Bandit optimisation of functions in the Matérn kernel RKHS

David Janz
University of Cambridge

David R. Burt
University of Cambridge

Javier González
Amazon

Abstract

We consider the problem of optimising functions in the reproducing kernel Hilbert space (RKHS) of a Matérn kernel with smoothness parameter ν over the domain $[0,1]^d$ under noisy bandit feedback. Our contribution, the π-GP-UCB algorithm, is the first practical approach with guaranteed sublinear regret for all $\nu > 1$ and $d \geq 1$. Empirical validation suggests better performance and drastically improved computational scalablity compared with its predecessor, Improved GP-UCB.

1 Introduction

We consider the black-box function optimisation problem using the stochastic bandit formalism of sequential decision making (Robbins, 1952). Under this model, we consider an agent that sequentially selects an action x_t from an action set $\mathcal{X}$ at each time step $t = 1, \ldots, T$ for $T \in \mathbb{N}$ and observes

$$y(x_t) = f(x_t) + \epsilon_t,$$

where y is the reward function, f is the expected reward function and we assume ϵ_t is conditionally sub-Gaussian given x_t. The agent's goal is to minimise regret, given by

$$R_T = \sum_{t=1}^{T} f(x^\star) - f(x_t),$$

where $f(x^\star)$ is the reward associated with an optimal arm. Regret is closely linked to bounds on the convergence of black-box optimisation in the presence of noise: bounding R_T by a quantity sublinear in T implies $f(x^\star) - \max_{t \leq T} f(x_t) \leq R_T/T \to 0$ as $T \to \infty$,

and so an algorithm achieving such a bound will converge to a subset of the global optima of f.

A standard approach to the bandit problem is to construct a $1 - \delta$ confidence upper bound for f, of the form

$$\mu_{t-1}(x) + \beta_t(\delta)\sigma_{t-1}(x),$$

where μ_{t-1} and σ_{t-1} are the mean and standard deviation predictors for f given by a suitable estimator based on observations prior to time t, and $\beta_t(\delta)$ is a confidence width multiplier that ensures the expression is an upper bound on f with probability at least $1 - \delta$. Algorithms then select $x_t \in \mathcal{X}$ that maximises this bound (Auer, 2002; Auer et al., 2002; Auer and Ortner, 2010). This upper confidence bound (UCB) strategy naturally balances exploration, sampling in regions where the uncertainty is large, with exploitation, focusing on regions where the mean is large, and leads to algorithms with minimax optimal regret bounds for the case where $\mathcal{X}$ is finite (Audibert and Bubeck, 2009).

From the perspective of black-box function optimisation, a particularly interesting bandit problem is the kernelised continuum-armed bandit (Srinivas et al., 2010). Here, f is assumed to be in the closure of functions on $[0,1]^d$ expressible as a linear combination of a feature embedding parameterised by a kernel k. The properties of the functions in the resulting space, referred to as the RKHS of k, are determined by the choice of the kernel. For example, the RKHS corresponding to the linear kernel contains linear functions, and in this case existing kernelised bandit algorithms recover bounds that match those of the relevant stochastic linear bandit algorithms (Chowdhury and Gopalan, 2017; Abbasi-Yadkori et al., 2011). For a squared exponential kernel, the corresponding RKHS contains only infinitely differentiable functions, and here the existing methods match known lower bounds up to polylogarithmic factors (Srinivas et al., 2010; Scarlett et al., 2017).

In this work, we focus on the RKHS associated with a Matérn kernel, parameterised by a smoothness parameter ν (Stein, 2012). For a given ν, the Matérn RKHS contains all $\lfloor \nu \rfloor$-times continuously differentiable func-

tions, and therefore for any $\nu < \infty$, contains as a strict subset the RKHS of both the linear and the squared exponential kernels. The Matérn RKHS is of particular practical significance, since it offers a more suitable set of assumptions for the modelling and optimisation of physical quantities (Stein, 2012; Rasmussen and Williams, 2006). However, the theoretical guarantees offered for this class of functions by existing kernelised algorithms, such as KernelUCB (Valko et al., 2013), GP-UCB (Srinivas et al., 2010) and Improved GP-UCB (Chowdhury and Gopalan, 2017), are limited. Specifically, these guarantee that the regret after T steps, R_T, is bounded with high probability as[1]

$$R_T = \widetilde{O}\left(T^{\min\left\{1, \frac{3d^2+2\nu}{2d^2+4\nu}\right\}}\right),$$

leaving a large gap to the $\Omega(T^{\frac{d+\nu}{d+2\nu}})$ algorithm-agnostic lower bound for this problem (Scarlett et al., 2017). Since this bound is linear for $2\nu \leq d^2$, existing practical kernel-based algorithms are not guaranteed to converge for functions with fewer than $\lfloor d^2/2 \rfloor$ derivatives.

Our main contribution is an algorithm that successfully tackles the Matérn RKHS bandit problem for all $\nu > 1$ and all $d \geq 1$. The algorithm, Partitioned Improved GP-UCB (π-GP-UCB), offers a high probability regret bound of order

$$R_T = \widetilde{O}\left(T^{\frac{d(2d+3)+2\nu}{d(2d+4)+4\nu}}\right),$$

and therefore guarantees convergence for once differentiable functions of any finite dimensionality. Our contribution is not limited to theory: π-GP-UCB shows strong empirical performance on 3-dimensional functions in the Matérn $\nu = 3/2$ RKHS (in contrast, Improved GP-UCB barely outperforms uniformly sampling of arms). Moreover, our experiments show that despite using Gaussian process regression to construct confidence intervals, π-GP-UCB achieves an empirically near-linear runtime.

Our analysis also results in tighter bounds on the effective dimension associated with the Matérn kernel RKHS, an important quantity in the context of kernelised bandit problems which immediately improves bounds for a range of existing algorithms.

2 Background

Our work builds on Improved GP-UCB (IGP-UCB), and like IGP-UCB, it uses Gaussian process regression to construct confidence intervals. We briefly outline these topics and introduce the required notation.

---

[1]$f(t) = \widetilde{\mathcal{O}}(t^a) \iff \forall \epsilon > 0, \ f(t) = \mathcal{O}(t^{a+\epsilon}).$

Gaussian process regression We use $D_t^{\mathcal{X}} = \{(x_i, y_i): i \leq t\}$ to denote a set of t observations with $x_i \in \mathcal{X}$, and

$$D_t^A = \{(x_i^A, y_i^A) \in D_t^{\mathcal{X}}: x_i^A \in A\}$$

for $A \subset \mathcal{X}$ to denote the subset of these located in A. We denote the sequence of input locations within A by X_t^A and by $(y_1^A, \ldots, y_N^A)$ the associated sequence of observations.

For D_t^A of cardinality N, $x \in A$, a kernel k, assumed to be normalised such that $k(x, x) = 1$, we define

$$k_t^A(x) = [k(x_1^A, x), \ldots, k(x_N^A, x)]^T,$$

as well as

$$K_t^A = [k(x, x')]_{x, x' \in X_t^A} \quad \text{and} \quad y_{1:t}^A = [y_1^A, \ldots, y_N^A]^T.$$

For a regularisation parameter $\alpha > 0$, we define the Gaussian process regressor on $A \subset \mathcal{X}$ by a mean,

$$\mu_t^A(x) = k_t^A(x)^T (K_t^A + \alpha I)^{-1} y_{1:t}^A,$$

and an associated predictive standard deviation,

$$\sigma_t^A(x) = \sqrt{k(x, x) - k_t^A(x)^T (K_t^A + \alpha I)^{-1} k_t^A(x)},$$

for each $x \in A$. Note that $\sigma_t^A(x)$ is monotone decreasing in both t and A, meaning $\sigma_t^A(x) \leq \sigma_t^{A'}(x)$ if $A' \subset A$ and $\sigma_t^A(x) \leq \sigma_{t+1}^A(x)$ (Vivarelli, 1998).

Effective dimension While Gaussian process regressors are often defined through an infinite dimensional feature embedding, due to finite data and regularisation, the number of features that have a noticeable impact on the regression model can be small. The effective dimension of the Gaussian process regressor on the set $A \subset \mathcal{X}$, defined as

$$\widetilde{d}^A = \text{Tr}(K_T^A (K_T^A + \alpha I)^{-1}) = \alpha^{-1} \sum_{t=1}^{T} (\sigma_T^A(x_t))^2,$$

provides an estimate of the number of relevant features used in the regression problem (Zhang, 2005; Valko et al., 2013), and frequently appears in the bounds on kernelised bandit algorithms. It is closely related to information gain, defined

$$\gamma_t^A = \frac{1}{2} \log|I + \alpha^{-1} K_t^A|$$

for $A \subset \mathcal{X}$, with the two of the same order up to polylogarithmic factors (Calandriello et al., 2019, proposition 5). We shall use this relationship throughout.

Matérn kernel family The Matérn family consists of kernels of the form $k(x, x') = \kappa(r)$ where $r = x - x'$ and κ is the Fourier transform of a Student's t-density,

$$\lambda(\omega) = C_{\ell,d} \left(1 + (\ell \|\omega\|_2)^2\right)^{-\nu - d/2}, \qquad (1)$$

for ω the d-dimensional frequency vector and

$$C_{\ell,d} = \ell^d \frac{\Gamma(\nu + d/2)}{\pi^{d/2} \Gamma(\nu)}$$

for $\ell, \nu > 0$ parameters of the kernel.

The Improved GP-UCB algorithm The IGP-UCB algorithm is an approach to the kernelised bandit problem based on the classic GP-UCB Bayesian optimisation algorithm. For f in the RKHS of a known kernel, it improves on the regret bounds of GP-UCB by a polylogarithmic factor and significantly improves empirical performance. For readers familiar with stochastic linear bandits, the step from GP-UCB to IGP-UCB mirrors that from ConfidenceBall (Dani et al., 2008) to OFUL (Abbasi-Yadkori et al., 2011).

IGP-UCB assumes a known bound B on the RKHS norm of the function f and a known bound on the subGaussianity constant L of ϵ_t. Then, for a chosen $\delta \in (0, 1)$ and $Z \subset \mathbb{R}^d$, IGP-UCB uses a Gaussian process regressor mean $\mu_{t-1}^Z(x)$ and standard deviation $\sigma_{t-1}^Z(x)$, along with a confidence width multiplier

$$\beta_t^Z(\delta) = B + L\sqrt{2\left(\gamma_t^Z + 1 + \log(1/\delta)\right)}, \qquad (2)$$

to construct $1 - \delta$ probability confidence intervals for the restriction of f to Z. Chowdhury and Gopalan (2017) choose to use $Z = \mathcal{X}$, that is, consider the whole domain, and show that selecting observations that maximise this upper confidence bound leads to regret bounded as $\mathcal{O}(\gamma_T^{\mathcal{X}}\sqrt{T})$ with probability $1 - \delta$.

Intuitively, the IGP-UCB algorithm needs strong regularity assumptions to guarantee sublinear regret because the information gain term, $\gamma_t^{\mathcal{X}}$, contained within the confidence width parameter $\beta_t^{\mathcal{X}}(\delta)$, grows too quickly with t otherwise. Specifically, in the case of the Matérn kernel, $\gamma_t^{\mathcal{X}}$ can be lower and upper bounded as

$$\widetilde{\mathcal{O}}(T^{\frac{d(d+1)}{d(d+1)+2\nu}}) \quad \text{and} \quad \Omega(T^{\frac{d}{d+2\nu}}) \qquad (3)$$

respectively, where the former is given in Srinivas et al. (2010) and the latter can be deduced from combining the bounds in Valko et al. (2013) and Scarlett et al. (2017). Our main contribution can be understood as addressing this issue. We provide a construction which leads to confidence intervals that grow only polylogarithmically with T.

3 Main contribution

We introduce *Partitioned Improved GP-UCB* (π-GP-UCB), an algorithm that at each time step constructs a closed cover of the domain $\mathcal{X}$ and selects points by taking a maximiser of the IGP-UCB upper confidence bound constructed independently on each cover element. Throughout the paper, we will make use of the following two constants,

$$b = \frac{d+1}{d+2\nu} \quad \text{and} \quad q = \frac{d(d+1)}{d(d+2)+2\nu},$$

which depend on the RKHS only. Also, for a hypercube $A \subset \mathcal{X}$ we will use ρ_A denote its ℓ^∞-diameter, i.e. the length of any of its sides.

The π-GP-UCB algorithm Choose a $1 - \delta$ confidence level. Let $\mathcal{A}_1$ be any set of closed hypercubes overlapping at edges only, of cardinality at most $\mathcal{O}(T^q)$, that covers the domain $\mathcal{X}$.[2] At each time step t select a query location x_t and then construct a new cover $\mathcal{A}_t$ as follows:

Point selection. Fit an independent Gaussian process with $\alpha = 1 + 2/T$ on each cover element $A \in \mathcal{A}_t$, conditioned only on data within A, and select the next point to evaluate by maximising

$$\mathrm{UCB}_t(x) = \max_{A \in \mathcal{A}_t : x \in A} \mu_{t-1}^A(x) + \widehat{\beta}_t^A \sigma_{t-1}^A(x),$$

where $\widehat{\beta}_t^A = \beta_t^A(\delta/\widetilde{N}_t)$, with $\widetilde{N}_t = 4(t+1)^{bd}$.

Splitting rule. Split any element $A \in \mathcal{A}_{t-1}$ for which $\rho_A^{-1/b} < |D_t^A| + 1$ along the middle of each side, resulting in 2^d new hypercubes. Let $\mathcal{A}_t$ be the set of the newly created hypercubes and the elements of $\mathcal{A}_{t-1}$ which were not split.

Properties of algorithm The construction of the cover $\mathcal{A}_t$ ensures the following two properties hold:

Lemma 1. *Let A be a subset of $\mathcal{X}$ and suppose there exists a τ such that $A \in \mathcal{A}_\tau$. Let $\tau'(A) = \max\{t : A \in \mathcal{A}_t\}$. Then, for some $C > 0$, $\gamma_{\tau'(A)}^A \leq C \log T \log \log T$.*

Lemma 2. *Let $\mathcal{A}_t$ be the covering set at time t. Suppose $|\mathcal{A}_1| = \mathcal{O}(T^q)$. Then, for T sufficiently large,*

$$|\cup_{t \leq T} \mathcal{A}_t| \leq C_{d,\nu} T^q,$$

where $C_{d,\nu} > 0$ depends on d and ν only.

That is, on all cover elements, information gain can be bounded polylogarithmically for all $t \geq 1$, and the cardinality of all the covering sets generated up to time

---

[2]For example, $\mathcal{A}_1 = \{\mathcal{X}\}$.

T is sublinear in T. We will show that the regret of π-GP-UCB after T steps is bounded by

$$R_T = \mathcal{O}(\gamma_T^A \sqrt{T|\cup_{t \leq T} \mathcal{A}_t|}) = \widetilde{\mathcal{O}}(\sqrt{T|\cup_{t \leq T} \mathcal{A}_t|}) = o(T).$$

The formal statement of this result is:

Theorem 3. *Let $\mathcal{X} = [0,1]^d$, let $\mathcal{H}_k(\mathcal{X})$ be the RKHS of a Matérn kernel k with parameter $\nu > 1$ such that $k \leq 1$. Suppose $f : \mathcal{X} \to \mathbb{R}$ satisfies $\|f\|_{\mathcal{H}_k} \leq B$ for a known B and we observe $y(x_t) = f(x_t) + \epsilon_t$, where ϵ_t is L sub-Gaussian. Then for any fixed $\delta \in (0,1)$, with probability at least $1 - \delta$ the regret incurred by π-GP-UCB is bounded by $\widetilde{\mathcal{O}}(T^{\frac{d(2d+3)+2\nu}{d(2d+4)+4\nu}})$.*

The properties of the cover also allows us to improve upon existing upper bounds for the information gain associated with a Matérn kernel. To do this, we bound the information gain as the sum of the information gain on each cover element $A \in \mathcal{A}_t$, and therefore

$$\gamma_T^{\mathcal{X}} = \mathcal{O}(\gamma_T^A |\cup_{t \leq T} \mathcal{A}_t|) = \widetilde{\mathcal{O}}(|\cup_{t \leq T} \mathcal{A}_t|),$$

which translates into the following bound:

Theorem 4. *Information gain associated with the Matérn kernel with parameter $\nu > 1$ after T steps can be bounded as $\gamma_T^{\mathcal{X}} = \widetilde{\mathcal{O}}(T^{\frac{d(d+1)}{d(d+2)+2\nu}})$.*

This is a strict improvement over the bound presented in Srinivas et al. (2010), given here in equation (3).

Practical considerations While the main purpose of the cover construction within π-GP-UCB is to provide strong theoretical guarantees on regret, it also results in a significantly more scalable algorithm.

To see this, first note that fitting a Gaussian process requires updating the inverse of a kernel matrix K_t. Our cover construction means that inverses are only performed on a kernel matrix of a subset of the data, reducing both compute and memory costs. Empirically, this allows π-GP-UCB to scale much more favourably than GP-UCB. However, the construction offers no asymptotic improvement on computational complexity: once the algorithm comes close to convergence (enters the asymptotic regime), points are expected to land within small neighbourhoods of the optima, and therefore likely within the same cover element.

Moreover, running Gaussian process UCB algorithms requires maximising the UCB index across the domain. Whereas the classic GP-UCB algorithms require the maximiser to be recomputed across all of the domain, under π-GP-UCB, the maximisation procedure need only be carried out over the cover elements containing the most recent observation and any newly created cover elements.

4 Proof of results

In order to prove theorem 3, we will use the following concentration inequality:

Lemma 5. *Given $\delta \in (0,1)$, for all $t \leq T$, for all $A \in \bigcup_{t \leq T} \mathcal{A}_t =: \widetilde{\mathcal{A}}_T$, for all $x \in A$, we have*

$$|\mu_{t-1}^A(x) - f(x)| \leq \widehat{\beta}_t^A \sigma_{t-1}^A(x),$$

with probability $1 - \delta$.

To prove lemma 5, we will use the following result, proven in section 4.3, which bounds the number of all cover elements that could ever be created within t steps of running the algorithm by $\widetilde{N}_t = 4(t+1)^{bd}$.

Lemma 6. *There exists a set B_t such that $|B_t| \leq \widetilde{N}_t$ and $\mathcal{A}_t \subset B_t$ with probability one.*

We present the full proof of lemma 5 in appendix A. Here, we prove a weaker result that admits a much shorter proof.

By theorem 2 in Chowdhury and Gopalan (2017), under the conditions of our theorem 3,

$$|\mu_{t-1}^A(x) - f(x)| \leq \beta_t^A(\delta) \sigma_{t-1}^A(x)$$

with probability $1 - \delta$ for any $A \subset \mathbb{R}^d$ compact. The weaker result then follows by taking a union bound over all $A \in B_T$, resulting in a confidence width multiplier of $\beta_t^A(\delta/\widetilde{N}_T)$. The full proof in the appendix allows us to use $\widehat{\beta}_t^A = \beta_t^A(\delta/\widetilde{N}_t)$ instead.

To prove theorem 3 we will also need the following bound relating the sum of predictive variances on a subset of the domain to information gain:

Lemma 7. *For any $\tau \geq 1$, sequence $(x_t : x_t \in \mathcal{X}, 1 \leq t \leq \tau)$, set $A \subset \mathcal{X}$ and a Gaussian process estimator with $\alpha > 1$, $\sum_{t=1}^{\tau} 1\{x_t \in A\}(\sigma_{t-1}^A(x_t))^2 \leq 4\alpha\gamma_\tau^A$.*

Proof. Since $0 \leq (\sigma_t^A(x))^2, \alpha^{-1} \leq 1$ for all $x \in \mathcal{X}$, we have that $\alpha^{-1}(\sigma_t^A(x))^2 \leq 2\log(1 + \alpha^{-1}(\sigma_t^A(x))^2)$, because for any $0 \leq \sigma \leq 1$, $\sigma \leq 2\log(1+\sigma)$. Summing over t, we have

$$\sum_{t=1}^{\tau} 1\{x_t \in A\}(\sigma_{t-1}^A(x_t))^2$$

$$\leq \sum_{\substack{t=1, \\ x_t \in A}}^{\tau} 2\alpha \log\left(1 + \alpha^{-1}(\sigma_{t-1}^A(x_t))^2\right) = 4\alpha\gamma_\tau^A,$$

where the final equality is by Lemma 5.3 in Srinivas et al. (2010). $\qquad\square$

Proof of theorem 3. By Cauchy-Schwarz, $R_T \leq (T\sum_{t=1}^{T} r_t^2)^{1/2}$. It therefore suffices to bound $\sum_{t=1}^{T} r_t^2$.

For each $x \in \mathcal{X}$ let $A_t(x)$ be an element of $\mathcal{A}_t$ such that

$$A_t(x) \in \underset{A \in \mathcal{A}_t : x \in A}{\operatorname{argmax}} \mu_{t-1}^A(x) + \widehat{\beta}_t^A \sigma_{t-1}^A(x).$$

That is, $A_t(x)$ is an element of $\mathcal{A}_t$ on which the upper confidence bound associated with x is the highest.

For any $x^\star \in \operatorname{argmax}_{x \in \mathcal{X}} f(x)$, we have that $\mathrm{UCB}_t(x_t) \geq \mathrm{UCB}_t(x^\star)$ due to the manner in which points x_t are selected. Expanding this expression and applying lemma 5, we bound r_t as

$$r_t = f(x^\star) - f(x_t) \leq 2\widehat{\beta}_t^{A_t(x_t)} \sigma_{t-1}^{A_t(x_t)}(x_t)$$

with probability $1 - \delta$ for all $t \leq T$.

Denote $\widetilde{\mathcal{A}}_T = \cup_{t \leq T} \mathcal{A}_t$, the set of all cover elements created until time T, and define the initial time for an element $A \in \widetilde{\mathcal{A}}_T$, as $\tau(A) = \min\{t : A \in \mathcal{A}_t\}$ and the terminal time as $\tau'(A) = \max\{t : A \in \mathcal{A}_t\}$. We have

$$\sum_{t=1}^T r_t^2 \leq 4 \sum_{t=1}^T (\widehat{\beta}_t^{A_t(x_t)})^2 (\sigma_{t-1}^{A_t(x_t)}(x_t))^2$$

$$\leq 4 \sum_{t=1}^T \sum_{A \in \mathcal{A}_t} \mathbb{1}\{x_t \in A\} (\widehat{\beta}_t^A)^2 (\sigma_{t-1}^A(x_t))^2$$

$$= 4 \sum_{A \in \widetilde{\mathcal{A}}_T} \sum_{t=\tau(A)}^{\tau'(A)} \mathbb{1}\{x_t \in A\} (\widehat{\beta}_t^A)^2 (\sigma_{t-1}^A(x_t))^2$$

$$\leq 48 \sum_{A \in \widetilde{\mathcal{A}}_T} (\widehat{\beta}_{\tau'(A)}^A)^2 \gamma_{\tau'(A)}^A \qquad (*)$$

The final inequality uses monotonicity of $\widehat{\beta}_t^A$, lemma 7 and $\alpha < 3$. By lemma 2, the number of summands in $(*)$ is $\mathcal{O}(T^q)$. As $(\widehat{\beta}_{\tau'(A)}^A)^2 = \mathcal{O}(\gamma_{\tau'(A)}^A + \log(T))$ for all $A \in \widetilde{\mathcal{A}}_T$, lemma 1 completes the proof. $\qquad \square$

Proof of theorem 4. From proposition 5 in Calandriello et al. (2019),

$$\widetilde{d}^A \leq \log|I + \alpha^{-1} K_T^A| \leq \widetilde{d}^A (1 + \log(\alpha^{-1} \|K_T^A\|_2 + 1)).$$

Noting that $\|K_T^A\|_2 = \lambda_1^A \leq \sum_{t=1}^{|X_T^A|} \lambda_t^A = \mathrm{Tr}(K_T^A) = |X_T^A| \leq T$, and taking $A = \mathcal{X}$ we have,

$$\widetilde{d}^{\mathcal{X}} \leq 2\gamma_T^{\mathcal{X}} \leq \widetilde{d}^{\mathcal{X}} (1 + \log(\alpha^{-1} T + 1)).$$

We now proceed to bound $\widetilde{d}^{\mathcal{X}}$. At each time step t choose $\mathcal{A}_t$ to be as in the closed cover of $\mathcal{X}$ from π-GP-UCB. Let $\widetilde{\mathcal{A}}_T = \cup_{t \leq T} \mathcal{A}_t$, as before.

For any $A \in \widetilde{\mathcal{A}}_T$, let $\tau'(A) = \max\{t : A \in \mathcal{A}_t\}$. Then, because x_t is in at least one of the A_t for all t and by monotonicity of predictive variance, we have

$$\widetilde{d}^{\mathcal{X}} \leq \frac{1}{\alpha} \sum_{t=1}^T \sum_{A \in \mathcal{A}_t} \mathbb{1}\{x_t \in A\} (\sigma_{\tau'(A)-1}^A(x_t))^2.$$

Interchanging the order of summation and applying lemma 7, we have that this is upper bounded by $4 \sum_{A \in \widetilde{\mathcal{A}}_T} \gamma_{\tau'(A)}^A$. Then, by lemma 1, we have that

$$\widetilde{d}^{\mathcal{X}} \leq 4C |\widetilde{\mathcal{A}}_T| \log T \log \log T.$$

By lemma 2, $|\widetilde{\mathcal{A}}_T| = \mathcal{O}(T^q)$. Therefore

$$\gamma_T^{\mathcal{X}} \leq C' T^q \log T (\log \log T)(1 + \log(\alpha^{-1} T + 1)),$$

for some C' independent of T. $\qquad \square$

4.1 Proof of lemma 1, information gain on a cover element

To prove lemma 1, we rely on the following theorem from Srinivas et al. (2010). It provides a bound on the maximum information gain after N samples, $\hat{\gamma}(A, N)$, in terms of the operator spectrum of the kernel k with respect to a uniform covariate distribution.[3]

Theorem 8. *(Srinivas et al., 2010, Theorem 8) Suppose that $A \subset \mathbb{R}^d$ is compact, and k is kernel continuously differentiable in a neighbourhood of A. Let $\mathcal{S}(s_0) = \sum_{s > s_0} \lambda_s$ where $(\lambda_s)_s$ is the operator spectrum of k with respect to the uniform distribution over A. Pick $\zeta > 0$ and let $n_N = (4\zeta + 2)\mathcal{V}_A N^\zeta \log N$, where $\mathcal{V}_A$ is the volume of A. Then $\hat{\gamma}(A, N)$ is bounded by*

$$C \max_{r=1,\dots,N} \left[s_0 \log \frac{r n_N}{\alpha} + (4\zeta + 2) \mathcal{V}_A \log N \alpha^{-1} \right.$$
$$\left. \times \left(1 - \frac{r}{N} \right) \left(N^{\zeta+1} \mathcal{S}(s_0) + 1 \right) \right] + \mathcal{O}\left(N^{1-\zeta/d} \right), \quad (4)$$

where $C = \frac{1}{2}/(1 - \frac{1}{e})$ for any $s_0 \in \mathbb{N} \cap [1, n_N]$.

The operator spectrum of the Matérn kernel, required to use theorem 8, can be bounded using the following.

Theorem 9. *(Seeger et al., 2008, Theorem 2) Let $K(r)$ be an isotropic covariance function on $\mathbb{R}^d$ satisfying the conditions of Widom's theorem (Widom, 1963), with a spectral density $\lambda(\cdot)$. Suppose that the covariate distribution μ has bounded support and a bounded density, in that $\mu(x) \leq D$ for all x and $\mu(x) = 0$ for all $\|x\|_2 > R$. Then,*

$$\lambda_s \leq D(2\pi)^d \lambda \left(C_d R^{-1} s^{1/d} \right) (1 + o(1))$$

asymptotically as $s \to \infty$, where $C_d > 0$.

The required spectral density of a Matérn kernel is given in equation (1). By Seeger et al. (2008), it satisfies the conditions of Widom's theorem.

---

[3] The original statement of Theorem 8 in Srinivas et al. (2010) assumed strong conditions on the sample paths of the Gaussian process associated with the kernel. These assumptions were not necessary for the theorem itself, and were present due to the later use of the bounds for optimising Gaussian process samples.

Proof of lemma 1. As the information gain is a sum of non-negative elements, we have for any $M \leq N$,

$$\hat{\gamma}(A, M) \leq \hat{\gamma}(A, N).$$

It therefore suffices to bound the maximum number of points that can fall in a partition A before it would split. Let N_A denote this quantity. The proof proceeds in two parts: first we bound $\hat{\gamma}(A, N_A)$ as a function of N_A, then we bound N_A in terms of the horizon.

From equation (1), we have that the spectral density for the Matérn kernel satisfies

$$\lambda(\omega) = C_{\ell,d}(1 + (\ell\omega)^2)^{-\nu - d/2} \leq C'_{\ell,d}\omega^{-(2v+d)}.$$

where $C'_{\ell,d} = C_{\ell,d}\ell^{-(2v+d)}$. Utilising this within theorem 9 with a uniform covariate distribution,

$$\lambda_s \leq C\rho_A^{-d}(s^{1/d}\rho_A^{-1})^{-(2v+d)} = C\rho_A^{2\nu}s^{-(2\nu+d)/d},$$

for some constant $C > 0$, where we used $\mu(x) = \mathcal{V}_A^{-1} = \rho_A^{-d}$ for all $x \in A$, as A is a d-dimensional cube.

As the bound on λ_s for large s is monotonically decreasing, we can bound the tail of the Matérn kernel operator spectrum $\mathcal{S}(s_0)$ as

$$\sum_{s>s_0} \lambda_s \leq C\rho_A^{2\nu}\int_{s=s_0}^{\infty} s^{-(2\nu+d)/d} = \mathcal{O}(\rho_A^{2\nu}).$$

We now apply theorem 8 in order to bound $\hat{\gamma}(A, N_A)$. Choose $\zeta = d$, then

$$N_A^{\frac{d(2\nu-1)}{2\nu+d}} < n_{N_A} < N_A^d$$

for N_A sufficiently large. Choose $s_0 = \lfloor \log \log N \rfloor$.

For these parameter choices and for any $r \leq N_A$,

$$s_0 \log \frac{rn_{N_A}}{\alpha} = \mathcal{O}(\log N_A \log \log N_A).$$

As $1 - r/N_A \leq 1$, the second term in the maximum in equation (4) is $\mathcal{O}(\mathcal{V}_A N_A^{d+1}\rho_A^{2\nu} \log N_A)$. As the diameter satisfies $N_A \leq \rho_A^{-1/b} < N_A + 1$ by the definition of N_A this term is $\mathcal{O}(\log N_A)$. The final term in equation (4) is $\mathcal{O}(1)$.

All that remains is to bound N_A. We consider two cases: first, if $A \in \mathcal{A}_1$ then $N_A \leq \rho_A^{-1/b} = \mathcal{O}(T^{\frac{q}{bd}})$. Otherwise, A was created by some set A' splitting, for which $N_{A'} < T$. Then,

$$N_A \leq \rho_A^{-1/b} = 2^{-1/b}\rho_{A'}^{-1/b}$$
$$\leq 2^{-1/b}(N_{A'} + 1) < 2^{-1/b}(T+1).$$

In both cases, $N_A = \mathcal{O}(T)$. $\qquad\square$

4.2 Proof of lemma 2, bound on size of cover

Proof. First, since any element $A \in \widetilde{\mathcal{A}}_T$ was either in $\mathcal{A}_1$ or was created by the splitting of a cover element $A' \in \widetilde{\mathcal{A}}_T \setminus \mathcal{A}_T$ into 2^d elements, we have that

$$|\widetilde{\mathcal{A}}_T| = |\mathcal{A}_1| + 2^d|\widetilde{\mathcal{A}}_T \setminus \mathcal{A}_T|.$$

By assumption $|\mathcal{A}_1| = \mathcal{O}(T^q)$. Denote $\Theta_T = \widetilde{\mathcal{A}}_T \setminus \mathcal{A}_T$. We now upper bound $|\Theta_T|$. Take the diameter of any element $A \in \mathcal{A}_1$ to be $\rho_0 \in (0, 1]$ (e.g. $\rho_0 = 1$ if $\mathcal{A}_1 = \{\mathcal{X}\}$). For any given T, we have that

$$|\Theta_T| \leq \max_{x_1,\dots,x_T}|\Theta_T| = \max_{x_1,\dots,x_T}\sum_{i=0}^{\infty}\sum_{A \in \Theta_T}1\{\rho_A = 2^{-i}\rho_0\}.$$

We upper bound the solution to this maximisation problem by considering just a subset of the constraints imposed by the splitting procedure.

First constraint: we have a *budget constraint* derived from placing T points. Let $\tau(A) = \min\{t: A \in \mathcal{A}_t\}$ and $\tau'(A) = \max\{t: A \in \mathcal{A}_t\}$, and suppose there exists an $M(\rho_A) \leq |X_{\tau'(A)}^A| - |X_{\tau(A)}^A|$ for all $A \in \Theta_T$. Then

$$\sum_{i=0}^{\infty}M(2^{-i}\rho_0)\sum_{A \in \Theta_T}1\{\rho_A = 2^{-i}\rho_0\}$$
$$\leq \sum_{i=0}^{\infty}\sum_{A \in \Theta_T}(|X_{\tau'(A)}^A| - |X_{\tau(A)}^A|)1\{\rho_A = 2^{-i}\rho_0\}$$
$$= \sum_{A \in \Theta_T}|X_{\tau'(A)}^A| - |X_{\tau(A)}^A| \leq \sum_{A \in \widetilde{\mathcal{A}}_T}|X_{\tau'(A)}^A| - |X_{\tau(A)}^A|$$
$$= \sum_{A \in \widetilde{\mathcal{A}}_T}\sum_{t=\tau(A)}^{\tau'(A)}1\{x_t \in A\} = \sum_{t=1}^{T}\sum_{A \in \mathcal{A}_t}1\{x_t \in A\}$$
$$= \sum_{t=1}^{T}|\{A \in \mathcal{A}_t: x_t \in A\}| \leq 2^dT.$$

Now we find a suitable $M(\cdot)$, which we shall refer to as the *cost* of splitting an element $A \in \Theta_T$. Because A split,

$$|X_{\tau'(A)}^A| + 1 > \rho_A^{-1/b} \geq |X_{\tau'(A)}^A|$$

Suppose that A' is the element that split to create A. Then $\rho_{A'} = 2\rho_A$ and $\tau'(A') + 1 = \tau(A)$. Therefore

$$|X_{\tau'(A)}^A| - |X_{\tau(A)}^A| \geq |X_{\tau'(A)}^A| - |X_{\tau'(A')}^{A'}| - 1$$
$$\geq \rho_A^{-1/b}(1 - 2^{-1/b}) - 2 = M(\rho_A).$$

Second constraint: a *supply constraint*. There are at most $\lceil \rho_0^{-d} \rceil$ elements of diameter ρ_0, and therefore at
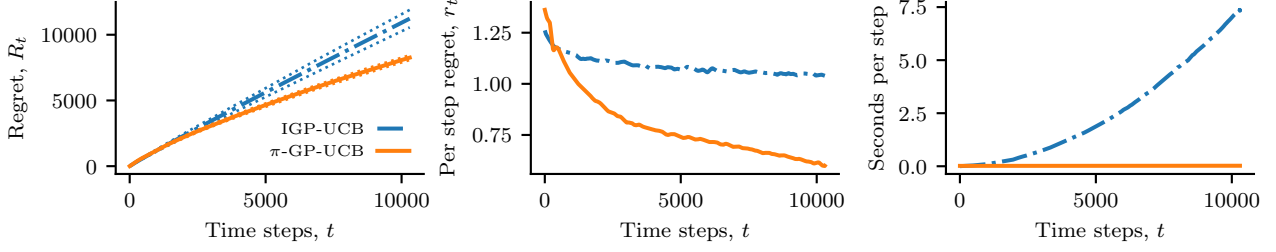
Figure 1: Comparison of performance of π-GP-UCB and IGP-UCB on the synthetic benchmark. Left, regret, 95% confidence interval dotted. Centre, per step regret, smoothed using a convolution with a top hat of length 200. Right, wall-clock time per step, in seconds.

most $\lceil \rho_0^{-d} \rceil 2^{di}$ elements of diameter $2^{-i}\rho_0$ can be split, leading to

$$\sum_{A \in \Theta_T} 1\{\rho_A = 2^{-i}\} \leq \lceil \rho_0^{-d} \rceil 2^{di}.$$

Since $M(2^{-i}\rho_0)$ increases with i, the solution to the relaxed optimisation problem will be to buy all the available A with smallest diameter, subject to the supply and budget constraints. Suppose the smallest A split with this strategy has a diameter $2^{-z}\rho_0$ for some $z \in \mathbb{N}$. Then, since the supply constraint is binding and budget constraint is satisfied, we have that

$$\sum_{i=0}^{z-1} M(2^{-i}\rho_0) 2^{di} \lceil \rho_0^{-d} \rceil \leq 2^d T.$$

Writing $\rho_0 = CT^\alpha$ for some $C > 0$ and $\alpha \leq q/d$ and using the geometric series formula to solve for 2^z, we obtain $2^z = \mathcal{O}(T^{\frac{b}{bd+1}})$, a quantity independent of α. Counting all the cover elements of diameters $2^0\rho_0, \ldots, 2^z\rho_0$, we have that $\sum_{i=0}^{z} 2^{di} = \mathcal{O}(2^{dz})$, which is $\mathcal{O}(T^q)$. Since this was a construction that maximises Θ_T, we have that $\widetilde{\mathcal{A}}_T = \mathcal{O}(T^q)$. □

4.3 Proof of Lemma 6

Proof. Let $B^0 = \mathcal{A}_1$ and define recursively B^{i+1} to be the set of the hypercubes created by splitting each element in B^i into 2^d hypercubes. Let $\rho_0 \in (0, 1]$ be the diameter of elements in $\mathcal{A}_1$. We have that

$$\mathcal{A}_t \subset \bigcup_{i \geq 0} B^i \implies \mathcal{A}_t \subset \bigcup_{i \geq 0} (B^i \cap \mathcal{A}_t).$$

Suppose there exists $Z \in B^i \cap \mathcal{A}_t$ for some $i \geq 0$. By the splitting condition,

$$Z \in \mathcal{A}_t \setminus B^0 \implies \rho_Z > (|X_t^Z|+1)^{-b} \geq (t+1)^{-b}.$$

Also, $Z \in B^i$ implies $\rho_Z \leq 2^{-i}\rho_0$. Therefore, $(t+1)^{-b} \leq \rho_Z \leq 2^{-i}\rho_0$ so $2^i \leq (t+1)^b \rho_0 = 2^{J_t}$. Let

$B_t = \bigcup_{i \leq J_t} B^i$. We have $B^i \cap \mathcal{A}_t = \emptyset$ for all $i > J_t$ and hence $\mathcal{A}_t \subset B_t$. We now bound the cardinality of B_t. We have $|B^i| = \lceil \rho_0^{-d} \rceil 2^{di}$, so

$$|B_t| = \sum_{i=0}^{J_t} |B^i| = \lceil \rho_0^{-d} \rceil \sum_{i=0}^{\lfloor J_t \rfloor} 2^{di} \leq \lceil \rho_0^{-d} \rceil 2^{dJ_t+1}$$

$$\leq 4(t+1)^{db}. \qquad \square$$

5 Empirical validation

We present an empirical comparison of π-GP-UCB and IGP-UCB on two types of functions: first, synthetic functions in the Matérn kernel RKHS, where the conditions of the theory for both algorithms are met; second, standard global optimisation baselines, corresponding to a more realistic setting where the RKHS norm of the target function is not known.

We run both IGP-UCB and π-GP-UCB using a Matérn kernel with parameters $\nu = 3/2$ and $\ell = 1/5$, and use a regularisation parameter $\alpha = 1$. For simplicity, the problems are discretised onto a regular grid, such that $\mathcal{X} = \{x_1, \ldots, x_n : n = 30^d\}$. We use a confidence parameter $\delta = 1/10$, and compute the quantities $\gamma_t^{\mathcal{X}}$ and γ_t^A exactly at each time step. For our method, π-GP-UCB, we use an initial partition $\mathcal{A}_1$ of cardinality approximately T^q.

Synthetic functions We benchmark on a set of synthetic functions which satisfy the assumptions behind the regret bounds of both algorithms. We construct each function f by sampling $m = 30d$ points, $\hat{x}_1, \ldots, \hat{x}_m$, uniformly on $[0, 1]^d$, and $\hat{a}_1, \ldots, \hat{a}_m$ each independent uniform on $[-1, 1]$ and defining $f(x) = \sum_{i=1}^{m} \hat{a}_j k(\hat{x}_j, x)$ for all $x \in \mathcal{X}$, where k is a Matérn $\nu = 3/2$ kernel with lengthscale $\ell = 1/5$. Both algorithms are given access to the exact RKHS norm of this function, computed as $\|f\|_k^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \hat{a}_i \hat{a}_j k(\hat{x}_j, \hat{x}_i)$. Evaluations of these functions are corrupted with independent additive noise ϵ_t sampled uniformly on $[-1, 1]$, and both algorithms use $L = 1$, the corresponding sub-Gaussianity constant.
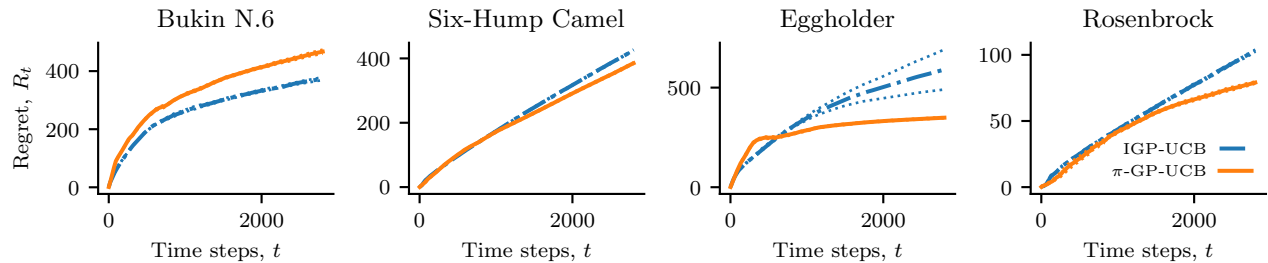
Figure 2: Comparison of performance of π-GP-UCB and IGP-UCB on the common two-dimensional global optimisation benchmark problems. Regret and 95% confidence intervals plotted.

d	IGP-UCB		π-GP-UCB	
	regret	runtime	regret	runtime
1	0.11	82 min	0.09	12 min
2	0.71	118 min	0.52	97 s
3	0.97	30 hours	0.77	29 min

Table 1: Tabulation of results for synthetic task with $d = 1, 2, 3$. Averages values over 12 runs for a horizon $T = 10000$. Regret is expressed as fraction of the regret incurred by pulling the arms uniformly.

We present regret and run-times on this set of synthetic functions for $d = 1, 2, 3$ in table 1 and plot the results for $d = 2$ in figure 1. The results suggest that π-GP-UCB not only provides an improved worst-case analysis, but may also yield improved empirical performance. Moreover, we see that π-GP-UCB is significantly more scalable. We note that the runtime of π-GP-UCB is high in the case $d = 1$. This is because in one dimension, π-GP-UCB converges rapidly, losing its computational advantages thereafter.

Global optimisation baselines We also provide a comparison of the empirical performance of π-GP-UCB and IGP-UCB on four common two-dimensional global optimisation baselines,[4] with results plotted in figure 2. The function outputs are scaled to $[-1, 1]$, and corrupted with additive noise uniform on $[-0.1, 0.1]$. Whilst we provide the tight subGaussianity parameter to both algorithms, we run both with the RKHS norm parameter $B = 1$, corresponding to a realistic setting where the RKHS norm is not known in advance. Our algorithm, π-GP-UCB, performs competitively across this set of problems.

6 Related work

While our theoretical guarantees are much stronger than those of existing practical kernelised methods like GP-UCB, IGP-UCB and KernelUCB, two exist-

ing methods achieve similar or better guarantees:

SupKernelUCB. Introduced in Valko et al. (2013), SupKernelUCB uses a phased elimination procedure to create batches of observations that are independent of previous observations. This allows for the use of the stronger concentration inequalities that apply to i.i.d. sequences, yielding a $\widetilde{\mathcal{O}}((\widetilde{d}^{\mathcal{X}}T)^{1/2})$ regret bound. While the algorithm is introduced for the case of a finite-armed bandit it can be extended to a continuum-armed bandit via a discretisation argument. SupKernelUCB is the kernelised version of stochastic linear bandit algorithm SupLinUCB (Auer, 2002; Chu et al., 2011). However, much like SupLinUCB (Lattimore and Szepesvri, 2020, Remarks 22.2), it fails to achieve empirically sublinear regret even on very simple problems (Calandriello et al., 2019).

Hierarchical optimisation. Extensions and generalisations of classic Lipschitz-continuity based methods enjoy strong regret guarantees under assumptions that are broadly similar to those in our work (Jones et al., 1993; Munos, 2011; Bubeck et al., 2011). Current upper bounds suggests for hierarchical methods are better for problems with a low degree of smoothness than the kernelised counterparts. It is an open question whether this holds in general, or whether the analysis of kernelised methods can be further improved.

7 Discussion

We have presented an algorithm for optimising functions in the RKHS of a Matérn family kernel, with a sublinear bound on regret for all smoothness parameters $\nu > 1$ and demonstrated the practical effectiveness and scalability of the proposed algorithm. The empirical performance of π-GP-UCB might be improved by using the actual information gain, as opposed to an upper bound on the information gain, in determining when to split a set.

---

[4]See `http://www.sfu.ca/~ssurjano` (Surjanovic and Bingham, 2013) for information on these functions.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

Jean-Yves Audibert and Sébastien Bubeck. Minimax policies for adversarial and stochastic bandits. In *COLT*, pages 217–226, 2009.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Peter Auer and Ronald Ortner. UCB revisited: Improved regret bounds for the stochastic multi-armed bandit problem. *Periodica Mathematica Hungarica*, 61(1-2):55–65, 2010.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(May):1655–1695, 2011.

Daniele Calandriello, Luigi Carratino, Alessandro Lazaric, Michal Valko, and Lorenzo Rosasco. Gaussian process optimization with adaptive sketching: Scalable and no regret. In *Conference on Learning Theory (COLT)*, pages 1–25, 2019.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 844–853, 2017.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 208–214, 2011.

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *COLT*, 2008.

Donald R Jones, Cary D Perttunen, and Bruce E Stuckman. Lipschitzian optimization without the lipschitz constant. *Journal of optimization Theory and Applications*, 79(1):157–181, 1993.

Tor Lattimore and Csaba Szepesvri. *Bandit Algorithms*. Cambridge University Press, 2020.

Rémi Munos. Optimistic optimization of a deterministic function without the knowledge of its smoothness. In *Advances in Neural Information Processing Systems*, pages 783–791, 2011.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2006. ISBN 026218253X.

Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy Gaussian process bandit optimization. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65, pages 1723–1742, 2017.

Matthias W Seeger, Sham M Kakade, and Dean P Foster. Information consistency of nonparametric Gaussian process methods. *IEEE Transactions on Information Theory*, 54(5):2376–2382, 2008.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 1015–1022, 2010.

Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

S. Surjanovic and D. Bingham. Virtual library of simulation experiments: Test functions and datasets, 2013. URL http://www.sfu.ca/~ssurjano.

Michal Valko, Nathan Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. Finite-time analysis of kernelised contextual bandits. In *Uncertainty in Artificial Intelligence*, 2013.

Francesco Vivarelli. *Studies on the generalisation of Gaussian processes and Bayesian neural networks*. PhD thesis, Aston University, 1998.

Harold Widom. Asymptotic behavior of the eigenvalues of certain integral equations. *Transactions of the American Mathematical Society*, 109(2):278–295, 1963.

Tong Zhang. Learning bounds for kernel regression using effective data dimensionality. *Neural Computation*, 17(9):2077–2098, 2005.