamma_t)$, and hence, the condition defining $N_{\max}$ in (12) weakens to $\frac{N}{(\log N)^d} \le O^*\big(\frac{(\log T)^d}{\Delta^2}\big)$. $N_{\max}$ is upper bounded by the N for which this expression holds with equality; from this fact, we can deduce that $\log N = \Theta\big(\log \frac{(\log T)^d}{\Delta^2}\big) = O\big(\log \frac{1}{\Delta} + \log\log T\big)$, and hence $N_{\max} \le O^*\big(\frac{(\log T \cdot \log \frac{1}{\Delta})^d}{\Delta^2}\big)$.

For the Matérn-ν kernel, assuming $d < 2\nu$, the condition defining $N_{\max}$ in (12) weakens to $N^{\frac{2\nu}{2\nu+d}} \le O^*\big(\frac{T^{\frac{d}{2\nu+d}}}{\Delta^2}\big)$, and we obtain $N_{\max} \le O^*\big(\frac{T^{\frac{d}{2\nu}}}{\Delta^{2+\frac{d}{\nu}}}\big)$.

These bounds on $N_{\max}$ directly bound $\widetilde{R}_T^{\mathrm{ind}}$, and can also be substituted into Theorem 1 to deduce similar (albeit more complicated) bounds on $\widetilde{R}_T^{\mathrm{hinge}}$ and $\widetilde{R}_T^{\mathrm{gap}}$; in particular, the dominant term is $\frac{(\log T)^d}{\Delta^2}$ for the SE kernel.

Comparison to standard regret bounds. The lenient regret bounds can be considerably smaller than the $O(\sqrt{T\gamma_T\beta_T})$ standard cumulative regret bounds for GP-UCB (Chowdhury & Gopalan, 2017). For instance, for the SE kernel, the reduction is from $\sqrt{T}\mathrm{poly}(\log T)$ to simply $\mathrm{poly}(\log \mathrm{T})$. More generally, we notice that the standard regret bound is only sublinear when $\gamma_T\beta_T = o(T)$, and limiting our attention to this regime along with $\Delta = \Theta(1)$, we immediately deduce from (12) that $N_{\max} = o(T)$, which in turn implies that the bound $\widetilde{R}_T^{\mathrm{gap}} \le \frac{C_1 \gamma_{N_{\max}} \beta_T}{\Delta}$ is at most $O(\gamma_T\beta_T)$ (and possibly much smaller), which is itself much smaller than $\sqrt{T\gamma_T\beta_T}$ (since $\sqrt{\gamma_T\beta_T} = o(\sqrt{T})$).

Discussion. While Theorem 1 indicates that the lenient regret of GP-UCB can be much smaller than the standard regret, it still grows unbounded as $T \to \infty$, due to the presence of β_T. It is conceivable that an algorithm could have *bounded* lenient regret with high probability, if it manages to find a region of points within Δ of the optimum and subsequently only samples in that region. However, GP-UCB will not satisfy such a property when $\lim_{t\to\infty} \beta_t = \infty$ (as is the case for all known variants with theoretical guarantees), since the growing exploration constant ensures that even suboptimal regions are returned to after long enough.[3]

3.2. Lenient Regret of the Elimination Algorithm

In light of the limitations of Theorem 1 discussed above, we present the following improved lenient regret bounds for the elimination algorithm.

[3] In Appendix A, we discuss the possibility of using GP-UCB with confidence bounds intersected across time.

Theorem 2. (Lenient Regret of the Elimination Algorithm) *Define*

$$N'_{\max} = \max\Big\{N \,:\, N \le \frac{4C_1\gamma_N\beta_N}{\Delta^2}\Big\}. \qquad (13)$$

For any $f \in \mathcal{F}_k(B)$ and any $\delta \in (0,1)$, $\lambda > 0$, and $\Delta > 0$, the elimination algorithm in Section 2.4 run with the UCB and LCB functions in Lemma 1 satisfies the following lenient regret bounds with probability at least $1-\delta$:

(i) $\widetilde{R}_T^{\rm ind} \le N'_{\max}$;

(ii) $\widetilde{R}_T^{\rm hinge} \le \widetilde{R}_T^{\rm gap} \le 2B + \frac{8C_1\gamma_{N'_{\max}}\beta_{N'_{\max}}}{\Delta}$;

where $C_1 = \frac{8\lambda^{-1}}{\log(1+\lambda^{-1})}$.

The main difference compared to Theorem 1 is that β_T in (12) is replaced by $4\beta_N$. The latter is highly preferable, since we have $N_{\max} = o(T)$ in the scaling regimes of interest, as discussed following Theorem 1. In particular, the regret bounds are now independent of T, with the intuition being that all bad actions are eventually eliminated.

However, this improvement has an important practical caveat, namely, the algorithm may degrade much less gracefully than GP-UCB when the kernel is unknown or learned online. This is because kernel mismatch in the earlier rounds may lead to $\mathbf{x}^*$ being eliminated, and in principle even the entire domain could get eliminated. In view of this trade-off, better understanding the interaction between kernel uncertainty and lenient regret remains an interesting direction for future work.

Specialization to the SE and Matérn kernels. Following a similar argument to the one following Theorem 1, we find that $N_{\max} \le O^*\big(\frac{(\log\frac{1}{\Delta})^{2d}}{\Delta^2}\big)$. For the Matérn kernel, we require $\frac{d}{2\nu+d} < \frac{1}{2}$ (or equivalently, $d < 2\nu$) for $N'_{\max}$ to be finite; note that analogous constraints are also required for the optimization regret bounds in (Chowdhury & Gopalan, 2017; Srinivas et al., 2010) to be non-trivial. When $d < 2\nu$, some simple manipulations give $N'_{\max} \le O^*\big(\frac{1}{\Delta^{2(1+\frac{d}{2\nu-d})}}\big)$, in particular becoming closer to $\frac{1}{\Delta^2}$ as ν increases.

3.3. Algorithm-Independent Lower Bounds

Lower bounds on the standard regret for noisy GP bandit optimization were introduced in (Scarlett et al., 2017), and were refined in (Cai & Scarlett, 2021) via a distinct but related analysis. The idea is to consider functions with a small "bump" that is hard for the algorithm to locate, with the height of the bump being tuned to attain the best possible cumulative regret lower bound. It turns out that the analysis techniques of (Cai & Scarlett, 2021) readily transfer to the setting of lenient regret, but with a larger bump height (namely, $O(\Delta)$) in order to prevent the scenario of trivially having zero lenient regret regardless of the points chosen. This yields the following.

Theorem 3. (Lower Bounds on the Lenient Regret) *Fix $\delta \in \big(0, \frac{1}{3}\big)$, $\Delta \in \big(0, \frac{1}{2}\big)$, $B > 0$, and $T \in \mathbb{Z}$, and suppose that $\frac{\Delta}{B} = O(1)$ with a sufficiently small implied constant,*[4] *and that the dimension d and kernel parameters are constant. Then, for any algorithm, the lenient regret must be lower bounded as follows:*

- *For the SE kernel, there exists $f \in \mathcal{F}_k(B)$ such that the following holds with probability at least δ:*[5]
  - *(i)* $\widetilde{R}_T^{\rm ind} \ge \Omega\big(\min\big\{T, \frac{\sigma^2}{\Delta^2}\big(\log\frac{B}{\Delta}\big)^{d/2}\log\frac{1}{\delta}\big\}\big)$;
  - *(ii)* $\widetilde{R}_T^{\rm hinge} \ge \Omega\big(\min\big\{T\Delta, \frac{\sigma^2}{\Delta}\big(\log\frac{B}{\Delta}\big)^{d/2}\log\frac{1}{\delta}\big\}\big)$ *(and $\widetilde{R}_T^{\rm gap} \ge \widetilde{R}_T^{\rm hinge}$).*
- *For the Matérn kernel, there exists $f \in \mathcal{F}_k(B)$ such that the following holds with probability at least δ:*
  - *(i)* $\widetilde{R}_T^{\rm ind} \ge \Omega\big(\min\big\{T, \frac{\sigma^2}{\Delta^2}\big(\frac{B}{\Delta}\big)^{d/\nu}\log\frac{1}{\delta}\big\}\big)$;
  - *(ii)* $\widetilde{R}_T^{\rm hinge} \ge \Omega\big(\min\big\{T\Delta, \frac{\sigma^2}{\Delta}\big(\frac{B}{\Delta}\big)^{d/\nu}\log\frac{1}{\delta}\big\}\big)$ *(and $\widetilde{R}_T^{\rm gap} \ge \widetilde{R}_T^{\rm hinge}$).*

To compare with the upper bounds in Theorem 2, we again treat B, σ^2, and δ as constants, focusing on the dependence on Δ. In addition, we focus on the scaling regimes of primary interest in which each $\min\{\cdot,\cdot\}$ is achieved by the second term (in the other case, there are $\Theta(T)$ bad arm pulls, which is analogous to the standard regret being linear in T).

For the SE kernel, the upper and lower bounds match up to the replacement of $d/2$ by $2d$ in the exponent, and thus, we have proved that $\frac{1}{\Delta^2}$ (for $\widetilde{R}_T^{\rm ind}$) or $\frac{1}{\Delta}$ (for $\widetilde{R}_T^{\rm gap}$ and $\widetilde{R}_T^{\rm hinge}$) is indeed the correct leading term.

For the Matérn kernel, wider gaps remain between the upper and lower bounds, as is also the case for the standard cumulative regret of GP-UCB (Chowdhury & Gopalan, 2017) and arm elimination (Contal et al., 2013) compared to the lower bounds (Scarlett et al., 2017). These gaps for the standard cumulative regret can be closed using the impractical SupKernelUCB algorithm (Valko et al., 2013), or partially closed using covering techniques that remain effective in practice (Janz et al., 2020). However, these algorithms are also more difficult to analyze, and would likely need further modifications to remove the dependence on T in the same way as Theorem 2. Hence, the analysis of their lenient regret is left for possible future work.

[4] Note that if $\Delta > 2B$ then the lenient regret is trivially zero, since any $f \in \mathcal{F}_k(B)$ must have $\max_x |f(x)| \le B$.

[5] We state our lower bounds as failure events that hold with probability at least δ, which is equivalent to saying that all algorithms are unable to attain a success probability of $1-\delta$.

3.4. Upper Bounds for the Bayesian Setting

Throughout the paper, we have focused on the non-Bayesian setting in which $f \in \mathcal{F}_k(B)$. However, since our upper bounds are centered around the validity of the confidence bounds in Lemma 1, they also naturally extend to the Bayesian setting in which $f \sim \mathrm{GP}(\mathbf{0}, k)$ and the exploration constants β_t are suitably modified. This is most straightforward in the finite-domain setting, in which we can set $\beta_t = 2 \log \frac{|D| t^2 \pi^2}{6\delta}$ (Srinivas et al., 2010), along with $\lambda = \sigma^2$ in (1)–(2).

In the continuous-domain Bayesian setting, the changes are slightly less straightforward, but we can again follow (Srinivas et al., 2010) under the assumption of the sample paths being Lipschitz-continuous with high probability. The analysis (but not the algorithm) then makes use of a discretization argument that slightly increases the uncertainty of any given point in the analysis. This added uncertainty amounts to replacing Δ by $\Delta - \epsilon$ for arbitrarily small $\epsilon > 0$ in the bounds, having a negligible impact for any fixed $\Delta > 0$. If Δ is considered to be decreasing as T increases, then the analysis can additionally be modified so that ϵ decreases. The details are omitted for the sake of brevity.

4. Good-Action Identification Algorithms

Our theory suggests that the GP-UCB algorithm, which was introduced for studying the standard regret notion (Srinivas et al., 2010), is also effective in finding "good enough" actions, either according to the lenient regret with parameter Δ or the fixed-threshold setting with parameter η. In this section, we complement our theory by introducing additional practical algorithms that are specifically geared towards the fixed-threshold setting, and explicitly incorporate knowledge of the threshold η with the goal of finding a point satisfying $f(x) \geq \eta$. Experimental evaluations will be performed in Section 5.

4.1. Probability of Being Good (PG)

The early work of (Kushner, 1964) suggested to choose the next query point as the one which has the highest *probability of improvement* (PI) over the current maximum $f(\mathbf{x}^+)$, where $\mathbf{x}^+ = \arg\max_{\mathbf{x} \in \{\mathbf{x}_1, \dots, \mathbf{x}_{t-1}\}} f(\mathbf{x})$. Motivated by this idea, we consider choosing the action as the one having the highest *probability of being good* (PG):

$$\alpha_t^{\mathrm{PG}}(\mathbf{x}) = \mathbb{P}_{t-1}[f(\mathbf{x}) \geq \eta] = \Phi\Big(\frac{\mu_{t-1}(\mathbf{x}) - \eta}{\sigma_{t-1}(\mathbf{x})}\Big), \quad (14)$$

where $\mathbb{P}_{t-1}[\cdot]$ denotes the posterior probability after $t-1$ queries (and subsequently similarly for $\mathbb{E}_{t-1}[\cdot]$), and $\Phi(\cdot)$ denotes the cumulative density function (CDF) of the standard Gaussian distribution.

Since $\Phi(\cdot)$ is an increasing function, we can equivalently maximize the argument $\frac{\mu_{t-1}(\mathbf{x}) - \eta}{\sigma_{t-1}(\mathbf{x})}$ in (14); this is more numerically stable due to avoiding very small $\Phi(\cdot)$ values.

4.2. Expected Improvement Over Good (EG)

By choosing the next query point as the one having the highest *expected improvement* (EI) over the current maximum $f(\mathbf{x}^+)$, one can account for the *amount* of improvement into consideration, rather than just the probability of improvement (Mockus et al., 1978). While any good action is considered sufficient in our setting, it is still natural to analogously consider the *expected improvement over good* (EG) selection rule:

$$\begin{aligned} \alpha_t^{\mathrm{EG}}(\mathbf{x}) &= \mathbb{E}_{t-1}[\max\{0, f(\mathbf{x}) - \eta\}] \\ &= (\mu_{t-1}(\mathbf{x}) - \eta)\Phi(u_{\mathbf{x}}) + \sigma_{t-1}(\mathbf{x})\phi(u_{\mathbf{x}}), \quad (15) \end{aligned}$$

where $u_{\mathbf{x}} = \frac{\mu_{t-1}(\mathbf{x}) - \eta}{\sigma_{t-1}(\mathbf{x})}$, and ϕ denotes the probability density function (PDF) of the standard Gaussian distribution.

4.3. Good-Action Search (GS)

Motivated by the success of entropy search and its variants (Hennig & Schuler, 2012; Hernández-Lobato et al., 2014; Wang & Jegelka, 2017), we can consider being "less myopic" and looking forward one step based on the current posterior. Specifically, if we consider choosing $\mathbf{x}$ as the next point, then the resulting y will be random, and appending $(\mathbf{x}, y)$ to the data set will form a new posterior (μ_t, σ_t). We can then consider seeking to maximize $\mathbb{E}_y\big[\mathbb{P}_t[y_0^* \geq \eta]\big]$, where $y_0^* = \max_{\mathbf{x}} f(\mathbf{x})$,[6] y is distributed according to the current posterior, and $\mathbb{P}_t[\cdot]$ implicitly depends on $(\mathbf{x}, y)$ and represents the updated posterior.

Since the exact computation of $\mathbb{E}_y\big[\mathbb{P}_t[y^* \geq \eta]\big]$ is difficult, we can instead use a surrogate based on randomly-drawn samples as follows:

$$\alpha_t^{\mathrm{GS}}(\mathbf{x}) = \frac{1}{K} \sum_{y_0^* \in Y_0^*} \mathbb{1}\{y_0^* \geq \eta\}, \quad (16)$$

where Y_0^* a set of K samples of maximum function values upon choosing $\mathbf{x}$, which can be generated in an identical manner to max-value entropy search (MES) via a Gumbel distribution approximation (Wang & Jegelka, 2017).

It may be the case that all of the $\mathbf{x}$ lead to a set Y_0^* in which all of the points are below η; when this occurs, we choose the $\mathbf{x}$ that produced the highest value of $\max_{y_0^* \in Y_0^*} y_0^*$.

4.4. Other Algorithms

In Appendix C, we additionally present two good-action identification algorithms that build on (i) Thompson sampling and (ii) action elimination. However, as discussed

[6] The subscript of 0 is used to emphasize representing a function value *before* adding noise.

therein, these algorithms appear to rely more heavily on prior knowledge that is typically unavailable, and so we omit them from our experiments in the following section.

5. Experiments

In this section, we experimentally evaluate our proposed algorithms alongside several standard baselines.[7] We first provide a simple proof-of-concept experiment to support our theoretical findings on the lenient regret, but we pay significantly more attention to evaluating the good-action identification algorithms proposed in Section 4, since these are designed for practical (rather than theoretical) purposes.

5.1. Behavior of the Lenient Regret

In this experiment (but not later ones), we consider the case of *fixed and known* kernel hyperparameters, since our theory assumes this. Since the theoretical choice of β_t is known to be overly conservative (Srinivas et al., 2010), we manually set $\beta_t^{1/2} = \sqrt{\log(2t)^3}$ in both algorithms. We fix $|D| = 2500$ points by discretizing $[0,1]^2$ to a 50×50 grid.

Figure 2 plots the standard and lenient regret for a 2D synthetic GP function drawn using the SE kernel with parameter $l = 0.1$ and $\sigma^{\mathrm{SE}} = 1$. We set the noise level to $\sigma = 0.02$, and the lenient regret parameter as $\Delta = 0.6$, with the latter choice being made in order to form two disjoint regions of good actions. We see that GP-UCB and the elimination algorithm initially behave similarly, but the lenient regret for the latter completely flattens out by time 700, whereas the lenient regret GP-UCB only remains gradually increasing, and the standard regret remains more significantly increasing. This behavior is consistent with Theorems 1 and 2.

We emphasize that elimination crucially depends on having strong prior knowledge of the kernel, hence performing slightly better here. However, we will see in the following sections that GP-UCB remains effective even without such prior knowledge.

5.2. Good-Action Identification Setup

GP model. We adopt the SE kernel with tunable hyperparameters (lengthscale l and scale σ^{SE}).[8] The hyperparameters are updated every 3 iterations by optimizing the log-likelihood (Rasmussen, 2006) within the range $l \in [10^{-3}, 1]$ and $\sigma^{\mathrm{SE}} \in [5 \times 10^{-2}, 1.5]$ using the built-in SciPy optimizer based on L-BFGS-B.

Choice of good-action threshold. In certain cases, we manually set η and specify its value, whereas in other cases,

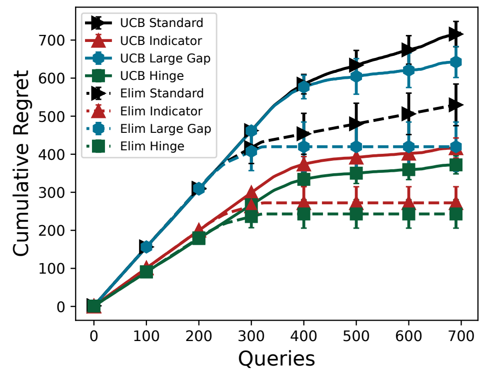


Figure 2. Standard and lenient regret for a 2D synthetic GP.

we select η such that roughly a fraction $\xi \in (0,1)$ of the domain lies above the threshold. To do so, we uniformly sample 10,000 actions and take the empirical ξ-quantile of their function values.

Optimization algorithms. Along with the good-action identification algorithms introduced in Section 4, we evaluate the performance of several optimization baselines (Shahriari et al., 2016; Wang & Jegelka, 2017), namely, GP-UCB, PI, EI, Thompson sampling (TS), and MES. For GP-UCB, we set $\beta_t^{1/2} = \sqrt{\log t}$,[9] which we found to provide a suitable exploration/exploitation trade-off.

Other details. To simplify the experimental evaluation, we focus primarily on noiseless function evaluations, but a noisy setting will also be considered in Section 5.5. We optimize the acquisition functions using the built-in SciPy optimizer with 10 random restarts. In the case of integer-valued variables, we work on the continuous space and round the decimal to the nearest integer.

Evaluation. Except where stated otherwise, we evaluate the performance by computing the proportion of runs for which a good action was found up to the indicated time. We perform 25 trials with 10 experiments each, with each experiment generating a fresh random initial set of 3 points to sample (common to all algorithms). The mean and standard deviation are then computed across trials, with error bars indicating half of a standard deviation.

5.3. Noiseless Synthetic Functions

We consider a variety of widely-used synthetic functions whose descriptions can be found at (Bingham, 2021). Here the threshold η is chosen so that (roughly) a $\xi = \frac{1}{100}$ fraction of points are good; the effect of varying ξ is explored in Appendix D.1. The results are shown in Figure 3.

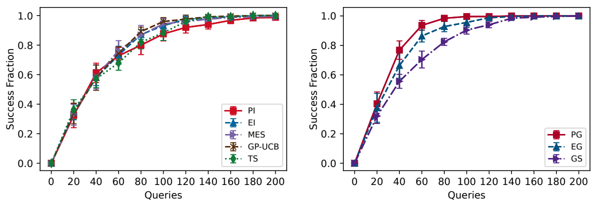
These experiments indicate that both optimization-based and

[7] The code can be found at https://github.com/caitree/GoodAction.

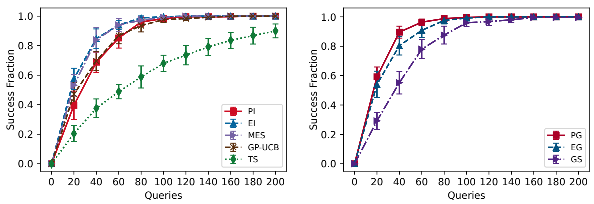
[8] The implementation of the GP model comes from https://github.com/ntienvu/MiniBO/

[9] This is lower than in Section 5.1, since there we wanted to be confident that the elimination algorithm eliminates correctly.

good-action based algorithms can perform well in terms of finding good actions, but the latter does so slightly faster in these experiments. In particular, the PG and EG algorithms appear to be most effective. We believe that GS is slightly slower here due to increased exploration, which may be of less benefit for good-action identification compared to regular optimization.

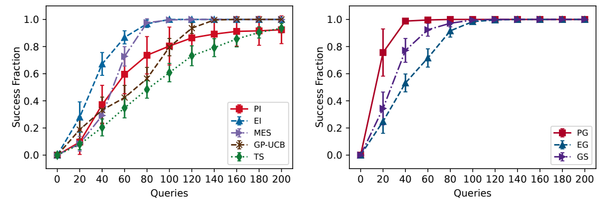


(a) Eggholder 2D

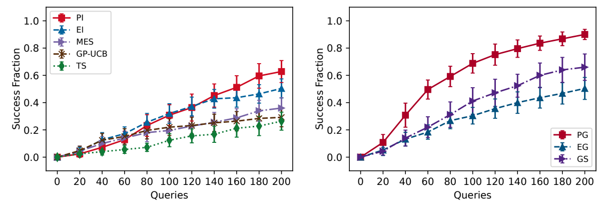


(b) Alpine 6D

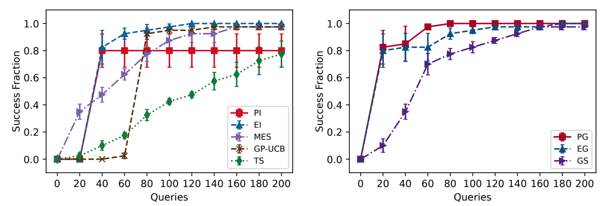
Figure 3. Results for noiseless synthetic functions with $\xi = \frac{1}{100}$.



(a) Robot Pushing 3D



(b) Robot Pushing 4D



(c) XGBoost Boston

Figure 4. Performance comparison on non-synthetic datasets.

5.4. Noiseless Non-Synthetic Functions

Robot pushing. We consider the robot pushing objective on a two-dimensional plane from (Wang & Jegelka, 2017), where the goal is to find a good enough pre-image for pushing an object to a fixed target location r_g. The 3-dimensional function takes robot location (r_x, r_y) and pushing duration r_t as input (the pushing angle is fixed to be $\arctan \frac{r_y}{r_x}$), then outputs the reversed gap between the final location and the target location, $5 - \|c(r_x, r_y, r_t) - r_g\|$, where $c(\cdot)$ calculates the robot final location. The 4-dimensional function takes an additional input r_θ specifying the angle to be pushed. The maximum function value is 5, and we set $\eta = 4.75$.

Hyperparameter tuning. We consider tuning a regression task using XGBoost (Chen & Guestrin, 2016) on the well-known Boston housing dataset. We perform 3-fold cross-validation, using a fixed seed in order to provide deterministic behavior. The five parameters that we tune are the maximum tree depth, the learning rate, the maximum delta step for each leaf output, the subsampling ratio of features, and the subsampling ratio of training instances. We take the objective function to be 10 minus the root-mean-square error (RMSE) on the test fold, and set $\eta = 7$.

Results. The results are shown in Figure 4. We observe similar overall behavior to the above synthetic functions, with PG performing best, and particularly noticeable improvements in the robot pushing experiment.

5.5. The Effect of Noise

In this experiment, we add zero-mean Gaussian noise with standard deviation $\sigma = 0.05$ to each evaluation. Due to the noise, the algorithm can no longer simply stop when a good action is sampled. Instead, we continue every algorithm up to the maximum time $T = 200$, and at each time instant, we plot the fraction of runs for which the algorithm's *best estimate* is a good action. We take the best estimate to be the point with the highest posterior mean.

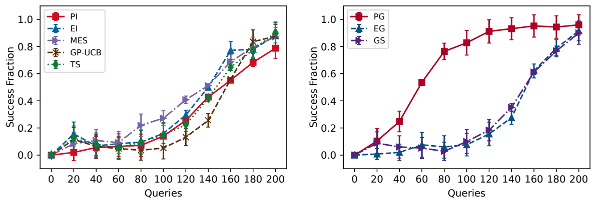
The results for this setting are shown in Figure 5. Unsurprisingly, the noise makes the curves more erratic overall, and sometimes even non-monotone. Interestingly, the gains offered by PG are considerable for the Keane function, and also marginally visible for the Ackley function.

5.6. Additional Experiments

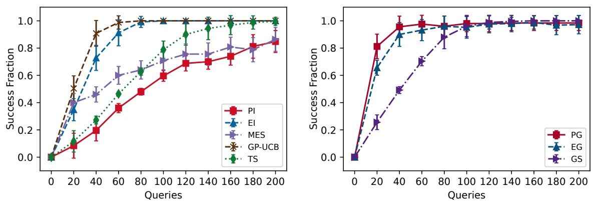
In Appendix D, we provide additional experiments exploring (i) the effect of varying η so that the space of good actions grows or shrinks, and (ii) the robustness of our algorithms when no good action exists (i.e., $\eta > f(\mathbf{x}^*)$).

5.7. Summary

Overall, we believe that our experiments indicate PG to be a highly effective algorithm for good-action identification, with EG typically also being competitive. While GS was



(a) Keane



(b) Ackley 6D

Figure 5. Results for the noisy setting with $\xi = \frac{1}{100}$.

typically less effective in the experiments that we ran, it may still be interest to further investigate further when non-myopic methods can help more significantly.

6. Conclusion

We have established theoretical bounds on the lenient regret for Gaussian process bandits, indicating a significant reduction compared to the standard notion of cumulative regret. In addition, in the fixed-threshold good-action identification problem, we provided several algorithms that exploit knowledge of the threshold, and provided experimental evidence that PG is particularly effective in practice.

Acknowledgement

This work was supported by the Singapore National Research Foundation (NRF) under grant number R-252-000-A74-281.

References

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

Bingham, D. Virtual library of simulation experiments: Test functions and datasets. https://www.sfu.ca/~ssurjano/index.html, 2021.

Bogunovic, I., Scarlett, J., Krause, A., and Cevher, V. Truncated variance reduction: A unified approach to Bayesian optimization and level-set estimation. In *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, 2016.

Bogunovic, I., Krause, A., and Scarlett, J. Corruption-tolerant Gaussian process bandit optimization. In *Int. Conf. Art. Intel. Stats. (AISTATS)*, 2020.

Bryan, B., Nichol, R. C., Genovese, C. R., Schneider, J., Miller, C. J., and Wasserman, L. Active learning for identifying function threshold boundaries. In *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, 2006.

Bull, A. D. Convergence rates of efficient global optimization algorithms. *J. Mach. Learn. Res.*, 12(Oct.):2879–2904, 2011.

Cai, X. and Scarlett, J. On lower bounds for standard and robust Gaussian process bandit optimization. In *Int. Conf. Mach. Learn. (ICML)*, 2021.

Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. In *ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining*, pp. 785–794, 2016.

Chowdhury, S. R. and Gopalan, A. On kernelized multi-armed bandits. In *Int. Conf. Mach. Learn. (ICML)*, 2017.

Contal, E., Buffoni, D., Robicquet, A., and Vayatis, N. *Machine Learning and Knowledge Discovery in Databases*, chapter Parallel Gaussian Process Optimization with Upper Confidence Bound and Pure Exploration, pp. 225–240. Springer Berlin Heidelberg, 2013.

Gotovos, A., Casati, N., Hitz, G., and Krause, A. Active learning for level set estimation. In *Int. Joint. Conf. Art. Intel. (IJCAI)*, 2013.

Grünewälder, S., Audibert, J.-Y., Opper, M., and Shawe-Taylor, J. Regret bounds for Gaussian process bandit problems. In *Int. Conf. Art. Intel. Stats. (AISTATS)*, pp. 273–280, 2010.

Hennig, P. and Schuler, C. J. Entropy search for information-efficient global optimization. *J. Mach. Learn. Research*, 13(1):1809–1837, 2012.

Hernández-Lobato, J. M., Hoffman, M. W., and Ghahramani, Z. Predictive entropy search for efficient global optimization of black-box functions. In *Conf. Neur. Inf. Proc. Sys. (NeurIPS)*, 2014.

Janz, D., Burt, D. R., and González, J. Bandit optimisation of functions in the Matérn kernel RKHS. In *Int. Conf. Art. Intel. Stats. (AISTATS)*, 2020.

Kano, H., Honda, J., Sakamaki, K., Matsuura, K., Nakamura, A., and Sugiyama, M. Good arm identification via bandit feedback. *Machine Learning*, 108(5):721–745, 2019.

Katz-Samuels, J. and Jamieson, K. The true sample complexity of identifying good arms. In *Int. Conf. Art. Intel. Stats. (AISTATS)*, 2020.

Kushner, H. J. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J. Fluids Eng.*, 86(1):97–106, 1964.

Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, 2020.

Li, C., Rana, S., Gupta, S., Nguyen, V., and Venkatesh, S. Bayesian optimization with monotonicity information. In *NeurIPS Workshop on Bayesian Optimization*, 2017.

Merlis, N. and Mannor, S. Lenient regret for multi-armed bandits. In *AAAI Conf. Art. Intel.*, 2021.

Mockus, J., Tiesis, V., and Zilinskas, A. The application of bayesian methods for seeking the extremum. *Towards Global Optimization*, 2(117-129):2, 1978.

Nguyen, V. and Osborne, M. A. Knowing the what but not the where in Bayesian optimization. In *Int. Conf. Mach. Learn. (ICML)*. PMLR, 2020.

Rasmussen, C. E. Gaussian processes for machine learning. MIT Press, 2006.

Russo, D. and Van Roy, B. Satisficing in time-sensitive bandit learning. https://arxiv.org/abs/1803.02855, 2018.

Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., and Wen, Z. A tutorial on thompson sampling. *Found. Trends Mach. Learn.*, 11(1):1–96, 2018.

Scarlett, J. Tight regret bounds for Bayesian optimization in one dimension. In *Int. Conf. Mach. Learn. (ICML)*, 2018.

Scarlett, J., Bogunovic, I., and Cevher, V. Lower bounds on regret for noisy Gaussian process bandit optimization. In *Conf. Learn. Theory (COLT)*. 2017.

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

Shekhar, S. and Javidi, T. Gaussian process bandits with adaptive discretization. *Elec. J. Stats.*, 12(2):3829–3874, 2018.

Shekhar, S. and Javidi, T. Multiscale Gaussian process level set estimation. In *Proc. Mach. Learn. Research*, volume 89, pp. 3283–3291, April 2019.

Shekhar, S. and Javidi, T. Multi-scale zero-order optimization of smooth functions in an RKHS. https://arxiv.org/abs/2005.04832, 2020.

Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Int. Conf. Mach. Learn. (ICML)*, 2010.

Sui, Y., Gotovos, A., Burdick, J. W., and Krause, A. Safe exploration for optimization with Gaussian processes. In *Int. Conf. Mach. Learn. (ICML)*, 2015.

Vakili, S., Khezeli, K., and Picheny, V. On information gain and regret bounds in Gaussian process bandits. In *Int. Conf. Art. Intel. Stats. (AISTATS)*, 2021.

Valko, M., Korda, N., Munos, R., Flaounas, I., and Cristianini, N. Finite-time analysis of kernelised contextual bandits. In *Conf. Uncertainty in AI (UAI)*, 2013.

Wang, Z. and Jegelka, S. Max-value entropy search for efficient Bayesian optimization. In *Int. Conf. Mach. Learn. (ICML)*, pp. 3627–3635, 2017.