ANALYSIS OF THOMPSON SAMPLING FOR GAUSSIAN PROCESS OPTIMIZATION IN THE BANDIT SETTING

By Kinjal Basu and Souvik Ghosh

LinkedIn Corporation

We consider the global optimization of a function over a continuous domain. At every evaluation attempt, we can observe the function at a chosen point in the domain and we reap the reward of the value observed. We assume that drawing these observations are expensive and noisy. We frame it as a continuum-armed bandit problem with a Gaussian Process prior on the function. In this regime, most algorithms have been developed to minimize some form of regret. Contrary to this popular norm, in this paper, we study the convergence of the sequential point $\boldsymbol{x}^t$ to the global optimizer $\boldsymbol{x}^*$ for the Thompson Sampling approach. Under some assumptions and regularity conditions, we show an exponential rate of convergence to the true optimal.

1. Introduction. Let $f : \mathcal{X} \to \mathbb{R}$ be an unknown function defined on a compact set $\mathcal{X} \subset \mathbb{R}^d$. We are interested in solving the global maximization problem and obtaining the global maximizer

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in \mathcal{X}}{\operatorname{argmax}} f(\boldsymbol{x}).$$

For simplicity, we assume that $\boldsymbol{x}^*$ is unique, i.e. the function f has a unique global maximizer. We further assume that the space $\mathcal{X}$ is continuous. Such optimization problems are common in scientific and engineering fields. Examples include learning continuous valuation models (Eric, Freitas and Ghosh, 2008), automatic gait optimization for both quadrupedal and bipedal robots (Lizotte et al., 2007), choosing the optimal derivative of a molecule that best treats a disease (Negoescu, Frazier and Powell, 2011), tuning Hamiltonian based Monte Carlo Samplers (Wang, Mohamed and de Freitas, 2013), etc. A good survey of the problem in practical machine learning applications is presented in Snoek, Larochelle and Adams (2012). We were motivated to study this problem with the application of ranking multiple items on a webpage so as to optimize a diverse range of business metrics like user engagement and revenue from advertisements. In our example, the function $f(\boldsymbol{x})$ is a utility function composed of various business metrics and $\boldsymbol{x}$ are parameters or knobs that control the relative frequency of different types of items we show on the webpage.

Throughout this paper, we assume that function evaluations are noisy. At every attempt, we choose $\boldsymbol{x} \in \mathcal{X}$ and observe $y = f(\boldsymbol{x}) + \epsilon$, where ϵ are independent errors in each observation with $\epsilon \sim N(0, \sigma^2)$ and σ^2 is unknown (but fixed). In many applications, y is the reward reaped with every attempt and the goal is to maximize reward over time. This naturally leads to *explore-exploit* type of algorithms and regret analysis for such algorithms. In some applications, *exploration* is expensive and it is desirable to completely move to the *exploitation* stage after a while. With that motivation, we study the convergence of the explore-exploit algorithms.

Global optimization of such functions is close to impossible without any further assumption on f. A common assumption is a Gaussian Process (GP) prior on the function f, which ensures a degree of smoothness of f. This assumption helps formulate algorithms such as GP-UCB and its variants for explore-exploit. Many such variants have been well studied Auer, Cesa-Bianchi and Fischer (2002); Garivier and Cappé (2011); Hernández-Lobato, Hoffman and Ghahramani (2014); Kaufmann, Cappé and Garivier (2012); Lai, Tze Leung and Robbins, Herbert (1985); Maillard et al. (2011). Some theoretical properties are also known for such algorithms (Srinivas et al., 2010, 2012). The main idea is to optimize an *acquisition function* to find the next point where we evaluate the function. Most analysis of such algorithms give an upper bound to the average *cumulative regret*,

$$(1) \qquad \frac{R_T}{T} = \frac{1}{T} \sum_{t=1}^{T} \left(f(\boldsymbol{x}^*) - f(\boldsymbol{x}_t) \right).$$

We focus on an approach known as Thompson Sampling (TS). Although this is an old idea dating back to Thompson (1933), there has been considerable attention in the recent past (Bijl et al., 2016; Granmo, 2010; May and Leslie, 2011). Several studies have shown good empirical evidence of TS (Chapelle and Li, 2011) and more recently, theoretical proofs have been obtained for the multi-arm bandit setting and some generalizations. Agrawal and Goyal (2012) showed for the first time that TS achieves logarithmic expected regret for the stochastic multi-armed bandit problem. The same authors (Agrawal and Goyal, 2013a) provided a near-optimal bound of $O(\sqrt{NT \log T})$ for expected regret of TS for the $N-$ armed bandit problem. Agrawal and Goyal (2013b) gave further results on contextual multi-armed bandits with linear payoffs. Analysis for the infinite armed bandit on a continuous space was missing, until Russo and Van Roy (2014) gave an overview of how to bound the regret by drawing an analogy between TS and Upper Confidence Bound (UCB) algorithms.

In this paper, we study a success metric which is different from regret or expected regret. All previous analyses of TS have tried to bound the average cumulative regret. However, in many practical situations, we would also like to know how fast $\boldsymbol{x}^t \to \boldsymbol{x}^*$ as the number of steps increases, and not just how fast the average regret decays. Our motivating example is the problem of ranking news items on a social network feed. A social network feed typically comprises of different types of items that come from different sources or systems. There is typically an algorithm (blender) that blends the different items together and tries to optimize for various business metrics like user engagement and revenue. The different systems are often updated independently and the blender has to adapt to find the optimal blending parameters after a system has been modified or improved. Systems like these typically handle a large throughput and work under strict latency constraints. It is advisable to turn off processes that are not absolutely necessary. In our case, we want to stop the exploration if and when the points $\boldsymbol{x}^t$ have converged to $\boldsymbol{x}^*$. In order to answer this question, we would need to know the rate of convergence for $\boldsymbol{x}^t \to \boldsymbol{x}^*$. The answer to this question is the main focus of this paper. To the best of our knowledge, this is the first result that proves the exponential rate of decay for an infinite-armed bandit where there is a dependency structure between the utility of each arm. The main result is stated in Theorem 1. Under some assumptions, we can show an exponential rate of convergence of $\boldsymbol{x}^t$ to $\boldsymbol{x}^*$. The main idea of the proof relies on breaking down the continuous domain into discrete regions and bounding the error on each discrete region, which can then be combined by the union bound.

The rest of the paper is organized as follows. In Section 2 we formally introduce the problem, the Thompson Sampling algorithm and the main result. We describe some preliminary results in Section 3 and prove the main result in Section 4. Simulation studies are shown in Section 5 which highlight the convergence without the explicit assumptions required for the proof. We discuss some generalizations and concluding remarks in Section 6. The proofs of all preliminary and supporting results in the given supplementary material.

2. Thompson Sampling Algorithm and The Main Result. We consider the problem of sequentially maximizing a black box function $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X}$ is a compact proper subset of $\mathbb{R}^d$. At every stage t we can sample $\boldsymbol{x}^t \in \mathcal{X}$ and observe y^t, where $y^t|\boldsymbol{x}^t \sim N(f(\boldsymbol{x}^t), \sigma^2)$.

2.1. *Gaussion Processes and Kernel Functions.* To solve the global optimization problem, we need sufficient smoothness assumptions on f. This is enforced by modeling f as a sample from a Gaussian Process (GP) with

mean 0 and kernel $k(\boldsymbol{x}, \boldsymbol{x}')$. For any $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, let $\boldsymbol{f}$ denote the vectorized version of the function values obtained at the n points. That is, $\boldsymbol{f} = (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n))^T$. Then, $\boldsymbol{f}$ is multivariate normal with mean 0 and covariance K, where $K_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. We assume that k is a Mercer kernel on the space $\mathcal{X}$ with respect to the uniform measure on $\mathcal{X}$. That is, we can write k as,

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{\infty} \lambda_i \psi_i(\boldsymbol{x}) \psi_i(\boldsymbol{y}),$$

where $(\lambda_i)_{i \in \mathbb{N}}$ is a sequence of non-negative, non-increasing numbers, which are summable and $(\psi_i)_{i \in \mathbb{N}}$ are a collection of mutually orthonormal functions with respect to the L^2 norm on $\mathcal{X}$. We can consider λ_i's to be the eigenvalues corresponding to the eigenfunctions ψ_i. A common example of a kernel is the Gaussian RBF-kernel, which can be parametrized by η, i.e.

$$k_\eta(\boldsymbol{x}, \boldsymbol{x}') = \zeta^2 \exp\left(-\frac{1}{2} \sum_{i=1}^{d} \frac{(\boldsymbol{x}_i - \boldsymbol{x}'_i)^2}{\ell_i^2}\right),$$

with $\eta = (\zeta, \ell_1, \ldots, \ell_d)$. See Minh, Niyogi and Yao (2006) for more details on Mercer's Theorem, kernel smoothing and many other examples.

2.2. *Thompson Sampling.* Suppose D_t denotes the data we have till iteration $t - 1$ and $\mathcal{F}_t$ denotes the posterior of the maximizer of f given D_t. The Thompson Sampling approach samples a new data point $\boldsymbol{x}_t$ at iteration t from $\mathcal{F}_t$. We observe the data $y^t = f(\boldsymbol{x}^t) + \epsilon^t$, where $\epsilon^t \sim N(0, \sigma^2)$ and update $D_{t+1} = D_t \cup \{(\boldsymbol{x}^t, y^t)\}$. We initialize the process by assuming a non-informative prior on the distribution of the maximizer, i.e., $\mathcal{F}_0 = U(\mathcal{X})$, the uniform distribution on $\mathcal{X}$. We stop the procedure when the variance of the distribution of $\mathcal{F}_t$ becomes considerably small and we return $\boldsymbol{x}^* = mode(\mathcal{F}_t)$ as the estimate of the global maximizer of f.

In some cases, especially when σ^2 is large, this process might converge to a local optimum. Since we sample $\boldsymbol{x}_t$ from $\mathcal{F}_t$, we might get stuck in one place and not explore the entire space. To ensure the convergence to the global maximum, we consider an ξ-greedy approach. That is, with some probability $\xi > 0$, we explore the entire region $\mathcal{X}$ uniformly at every stage t. Thus, we sample $\boldsymbol{x}^t \sim \mathcal{F}_t$ with probability $1 - \xi$ and we sample $\boldsymbol{x}^t \sim U(\mathcal{X})$ with probability ξ. We state the detailed steps in Algorithm 1.

2.3. *Estimation of Hyper-Parameters.* There are several methods known in literature for estimating the hyper parameters in this set up. We focus

on the maximum a posteriori (MAP) estimation. For other methods see Vanhatalo et al. (2012). Here we use,

$$\{\hat{\eta}_t, \hat{\sigma}_t\} = \operatorname*{argmax}_{\eta,\sigma} p(\eta, \sigma|D_t) = \operatorname*{argmin}_{\eta,\sigma} \left(-\log p(D_t|\eta, \sigma) - \log p(\eta, \sigma)\right),$$

where $p(\cdot)$ denotes the likelihood function. For the Gaussian RBF kernel we can write the marginal likelihood given the parameters, $p(D_t|\eta, \sigma) = \int p(\boldsymbol{y}|\boldsymbol{f}, \sigma)p(\boldsymbol{f}|\boldsymbol{x}, \eta)\mathrm{d}\boldsymbol{f}$ in a closed form,

$$\log p(D_t|\eta, \sigma) = C - \frac{1}{2}\log \left|K_\eta + \sigma^2 I\right| - \frac{1}{2}\boldsymbol{y}^T \left(K_\eta + \sigma^2 I\right)^{-1}\boldsymbol{y},$$

where $\boldsymbol{y}$ denotes the vectorized version of our observed function values. Since this function is easily differentiable, we can find the optimum using any gradient descent algorithm (Boyd and Vandenberghe, 2004). In situations where, a closed form expression cannot be found, we can resort to Laplace Approximations or EP's marginal likelihood approximation (Vanhatalo et al., 2012).

Note that the MAP estimator converges to the maximum likelihood estimator as we sample more and more points. Moreover, since the maximum likelihood estimator (MLE) is a consistent estimator, we assume that the regularity conditions hold such that $\eta_t \to \eta^*$ and $\sigma_t \to \sigma^*$ almost surely (Lehmann and Casella, 2006), where η^*, σ^* denotes the true optimal parameters.

REMARK. Although the usual result for consistency of the MLE only gives us convergence in probability, it is not hard to see that if we follow the proof in Lehmann and Casella (2006) we can get almost sure convergence under the extra condition that,

$$(2) \qquad \sup_{\theta \in \Theta} \left\|\hat{\ell}(\boldsymbol{x}|\theta) - \ell(\theta)\right\| \xrightarrow{\text{a.s.}} 0,$$

where $\ell, \hat{\ell}$ denotes the expected log-likelihood function and its estimate, and θ is the parameter of interest.

2.4. *Sampling from the Posterior Distribution of the maximizer.* We follow the approach in Section 2.1 of Hernández-Lobato, Hoffman and Ghahramani (2014) to sample from the distribution of the maximum given the data D_t. For sake of the proof of convergence, we choose a different feature map than what is used in Hernández-Lobato, Hoffman and Ghahramani (2014).

Given any Mercer kernel k_η, there exists a feature map $\phi(\boldsymbol{x})$ such that, $k_\eta(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x})^T \phi(\boldsymbol{x}')$ where

$$\phi(\boldsymbol{x}) = (\sqrt{\lambda_1}\psi_1(\boldsymbol{x}), \sqrt{\lambda_2}\psi_2(\boldsymbol{x}), \ldots)^T$$

Note that for any given η_t, we can identify the eigenvalue sequence $(\lambda_i^t)_{i\in\mathbb{N}}$. If it has infinite non-zero eigenvalues then we can pick a large enough truncation point m^*. Otherwise, if it has finitely many positive eigenvalues, pick m_t to be the largest integer such that $\lambda_{m_t}^t > 0$. That is,

$$(3) \qquad m_t = \min\left\{m^*, \max\left\{m \in \mathbb{N} | \lambda_m^t > 0\right\}\right\}$$

This enables us to approximate the kernel as

$$k_{\eta_t}(\boldsymbol{x}, \boldsymbol{x}') \approx \phi^t(\boldsymbol{x})^T \phi^t(\boldsymbol{x}')$$

where

$$(4) \qquad \phi^t(\boldsymbol{x}) = \left(\sqrt{\lambda_1^t}\psi_1^t(\boldsymbol{x}), \sqrt{\lambda_2^t}\psi_2^t(\boldsymbol{x}), \ldots, \sqrt{\lambda_{m_t}^t}\psi_{m_t}^t(\boldsymbol{x})\right)^T.$$

Since f is modeled as a sample from a Gaussian process, we can write $f(\cdot) = \phi(\cdot)^T\boldsymbol{\theta}$, where $\boldsymbol{\theta} \sim N(0, \mathbf{I})$. Thus, to draw a sample from $\mathcal{F}_t$, we draw a random function $f^t(\cdot) = \phi^t(\cdot)^T\boldsymbol{\theta}^t$, where ϕ^t is given by (4) and $\boldsymbol{\theta}^t$ is a random vector drawn from the posterior distribution of $\boldsymbol{\theta}|(D_t, \phi^t)$, i.e.

$$(5) \qquad \boldsymbol{\theta}^t \sim \boldsymbol{\theta}|(D_t, \phi^t) = N\left(\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{y}, \sigma_t^2\boldsymbol{A}^{-1}\right)$$

where $\boldsymbol{A} = \boldsymbol{\Phi}^T\boldsymbol{\Phi} + \sigma_t^2\boldsymbol{I}$ and

$$(6) \qquad \boldsymbol{\Phi}^T = [\phi^t(\boldsymbol{x}^0), \ldots, \phi^t(\boldsymbol{x}^{t-1})].$$

This $f^t(\cdot)$ is an approximation to the true f after observing the data D_t. Thus, we get $\operatorname{argmax}_{\boldsymbol{x}\in\mathcal{X}} \phi^t(\boldsymbol{x})^T\boldsymbol{\theta}^t \sim \mathcal{F}_t$.

REMARK. In theory, we can start with a kernel which has a known orthonormal expansion in which case, it is easy to identify these functions ϕ^t. Sometimes in practice, if we do not know the exact form of these orthonormal functions, we can use the spectral decomposition of the matrix $K_{t\times t} = \boldsymbol{\Phi}\boldsymbol{\Phi}^T$. There are many algorithms readily available to identify the top m_t eigenvalues and eigenvectors of a symmetric matrix. We refer the reader to Parlett (1998) and Yuan and Zhang (2013) for more details.

Algorithm 1 Thompson Sampling for Infinite-Armed Bandits

1: Input : Function f, Kernel k_η, Domain $\mathcal{X}$, Parameter ξ
2: Output : $\boldsymbol{x}^*$, the global maximum of f
3: Sample $\boldsymbol{x}^0$ uniformly from $\mathcal{X}$
4: Observe $y^0 = f(\boldsymbol{x}^0) + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$.
5: Put $D_1 = \{(\boldsymbol{x}^0, y^0)\}$
6: **for** $t = 1, 2, \ldots$ **do**
7:     Estimate the hyper-parameters η_t, σ_t using D_t as given in Section 2.3.
8:     Sample a random function ϕ according to (4) corresponding to k_{η_t}
9:     Sample $\boldsymbol{\theta}^t$ from $\boldsymbol{\theta}|(D_t, \phi)$ according to (5)
10:     $\boldsymbol{x}^t = \begin{cases} \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}} \phi(\boldsymbol{x})^T \boldsymbol{\theta}^t & \text{w. p. } 1 - \xi \\ U(\mathcal{X}) & \text{w. p. } \xi \end{cases}$
11:     Observe $y^t = f(\boldsymbol{x}^t) + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$.
12:     Set $D_{t+1} = D_t \cup \{(\boldsymbol{x}^t, y^t)\}$
13:     Break from the loop when $\boldsymbol{x}^t$ chosen as the maximizer of $\phi(\boldsymbol{x})^T \boldsymbol{\theta}^t$ converges to $\boldsymbol{x}^*$.
14: **end for**
15: Return $\boldsymbol{x}^*$

2.5. *Main Result.* The main aim of the paper is to prove the exponential rate of convergence for the Thompson Sampling approach from Algorithm 1. We need some further assumptions and regularity conditions to achieve that.

ASSUMPTION 1. *Let ϕ^t be the feature map for k_{η_t}. Then, there exists a $\boldsymbol{\theta}_t^*$ with $\|\boldsymbol{\theta}_t^*\| \leq \sqrt{t}M$ and $\delta_0(t) \leq c^*/\sqrt{t}$ such that,*

$$\lim_{t \to \infty} \frac{1}{\delta_0(t)} \sup_{\boldsymbol{x} \in \mathcal{X}} \left| f(x) - \phi^t(\boldsymbol{x})^T \boldsymbol{\theta}_t^* \right| \leq 1$$

almost surely. Here M and c^* are positive constants.

The above assumption says that outside of a measure zero set, for large enough t we have, $\sup_{\boldsymbol{x} \in \mathcal{X}} \left| f(x) - \phi^t(\boldsymbol{x})^T \boldsymbol{\theta}_t^* \right| < \delta_0(t)$. It gives a bound on the difference of f and the projection of f on the column space of the feature map from the kernel. If the kernel hyperparameters η can be learned well, then this assumption intuitively holds.

ASSUMPTION 2. *The kernel must belong to either of the following two classes.*

(a) *Bounded eigen functions. For this class of kernels, there exists an M such that*

$$|\psi_i(\boldsymbol{x})| \leq M \quad \text{for all } i.$$

An example of this is when ψ_i form a sine basis on $\mathcal{X} = [0, 2\pi]$ (Braun, 2006).

(b) *Bounded kernel functions. For this class of kernels, there exists an M such that*

$$k(\boldsymbol{x}, \boldsymbol{x}) \leq M < \infty \quad \text{for all } \boldsymbol{x} \in \mathcal{X}$$

A very typical example in this class is the RBF kernel, or the squared exponential kernel. All shift-invariant kernels fall in this category.

ASSUMPTION 3. *All eigenfunctions are Lipschitz continuous. That is, there exists a B such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$,*

$$|\psi_i(\boldsymbol{x}) - \psi_i(\boldsymbol{y})| \leq B\|\boldsymbol{x} - \boldsymbol{y}\| \text{ for all } i.$$

Most common kernels usually have continuous eigenfunctions with bounded derivatives and hence are Lipschitz. For a thorough list and more examples see Minh, Niyogi and Yao (2006).

With these assumptions we can now state our main result.

THEOREM 1. *Let f be a smooth continuous function on a compact set $\mathcal{X}$ having a global unique maximum at $\boldsymbol{x}^*$. Then, under Assumptions 1 - 3, for any $\epsilon > 0$, if we follow the Thompson Sampling procedure as given in Algorithm 1, there exists a T such that for all $t > T$,*

$$P(\|\boldsymbol{x}^t - \boldsymbol{x}^*\| > \epsilon) \leq C_\epsilon \exp(-c_\epsilon t),$$

where $\boldsymbol{x}^t$ is the maximizer at step t, and C_ϵ, c_ϵ are positive constants.

2.6. *Remarks on Theorem 1.* Theorem 1 gives us an explicit rate of convergence of $\boldsymbol{x}^t$ to $\boldsymbol{x}^*$. However, it does not characterize how fast we realize this rate in practice since it holds for large enough t. The function structure plays an important role in determining this time. We show a simple simulation example in Section 5 to see this decay rate as a function degenerates into a flat function.

3. Preliminaries. We now state some preliminary results, which will be used throughout the rest of the proof. For the rest of the paper, we denote the changing constant as c. Also, throughout the paper we make statements for Algorithm 1 under the Assumptions 1 - 3 without explicitly stating it every time. The first result gives a bound on the minimum and maximum eigenvalues of the matrix $\boldsymbol{A}/t$.

LEMMA 1. *Let* $\boldsymbol{A} = \boldsymbol{\Phi}^T\boldsymbol{\Phi} + \sigma_t^2\boldsymbol{I}$, *where* $\boldsymbol{\Phi}$ *is defined* (6). *Then,*

$$\lim_{t\to\infty} \lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right) \geq \xi c > 0 \quad a.\ s.,$$

and

$$\lim_{t\to\infty} \lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right) \leq \xi C + \alpha(1-\xi) + 1 < \infty \quad a.\ s.,$$

where $\alpha = k_\eta(\boldsymbol{x}, \boldsymbol{x})$ *and* C, c *are constants.*

REMARK. Note that, Lemma 1 and Assumption 1 make almost sure statements. If we denote our probability space by $(\Omega, \mathcal{F}, \mathcal{P})$, then, throughout the rest of this paper, we only work over those set of $\omega \in \Omega$ where the statements in Lemma 1 and Assumption 1 hold.

As a corollary to Lemma 1 we can show upper bounds to much more complicated matrix forms involving $\boldsymbol{A}$. Two such results, which will be used later are as follows.

LEMMA 2. *There exists a constant* $c > 0$ *such that for large enough* t,

$$\lambda_{\max}\left(\left(\frac{\boldsymbol{A}}{t}\right)^{-1} \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{t} \left(\frac{\boldsymbol{A}}{t}\right)^{-1}\right) \leq c.$$

LEMMA 3. *The following bounds hold:*

(a) $\left\|\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi} - \boldsymbol{I}\right\| \leq c/t$
(b) $\left\|\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\right\| \leq C/\sqrt{t}$

where c, C *are constants and* $\|\cdot\|$ *denotes the spectral norm of the matrix.*

Before we introduce the next result, let us introduce some more notation that we use throughout the proofs. Let us define δ_ϵ as the minimum difference in the function values between the optimal and any $\boldsymbol{x}$ which is at least ϵ distance away from the optimal. Formally,

(7) $$\delta_\epsilon := \inf_{\boldsymbol{x}:\|\boldsymbol{x}-\boldsymbol{x}^*\|>\epsilon} f(\boldsymbol{x}^*) - f(\boldsymbol{x}).$$

We know that $\delta_\epsilon > 0$ since f has an unique maximum.

The following result quantifies that if f can be approximated well, then there is a positive difference between $f(\boldsymbol{x}^*)$ and $f(\boldsymbol{x})$ for any $\boldsymbol{x}$ which is at least ϵ distance away from the optimal $\boldsymbol{x}^*$. Formally, we show the following.

LEMMA 4.   *Given $\epsilon > 0$, for any $\boldsymbol{x}$ such that $\|\boldsymbol{x} - \boldsymbol{x}^*\| > \epsilon$ and large enough t,*

$$\left(\phi^t(\boldsymbol{x}^*) - \phi^t(\boldsymbol{x})\right)^T \boldsymbol{A}^{-1}\boldsymbol{\Phi}^T \boldsymbol{f} \geq \frac{\delta_\epsilon}{2} > 0.$$

Our last preliminary result shows a concentration bound on Chi-square random variables, which will be needed for subsequent proofs.

LEMMA 5.   *Let $Z \sim \chi_m^2$. Then for any $\delta > 0$,*

$$P(Z > m + \delta) \leq \exp\left(-\frac{1}{2}\left(\delta + m + \sqrt{2\delta m + m^2}\right)\right).$$

We set one last notation that we use in the proofs below. By Lemma 1 and Lemma 2, for large enough t we have,

(8)
$$\lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right) \geq c_1$$
$$\lambda_{\max}\left(\left(\frac{\boldsymbol{A}}{t}\right)^{-1}\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{t}\left(\frac{\boldsymbol{A}}{t}\right)^{-1}\right) \leq c_2$$

where c_1 and c_2 are constants. Furthermore, using Lemma 3 and Assumption 1 we can write,

$$\begin{aligned}
\|\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{f}\| &= \|\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\theta}^* - \delta_0(t)\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{1}\| \\
&= \|\boldsymbol{\theta}^* + \boldsymbol{E}\boldsymbol{\theta}^* - \delta_0(t)\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{1}\| \\
&\leq \|\boldsymbol{\theta}^*\| + \|\boldsymbol{E}\boldsymbol{\theta}^*\| + \|\delta_0(t)\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{1}\| \\
&\leq \|\boldsymbol{\theta}^*\|(1 + \|\boldsymbol{E}\|) + \delta_0(t)\sqrt{t}\|\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\| \\
(9)\qquad &\leq \sqrt{c_3 t}
\end{aligned}$$

where $\boldsymbol{E} = \boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi} - \boldsymbol{I}$.

4. Proof of Theorem 1.

4.1. *Outline.* The main idea of the proof relies on breaking down the continuous domain into small regions and bounding the errors on each region, which are then combined using the union bound. To bound the error on each small region, we compare the function values at a single point within the region and bound the error appropriately. In order to first compare the function values at a single point, we rely on the following Lemma.

LEMMA 6. *For any $\boldsymbol{x}$ such that $\|\boldsymbol{x} - \boldsymbol{x}^*\| > \epsilon$, any $0 < \epsilon' \leq \delta_\epsilon/4$, for t large enough*

$$P\left(\phi^t(\boldsymbol{x}^*)^T \boldsymbol{\theta}^t < \phi^t(\boldsymbol{x})^T \boldsymbol{\theta}^t + \epsilon'\right) \leq 2\exp(-c_\epsilon t)$$

where δ_ϵ is defined as in (7) and c_ϵ is positive constants that depends on ϵ.

Once we have this control, in order to bound the supremum in the sub-region, we a use a truncation argument. We truncate on $\|\boldsymbol{\theta}^t\|$ and bound the probability of $\|\boldsymbol{\theta}^t\|$ exceeding the truncation value. This is done, using the following result.

LEMMA 7. *For any $\delta > 0$ and t large enough, let*

$$L_t := \sqrt{\sigma_t^2 \left(\frac{m_t}{t} + \delta\right)\left(\frac{1}{c_1} + \frac{1}{c_2}\right) + c_3 t},$$

where c_1, c_2, c_3 are defined in (8) and (9). Then,

$$P\left(\|\boldsymbol{\theta}^t\| > L_t\right) \leq 2\exp\left(-\frac{\delta t}{2}\right).$$

Finally, we show that we can appropriately choose the number of discrete sub-regions such that the union bound converges, which will be enough for the proof. The details are now given below.

4.2. *Proof.* We begin by observing that for any $\epsilon > 0$,

$$P(\|\boldsymbol{x}^t - \boldsymbol{x}^*\| > \epsilon) = P\left(\sup_{\boldsymbol{x} \in \mathcal{B}_\epsilon(\boldsymbol{x}^*)} \phi^t(\boldsymbol{x})^T \boldsymbol{\theta}^t < \sup_{\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{B}_\epsilon(\boldsymbol{x}^*)} \phi^t(\boldsymbol{x})^T \boldsymbol{\theta}^t\right)$$

$$\leq P\left(\phi^t(\boldsymbol{x}^*)^T \boldsymbol{\theta}^t < \sup_{\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{B}_\epsilon(\boldsymbol{x}^*)} \phi^t(\boldsymbol{x})^T \boldsymbol{\theta}^t\right),$$

where $\mathcal{B}_\epsilon(\boldsymbol{x}^*)$ denotes a ball of radius ϵ around $\boldsymbol{x}^*$. Now since $\mathcal{X} \setminus \mathcal{B}_\epsilon(\boldsymbol{x}^*)$ is a compact set, and a metric space with respect to the Euclidean norm, we can cover it with an ϵ-Net Vershynin (2010). Specifically, for any $\epsilon_t > 0$, there exists a subset $\mathcal{N}_{\epsilon_t}$ of $\mathcal{X} \setminus \mathcal{B}_\epsilon(\boldsymbol{x}^*)$ such that given any $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{B}_\epsilon(\boldsymbol{x}^*)$, there exists a $\boldsymbol{y} \in \mathcal{N}_{\epsilon_t}$ such that $\|\boldsymbol{x} - \boldsymbol{y}\| < \epsilon_t$. Moreover, it is known that

$$|\mathcal{N}_{\epsilon_t}| \leq \left(\frac{3}{\epsilon_t}\right)^d.$$

For simplicity, let $\{\boldsymbol{x}_i\}_{i=1}^{\mathcal{N}_{\epsilon_t}}$ denotes the set of points in the ϵ_t-Net. Thus, we can write,

$$P(\|\boldsymbol{x}^t - \boldsymbol{x}^*\| > \epsilon) \leq P\left(\phi^t(\boldsymbol{x}^*)^T\boldsymbol{\theta}^t < \max_{i=1,\ldots,\mathcal{N}_{\epsilon_t}} \sup_{\boldsymbol{x} \in \mathcal{B}_{\epsilon_t}(\boldsymbol{x}_i)} \phi^t(\boldsymbol{x})^T\boldsymbol{\theta}^t\right)$$

Now for any $\boldsymbol{x} \in \mathcal{B}_{\epsilon_t}(\boldsymbol{x}_i)$, we have

$$\left\|\phi^t(\boldsymbol{x}) - \phi^t(\boldsymbol{x}_i)\right\|^2 = \sum_{j=1}^{m_t} \lambda_j^t \left[\psi_j^t(\boldsymbol{x})^2 + \psi_j^t(\boldsymbol{x}_i)^2 - 2\psi_j^t(\boldsymbol{x})\psi_j^t(\boldsymbol{x}_i)\right]$$

$$= \sum_{j=1}^{m_t} \lambda_j^t \left(\psi_j^t(\boldsymbol{x}) - \psi_j^t(\boldsymbol{x}_i)\right)^2$$

$$\leq \sum_{j=1}^{m_t} \lambda_j^t B^2 \|\boldsymbol{x} - \boldsymbol{x}_i\|^2 \leq C^2 \epsilon_t^2$$

where first inequality follows using the Lipschitz condition for ψ_j^t and the last inequality follows since m_t is finite and each λ_j^t is positive. Thus, for any $\delta > 0$, from Lemma 7 choosing

$$L_t = \sqrt{\sigma_t^2 \left(\frac{m_t}{t} + \delta\right)\left(\frac{1}{c_1} + \frac{1}{c_2}\right) + c_3 t} = O\left(\sqrt{t}\right)$$

and letting $\|\boldsymbol{\theta}^t\| \leq L_t$, we get,

$$\left\|(\phi^t(\boldsymbol{x}) - \phi^t(\boldsymbol{x}_i))^T\boldsymbol{\theta}^t\right\| \leq \|\boldsymbol{\theta}^t\| \left\|\phi^t(\boldsymbol{x}) - \phi^t(\boldsymbol{x}_i)\right\| \leq CL_t\epsilon_t.$$

Thus conditioning on $\|\boldsymbol{\theta}^t\| \leq L_t$, we get,

$$\sup_{\boldsymbol{x} \in \mathcal{B}_{\epsilon_t}(\boldsymbol{x}_i)} \phi^t(\boldsymbol{x})^T\boldsymbol{\theta}^t \leq \phi^t(\boldsymbol{x}_i)^T\boldsymbol{\theta}^t + CL_t\epsilon_t.$$

Using this we get,

$$P(\left\|\boldsymbol{x}^t - \boldsymbol{x}^*\right\| > \epsilon)$$

$$\leq P\left(\phi^t(\boldsymbol{x}^*)^T\boldsymbol{\theta}^t < \max_{i=1,\ldots,\mathcal{N}_{\epsilon_t}} \phi^t(\boldsymbol{x}_i)^T\boldsymbol{\theta}^t + CL_t\epsilon_t\right) + P(\|\boldsymbol{\theta}^t\| > L_t)$$

$$\leq \sum_{i=1}^{\mathcal{N}_{\epsilon_t}} P\left(\phi^t(\boldsymbol{x}^*)^T\boldsymbol{\theta}^t < \phi^t(\boldsymbol{x}_i)^T\boldsymbol{\theta}^t + CL_t\epsilon_t\right) + P(\|\boldsymbol{\theta}^t\| > L_t)$$

$$\leq 2\left|\mathcal{N}_{\epsilon_t}\right| \exp(-c_\epsilon t) + 2\exp\left(-\frac{\delta t}{2}\right)$$

where the first inequality follows by conditioning on $\left\|\boldsymbol{\theta}^t\right\| \leq L_t$, the second from the union bound and the last inequality by using Lemmas 6 and 7.

To satisfy Lemma 6 we choose $\epsilon_t = \delta_\epsilon/16CL_t$. This gives us,

$$|\mathcal{N}_{\epsilon_t}| \leq \left(\frac{3}{\epsilon_t}\right)^d = \left(\frac{48CL_t}{\delta_\epsilon}\right)^d = O\left(t^{\frac{d}{2}}\right).$$

Now, since d is finite, $|\mathcal{N}_{\epsilon_t}|$ grows to infinity much slower than exponential. Hence, we get for some positive constants C, c and t large enough,

$$P(\left\|\boldsymbol{x}^t - \boldsymbol{x}^*\right\| > \epsilon) \leq C \exp(-ct) + 2 \exp\left(-\frac{\delta t}{2}\right).$$

Finally, since the choice of δ was arbitrary, we have for constants C_ϵ, c_ϵ,

$$P(\left\|\boldsymbol{x}^t - \boldsymbol{x}^*\right\| > \epsilon) \leq C_\epsilon \exp(-c_\epsilon t)$$

This concludes the proof of Theorem 1.

5. Simulation Study. We have proved an exponential rate of convergence for the ξ-greedy Thompson Sampling algorithm, under some assumptions. Since Assumption 1 is hard to verify, we show a simulation study on a univariate and bivariate function to empirically validate that it usually holds in practice. Moreover, we present a different simulation study to capture how the structure of the function affects its rate of convergence as explained in Section 2.6.

5.1. *Assumption Validation.* In order to empirically validate Assumption 1, we present two different examples. Throughout this study, we have considered the kernel to be the RBF-kernel, which satisfies all the regularity conditions. Moreover, we have taken a batch approach, where at each stage instead of drawing a single $\boldsymbol{x}^t$, we draw 30 points from the distribution of the maximum. This is done to increase the exploration part of the algorithm while keeping the running time constant.

For the 1-dimensional example, we consider a bimodal function which is a mixture of two Gaussian probability density functions. Specifically, we choose,

$$f^1(x) = \frac{5}{\sqrt{2\pi}} \exp\left(-\frac{(x-2)^2}{2}\right) + \frac{10}{\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2}\right),$$

which has a local maximum at 2 and a global maximum at 5. We consider a similar function in a 2-dimensional space as well. In particular we pick,

$$f^2(x) = \frac{5}{2\pi} \exp\left(-\frac{1}{2}\left\|\boldsymbol{x} - \mu_1\right\|^2\right) + \frac{10}{2\pi} \exp\left(-\frac{1}{2}\left\|\boldsymbol{x} - \mu_2\right\|^2\right),$$

where $\mu_1 = (2, 2)$ is the local maximum and $\mu_2 = (5, 5)$ is the global maximum. Since in practice we usually have a high variance in the observations, we add Gaussian random errors with $\sigma \in [0.1, 5]$ to both the functions while observation the data. Figure 1 shows the univariate function, along with the 95% confidence bands for the sampling error. It also shows a sample of points obtained when using different standard deviations in the error generation mechanism.
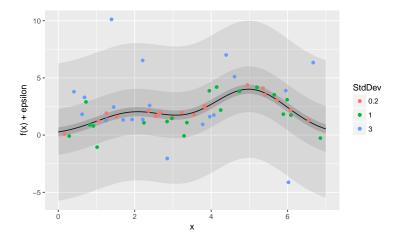

Fig 1: $f^1(x)$ with some observed realizations.

Note that, as the error increases, it becomes increasingly hard to identify the true function. Although here we have only shown three different values of σ, we work with a wider range and the results are shown in the subsequent figures. Figure 2 shows the bivariate function $f^2(\boldsymbol{x})$. For simplicity, we do not add the confidence bands and sample points in the figure. However, here too we work with a wide range of the error.

Our setup does not artificially enforce the assumption on the function, but we still observe an exponential rate of convergence. Figure 3 and 4 show the decays in the relative squared error for each of the different standard deviations. Specifically, we plot $2 \log_{10}((\boldsymbol{x}_t - \boldsymbol{x}^*)/\boldsymbol{x}^*)$ vs iteration t. For smaller σ, we see very quick convergence for both of the example functions. As the errors increase we see that the algorithm takes a longer time to converge, but it is still exponential in t. This phenomenon is common to both the one and the two-dimensional example. The sudden spikes in the error plot are because of the iterations where we do random sampling instead of sampling from the maximum. Moreover, we notice that in general, the number of iterations required to converge for a 2-dimensional function is larger than
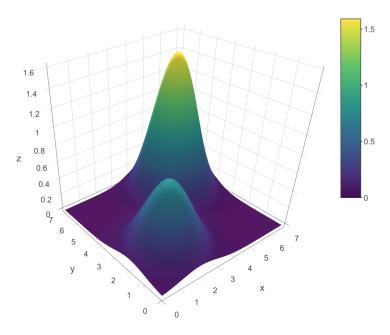
Fig 2: 2-dimentional function $f^2(\boldsymbol{x})$

that for a 1-dimensional case, especially when the errors increase in the observations.

To get a better idea of the algorithm in practice, we show Figure 5. It captures the distribution of the maximum across different iterations when we have $\sigma = 5$. We see that as the iterations increase, we inch closer and closer to the Dirac delta measure at $\boldsymbol{x}^*$. At each stage, sampling from the probability measure brings us closer to the true maximum while allowing some room to explore.

These examples show that without explicit assumptions on the function we converge to the true global maximum even when there is a large level of noise in the data. Thereby strengthening the heuristic that we only need Assumption 1 for the rigorous proof.

5.2. *Function Structure.*   Although we have seen an exponential rate of decay, it is a difficult problem to characterize how quickly we start to see that rate of decay, since it depends on the structure of the underlying function. To observe how the decay rate actually changes in practice, we consider the simple example of the following one dimensional function, where β acts as
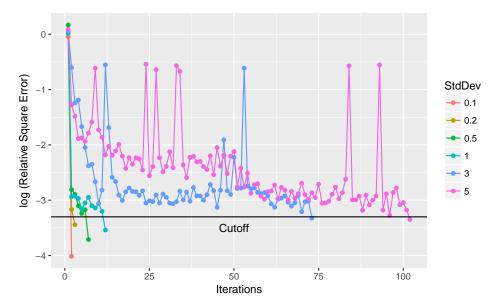
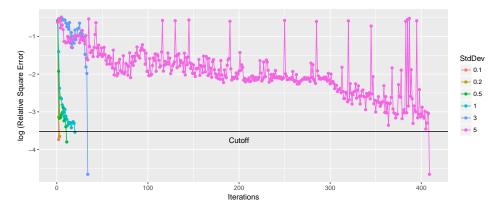Fig 3: Decay rate for different σ for function $f^1(x)$

Fig 4: Decay rate for different σ for function $f^2(x)$

a scale weight.

$$f_\beta(x) = \beta \exp\left(-\frac{(x-5)^2}{8}\right)$$

This is an unimodal function with a global maximum at 5 and as we decrease the value of β the function becomes more and more flat. Figure 6 shows this behavior across different value of β. While trying to estimate the maximum
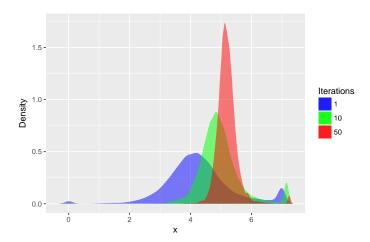
Fig 5: Sequence of estimated distributions of the maximum across iterations for $\sigma = 5$ when trying to estimate the maximum of $f^1(\boldsymbol{x})$. A similar figure is seen for the bivariate function, $f^2(\boldsymbol{x})$ as well.
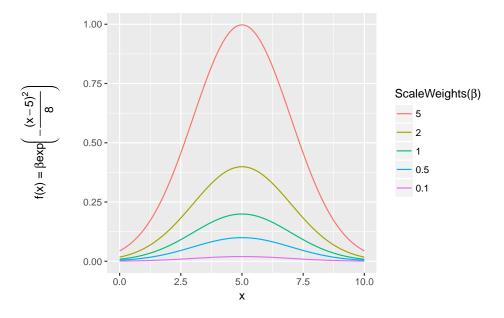

Fig 6: The change in function $f_\beta(x)$ as β decreases.

of each of the above functions, we keep a constant error rate of $\sigma = 0.1$. In this setup, we observe the decay rate as shown in Figure 7. Given, a fixed error rate σ, the rate slows down as the function becomes more and more
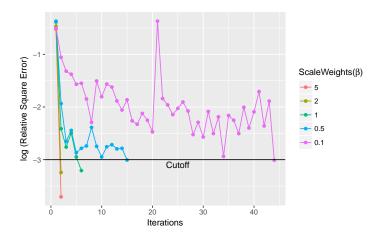
flat.

Fig 7: Error decay for function $f_\beta(x)$ as β decreases.

6. Conclusion. We have proved an exponential rate of convergence for the ξ-Greedy Thompson Sampling Algorithm in the case of an infinite armed bandit with a Gaussian Process prior on the reward. As far as we know, this is the first result which quantifies the rate of decay of the sequential point $\boldsymbol{x}^t$ to the true optimal $\boldsymbol{x}^*$. Although we have proved the result where at every stage we are sampling only one point, it should be easy to generalize to more points and we leave it as a future work. While actually running the algorithm in practice, we can repeat steps 8-11 of Algorithm 1 to sample more points at every stage as done throughout our simulations in the Section 5. By doing so, we can explore the function better in parallel within the same running time as a single point evaluation. The simulation study under this regime shows quick convergence as well as the fact that the assumptions are not too restrictive. This novel proof technique that we have presented here, can be now used to solve a variety of problems and opens up a new direction of research. Using this technique explicit convergence rates can be shown for most of the UCB type algorithms. We leave such generalizations as future work.

Acknowledgment. We would like to thank Deepak Agarwal, Liang Zhang, Yang Yang, Ying Xuan and Rajarshi Mukherjee for the several fruitful discussions regarding this paper.

References.

AGRAWAL, S. and GOYAL, N. (2012). Analysis of Thompson Sampling for the Multi-Armed Bandit Problem. In *COLT*.

AGRAWAL, S. and GOYAL, N. (2013a). Further Optimal Regret Bounds for Thompson Sampling. In *AISTATS* 99–107.

AGRAWAL, S. and GOYAL, N. (2013b). Thompson Sampling for Contextual Bandits with Linear Payoffs. In *ICML* 127–135.

AUER, P., CESA-BIANCHI, N. and FISCHER, P. (2002). Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning* **47** 235–256.

BHATIA, R. (2013). *Matrix analysis* **169**. Springer Science & Business Media.

BIJL, H., SCHÖN, T. B., VAN WINGERDEN, J.-W. and VERHAEGEN, M. (2016). A sequential Monte Carlo approach to Thompson sampling for Bayesian optimization. *arXiv:1604.00169.*

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization.* Cambridge University Press.

BRAUN, M. L. (2006). Accurate error bounds for the eigenvalues of the kernel matrix. *Journal of Machine Learning Research* **7** 2303–2328.

CHAPELLE, O. and LI, L. (2011). An empirical evaluation of Thompson Sampling. In *NIPS* 2249–2257.

ERIC, B., FREITAS, N. D. and GHOSH, A. (2008). Active preference learning with discrete choice data. In *NIPS* 409–416.

GARIVIER, A. and CAPPÉ, O. (2011). The KL-UCB Algorithm for Bounded Stochastic Bandits and Beyond. In *COLT* 359–376.

GRANMO, O.-C. (2010). Solving two-armed Bernoulli bandit problems using a Bayesian learning automaton. *International Journal of Intelligent Computing and Cybernetics* **3** 207–234.

HERNÁNDEZ-LOBATO, J. M., HOFFMAN, M. W. and GHAHRAMANI, Z. (2014). Predictive entropy search for efficient global optimization of black-box functions. In *NIPS* 918–926.

KAUFMANN, E., CAPPÉ, O. and GARIVIER, A. (2012). On Bayesian Upper Confidence Bounds for Bandit Problems. In *AISTATS* 592–600.

LAI, TZE LEUNG and ROBBINS, HERBERT (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics* **6** 4–22.

LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics* 1302–1338.

LEHMANN, E. L. and CASELLA, G. (2006). *Theory of point estimation.* Springer Science & Business Media.

LIZOTTE, D. J., WANG, T., BOWLING, M. H. and SCHUURMANS, D. (2007). Automatic Gait Optimization with Gaussian Process Regression. In *IJCAI* **7** 944–949.

MAILLARD, O.-A., MUNOS, R., STOLTZ, G. et al. (2011). A Finite-Time Analysis of Multi-armed Bandits Problems with Kullback-Leibler Divergences. In *COLT*.

MAY, B. C. and LESLIE, D. S. (2011). Simulation studies in optimistic Bayesian sampling in contextual-bandit problems. *Technical Report 11:02. Statistics Group, Department of Mathematics, University of Bristol.*

MINH, H. Q., NIYOGI, P. and YAO, Y. (2006). Mercers Theorem, Feature Maps, and Smoothing. In *International Conference on Computational Learning Theory* 154–168. Springer.

NEGOESCU, D. M., FRAZIER, P. I. and POWELL, W. B. (2011). The knowledge-gradient algorithm for sequencing experiments in drug discovery. *INFORMS Journal on Computing* **23** 346–363.

PARLETT, B. N. (1998). *The symmetric eigenvalue problem.* SIAM.

Russo, D. and Van Roy, B. (2014). Learning to optimize via posterior sampling. *Mathematics of Operations Research* **39** 1221–1243.

Snoek, J., Larochelle, H. and Adams, R. P. (2012). Practical Bayesian optimization of Machine Learning algorithms. In *NIPS* 2951–2959.

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

Srinivas, N., Krause, A., Kakade, S. M. and Seeger, M. W. (2012). Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory* **58** 3250–3265.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25** 285–294.

Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jyländki, P., Tolvanen, V. and Vehtari, A. (2012). Bayesian modeling with Gaussian processes using the GPstuff toolbox. *arXiv:1206.5754*.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv:1011.3027*.

Wang, Z., Mohamed, S. and de Freitas, N. (2013). Adaptive Hamiltonian and Riemann Monte Carlo samplers. In *ICML*.

Yuan, X.-T. and Zhang, T. (2013). Truncated power method for sparse eigenvalue problems. *Journal of Machine Learning Research* **14** 899–925.

7. Appendix. We collect the proofs of all the preliminary and supporting Lemmas here.

Proof of Lemma 1. We begin with the lower bound. Observe that,

$$
\lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right) = \lambda_{\min}\left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{t} + \frac{\sigma_t^2\boldsymbol{I}}{t}\right)
$$
$$
\geq \lambda_{\min}\left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{t}\right) = \lambda_{\min}\left(\frac{1}{t}\sum_{i=0}^{t-1}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T\right)
$$

We separate out the sum into cases where $\boldsymbol{x}^i \sim U(\mathcal{X})$ and where $\boldsymbol{x}^i$ is generated as the maximizer. Using a simple ordering of the $\boldsymbol{x}$ and appropriate change of notation we have,

$$
\lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right) \geq \lambda_{\min}\left(\frac{1}{t}\sum_{i=1}^{t\xi}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T + \frac{1}{t}\sum_{i=t\xi+1}^{t}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T\right)
$$

$$
(10) \qquad\qquad \geq \xi\lambda_{\min}\left(\frac{1}{t\xi}\sum_{i=1}^{t\xi}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T\right) = \xi\lambda_{(m_t)}\left(K_{t\xi}^{m_t}\right)
$$

where $\lambda_{(m_t)}$ denotes the m_t-th largest eigenvalue and $K_{t\xi}^{m_t}$ is the $t\xi \times t\xi$ matrix, whose (i,j) entry is the m_t level approximation of $\hat{k}_{\eta_t}$. That is,

$$\left[K_{t\xi}^{m_t}\right]_{i,j} = \frac{1}{t\xi} k_{\eta_t}^{m_t}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{1}{t\xi} \sum_{\ell=1}^{m_t} \lambda_\ell^t \psi_\ell^t(\boldsymbol{x}_i) \psi_\ell^t(\boldsymbol{x}_j).$$

Now note that using the finite sample error bounds from Braun (2006) we have,

$$\lambda_{(m_t)}\left(K_{t\xi}^{m_t}\right) \geq \lambda_{m_t}^t \left(1 - C(t\xi, m_t)\right).$$

Now from the results in Braun (2006) if the kernel has bounded eigenfunctions, i.e., $|\psi_i(\boldsymbol{x})| \leq M$, we have probability larger than $1 - \epsilon_1$,

$$C(t\xi, m_t) < M^2 m_t \sqrt{\frac{2}{t\xi} \log \frac{m_t(m_t + 1)}{\epsilon_1}}.$$

On the other hand, if the kernel is bounded, i.e. $k(\boldsymbol{x}, \boldsymbol{x}) \leq M$ then with probability larger than $1 - \epsilon_1$

$$C(t\xi, m_t) < m_t \sqrt{\frac{2M}{t\xi \lambda_{m_t}^t} \log \frac{2m_t(m_t + 1)}{\epsilon_1}} + \frac{4M}{3t\xi \lambda_{m_t}^t} \log \frac{2m_t(m_t + 1)}{\epsilon_1}$$

Choosing $\epsilon_1 = \epsilon^*/t^{1+\delta}$ for some $\delta > 0$, we see that in both cases,

$$C(t\xi, m_t) = O\left(\sqrt{\frac{\log t}{t}}\right)$$

Thus, if we denote an event as,

$$E_t = \left\{\lambda_{(m_t)}\left(K_{t\xi}^{m_t}\right) \geq \lambda_{m_t}^t \left(1 - c\sqrt{\frac{\log t}{t}}\right)\right\}$$

Then, $P(E_t^c) \leq \epsilon^*/t^{1+\delta}$ and hence, $\sum_{t=1}^{\infty} P(E_t^c) < \infty$. Thus, by the Borel Cantelli Lemma, $P(E_t^c$ occurs infinitely often $) = 0$. Thus, outside a set of measure zero, for any $\omega \in \Omega$, and exists a $t(\omega)$ such that for all $t > t(\omega)$,

$$\lambda_{(m_t)}\left(K_{t\xi}^{m_t}\right) \geq \lambda_{m_t}^t \left(1 - c\sqrt{\frac{\log t}{t}}\right).$$

Now for large enough t, $\lambda_{m_t}^t \to \lambda_{m^*} > 0$, where λ_{m^*} is the m^* largest eigenvalue of the optimal kernel k_{η^*}. If, k_{η^*} has finitely many positive eigenvalues, then m^* denotes the smallest positive value. Otherwise m^* is some

finite large integer. Moreover, the multiplier term can be bounded by a constant c^*. Thus, outside a set of measure zero, for any $\omega \in \Omega$, there exists a $t(\omega)$ such that for all $t > t(\omega)$

$$\lambda_{(m_t)}\left(K_{t\xi}^{m_t}\right) \geq c^*\lambda_{m^*}.$$

Hence, there exists a constant c such that

$$\lim_{t \to \infty} \lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right) > \xi c > 0 \quad \text{almost surely.}$$

Similarly for the upper bound, we see for t large enough,

$$\lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right) = \lambda_{\max}\left(\frac{\xi}{t\xi}\sum_{i=1}^{t\xi}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T + \frac{1-\xi}{t(1-\xi)}\sum_{i=t\xi+1}^{t}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T + \frac{\sigma_t^2}{t}\boldsymbol{I}\right)$$

$$\leq \lambda_{\max}\left(\frac{\xi}{t\xi}\sum_{i=1}^{t\xi}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T\right)$$

(11)

$$+ \lambda_{\max}\left(\frac{1-\xi}{t(1-\xi)}\sum_{i=t\xi+1}^{t}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T\right) + 1.$$

where we have used $\sigma_t^2/t \leq 1$ for large t. This is because for large t, $\sigma_t \to \sigma^*$ by the consistency of the maximum likelihood estimator. Now consider the second term.

$$\lambda_{\max}\left(\frac{1-\xi}{t(1-\xi)}\sum_{i=t\xi+1}^{t}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T\right) = \frac{1-\xi}{t(1-\xi)}\max_{\|x\|=1}\sum_{i=t\xi+1}^{t}x^T\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^Tx$$

$$\leq \frac{1-\xi}{t(1-\xi)}\max_{\|x\|=1}\sum_{i=t\xi+1}^{t}\|x\|^2\|\phi^t(\boldsymbol{x}^i)\|^2 \leq (1-\xi)\alpha$$

where we have used the Cauchy-Schwarz inequality and the regularity conditions on the kernel which gives us,

$$\|\phi^t(\boldsymbol{x}^i)\|^2 = k_{\eta_t}^{m_t}(\boldsymbol{x}^i, \boldsymbol{x}^i) \leq k_{\eta_t}(\boldsymbol{x}^i, \boldsymbol{x}^i) = \alpha \qquad \text{for all } i.$$

Lastly, to control the first term we have

$$\lambda_{\max}\left(\frac{1}{t\xi}\sum_{i=1}^{t\xi}\phi^t(\boldsymbol{x}^i)\phi^t(\boldsymbol{x}^i)^T\right) = \lambda_{(1)}\left(K_{t\xi}^{mt}\right).$$

Similar to the above proof and using the results from Braun (2006), we have with probability larger than $1 - \epsilon^*/t^{1+\delta}$,

$$\lambda_{(1)}\left(K_{t\xi}^{m_t}\right) \leq \lambda_1^t \left(1 + c\sqrt{\frac{\log t}{t}}\right)$$

Applying the Borel-Cantelli Lemma, we have outside a set of measure zero, for any $\omega \in \Omega$, there exists a $t(\omega)$ such that for all $t > t(\omega)$

$$\lambda_{(1)}\left(K_{t\xi}^{m_t}\right) \leq \lambda_1 \tilde{c}^*$$

where λ_1 denotes the maximum eigenvalue of the optimal kernel k_{η^*}. Thus, there exists a constant C, such that

$$\lim_{t \to \infty} \lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right) \leq \xi C + \alpha(1 - \xi) + 1 \quad \text{almost surely,}$$

which completes the proof of the lemma. ∎

Proof of Lemma 2. Observe that,

$$\lambda_{\max}\left(\left(\frac{\boldsymbol{A}}{t}\right)^{-1} \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{t}\left(\frac{\boldsymbol{A}}{t}\right)^{-1}\right) \leq \lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right)^{-1} + \lambda_{\max}\left[\left(\frac{\boldsymbol{A}}{t}\right)^{-1}\boldsymbol{E}\right]$$

where $\boldsymbol{E} = \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{A}^{-1} - \boldsymbol{I}$. Now, it is easy to see that $\boldsymbol{E}$ is a negative definite matrix. Let $\xi_1^t \geq, \ldots, \geq \xi_{m_t}^t$ denote the eigenvalues of $\boldsymbol{\Phi}^T\boldsymbol{\Phi}/t$. Thus, using the spectral expansion there exists orthonormal eigenvectors u_i such that,

$$\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{A}^{-1} = \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{t}\left(\frac{\boldsymbol{A}}{t}\right)^{-1} = \sum_{i=1}^{m_t} \frac{\xi_i^t}{\xi_i^t + \sigma_t^2/t} u_i u_i^T \preccurlyeq \sum_{i=1}^{m_t} u_i u_i^T = \boldsymbol{I}.$$

where $\boldsymbol{A} \preccurlyeq \boldsymbol{B}$ implied, $\boldsymbol{B} - \boldsymbol{A}$ is positive definite. Thus, we get that $\boldsymbol{E}$ is negative definite. Therefore, using this and Lemma 1 we have

$$\lambda_{\max}\left(\left(\frac{\boldsymbol{A}}{t}\right)^{-1} \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{t}\left(\frac{\boldsymbol{A}}{t}\right)^{-1}\right) \leq \lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right)^{-1} - \lambda_{\min}\left[-\left(\frac{\boldsymbol{A}}{t}\right)^{-1}\boldsymbol{E}\right]$$

$$\leq \lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right)^{-1} \leq c$$

where we have used the result that for two positive definite matrices $\boldsymbol{A}, \boldsymbol{B}$, $\lambda_{\min}(\boldsymbol{A}\boldsymbol{B}) \geq \lambda_{\min}(\boldsymbol{A})\lambda_{\min}(\boldsymbol{B}) > 0$. This completes the proof of the Lemma. ∎

Proof of Lemma 3. As in the proof of Lemma 2, let $\xi_1^t \geq, \ldots, \geq \xi_{m_t}^t$ denote the eigenvalues of $\boldsymbol{\Phi}^T \boldsymbol{\Phi}/t$. Thus, using the spectral decomposition,

$$
\begin{aligned}
\left\| \left(\frac{\boldsymbol{A}}{t} \right)^{-1} \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} - \boldsymbol{I} \right\| &= \sqrt{ \lambda_{\max} \left[\left(\left(\frac{\boldsymbol{A}}{t} \right)^{-1} \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} - \boldsymbol{I} \right)^T \left(\left(\frac{\boldsymbol{A}}{t} \right)^{-1} \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} - \boldsymbol{I} \right) \right] } \\
&= \sqrt{ \lambda_{\max} \left[\left(\frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} \left(\frac{\boldsymbol{A}}{t} \right)^{-1} - \boldsymbol{I} \right) \left(\left(\frac{\boldsymbol{A}}{t} \right)^{-1} \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} - \boldsymbol{I} \right) \right] } \\
&= \sqrt{ \lambda_{\max} \left(\frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} \left(\frac{\boldsymbol{A}}{t} \right)^{-2} \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} - \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} \left(\frac{\boldsymbol{A}}{t} \right)^{-1} - \left(\frac{\boldsymbol{A}}{t} \right)^{-1} \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} + \boldsymbol{I} \right) } \\
&= \sqrt{ \lambda_{\max} \left[\sum_{i=1}^{m_t} \left(\left(\frac{\xi_i^t}{\xi_i^t + \sigma_t^2/t} \right)^2 - \frac{2 \xi_i^t}{\xi_i^t + \sigma_t^2/t} + 1 \right) u_i u_i^T \right] } \\
&= \sqrt{ \max_{i=1,\ldots,m_t} \left(1 - \frac{\xi_i^t}{\xi_i^t + \sigma_t^2/t} \right)^2 }.
\end{aligned}
$$

Now for each i, and large enough t, it easily follows from Lemma 1 that the eigenvalues ξ_i^t are bounded. Specifically,

$$
0 < c \leq \lambda_{\min}(\boldsymbol{\Phi}^T \boldsymbol{\Phi}/t) \leq \xi_i^t \leq \lambda_{\max}(\boldsymbol{\Phi}^T \boldsymbol{\Phi}/t) \leq C < \infty.
$$

and $\sigma_t \to \sigma^*$ almost surely by the consistency of the maximum likelihood estimator. Thus, we get for some constant c and large enough t,

$$
\left\| \left(\frac{\boldsymbol{A}}{t} \right)^{-1} \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} - \boldsymbol{I} \right\| = \| \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{\Phi} - \boldsymbol{I} \| \leq \frac{c}{t}
$$

which completes the first result. To prove the second result, note that using Lemma 2,

$$
\begin{aligned}
\| \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \| &= \sqrt{ \lambda_{\max}(\boldsymbol{\Phi} \boldsymbol{A}^{-1} \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T) } = \sqrt{ \lambda_{\max}(\boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{A}^{-1}) } \\
&= \sqrt{ \frac{1}{t} \lambda_{\max} \left(\left(\frac{\boldsymbol{A}}{t} \right)^{-1} \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{t} \left(\frac{\boldsymbol{A}}{t} \right)^{-1} \right) } \\
&\leq \frac{c}{\sqrt{t}},
\end{aligned}
$$

which completes the proof. ∎

Proof of Lemma 4. Using Assumption 1 and Lemma 3 observe that,

$$\|\boldsymbol{u}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{f} - \boldsymbol{u}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{\theta}^*\| \leq \|\boldsymbol{u}\| \|\boldsymbol{A}^{-1} \boldsymbol{\Phi}^T\| \|\boldsymbol{f} - \boldsymbol{\Phi} \boldsymbol{\theta}^*\|$$
$$\leq 2\sqrt{\alpha t} \delta_0(t) \|\boldsymbol{A}^{-1} \boldsymbol{\Phi}^T\|$$
$$\leq c \delta_0(t),$$

where we have used,

$$(12) \qquad \|\boldsymbol{u}\|^2 = \|\phi^t(\boldsymbol{x}^*) - \phi^t(\boldsymbol{x})\|^2 \leq \left(\|\phi^t(\boldsymbol{x}^*)\| + \|\phi^t(\boldsymbol{x})\|\right)^2 \leq 4\alpha$$

and $\|\phi^t(\boldsymbol{x})\|^2 = k^{m_t}_{\eta_t}(\boldsymbol{x}, \boldsymbol{x}) \leq k_{\eta_t}(\boldsymbol{x}, \boldsymbol{x}) = \alpha$. Thus we get,

$$\boldsymbol{u}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{f} \geq \boldsymbol{u}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{\theta}^* - c \delta_0(t)$$
$$(13) \qquad\qquad\qquad = \boldsymbol{u}^T \boldsymbol{\theta}^* - \boldsymbol{u}^T \boldsymbol{E} \boldsymbol{\theta}^* - c \delta_0(t).$$

We can now bound each of the terms on the right hand side of (13). Note that using Lemma 3,

$$(14) \qquad\qquad \|\boldsymbol{u}^T \boldsymbol{E} \boldsymbol{\theta}^*\| \leq \|\boldsymbol{u}\| \|\boldsymbol{E}\| \|\boldsymbol{\theta}^*\| \leq 2\sqrt{\alpha} c M \frac{1}{\sqrt{t}}.$$

Combining all the above bounds and using Assumption 1 we have for large enough t,

$$\boldsymbol{u}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{f} \geq \boldsymbol{u}^T \boldsymbol{\theta}^* - c\left(\frac{1}{\sqrt{t}} + \delta_0(t)\right) \geq f(\boldsymbol{x}^*) - f(\boldsymbol{x}) - c\left(\frac{1}{\sqrt{t}} + \delta_0(t)\right)$$
$$\geq \delta_\epsilon - c\left(\frac{1}{\sqrt{t}} + \delta_0(t)\right) \geq \frac{\delta_\epsilon}{2}$$

where the last inequality follows for large enough t since $\delta_0(t)$ is a decreasing function of t. This completes the proof. ∎

Proof of Lemma 5. The result follows as an consequence of Lemma 1 from Laurent and Massart (2000). ∎

Proof of Lemma 6. For any $\epsilon' < \delta_\epsilon/4$ and for any $\boldsymbol{x}$ such that $\|\boldsymbol{x} - \boldsymbol{x}^*\| > \epsilon$, we have

$$P\left(\phi^t(\boldsymbol{x}^*)^T \boldsymbol{\theta}^t < \phi^t(\boldsymbol{x})^T \boldsymbol{\theta}^t + \epsilon'\right)$$
$$= E_{\phi^t, D_t}\left(P\left(\phi^t(\boldsymbol{x}^*)^T \boldsymbol{\theta}^t < \phi^t(\boldsymbol{x})^T \boldsymbol{\theta}^t + \epsilon' \Big| \phi^t, D_t\right)\right).$$

For notational simplicity we hide the variables we are conditioning on, specifically, ϕ^t, D_t. Moreover, let $\boldsymbol{v}$ denote $\phi^t(\boldsymbol{x}) - \phi^t(\boldsymbol{x}^*)$. Under this simplified notation, let us define $X = \boldsymbol{v}^T \boldsymbol{\theta}^t$, which, given ϕ^t, D_t, follows $N(\mu, \gamma^2)$ with $\mu = \boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{y}$ and $\gamma^2 = \sigma_t^2 \boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{v}$. Thus, we have

$$(15) \qquad E\left(P\left(\boldsymbol{v}^T \boldsymbol{\theta}^t > -\epsilon'\right)\right) \leq E\left(\exp\left(-\frac{(\mu+\epsilon')^2}{2\gamma^2}\right)\right) + P(\mu + \epsilon' > 0)$$

where the last inequality follows by conditioning on the sign on $\mu + \epsilon'$ and appropriately applying the tail bound for Gaussian random variables. Now, we separately consider the two terms in (15). For the first term, conditioning on $\boldsymbol{x}^0, \dots, \boldsymbol{x}^{t-1}$ and ϕ^t define,

$$\zeta = \frac{\mu+\epsilon'}{\gamma}\bigg|\{\boldsymbol{x}^i\}_{i=0}^{t-1}, \phi^t \sim N(\tilde{\mu}, \tilde{\sigma}^2) \qquad \text{where}$$

$$\tilde{\mu} = \frac{\boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{f} + \epsilon'}{\sigma_t \sqrt{\boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{v}}} \qquad \text{and} \qquad \tilde{\sigma}^2 = \frac{\boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{A}^{-1} \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{v}}$$

where $\boldsymbol{f} = (f(\boldsymbol{x}^0), \dots, f(\boldsymbol{x}^{t-1}))^T$. Thus using the above notation we can write the first term as

$$E\left(\exp\left(-\frac{(\mu+\epsilon')^2}{2\gamma^2}\right)\right) = E_{\boldsymbol{x}, \phi^t}\left(E\left(\exp\left(-\frac{\tilde{\sigma}^2}{2}\frac{\zeta^2}{\tilde{\sigma}^2}\right)\right)\right)$$

$$= E_{\boldsymbol{x}, \phi^t}\left(\frac{1}{\sqrt{1-2\tau}}\exp\left(\frac{\lambda\tau}{1-2\tau}\right)\right)$$

where the last equality follows from the moment generating function of a non-central chi-squares distribution with non-centrality parameter $\lambda = \tilde{\mu}^2$ and $\tau = -\tilde{\sigma}^2/2$. Thus, we have,

$$E\left(\exp\left(-\frac{(\mu+\epsilon')^2}{2\gamma^2}\right)\right) = E_{\boldsymbol{x}, \phi^t}\left(\frac{1}{\sqrt{1+\tilde{\sigma}^2}}\exp\left(\frac{-\tilde{\sigma}^2\tilde{\mu}^2}{2(1+\tilde{\sigma}^2)}\right)\right)$$

$$\leq E_{\boldsymbol{x}, \phi^t}\left(\exp\left(\frac{-\tilde{\sigma}^2\tilde{\mu}^2}{2(1+\tilde{\sigma}^2)}\right)\right).$$

To give a lower bound to $\tilde{\sigma}^2$, observe that

$$\tilde{\sigma}^2 \geq \frac{\lambda_{\min}\left(\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{A}^{-1}\right)}{\lambda_{\max}\left(\boldsymbol{A}^{-1}\right)} = \lambda_{\min}\left(\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{A}^{-1}\right)\lambda_{\min}\left(\boldsymbol{A}\right)$$

$$= \lambda_{\min}\left(\frac{\boldsymbol{A}^{-1}}{t}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{A}^{-1}\right)\lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right) = \lambda_{\min}\left(\frac{\boldsymbol{A}^{-1}}{t}\boldsymbol{E} + \frac{\boldsymbol{A}^{-1}}{t}\right)\lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right)$$

$$\geq \left[\lambda_{\min}\left(\frac{\boldsymbol{A}^{-1}}{t}\right) + \lambda_{\min}\left(\frac{\boldsymbol{A}^{-1}}{t}\boldsymbol{E}\right)\right]\lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right)$$

where $\boldsymbol{E} = \boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{A}^{-1} - \boldsymbol{I}$. From the proof of Lemma 3, $\boldsymbol{E}$ is a negative definite matrix. Thus, we can write,

$$\tilde{\sigma}^2 \geq \left[\lambda_{\min}\left(\frac{\boldsymbol{A}}{t}^{-1}\right) - \lambda_{\max}\left(-\frac{\boldsymbol{A}}{t}^{-1}\boldsymbol{E}\right) \right] \lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right)$$

Moreover, using the results in Bhatia (2013),

$$\lambda_{\max}\left(-\frac{\boldsymbol{A}}{t}^{-1}\boldsymbol{E}\right) \leq \lambda_{\max}\left(\frac{\boldsymbol{A}}{t}^{-1}\right) \lambda_{\max}(-\boldsymbol{E})$$

Thus, we have,

$$\tilde{\sigma}^2 \geq \lambda_{\min}\left(\frac{\boldsymbol{A}}{t}^{-1}\right) \lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right) - \lambda_{\max}\left(-\boldsymbol{E}\right) = \frac{\lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right)}{\lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right)} - \lambda_{\max}\left(-\boldsymbol{E}\right)$$

Now using Lemma 1 and Lemma 3, there exists a c^1, c^2 and c^3 such that for large enough t,

$$\lambda_{\max}(-\boldsymbol{E}) \leq \frac{c^1}{t}, \quad \lambda_{\min}\left(\frac{\boldsymbol{A}}{t}\right) \geq c^2 \text{ and } \lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right) \leq c^3$$

Thus, for large enough t there exists a $c_4 > 0$ such that

$$\tilde{\sigma}^2 \geq c_4.$$

Denoting, $c = 1/(2 + 2/c_4)$ we have

$$E\left(\exp\left(-\frac{(\mu + \epsilon')^2}{2\gamma^2}\right)\right) \leq E_{\boldsymbol{x}, \phi^t}\left(\exp\left(-c\tilde{\mu}^2\right)\right).$$

To give a lower bound to $\tilde{\mu}^2$, we separately bound the numerator and the denominator. Using Lemma 4 we can give a lower bound to numerator of $\tilde{\mu}^2$. Specifically, for large enough t we get

$$\left(\boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{f} + \epsilon'\right)^2 = \left|\boldsymbol{u}^T \boldsymbol{A}^{-1} \boldsymbol{\Phi}^T \boldsymbol{f} - \epsilon'\right|^2 \geq \frac{\delta_\epsilon^2}{16},$$

where $\boldsymbol{u} = -\boldsymbol{v} = \phi^t(\boldsymbol{x}^*) - \phi^t(\boldsymbol{x})$. We now give an upper bound on the denominator of $\tilde{\mu}^2$. Note that using (12)

$$\|\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 \leq 4\alpha,$$

Hence we have,

$$\frac{\sigma_t^2}{t} \boldsymbol{v}^T \left(\frac{\boldsymbol{A}}{t}\right)^{-1} \boldsymbol{v} \leq \frac{4\alpha \sigma_t^2}{t} \lambda_{\max}\left(\frac{\boldsymbol{A}}{t}\right)^{-1} \leq \frac{c}{t}.$$

where the last inequality follows from Lemma 1 and the consistency of σ_t to σ^*, for large enough t. Combining the bounds for the numerator and the denominator we get for some constant c_ϵ,

$$E\left(\exp\left(-\frac{(\mu + \epsilon')^2}{2\gamma^2}\right)\right) \leq E_{\boldsymbol{x},\phi^t}\left(\exp\left(-c_\epsilon t\right)\right)$$
$$= \exp\left(-c_\epsilon t\right)$$

Using a very similar proof technique as above, we can show,

$$P(\mu + \epsilon' > 0) \leq \exp\left(-c_\epsilon t\right).$$

Here we use c_ϵ as a generic constant. Thus, there exists a c_ϵ such that for large enough t, we have

$$P\left(\phi^t(\boldsymbol{x}^*)^T\boldsymbol{\theta}^t < \phi^t(\boldsymbol{x})^T\boldsymbol{\theta}^t + \epsilon'\right) \leq 2\exp(-c_\epsilon t).$$

This completes the proof of the lemma. ∎

Proof of Lemma 7. Let $\mu_{\phi^t,D_t} = \boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{y}$, $\Sigma_{\phi^t,D_t} = \sigma_t^2\boldsymbol{A}^{-1}$ and

$$L_t = \sqrt{\sigma_t^2\left(\frac{m_t}{t} + \delta\right)\left(\frac{1}{c_1} + \frac{1}{c_2}\right) + c_3 t}.$$

Then,

$$P\left(\|\boldsymbol{\theta}^t\| > L_t\right) \leq P\left(\|\boldsymbol{\theta}^t\|^2 > L_t^2\right)$$
$$= E_{D_t,\phi^t}\left(P\left(\|\boldsymbol{\theta}^t\|^2 - \|\mu_{\phi^t,D_t}\|^2 > L_t^2 - \|\mu_{\phi^t,D_t}\|^2 \middle| D_t, \phi^t\right)\right)$$
$$\leq E_{D_t,\phi^t}\left(P\left(\|\boldsymbol{\theta}^t - \mu_{\phi^t,D_t}\|^2 > L_t^2 - \|\mu_{\phi^t,D_t}\|^2 \middle| D_t, \phi^t\right)\right)$$

(16)

$$\leq E_{D_t,\phi^t}\left(P\left(\|\boldsymbol{\theta}^t - \mu_{\phi^t,D_t}\|_{\Sigma_{\phi^t,D_t}^{-1}}^2 > \lambda_{\min}(\Sigma_{\phi^t,D_t}^{-1})\left(L_t^2 - \|\mu_{\phi^t,D_t}\|^2\right) \middle| D_t, \phi^t\right)\right)$$

Now, we condition on

(17) $$\|\mu_{\phi^t,D_t}\|^2 \leq c_3 t + \left(\frac{m_t}{t} + \delta\right)\frac{\sigma_t^2}{c_2}.$$

Using (17) we have,

$$\lambda_{\min}(\Sigma_{\phi^t,D_t}^{-1}) \left(L_t^2 - \|\mu_{\phi^t,D_t}\|^2 \right) \geq \lambda_{\min}(\Sigma_{\phi^t,D_t}^{-1}) \frac{(m_t + \delta t)\sigma_t^2}{c_1 t}$$
$$= \frac{t}{\sigma_t^2} \lambda_{\min}\left(\frac{\boldsymbol{A}}{t} \right) \frac{(m_t + \delta t)\sigma_t^2}{c_1 t}$$
$$\geq m_t + \delta t.$$

Plugging this into (16) and using Lemma 5 we get,

$$(18) \quad P\left(\|\boldsymbol{\theta}^t\| > L_t \right) \leq \exp\left(-\frac{\delta t}{2} \right) + P\left(\|\mu_{\phi^t,D_t}\|^2 \geq c_3 t + \left(\frac{m_t}{t} + \delta \right) \frac{\sigma_t^2}{c_2} \right)$$

Similar to the above analysis if we let $\mu_{\phi^t} = \boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{f}$, $\Sigma_{\phi^t} = \sigma_t^2 \boldsymbol{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{A}^{-1}$ and

$$\tilde{L}_t^2 = c_3 t + \left(\frac{m_t}{t} + \delta \right) \frac{\sigma_t^2}{c_2}$$

then

$$P\left(\|\mu_{\phi^t,D_t}\|^2 \geq \tilde{L}_t^2 \right)$$
$$(19)$$
$$\leq E_{X_{t-1},\phi^t}\left(P\left(\|\mu_{\phi^t,D_t} - \mu_{\phi^t}\|_{\Sigma_{\phi^t}^{-1}}^2 > \lambda_{\min}\left(\Sigma_{\phi^t}^{-1} \right) \left(\tilde{L}_t^2 - \|\mu_{\phi^t}\|^2 \right) \middle| X_{t-1}, \phi^t \right) \right)$$

Now note that using Lemma 2 and (9) we have for large enough t,

$$\lambda_{\min}(\Sigma_{\phi^t}^{-1}) \left(\tilde{L}_t^2 - \|\mu_{\phi^t}\|^2 \right) = \frac{1}{\sigma_t^2 \lambda_{\max}\left(\boldsymbol{A}^{-1}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})\boldsymbol{A}^{-1} \right)} \left(\frac{m_t}{t} + \delta \right) \frac{\sigma_t^2}{c_2}$$
$$\geq m_t + \delta t$$

This gives us,

$$(20) \qquad\qquad P\left(\|\mu_{\phi^t,D_t}\|^2 \geq \tilde{L}_t^2 \right) \leq \exp\left(-\frac{\delta t}{2} \right).$$

Using the (18) and (20) for large enough t we get,

$$P\left(\|\boldsymbol{\theta}^t\| > \sqrt{\sigma_t^2 \left(\frac{m_t}{t} + \delta \right) \left(\frac{1}{c_1} + \frac{1}{c_2} \right) + c_3 t} \right) \leq 2\exp\left(-\frac{\delta t}{2} \right).$$

which completes the proof of the Lemma. ∎

700 E. Middlefield Road,
Mountain View, CA 94043
E-mail: kbasu@linkedin.com
         sghosh@linkedin.com