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

Conference Paper · July 2010

Source: DBLP

4 authors, including:

Andreas Krause
ETH Zurich
285 PUBLICATIONS **17,143** CITATIONS
SEE PROFILE

Matthias Seeger
École Polytechnique Fédérale de Lausanne
48 PUBLICATIONS **4,051** CITATIONS
SEE PROFILE

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

Niranjan Srinivas NIRANJAN@CALTECH.EDU
Andreas Krause KRAUSEA@CALTECH.EDU
California Institute of Technology, Pasadena, CA, USA
Sham Kakade SKAKADE@WHARTON.UPENN.EDU
University of Pennsylvania, Philadelphia, PA, USA
Matthias Seeger MSEEGER@MMCI.UNI-SAARLAND.DE
Saarland University, Saarbrücken, Germany

Abstract

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

1. Introduction

In most stochastic optimization settings, evaluating the unknown function is expensive, and sampling is to be minimized. Examples include choosing advertisements in sponsored search to maximize profit in a click-through model (Pandey & Olston, 2007) or learning optimal control strategies for robots (Lizotte et al., 2007). Predominant approaches to this problem include the multi-armed bandit paradigm (Robbins, 1952), where the goal is to maximize cumulative reward by optimally balancing exploration and exploitation, and experimental design (Chaloner & Verdinelli, 1995), where the function is to be explored globally with as few evaluations as possible, for example by maximizing information gain. The challenge in both approaches is twofold: we have to estimate an unknown function f from noisy samples, and we must optimize our estimate over some high-dimensional input space. For the former, much progress has been made in machine learning through kernel methods and Gaussian process (GP) models (Rasmussen & Williams, 2006), where smoothness assumptions about f are encoded through the choice of kernel in a flexible nonparametric fashion. Beyond Euclidean spaces, kernels can be defined on diverse domains such as spaces of graphs, sets, or lists.

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

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

Appearing in *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, 2010.

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 (Nemhauser et al., 1978) and knowledge about kernel operator spectra.

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

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

Summary. Our main contributions are:

- We analyze GP-UCB, an intuitive algorithm for GP optimization, when the function is either sampled from a known GP, or has low RKHS norm.

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

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

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

2. Problem Statement and Background

Consider the problem of sequentially optimizing an unknown reward function $f : D \to \mathbb{R}$: 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: 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.

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

2.1. Gaussian Processes and RKHS's

Gaussian Processes. Some assumptions on f are required to guarantee no-regret. While rigid parametric assumptions such as linearity may not hold in practice, a certain degree of smoothness is often warranted. In our sensor network, temperature readings at closeby locations are highly correlated (see Figure 2(a)). We can enforce implicit properties like smoothness without relying on any parametric assumptions, modeling f as a sample from a *Gaussian process* (GP): a collection of dependent random variables, one for each $\boldsymbol{x} \in D$, every finite subset of which is multivariate Gaussian distributed in an overall consistent way (Rasmussen & Williams, 2006). A $GP(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$ 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 $k(\boldsymbol{x}, \boldsymbol{x}) \leq 1$, $\boldsymbol{x} \in D$, 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 5.2.

In this work, GPs play multiple roles. First, some of our results hold when the unknown target function is a sample from a known GP distribution $\mathrm{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$. 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 formulae for mean and covariance of the posterior distribution, which allows easy implementation of algorithms. For a noisy sample $\boldsymbol{y}_T = [y_1 \dots y_T]^T$ at points $A_T = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_T\}$, $y_t = f(\boldsymbol{x}_t) + \epsilon_t$ with $\epsilon_t \sim N(0, \sigma^2)$ i.i.d. Gaussian noise, the posterior over f is a GP distribution again, with mean $\mu_T(\boldsymbol{x})$, 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 $\boldsymbol{k}_T(\boldsymbol{x}) = [k(\boldsymbol{x}_1, \boldsymbol{x}) \dots k(\boldsymbol{x}_T, \boldsymbol{x})]^T$ 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, Wahba 1990) 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 $f \in \mathcal{H}_k(D)$. The induced RKHS norm $\|f\|_k = \sqrt{\langle f, f \rangle_k}$ measures smoothness of f w.r.t. k: in much the same way as k_1 would generate smoother samples than k_2 as GP covariance functions, $\|\cdot\|_{k_1}$ assigns larger penalties than $\|\cdot\|_{k_2}$. $\langle \cdot, \cdot \rangle_k$ can be extended to all of $L_2(D)$, in which case $\|f\|_k < \infty$ iff $f \in \mathcal{H}_k(D)$. For most kernels discussed in Section 5.2, members of $\mathcal{H}_k(D)$ can uniformly approximate any continuous function on any compact subset of D.

[2] This is w.l.o.g. (Rasmussen & Williams, 2006).

2.2. Information Gain & Experimental Design

One approach to maximizing f is to first choose points $\boldsymbol{x}_t$ so as to estimate the function globally well, then play the maximum point of our estimate. How can we learn about f as rapidly as possible? This question comes down to Bayesian Experimental Design (henceforth "ED"; see Chaloner & Verdinelli 1995), where the informativeness of a set of sampling points $A \subset D$ about f is measured by the *information gain* (c.f., Cover & Thomas 1991), which is the mutual information between f and observations $\boldsymbol{y}_A = \boldsymbol{f}_A + \epsilon_A$ 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 $\boldsymbol{\varepsilon}_A \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. For a Gaussian, $\mathrm{H}(N(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = \frac{1}{2} \log |2\pi e \boldsymbol{\Sigma}|$, 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 $A \subset D$, $|A| \leq T$ is NP-hard (Ko et al., 1995), it can be approximated by an efficient greedy algorithm. If $F(A) = \mathrm{I}(\boldsymbol{y}_A; f)$, this algorithm picks $\boldsymbol{x}_t = \mathrm{argmax}_{\boldsymbol{x} \in D} F(A_{t-1} \cup \{\boldsymbol{x}\})$ 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 $A_{t-1} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1}\}$. Importantly, this simple algorithm is guaranteed to find a near-optimal solution: for the set A_T obtained after T rounds, we have that

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

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

While sequentially optimizing Eq. 4 is a provably good way to *explore* f globally, it is not well suited for function optimization. For the latter, we only need to identify points $\boldsymbol{x}$ where $f(\boldsymbol{x})$ is large, in order to concen-

trate sampling there as rapidly as possible, thus *exploit* our knowledge about maxima. In fact, the ED rule (4) does not even depend on observations y_t obtained along the way. Nevertheless, the maximum information gain after T rounds will play a prominent role in our regret bounds, forging an important connection between GP optimization and experimental design.

3. GP-UCB Algorithm

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

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

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 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 *Gaussian process upper confidence bound* rule (GP-UCB), where β_t is specified depending on the context (see Section 4). Pseudocode for the GP-UCB algorithm is provided in Algorithm 1. Figure 2 illustrates two subsequent iterations, where GP-UCB both explores (Figure 2(b)) by sampling an input $\boldsymbol{x}$ with large $\sigma^2_{t-1}(\boldsymbol{x})$ and exploits (Figure 2(c)) by sampling $\boldsymbol{x}$ with large $\mu_{t-1}(\boldsymbol{x})$.

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

If D is infinite, finding $\boldsymbol{x}_t$ in (6) may be hard: the upper confidence index is multimodal in general. However, global search heuristics are very effective in practice (Brochu et al., 2009). It is generally assumed that evaluating f is more costly than maximizing the 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 (Kocsis & Szepesvári, 2006; Pandey & Olston, 2007; Lizotte et al., 2007). Their performance is well characterized in both the finite arm setting and the linear optimization setting, but no convergence rates for GP optimization are known.

4. Regret Bounds

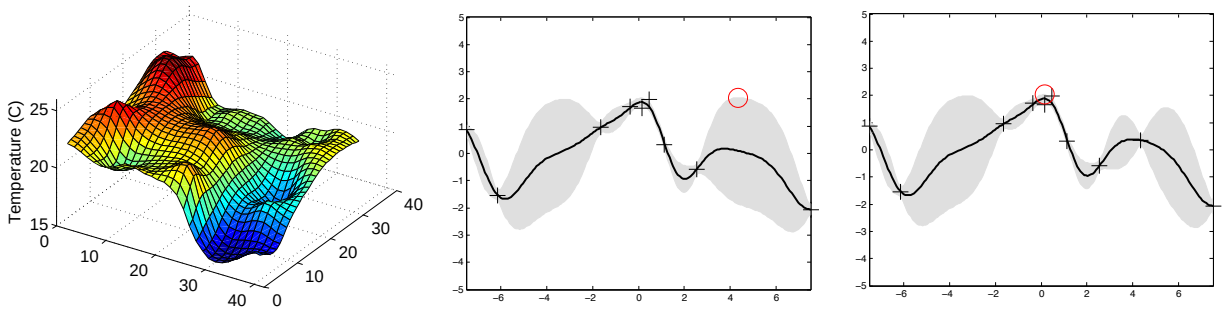
We now establish cumulative regret bounds for GP optimization, treating a number of different settings: $f \sim \mathrm{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$ for finite D, $f \sim \mathrm{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$ for general compact D, and the agnostic case of arbitrary f with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

GP optimization generalizes stochastic linear optimization, where a function f from a finite-dimensional linear space is optimized over. For the linear case, Dani et al. (2008) 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:

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

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. (2008) are of the form $\mathcal{O}^*(\sqrt{T\beta_T d})$ (d the dimensionality of the linear function space). Here and below, the $\mathcal{O}^*$ notation is a variant of $\mathcal{O}$, where log factors are suppressed. While our proofs – all provided in the longer version (Srinivas et al., 2009) – use techniques similar to those of Dani et al. (2008), we face a number of additional significant technical challenges. Besides avoiding the finite-dimensional analysis, we must handle confidence issues, which are more

[3] In general, d is the dimensionality of the input space D, which in the finite-dimensional linear case coincides with the feature space.



(a) *Temperature data* (b) *Iteration t* (c) *Iteration t + 1*

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

delicate for nonlinear random functions.

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

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

Theorem 1 *Let* $\delta \in (0,1)$ *and* $\beta_t = 2\log(|D|t^2\pi^2/6\delta)$. *Running GP-UCB with* 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 essentially relates the regret to the growth of the log volume of the confidence ellipsoid, and in a novel manner, shows how this growth is characterized by the information gain.

This theorem shows that, with high probability over samples from the GP, the cumulative regret is bounded in terms of the maximum information gain, forging a novel connection between GP optimization and experimental design. This link is of fundamental technical importance, allowing us to generalize Theorem 1 to infinite decision spaces. Moreover, the submodularity of $\mathrm{I}(\boldsymbol{y}_A;\boldsymbol{f}_A)$ allows us to derive sharp a priori bounds, depending on choice and parameterization of k (see Section 5). In the following theorem, we generalize our result to any compact and convex $D \subset \mathbb{R}^d$ under mild assumptions on the kernel function k.

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

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

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 the 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 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 (Theorem 5 of Ghosal & Roy (2006)), such as the Squared Exponential and Matérn kernels with $\nu > 2$ (see Section 5.2), while it is violated for the Ornstein-Uhlenbeck kernel (Matérn with $\nu = 1/2$; a stationary variant of the Wiener process). For the latter, sample paths f are nondifferentiable almost everywhere with probability one and come with independent increments. We conjecture that a result of the form of Theorem 2 does not hold in this case.

Bounds for Arbitrary f in the RKHS. Thus far, we have assumed that the target function f is sampled from a GP prior and that the noise is $N(0,\sigma^2)$ with known variance 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

$\mathbb{E}[\varepsilon_t \,|\, \boldsymbol{\varepsilon}_{<t}] = 0$ 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. Our following result shows that GP-UCB attains sublinear regret even in the agnostic setting.

Theorem 3 *Let $\delta \in (0,1)$. Assume that the true underlying f lies in the RKHS $\mathcal{H}_k(D)$ corresponding to the kernel $k(\boldsymbol{x},\boldsymbol{x}')$, 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 $GP(0, k(\boldsymbol{x},\boldsymbol{x}'))$ and noise model $N(0,\sigma^2)$, we obtain a regret bound of $\mathcal{O}^*(\sqrt{T}(B\sqrt{\gamma_T}+\gamma_T))$ with high probability (over the noise). Precisely,*

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

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

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

5. Bounding the Information Gain

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

5.1. 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 2, $F(A)$ is a submodular function, which implies the performance guarantee (5) for maximizing F sequentially by the greedy ED rule (4). Dividing both sides of (5) by $1-1/e$, we can upper-bound 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 2, the ED rule does not depend on observations y_t and can be run without evaluating f.

The importance of this greedy bound is twofold. First, it allows us to numerically compute highly problem-specific bounds on γ_T, which can be plugged into our results in Section 4 to obtain high-probability bounds on R_T. This being a laborious procedure, one would prefer *a priori* bounds for γ_T in practice which are simple analytical expressions of T and parameters of k. In this section, we sketch a general procedure for obtaining such expressions, instantiating them for a number of commonly used covariance functions, once more relying crucially on the greedy ED rule upper bound. Suppose that D is finite for now, and let $\boldsymbol{f} = [f(\boldsymbol{x})]_{\boldsymbol{x}\in D}$, $\boldsymbol{K}_D = [k(\boldsymbol{x},\boldsymbol{x}')]_{\boldsymbol{x},\boldsymbol{x}'\in D}$. 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 $\mathrm{spec}(\boldsymbol{K}_D) = \{\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots\}$. We can split the sum into two parts in order to obtain a bound to leading order. The following Theorem captures this intuition:

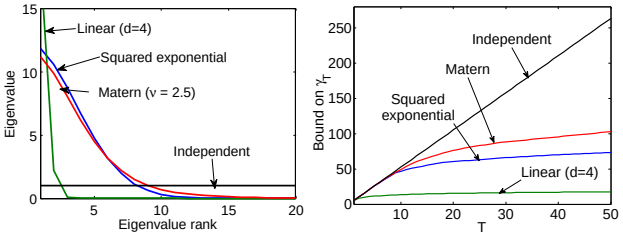


Figure 3. Spectral decay (left) and information gain bound (right) for independent (diagonal), linear, squared exponential and Matérn kernels ($\nu = 2.5$.) with equal trace.

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. Figure 3 illustrates this intuition.

5.2. 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 4, 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 a lengthscale parameter. Sample functions are differentiable to any order almost surely (Rasmussen & Williams, 2006).

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.

Theorem 5 *Let $D \subset \mathbb{R}^d$ be compact and convex, $d \in \mathbb{N}$. Assume 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 $\nu > 1$: $\gamma_T = \mathcal{O}\big(T^{d(d+1)/(2\nu+d(d+1))}(\log T)\big)$.*

We now provide a sketch of the proof. γ_T is bounded by Theorem 4 in terms the eigendecay of the kernel matrix $\boldsymbol{K}_D$. If D is infinite or very large, we can use the operator spectrum of $k(\boldsymbol{x}, \boldsymbol{x}')$, which likewise decays rapidly. For the kernels of interest here, asymptotic expressions for the operator eigenvalues are given in Seeger et al. (2008), who derived bounds on the information gain for fixed and random designs (in contrast to the worst-case information gain considered here, which is substantially more challenging to bound). The main challenge in the proof is to ensure the existence of discretizations $D_T \subset D$, dense in the limit, for which tail sums $B(T_*)/n_T$ in Theorem 4 are close to corresponding operator spectra tail sums.

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

6. Experiments

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

For synthetic data, we sample random functions from a squared exponential kernel with lengthscale parameter 0.2. The sampling noise variance 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). 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.

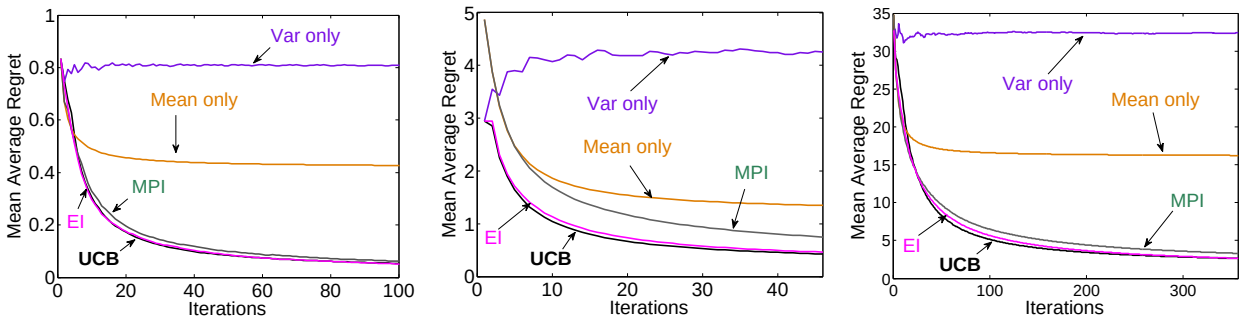
Next, we use temperature data collected from 46 sensors deployed at Intel Research Berkeley over 5 days at 1 minute intervals, pertaining to the example in Section 2. We take the first two-thirds of the data set to compute the empirical covariance of the sensor readings, and use it as the kernel matrix. The functions f for optimization consist of one set of observations from all the sensors taken from the remaining third of the data set, and the results (for $T = 46$, $\sigma^2 = 0.5$ or 5% noise, $\delta = 0.1$) were averaged over 2000 possible choices of the objective function.

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

Figure 4 compares the mean average regret incurred by the different heuristics and the GP-UCB algorithm on synthetic and real data. For temperature data, the GP-UCB algorithm and EI heuristic clearly outperform the others, and do not exhibit significant difference between each other. On synthetic and traffic data MPI does equally well. In summary, GP-UCB performs at least on par with the existing approaches which are not equipped with regret bounds.

7. Conclusions

We prove the first sublinear regret bounds for GP optimization with commonly used kernels (see Figure 1), both for f sampled from a known GP and f of low RKHS norm. We analyze GP-UCB, an intuitive, Bayesian upper confidence bound based sampling rule. Our regret bounds crucially depend on the information gain due to sampling, establishing a novel connection between bandit optimization and experimental design.



(a) *Squared exponential* (b) *Temperature data* (c) *Traffic data*

Figure 4. Comparison of performance: GP-UCB and various heuristics on synthetic (a), and sensor network data (b, c).

We bound the information gain in terms of the kernel spectrum, providing a general methodology for obtaining regret bounds with kernels of interest. Our experiments on real sensor network data indicate that GP-UCB performs at least on par with competing criteria for GP optimization, for which no regret bounds are known at present. Our results provide an interesting step towards understanding exploration–exploitation tradeoffs with complex utility functions.

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

References

Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *JMLR*, 3:397–422, 2002.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2-3):235–256, 2002.

Brochu, E., Cora, M., and de Freitas, N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. In *TR-2009-23, UBC*, 2009.

Bubeck, S., Munos, R., Stoltz, G., and Szepesvári, C. Online optimization in X-armed bandits. In *NIPS*, 2008.

Chaloner, K. and Verdinelli, I. Bayesian experimental design: A review. *Stat. Sci.*, 10(3):273–304, 1995.

Cover, T. M. and Thomas, J. A. *Elements of Information Theory*. Wiley Interscience, 1991.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. In *COLT*, 2008.

Dorard, L., Glowacka, D., and Shawe-Taylor, J. Gaussian process modelling of dependencies in multi-armed bandit problems. In *Int. Symp. Op. Res.*, 2009.

Ghosal, S. and Roy, A. Posterior consistency of Gaussian process prior for nonparametric binary regression. *Ann. Stat.*, 34(5):2413–2429, 2006.

Grünewälder, S., Audibert, J-Y., Opper, M., and Shawe-Taylor, J. Regret bounds for gaussian process bandit problems. In *AISTATS*, 2010.

Huang, D., Allen, T. T., Notz, W. I., and Zeng, N. Global optimization of stochastic black-box systems via sequential kriging meta-models. *J Glob. Opt.*, 34:441–466, '06.

Jones, D. R., Schonlau, M., and Welch, W. J. Efficient global optimization of expensive black-box functions. *J Glob. Opti.*, 13:455–492, 1998.

Kleinberg, R., Slivkins, A., and Upfal, E. Multi-armed bandits in metric spaces. In *STOC*, pp. 681–690, 2008.

Ko, C., Lee, J., and Queyranne, M. An exact algorithm for maximum entropy sampling. *Ops Res*, 43(4):684–691, 1995.

Kocsis, L. and Szepesvári, C. Bandit based monte-carlo planning. In *ECML*, 2006.

Krause, A. and Guestrin, C. Near-optimal nonmyopic value of information in graphical models. In *UAI*, 2005.

Lizotte, D., Wang, T., Bowling, M., and Schuurmans, D. Automatic gait optimization with Gaussian process regression. In *IJCAI*, pp. 944–949, 2007.

Mockus, J. *Bayesian Approach to Global Optimization*. Kluwer Academic Publishers, 1989.

Mockus, J., Tiesis, V., and Zilinskas, A. *Toward Global Optimization*, volume 2, chapter Bayesian Methods for Seeking the Extremum, pp. 117–128. 1978.

Nemhauser, G., Wolsey, L., and Fisher, M. An analysis of the approximations for maximizing submodular set functions. *Math. Prog.*, 14:265–294, 1978.

Pandey, S. and Olston, C. Handling advertisements of unknown quality in search advertising. In *NIPS*. 2007.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Robbins, H. Some aspects of the sequential design of experiments. *Bul. Am. Math. Soc.*, 58:527–535, 1952.

Seeger, M. W., Kakade, S. M., and Foster, D. P. Information consistency of nonparametric Gaussian process methods. *IEEE Tr. Inf. Theo.*, 54(5):2376–2382, 2008.

Srinivas, N., Krause, A., Kakade, S., and Seeger, M. Gaussian process optimization in the bandit setting: No regret and experimental design. arXiv:0912.3995, 2009.

Vazquez, E. and Bect, J. Convergence properties of the expected improvement algorithm, 2007.

Wahba, G. *Spline Models for Observational Data*. SIAM, 1990.