Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger

***Abstract*—Many applications require optimizing an unknown, noisy function that is expensive to evaluate. We formalize this task as a multiarmed bandit problem, where the payoff function is either sampled from a Gaussian process (GP) or has low norm in a reproducing kernel Hilbert space. We resolve the important open problem of deriving regret bounds for this setting, which imply novel convergence rates for GP optimization. We analyze an intuitive Gaussian process upper confidence bound (GP-UCB) 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.**

***Index Terms*—Bandit problems, Bayesian prediction, experimental design, Gaussian process (GP), information gain, nonparametric statistics, online learning, regret bound, statistical learning.**

I. 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 [2] or learning optimal control strategies for robots [3]. Predominant approaches to this problem include the multiarmed bandit paradigm [4], where the goal is to maximize cumulative reward by optimally balancing exploration and exploitation, and experimental design [5], 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 [6], 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 multiarmed bandit setting, where f is sampled from a GP distribution or has low "complexity" measured in terms of its reproducing kernel Hilbert space (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 [7], [9]. While objectives are different in the multiarmed bandit and experimental design paradigms, 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.* [10] provide a near-complete characterization 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 [11] and knowledge about kernel operator spectra.

Compared to an earlier version of [1], this paper is significantly expanded, including detailed proofs, additional explanations (e.g., Fig. 3), and more comprehensive experimental demonstration of the performance of the GP-UCB algorithm.

Kleinberg *et al.* [12] provide regret bounds under weaker and less configurable assumptions (only Lipschitz continuity w.r.t.

Manuscript received October 17, 2010; accepted September 27, 2011. Date of publication January 24, 2012; date of current version April 17, 2012. This work was supported in part by the Office of Naval Research under Grant N00014-09-1-1044, in part by the National Science Foundation under Grants CNS-0932392 and IIS-0953413, in part by a gift from Microsoft Corporation, and in part by the Excellence Initiative of the German Research Foundation (DFG). This manuscript is an extended version of our paper that appeared in ICML 2010 [1].

N. Srinivas is with the California Institute of Technology, Pasadena, CA 91125 USA (e-mail: niranjan@caltech.edu).

A. Krause is with the Swiss Federal Institute of Technology, Zürich 8006, Switzerland (e-mail: krausea@ethz.ch).

S. M. Kakade is with Microsoft Research, New England, Cambridge, MA 02142 USA, and also with the Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104-6340 USA (e-mail: skakade@wharton.upenn.edu).

M. Seeger is with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland (e-mail: matthias.seeger@epfl.ch).

Communicated by D. Palomar, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2182033

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

Fig. 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 ν is the Matérn parameter.

a metric is assumed; Bubeck *et al.* [13] 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). In addition, the GP approach allows for natural extensions. Subsequently, to the initial version of this paper [1], Krause and Ong [14] show how the approach can be extended to address the GP contextual bandit setting, in which the decision maker is provided with context, and needs to learn an optimal mapping from context to action. They further show how the information-theoretic bounds extend to the stronger notion of contextual regret.

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 (EI)[15], most probable improvement (MPI)[16], and upper confidence sampling [7]) and successfully applied in practice (cf., [3]). Brochu *et al.* [17] provide a comprehensive review of and motivation for Bayesian optimization using GPs. The efficient global optimization (EGO) algorithm for optimizing expensive black-box functions was proposed by Jones *et al.* [18] and extended to GPs by Huang *et al.*[19]. Little is known about theoretical performance of GP optimization. While convergence of EGO is established by Vazquez and Bect [20], convergence rates have remained elusive. Grünewälder *et al.* [21] 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.

In summary, our main contributions are as follows.

1) We analyze GP-UCB, an intuitive algorithm for GP optimization, when the function is either sampled from a known GP, or has low RKHS norm.
2) 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.
3) 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.
4) We evaluate GP-UCB on sensor network data, demonstrating that it compares favorably to existing algorithms for GP optimization.

II. Problem Statement and Background

Consider the problem of sequentially optimizing an unknown reward function $f : D \to \mathbb{R}$. In each round t; we choose a point $\boldsymbol{x}_t \in D$ and get to see the function value there, perturbed by noise: $y_t = f(\boldsymbol{x}_t) + \epsilon_t$. Our goal is to maximize the sum of rewards $\sum_{t=1}^{T} f(\boldsymbol{x}_t)$, thus to perform essentially as well as $\boldsymbol{x}^* = \operatorname{argmax}_{\boldsymbol{x}\in D} f(\boldsymbol{x})$ (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, $f(\boldsymbol{x})$ is the temperature at $\boldsymbol{x}$, 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 [1]$\boldsymbol{x}^* = \operatorname{argmax}_{\boldsymbol{x}\in D} f(\boldsymbol{x})$. For our choice $\boldsymbol{x}_t$ in round t, we incur instantaneous regret $r_t = f(\boldsymbol{x}^*) - f(\boldsymbol{x}_t)$. 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*: $\lim_{T\to\infty} R_T/T = 0$. 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, since the maximum $\max_{t\le T} f(\boldsymbol{x}_t)$ in the first T rounds is no further from $f(\boldsymbol{x}^*)$ than the average.

A. Gaussian Processes and RKHS's

GPs: 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 Fig. 2(a)]. We can enforce implicit properties like smoothness without relying on any parametric assumptions, modeling f as a sample from a GP: a collection of dependent random variables, one for each $\boldsymbol{x} \in D$, every finite subset of which is multivariate Gaussian distributed in an overall consistent way [6]. A $\mathrm{GP}(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$ is specified by its mean function $\mu(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})]$ and covariance (or kernel) function $k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(f(x) - \mu(\boldsymbol{x}))(f(x') - \mu(\boldsymbol{x}'))]$. For GPs not conditioned on data, we assume[2] that $\mu \equiv 0$. Moreover, we restrict $k(\boldsymbol{x}, \boldsymbol{x}) \le 1, \boldsymbol{x} \in D$, 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 V-B.

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 $\mathrm{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$. Second, the Bayesian algorithm we analyze generally uses $\mathrm{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$ as prior distribution over f. A major advantage of working with GPs is the existence of simple analytic formulas for mean and covariance of the posterior distribution, which allows easy implementation of algorithms. For a noisy sample $\boldsymbol{y}_T = [y_1 \cdots y_T]^T$ at points $A_T = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$, $y_t = f(\boldsymbol{x}_t) + \epsilon_t$ with $\epsilon_t \sim$

[1]$\boldsymbol{x}^*$ need not be unique; only $f(\boldsymbol{x}^*)$ occurs in the regret.

[2]This is w.l.o.g. [6].

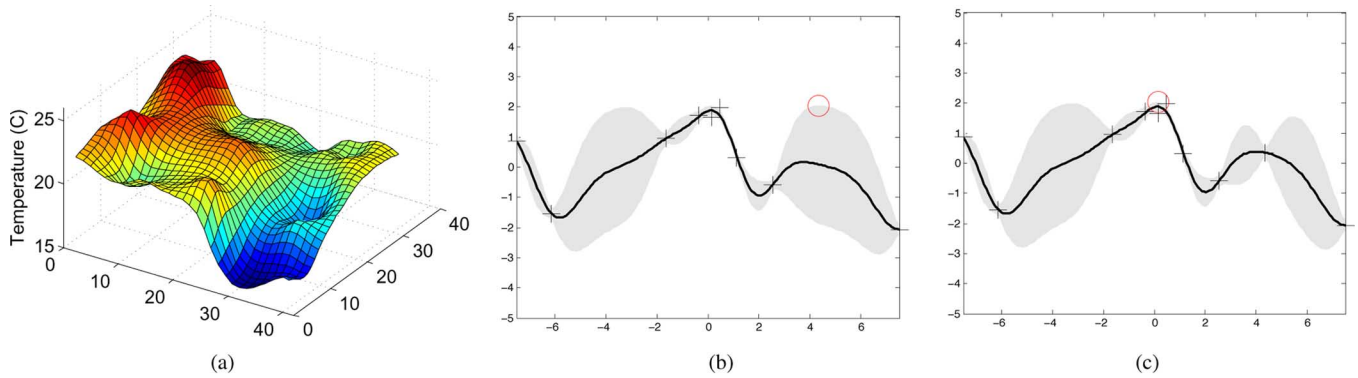

Fig. 2. (a) Example of temperature data collected by a network of 46 sensors at Intel Research Berkeley. (b) and (c) Two iterations of the GP-UCB algorithm. The dark curve indicates the current posterior mean, while the gray bands represent the upper and lower confidence bounds which contain the function with high probability. The "+" mark indicates points that have been sampled before, while the "o" mark shows the point chosen by the GP-UCB algorithm to sample next. It samples points that are either (b) uncertain or have (c) high posterior mean.

$N(0,\sigma^2)$ i.i.d. Gaussian noise, the posterior over f is a GP distribution again, with mean $\mu_T(\boldsymbol{x})$, covariance $k_T(\boldsymbol{x},\boldsymbol{x}')$, and variance $\sigma_T^2(\boldsymbol{x})$:

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

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

$$\sigma_T^2(\boldsymbol{x}) = k_T(\boldsymbol{x},\boldsymbol{x}) \tag{2}$$

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

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, [22]) is intimately related to GPs and their covariance functions $k(\boldsymbol{x},\boldsymbol{x}')$. 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: $\langle f, k(\boldsymbol{x},\cdot)\rangle_k = f(\boldsymbol{x})$ for all $f\in\mathcal{H}_k(D)$. It is literally constructed by completing the set of mean functions μ_T for all possible T, $\{\boldsymbol{x}_t\}$, and $\boldsymbol{y}_T$. 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 V-B, members of $\mathcal{H}_k(D)$ can uniformly approximate any continuous function on any compact subset of D.

B. Information Gain and Experimental Design

One approach to maximizing f is to first choose points $\boldsymbol{x}_t$ so as to estimate the function globally well, and 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 [5]), where the informativeness of a set of sampling points $A\subset D$ about f is measured by the *information gain* (cf., [23]), which is the mutual information between f and observations $\boldsymbol{y}_A = \boldsymbol{f}_A+\epsilon_A$ at these points:

$$\mathrm{I}(\boldsymbol{y}_A; f) = \mathrm{H}(\boldsymbol{y}_A) - \mathrm{H}(\boldsymbol{y}_A|f) \tag{3}$$

quantifying the reduction in uncertainty about f from revealing $\boldsymbol{y}_A$. Here, $\boldsymbol{f}_A = [f(\boldsymbol{x})]_{\boldsymbol{x}\in A}$ and $\varepsilon_A\sim N(\boldsymbol{0},\sigma^2\boldsymbol{I})$. For a Gaussian, $\mathrm{H}(N(\mu,\boldsymbol{\Sigma})) = \frac{1}{2}\log|2\pi e\boldsymbol{\Sigma}|$, so that in our setting $\mathrm{I}(\boldsymbol{y}_A;f) = \mathrm{I}(\boldsymbol{y}_A;\boldsymbol{f}_A) = \frac{1}{2}\log|\boldsymbol{I}+\sigma^{-2}\boldsymbol{K}_A|$, where $\boldsymbol{K}_A = [k(\boldsymbol{x},\boldsymbol{x}')]_{\boldsymbol{x},\boldsymbol{x}'\in A}$. While finding the information gain maximizer among $A\subset D$, $|A|\le T$ is NP-hard [24], it can be approximated by an efficient greedy algorithm. If $F(A) = \mathrm{I}(\boldsymbol{y}_A;f)$, this algorithm picks $\boldsymbol{x}_t = \operatorname{argmax}_{\boldsymbol{x}\in D} F(A_{t-1}\cup\{\boldsymbol{x}\})$ in round t, which can be shown to be equivalent to

$$\boldsymbol{x}_t = \operatorname*{argmax}_{\boldsymbol{x}\in D}\sigma_{t-1}(\boldsymbol{x}) \tag{4}$$

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

$$F(A_T)\ge(1-1/e)\max_{|A|\le T}F(A) \tag{5}$$

at least a constant fraction of the optimal information gain value. This is because $F(A)$ satisfies a diminishing returns property called *submodularity* [25], and the greedy approximation guarantee (5) holds for any submodular function [11].

While sequentially optimizing (4) is a provably good way to *explore* f globally, it is not well suited for function optimization. For the latter, we only need to identify point $\boldsymbol{x}$ where $f(\boldsymbol{x})$ is large, in order to concentrate sampling there as rapidly as possible, thus *exploiting* 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.

III. 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 $\boldsymbol{x}_t = \operatorname{argmax}_{\boldsymbol{x}\in D}\mu_{t-1}(\boldsymbol{x})$, maximizing the expected reward based on the posterior so far. However, this rule is too greedy

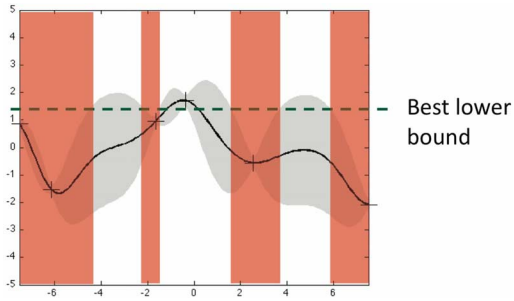

Fig. 3. GP-UCB selection rule implicitly rules out regions of the decision set where the upper confidence bound is less than the maximum lower confidence bound, thus eliminating regions where the function value is suboptimal with high probability.

too soon and tends to get stuck in shallow local optima. A combined strategy is to choose

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

where β_t are appropriate constants. This latter objective prefers both points $\boldsymbol{x}$ 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, which will give insight into the choice of β_t, is that it greedily selects points $\boldsymbol{x}$ such that $f(\boldsymbol{x})$ should be a reasonable upper bound on $f(\boldsymbol{x}^*)$, since the argument in (6) is an upper quantile of the marginal posterior $P(f(\boldsymbol{x})|\boldsymbol{y}_{t-1})$. We call this choice the GP-UCB index, where β_t is specified depending on the context (see Section IV). Pseudocode for the GP-UCB algorithm is provided in Algorithm 1. Fig. 2 illustrates two subsequent iterations, where GP-UCB both explores [see Fig. 2(b)] by sampling an input $\boldsymbol{x}$ with large $\sigma^2_{t-1}(\boldsymbol{x})$ and exploits [see Fig. 2(c)] by sampling $\boldsymbol{x}$ with large $\mu_{t-1}(\boldsymbol{x})$.

Another intuition about the GP-UCB selection rule is presented in Fig. 3. Since the upper and lower confidence bounds correspond to percentile points for f, at points where even the UCB is smaller than the highest lower confidence bound, the function values are suboptimal with high probability. The GP-UCB selection rule picks the point of highest UCB and therefore avoids these regions of the decision set.

The GP-UCB selection rule (6) is motivated by the UCB algorithm for the classical multiarmed bandit problem [8], [26]. Among competing criteria for GP optimization (see Section I), variants of the GP-UCB rule (with only heuristically defined β_t) have been introduced and demonstrated to be effective for this application [7], [27]. 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 VI, we show that in practice GP-UCB compares favorably with these alternatives.

If D is infinite, finding $\boldsymbol{x}_t$ in (6) may be hard: the upper confidence index is multimodal in general. However, global search heuristics are very effective in practice [17]. It is generally assumed that evaluating f is more costly than maximizing the GP-UCB index.

Algorithm 1 The GP-UCB algorithm.

Input: Input space D; GP Prior $\mu_0 = 0, \sigma_0, k$

for $t = 1, 2, \ldots$ **do**

Choose $\boldsymbol{x}_t = \operatorname*{argmax}_{\boldsymbol{x}\in D} \mu_{t-1}(\boldsymbol{x}) + \sqrt{\beta_t}\sigma_{t-1}(\boldsymbol{x})$

Sample $y_t = f(\boldsymbol{x}_t) + \epsilon_t$

Perform Bayesian update to obtain μ_t and σ_t

end for

UCB algorithms (and GP optimization techniques in general) have been applied to a large number of problems in practice [26], [2], [3]. 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.

IV. Regret Bounds

We now establish cumulative regret bounds for GP optimization, treating a number of different settings: $f \sim \mathrm{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$ for finite D, $f \sim \mathrm{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$ 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. For the linear case, Dani *et al.* [10] provide regret bounds that explicitly depend on the dimensionality[3] 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 γ_T after T rounds, defined as

$$\gamma_T := \max_{A\subset D:|A|=T} \mathrm{I}(\boldsymbol{y}_A; \boldsymbol{f}_A) \tag{7}$$

where $\mathrm{I}(\boldsymbol{y}_A; \boldsymbol{f}_A) = \mathrm{I}(\boldsymbol{y}_A; f)$ is defined in (3). Recall that $\mathrm{I}(\boldsymbol{y}_A; \boldsymbol{f}_A) = \frac{1}{2}\log|\boldsymbol{I} + \sigma^{-2}\boldsymbol{K}_A|$, where $\boldsymbol{K}_A = [k(\boldsymbol{x}, \boldsymbol{x}')]_{\boldsymbol{x},\boldsymbol{x}'\in A}$ is the covariance matrix of $\boldsymbol{f}_A = [f(\boldsymbol{x})]_{\boldsymbol{x}\in A}$ associated with the samples A. Our regret bounds are of the form $\mathcal{O}^*(\sqrt{T\beta_T\gamma_T})$, where β_T is the confidence parameter in Algorithm 1, while the bounds of Dani *et al.* [10] are of the form $\mathcal{O}^*(\sqrt{T\beta_T d})$ (d the dimensionality of the linear function space). Here and later, 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.* [10], we face a number of additional 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 V, we provide general methods for bounding γ_T, either by efficient auxiliary computations or by direct expressions for specific kernels of

[3]In general, d is the dimensionality of the input space D, which in the finite-dimensional linear case coincides with the feature space. By this, we mean that $D \subset [0, r]^d \subset \mathbb{R}^d$ for some $r > 0$.

interest. Our results match known lower bounds (up to log factors) in both the K-armed bandit and the d-dimensional linear optimization case.

A. 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 β_t for a sample f of a GP with mean function zero and covariance function $k(\boldsymbol{x},\boldsymbol{x}')$, 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 is provided in the Appendix.

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 $\mathrm{I}(\boldsymbol{y}_A;\boldsymbol{f}_A)$ allows us to derive sharp *a priori* bounds, depending on choice and parameterization of k (see Section V). 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 $k(\boldsymbol{x},\boldsymbol{x}')$ satisfies the following high probability bound on the derivatives of GP sample paths f: for some constants a, $b > 0$:

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

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

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

Running GP-UCB with β_t for a sample f of a GP with mean function zero and covariance function $k(\boldsymbol{x},\boldsymbol{x}')$, 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 $k(\boldsymbol{x},\boldsymbol{x}')$ disqualifies GPs with highly erratic sample paths. It holds for stationary kernels $k(\boldsymbol{x},\boldsymbol{x}') = k(\boldsymbol{x}-\boldsymbol{x}')$ which are four times differentiable ([28, Theorem 5])., such as the squared exponential and Matérn kernels with $\nu > 2$ (see Section V-B), 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 1 and come with independent increments. We conjecture that a result of the form of Theorem 2 does not hold in this case.

B. 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 σ^2. We now analyze GP-UCB in an agnostic setting, where f is an arbitrary function from the RKHS corresponding to kernel $k(\boldsymbol{x},\boldsymbol{x}')$. Moreover, we allow the noise variables ε_t to be an arbitrary martingale difference sequence (meaning that $\mathbb{E}[\varepsilon_t|\varepsilon_{<t}] = 0$ for all $t \in \mathbb{N}$), uniformly bounded by σ. Note that we still run the same GP-UCB algorithm, whose prior and noise model are misspecified in this case. The 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 $k(\boldsymbol{x},\boldsymbol{x}')$, and that the noise ε_t has zero mean conditioned on the history and is bounded by σ 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 β_t, prior $\mathrm{GP}(0,k(\boldsymbol{x},\boldsymbol{x}'))$, and noise model $N(0,\sigma^2)$, we obtain a regret bound of $\mathcal{O}^*(\sqrt{TB}(\sqrt{\gamma_T}+\gamma_T))$ 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 Theorems 2 and 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 $f \sim \mathrm{GP}(0,k(\boldsymbol{x},\boldsymbol{x}'))$, then $\|f\|_k = \infty$ almost surely [22]: sample paths are rougher than RKHS functions. Neither Theorem 2 nor 3 encompasses the other.

V. Bounding the Information Gain

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

A. Submodularity and Greedy Maximization

In order to bound γ_T, we have to maximize the information gain $F(A) = \mathrm{I}(\boldsymbol{y}_A;f)$ over all subsets $A \subset D$ of size T: a combinatorial problem in general. However, as noted in Section II, $F(A)$ 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 γ_T by $(1-1/e)^{-1}\mathrm{I}(\boldsymbol{y}_{A_T};f)$, 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 γ_T is "near-greedy." As noted in Section II, 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 γ_T, which can be plugged into our results in Section IV to obtain high-probability bounds on R_T. This being a laborious procedure, one would prefer *a priori* bounds for γ_T in practice which are simple analytical expressions of T and parameters of

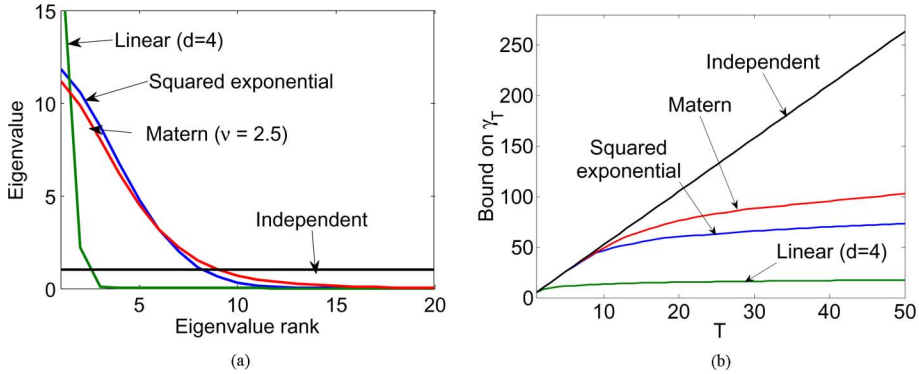

Fig. 4. (Left) Spectral decay and (right) information gain bound for independent (diagonal), linear, squared exponential, and Matérn kernels ($\nu = 2.5$) with equal trace.

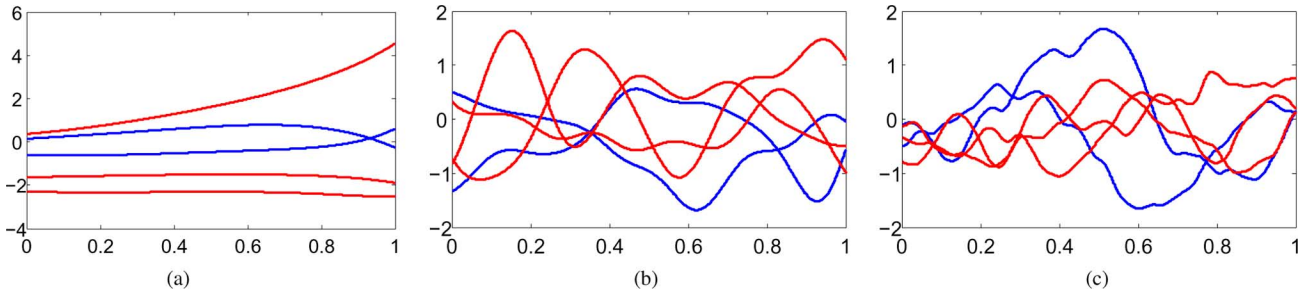

Fig. 5. Sample functions drawn from a GP with (a) Bayesian linear regression, (b) squared exponential, and (c) Matérn kernels ($\nu = 2.5$).

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 $\boldsymbol{f} = [f(\boldsymbol{x})]_{\boldsymbol{x}\in D}$, $\boldsymbol{K}_D = [k(\boldsymbol{x},\boldsymbol{x}')]_{\boldsymbol{x},\boldsymbol{x}'\in D}$. Sampling f at $\boldsymbol{x}_t$, we obtain $y_t \sim N(\boldsymbol{v}_t^T\boldsymbol{f},\sigma^2)$, where $\boldsymbol{v}_t \in \mathbb{R}^{|D|}$ is the indicator vector associated with $\boldsymbol{x}_t$. We can upper-bound the greedy maximum once more, by relaxing this constraint to $\|\boldsymbol{v}_t\| = 1$ in round t of the sequential method. For this relaxed greedy procedure, all $\boldsymbol{v}_t$ are leading eigenvectors of $\boldsymbol{K}_D$, since successive covariance matrices of $P(\boldsymbol{f}|\boldsymbol{y}_{t-1})$ share their eigenbasis with $\boldsymbol{K}_D$, 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 $\boldsymbol{K}_D$:

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

subject to $\sum_t m_t = T$ and $\operatorname{spec}(\boldsymbol{K}_D) = \{\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots\}$. 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 $T_* = 1,\ldots,T$:

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

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 $\boldsymbol{K}_D$ decays, the slower the growth of γ_T. Fig. 4 illustrates this intuition.

B. Bounds for Common Kernels

In this section, we bound γ_T for a range of commonly used covariance functions: finite-dimensional linear, squared exponential, and Matérn kernels. Together with our results in Section IV, these imply sublinear regret bounds for GP-UCB in all cases.

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

The squared exponential kernel is $k(\boldsymbol{x},\boldsymbol{x}') = \exp(-(2l^2)^{-1}\|\boldsymbol{x}-\boldsymbol{x}'\|^2)$, l being a length scale parameter. Sample functions are differentiable to any order almost surely [6].

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

Fig. 5 shows random functions drawn from GP distributions with the aforementioned kernels.

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

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

A proof of Theorem 5 is given in the Appendix; we only sketch the idea here. γ_T is bounded by Theorem 4 in terms of the eigendecay of the kernel matrix $\boldsymbol{K}_D$. If D is infinite or very

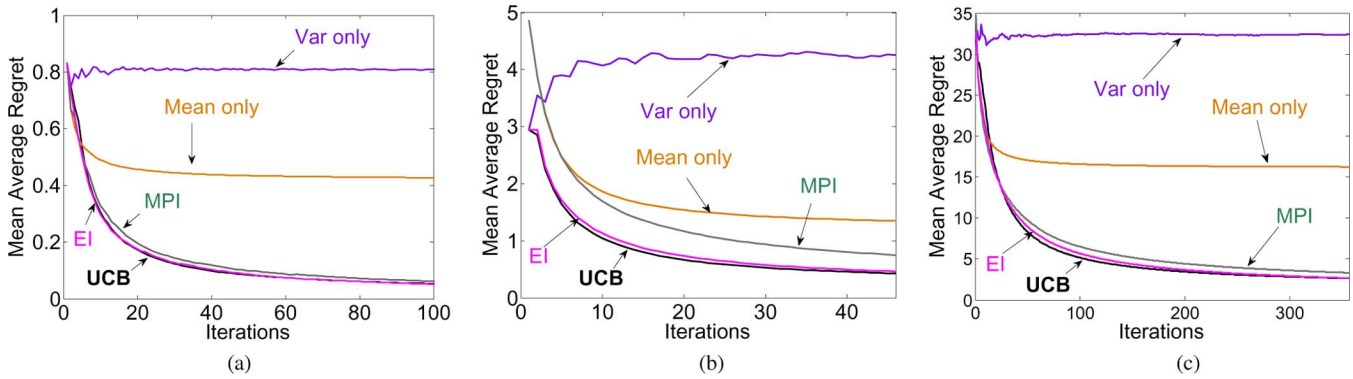

Fig. 6. Mean average regret: GP-UCB and various heuristics on (a) synthetic and (b, c) sensor network data.

large, we can use the operator spectrum of $k(x, x')$, which likewise decays rapidly. For the kernels of interest here, asymptotic expressions for the operator eigenvalues are given by Seeger *et al.* [29], 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. Our existence result relies on the probabilistic method.

Together with Theorems 2 and 3, this result guarantees sublinear regret of GP-UCB for any dimension (see Fig. 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.

VI. Experiments

We compare GP-UCB with commonly used heuristics such as the EI and MPI, and with naive methods which choose points of maximum mean or variance only, both on synthetic and real sensor network data, and also on the Branin and Goldstein–Price benchmark functions [30] for global optimization. The EI heuristic [15] chooses the point where the EI over the currently observed maximum value is highest, while the MPI heuristic [16] chooses the point where improvement over current maximum is most probable.

Experimental Setup: To generate synthetic test functions, we sample random functions from a GP with squared exponential and Matérn $(\nu = 1.5)$ kernels, using a length scale parameter 0.2. The sampling noise variance σ^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).

Next, we use temperature data collected from 46 sensors deployed at Intel Research Berkeley over 5 days at 1-min intervals, pertaining to the example in Section II. We take the first two-thirds of the dataset 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 dataset, 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.

We also use 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 A.M. to 11 A.M. for one month, from 357 sensors. We again use the covariance matrix from two-thirds of the dataset 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.

While the choice of β_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 β_t down by a factor 5. Note that we did not optimize constants in our regret bounds.

Exploration–Exploitation Performance: Fig. 6 compares the mean average regret $(\frac{1}{T}\sum_{t=1}^{T} r_t)$ incurred by the different heuristics and the GP-UCB algorithm on synthetic and real data, as a function of the number of iterations (samples) T. 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.

Performance in Search Problems: Fig. 7 compares the mean minimum regret $(\min_{0\le t\le T} r_t)$ incurred by the different heuristics and the GP-UCB algorithm on synthetic and real data. This measure is more relevant to pure search problems (i.e., no exploitation) and captures how quickly the algorithms find the optimal point. For temperature data, the GP-UCB algorithm and EI heuristic clearly outperform the MPI, and do not exhibit significant difference between each other. On synthetic and traffic data, the MPI does slightly better but still not as well as the GP-UCB algorithm and the EI heuristic. In contrast to the average regret (see Fig. 6), for the minimum regret the variance-only (information gain) criterion performs well, but (except for temperature data) still exhibits slower convergence than GP-UCB and EI. Mean-only performs very poorly due to convergence to local optima. In summary, GP-UCB performs at least on par with the existing approaches which are not equipped with regret bounds, even on the search (pure exploration) problem.

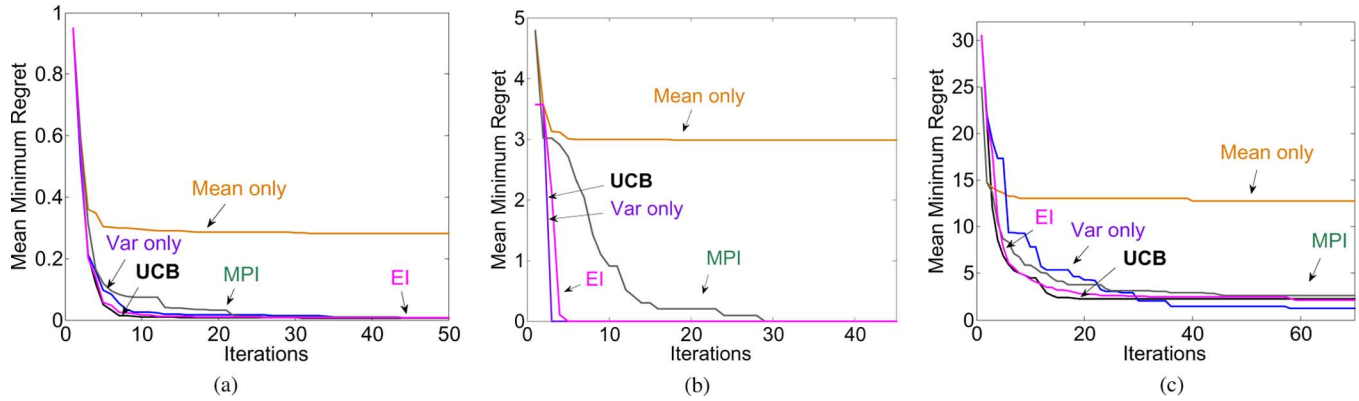

Fig. 7. Mean minimum regret: GP-UCB and various heuristics on (a) synthetic, and (b, c) sensor network data.

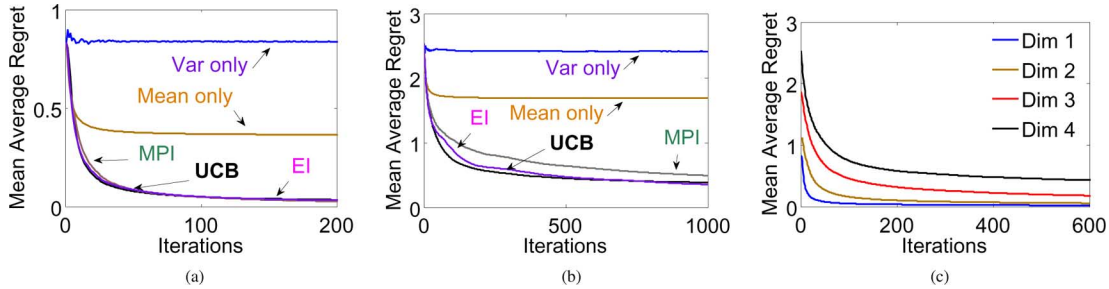

Fig. 8. Mean average regret: GP-UCB and various heuristics on synthetic data from the squared exponential kernel of (a) dimension 1 and (b) dimension 4; (c) comparison of GP-UCB performance with increasing dimensionality of the problem.

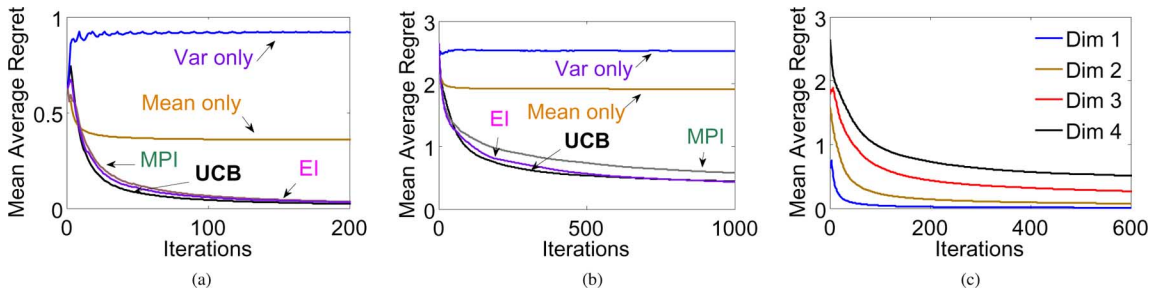

Fig. 9. Mean average regret: GP-UCB and various heuristics on synthetic data from the Matérn kernel ($\nu = 1.5$) of (a) dimension 1 and (b) dimension 4; (c) comparison of GP-UCB performance with increasing dimensionality of the problem.

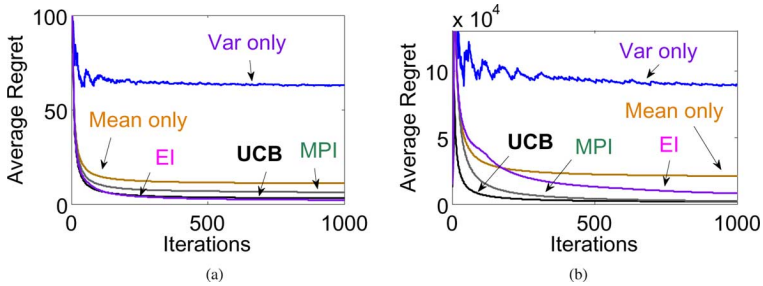

Fig. 10. Comparison of GP-UCB and various heuristics on (a) Branin and (b) Goldstein–Price benchmark functions.

Dependence on Dimensionality: We also compare the GP-UCB algorithm with the heuristics on synthetic test functions for decision sets of dimensionality varying between 1 and 4. In each case, the decision set $D = [0, 1]^d$ was discretized into nine points along each dimension, i.e., discretized uniformly into 9^d points. Fig. 8 compares the mean average regret incurred by the different heuristics and the GP-UCB algorithm on synthetic data from the squared exponential kernel of dimensions 1 [see Fig. 8(a)] and 4 [see Fig. 8(b)]. Fig. 8(c) illustrates how the performance of the GP-UCB algorithm scales with increasing dimensionality of the problem. These figures illustrate that the GP-UCB algorithm performs competitively with popular heuristics on problems of varying dimensionality, and that the relative performance scales well with dimension. Fig. 9 shows qualitatively similar results for the Matérn kernel with $\nu = 1.5$.

Synthetic Benchmarks: Finally, we compare (see Fig. 10) the performance of the GP-UCB algorithm with the heuristics on the Branin [see Fig. 10(a)] and Goldstein–Price [see Fig. 10(b)] benchmark functions [30] for global optimization. The respective domains of the functions were scaled and translated onto the

2-D unit square, which was uniformly discretized into 10 000 points. A Matérn kernel ($\nu = 2.5$ and length scale parameter 0.2) was used as the prior, and all other details were identical to the high-dimensional experiment case. The GP-UCB algorithm and EI heuristic seem to do better than the others on the Branin benchmark function, while the GP-UCB algorithm seems to outperform both the EI and MPI heuristics on the Goldstein–Price benchmark function.

Based on these experiments, we conclude that GP-UCB is at least competitive with other selection heuristics (which are not known to admit regret bounds) on several synthetic benchmarks and real-world sensor selection problems.

VII. Conclusion

We prove the first sublinear regret bounds for GP optimization with commonly used kernels (see Fig. 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.

We remark that while our bounds hold under weak regularity assumptions on the kernel, we do not have lower bounds showing that our conditions are necessary in order to achieve (sublinear) regret. However, we should note that our regret bounds match known lower bounds for both the k-arm bandit setting and the finite-dimensional linear kernel. An important open question is characterizing what the necessary conditions are for (sublinear) regret in the nonparametric case (along with characterizing the achievable rate of regret). Moreover, it is unclear whether the EI heuristic can be shown to achieve sublinear regret.

Overall, we believe that our results provide an interesting step toward understanding exploration–exploitation tradeoffs with complex utility functions.

Appendix I
Regret Bounds for f Sampled From GP

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

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,

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

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

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

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

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

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

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

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

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

∎

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

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

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

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

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

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

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

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

with $C_2 = \sigma^{-2}/\log(1 + \sigma^{-2}) \geq 1$, since $s^2 \leq C_2\log(1 + s^2)$ for $s^2 \in [0, \sigma^{-2}]$, and $\sigma^{-2}\sigma_{t-1}^2(\boldsymbol{x}_t) \leq \sigma^{-2}k(\boldsymbol{x}_t, \boldsymbol{x}_t) \leq \sigma^{-2}$. 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.

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 $k(\boldsymbol{x}, \boldsymbol{x}')$. For example, if $k(\boldsymbol{x}, \boldsymbol{x}') = e^{-\|\boldsymbol{x}-\boldsymbol{x}'\|}$ (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 IV.

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,

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

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

Proof: Fix $t \geq 1$ and $\boldsymbol{x} \in D$. Conditioned on $\boldsymbol{y}_{t-1} = (y_1, \ldots, y_{t-1})$, $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}\}$ are deterministic, and $f(\boldsymbol{x}) \sim N(\mu_{t-1}(\boldsymbol{x}), \sigma_{t-1}^2(\boldsymbol{x}))$. As before, $\Pr\{|f(\boldsymbol{x}_t) - \mu_{t-1}(\boldsymbol{x}_t)| > \beta_t^{1/2}\sigma_{t-1}(\boldsymbol{x}_t)\} \leq e^{-\beta_t/2}$. 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 $\boldsymbol{x}^*$. 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,

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

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

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

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

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

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

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

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

where $[\boldsymbol{x}]_t$ denotes the closest point in D_t to $\boldsymbol{x}$. A sufficient discretization has each coordinate with τ_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 $[\boldsymbol{x}^*]_t$ denotes the closest point in D_t to $\boldsymbol{x}^*$. Then,

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

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

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

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

Hence,

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

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

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

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 $[\boldsymbol{x}^*]_t$ (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:

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

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

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

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

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 $1-\delta$,

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

so that by Cauchy–Schwarz:

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

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 $j = 1, \ldots, d$. [28, Th. 5] states that if derivatives up to fourth order exist for $(\boldsymbol{x}, \boldsymbol{x}') \mapsto k(\boldsymbol{x}, \boldsymbol{x}')$, then f is almost surely continuously differentiable, with $\partial f/(\partial x_j)$ distributed as GPs again. Moreover, there are constants $a, b_j > 0$ such that

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

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

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 $\partial f/(\partial x_j)$ existing are continuous, and the complement of (10) holds for all j. [28, Th. 5], 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 $\nu > 2$ [31].

Appendix II
Regret Bound for the 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 II, conditioned on data $(\boldsymbol{x}_t, y_t)$, $t = 1, \ldots, T$. It is easy to see that the RKHS norm corresponding to k_T is given by

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

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 $\langle f(\cdot), k_T(\cdot, \boldsymbol{x}) \rangle_{k_T} = f(\boldsymbol{x})$ for any $f \in \mathcal{H}_{k_T}(D)$ by the reproducing property, then

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

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 $\varepsilon_t = y_t - f(\boldsymbol{x}_t)$ is drawn independently from $N(0, \sigma^2)$, the true sequence of noise variables ε_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 analog to Lemma 5.1, by way of the following concentration result.

Theorem 6: Let $\delta \in (0, 1)$. Assume that the noise variables ε_t are uniformly bounded by σ. Define

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

Then

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

1) Concentration of Martingales: In our analysis, we use the following Bernstein-type concentration inequality for martingale differences, due to Freedman [32] (see also Theorem 3.15 of [33, Th. 3.15]).

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

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

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

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

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 $\varepsilon_t = y_t - f(\boldsymbol{x}_t)$. We will analyze the quantity $Z_T = \|\mu_T - f\|_{k_T}^2$, measuring the error of μ_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

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

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: $\widetilde{\varepsilon}_t = \varepsilon_t/\sigma$, $\widetilde{f} = f/\sigma$, $\widetilde{\mu}_t = \mu_t/\sigma$, $\widetilde{\sigma}_t = \sigma_t/\sigma$. Also, to ease notation, we will use μ_{t-1}, σ_{t-1} as shorthand for $\mu_{t-1}(\boldsymbol{x}_t)$, $\sigma_{t-1}(\boldsymbol{x}_t)$.

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

$$\begin{aligned} Z_T \leq \|f\|_k^2 &+ 2 \sum_{t=1}^{T} \widetilde{\varepsilon}_t \frac{\widetilde{\mu}_{t-1} - \widetilde{f}(\boldsymbol{x}_t)}{1 + \widetilde{\sigma}_{t-1}^2} \\ &+ \sum_{t=1}^{T} \widetilde{\varepsilon}_t^2 \frac{\widetilde{\sigma}_{t-1}^2}{1 + \widetilde{\sigma}_{t-1}^2}. \end{aligned}$$

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

$$\begin{aligned} Z_T &= \|f\|_k^2 - 2\boldsymbol{f}_T^T\alpha_T + \boldsymbol{y}_T^T\alpha_T - \sigma^2\|\alpha_T\|^2 \\ &\quad + \sigma^{-2}\sum\nolimits_{t=1}^{T}(\varepsilon_t - \sigma^2\alpha_t)^2 = \|f\|_k^2 \\ &\quad - \boldsymbol{y}_T^T(\boldsymbol{K}_T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y}_T + \sigma^{-2}\|\varepsilon_T\|^2. \end{aligned}$$

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

$$\begin{aligned} &-\boldsymbol{y}_T^T(\boldsymbol{K}_T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y}_T = -\sum\nolimits_t \frac{(y_t - \mu_{t-1})^2}{\sigma^2 + \sigma_{t-1}^2} \\ &= 2\sum\nolimits_t \varepsilon_t \frac{\mu_{t-1} - f(\boldsymbol{x}_t)}{\sigma^2 + \sigma_{t-1}^2} - \sum\nolimits_t \frac{\varepsilon_t^2\widetilde{\sigma}_{t-1}^2}{\sigma^2 + \sigma_{t-1}^2} - R \end{aligned}$$

with $R = \sum_t(\mu_{t-1} - f(\boldsymbol{x}_t))^2/(\sigma^2 + \sigma_{t-1}^2) \ge 0$. 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

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

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

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

Now, since $\widetilde{\varepsilon}_t$ is a martingale difference sequence with respect to the histories $\mathcal{H}_{<t}$ and $M_t/\widetilde{\varepsilon}_t$ 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 β_t as defined in Theorem 6, we have

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

The proof is given in Section II-E in the Appendix. 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 β_1, we know that $Z_0 \le \|f\|_k \le \beta_1$. Hence, $E_0 = 1$ always. Now suppose the high-probability event of Lemma 7.3 holds, in particular $\sum_{t=1}^T M_t \le \beta_{T+1}/2$. For the inductive hypothesis, assume $E_{T-1} = 1$. Using this and Lemma 7.2:

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

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

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:

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

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

$$\begin{aligned} V_T &:= \sum\nolimits_{t=1}^{T}\mathbf{Var}\,(M_t \mid M_1 \ldots M_{t-1}) \\ &\le \sum\nolimits_{t=1}^{T} 4|\widetilde{\varepsilon}_t|^2 E_{t-1}\beta_t\min\{\widetilde{\sigma}_{t-1}^2, 1/4\} \\ &\le 4\beta_T\sum\nolimits_{t=1}^{T} E_{t-1}\min\{\widetilde{\sigma}_{t-1}^2, 1/4\} \qquad |\widetilde{\varepsilon}_t| \le 1 \\ &\le 9\beta_T\gamma_T. \end{aligned}$$

In the last line, we used Lemma 7.1 with $\alpha = 1/4$, noting that $8\alpha/\log(1 + \alpha) \le 9$. 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

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

Note that our choice of β_{T+1} satisfies

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

Therefore, the previous probability is bounded by δ/T^2, whereas the last inequality follows from the definition of β_{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. ∎

Appendix III
Bounds on Information Gain

In this section, we show how to bound γ_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[4] that $k(\boldsymbol{x},\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in D$.

The plan of attack is as follows. First, we note that the argument of γ_T, $\mathrm{I}(\boldsymbol{y}_A; \boldsymbol{f}_A)$ is a submodular function, so γ_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 neighbor distance $o(1)$, consider the kernel matrix $\boldsymbol{K}_{D_T} \in \mathbb{R}^{n_T \times n_T}$, and bound γ_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 [29], 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.

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^τ, such that

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

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

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

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

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

[4]Without loss in generality, we use this assumption later to ensure that $n_T^{-1}\mathrm{tr}\boldsymbol{K}_{D_T} = \int k(\boldsymbol{x},\boldsymbol{x})\, d\boldsymbol{x}$. If $k(\boldsymbol{x},\boldsymbol{x})$ is not constant, this is approximately true by the law of large numbers, and our result given later remains valid.

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

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

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

Proof: Fix $T \in \mathbb{N}$, and let $A = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ be a maximizer for γ_T. Consider neighbors $[\boldsymbol{x}_t]_T \in D_T$ according to (13), $[A]_T = \{[\boldsymbol{x}_t]_T\}$. Then,

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

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

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

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

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

Proof: As shown by Krause and Guestrin [25], the function $F(A) = \mathrm{I}(\boldsymbol{y}_A; \boldsymbol{f})$ is submodular. In the particular case considered here, this can be seen as follows: $F(A) = \mathrm{H}(\boldsymbol{y}_A) - \mathrm{H}(\boldsymbol{y}_A \mid \boldsymbol{f})$, where the entropy $\mathrm{H}(\boldsymbol{y}_A)$ is a (not-necessarily monotonic) submodular function in A, and since the noise is conditionally independent given $\boldsymbol{f}$, $\mathrm{H}(\boldsymbol{y}_A \mid \boldsymbol{f})$ 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 $A \subseteq B$) [23]. Thus, we can apply the result of Nemhauser *et al.* [11][5] 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 $\boldsymbol{v}_t = \delta_{\boldsymbol{x}_t} = [\mathrm{I}_{\{\boldsymbol{x}=\boldsymbol{x}_t\}}]$ in each round, $\boldsymbol{x}_t \in D_T$. We upper-bound the greedy maximum once more by relaxing these constraints to $\|\boldsymbol{v}_t\| = 1$ only. In the remainder of the proof, we concentrate on this relaxed greedy procedure. Suppose that up to round t, it chose $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{t-1}$.

[5]While the result of Nemhauser *et al.* [11] is stated in terms of finite sets, it extends to infinite sets as long as the greedy selection can be implemented efficiently.

The posterior $P(\boldsymbol{f}|\boldsymbol{y}_{t-1})$ has inverse covariance matrix $\boldsymbol{\Sigma}_{t-1}^{-1} = \boldsymbol{K}_{D_T}^{-1} + \sigma^{-2}\boldsymbol{V}_{t-1}\boldsymbol{V}_{t-1}^T$, $\boldsymbol{V}_{t-1} = [\boldsymbol{v}_1 \cdots \boldsymbol{v}_{t-1}]$, and the greedy procedure selects $\boldsymbol{v}$ so to maximize the variance $\boldsymbol{v}^T\boldsymbol{\Sigma}_{t-1}\boldsymbol{v}$: the eigenvector corresponding to $\boldsymbol{\Sigma}_{t-1}$'s largest eigenvalue (by the Rayleigh–Ritz theorem). Since $\boldsymbol{\Sigma}_0 = \boldsymbol{K}_{D_T}$, then $\boldsymbol{v}_1 = \boldsymbol{u}_1$. Moreover, if all $\boldsymbol{v}_{t'}$, $t' < t$, have been chosen among $\boldsymbol{U}$'s columns, then by the inverse covariance expression just given, $\boldsymbol{K}_{D_T}$ and $\boldsymbol{\Sigma}_{t-1}$ have the same eigenvectors, so that $\boldsymbol{v}_t$ is a column of $\boldsymbol{U}$ as well. For example, if $\boldsymbol{v}_t = \boldsymbol{u}_j$, then comparing $\boldsymbol{\Sigma}_{t-1}$ and $\boldsymbol{\Sigma}_t$, all eigenvalues other than the jth remain the same, while the latter is shrunk. Therefore, after T rounds of the relaxed greedy procedure, $\boldsymbol{v}_t \in \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{\min\{T,n_T\}}\}$, $t = 1, \ldots, T$: at most the leading T eigenvectors of $\boldsymbol{K}_{D_T}$ can have been selected (possibly multiple times). If m_t denotes the number that the tth column of $\boldsymbol{U}$ has been selected, we obtain the theorem statement by a final bounding step. ∎

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 $\mu(\boldsymbol{x}) = \mathcal{V}(D)^{-1}\mathrm{I}_{\{\boldsymbol{x}\in D\}}$ be the uniform distribution on D, $\mathcal{V}(D) = \int_{\boldsymbol{x}\in D} d\boldsymbol{x}$, and assume that k is continuous. Note that $\int k(\boldsymbol{x},\boldsymbol{x})\mu(\boldsymbol{x})\,d\boldsymbol{x} = 1$ by our assumption $k(\boldsymbol{x},\boldsymbol{x}) = 1$, so that k is Hilbert–Schmidt on $L_2(\mu)$. Then, Mercer's theorem [22] states that the corresponding kernel operator has a discrete eigenspectrum $\{(\lambda_s, \phi_s(\cdot))\}$, and

$$k(\boldsymbol{x},\boldsymbol{x}') = \sum\nolimits_{s\geq 1} \lambda_s\phi_s(\boldsymbol{x})\phi_s(\boldsymbol{x}')$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and $\mathbb{E}_\mu[\phi_s(\boldsymbol{x})\phi_t(\boldsymbol{x})] = \delta_{s,t}$. Moreover, $\sum_{s\geq 1}\lambda_s^2 < \infty$, and the expansion of k converges absolutely and uniformly on $D \times D$. Note that $\sum_{s\geq 1}\lambda_s = \sum_{s\geq 1}\lambda_s\mathbb{E}_\mu[\phi_s(\boldsymbol{x})^2] = \int K(\boldsymbol{x},\boldsymbol{x})\mu(\boldsymbol{x})\,d\boldsymbol{x} = 1$. 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 fulfills the following requirements.

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

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

Proof: First, if we draw n_T samples $\tilde{\boldsymbol{x}}_j \sim \mu(\boldsymbol{x})$ independently at random, then $D_T = \{\tilde{\boldsymbol{x}}_j\}$ is ε-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 ε. 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 δ 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.* [34] 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 $S \leq n_T^{-1}\mathrm{tr}\boldsymbol{K}_{D_T} = 1$ 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 hold. ∎

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 $T_* = 1, \ldots, \min\{T, n_T\}$:

$$\tilde{\gamma}_T \leq \frac{1/2}{1-e^{-1}}\max_{r=1,\ldots,T}\Big(T_*\log(rn_T/\sigma^2) + (T-r)\sigma^{-2}\sum\nolimits_{t=T_*+1}^{n_T}\hat{\lambda}_t\Big).$$

Proof: We split the right-hand side in Lemma 7.6 at $t = T_*$. Let $r = \sum_{t\leq T_*} m_t$. For $t \leq T_*$, $\log(1+m_t\hat{\lambda}_t/\sigma^2) \leq \log(rn_T/\sigma^2)$, 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 γ_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 $k(\boldsymbol{x},\boldsymbol{x}')$ 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 $n_T = C_4T^\tau(\log T)$ with $C_4 = 2\mathcal{V}(D)(2\tau+1)$. Then, the following bound holds true:

$$\gamma_T \leq \frac{1/2}{1-e^{-1}}\max_{r=1,\ldots,T}\Big(T_*\log(rn_T/\sigma^2) + C_4\sigma^{-2}(1-r/T)(\log T)\left(T^{\tau+1}B_k(T_*)+1\right)\Big) + \mathcal{O}(T^{1-\tau/d})$$

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

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 ε-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. ∎

3) Proof of Theorem 5: Here, we instantiate Theorem 8 in order to obtain bounds on γ_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.* [29]. While $\mu(\boldsymbol{x})$ was Gaussian there, the same decay rate holds for λ_s w.r.t. uniform $\mu(\boldsymbol{x})$, 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,

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

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

$$\max_{r=1,\ldots,T}\Big(T_*\log(rn_T/\sigma^2) + \sigma^{-2}(1-r/T)(C_5\beta^{d-1} + C_4(\log T))\Big)$$

with $n_T = C_4T^d(\log T)$. 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 $\mathbb{E}[\mathrm{I}(\boldsymbol{y}_T;\boldsymbol{f}_T)] = \mathcal{O}((\log T)^{d+1})$ given by Seeger *et al.* [29], where $\boldsymbol{x}_t$ 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 $\max_{\{\boldsymbol{x}_t\}}\mathrm{I}(\boldsymbol{y}_T;\boldsymbol{f}_T)$.

Matérn Kernels: For Matérn kernels k with roughness parameter ν, $B_k(T_*)$ is given by Seeger *et al.* [29] for the uniform base distribution $\mu(\boldsymbol{x})$ 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 $T_* = (Tn_T)^{d/(2\nu+d)}(\log(Tn_T))^{\kappa}$ (κ chosen next), so that the bound becomes

$$\max_{r=1,\ldots,T}\Big(T_*\log(rn_T/\sigma^2) + \sigma^{-2}(1-r/T) \times (C_5T_*(\log(Tn_T))^{-\kappa(2\nu+d)/d} + C_4(\log T))\Big) + \mathcal{O}(T^{1-\tau/d})$$

with $n_T = C_4T^{\tau}(\log T)$. 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 $\mathbb{E}[\mathrm{I}(\boldsymbol{y}_T;\boldsymbol{f}_T)]$ given by Seeger *et al.* [29]? Here, $\gamma_T = \mathcal{O}(T^{d(d+1)/(2\nu+d(d+1))}(\log T))$, while $\mathbb{E}[\mathrm{I}(\boldsymbol{y}_T;\boldsymbol{f}_T)] = \mathcal{O}(T^{d/(2\nu+d)}(\log T)^{2\nu/(2\nu+d)})$.

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

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

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

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

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

ACKNOWLEDGMENT

The authors thank M. Hutter for insightful comments on an earlier version of this paper. The authors also thank the anonymous reviewers for detailed feedback.

REFERENCES

[1] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," presented at the Int. Conf. Machine Learning, Haifa, Israel, 2010.

[2] S. Pandey and C. Olston, "Handling advertisements of unknown quality in search advertising," in *Conf. Neural Inf. Process. Systems*, 2007, pp. 1065–1072.

[3] D. Lizotte, T. Wang, M. Bowling, and D. Schuurmans, "Automatic gait optimization with Gaussian process regression," in *Proc. Int. Joint Conf. Artificial Intell.*, 2007, pp. 944–949.

[4] H. Robbins, "Some aspects of the sequential design of experiments," *Bul. Amer. Math. Soc.*, vol. 58, pp. 527–535, 1952.

[5] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Stat. Sci.*, vol. 10, no. 3, pp. 273–304, 1995.

[6] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[7] D. D. Cox and S. John, "SDO: A statistical method for global optimization," in *Multidisciplinary Design Optimization: State of the Art*. Philadelphia, PA: SIAM, 1997.

[8] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2–3, pp. 235–256, 2002.

[9] P. Auer, "Using confidence bounds for exploitation-exploration tradeoffs," *J. Mach. Learn. Res.*, vol. 3, pp. 397–422, 2002.

[10] V. Dani, T. P. Hayes, and S. M. Kakade, "Stochastic linear optimization under bandit feedback," in *Conf. Learn. Theory*, 2008, pp. 355–367.

[11] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of the approximations for maximizing submodular set functions," *Math. Prog.*, vol. 14, pp. 265–294, 1978.

[12] R. Kleinberg, A. Slivkins, and E. Upfal, "Multi-armed bandits in metric spaces," in *Symp. Theory of Comput.*, 2008, pp. 681–690.

[13] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári, "Online optimization in X-armed bandits," in *Conf. Neural Inf. Process. Syst.*, 2008, pp. 201–208.

[14] A. Krause and C. S. Ong, "Contextual Gaussian process bandit optimization," in *Conf. Neural Inf. Process. Systems*, 2011.

[15] J. Mockus, V. Tiesis, and A. Zilinskas, *Toward Global Optimization*. Amsterdam, The Netherlands, North-Holland: , 1978, vol. 2, ch. Bayesian methods for seeking the extremum, pp. 117–128.

[16] J. Mockus, *Bayesian Approach to Global Optimization*. Norwell, MA: Kluwer, 1989.

[17] E. Brochu, M. Cora, and N. de Freitas, A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning, 2009.

[18] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Glob. Opt.*, vol. 13, pp. 455–492, 1998.

[19] D. Huang, T. T. Allen, W. I. Notz, and N. Zeng, "Global optimization of stochastic black-box systems via sequential Kriging meta-models," *J. Glob. Opt.*, vol. 34, pp. 441–466, 2006.

[20] E. Vazquez and J. Bect, "Convergence properties of the expected improvement algorithm with fixed and mean covariance functions," *J. Stat. Planning Inference*, vol. 140, no. 11, pp. 3088–3095, 2010.

[21] S. Grünewälder, J.-Y. Audibert, M. Opper, and J. Shawe-Taylor, "Regret bounds for Gaussian process bandit problems," presented at the Proc. Int. Conf. Artificial Intelligence and Stat., Sardinia, Italy, 2010.

[22] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA: SIAM, 1990.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley Interscience, 1991.

[24] C. Ko, J. Lee, and M. Queyranne, "An exact algorithm for maximum entropy sampling," *Oper. Res.*, vol. 43, no. 4, pp. 684–691, 1995.

[25] A. Krause and C. Guestrin, "Near-optimal nonmyopic value of information in graphical models," in *Proc. Uncertainty in Artif. Intell.*, 2005, pp. 324–331.
[26] L. Kocsis and C. Szepesvári, "Bandit based Monte-Carlo planning," in *Proc. ECML Conf.*, 2006, pp. 282–293.
[27] L. Dorard, D. Glowacka, and J. Shawe-Taylor, "Gaussian process modelling of dependencies in multi-armed bandit problems," in *Proc. Int. Symp. Oper. Res.*, 2009, pp. 77–84.
[28] S. Ghosal and A. Roy, "Posterior consistency of Gaussian process prior for nonparametric binary regression," *Ann. Stat.*, vol. 34, no. 5, pp. 2413–2429, 2006.
[29] M. W. Seeger, S. M. Kakade, and D. P. Foster, "Information consistency of nonparametric Gaussian process methods," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2376–2382, May 2008.
[30] L. C. W. Dixon and G. P. Szego, "The global optimization problem: An introduction," *Towards Global Optim.*, vol. 2, pp. 1–15, 1978.
[31] M. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag, 1999.
[32] D. A. Freedman, "On tail probabilities for martingales," *Ann. Prob.*, vol. 3, no. 1, pp. 100–118, 1975.
[33] C. McDiarmid, *Concentration. In Probabilistic Methods for Algorithmic Discrete Mathematics*. New York: Springer-Verlag, 1998.
[34] J. Shawe-Taylor, C. Williams, N. Cristianini, and J. Kandola, "On the eigenspectrum of the Gram matrix and the generalization error of kernel-PCA," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2510–2522, Jul. 2005.

Niranjan Srinivas received his B.S. and M.S. degrees in Mathematics and Scientific Computing from the Indian Institute of Technology at Kanpur (2008) and is currently a Ph.D. candidate in the Computation and Neural Systems program at the California Institute of Technology (Caltech).

Andreas Krause received his Diploma in Computer Science and Mathematics from the Technical University of Munich, Germany (2004), and his Ph.D. in Computer Science from Carnegie Mellon University (2008). He joined the California Institute of Technology as an Assistant Professor of computer science in 2009, and is currently an Assistant Professor in the Department of Computer Science at the Swiss Federal Institute of Technology Zürich. His research is in adaptive systems that actively acquire information, reason and make decisions in large, distributed and uncertain domains, such as sensor networks and the Web. Dr. Krause received an NSF CAREER Award, the Okawa Foundation Research Grant recognizing top young researchers in telecommunications, as well as awards at several premier conferences (AAAI, KDD, IPSN, ICML, UAI) and the ASCE *Journal of Water Resources Planning and Management*.

Sham M. Kakade is currently a Senior Research Scientist at Microsoft Research, New England, and an Associate Professor of statistics at the Wharton School at the University of Pennsylvania. He received his B.A. degree in Physics from the California Institute of Technology and his Ph.D. degree from the Gatsby Computational Neuroscience Unit affiliated with University College London. He spent the following two years as a Postdoctoral Researcher at the Department of Computer and Information Science at the University of Pennsylvania. Subsequently, he joined the Toyota Technological Institute, where he was an Assistant Professor for four years. He is now an Associate Professor at the Wharton Statistics Department at UPenn. His research focuses on artificial intelligence and machine learning, and their connections to other areas such as game theory and economics.

Matthias W. Seeger received the M.S. degree in computer science from the University of Karlsruhe, Germany, in 1999, and the Ph.D. degree in informatics from the University of Edinburgh, U.K., in 2003. From 2003 to 2005, he was a Research Associate in the EECS Department, University of California at Berkeley, and a Postdoctoral Research Fellow at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, until 2008. From 2008 until 2010, he was an Assistant Professor at the University of Saarbrücken, Germany, after which he joined EPFL. His research interests include analysis of approximate Bayesian inference methods and nonparametric Bayesian statistics.