

# Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design

Niranjn Srinivas  
California Institute of Technology  
niranjn@caltech.edu

Sham M. Kakade  
University of Pennsylvania  
skakade@wharton.upenn.edu

Andreas Krause  
California Institute of Technology  
krausea@caltech.edu

Matthias Seeger  
Saarland University  
mseeger@mmci.uni-saarland.de

## Abstract

Many applications require optimizing an unknown, noisy function that is expensive to evaluate. We formalize this task as a multi-armed bandit problem, where the payoff function is either sampled from a Gaussian process (GP) or has low RKHS norm. We resolve the important open problem of deriving regret bounds for this setting, which imply novel convergence rates for GP optimization. We analyze GP-UCB, an intuitive upper-confidence based algorithm, and bound its cumulative regret in terms of maximal information gain, establishing a novel connection between GP optimization and experimental design. Moreover, by bounding the latter in terms of operator spectra, we obtain explicit sublinear regret bounds for many commonly used covariance functions. In some important cases, our bounds have surprisingly weak dependence on the dimensionality. In our experiments on real sensor data, GP-UCB compares favorably with other heuristical GP optimization approaches.

## 1. Introduction

In most stochastic optimization settings, evaluating the unknown function is expensive, and sampling is to be minimized. Examples include choosing advertisements in sponsored search to maximize profit in a click-through model (Pandey & Olston, 2007) or learning optimal control strategies for robots (Lizotte et al., 2007). Predominant approaches to this problem include the multi-armed bandit paradigm (Robbins, 1952), where the goal is to maximize cumulative reward by optimally balancing exploration and exploitation, and experimental design (Chaloner & Verdinelli, 1995), where the function is to be explored globally with as few evaluations as possible, for example by maximizing information

gain. The challenge in both approaches is twofold: we have to estimate an unknown function  $f$  from noisy samples, and we must optimize our estimate over some high-dimensional input space. For the former, much progress has been made in machine learning through kernel methods and Gaussian process (GP) models (Rasmussen & Williams, 2006), where smoothness assumptions about  $f$  are encoded through the choice of kernel in a flexible nonparametric fashion. Beyond Euclidean spaces, kernels can be defined on diverse domains such as spaces of graphs, sets, or lists.

We are concerned with GP optimization in the multi-armed bandit setting, where  $f$  is sampled from a GP distribution or has low “complexity” measured in terms of its RKHS norm under some kernel. We provide the first sublinear regret bounds in this nonparametric setting, which imply convergence rates for GP optimization. In particular, we analyze the Gaussian Process Upper Confidence Bound (GP-UCB) algorithm, a simple and intuitive Bayesian method (Auer et al., 2002; Auer, 2002; Dani et al., 2008). While objectives are different in the multi-armed bandit and experimental design paradigm, our results draw a close technical connection between them: our regret bounds come in terms of an *information gain* quantity, measuring how fast  $f$  can be learned in an information theoretic sense. The submodularity of this function allows us to prove sharp regret bounds for particular covariance functions, which we demonstrate for commonly used Squared Exponential and Matérn kernels.

**Related Work.** Our work generalizes stochastic *linear* optimization in a bandit setting, where the unknown function comes from a finite-dimensional linear space. GPs are nonlinear random functions, which can be represented in an infinite-dimensional linear space. For the standard linear setting, Dani et al. (2008) provide a near-complete characterization

<sup>1</sup>This is the longer version of our paper in ICML 2010; see Srinivas et al. (2010)

(also see Auer 2002; Dani et al. 2007; Abernethy et al. 2008; Rusmevichientong & Tsitsiklis 2008), explicitly dependent on the dimensionality. In the GP setting, the challenge is to characterize complexity in a different manner, through properties of the kernel function. Our technical contributions are twofold: first, we show how to analyze the nonlinear setting by focusing on the concept of information gain, and second, we explicitly bound this information gain measure using the concept of submodularity (Nemhauser et al., 1978) and knowledge about kernel operator spectra.

Kleinberg et al. (2008) provide regret bounds under weaker and less configurable assumptions (only Lipschitz-continuity w.r.t. a metric is assumed; Bubeck et al. 2008 consider arbitrary topological spaces), which however degrade rapidly with the dimensionality of the problem ( $\Omega(T^{\frac{d+1}{d+2}})$ ). In practice, linearity w.r.t. a fixed basis is often too stringent an assumption, while Lipschitz-continuity can be too coarse-grained, leading to poor rate bounds. Adopting GP assumptions, we can model levels of smoothness in a fine-grained way. For example, our rates for the frequently used Squared Exponential kernel, enforcing a high degree of smoothness, have weak dependence on the dimensionality:  $\mathcal{O}(\sqrt{T(\log T)^{d+1}})$  (see Fig. 1).

There is a large literature on GP (response surface) optimization. Several heuristics for trading off exploration and exploitation in GP optimization have been proposed (such as Expected Improvement, Mockus et al. 1978, and Most Probable Improvement, Mockus 1989) and successfully applied in practice (c.f., Lizotte et al. 2007). Brochu et al. (2009) provide a comprehensive review of and motivation for Bayesian optimization using GPs. The Efficient Global Optimization (EGO) algorithm for optimizing expensive black-box functions is proposed by Jones et al. (1998) and extended to GPs by Huang et al. (2006). Little is known about theoretical performance of GP optimization. While convergence of EGO is established by Vazquez & Bect (2007), convergence rates have remained elusive. Grünewälder et al. (2010) consider the pure exploration problem for GPs, where the goal is to find the optimal decision over  $T$  rounds, rather than maximize cumulative reward (with no exploration/exploitation dilemma). They provide sharp bounds for this exploration problem. Note that this methodology would not lead to bounds for minimizing the cumulative regret. Our cumulative regret bounds translate to the first performance guarantees (rates) for GP optimization.

**Summary.** Our main contributions are:

- We analyze GP-UCB, an intuitive algorithm for GP optimization, when the function is either sam-

Kernel	Linear	RBF	Matérn
Regret $R_T$	$d\sqrt{T}$	$\sqrt{T(\log T)^{d+1}}$	$T^{\frac{\nu+d(d+1)}{2\nu+d(d+1)}}$

Figure 1. Our regret bounds (up to polylog factors) for linear, radial basis, and Matérn kernels —  $d$  is the dimension,  $T$  is the time horizon, and  $\nu$  is a Matérn parameter.

pled from a known GP, or has low RKHS norm.

- We bound the cumulative regret for GP-UCB in terms of the information gain due to sampling, establishing a novel connection between experimental design and GP optimization.
- By bounding the information gain for popular classes of kernels, we establish sublinear regret bounds for GP optimization for the first time. Our bounds depend on kernel choice and parameters in a fine-grained fashion.
- We evaluate GP-UCB on sensor network data, demonstrating that it compares favorably to existing algorithms for GP optimization.

## 2. Problem Statement and Background

Consider the problem of sequentially optimizing an unknown reward function  $f : D \rightarrow \mathbb{R}$ : in each round  $t$ , we choose a point  $\mathbf{x}_t \in D$  and get to see the function value there, perturbed by noise:  $y_t = f(\mathbf{x}_t) + \epsilon_t$ . Our goal is to maximize the sum of rewards  $\sum_{t=1}^T f(\mathbf{x}_t)$ , thus to perform essentially as well as  $\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x} \in D} f(\mathbf{x})$  (as rapidly as possible). For example, we might want to find locations of highest temperature in a building by sequentially activating sensors in a spatial network and regressing on their measurements.  $D$  consists of all sensor locations,  $f(\mathbf{x})$  is the temperature at  $\mathbf{x}$ , and sensor accuracy is quantified by the noise variance. Each activation draws battery power, so we want to sample from as few sensors as possible.

**Regret.** A natural performance metric in this context is cumulative regret, the loss in reward due to not knowing  $f$ 's maximum points beforehand. Suppose the unknown function is  $f$ , its maximum point<sup>1</sup>  $\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x} \in D} f(\mathbf{x})$ . For our choice  $\mathbf{x}_t$  in round  $t$ , we incur instantaneous regret  $r_t = f(\mathbf{x}^*) - f(\mathbf{x}_t)$ . The *cumulative regret*  $R_T$  after  $T$  rounds is the sum of instantaneous regrets:  $R_T = \sum_{t=1}^T r_t$ . A desirable asymptotic property of an algorithm is to be *no-regret*:  $\lim_{T \rightarrow \infty} R_T/T = 0$ . Note that neither  $r_t$  nor  $R_T$  are ever revealed to the algorithm. Bounds on the average regret  $R_T/T$  translate to convergence rates for GP optimization: the maximum  $\max_{t \leq T} f(\mathbf{x}_t)$  in the first  $T$  rounds is no further from  $f(\mathbf{x}^*)$  than the average.

<sup>1</sup>  $\mathbf{x}^*$  need not be unique; only  $f(\mathbf{x}^*)$  occurs in the regret.

## 2.1. Gaussian Processes and RKHS's

**Gaussian Processes.** Some assumptions on  $f$  are required to guarantee no-regret. While rigid parametric assumptions such as linearity may not hold in practice, a certain degree of smoothness is often warranted. In our sensor network, temperature readings at closeby locations are highly correlated (see Figure 2(a)). We can enforce implicit properties like smoothness without relying on any parametric assumptions, modeling  $f$  as a sample from a *Gaussian process* (GP): a collection of dependent random variables, one for each  $\mathbf{x} \in D$ , every finite subset of which is multivariate Gaussian distributed in an overall consistent way (Rasmussen & Williams, 2006). A  $GP(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$  is specified by its mean function  $\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$  and covariance (or kernel) function  $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))]$ . For GPs not conditioned on data, we assume<sup>2</sup> that  $\mu \equiv 0$ . Moreover, we restrict  $k(\mathbf{x}, \mathbf{x}) \leq 1$ ,  $\mathbf{x} \in D$ , i.e., we assume bounded variance. By fixing the correlation behavior, the covariance function  $k$  encodes smoothness properties of sample functions  $f$  drawn from the GP. A range of commonly used kernel functions is given in Section 5.2.

In this work, GPs play multiple roles. First, some of our results hold when the unknown target function is a sample from a known GP distribution  $GP(0, k(\mathbf{x}, \mathbf{x}'))$ . Second, the Bayesian algorithm we analyze generally uses  $GP(0, k(\mathbf{x}, \mathbf{x}'))$  as prior distribution over  $f$ . A major advantage of working with GPs is the existence of simple analytic formulae for mean and covariance of the posterior distribution, which allows easy implementation of algorithms. For a noisy sample  $\mathbf{y}_T = [y_1 \dots y_T]^T$  at points  $A_T = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ ,  $y_t = f(\mathbf{x}_t) + \epsilon_t$  with  $\epsilon_t \sim N(0, \sigma^2)$  i.i.d. Gaussian noise, the posterior over  $f$  is a GP distribution again, with mean  $\mu_T(\mathbf{x})$ , covariance  $k_T(\mathbf{x}, \mathbf{x}')$  and variance  $\sigma_T^2(\mathbf{x})$ :

$$\mu_T(\mathbf{x}) = \mathbf{k}_T(\mathbf{x})^T (\mathbf{K}_T + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_T, \quad (1)$$

$$k_T(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}_T(\mathbf{x})^T (\mathbf{K}_T + \sigma^2 \mathbf{I})^{-1} \mathbf{k}_T(\mathbf{x}'),$$

$$\sigma_T^2(\mathbf{x}) = k_T(\mathbf{x}, \mathbf{x}), \quad (2)$$

where  $\mathbf{k}_T(\mathbf{x}) = [k(\mathbf{x}_1, \mathbf{x}) \dots k(\mathbf{x}_T, \mathbf{x})]^T$  and  $\mathbf{K}_T$  is the positive definite kernel matrix  $[k(\mathbf{x}, \mathbf{x}')]_{\mathbf{x}, \mathbf{x}' \in A_T}$ .

**RKHS.** Instead of the Bayes case, where  $f$  is sampled from a GP prior, we also consider the more agnostic case where  $f$  has low “complexity” as measured under an RKHS norm (and distribution free assumptions on the noise process). The notion of *reproducing kernel Hilbert spaces* (RKHS, Wahba 1990) is intimately related to GPs and their covariance functions  $k(\mathbf{x}, \mathbf{x}')$ . The RKHS  $\mathcal{H}_k(D)$  is a complete subspace of  $L_2(D)$  of nicely behaved functions, with an

inner product  $\langle \cdot, \cdot \rangle_k$  obeying the reproducing property:  $\langle f, k(\mathbf{x}, \cdot) \rangle_k = f(\mathbf{x})$  for all  $f \in \mathcal{H}_k(D)$ . It is literally constructed by completing the set of mean functions  $\mu_T$  for all possible  $T$ ,  $\{\mathbf{x}_t\}$ , and  $\mathbf{y}_T$ . The induced RKHS norm  $\|f\|_k = \sqrt{\langle f, f \rangle_k}$  measures smoothness of  $f$  w.r.t.  $k$ : in much the same way as  $k_1$  would generate smoother samples than  $k_2$  as GP covariance functions,  $\|\cdot\|_{k_1}$  assigns larger penalties than  $\|\cdot\|_{k_2}$ .  $\langle \cdot, \cdot \rangle_k$  can be extended to all of  $L_2(D)$ , in which case  $\|f\|_k < \infty$  iff  $f \in \mathcal{H}_k(D)$ . For most kernels discussed in Section 5.2, members of  $\mathcal{H}_k(D)$  can uniformly approximate any continuous function on any compact subset of  $D$ .

## 2.2. Information Gain & Experimental Design

One approach to maximizing  $f$  is to first choose points  $\mathbf{x}_t$  so as to estimate the function globally well, then play the maximum point of our estimate. How can we learn about  $f$  as rapidly as possible? This question comes down to Bayesian Experimental Design (henceforth “ED”; see Chaloner & Verdinelli 1995), where the informativeness of a set of sampling points  $A \subset D$  about  $f$  is measured by the *information gain* (c.f., Cover & Thomas 1991), which is the mutual information between  $f$  and observations  $\mathbf{y}_A = \mathbf{f}_A + \epsilon_A$  at these points:

$$I(\mathbf{y}_A; f) = H(\mathbf{y}_A) - H(\mathbf{y}_A | f), \quad (3)$$

quantifying the reduction in uncertainty about  $f$  from revealing  $\mathbf{y}_A$ . Here,  $\mathbf{f}_A = [f(\mathbf{x})]_{\mathbf{x} \in A}$  and  $\epsilon_A \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ . For a Gaussian,  $H(N(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = \frac{1}{2} \log |2\pi e \boldsymbol{\Sigma}|$ , so that in our setting  $I(\mathbf{y}_A; f) = I(\mathbf{y}_A; \mathbf{f}_A) = \frac{1}{2} \log |\mathbf{I} + \sigma^{-2} \mathbf{K}_A|$ , where  $\mathbf{K}_A = [k(\mathbf{x}, \mathbf{x}')]_{\mathbf{x}, \mathbf{x}' \in A}$ . While finding the information gain maximizer among  $A \subset D$ ,  $|A| \leq T$  is NP-hard (Ko et al., 1995), it can be approximated by an efficient greedy algorithm. If  $F(A) = I(\mathbf{y}_A; f)$ , this algorithm picks  $\mathbf{x}_t = \operatorname{argmax}_{\mathbf{x} \in D} F(A_{t-1} \cup \{\mathbf{x}\})$  in round  $t$ , which can be shown to be equivalent to

$$\mathbf{x}_t = \operatorname{argmax}_{\mathbf{x} \in D} \sigma_{t-1}(\mathbf{x}), \quad (4)$$

where  $A_{t-1} = \{\mathbf{x}_1, \dots, \mathbf{x}_{t-1}\}$ . Importantly, this simple algorithm is guaranteed to find a near-optimal solution: for the set  $A_T$  obtained after  $T$  rounds, we have that

$$F(A_T) \geq (1 - 1/e) \max_{|A| \leq T} F(A), \quad (5)$$

at least a constant fraction of the optimal information gain value. This is because  $F(A)$  satisfies a diminishing returns property called *submodularity* (Krause & Guestrin, 2005), and the greedy approximation guarantee (5) holds for any submodular function (Nemhauser et al., 1978).

While sequentially optimizing Eq. 4 is a provably good way to *explore*  $f$  globally, it is not well suited for func-

<sup>2</sup>This is w.l.o.g. (Rasmussen & Williams, 2006).

tion optimization. For the latter, we only need to identify points  $\mathbf{x}$  where  $f(\mathbf{x})$  is large, in order to concentrate sampling there as rapidly as possible, thus *exploit* our knowledge about maxima. In fact, the ED rule (4) does not even depend on observations  $y_t$  obtained along the way. Nevertheless, the maximum information gain after  $T$  rounds will play a prominent role in our regret bounds, forging an important connection between GP optimization and experimental design.

### 3. GP-UCB Algorithm

For sequential optimization, the ED rule (4) can be wasteful: it aims at decreasing uncertainty globally, not just where maxima might be. Another idea is to pick points as  $\mathbf{x}_t = \operatorname{argmax}_{\mathbf{x} \in D} \mu_{t-1}(\mathbf{x})$ , maximizing the expected reward based on the posterior so far. However, this rule is too greedy too soon and tends to get stuck in shallow local optima. A combined strategy is to choose

$$\mathbf{x}_t = \operatorname{argmax}_{\mathbf{x} \in D} \mu_{t-1}(\mathbf{x}) + \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}), \quad (6)$$

where  $\beta_t$  are appropriate constants. This latter objective prefers both points  $\mathbf{x}$  where  $f$  is uncertain (large  $\sigma_{t-1}(\cdot)$ ) and such where we expect to achieve high rewards (large  $\mu_{t-1}(\cdot)$ ): it implicitly negotiates the exploration–exploitation tradeoff. A natural interpretation of this sampling rule is that it greedily selects points  $\mathbf{x}$  such that  $f(\mathbf{x})$  should be a reasonable upper bound on  $f(\mathbf{x}^*)$ , since the argument in (6) is an upper quantile of the marginal posterior  $P(f(\mathbf{x})|\mathbf{y}_{t-1})$ . We call this choice the *Gaussian process upper confidence bound* rule (GP-UCB), where  $\beta_t$  is specified depending on the context (see Section 4). Pseudocode for the GP-UCB algorithm is provided in Algorithm 1. Figure 2 illustrates two subsequent iterations, where GP-UCB both explores (Figure 2(b)) by sampling an input  $\mathbf{x}$  with large  $\sigma_{t-1}^2(\mathbf{x})$  and exploits (Figure 2(c)) by sampling  $\mathbf{x}$  with large  $\mu_{t-1}(\mathbf{x})$ .

The GP-UCB selection rule Eq. 6 is motivated by the UCB algorithm for the classical multi-armed bandit problem (Auer et al., 2002; Kocsis & Szepesvári, 2006). Among competing criteria for GP optimization (see Section 1), a variant of the GP-UCB rule has been demonstrated to be effective for this application (Dorard et al., 2009). To our knowledge, strong theoretical results of the kind provided for GP-UCB in this paper have not been given for any of these search heuristics. In Section 6, we show that in practice GP-UCB compares favorably with these alternatives.

If  $D$  is infinite, finding  $\mathbf{x}_t$  in (6) may be hard: the upper confidence index is multimodal in general. However, global search heuristics are very effective in practice (Brochu et al., 2009). It is generally assumed

---

#### Algorithm 1 The GP-UCB algorithm.

---

**Input:** Input space  $D$ ; GP Prior  $\mu_0 = 0$ ,  $\sigma_0$ ,  $k$   
**for**  $t = 1, 2, \dots$  **do**  
    Choose  $\mathbf{x}_t = \operatorname{argmax}_{\mathbf{x} \in D} \mu_{t-1}(\mathbf{x}) + \sqrt{\beta_t} \sigma_{t-1}(\mathbf{x})$   
    Sample  $y_t = f(\mathbf{x}_t) + \epsilon_t$   
    Perform Bayesian update to obtain  $\mu_t$  and  $\sigma_t$   
**end for**

---

that evaluating  $f$  is more costly than maximizing the UCB index.

UCB algorithms (and GP optimization techniques in general) have been applied to a large number of problems in practice (Kocsis & Szepesvári, 2006; Pandey & Olston, 2007; Lizotte et al., 2007). Their performance is well characterized in both the finite arm setting and the linear optimization setting, but no convergence rates for GP optimization are known.

### 4. Regret Bounds

We now establish cumulative regret bounds for GP optimization, treating a number of different settings:  $f \sim \text{GP}(0, k(\mathbf{x}, \mathbf{x}'))$  for finite  $D$ ,  $f \sim \text{GP}(0, k(\mathbf{x}, \mathbf{x}'))$  for general compact  $D$ , and the agnostic case of arbitrary  $f$  with bounded RKHS norm.

GP optimization generalizes stochastic linear optimization, where a function  $f$  from a finite-dimensional linear space is optimized over. For the linear case, Dani et al. (2008) provide regret bounds that explicitly depend on the dimensionality<sup>3</sup>  $d$ . GPs can be seen as random functions in some infinite-dimensional linear space, so their results do not apply in this case. This problem is circumvented in our regret bounds. The quantity governing them is the *maximum information gain*  $\gamma_T$  after  $T$  rounds, defined as:

$$\gamma_T := \max_{A \subset D: |A|=T} \mathbf{I}(\mathbf{y}_A; \mathbf{f}_A), \quad (7)$$

where  $\mathbf{I}(\mathbf{y}_A; \mathbf{f}_A) = \mathbf{I}(\mathbf{y}_A; f)$  is defined in (3). Recall that  $\mathbf{I}(\mathbf{y}_A; \mathbf{f}_A) = \frac{1}{2} \log |\mathbf{I} + \sigma^{-2} \mathbf{K}_A|$ , where  $\mathbf{K}_A = [k(\mathbf{x}, \mathbf{x}')]_{\mathbf{x}, \mathbf{x}' \in A}$  is the covariance matrix of  $\mathbf{f}_A = [f(\mathbf{x})]_{\mathbf{x} \in A}$  associated with the samples  $A$ . Our regret bounds are of the form  $\mathcal{O}^*(\sqrt{T\beta_T\gamma_T})$ , where  $\beta_T$  is the confidence parameter in Algorithm 1, while the bounds of Dani et al. (2008) are of the form  $\mathcal{O}^*(\sqrt{T\beta_T d})$  ( $d$  the dimensionality of the linear function space). Here and below, the  $\mathcal{O}^*$  notation is a variant of  $\mathcal{O}$ , where log factors are suppressed. While our proofs – all provided in the Appendix – use techniques similar to those of Dani et al. (2008), we face a number of additional

---

<sup>3</sup> In general,  $d$  is the dimensionality of the input space  $D$ , which in the finite-dimensional linear case coincides with the feature space.



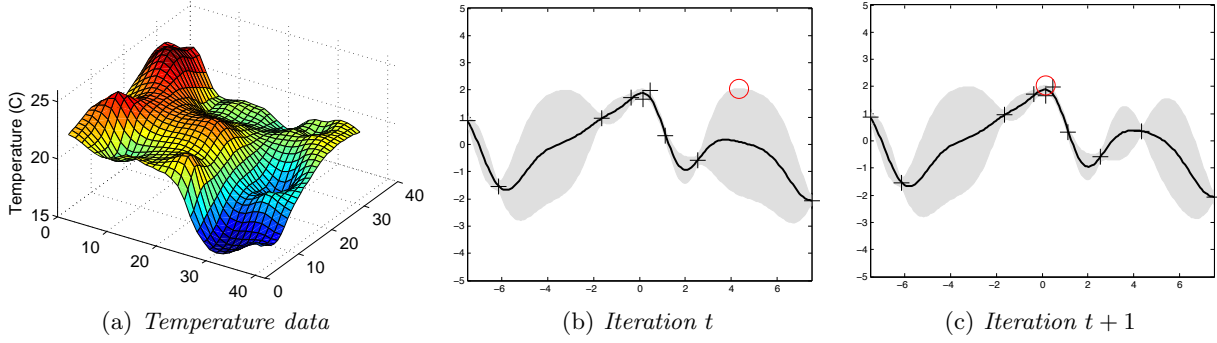


Figure 2. (a) Example of temperature data collected by a network of 46 sensors at Intel Research Berkeley. (b,c) Two iterations of the GP-UCB algorithm. It samples points that are either uncertain (b) or have high posterior mean (c).

significant technical challenges. Besides avoiding the finite-dimensional analysis, we must handle confidence issues, which are more delicate for nonlinear random functions.

Importantly, note that the information gain is a problem dependent quantity — properties of both the kernel and the input space will determine the growth of regret. In Section 5, we provide general methods for bounding  $\gamma_T$ , either by efficient auxiliary computations or by direct expressions for specific kernels of interest. Our results match known lower bounds (up to log factors) in both the  $K$ -armed bandit and the  $d$ -dimensional linear optimization case.

**Bounds for a GP Prior.** For finite  $D$ , we obtain the following bound.

**Theorem 1** Let  $\delta \in (0, 1)$  and  $\beta_t = 2 \log(|D|t^2\pi^2/6\delta)$ . Running GP-UCB with  $\beta_t$  for a sample  $f$  of a GP with mean function zero and covariance function  $k(\mathbf{x}, \mathbf{x}')$ , we obtain a regret bound of  $\mathcal{O}^*(\sqrt{T\gamma_T \log |D|})$  with high probability. Precisely,

$$\Pr \left\{ R_T \leq \sqrt{C_1 T \beta_T \gamma_T} \quad \forall T \geq 1 \right\} \geq 1 - \delta.$$

where  $C_1 = 8 / \log(1 + \sigma^{-2})$ .

The proof methodology follows Dani et al. (2007) in that we relate the regret to the growth of the log volume of the confidence ellipsoid — a novelty in our proof is showing how this growth is characterized by the information gain.

This theorem shows that, with high probability over samples from the GP, the cumulative regret is bounded in terms of the maximum information gain, forging a novel connection between GP optimization and experimental design. This link is of fundamental technical importance, allowing us to generalize Theorem 1 to infinite decision spaces. Moreover, the submodularity of  $I(\mathbf{y}_A; \mathbf{f}_A)$  allows us to derive sharp a priori bounds,

depending on choice and parameterization of  $k$  (see Section 5). In the following theorem, we generalize our result to any compact and convex  $D \subset \mathbb{R}^d$  under mild assumptions on the kernel function  $k$ .

**Theorem 2** Let  $D \subset [0, r]^d$  be compact and convex,  $d \in \mathbb{N}$ ,  $r > 0$ . Suppose that the kernel  $k(\mathbf{x}, \mathbf{x}')$  satisfies the following high probability bound on the derivatives of GP sample paths  $f$ : for some constants  $a, b > 0$ ,

$$\Pr \left\{ \sup_{\mathbf{x} \in D} |\partial f / \partial x_j| > L \right\} \leq ae^{-(L/b)^2}, \quad j = 1, \dots, d.$$

Pick  $\delta \in (0, 1)$ , and define

$$\beta_t = 2 \log(t^2 2\pi^2 / (3\delta)) + 2d \log \left( t^2 d b r \sqrt{\log(4da/\delta)} \right).$$

Running the GP-UCB with  $\beta_t$  for a sample  $f$  of a GP with mean function zero and covariance function  $k(\mathbf{x}, \mathbf{x}')$ , we obtain a regret bound of  $\mathcal{O}^*(\sqrt{dT\gamma_T})$  with high probability. Precisely, with  $C_1 = 8 / \log(1 + \sigma^{-2})$  we have

$$\Pr \left\{ R_T \leq \sqrt{C_1 T \beta_T \gamma_T} + 2 \quad \forall T \geq 1 \right\} \geq 1 - \delta.$$

The main challenge in our proof (provided in the Appendix) is to lift the regret bound in terms of the confidence ellipsoid to general  $D$ . The smoothness assumption on  $k(\mathbf{x}, \mathbf{x}')$  disqualifies GPs with highly erratic sample paths. It holds for stationary kernels  $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$  which are four times differentiable (Theorem 5 of Ghosal & Roy (2006)), such as the Squared Exponential and Matérn kernels with  $\nu > 2$  (see Section 5.2), while it is violated for the Ornstein-Uhlenbeck kernel (Matérn with  $\nu = 1/2$ ; a stationary variant of the Wiener process). For the latter, sample paths  $f$  are nondifferentiable almost everywhere with probability one and come with independent increments. We conjecture that a result of the form of Theorem 2 does not hold in this case.

**Bounds for Arbitrary  $f$  in the RKHS.** Thus far, we have assumed that the target function  $f$  is sampled

from a GP prior and that the noise is  $N(0, \sigma^2)$  with known variance  $\sigma^2$ . We now analyze GP-UCB in an agnostic setting, where  $f$  is an arbitrary function from the RKHS corresponding to kernel  $k(\mathbf{x}, \mathbf{x}')$ . Moreover, we allow the noise variables  $\varepsilon_t$  to be an arbitrary martingale difference sequence (meaning that  $\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 0$  for all  $t \in \mathbb{N}$ ), uniformly bounded by  $\sigma$ . Note that we still run the same GP-UCB algorithm, whose prior and noise model are misspecified in this case. Our following result shows that GP-UCB attains sublinear regret even in the agnostic setting.

**Theorem 3** *Let  $\delta \in (0, 1)$ . Assume that the true underlying  $f$  lies in the RKHS  $\mathcal{H}_k(D)$  corresponding to the kernel  $k(\mathbf{x}, \mathbf{x}')$ , and that the noise  $\varepsilon_t$  has zero mean conditioned on the history and is bounded by  $\sigma$  almost surely. In particular, assume  $\|f\|_k^2 \leq B$  and let  $\beta_t = 2B + 300\gamma_T \log^3(t/\delta)$ . Running GP-UCB with  $\beta_t$ , prior  $GP(0, k(\mathbf{x}, \mathbf{x}'))$  and noise model  $N(0, \sigma^2)$ , we obtain a regret bound of  $\mathcal{O}^*(\sqrt{T}(B\sqrt{\gamma_T} + \gamma_T))$  with high probability (over the noise). Precisely,*

$$\Pr\left\{R_T \leq \sqrt{C_1 T \beta_T \gamma_T} \quad \forall T \geq 1\right\} \geq 1 - \delta,$$

where  $C_1 = 8/\log(1 + \sigma^{-2})$ .

Note that while our theorem implicitly assumes that GP-UCB has knowledge of an upper bound on  $\|f\|_k$ , standard guess-and-doubling approaches suffice if no such bound is known a priori. Comparing Theorem 2 and Theorem 3, the latter holds uniformly over all functions  $f$  with  $\|f\|_k < \infty$ , while the former is a probabilistic statement requiring knowledge of the GP that  $f$  is sampled from. In contrast, if  $f \sim GP(0, k(\mathbf{x}, \mathbf{x}'))$ , then  $\|f\|_k = \infty$  almost surely (Wahba, 1990): sample paths are rougher than RKHS functions. Neither Theorem 2 nor 3 encompasses the other.

## 5. Bounding the Information Gain

Since the bounds developed in Section 4 depend on the information gain, the key remaining question is how to bound the quantity  $\gamma_T$  for practical classes of kernels.

### 5.1. Submodularity and Greedy Maximization

In order to bound  $\gamma_T$ , we have to maximize the information gain  $F(A) = \mathbf{I}(\mathbf{y}_A; f)$  over all subsets  $A \subset D$  of size  $T$ : a combinatorial problem in general. However, as noted in Section 2,  $F(A)$  is a submodular function, which implies the performance guarantee (5) for maximizing  $F$  sequentially by the greedy ED rule (4). Dividing both sides of (5) by  $1 - 1/e$ , we can upper-bound  $\gamma_T$  by  $(1 - 1/e)^{-1} \mathbf{I}(\mathbf{y}_{A_T}; f)$ , where  $A_T$  is constructed by the greedy procedure. Thus, somewhat counterintuitively, instead of using submodularity to prove that  $F(A_T)$  is near-optimal, we use it in order to show that

$\gamma_T$  is “near-greedy”. As noted in Section 2, the ED rule does not depend on observations  $y_t$  and can be run without evaluating  $f$ .

The importance of this greedy bound is twofold. First, it allows us to numerically compute highly problem-specific bounds on  $\gamma_T$ , which can be plugged into our results in Section 4 to obtain high-probability bounds on  $R_T$ . This being a laborious procedure, one would prefer *a priori* bounds for  $\gamma_T$  in practice which are simple analytical expressions of  $T$  and parameters of  $k$ . In this section, we sketch a general procedure for obtaining such expressions, instantiating them for a number of commonly used covariance functions, once more relying crucially on the greedy ED rule upper bound. Suppose that  $D$  is finite for now, and let  $\mathbf{f} = [f(\mathbf{x})]_{\mathbf{x} \in D}$ ,  $\mathbf{K}_D = [k(\mathbf{x}, \mathbf{x}')]_{\mathbf{x}, \mathbf{x}' \in D}$ . Sampling  $f$  at  $\mathbf{x}_t$ , we obtain  $y_t \sim N(\mathbf{v}_t^T \mathbf{f}, \sigma^2)$ , where  $\mathbf{v}_t \in \mathbb{R}^{|D|}$  is the indicator vector associated with  $\mathbf{x}_t$ . We can upper-bound the greedy maximum once more, by relaxing this constraint to  $\|\mathbf{v}_t\| = 1$  in round  $t$  of the sequential method. For this relaxed greedy procedure, all  $\mathbf{v}_t$  are leading eigenvectors of  $\mathbf{K}_D$ , since successive covariance matrices of  $P(\mathbf{f} | \mathbf{y}_{t-1})$  share their eigenbasis with  $\mathbf{K}_D$ , while eigenvalues are damped according to how many times the corresponding eigenvector is selected. We can upper-bound the information gain by considering the worst-case allocation of  $T$  samples to the  $\min\{T, |D|\}$  leading eigenvectors of  $\mathbf{K}_D$ :

$$\gamma_T \leq \frac{1/2}{1 - e^{-1}} \max_{(m_t)} \sum_{t=1}^{|D|} \log(1 + \sigma^{-2} m_t \hat{\lambda}_t), \quad (8)$$

subject to  $\sum_t m_t = T$ , and  $\text{spec}(\mathbf{K}_D) = \{\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots\}$ . We can split the sum into two parts in order to obtain a bound to leading order. The following Theorem captures this intuition:

**Theorem 4** *For any  $T \in \mathbb{N}$  and any  $T_* = 1, \dots, T$ :*

$$\gamma_T \leq \mathcal{O}(\sigma^{-2}[B(T_*)T + T_*(\log n_T T)]),$$

where  $n_T = \sum_{t=1}^{|D|} \hat{\lambda}_t$  and  $B(T_*) = \sum_{t=T_*+1}^{|D|} \hat{\lambda}_t$ .

Therefore, if for some  $T_* = o(T)$  the first  $T_*$  eigenvalues carry most of the total mass  $n_T$ , the information gain will be small. The more rapidly the spectrum of  $\mathbf{K}_D$  decays, the slower the growth of  $\gamma_T$ . Figure 3 illustrates this intuition.

### 5.2. Bounds for Common Kernels

In this section we bound  $\gamma_T$  for a range of commonly used covariance functions: finite dimensional linear, Squared Exponential and Matérn kernels. Together with our results in Section 4, these imply sublinear regret bounds for GP-UCB in all cases.

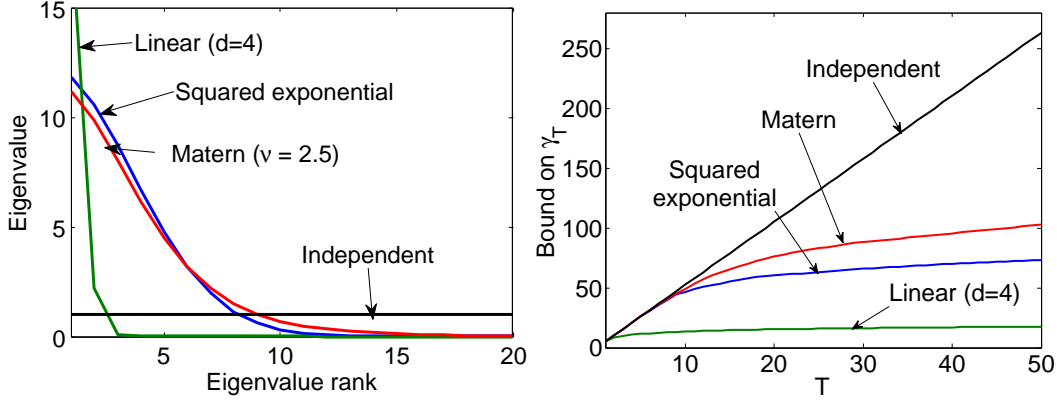


Figure 3. Spectral decay (left) and information gain bound (right) for independent (diagonal), linear, squared exponential and Matérn kernels ( $\nu = 2.5$ ) with equal trace.

*Finite dimensional linear* kernels have the form  $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$ . GPs with this kernel correspond to random linear functions  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$ ,  $\mathbf{w} \sim N(\mathbf{0}, \mathbf{I})$ .

The *Squared Exponential* kernel is  $k(\mathbf{x}, \mathbf{x}') = \exp(-(2l^2)^{-1} \|\mathbf{x} - \mathbf{x}'\|^2)$ ,  $l$  a lengthscale parameter. Sample functions are differentiable to any order almost surely (Rasmussen & Williams, 2006).

The *Matérn* kernel is given by  $k(\mathbf{x}, \mathbf{x}') = (2^{1-\nu}/\Gamma(\nu)) r^\nu B_\nu(r)$ ,  $r = (\sqrt{2\nu}/l) \|\mathbf{x} - \mathbf{x}'\|$ , where  $\nu$  controls the smoothness of sample paths (the smaller, the rougher) and  $B_\nu$  is a modified Bessel function. Note that as  $\nu \rightarrow \infty$ , appropriately rescaled Matérn kernels converge to the Squared Exponential kernel.

Figure 4 shows random functions drawn from GP distributions with the above kernels.

**Theorem 5** Let  $D \subset \mathbb{R}^d$  be compact and convex,  $d \in \mathbb{N}$ . Assume the kernel function satisfies  $k(\mathbf{x}, \mathbf{x}') \leq 1$ .

1. Finite spectrum. For the  $d$ -dimensional Bayesian linear regression case:  $\gamma_T = \mathcal{O}(d \log T)$ .
2. Exponential spectral decay. For the Squared Exponential kernel:  $\gamma_T = \mathcal{O}((\log T)^{d+1})$ .
3. Power law spectral decay. For Matérn kernels with  $\nu > 1$ :  $\gamma_T = \mathcal{O}(T^{d(d+1)/(2\nu+d(d+1))} (\log T))$ .

A proof of Theorem 5 is given in the Appendix, we only sketch the idea here.  $\gamma_T$  is bounded by Theorem 4 in terms of the eigendecay of the kernel matrix  $\mathbf{K}_D$ . If  $D$  is infinite or very large, we can use the operator spectrum of  $k(\mathbf{x}, \mathbf{x}')$ , which likewise decays rapidly. For the kernels of interest here, asymptotic expressions for the operator eigenvalues are given in Seeger et al. (2008), who derived bounds on the information gain for fixed and random designs (in contrast to the worst-case information gain considered here, which is substantially more challenging to bound). The main challenge in the proof is to ensure

the existence of discretizations  $D_T \subset D$ , dense in the limit, for which tail sums  $B(T_*)/n_T$  in Theorem 4 are close to corresponding operator spectra tail sums.

Together with Theorems 2 and 3, this result guarantees sublinear regret of GP-UCB for any dimension (see Figure 1). For the Squared Exponential kernel, the dimension  $d$  appears as exponent of  $\log T$  only, so that the regret grows at most as  $\mathcal{O}^*(\sqrt{T}(\log T)^{\frac{d+1}{2}})$  – the high degree of smoothness of the sample paths effectively combats the curse of dimensionality.

## 6. Experiments

We compare GP-UCB with heuristics such as the Expected Improvement (EI) and Most Probable Improvement (MPI), and with naive methods which choose points of maximum mean or variance only, both on synthetic and real sensor network data.

For synthetic data, we sample random functions from a squared exponential kernel with lengthscale parameter 0.2. The sampling noise variance  $\sigma^2$  was set to 0.025 or 5% of the signal variance. Our decision set  $D = [0, 1]$  is uniformly discretized into 1000 points. We run each algorithm for  $T = 1000$  iterations with  $\delta = 0.1$ , averaging over 30 trials (samples from the kernel). While the choice of  $\beta_t$  as recommended by Theorem 1 leads to competitive performance of GP-UCB, we find (using cross-validation) that the algorithm is improved by scaling  $\beta_t$  down by a factor 5. Note that we did not optimize constants in our regret bounds.

Next, we use temperature data collected from 46 sensors deployed at Intel Research Berkeley over 5 days at 1 minute intervals, pertaining to the example in Section 2. We take the first two-thirds of the data set to compute the empirical covariance of the sensor readings, and use it as the kernel matrix. The functions  $f$  for optimization consist of one set of observations from all the sensors taken from the remaining third of the

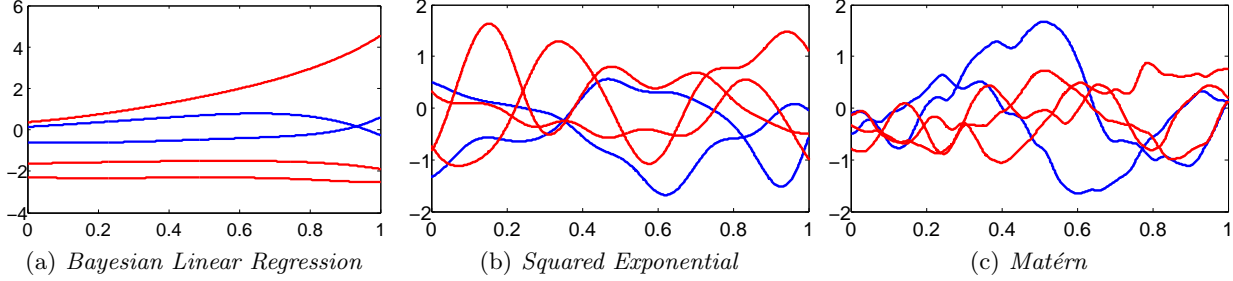


Figure 4. Sample functions drawn from a GP with linear, squared exponential and Matérn kernels ( $\nu = 2.5$ .)

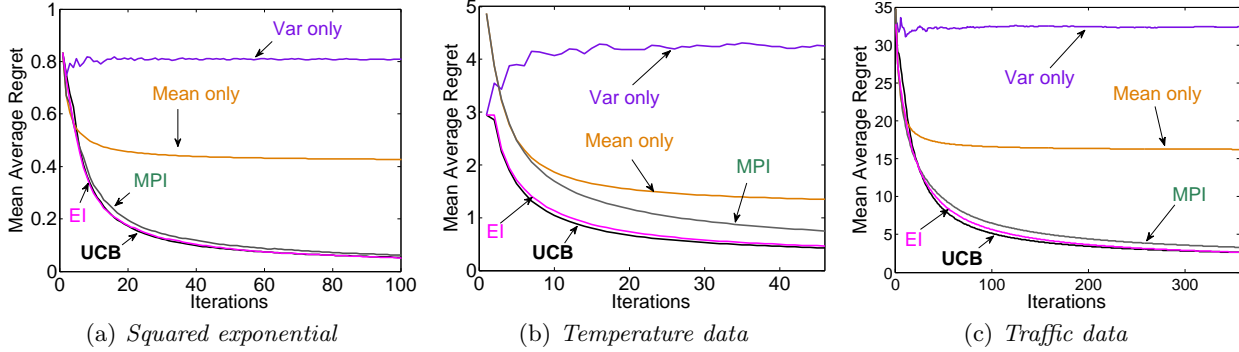


Figure 5. Comparison of performance: GP-UCB and various heuristics on synthetic (a), and sensor network data (b, c).

data set, and the results (for  $T = 46$ ,  $\sigma^2 = 0.5$  or 5% noise,  $\delta = 0.1$ ) were averaged over 2000 possible choices of the objective function.

Lastly, we take data from traffic sensors deployed along the highway I-880 South in California. The goal was to find the point of minimum speed in order to identify the most congested portion of the highway; we used traffic speed data for all working days from 6 AM to 11 AM for one month, from 357 sensors. We again use the covariance matrix from two-thirds of the data set as kernel matrix, and test on the other third. The results (for  $T = 357$ ,  $\sigma^2 = 4.78$  or 5% noise,  $\delta = 0.1$ ) were averaged over 900 runs.

Figure 5 compares the mean average regret incurred by the different heuristics and the GP-UCB algorithm on synthetic and real data. For temperature data, the GP-UCB algorithm and EI heuristic clearly outperform the others, and do not exhibit significant difference between each other. On synthetic and traffic data MPI does equally well. In summary, GP-UCB performs at least on par with the existing approaches which are not equipped with regret bounds.

## 7. Conclusions

We prove the first sublinear regret bounds for GP optimization with commonly used kernels (see Figure 1), both for  $f$  sampled from a known GP and  $f$  of low RKHS norm. We analyze GP-UCB, an intuitive,

Bayesian upper confidence bound based sampling rule. Our regret bounds crucially depend on the information gain due to sampling, establishing a novel connection between bandit optimization and experimental design. We bound the information gain in terms of the kernel spectrum, providing a general methodology for obtaining regret bounds with kernels of interest. Our experiments on real sensor network data indicate that GP-UCB performs at least on par with competing criteria for GP optimization, for which no regret bounds are known at present. Our results provide an interesting step towards understanding exploration–exploitation tradeoffs with complex utility functions.

## Acknowledgements

We thank Marcus Hutter for insightful comments on an earlier version of this paper. This research was partially supported by ONR grant N00014-09-1-1044, NSF grant CNS-0932392, a gift from Microsoft Corporation and the Excellence Initiative of the German research foundation (DFG).

## References

- Abernethy, J., Hazan, E., and Rakhlin, A. An efficient algorithm for linear bandit optimization, 2008. COLT.
- Auer, P. Using confidence bounds for exploitation–exploration trade-offs. *JMLR*, 3:397–422, 2002.
- Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time



- analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2-3):235–256, 2002.
- Brochu, E., Cora, M., and de Freitas, N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. In *TR-2009-23, UBC*, 2009.
- Bubeck, S., Munos, R., Stoltz, G., and Szepesvári, C. On-line optimization in X-armed bandits. In *NIPS*, 2008.
- Chaloner, K. and Verdinelli, I. Bayesian experimental design: A review. *Stat. Sci.*, 10(3):273–304, 1995.
- Cover, T. M. and Thomas, J. A. *Elements of Information Theory*. Wiley Interscience, 1991.
- Dani, V., Hayes, T. P., and Kakade, S. The price of bandit information for online optimization. In *NIPS*, 2007.
- Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. In *COLT*, 2008.
- Dorard, L., Glowacka, D., and Shawe-Taylor, J. Gaussian process modelling of dependencies in multi-armed bandit problems. In *Int. Symp. Op. Res.*, 2009.
- Freedman, D. A. On tail probabilities for martingales. *Ann. Prob.*, 3(1):100–118, 1975.
- Ghosal, S. and Roy, A. Posterior consistency of Gaussian process prior for nonparametric binary regression. *Ann. Stat.*, 34(5):2413–2429, 2006.
- Grünewälder, S., Audibert, J.-Y., Opper, M., and Shawe-Taylor, J. Regret bounds for gaussian process bandit problems. In *AISTATS*, 2010.
- Huang, D., Allen, T. T., Notz, W. I., and Zeng, N. Global optimization of stochastic black-box systems via sequential kriging meta-models. *J Glob. Opt.*, 34:441–466, 2006.
- Jones, D. R., Schonlau, M., and Welch, W. J. Efficient global optimization of expensive black-box functions. *J Glob. Opt.*, 13:455–492, 1998.
- Kleinberg, R., Slivkins, A., and Upfal, E. Multi-armed bandits in metric spaces. In *STOC*, pp. 681–690, 2008.
- Ko, C., Lee, J., and Queyranne, M. An exact algorithm for maximum entropy sampling. *Ops Res*, 43(4):684–691, 1995.
- Kocsis, L. and Szepesvári, C. Bandit based monte-carlo planning. In *ECML*, 2006.
- Krause, A. and Guestrin, C. Near-optimal nonmyopic value of information in graphical models. In *UAI*, 2005.
- Lizotte, D., Wang, T., Bowling, M., and Schuurmans, D. Automatic gait optimization with Gaussian process regression. In *IJCAI*, pp. 944–949, 2007.
- McDiarmid, C. *Concentration. In Probabilistic Methods for Algorithmic Discrete Mathematics*. Springer, 1998.
- Mockus, J. *Bayesian Approach to Global Optimization*. Kluwer Academic Publishers, 1989.
- Mockus, J., Tiesis, V., and Zilinskas, A. *Toward Global Optimization*, volume 2, chapter Bayesian Methods for Seeking the Extremum, pp. 117–128. 1978.
- Nemhauser, G., Wolsey, L., and Fisher, M. An analysis of the approximations for maximizing submodular set functions. *Math. Prog.*, 14:265–294, 1978.
- Pandey, S. and Olston, C. Handling advertisements of unknown quality in search advertising. In *NIPS*. 2007.
- Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- Robbins, H. Some aspects of the sequential design of experiments. *Bul. Am. Math. Soc.*, 58:527–535, 1952.
- Rusmevichientong, P. and Tsitsiklis, J. N. Linearly parameterized bandits. abs/0812.3465, 2008.
- Seeger, M. W., Kakade, S. M., and Foster, D. P. Information consistency of nonparametric Gaussian process methods. *IEEE Tr. Inf. Theo.*, 54(5):2376–2382, 2008.
- Shawe-Taylor, J., Williams, C., Cristianini, N., and Kandola, J. On the eigenspectrum of the Gram matrix and the generalization error of kernel-PCA. *IEEE Trans. Inf. Theo.*, 51(7):2510–2522, 2005.
- Srinivas, N., Krause, A., Kakade, S., and Seeger, M. Gaussian process optimization in the bandit setting: No regret and experimental design. In *ICML*, 2010.
- Stein, M. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999.
- Vazquez, E. and Bect, J. Convergence properties of the expected improvement algorithm, 2007.
- Wahba, G. *Spline Models for Observational Data*. SIAM, 1990.

## A. Regret Bounds for Target Function Sampled from GP

In this section, we provide details for the proofs of Theorem 1 and Theorem 2. In both cases, the strategy is to show that  $|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \leq \beta_t^{1/2} \sigma_{t-1}(\mathbf{x})$  for all  $t \in \mathbb{N}$  and all  $\mathbf{x} \in D$ , or in the infinite case, all  $\mathbf{x}$  in a discretization of  $D$  which becomes dense as  $t$  gets large.

### A.1. Finite Decision Set

We begin with the finite case,  $|D| < \infty$ .

**Lemma 5.1** *Pick  $\delta \in (0, 1)$  and set  $\beta_t = 2 \log(|D| \pi_t / \delta)$ , where  $\sum_{t \geq 1} \pi_t^{-1} = 1$ ,  $\pi_t > 0$ . Then,*

$$|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \leq \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}) \quad \forall \mathbf{x} \in D \quad \forall t \geq 1$$

*holds with probability  $\geq 1 - \delta$ .*

**Proof** Fix  $t \geq 1$  and  $\mathbf{x} \in D$ . Conditioned on  $\mathbf{y}_{t-1} = (y_1, \dots, y_{t-1})$ ,  $\{\mathbf{x}_1, \dots, \mathbf{x}_{t-1}\}$  are deterministic, and  $f(\mathbf{x}) \sim N(\mu_{t-1}(\mathbf{x}), \sigma_{t-1}^2(\mathbf{x}))$ . Now, if  $r \sim N(0, 1)$ , then

$$\begin{aligned} \Pr\{r > c\} &= e^{-c^2/2} (2\pi)^{-1/2} \int e^{-(r-c)^2/2 - c(r-c)} dr \\ &\leq e^{-c^2/2} \Pr\{r > 0\} = (1/2)e^{-c^2/2} \end{aligned}$$

for  $c > 0$ , since  $e^{-c(r-c)} \leq 1$  for  $r \geq c$ . Therefore,  $\Pr\{|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| > \beta_t^{1/2} \sigma_{t-1}(\mathbf{x})\} \leq e^{-\beta_t/2}$ , using  $r = (f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})) / \sigma_{t-1}(\mathbf{x})$  and  $c = \beta_t^{1/2}$ . Applying the union bound,

$$|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \leq \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}) \quad \forall \mathbf{x} \in D$$

holds with probability  $\geq 1 - |D|e^{-\beta_t/2}$ . Choosing  $|D|e^{-\beta_t/2} = \delta/\pi_t$  and using the union bound for  $t \in \mathbb{N}$ , the statement holds. For example, we can use  $\pi_t = \pi^2 t^2/6$ . ■

**Lemma 5.2** Fix  $t \geq 1$ . If  $|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \leq \beta_t^{1/2} \sigma_{t-1}(\mathbf{x})$  for all  $\mathbf{x} \in D$ , then the regret  $r_t$  is bounded by  $2\beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t)$ .

**Proof** By definition of  $\mathbf{x}_t$ :  $\mu_{t-1}(\mathbf{x}_t) + \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) \geq \mu_{t-1}(\mathbf{x}^*) + \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}^*) \geq f(\mathbf{x}^*)$ . Therefore,

$$\begin{aligned} r_t = f(\mathbf{x}^*) - f(\mathbf{x}_t) &\leq \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) + \mu_{t-1}(\mathbf{x}_t) - f(\mathbf{x}_t) \\ &\leq 2\beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t). \end{aligned}$$

■

**Lemma 5.3** The information gain for the points selected can be expressed in terms of the predictive variances. If  $\mathbf{f}_T = (f(\mathbf{x}_t)) \in \mathbb{R}^T$ :

$$I(\mathbf{y}_T; \mathbf{f}_T) = \frac{1}{2} \sum_{t=1}^T \log(1 + \sigma^{-2} \sigma_{t-1}^2(\mathbf{x}_t)).$$

**Proof** Recall that  $I(\mathbf{y}_T; \mathbf{f}_T) = H(\mathbf{y}_T) - (1/2) \log |2\pi e \sigma^2 \mathbf{I}|$ . Now,  $H(\mathbf{y}_T) = H(\mathbf{y}_{T-1}) + H(y_T | \mathbf{y}_{T-1}) = H(\mathbf{y}_{T-1}) + \log(2\pi e(\sigma^2 + \sigma_{t-1}^2(\mathbf{x}_T)))/2$ . Here, we use that  $\mathbf{x}_1, \dots, \mathbf{x}_T$  are deterministic conditioned on  $\mathbf{y}_{T-1}$ , and that the conditional variance  $\sigma_{T-1}^2(\mathbf{x}_T)$  does not depend on  $\mathbf{y}_{T-1}$ . The result follows by induction. ■

**Lemma 5.4** Pick  $\delta \in (0, 1)$  and let  $\beta_t$  be defined as in Lemma 5.1. Then, the following holds with probability  $\geq 1 - \delta$ :

$$\sum_{t=1}^T r_t^2 \leq \beta_T C_1 I(\mathbf{y}_T; \mathbf{f}_T) \leq C_1 \beta_T \gamma_T \quad \forall T \geq 1,$$

where  $C_1 := 8/\log(1 + \sigma^{-2}) \geq 8\sigma^2$ .

**Proof** By Lemma 5.1 and Lemma 5.2, we have that  $\{r_t^2 \leq 4\beta_t \sigma_{t-1}^2(\mathbf{x}_t) \mid \forall t \geq 1\}$  with probability  $\geq 1 - \delta$ . Now,  $\beta_t$  is nondecreasing, so that

$$\begin{aligned} 4\beta_t \sigma_{t-1}^2(\mathbf{x}_t) &\leq 4\beta_T \sigma^2(\sigma^{-2} \sigma_{t-1}^2(\mathbf{x}_t)) \\ &\leq 4\beta_T \sigma^2 C_2 \log(1 + \sigma^{-2} \sigma_{t-1}^2(\mathbf{x}_t)) \end{aligned}$$

with  $C_2 = \sigma^{-2}/\log(1 + \sigma^{-2}) \geq 1$ , since  $s^2 \leq C_2 \log(1 + s^2)$  for  $s \in [0, \sigma^{-2}]$ , and  $\sigma^{-2} \sigma_{t-1}^2(\mathbf{x}_t) \leq \sigma^{-2} k(\mathbf{x}_t, \mathbf{x}_t) \leq \sigma^{-2}$ . Noting that  $C_1 = 8\sigma^2 C_2$ , the result follows by plugging in the representation of Lemma 5.3. ■

Finally, Theorem 1 is a simple consequence of Lemma 5.4, since  $R_T^2 \leq T \sum_{t=1}^T r_t^2$  by the Cauchy-Schwarz inequality.

## A.2. General Decision Set

Theorem 2 extends the statement of Theorem 1 to the general case of  $D \subset \mathbb{R}^d$  compact. We cannot expect this generalization to work without any assumptions on the kernel  $k(\mathbf{x}, \mathbf{x}')$ . For example, if  $k(\mathbf{x}, \mathbf{x}') = e^{-\|\mathbf{x} - \mathbf{x}'\|}$  (Ornstein-Uhlenbeck), while sample paths  $f$  are a.s. continuous, they are still very erratic:  $f$  is a.s. nondifferentiable almost everywhere, and the process comes with independent increments, a stationary variant of Brownian motion. The additional assumption on  $k$  in Theorem 2 is rather mild and is satisfied by several common kernels, as discussed in Section 4.

Recall that the finite case proof is based on Lemma 5.1 paving the way for Lemma 5.2. However, Lemma 5.1 does not hold for infinite  $D$ . First, let us observe that we have confidence on all decisions actually chosen.

**Lemma 5.5** Pick  $\delta \in (0, 1)$  and set  $\beta_t = 2 \log(\pi_t/\delta)$ , where  $\sum_{t \geq 1} \pi_t^{-1} = 1$ ,  $\pi_t > 0$ . Then,

$$|f(\mathbf{x}_t) - \mu_{t-1}(\mathbf{x}_t)| \leq \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) \quad \forall t \geq 1$$

holds with probability  $\geq 1 - \delta$ .

**Proof** Fix  $t \geq 1$  and  $\mathbf{x} \in D$ . Conditioned on  $\mathbf{y}_{t-1} = (y_1, \dots, y_{t-1})$ ,  $\{\mathbf{x}_1, \dots, \mathbf{x}_{t-1}\}$  are deterministic, and  $f(\mathbf{x}) \sim N(\mu_{t-1}(\mathbf{x}), \sigma_{t-1}^2(\mathbf{x}))$ . As before,  $\Pr\{|f(\mathbf{x}_t) - \mu_{t-1}(\mathbf{x}_t)| > \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t)\} \leq e^{-\beta_t/2}$ . Since  $e^{-\beta_t/2} = \delta/\pi_t$  and using the union bound for  $t \in \mathbb{N}$ , the statement holds. ■

Purely for the sake of analysis, we use a set of discretizations  $D_t \subset D$ , where  $D_t$  will be used at time

$t$  in the analysis. Essentially, we use this to obtain a valid confidence interval on  $\mathbf{x}^*$ . The following lemma provides a confidence bound for these subsets.

**Lemma 5.6** *Pick  $\delta \in (0, 1)$  and set  $\beta_t = 2\log(|D_t|/\pi_t/\delta)$ , where  $\sum_{t \geq 1} \pi_t^{-1} = 1$ ,  $\pi_t > 0$ . Then,*

$$|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \leq \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}) \quad \forall \mathbf{x} \in D_t, \forall t \geq 1$$

*holds with probability  $\geq 1 - \delta$ .*

**Proof** The proof is identical to that in Lemma 5.1, except now we use  $D_t$  at each timestep. ■

Now by assumption and the union bound, we have that

$$\Pr\{\forall j, \forall \mathbf{x} \in D, |\partial f/(\partial x_j)| < L\} \geq 1 - dae^{-L^2/b^2}.$$

which implies that, with probability greater than  $1 - dae^{-L^2/b^2}$ , we have that

$$\forall \mathbf{x} \in D, |f(\mathbf{x}) - f(\mathbf{x}')| \leq L\|\mathbf{x} - \mathbf{x}'\|_1. \quad (9)$$

This allows us to obtain confidence on  $\mathbf{x}^*$  as follows.

Now let us choose a discretization  $D_t$  of size  $(\tau_t)^d$  so that for all  $\mathbf{x} \in D_t$

$$\|\mathbf{x} - [\mathbf{x}]_t\|_1 \leq rd/\tau_t$$

where  $[\mathbf{x}]_t$  denotes the closest point in  $D_t$  to  $\mathbf{x}$ . A sufficient discretization has each coordinate with  $\tau_t$  uniformly spaced points.

**Lemma 5.7** *Pick  $\delta \in (0, 1)$  and set  $\beta_t = 2\log(2\pi_t/\delta) + 4d\log(dtbr\sqrt{\log(2da/\delta)})$ , where  $\sum_{t \geq 1} \pi_t^{-1} = 1$ ,  $\pi_t > 0$ . Let  $\tau_t = dt^2br\sqrt{\log(2da/\delta)}$ . Let  $[\mathbf{x}^*]_t$  denotes the closest point in  $D_t$  to  $\mathbf{x}^*$ . Hence, Then,*

$$|f(\mathbf{x}^*) - \mu_{t-1}([\mathbf{x}^*]_t)| \leq \beta_t^{1/2} \sigma_{t-1}([\mathbf{x}^*]_t) + \frac{1}{t^2} \quad \forall t \geq 1$$

*holds with probability  $\geq 1 - \delta$ .*

**Proof** Using (9), we have that with probability greater than  $1 - \delta/2$ ,

$$\forall \mathbf{x} \in D, |f(\mathbf{x}) - f(\mathbf{x}')| \leq b\sqrt{\log(2da/\delta)}\|\mathbf{x} - \mathbf{x}'\|_1.$$

Hence,

$$\forall \mathbf{x} \in D_t, |f(\mathbf{x}) - f([\mathbf{x}]_t)| \leq rdb\sqrt{\log(2da/\delta)}/\tau_t.$$

Now by choosing  $\tau_t = dt^2br\sqrt{\log(2da/\delta)}$ , we have that

$$\forall \mathbf{x} \in D_t, |f(\mathbf{x}) - f([\mathbf{x}]_t)| \leq \frac{1}{t^2}$$

This implies that  $|D_t| = (dt^2br\sqrt{\log(2da/\delta)})^d$ . Using  $\delta/2$  in Lemma 5.6, we can apply the confidence bound to  $[\mathbf{x}^*]_t$  (as this lives in  $D_t$ ) to obtain the result. ■

Now we are able to bound the regret.

**Lemma 5.8** *Pick  $\delta \in (0, 1)$  and set  $\beta_t = 2\log(4\pi_t/\delta) + 4d\log(dtbr\sqrt{\log(4da/\delta)})$ , where  $\sum_{t \geq 1} \pi_t^{-1} = 1$ ,  $\pi_t > 0$ . Then, with probability greater than  $1 - \delta$ , for all  $t \in \mathbb{N}$ , the regret is bounded as follows:*

$$r_t \leq 2\beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) + \frac{1}{t^2}.$$

**Proof** We use  $\delta/2$  in both Lemma 5.5 and Lemma 5.7, so that these events hold with probability greater than  $1 - \delta$ . Note that the specification of  $\beta_t$  in the above lemma is greater than the specification used in Lemma 5.5 (with  $\delta/2$ ), so this choice is valid.

By definition of  $\mathbf{x}_t$ :  $\mu_{t-1}(\mathbf{x}_t) + \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) \geq \mu_{t-1}([\mathbf{x}^*]_t) + \beta_t^{1/2} \sigma_{t-1}([\mathbf{x}^*]_t)$ . Also, by Lemma 5.7, we have that  $\mu_{t-1}([\mathbf{x}^*]_t) + \beta_t^{1/2} \sigma_{t-1}([\mathbf{x}^*]_t) + 1/t^2 \geq f(\mathbf{x}^*)$ , which implies  $\mu_{t-1}(\mathbf{x}_t) + \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) \geq f(\mathbf{x}^*) - 1/t^2$ . Therefore,

$$\begin{aligned} r_t &= f(\mathbf{x}^*) - f(\mathbf{x}_t) \\ &\leq \beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) + 1/t^2 + \mu_{t-1}(\mathbf{x}_t) - f(\mathbf{x}_t) \\ &\leq 2\beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) + 1/t^2. \end{aligned}$$

which completes the proof. ■

Now we are ready to complete the proof of Theorem 2. As shown in the proof of Lemma 5.4, we have that with probability greater than  $1 - \delta$ ,

$$\sum_{t=1}^T 4\beta_t \sigma_{t-1}^2(\mathbf{x}_t) \leq C_1 \beta_T \gamma_T \quad \forall T \geq 1,$$

so that by Cauchy-Schwarz:

$$\sum_{t=1}^T 2\beta_t^{1/2} \sigma_{t-1}(\mathbf{x}_t) \leq \sqrt{C_1 T \beta_T \gamma_T} \quad \forall T \geq 1,$$

Hence,

$$\sum_{t=1}^T r_t \leq \sqrt{C_1 T \beta_T \gamma_T} + \pi^2/6 \quad \forall T \geq 1,$$

(since  $\sum 1/t^2 = \pi^2/6$ ). Theorem 2 now follows.

Finally, we now discuss the additional assumption on  $k$  in Theorem 2. For samples  $f$  of the GP, consider partial derivatives  $\partial f/(\partial x_j)$  of this sample path for  $j = 1, \dots, d$ . Theorem 5 of Ghosal & Roy (2006)

states that if derivatives up to fourth order exists for  $(\mathbf{x}, \mathbf{x}') \mapsto k(\mathbf{x}, \mathbf{x}')$ , then  $f$  is almost surely continuously differentiable, with  $\partial f / (\partial x_j)$  distributed as Gaussian processes again. Moreover, there are constants  $a, b_j > 0$  such that

$$\Pr \left\{ \sup_{\mathbf{x} \in D} |\partial f / (\partial x_j)| > L \right\} \leq a e^{-b_j L^2}. \quad (10)$$

Picking  $L = [\log(da2/\delta)/\min_j b_j]^{1/2}$ , we have that  $a e^{-b_j L^2} \leq \delta/(2d)$  for all  $j = 1, \dots, d$ , so that for  $K_1 = d^{1/2}L$ , by the mean value theorem, we have  $\Pr\{|f(\mathbf{x}) - f(\mathbf{x}')| \leq K_1 \|\mathbf{x} - \mathbf{x}'\| \forall \mathbf{x}, \mathbf{x}' \in D\} \geq 1 - \delta/2$ .

Also, note that  $K_1 = \mathcal{O}((\log \delta^{-1})^{1/2})$ .

This statement is about the joint distribution of  $f(\cdot)$  and its partial derivatives w.r.t. each component. For a certain event in this sample space, all  $\partial f / (\partial x_j)$  exist, are continuous, and the complement of (10) holds for all  $j$ . Theorem 5 of Ghosal & Roy (2006), together with the union bound, implies that this event has probability  $\geq 1 - \delta/2$ . Derivatives up to fourth order exist for the Gaussian covariance function, and for Matérn kernels with  $\nu > 2$  (Stein, 1999).

## B. Regret Bound for Target Function in RKHS

In this section, we detail a proof of Theorem 3. Recall that in this setting, we do not know the generator of the target function  $f$ , but only a bound on its RKHS norm  $\|f\|_k$ .

Recall the posterior mean function  $\mu_T(\cdot)$  and posterior covariance function  $k_T(\cdot, \cdot)$  from Section 2, conditioned on data  $(\mathbf{x}_t, y_t)$ ,  $t = 1, \dots, T$ . It is easy to see that the RKHS norm corresponding to  $k_T$  is given by

$$\|f\|_{k_T}^2 = \|f\|_k^2 + \sigma^{-2} \sum_{t=1}^T f(\mathbf{x}_t)^2.$$

This implies that  $\mathcal{H}_k(D) = \mathcal{H}_{k_T}(D)$  for any  $T$ , while the RKHS inner products are different:  $\|f\|_{k_T} \geq \|f\|_k$ . Since  $\langle f(\cdot), k_T(\cdot, \mathbf{x}) \rangle_{k_T} = f(\mathbf{x})$  for any  $f \in \mathcal{H}_{k_T}(D)$  by the reproducing property, then

$$\begin{aligned} |\mu_t(\mathbf{x}) - f(\mathbf{x})| &\leq k_T(\mathbf{x}, \mathbf{x})^{1/2} \|\mu_t - f\|_{k_T} \\ &= \sigma_T(\mathbf{x}) \|\mu_t - f\|_{k_T} \end{aligned} \quad (11)$$

by the Cauchy-Schwarz inequality.

Compared to our other results, Theorem 3 is an agnostic statement, in that the assumptions the Bayesian UCB algorithm bases its predictions on differ from how  $f$  and data  $y_t$  are generated. First,  $f$  is not drawn from a GP, but can be an arbitrary function

from  $\mathcal{H}_k(D)$ . Second, while the UCB method assumes that the noise  $\varepsilon_t = y_t - f(\mathbf{x}_t)$  is drawn independently from  $N(0, \sigma^2)$ , the true sequence of noise variables  $\varepsilon_t$  can be a uniformly bounded martingale difference sequence:  $\varepsilon_t \leq \sigma$  for all  $t \in \mathbb{N}$ . All we have to do in order to lift the proof of Theorem 1 to the agnostic setting is to establish an analogue to Lemma 5.1, by way of the following concentration result.

**Theorem 6** *Let  $\delta \in (0, 1)$ . Assume the noise variables  $\varepsilon_t$  are uniformly bounded by  $\sigma$ . Define:*

$$\beta_t = 2\|f\|_k^2 + 300\gamma_t \ln^3(t/\delta),$$

*Then*

$$\Pr \left\{ \forall T, \forall \mathbf{x} \in D, |\mu_T(\mathbf{x}) - f(\mathbf{x})| \leq \beta_{T+1}^{1/2} \sigma_T(\mathbf{x}) \right\} \geq 1 - \delta.$$

### B.1. Concentration of Martingales

In our analysis, we use the following Bernstein-type concentration inequality for martingale differences, due to Freedman (1975) (see also Theorem 3.15 of McDiarmid 1998).

**Theorem 7 (Freedman)** *Suppose  $X_1, \dots, X_T$  is a martingale difference sequence, and  $b$  is an uniform upper bound on the steps  $X_i$ . Let  $V$  denote the sum of conditional variances,*

$$V = \sum_{i=1}^n \mathbf{Var}(X_i | X_1, \dots, X_{i-1}).$$

*Then, for every  $a, v > 0$ ,*

$$\Pr \left\{ \sum X_i \geq a \text{ and } V \leq v \right\} \leq \exp \left( \frac{-a^2}{2v + 2ab/3} \right).$$

### B.2. Proof of Theorem 6

We will show that:

$$\Pr \left\{ \forall T, \|\mu_T - f\|_{k_T}^2 \leq \beta_{T+1} \right\} \geq 1 - \delta.$$

Theorem 6 then follows from (11). Recall that  $\varepsilon_t = y_t - f(\mathbf{x}_t)$ . We will analyze the quantity  $Z_T = \|\mu_T - f\|_{k_T}^2$ , measuring the error of  $\mu_T$  as approximation to  $f$  under the RKHS norm of  $\mathcal{H}_{k_T}(D)$ . The following lemma provides the connection with the information gain. This lemma is important since our concentration argument is an inductive argument — roughly speaking, we condition on getting concentration in the past, in order to achieve good concentration in the future.

**Lemma 7.1** *We have that*

$$\sum_{t=1}^T \min\{\sigma^{-2} \sigma_{t-1}^2(\mathbf{x}_t), \alpha\} \leq \frac{2\alpha}{\log(1 + \alpha)} \gamma_T, \quad \alpha > 0.$$



**Proof** We have that  $\min\{r, \alpha\} \leq (\alpha/\log(1 + \alpha))\log(1 + r)$ . The statement follows from Lemma 5.3. ■

The next lemma bounds the growth of  $Z_T$ . It is formulated in terms of normalized quantities:  $\tilde{\varepsilon}_t = \varepsilon_t/\sigma$ ,  $\tilde{f} = f/\sigma$ ,  $\tilde{\mu}_t = \mu_t/\sigma$ ,  $\tilde{\sigma}_t = \sigma_t/\sigma$ . Also, to ease notation, we will use  $\mu_{t-1}$ ,  $\sigma_{t-1}$  as shorthand for  $\mu_{t-1}(\mathbf{x}_t)$ ,  $\sigma_{t-1}(\mathbf{x}_t)$ .

**Lemma 7.2** For all  $T \in \mathbb{N}$ ,

$$Z_T \leq \|f\|_k^2 + 2 \sum_{t=1}^T \tilde{\varepsilon}_t \frac{\tilde{\mu}_{t-1} - \tilde{f}(\mathbf{x}_t)}{1 + \tilde{\sigma}_{t-1}^2} + \sum_{t=1}^T \tilde{\varepsilon}_t^2 \frac{\tilde{\sigma}_{t-1}^2}{1 + \tilde{\sigma}_{t-1}^2}.$$

**Proof** If  $\alpha_t = (\mathbf{K}_t + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_t$ , then  $\mu_t(\mathbf{x}) = \alpha_t^T \mathbf{k}_t(\mathbf{x})$ . Then,  $\langle \mu_T, f \rangle_k = \mathbf{f}_T^T \alpha_T$ ,  $\|\mu_T\|_k^2 = \mathbf{y}_T^T \alpha_T - \sigma^2 \|\alpha_T\|^2$ . Moreover, for  $t \leq T$ ,  $\mu_T(\mathbf{x}_t) = \delta_t^T \mathbf{K}_T (\mathbf{K}_T + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_T = y_t - \sigma^2 \alpha_t$ . Since  $Z_T = \|\mu_T - f\|_k + \sigma^{-2} \sum_{t \leq T} (\mu_T(\mathbf{x}_t) - f(\mathbf{x}_t))^2$ , we have that

$$Z_T = \|f\|_k^2 - 2 \mathbf{f}_T^T \alpha_T + \mathbf{y}_T^T \alpha_T - \sigma^2 \|\alpha_T\|^2 + \sigma^{-2} \sum_{t=1}^T (\varepsilon_t - \sigma^2 \alpha_t)^2 = \|f\|_k^2 - \mathbf{y}_T^T (\mathbf{K}_T + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_T + \sigma^{-2} \|\varepsilon_T\|^2.$$

Now,  $-\mathbf{y}_T^T (\mathbf{K}_T + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_T \doteq 2 \log P(\mathbf{y}_T)$ , where “ $\doteq$ ” means that we drop determinant terms, thus concentrate on quadratic functions. Since  $\log P(\mathbf{y}_T) = \sum_t \log P(y_t | \mathbf{y}_{<t}) = \sum_t \log N(y_t | \mu_{t-1}(\mathbf{x}_t), \sigma_{t-1}^2(\mathbf{x}_t) + \sigma^2)$ , we have that

$$-\mathbf{y}_T^T (\mathbf{K}_T + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_T = - \sum_t \frac{(y_t - \mu_{t-1})^2}{\sigma^2 + \sigma_{t-1}^2} = 2 \sum_t \varepsilon_t \frac{\mu_{t-1} - f(\mathbf{x}_t)}{\sigma^2 + \sigma_{t-1}^2} - \sum_t \frac{\varepsilon_t^2 \tilde{\sigma}_{t-1}^2}{\sigma^2 + \sigma_{t-1}^2} - R$$

with  $R = \sum_t (\mu_{t-1} - f(\mathbf{x}_t))^2 / (\sigma^2 + \sigma_{t-1}^2) \geq 0$ . Dropping  $-R$  and changing to normalized quantities concludes the proof. ■

We now define a useful martingale difference sequence. First, it is convenient to define an “escape event”  $E_T$  as:

$$E_T = \mathbb{I}\{Z_t \leq \beta_{t+1} \text{ for all } t \leq T\}$$

where  $\mathbb{I}\{\cdot\}$  is the indicator function. Define the random variables  $M_t$  by

$$M_t = 2\tilde{\varepsilon}_t E_{t-1} \frac{\tilde{\mu}_{t-1} - \tilde{f}(\mathbf{x}_t)}{1 + \tilde{\sigma}_{t-1}^2}.$$

Now, since  $\tilde{\varepsilon}_t$  is a martingale difference sequence with respect to the histories  $\mathcal{H}_{<t}$  and  $M_t/\tilde{\varepsilon}_t$  is deterministic given  $\mathcal{H}_{<t}$ ,  $M_t$  is a martingale difference sequence as well. Next, we show that with high probability, the associated martingale  $\sum_{t=1}^T M_t$  does not grow too large.

**Lemma 7.3** Given  $\delta \in (0, 1)$  and  $\beta_t$  as defined in in Theorem 6, we have that

$$\Pr \left\{ \forall T, \sum_{t=1}^T M_t \leq \beta_{T+1}/2 \right\} \geq 1 - \delta,$$

The proof is given below in Section B.3. Equipped with this lemma, we can prove Theorem 6.

**Proof** [of Theorem 6] It suffices to show that the high-probability event described in Lemma 7.3 is contained in the support of  $E_T$  for every  $T$ . We prove the latter by induction on  $T$ .

By Lemma 7.2 and the definition of  $\beta_1$ , we know that  $Z_0 \leq \|f\|_k \leq \beta_1$ . Hence  $E_0 = 1$  always. Now suppose the high-probability event of Lemma 7.3 holds, in particular  $\sum_{t=1}^T M_t \leq \beta_{T+1}/2$ . For the inductive hypothesis, assume  $E_{T-1} = 1$ . Using this and Lemma 7.2:

$$\begin{aligned} Z_T &\leq \|f\|_k^2 + 2 \sum_{t=1}^T \frac{\tilde{\varepsilon}_t (\tilde{\mu}_{t-1} - \tilde{f}(\mathbf{x}_t))}{1 + \tilde{\sigma}_{t-1}^2} + \sum_{t=1}^T \frac{\tilde{\varepsilon}_t^2 \tilde{\sigma}_{t-1}^2}{1 + \tilde{\sigma}_{t-1}^2} \\ &= \|f\|_k^2 + \sum_{t=1}^T M_t + \sum_{t=1}^T \tilde{\varepsilon}_t^2 \frac{\tilde{\sigma}_{t-1}^2}{1 + \tilde{\sigma}_{t-1}^2} \\ &\leq \|f\|_k^2 + \beta_{T+1}/2 + \sum_{t=1}^T \min\{\tilde{\sigma}_{t-1}^2, 1\} \\ &\leq \|f\|_k^2 + \beta_{T+1}/2 + (2/\log 2)\gamma_T \leq \beta_{T+1}. \end{aligned}$$

The equality in the second step uses the inductive hypothesis. Thus we have shown  $E_T = 1$ , completing the induction. ■

### B.3. Concentration

What remains to be shown is Lemma 7.3. While the step sizes  $|M_t|$  are uniformly bounded, a standard application of the Hoeffding-Azuma inequality leads to a bound of  $T^{3/4}$ , too large for our purpose. We use the more specific Theorem 7 instead, which requires to control the conditional variances rather than the marginal variances which can be much larger.

**Proof** [of Lemma 7.3] Let us first obtain upper bounds

on the step sizes of our martingale.

$$\begin{aligned}
|M_t| &= 2|\tilde{\varepsilon}_t|E_{t-1} \frac{|\tilde{\mu}_{t-1} - \tilde{f}(\mathbf{x}_t)|}{1 + \tilde{\sigma}_{t-1}^2} \\
&\leq 2|\tilde{\varepsilon}_t|E_{t-1} \frac{\beta_t^{1/2}\tilde{\sigma}_{t-1}}{1 + \tilde{\sigma}_{t-1}^2} \\
&\leq 2|\tilde{\varepsilon}_t|E_{t-1}\beta_t^{1/2} \min\{\tilde{\sigma}_{t-1}, 1/2\}, \quad (12)
\end{aligned}$$

where the first inequality follows from the definition of  $E_t$ . Moreover,  $r/(1+r^2) \leq \min\{r, 1/2\}$  for  $r \geq 0$ . Therefore,  $|M_t| \leq \beta_t^{1/2}$ , since  $|\tilde{\varepsilon}_t| \leq 1$  and  $\beta_t$  is nondecreasing. Next, we bound the sum of the conditional variances of the martingale:

$$\begin{aligned}
V_T &:= \sum_{t=1}^T \mathbf{Var}(M_t | M_1 \dots M_{t-1}) \\
&\leq \sum_{t=1}^T 4|\tilde{\varepsilon}_t|^2 E_{t-1} \beta_t \min\{\tilde{\sigma}_{t-1}^2, 1/4\} \\
&\leq 4\beta_T \sum_{t=1}^T E_{t-1} \min\{\tilde{\sigma}_{t-1}^2, 1/4\} \quad |\tilde{\varepsilon}_t| \leq 1 \\
&\leq 9\beta_T \gamma_T.
\end{aligned}$$

In the last line, we used Lemma 7.1 with  $\alpha = 1/4$ , noting that  $8\alpha/\log(1+\alpha) \leq 9$ . Since we have established that the sum of conditional variances,  $V_T$ , is always bounded by  $9\beta_T \gamma_T$ , we can apply Theorem 7 with parameters  $a = \beta_{T+1}/2$ ,  $b = \beta_{T+1}^{1/2}$  and  $v = 9\beta_T \gamma_T$  to get

$$\begin{aligned}
&\Pr \left\{ \sum_{t=1}^T M_t \geq \beta_{T+1}/2 \right\} \\
&= \Pr \left\{ \sum_{t=1}^T M_t \geq \beta_{T+1}/2 \text{ and } V_T \leq 9\beta_T \gamma_T \right\} \\
&\leq \exp \left( \frac{-(\beta_{T+1}/2)^2}{2(9\beta_T \gamma_T) + \frac{2}{3}(\beta_{T+1}/2)\beta_{T+1}^{1/2}} \right) \\
&= \exp \left( \frac{-\beta_{T+1}}{72\gamma_T + \frac{4}{3}\beta_{T+1}^{1/2}} \right) \\
&\leq \max \left\{ \exp \left( \frac{-\beta_{T+1}}{144\gamma_T} \right), \exp \left( \frac{-3\beta_{T+1}^{1/2}}{8} \right) \right\}.
\end{aligned}$$

Note that our choice of  $\beta_{T+1}$  satisfies:

$$\max \left\{ 144\gamma_T \log(T^2/\delta), ((8/3) \log(T^2/\delta))^2 \right\} \leq \beta_{T+1}.$$

Therefore, the previous probability is bounded by  $\delta/T^2$ , whereas the last inequality follows from the definition of  $\beta_{T+1}$ . With a final application of the union

bound:

$$\begin{aligned}
&\Pr \left\{ \sum_{t=1}^T M_t \geq \beta_{T+1}/2 \text{ for some } T \right\} \\
&\leq \sum_{T \geq 1} \Pr \left\{ \sum_{t=1}^T M_t \geq \beta_{T+1}/2 \right\} \\
&\leq \sum_{T \geq 2} \delta/T^2 \leq \delta(\pi^2/6 - 1) \leq \delta,
\end{aligned}$$

completing the proof of Lemma 7.3.  $\blacksquare$

## C. Bounds on Information Gain

In this section, we show how to bound  $\gamma_T$ , the maximum information gain after  $T$  rounds, for compact  $D \subset \mathbb{R}^d$  (assumptions of Theorem 2) and several commonly used covariance functions. In this section, we assume<sup>4</sup> that  $k(\mathbf{x}, \mathbf{x}) = 1$  for all  $\mathbf{x} \in D$ .

The plan of attack is as follows. First, we note that the argument of  $\gamma_T$ ,  $I(\mathbf{y}_A; \mathbf{f}_A)$  is a submodular function, so  $\gamma_T$  can be bounded by the value obtained by greedy maximization. Next, we use a discretization  $D_T \subset D$  with  $n_T = |D_T| = T^\tau$  with nearest neighbour distance  $o(1)$ , consider the kernel matrix  $\mathbf{K}_{D_T} \in \mathbb{R}^{n_T \times n_T}$ , and bound  $\gamma_T$  by an expression involving the eigenvalues  $\{\hat{\lambda}_t\}$  of this matrix, which is done by a further relaxation of the greedy procedure. Finally, we bound this empirical expression in terms of the kernel operator eigenvalues of  $k$  w.r.t. the uniform distribution on  $D$ . Asymptotic expressions for the latter are reviewed in Seeger et al. (2008), which we plug in to obtain our results. A key step in this argument is to ensure the existence of a discretization  $D_T$ , for which tails of the empirical spectrum can be bounded by tails of the process spectrum. We will invoke the probabilistic method for that.

### C.1. Greedy Maximization and Discretization

In this section, we fix  $T \in \mathbb{N}$  and assume the existence of a discretization  $D_T \subset D$ ,  $n_T = |D_T|$  on the order of  $T^\tau$ , such that:

$$\forall \mathbf{x} \in D \exists [\mathbf{x}]_T \in D_T : \|\mathbf{x} - [\mathbf{x}]_T\| = \mathcal{O}(T^{-\tau/d}). \quad (13)$$

We come back to the choice of  $D_T$  below. We restrict the information gain to subsets  $A \subset D_T$ :

$$\tilde{\gamma}_T = \max_{A \subset D_T, |A|=T} I(\mathbf{y}_A; \mathbf{f}_A).$$

Of course,  $\tilde{\gamma}_T \leq \gamma_T$ , but we can bound the slack.

<sup>4</sup> Without loss in generality. We use this assumption below to ensure that  $n_T^{-1} \text{tr} \mathbf{K}_{D_T} = \int k(\mathbf{x}, \mathbf{x}) d\mathbf{x}$ . If  $k(\mathbf{x}, \mathbf{x})$  is not constant, this is approximately true by the law of large numbers, and our result below remains valid.

**Lemma 7.4** *Under the assumptions of Theorem 2, the information gain  $F_T(\{\mathbf{x}_t\}) = (1/2) \log |\mathbf{I} + \sigma^{-2} \mathbf{K}_{\{\mathbf{x}_t\}}|$  is uniformly Lipschitz-continuous in each component  $\mathbf{x}_t \in D$ .*

**Proof** The assumptions of Theorem 2 imply that the kernel  $K(\mathbf{x}, \mathbf{x}')$  is continuously differentiable. The result follows from the fact that  $F_T(\{\mathbf{x}_t\})$  is continuously differentiable in the kernel matrix  $\mathbf{K}_{\{\mathbf{x}_t\}}$ . ■

**Lemma 7.5** *Let  $D_T$  be a discretization of  $D$  such that (13) holds. Under the assumptions of Theorem 2, we have that*

$$0 \leq \gamma_T - \tilde{\gamma}_T = \mathcal{O}(T^{1-\tau/d}).$$

**Proof** Fix  $T \in \mathbb{N}$ , and let  $A = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$  be a maximizer for  $\gamma_T$ . Consider neighbours  $[\mathbf{x}_t]_T \in D_T$  according to (13),  $[A]_T = \{[\mathbf{x}_t]_T\}$ . Then,

$$0 \leq \gamma_T - \tilde{\gamma}_T \leq \gamma_T - \mathbf{I}(\mathbf{y}_{[A]_T}; \mathbf{f}_{[A]_T}) = F_T(A) - F_T([A]_T),$$

where  $F_T(\{\mathbf{x}_t\}) = (1/2) \log |\mathbf{I} + \sigma^{-2} \mathbf{K}_{\{\mathbf{x}_t\}}|$ . By Lemma 7.4,  $F_T$  is uniformly Lipschitz-continuous in each component, so that  $|\gamma_T - \mathbf{I}(\mathbf{y}_{[A]_T}; \mathbf{f}_{[A]_T})| = \mathcal{O}(T \max_t \|\mathbf{x}_t - [\mathbf{x}_t]_T\|) = \mathcal{O}(T^{1-\tau/d})$  by (13) and the mean value theorem. ■

We concentrate on  $\tilde{\gamma}_T$  in the sequel. Let  $\mathbf{K}_{D_T} = [k(\mathbf{x}, \mathbf{x}')]_{\mathbf{x}, \mathbf{x}' \in D_T}$  be the kernel matrix over the entire  $D_T$ , and  $\mathbf{K}_{D_T} = \mathbf{U} \hat{\Lambda} \mathbf{U}^T$  its eigendecomposition, with  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$  and  $\mathbf{U} = [\mathbf{u}_1 \mathbf{u}_2 \dots]$  orthonormal. Here, if  $T > n_T$ , define  $\hat{\lambda}_t = 0$  for  $t = n_T + 1, \dots, T$ . Information gain maximization over a finite  $D_T$  can be described in terms of a simple linear-Gaussian model over the unknown  $\mathbf{f} \in \mathbb{R}^{n_T}$ , with prior  $P(\mathbf{f}) = N(\mathbf{0}, \mathbf{K}_{D_T})$  and likelihood potentials  $P(y_t | \mathbf{f}) = N(\mathbf{v}_t^T \mathbf{f}, \sigma^2)$  with unit-norm features,  $\|\mathbf{v}_t\| = 1$ . With the following lemma, we upper-bound  $\tilde{\gamma}_T$  by way of two relaxations.

**Lemma 7.6** *For any  $T \geq 1$ , we have that*

$$\tilde{\gamma}_T \leq \frac{1/2}{1 - e^{-1}} \max_{m_1, \dots, m_T} \sum_{t=1}^T \log(1 + \sigma^{-2} m_t \hat{\lambda}_t),$$

*subject to  $m_t \in \mathbb{N}$ ,  $\sum_t m_t = T$ , where  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots$  is the spectrum of the kernel matrix  $\mathbf{K}_{D_T}$ . Here, if  $T > n_T$ , then  $m_t = 0$  for  $t > n_T$ .*

**Proof** As shown by Krause & Guestrin (2005), the function  $F(A) = \mathbf{I}(\mathbf{y}_A; \mathbf{f})$  is submodular. In

the particular case considered here, this can be seen as follows:  $F(A) = H(\mathbf{y}_A) - H(\mathbf{y}_A | \mathbf{f})$ , where the entropy  $H(\mathbf{y}_A)$  is a (not-necessarily monotonic) submodular function in  $A$ , and since the noise is conditionally independent given  $\mathbf{f}$ ,  $H(\mathbf{y}_A | \mathbf{f})$  is an additive (modular) function in  $A$ . Subtracting a modular function preserves submodularity, thus  $F(A)$  is submodular. Furthermore, the information gain is monotonic in  $A$  (i.e.,  $F(A) \leq F(B)$  whenever  $A \subseteq B$ ) (Cover & Thomas, 1991). Thus, we can apply the result of Nemhauser et al. (1978)<sup>5</sup> which guarantees that  $\tilde{\gamma}_T$  is upper-bounded by  $1/(1 - 1/e)$  times the value the greedy maximization algorithm attains. The latter chooses features of the form  $\mathbf{v}_t = \delta_{\mathbf{x}_t} = [\mathbf{I}_{\{\mathbf{x}=\mathbf{x}_t\}}]$  in each round,  $\mathbf{x}_t \in D_T$ . We upper-bound the greedy maximum once more by relaxing these constraints to  $\|\mathbf{v}_t\| = 1$  only. In the remainder of the proof, we concentrate on this relaxed greedy procedure. Suppose that up to round  $t$ , it chose  $\mathbf{v}_1, \dots, \mathbf{v}_{t-1}$ . The posterior  $P(\mathbf{f} | \mathbf{y}_{t-1})$  has inverse covariance matrix  $\Sigma_{t-1}^{-1} = \mathbf{K}_{D_T}^{-1} + \sigma^{-2} \mathbf{V}_{t-1} \mathbf{V}_{t-1}^T$ ,  $\mathbf{V}_{t-1} = [\mathbf{v}_1 \dots \mathbf{v}_{t-1}]$ , and the greedy procedure selects  $\mathbf{v}$  so to maximize the variance  $\mathbf{v}^T \Sigma_{t-1} \mathbf{v}$ : the eigenvector corresponding to  $\Sigma_{t-1}$ 's largest eigenvalue (by the Rayleigh-Ritz theorem). Since  $\Sigma_0 = \mathbf{K}_{D_T}$ , then  $\mathbf{v}_1 = \mathbf{u}_1$ . Moreover, if all  $\mathbf{v}_{t'}, t' < t$ , have been chosen among  $\mathbf{U}$ 's columns, then by the inverse covariance expression just given,  $\mathbf{K}_{D_T}$  and  $\Sigma_{t-1}$  have the same eigenvectors, so that  $\mathbf{v}_t$  is a column of  $\mathbf{U}$  as well. For example, if  $\mathbf{v}_t = \mathbf{u}_j$ , then comparing  $\Sigma_{t-1}$  and  $\Sigma_t$ , all eigenvalues other than the  $j$ -th remain the same, while the latter is shrunk. Therefore, after  $T$  rounds of the relaxed greedy procedure:  $\mathbf{v}_t \in \{\mathbf{u}_1, \dots, \mathbf{u}_{\min\{T, n_T\}}\}$ ,  $t = 1, \dots, T$ : at most the leading  $T$  eigenvectors of  $\mathbf{K}_{D_T}$  can have been selected (possibly multiple times). If  $m_t$  denotes the number that the  $t$ -th column of  $\mathbf{U}$  has been selected, we obtain the theorem statement by a final bounding step. ■

## C.2. From Empirical to Process Eigenvalues

The final step will be to relate the empirical spectrum  $\{\hat{\lambda}_t\}$  to the kernel operator spectrum. Since  $\log(1 + \sigma^{-2} m_t \hat{\lambda}_t) \leq \sigma^{-2} m_t \hat{\lambda}_t$  in Theorem 7.6, we will mainly be interested in relating the tail sums of the spectra. Let  $\mu(\mathbf{x}) = \mathcal{V}(D)^{-1} \mathbf{I}_{\{\mathbf{x} \in D\}}$  be the uniform distribution on  $D$ ,  $\mathcal{V}(D) = \int_{\mathbf{x} \in D} d\mathbf{x}$ , and assume that  $k$  is continuous. Note that  $\int k(\mathbf{x}, \mathbf{x}) \mu(\mathbf{x}) d\mathbf{x} = 1$  by our assumption  $k(\mathbf{x}, \mathbf{x}) = 1$ , so that  $k$  is Hilbert-

<sup>5</sup>While the result of Nemhauser et al. (1978) is stated in terms of finite sets, it extends to infinite sets as long as the greedy selection can be implemented efficiently.

Schmidt on  $L_2(\mu)$ . Then, Mercer's theorem (Wahba, 1990) states that the corresponding kernel operator has a discrete eigenspectrum  $\{(\lambda_s, \phi_s(\cdot))\}$ , and

$$k(\mathbf{x}, \mathbf{x}') = \sum_{s \geq 1} \lambda_s \phi_s(\mathbf{x}) \phi_s(\mathbf{x}'),$$

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ , and  $\mathbb{E}_\mu[\phi_s(\mathbf{x})\phi_t(\mathbf{x})] = \delta_{s,t}$ . Moreover,  $\sum_{s \geq 1} \lambda_s^2 < \infty$ , and the expansion of  $k$  converges absolutely and uniformly on  $D \times D$ . Note that  $\sum_{s \geq 1} \lambda_s = \sum_{s \geq 1} \lambda_s \mathbb{E}_\mu[\phi_s(\mathbf{x})^2] = \int K(\mathbf{x}, \mathbf{x}) \mu(\mathbf{x}) d\mathbf{x} = 1$ . In order to proceed from Theorem 7.6, we have to pick a discretization  $D_T$  for which (13) holds, and for which  $\sum_{t > T_*} \hat{\lambda}_t$  is not much larger than  $\sum_{t > T_*} \lambda_t$ . With the following lemma, we determine sizes  $n_T$  for which such discretizations exist.

**Lemma 7.7** Fix  $T \in \mathbb{N}$ ,  $\delta > 0$  and  $\varepsilon > 0$ . There exists a discretization  $D_T \subset D$  of size

$$n_T = \mathcal{V}(D)(\varepsilon/\sqrt{d})^{-d} [\log(1/\delta) + d \log(\sqrt{d}/\varepsilon) + \log \mathcal{V}(D)]$$

which fulfils the following requirements:

- $\varepsilon$ -denseness: For any  $\mathbf{x} \in D$ , there exists  $[\mathbf{x}]_T \in D_T$  such that  $\|\mathbf{x} - [\mathbf{x}]_T\| \leq \varepsilon$ .
- If  $\text{spec}(\mathbf{K}_{D_T}) = \{\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots\}$ , then for any  $T_* = 1, \dots, n_T$ :

$$n_T^{-1} \sum_{t=1}^{T_*} \hat{\lambda}_t \geq \sum_{t=1}^{T_*} \lambda_t - \delta.$$

**Proof** First, if we draw  $n_T$  samples  $\tilde{\mathbf{x}}_j \sim \mu(\mathbf{x})$  independently at random, then  $D_T = \{\tilde{\mathbf{x}}_j\}$  is  $\varepsilon$ -dense with probability  $\geq 1 - \delta$ . Namely, cover  $D$  with  $N = \mathcal{V}(D)(\varepsilon/\sqrt{d})^{-d}$  hypercubes of sidelength  $\varepsilon/\sqrt{d}$ , within which the maximum Euclidean distance is  $\varepsilon$ . The probability of not hitting at least one cell is upper-bounded by  $N(1 - 1/N)^{n_T}$ . Since  $\log(1 - 1/N) \leq -1/N$ , this is upper-bounded by  $\delta$  if  $n_T \geq N \log(N/\delta)$ .

Now, let  $S = n_T^{-1} \sum_{t=1}^{T_*} \hat{\lambda}_t$ . Shawe-Taylor et al. (2005) show that  $\mathbb{E}[S] \geq \sum_{t=1}^{T_*} \lambda_t$ . If  $\mathcal{C}$  is the event  $\{D_T \text{ is } \varepsilon\text{-dense}\}$ , then  $\Pr(\mathcal{C}) \geq 1 - \delta$ . Since  $S \leq n_T^{-1} \text{tr} \mathbf{K}_{D_T} = 1$  in any case, we have that  $\mathbb{E}[S|\mathcal{C}] \geq \mathbb{E}[S] - \Pr(\mathcal{C}^c) \geq \sum_{t=1}^{T_*} \lambda_t - \delta$ . By the probabilistic method, there must exist some  $D_T$  for which  $\mathcal{C}$  and the latter inequality holds. ■

The following lemma, the equivalent of Theorem 4 in the context here, is a direct consequence of Lemma 7.6.

**Lemma 7.8** Let  $D_T$  be some discretization of  $D$ ,

$n_T = |D_T|$ . Then, for any  $T_* = 1, \dots, \min\{T, n_T\}$ :

$$\tilde{\gamma}_T \leq \frac{1/2}{1 - e^{-1}} \max_{r=1, \dots, T} \left( T_* \log(r n_T / \sigma^2) + (T - r) \sigma^{-2} \sum_{t=T_*+1}^{n_T} \hat{\lambda}_t \right).$$

**Proof** We split the right hand side in Lemma 7.6 at  $t = T_*$ . Let  $r = \sum_{t \leq T_*} m_t$ . For  $t \leq T_*$ :  $\log(1 + m_t \hat{\lambda}_t / \sigma^2) \leq \log(r n_T / \sigma^2)$ , since  $\hat{\lambda}_t \leq n_T$ . For  $t > T_*$ :  $\log(1 + m_t \hat{\lambda}_t / \sigma^2) \leq m_t \hat{\lambda}_t / \sigma^2 \leq (T - r) \hat{\lambda}_t / \sigma^2$ . ■

The following theorem describes our “recipe” for obtaining bounds on  $\gamma_T$  for a particular kernel  $k$ , given that tail bounds on  $B_k(T_*) = \sum_{s > T_*} \lambda_s$  are known.

**Theorem 8** Suppose that  $D \subset \mathbb{R}^d$  is compact, and  $k(\mathbf{x}, \mathbf{x}')$  is a covariance function for which the additional assumption of Theorem 2 holds. Moreover, let  $B_k(T_*) = \sum_{s > T_*} \lambda_s$ , where  $\{\lambda_s\}$  is the operator spectrum of  $k$  with respect to the uniform distribution over  $D$ . Pick  $\tau > 0$ , and let  $n_T = C_4 T^\tau (\log T)$  with  $C_4 = 2\mathcal{V}(D)(2\tau + 1)$ . Then, the following bound holds true:

$$\begin{aligned} \gamma_T \leq & \frac{1/2}{1 - e^{-1}} \max_{r=1, \dots, T} \left( T_* \log(r n_T / \sigma^2) \right. \\ & \left. + C_4 \sigma^{-2} (1 - r/T) (\log T) (T^{\tau+1} B_k(T_*) + 1) \right) \\ & + \mathcal{O}(T^{1-\tau/d}) \end{aligned}$$

for any  $T_* \in \{1, \dots, n_T\}$ .

**Proof** Let  $\varepsilon = d^{1/2} T^{-\tau/d}$  and  $\delta = T^{-(\tau+1)}$ . Lemma 7.7 provides the existence of a discretization  $D_T$  of size  $n_T$  which is  $\varepsilon$ -dense, and for which  $n_T^{-1} \sum_{t=1}^{T_*} \hat{\lambda}_t \geq \sum_{t=1}^{T_*} \lambda_t - \delta$ . Since  $n_T^{-1} \sum_{t=1}^{n_T} \hat{\lambda}_t = 1 = \sum_{t \geq 1} \lambda_t$ , then  $\sum_{t > T_*} \hat{\lambda}_t \leq B_k(T_*) + \delta$ . The statement follows by using Lemma 7.8 with these bounds, and finally employing Lemma 7.5. ■

### C.3. Proof of Theorem 5

In this section, we instantiate Theorem 8 in order to obtain bounds on  $\gamma_T$  for Squared Exponential and Matérn kernels, results which are summarized in Theorem 5.

#### SQUARED EXPONENTIAL KERNEL

For the Squared Exponential kernel  $k$ ,  $B_k(T_*)$  is given by Seeger et al. (2008). While  $\mu(\mathbf{x})$  was Gaussian



there, the same decay rate holds for  $\lambda_s$  w.r.t. uniform  $\mu(\mathbf{x})$ , while constants might change. In hindsight, it turns out that  $\tau = d$  is the optimal choice for the discretization size, rendering the second term in Theorem 5 to be  $\mathcal{O}(1)$ , which is subdominant and will be neglected in the sequel. We have that  $\lambda_s \leq cB^{s^{1/d}}$  with  $B < 1$ . Following their analysis,

$$B_k(T_*) \leq c(d!) \alpha^{-d} e^{-\beta} \sum_{j=0}^{d-1} (j!)^{-1} \beta^j,$$

where  $\alpha = -\log B$ ,  $\beta = \alpha T_*^{1/d}$ . Therefore,  $B_k(T_*) = \mathcal{O}(e^{-\beta} \beta^{d-1})$ ,  $\beta = \alpha T_*^{1/d}$ .

We have to pick  $T_*$  such that  $e^{-\beta}$  is not much larger than  $(Tn_T)^{-1}$ . Suppose that  $T_* = [\log(Tn_T)/\alpha]^d$ , so that  $e^{-\beta} = (Tn_T)^{-1}$ ,  $\beta = \log(Tn_T)$ . The bound becomes

$$\begin{aligned} \max_{r=1, \dots, T} & \left( T_* \log(rn_T/\sigma^2) \right. \\ & \left. + \sigma^{-2}(1 - r/T)(C_5 \beta^{d-1} + C_4(\log T)) \right) \end{aligned}$$

with  $n_T = C_4 T^d (\log T)$ . The first part dominates, so that  $r = T$  and  $\gamma_T = \mathcal{O}([\log(T^{d+1}(\log T))]^{d+1}) = \mathcal{O}((\log T)^{d+1})$ . This should be compared with  $\mathbb{E}[\mathbf{I}(\mathbf{y}_T; \mathbf{f}_T)] = \mathcal{O}((\log T)^{d+1})$  given by Seeger et al. (2008), where the  $\mathbf{x}_t$  are drawn independently from a Gaussian base distribution. At least restricted to a compact set  $D$ , we obtain the same expression to leading order for  $\max_{\{\mathbf{x}_t\}} \mathbf{I}(\mathbf{y}_T; \mathbf{f}_T)$ .

#### MATÉRN KERNELS

For Matérn kernels  $k$  with roughness parameter  $\nu$ ,  $B_k(T_*)$  is given by Seeger et al. (2008) for the uniform base distribution  $\mu(\mathbf{x})$  on  $D$ . Namely,  $\lambda_s \leq cs^{-(2\nu+d)/d}$  for almost all  $s \in \mathbb{N}$ , and  $B_k(T_*) = \mathcal{O}(T_*^{1-(2\nu+d)/d})$ . To match terms in the  $\tilde{\gamma}_T$  bound, we choose  $T_* = (Tn_T)^{d/(2\nu+d)} (\log(Tn_T))^\kappa$  ( $\kappa$  chosen below), so that the bound becomes

$$\begin{aligned} \max_{r=1, \dots, T} & \left( T_* \log(rn_T/\sigma^2) + \sigma^{-2}(1 - r/T) \right. \\ & \times (C_5 T_* (\log(Tn_T))^{-\kappa(2\nu+d)/d} + C_4(\log T)) \Big) \\ & + \mathcal{O}(T^{1-\tau/d}) \end{aligned}$$

with  $n_T = C_4 T^\tau (\log T)$ . For  $\kappa = -d/(2\nu+d)$ , we obtain that the maximum over  $r$  is  $\mathcal{O}(T_* \log(Tn_T)) = \mathcal{O}(T^{(\tau+1)d/(2\nu+d)} (\log T))$ . Finally, we choose  $\tau = 2\nu d/(2\nu+d(d+1))$  to match this term with  $\mathcal{O}(T^{1-\tau/d})$ . Plugging this in, we have  $\gamma_T = \mathcal{O}(T^{1-2\eta} (\log T))$ ,  $\eta = \frac{\nu}{2\nu+d(d+1)}$ . Together with Theorem 2 (for  $\nu > 2$ ), we have that  $R_T = \mathcal{O}^*(T^{1-\eta})$  (suppressing log factors): for any  $\nu > 2$  and any dimension  $d$ , the GP-

UCB algorithm is guaranteed to be no-regret in this case with arbitrarily high probability.

How does this bound compare to the bound on  $\mathbb{E}[\mathbf{I}(\mathbf{y}_T; \mathbf{f}_T)]$  given by Seeger et al. (2008)? Here,  $\gamma_T = \mathcal{O}(T^{d(d+1)/(2\nu+d(d+1))} (\log T))$ , while  $\mathbb{E}[\mathbf{I}(\mathbf{y}_T; \mathbf{f}_T)] = \mathcal{O}(T^{d/(2\nu+d)} (\log T)^{2\nu/(2\nu+d)})$ .

#### LINEAR KERNEL

For linear kernels  $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$ ,  $\mathbf{x} \in \mathbb{R}^d$  with  $\|\mathbf{x}\| \leq 1$ , we can bound  $\gamma_T$  directly. Let  $\mathbf{X}_T = [\mathbf{x}_1 \dots, \mathbf{x}_T] \in \mathbb{R}^{d \times T}$  with all  $\|\mathbf{x}_t\| \leq 1$ . Now,

$$\begin{aligned} \log |\mathbf{I} + \sigma^{-2} \mathbf{X}_T^T \mathbf{X}_T| &= \log |\mathbf{I} + \sigma^{-2} \mathbf{X}_T \mathbf{X}_T^T| \\ &\leq \log |\mathbf{I} + \sigma^{-2} \mathbf{D}| \end{aligned}$$

with  $\mathbf{D} = \text{diag} \text{diag}^{-1}(\mathbf{X}_T \mathbf{X}_T^T)$ , by Hadamard's inequality. The largest eigenvalue  $\hat{\lambda}_1$  of  $\mathbf{X}_T \mathbf{X}_T^T$  is  $\mathcal{O}(T)$ , so that

$$\log |\mathbf{I} + \sigma^{-2} \mathbf{X}_T^T \mathbf{X}_T| \leq d \log(1 + \sigma^{-2} \hat{\lambda}_1),$$

and  $\gamma_T = \mathcal{O}(d \log T)$ .