Improved Self-Normalized Concentration in Hilbert Spaces: Sublinear Regret for GP-UCB

Justin Whitehouse
Carnegie Mellon University
jwhiteho@andrew.cmu.edu

Zhiwei Steven Wu
Carnegie Mellon University
zstevenwu@cmu.edu

Aaditya Ramdas
Carnegie Mellon University
aramdas@cmu.edu

Abstract

In the kernelized bandit problem, a learner aims to sequentially compute the optimum of a function lying in a reproducing kernel Hilbert space given only noisy evaluations at sequentially chosen points. In particular, the learner aims to minimize regret, which is a measure of the suboptimality of the choices made. Arguably the most popular algorithm is the Gaussian Process Upper Confidence Bound (GP-UCB) algorithm, which involves acting based on a simple linear estimator of the unknown function. Despite its popularity, existing analyses of GP-UCB give a suboptimal regret rate, which fails to be sublinear for many commonly used kernels such as the Matérn kernel. This has led to a longstanding open question: are existing regret analyses for GP-UCB tight, or can bounds be improved by using more sophisticated analytical techniques? In this work, we resolve this open question and show that GP-UCB enjoys nearly optimal regret. In particular, our results directly imply sublinear regret rates for the Matérn kernel, improving over the state-of-the-art analyses and partially resolving a COLT open problem posed by Vakili et al. Our improvements rely on two key technical results. First, we use modern supermartingale techniques to construct a novel, self-normalized concentration inequality that greatly simplifies existing approaches. Second, we address the importance of regularizing in proportion to the smoothness of the underlying kernel k. Together, these new technical tools enable a simplified, tighter analysis of the GP-UCB algorithm.

1 Introduction

An essential problem in areas such as econometrics [9, 10], medicine [17, 18], optimal control [3, 2], and advertising [16] is to optimize an unknown function given *bandit feedback*, in which algorithms only get to observe the outcomes for the chosen actions. Due to the bandit feedback, there is a fundamental tradeoff between *exploiting* what has been observed about the local behavior of the function and *exploring* to learn more about the function's global behavior. There has been a long line of work on bandit learning that investigates this tradeoff across different settings, including multi-armed bandits [24, 14, 32], linear bandits [1, 25], and kernelized bandits [4, 21, 27].

In this work, we focus on the kernelized bandit framework, which can be viewed as an extension of the well-studied linear bandit setting to an infinite-dimensional reproducing kernel Hilbert space (or RKHS) $(H, \langle \cdot, \cdot \rangle_H)$. In this problem, there is some unknown function $f^* : \mathcal{X} \to \mathbb{R}$ of bounded norm in H, where $\mathcal{X} \subset \mathbb{R}^d$ is a bounded set. In each round $t \in [T]$, the learner uses previous observations to select an action $X_t \in \mathcal{X}$, and then observes feedback $Y_t := f^*(X_t) + \epsilon_t$, where ϵ_t is

Preprint. Under review.

a zero-mean noise variable. The learner aims to minimize (with high probability) the regret at time T, which is defined as

$$R_T := \sum_{t=1}^{T} f^*(x^*) - f^*(X_t)$$

where $x^* := \arg\max_{x \in \mathcal{X}} f^*(x)$. The goal is to develop simple, efficient algorithms for the kernelized bandit problem that minimize regret R_T. We make the following standard assumption. We also make assumptions on the underlying kernel k, which we discuss in Section 2.

Assumption 1. *We assume that (a) there is some constant $D > 0$ known to the learner such that $\|f^*\|_H \leq D$ and (b) for every $t \geq 1$, ϵ_t is σ-subGaussian conditioned on $\sigma(Y_{1:t-1}, X_{1:t})$.*

Arguably the simplest algorithm for the kernelized bandit problem is GP-UCB (Gaussian process upper confidence bound) [26, 4]. GP-UCB works by maintaining a kernel ridge regression estimator of the unknown function f^* alongside a confidence ellipsoid, optimistically selecting in each round the action that provides the maximal payoff over all feasible functions. Not only is GP-UCB efficiently computable thanks to the kernel trick, but it also offers strong empirical guarantees [4]. The only seeming deficit of GP-UCB is its regret guarantee, as existing analyses only show that, with high probability, $R_T = \widetilde{O}(\gamma_T\sqrt{T})$, where γ_T is a kernel-dependent measure of complexity known as the maximum information gain [26, 5]. In contrast, more complicated, less computationally efficient algorithms such as SupKernelUCB [29, 20] have been shown to obtain regret bounds of $\widetilde{O}(\sqrt{\gamma_T T})$, improving over the analysis of GP-UCB by a multiplicative factor of $\sqrt{\gamma_T}$. This gap is stark as the bound $\widetilde{O}(\gamma_T\sqrt{T})$ fails, in general, to be sub-linear for the practically relevant Matérn kernel, whereas $\widetilde{O}(\sqrt{\gamma_T T})$ is sublinear for *any* kernel experiencing polynomial eigendecay [27].

This discrepancy has prompted the development of many variants of GP-UCB that, while less computationally efficient, offer better, regret guarantees in some situations [13, 22, 23]. (See a detailed discussion of these algorithms along with other related work in Appendix A.) However, the following question remains an open problem in online learning [28]: are existing analyses of vanilla GP-UCB tight, or can an improved analysis show GP-UCB enjoys sublinear regret?

1.1 Contributions

In this work, we show that GP-UCB obtains almost optimal, sublinear regret for any kernel experiencing polynomial eigendecay. This, in particular, implies for the first time that GP-UCB obtains sublinear regret for the commonly used Matérn family of kernels. In more detail, our two main contributions are as follows.

1. In Section 3, we construct a novel, time-uniform confidence bound for controlling the growth of self-normalized processes in separable Hilbert spaces. As opposed to the existing bound of Chowdhury and Gopalan [4], which involves employing a complicated "double mixture" argument, our bound follows directly from applying the well-studied finite-dimensional method of mixtures alongside a simple truncation argument [6, 7, 1]. These bounds are clean and show simple dependence on the regularization parameter. We believe our bounds and proofs may be applicable in (or easily adapted to) other kernelized learning problems as well.
2. In Section 4, we use our new concentration inequalities to provide an improved regret analysis for GP-UCB. By carefully choosing regularization parameters based on the smoothness of the underlying kernel, we demonstrate that GP-UCB enjoys sublinear regret of $\widetilde{O}\left(T^{\frac{3+\beta}{2+2\beta}}\right)$ for any kernel experiencing (C, β)-polynomial eigendecay. As a special case of this result, we obtain regret bounds of $\widetilde{O}\left(T^{\frac{\nu+2d}{2\nu+2d}}\right)$ for the commonly used Matérn kernel with smoothness ν in dimension d. Our new analysis improves over existing state-of-the-art analysis for GP-UCB, which fails to guarantee sublinear regret in general for the Matérn kernel family [4], and thus partially resolves an open problem posed by [28].

In sum, our results show that GP-UCB, the go-to algorithm for the kernelized bandit problem, is nearly optimal, coming close to the algorithm-independent lower bounds of Scarlett et al. [20]. Our work thus can be seen as providing theoretical justification for the strong empirical performance of GP-UCB [26]. Perhaps the most important message of our work is the importance of careful regularization in online learning problems. While many existing bandit works treat the regularization parameter as a small, kernel-independent constant, we are able to obtain significant improvements by carefully selecting the regularization parameter. We hope our work will encourage others to pay close attention to the selection of regularization parameters in future works.

2 Background and Problem Statement

Notation. We briefly touch on basic definitions and notational conveniences that will be used throughout our work. If $a_1, \dots, a_t \in \mathbb{R}$, we let $a_{1:t} := (a_1, \dots, a_t)^\top$. Let $(H, \langle \cdot, \cdot \rangle_H)$ be a reproducing kernel Hilbert space associated with a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. We refer to the identity operator on H as id_H. This is distinct from the identity mapping on $\mathbb{R}^d$, which we will refer to as I_d. For elements $f, g \in H$, we define their outer product as $fg^\top := f\langle g, \cdot \rangle_H$ and inner product as $f^\top g := \langle f, g \rangle_H$. For any $t \geq 1$ and sequence of points $x_1, \dots, x_t \in \mathcal{X}$ (which will typically be understood from context), let $\Phi_t := (k(\cdot, x_1), \dots, k(\cdot, x_t))^\top$. We can respectively define the Gram matrix $K_t : \mathbb{R}^t \to \mathbb{R}^t$ and covariance operator $V_t : H \to H$ as $K_t := (k(x_i, x_j))_{i,j \in [t]} = \Phi_t \Phi_t^\top$ and $V_t := \sum_{s=1}^t k(\cdot, x_s)k(\cdot, x_s)^\top = \Phi_t^\top \Phi_t$. These two operators essentially encode the same information about the observed data points, the former being easier to work with when actually performing computations (by use of the well known kernel trick) and latter being easier to algebraically manipulate.

Suppose $A : H \to H$ is a Hermitian operator of finite rank; enumerate its non-zero eigenvalues as $\lambda_1(A), \dots, \lambda_k(A)$. We can define the Fredholm determinant of $I + A$ as $\det(I + A) := \prod_{m=1}^k (1 + \lambda_i(A))$ [15]. For any $t \geq 1, \rho > 0$, and $x_1, \dots, x_t \in \mathcal{X}$, one can check via a straightforward computation that $\det(I_t + \rho^{-1}K_t) = \det(\mathrm{id}_H + \rho^{-1}V_t)$, where K_t and V_t are the Gram matrix and covariance operator defined above. We, again, will use these two quantities interchangeably in the sequel, but will typically prefer the latter in our proofs.

If $(H, \langle \cdot, \cdot \rangle_H)$ is a (now general) separable Hilbert space and $(\varphi_n)_{n \geq 1}$ is an orthonormal basis for H, for any $N \geq 1$ we can define the orthogonal projection operator $\pi_N : H \to \mathrm{span}\{\varphi_1, \dots, \varphi_N\} \subset H$ by $\pi_N f := \sum_{n=1}^N \langle f, \varphi_n \rangle_H \varphi_n$. We can correspondingly the define the projection onto the remaining basis functions to be the map $\pi_N^\perp : H \to \mathrm{span}\{\varphi_1, \dots, \varphi_N\}^\perp$ given by $\pi_N^\perp f := f - \pi_N f$. Lastly, if $A : H \to H$ is a symmetric, bounded linear operator, we let $\lambda_{\max}(A)$ denote the maximal eigenvalue of A, when such a value exists. In particular, $\lambda_{\max}(A)$ will exist whenever A has a finite rank, as will typically be the case considered in this paper.

Basics on RKHSs. Let $\mathcal{X} \subset \mathbb{R}^d$ be some domain. A *kernel* is a positive semidefinite map $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that is square-integrable, i.e. $\int_{\mathcal{X}} \int_{\mathcal{X}} |k(x, y)|^2 dxdy < \infty$. Any kernel k has an associated *reproducing kernel Hilbert space* or *RKHS* $(H, \langle \cdot, \cdot \rangle_H)$ containing the closed span of all partial kernel evaluations $k(\cdot, x), x \in \mathcal{X}$. In particular, the inner product $\langle \cdot, \cdot \rangle_H$ on H satisfies the reproducing relationship $f(x) = \langle f, k(\cdot, x) \rangle_H$ for all $x \in \mathcal{X}$.

A kernel k can be associated with a corresponding *Hilbert-Schmidt operator*, which is the Hermitian operator $T_k : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ given by $(T_k f)(x) := \int_{\mathcal{X}} f(y)k(x, y)dy$ for any $x \in \mathcal{X}$. In short, T_k can be thought of as "smoothing out" or "mollifying" a function f according to the similarity metric induced by k. T_k plays a key role in kernelized learning through *Mercer's Theorem*, which gives an explicit representation for H in terms of the eigenvalues and eigenfunctions of T_k.

Fact 1 (Mercer's Theorem). *Let $(H, \langle \cdot, \cdot \rangle_H)$ be the RKHS associated with kernel k, and let $(\mu_n)_{n \geq 1}$ and $(\phi_n)_{n \geq 1}$ be the sequence of non-increasing eigenvalues and corresponding eigenfunctions for T_k. Let $(\varphi_n)_{n \geq 1}$ be the sequence of rescaled functions $\varphi_n := \sqrt{\mu}_n \phi_n$. Then,*

$$H = \left\{ \sum_{n=1}^\infty \theta_n \varphi_n : \sum_{n=1}^\infty \theta_n^2 < \infty \right\},$$

and $(\varphi_n)_{n \geq 1}$ forms an orthonormal basis for $(H, \langle \cdot, \cdot \rangle_H)$.

We make the following assumption throughout the remainder of our work, which is standard and comes from Vakili et al. [27].

Assumption 2 (**Assumption on kernel** k)**.** *The kernel* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *satisfies (a)* $|k(x, y)| \leq L$ *for all* $x, y \in \mathcal{X}$*, for some constant* $L > 0$ *and (b)* $|\phi_n(x)| \leq B$ *for all* $x \in \mathcal{X}$*, for some* $B > 0$*.*

"Complexity" of RKHS's. By the eigendecay of a kernel k, we really mean the rate of decay of the sequence of eigenvalues $(\mu_n)_{n\geq 1}$. In the literature, there are two common paradigms for studying the eigendecay of k: (C_1, C_2, β)-exponential eigendecay, under which $\forall n \geq 1, \mu_n \leq C_1 \exp(-C_2 n^\beta)$, and (C, β)-polynomial eigendecay, under which $\forall n \geq 1, \mu_n \leq C n^{-\beta}$. For kernels experiencing exponential eigendecay, of which the squared exponential is the most important example, GP-UCB is known to be optimal up to poly-logarithmic factors. However, for kernels experiencing polynomial eigendecay, of which the Matérn family is a common example, existing analyses of GP-UCB fail to yield sublinear regret. It is this latter case we focus on in this work.

Given the above representation in Fact 1, it is clear that the eigendecay of the kernel k governs the "complexity" or "size" of the RKHS H. We make this notion of complexity precise by discussing *maximum information gain*, a sequential, kernel-dependent quantity governing concentration and hardness of learning in RKHS's [5, 26, 27].

Let $t \geq 1$ and $\rho > 0$ be arbitrary. The maximum information gain at time t with regularization ρ is the scalar $\gamma_t(\rho)$ given by

$$\gamma_t(\rho) := \sup_{x_1,\dots,x_t \in \mathcal{X}} \frac{1}{2} \log\det\left(\mathrm{id}_H + \rho^{-1} V_t\right) = \sup_{x_1,\dots,x_t \in \mathcal{X}} \frac{1}{2} \log\det\left(I_t + \rho^{-1} K_t\right).$$

Our presentation of maximum information gain differs from some previous works in that we encode the regularization parameter ρ into our notation. This inclusion is key for our results, as we obtain improvements by carefully selecting ρ. Vakili et al. [27] bound the rate of growth of $\gamma_t(\rho)$ in terms of the rate of eigendecay of the kernel k. We leverage the following fact in our main results.

Fact 2 (**Corollary 1 in Vakili et al. [27]**)**.** *Suppose that kernel* k *satisfies Assumption 2 and experiences* (C, β)*-polynomial eigendecay. Then, for any* $t \geq 1$*, we have*

$$\gamma_t(\rho) \leq \left(\left(\frac{CB^2 t}{\rho}\right)^{1/\beta} \log^{-1/\beta}\left(1 + \frac{Lt}{\rho}\right) + 1\right) \log\left(1 + \frac{Lt}{\rho}\right).$$

We last define the practically relevant Matérn kernel and discuss its eigendecay.

Definition/Fact 3. The Matérn kernel with bandwidth $\sigma > 0$ and smoothness $\nu > 1/2$ is given by

$$k_{\nu,\sigma}(x, y) := \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{\sqrt{2\nu}\|x - y\|_y}{\sigma}\right)^\nu B_\nu\left(\frac{\sqrt{2\nu}\|x - y\|_2}{\sigma}\right),$$

where Γ is the gamma function and B_ν is the modified Bessel function of the second kind. It is known that there is some constant $C > 0$ that may depend on σ but not on d or ν such that $k_{\nu,\sigma}$ experiences $\left(C, \frac{2\nu+d}{d}\right)$-eigendecay [19, 27].

Basics on martingale concentration: A filtration $(\mathcal{F}_t)_{t\geq 0}$ is a sequence of σ-algebras satisfying $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for all $t \geq 1$. If $(M_t)_{t\geq 0}$ is a H-valued process, we say $(M_t)_{t\geq 0}$ is a martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$ if (a) $(M_t)_{t\geq 0}$ is $(\mathcal{F}_t)_{t\geq 0}$-adapted, and (b) $\mathbb{E}(M_t \mid \mathcal{F}_{t-1}) = M_{t-1}$ for all $t \geq 1$. An $\mathbb{R}$-valued process is called a supermartingale if the equality in (b) is replaced with "$\leq$", i.e. supermartingales tend to decrease. Martingales are useful in many statistical applications due to their strong concentration of measure properties [12, 31]. The follow fact can be leveraged to provide time-uniform bounds on the growth of any non-negative supermartingale.

Fact 4 (**Ville's Inequality**)**.** *Let* $(M_t)_{t\geq 0}$ *be a non-negative supermartingale with respect to some filtration. Suppose* $\mathbb{E}M_0 = 1$*. Then, for any* $\delta \in (0, 1)$*, we have*

$$\mathbb{P}\left(\exists t \geq 0 : M_t \geq \frac{1}{\delta}\right) \leq \delta.$$

See Howard et al. [11] for a self-contained proof of Ville's inequality, and many applications.

If $\mathcal{F}$ is a σ-algebra, and ϵ is an $\mathbb{R}$-valued random variable, we say ϵ is σ-subGaussian conditioned on $\mathcal{F}$ if, for any $\lambda \in \mathbb{R}$, we have $\log \mathbb{E}\left(e^{\lambda\epsilon} \mid \mathcal{F}\right) \leq \frac{\lambda^2\sigma^2}{2}$; in particular this condition implies that

ϵ is mean zero. With this, we state the following result on self-normalized processes from Abbasi-Yadkori et al. [1] (based off of results in de la Peña et al. [6]) which is commonly leveraged to construct confidence ellipsoids in the linear bandit setting.

Fact 5 (**Theorem 1 from Abbasi-Yadkori et al. [1]**). *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration, let $(X_t)_{t\geq 1}$ be an $(\mathcal{F}_t)_{t\geq 0}$-predictable sequence in $\mathbb{R}^d$, and let $(\epsilon_t)_{t\geq 1}$ be a real-valued $(\mathcal{F}_t)_{t\geq 1}$-adapted sequence such that conditional on $\mathcal{F}_{t-1}$, ϵ_t is mean zero and σ-subGaussian. Then, for any $\rho > 0$, the process $(M_t)_{t\geq 0}$ given by*

$$M_t := \frac{1}{\sqrt{\det(I_d + \rho^{-1}V_t)}} \exp\left\{\frac{1}{2}\left\|(\rho I_d + V_t)^{-1/2}S_t/\sigma\right\|_2^2\right\}$$

is a non-negative supermartingale with respect to $(\mathcal{F}_t)_{t\geq 0}$, where $S_t := \sum_{s=1}^t \epsilon_s X_s$ and $V_t := \sum_{s=1}^t X_s X_s^\top$. Consequently, by Fact 4, for any confidence $\delta \in (0,1)$, the following holds: with probability at least $1-\delta$, simultaneously for all $t \geq 1$, we have

$$\left\|(V_t + \rho I_d)^{-1/2}S_t\right\|_2 \leq \sigma\sqrt{2\log\left(\frac{1}{\delta}\sqrt{\det(I_d + \rho^{-1}V_t)}\right)}.$$

Note the simple dependence on the regularization parameter $\rho > 0$ in the above bound. While the regularization parameter ρ doesn't prove important in regret analysis for linear bandits (where ρ is treated as constant), the choice for ρ will be critical in our setting. Our main result in the following section will extend Fact 5 to the setting of Hilbert spaces essentially verbatim.

3 Improved Self-Normalized Concentration in Hillbert Spaces

We now discuss the first of our main contributions: a novel time-uniform, self-normalized concentration inequality for processes taking value in any separable Hilbert space. We use this bound in the sequel to construct simpler, more flexible confidence ellipsoids than currently exist for GP-UCB. The idea behind our contributions is straightforward — by using the existing, finite-dimensional bounds of [1] along with a careful limiting argument, we can show and identical result holds in infinite-dimensional Hilbert spaces. We start by presenting our new concentration inequality below.

Theorem 1 (**Self-normalized concentration in Hilbert spaces**). *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration, $(f_t)_{t\geq 1}$ be an $(\mathcal{F}_t)_{t\geq 0}$-predictable sequence in a separable Hilbert space H such that $\|f_t\|_H < \infty$ a.s. for all $t \geq 0$, and $(\epsilon_t)_{t\geq 1}$ be an $(\mathcal{F}_t)_{t\geq 1}$-adapted sequence in $\mathbb{R}$ such that conditioned on $\mathcal{F}_{t-1}$, ϵ_t is mean zero and σ-subGaussian. Defining $S_t := \sum_{s=1}^t \epsilon_s f_s$ and $V_t := \sum_{s=1}^t f_s f_s^\top$, we have that for any $\rho > 0$, the process $(M_t)_{t\geq 0}$ defined by*

$$M_t := \frac{1}{\sqrt{\det(\mathrm{id}_H + \rho^{-1}V_t)}} \exp\left\{\frac{1}{2}\left\|(\rho\,\mathrm{id}_H + V_t)^{-1/2}S_t/\sigma\right\|_H^2\right\}$$

is a nonnegative supermartingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. Consequently, by Fact 4, for any $\delta \in (0,1)$, with probability at least $1-\delta$, simultaneously for all $t \geq 1$, we have

$$\left\|(V_t + \rho I_d)^{-1/2}S_t\right\|_H \leq \sigma\sqrt{2\log\left(\frac{1}{\delta}\sqrt{\det(\mathrm{id}_H + \rho^{-1}V_t)}\right)}.$$

We provide a full proof of Theorem 1 in Appendix B and provide a sketch below. We can summarize our proof in two steps. First, following from Fact 5, the bound in Theorem 1 holds when we project S_t and V_t onto a finite number N of coordinates, defining a "truncated" non-negative supermartingale $M_t^{(N)}$. Secondly, we can make a limiting arugment, showing $M_t^{(N)}$ is "essentially" M_t for large values of N. In other words, we develop a method of mixtures for separable Hilbert spaces[1].

Proof Sketch for Theorem 1. Let $(\varphi_n)_{n\geq 1}$ be an orthonormal basis for H, and, for any $N \geq 1$, let π_N denote the projection operator onto $H_N := \mathrm{span}\{\varphi_1, \dots, \varphi_N\}$. Note that the projected process

[1] A space is separable if it has a countable, dense set. Separability is key, because it means we have a countable basis, whose first N elements we project onto.

$(\pi_N S_t)_{t\geq 1}$ is an H-valued martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. Further, note that the projected variance process $(\pi_N V_t \pi_N^\top)_{t\geq 0}$ satisfies

$$\pi_N V_t \pi_N^\top = \sum_{s=1}^{t} (\pi_N f_s)(\pi_N f_s)^\top.$$

Since, for any $N \geq 1$, H_N is a finite-dimensional Hilbert space, it follows from Lemma 1 that the process $(M_t^{(N)})_{t\geq 0}$ given by

$$M_t^{(N)} := \frac{1}{\sqrt{\det(\mathrm{id}_H + \rho^{-1}\pi_N V_t \pi_N^\top)}} \exp\left\{\frac{1}{2}\left\|(\rho\,\mathrm{id}_H + \pi_N V_t \pi_N^\top)^{-1/2}\pi_N S_t\right\|_H^2\right\},$$

is a non-negative supermartingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. One can check that, for any $t \geq 0$, $M_t^{(N)} \xrightarrow[N\to\infty]{} M_t$. Thus, Fatou's Lemma implies

$$\begin{aligned}
\mathbb{E}\left(M_t \mid \mathcal{F}_{t-1}\right) &= \mathbb{E}\left(\liminf_{N\to\infty} M_t^{(N)} \mid \mathcal{F}_{t-1}\right) \\
&\leq \liminf_{N\to\infty} \mathbb{E}\left(M_t^{(N)} \mid \mathcal{F}_{t-1}\right) \\
&\leq \liminf_{N\to\infty} M_{t-1}^{(N)} \\
&= M_{t-1},
\end{aligned}$$

which proves the first part of the claim. The second part of the claim follows from applying Fact 4 to the defined non-negative supermartingale and rearranging. See Appendix B for details. ■

The following corollary specializes Theorem 1 to the case where H is a RKHS and $f_t = k(\cdot, X_t)$, for all $t \geq 1$. In this special case, we can put our results in terms of the familiar Gram matrix K_t, assuming the quantity is invertible. While we prefer the simplicity and elegance of working directly in the RKHS H in the sequel, the follow corollary allows us to present our bounds in a way that is computationally tractable.

Corollary 1. *Let us assume the same setup as Theorem 1, and additionally assume that (a) $(H, \langle\cdot,\cdot\rangle_H)$ is a RKHS associated with some kernel k, and (b) there is some $\mathcal{X}$-valued $(\mathcal{F}_t)_{t\geq 0}$-predictable process $(X_t)_{t\geq 1}$ such that $(f_t)_{t\geq 1} = (k(\cdot, X_t))_{t\geq 1}$. Then, for any $\rho > 0$ and $\delta \in (0,1)$, we have that, with probability at least $1-\delta$, simultaneously for all $t \geq 0$,*

$$\left\|(V_t + \rho\,\mathrm{id}_H)^{-1/2} S_t\right\|_H \leq \sigma\sqrt{2\log\left(\sqrt{\frac{1}{\delta}\det(I_t + \rho^{-1}K_t)}\right)}.$$

If, in addition, the Gram matrix $K_t = (k(X_i, X_j))_{i,j\in[t]}$ is invertible, we have the equality

$$\|(I_t + \rho K_t^{-1})^{-1/2}\epsilon_{1:t}\|_2 = \|(\rho\,\mathrm{id}_H + V_t)^{-1/2} S_t\|_H.$$

We prove Corollary 1 in Appendix B. We compare our bound to the following result from Chowdhury and Gopalan [4].

Fact 6 (Theorem 1 from Chowdhury and Gopalan [4]). *Assume the same setup as Fact 5. Let $\eta > 0$ be arbitrary, and let $K_t := (k(X_i, X_j))_{i,j\in[t]}$ be the Gram matrix corresponding to observations made by time $t \geq 1$. Then, with probability at least $1-\delta$, simultaneously for all $t \geq 1$, we have*

$$\left\|\left((K_t + \eta I_t)^{-1} + I_t\right)^{-1/2}\epsilon_{1:t}\right\|_2 \leq \sigma\sqrt{2\log\left(\frac{1}{\delta}\sqrt{\det\left((1+\eta)I_t + K_t\right)}\right)}.$$

To the best of our knowledge, Fact 6 is the only existing result on self-normalized, time-uniform concentration in RKHS's. We parameterize the bounds in the above fact in terms of $\eta > 0$ instead of $\rho > 0$ to emphasize the following difference: both sides of our bounds shrink as ρ is increased, whereas both sides of the bound in Fact 6 increase as η grows. Thus, increasing ρ in our bound should be seen as decreasing η in the bound of Chowdhury and Gopalan [4]. The bounds in Corollary 1 and Fact 6 coincide when $\rho = 1$ and $\eta \downarrow 0$ (per Lemma 1 in Chowdhury and Gopalan [4]), but are otherwise not equivalent for other choices of ρ and η.

We believe our bounds to be more usable than those of Chowdhury and Gopalan [4] for several reasons. First, our bounds *directly* extend the method of mixtures (in particular, Fact 5) to potentially infinite-dimensional Hilbert spaces. This allows us to directly leverage existing analysis from Abbasi-Yadkori et al. [1] to prove our regret bounds, with only slight modifications. This is in contrast to the more cumbersome regret analysis that leverages Fact 6, which is not only more difficult to follow, but also obtains inferior regret guarantees.

Second, we note that our bound has a simple dependence on $\rho > 0$. In more detail, directly as a byproduct of our simpler bounds, Theorem 2 offers a regret bound that can readily be tuned in terms of ρ. Due to their use of a "double mixture" technique in proving Fact 6, Chowdhury and Gopalan [4] essentially wind up with a nested, doubly-regularized matrix $((K_t + \eta I_t)^{-1} + I_t)^{-1/2}$ with which they normalize the residuals $\epsilon_{1:t}$. In particular, this more complicated normalization make it difficult to understand how varying η impacts regret guarantees, which we find to be essential for proving improved regret guarantees.

4 An Improved Regret Analysis of GP-UCB

In this section, we provide the second of our main contributions, which is an improved regret analysis for the GP-UCB algorithm. We provide a description of GP-UCB in Algorithm 1. While we state the algorithm directly in terms of quantities in the RKHS H, these quantities can be readily converted to those involving Gram matrices or Gaussian processes for those who prefer that perspective [4, 33].

As seen in Section 3, by carefully extending the "method of mixtures" bounds of Abbasi-Yadkori et al. [1] and de la Peña et al. [6, 7] to Hilbert spaces, we can construct self-normalized concentration inequalities that have simple dependence on the regularization parameter ρ. These simplified bounds, in conjunction with information about the eigendecay of the kernel k [27], can be combined to carefully choose ρ to obtain improved regret. We now present our main result.

Algorithm 1 Gaussian Process Upper Confidence Bound (GP-UCB)

Input: Regularization parameter $\rho > 0$, norm bound D, confidence bounds $(U_t)_{t \geq 1}$, and time horizon T.
  Set $V_0 := \rho \mathrm{id}_H$, $f_0 := 0$, $\mathcal{E}_0 := \{f \in H : \|f\|_H \leq D\}$
  for $t = 1, \ldots, T$ **do**
    Let $(X_t, \widetilde{f}_t) := \arg\max_{x \in \mathcal{X}, f \in \mathcal{E}_{t-1}} (f, k(\cdot, x))_H$
    Play action X_t and observe reward $Y_t := f^*(X_t) + \epsilon_t$
    Set $V_t := V_{t-1} + k(\cdot, X_t)k(\cdot, X_t)^\top$ and $f_t := (V_t + \rho \mathrm{id}_H)^{-1}\Phi_t^\top Y_{1:t}$
    Set $\mathcal{E}_t := \left\{f \in H : \left\|(V_t + \rho \mathrm{id}_H)^{1/2}(f_t - f)\right\|_H \leq U_t\right\}$

Theorem 2. *Let $T > 0$ be a fixed time horizon, $\rho > 0$ a regularization parameter, and assume Assumptions 2 and 1 hold. Let $\delta \in (0, 1)$, and for $t \geq 1$ define*

$$U_t := \sigma \sqrt{2 \log\left(\frac{1}{\delta}\sqrt{\det(\mathrm{id}_H + \rho^{-1}V_t)}\right)} + \rho^{1/2} D.$$

Then, with probability at least $1 - \delta$, the regret of Algorithm 1 run with parameters $\rho, (U_t)_{t \geq 1}, D$ satisfies

$$R_T = O\left(\gamma_T(\rho)\sqrt{T} + \sqrt{\rho \gamma_T(\rho) T}\right),$$

where in the big-Oh notation above we treat $\delta, D, \sigma, B,$ and L as being held constant. If the kernel k experiences (C, β)-polynomial eigendecay for some $C > 0$ and $\beta > 1$, taking $\rho = T^{\frac{1}{1+\beta}}$ yields $R_T = \widetilde{O}\left(T^{\frac{3+\beta}{2+2\beta}}\right)$[2], which is always sub-linear in T.

We specialize the above theorem to the case of the Matérn kernel in the following corollary.

[2]The notation $\widetilde{O}$ suppresses multiplicative, poly-logarithmic factors in T

Corollary 2. *Definition 3 states that the Matérn kernel with smoothness $\nu > 1/2$ in dimension d experiences $(C, \frac{2\nu+d}{d})$-eigendecay, for some constnat $C > 0$. Thus, GP-UCB obtains a regret rate of $R_T = \widetilde{O}\left(T^{\frac{\nu+2d}{2\nu+2d}}\right)$.*

We note that our regret analysis is the first to show that GP-UCB attains sublinear regret for general kernels experiencing polynomial eigendecay. Of particular import is that Corollary 2 of Theorem 2 yields the first analysis of GP-UCB that implies sublinear regret for the Matérn kernel under general settings of ambient dimension d and smoothness ν.

We note that our analysis does not obtain optimal regret, as the theoretically interesting but computationally cumbersome SupKernelUCB algorithm [20, 29] obtains a slightly improved regret bound of $\widetilde{O}\left(T^{\frac{\beta+1}{2\beta}}\right)$ for (C, β)-polynomial eigendecay and $\widetilde{O}\left(T^{\frac{\nu+d}{2\nu+d}}\right)$ for the Matérn kernel with smoothness ν in dimension d. We hypothesize that it may not be possible to construct self-normalized concentration inequalities under $\|\cdot\|_H$ that obtain this bound, and supply a heuristic justification in the conclusion.

We now provide a sketch of the proof of Theorem 2. The entire proof, along with full statements and proofs of the technical lemmas, can be found in Appendix C.

Proof Sketch for Theorem 2. Letting, for any $t \in [T]$, the "instantaneous regret" be defined as $r_t := f^*(x^*) - f^*(X_t)$, a standard argument yields that, with probability at least $1 - \delta$, simultaneously for all $t \in [T]$,

$$r_t \leq 2U_{t-1} \left\| (\rho \mathrm{id}_H + V_{t-1})^{-1/2} k(\cdot, X_t) \right\|_H .$$

A further standard argument using Cauchy-Schwarz and an elliptical potential argument yields

$$\begin{aligned}
R_T &= \sum_{t=1}^T r_t \leq U_T \sqrt{2T \log\det(\mathrm{id}_H + \rho^{-1} V_T)} \\
&= \left(\sigma \sqrt{2 \log\left(\frac{1}{\delta}\sqrt{\det(\mathrm{id}_H + \rho^{-1} V_T)}\right)} + \rho^{1/2} D\right) \sqrt{2T \log\det(\mathrm{id}_H + \rho^{-1} V_T)} \\
&\leq \left(\sigma\sqrt{2\log(1/\delta)} + \sigma\sqrt{2\gamma_T(\rho)} + \rho^{1/2} D\right) \sqrt{4T\gamma_T(\rho)} = O\left(\gamma_T(\rho)\sqrt{T} + \sqrt{\rho\gamma_T(\rho)T}\right),
\end{aligned}$$

which proves the first part of the claim. If, additionally, k experiences (C, β)-polynomial eigendecay, we know that $\gamma_T(\rho) = \widetilde{O}\left(\left(\frac{T}{\rho}\right)^{1/\beta}\right)$ by Fact 2. Setting $\rho := T^{\frac{1}{1+\beta}}$ thus yields

$$R_T = O\left(\gamma_T(\rho)\sqrt{T} + \sqrt{\rho\gamma_T(\rho)T}\right) = \widetilde{O}\left(T^{\frac{3+\beta}{2+2\beta}}\right),$$

proving the second part of the claim. ■

5 Conclusion

In this work, we present an improved analysis for the GP-UCB algorithm in the kernelized bandit problem. We provide the first analysis showing that GP-UCB obtains sublinear regret when the underlying kernel k experiences polynomial eigendecay, which in particular implies sublinear regret rates for the practically relevant Matérn kernel. In particular, we show GP-UCB obtains regret $\widetilde{O}\left(T^{\frac{3+\beta}{2+2\beta}}\right)$ when k experiences (C, β)-polynomial eigendecay, and regret $\widetilde{O}\left(T^{\frac{\nu+2d}{2\nu+2d}}\right)$ for the Matérn kernel with smoothness ν in dimension d.

Our technical contributions are twofold. First, we show how to extend self-normalized concentration inequalities for finite-dimensional, Euclidean spaces directly to the case of Hilbert spaces through carefully making a truncation argument. Second, we demonstrate the importance of regularization in the kernelized bandit problem. In particular, since the smoothness of the kernel k governs the hardness of learning, by regularizing in proportion to the rate of eigendecay of k, one can obtain significantly improved regret bounds.

A shortcoming of our work is that, despite obtaining the first generally sublinear regret bounds for GP-UCB, our rates are not optimal. In particular, there are discretization-based algorithms, such as SupKernelUCB [29], which obtain slightly better regret bounds of $\widetilde{O}\left(T^{\frac{1+\beta}{2\beta}}\right)$ for (C, β)-polynomial eigendecay. We hypothesize that the vanilla GP-UCB algorithm, which involves constructing confidence ellipsoids directly in the RKHS H, cannot obtain this rate.

The common line of reasoning [28] is that because the Lin-UCB (the equivalent algorithm in $\mathbb{R}^d$) obtains the optimal regret rate of $\widetilde{O}(d\sqrt{T})$ in the linear bandit problem setting, then GP-UCB should attain optimal regret as well. In the linear bandit setting, there is no subtlety between estimating the optimal action and unknown slope vector, as these are one and the same. In the kernel bandit setting, estimating the function and optimal action are not equivalent tasks. In particular, the former serves in essence as a nuisance parameter in estimating the latter: tight estimation of unknown function under the Hilbert space norm implies tight estimation of the optimal action, but not the other way around. Existing optimal algorithms are successful because they discretize the input domain, which has finite metric dimension [21], and make no attempts to estimate the unknown function in RKHS norm. Since compact sets in RKHS's do not, in general, have finite metric dimension [30], this makes estimation of the unknown function a strictly more difficult task. We leave the verification of this hypothesis as future work.

6 Acknowledgements

AR acknowledges support from NSF DMS 1916320 and an ARL IoBT CRA grant. Research reported in this paper was sponsored in part by the DEVCOM Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196 (ARL IoBT CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

ZSW and JW were supported in part by the NSF CNS2120667, a CyLab 2021 grant, a Google Faculty Research Award, and a Mozilla Research Grant.

JW acknowledges support from NSF GRFP grants DGE1745016 and DGE2140739.

References

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24, 2011.

[2] Rajeev Agrawal. The continuum-armed bandit problem. *SIAM Journal on Control and Optimization*, 33(6):1926–1951, 1995.

[3] Andrew G Barto. Reinforcement learning control. *Current Opinion in Neurobiology*, 4(6): 888–893, 1994.

[4] Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, pages 844–853. PMLR, 2017.

[5] Thomas M Cover and Joy A Thomas. Information theory and statistics. *Elements of Information Theory*, 1(1):279–335, 1991.

[6] Victor de la Peña, Michael J Klass, and Tze Leung Lai. Self-normalized processes: Exponential inequalities, moment bounds and iterated logarithm laws. *The Annals of Probability*, 32, 07 2004. doi: 10.1214/009117904000000397.

[7] Victor de la Peña, Michael J Klass, and Tze Leung Lai. Pseudo-maximization and self-normalized processes. *Probability Surveys Vol*, 4:172–192, 09 2007. doi: 10.1214/07-PS119.

[8] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[9] Vivek Farias, Ciamac Moallemi, Tianyi Peng, and Andrew Zheng. Synthetically controlled bandits. *arXiv preprint arXiv:2202.07079*, 2022.

[10] Matthew Hoffman, Eric Brochu, and Nando De Freitas. Portfolio allocation for Bayesian optimization. In *UAI*, pages 327–336, 2011.

[11] Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020.

[12] Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2), 2021.

[13] David Janz, David Burt, and Javier González. Bandit optimisation of functions in the matérn kernel RKHS. In *International Conference on Artificial Intelligence and Statistics*, pages 2486–2495. PMLR, 2020.

[14] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[15] Peter D Lax. *Functional Analysis*, volume 55. John Wiley & Sons, 2002.

[16] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide web*, pages 661–670, 2010.

[17] Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. *Operations Research*, 68(4):1132–1161, 2020.

[18] Aditya Mate, Jackson Killian, Haifeng Xu, Andrew Perrault, and Milind Tambe. Collapsing bandits and their application to public health intervention. *Advances in Neural Information Processing Systems*, 33:15639–15650, 2020.

[19] Gabriele Santin and Robert Schaback. Approximation of eigenfunctions in kernel-based spaces. *Advances in Computational Mathematics*, 42(4):973–993, 2016.

[20] Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy Gaussian process bandit optimization. In *Conference on Learning Theory*, pages 1723–1742. PMLR, 2017.

[21] Shubhanshu Shekhar and Tara Javidi. Gaussian process bandits with adaptive discretization. *Electronic Journal of Statistics*, 12(2):3829 – 3874, 2018.

[22] Shubhanshu Shekhar and Tara Javidi. Multi-scale zero-order optimization of smooth functions in an RKHS. *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 288–293, 2020.

[23] Shubhanshu Shekhar and Tara Javidi. Instance dependent regret analysis of kernelized bandits. In *International Conference on Machine Learning*, pages 19747–19772. PMLR, 2022.

[24] Aleksandrs Slivkins et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

[25] Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 27, 2014.

[26] Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *International Conference on Machine Learning*, 2009.

[27] Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in Gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 82–90. PMLR, 2021.

[28] Sattar Vakili, Jonathan Scarlett, and Tara Javidi. Open problem: Tight online confidence intervals for RKHS elements. In *Conference on Learning Theory*, pages 4647–4652. PMLR, 2021.

[29] Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nelo Cristianini. Finite-time analysis of kernelised contextual bandits. *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence*, 2013.

[30] Martin J Wainwright. *High-dimensional Statistics: A Non-asymptotic Viewpoint*, volume 48. Cambridge university press, 2019.

[31] Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society, Series B*, 2023.

[32] Peter Whittle. Multi-armed bandits and the Gittins index. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):143–149, 1980.

[33] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

A Related Work

The kernelized bandit problem was first studied by Srinivas et al. [26], who introduce the GP-UCB algorithm and characterize its regret in both the Bayesian and Frequentist setting. While the authors demonstrate that GP-UCB obtains sublinear regret in the Bayesian setting for the commonly used kernels, their bounds fail to be sublinear in general in the frequentist setting for the Matérn kernel, one of the most popular kernel choices in practice. Chowdhury and Gopalan [4] further study the performance of GP-UCB in the frequentist setting. In particular, by leveraging a martingale-based "double mixture" argument, the authors are able to significantly simplify the confidence bounds presented in Srinivas et al. [26]. Unfortunately, the arguments introduced by Chowdhury and Gopalan [4] did not improve regret bounds beyond logarithmic factors, and thus GP-UCB continued to fail to obtain sublinear regret for certain kernels in their work.

There are many other algorithms that have been created for kernelized bandits. Janz et al. [13] introduce an algorithm specific to the Matérn kernel that obtains significantly improved regret over GP-UCB. This algorithm adaptively partitions the input domain into small hypercubes and running an instance of GP-UCB in each element of the discretized domain. Shekhar and Javidi [23] introduce an algorithm called LP-GP-UCB, which augments the GP-UCB estimator with local polynomial corrections. While in the worst case this algorithm recovers the regret bound of Chowdhury and Gopalan [4], if additional information is known about the unknown function f^* (e.g. it is Holder continuous), it can provide improved regret guarantees. Perhaps the most important non-GP-UCB algorithm in the literature is the SupKernel algorithm introduced by Valko et al. [29], which discretizes the input domain and successively eliminates actions from play. This algorithm is signficant because, despite its complicated nature, it obtains regret rates that match known lower bounds provided by Scarlett et al. [20] up to logarithmic factors.

Intimately tied to the kernelized bandit problem is the information-theoretic quantity of maximum information gain [5, 26], which is a sequential, kernel-specific measure of hardness of learning. Almost all preceding algorithms provide regret bounds in terms of the max information gain. Of particular import for our paper is the work of Vakili et al. [27]. In this work, the authors use a truncation argument to upper bound the maximum information gain of kernels in terms of their eigendecay. We directly employ these bounds in our improved analysis of GP-UCB. The max-information gain bounds presented in Vakili et al. [27] can be coupled with the regret analysis in Chowdhury and Gopalan [4] to yield a regret bound of $\widetilde{O}\left(T^{\frac{\nu+3d/2}{2\nu+d}}\right)$ in the case of the Matérn kernel with smoothness ν in dimension d. In particular, when $\nu \leq \frac{d}{2}$, this regret bound fails to be sublinear. In practical setting, d is viewed as large and ν is taken to be $3/2$ or $5/2$, making these bounds vacuous [21, 33] The regret bounds in this paper are sublinear for *any* selection of smoothness $\nu > \frac{1}{2}$ and $d \geq 1$. Moreover, a simple computation yields that our regret bounds strictly improve over (in terms of d and ν) those implied by Vakili et al. [27].

Last, we touch upon the topic of self-normalized concentration, which is an integral tool for constructing confidence bounds in UCB-like algorithms. Heuristically, self-normalized aims to sequentially control the growth of processes that have been rescaled by their variance to look, roughly speaking, normally (or subGaussian) distributed. The prototypical example of self-normalized concentration in the bandit literature comes from Abbasi-Yadkori et al. [1], wherein the authors use a well known technique called the "method of mixtures" to construct confidence ellipsoids for finite dimensional online regression estimates. The concentration result in the aforementioned work is a specialization of results in de la Peña et al. [6], which provide self-normalized concentration for a wide variety of martingale-related processes, several of which have been recently improved [11]. Chowdhury and Gopalan [4] extend the results of Abbasi-Yadkori et al. [1] to the kernel setting using a "double mixture" technique, allowing them to construct self-normalized concentration inequalities for infinite-dimensional processes. We also use the method of mixtures, but a much simpler finite-dimensional version, as we explain next.

B Technical Lemmas for Theorem 1

In this appendix, prove Theorem 1 along with several corresponding technical lemmas. While many of the following results are intuitively true, we provide their proofs in full rigor, as there can be subtleties when working in infinite-dimensional spaces. Throughout, we assume that the subGaussian noise parameter is $\sigma = 1$. The general case can readily be recovered by considering the rescaled process $(S_t/\sigma)_{t\geq 0}$.

The first lemma we present is a restriction of Theorem 1 to the case where the underlying Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$ is finite dimensional, say of dimension N. In this setting, the result essentially follows immediately from Fact 5. All we need to do is construct a natural isometric isomorphism between the spaces H and $\mathbb{R}^N$, and then argue that applying such a mapping doesn't alter the norm of the self-normalized process.

Lemma 1. *Theorem 1 holds if we additionally assume that H is finite dimensional, i.e. if there exists $N \geq 1$ and orthonormal functions $\varphi_1, \ldots, \varphi_N$ such that*

$$H := \mathrm{span}\{\varphi_1, \ldots, \varphi_N\}.$$

Proof. Let $\tau : H \to \mathbb{R}^N$ be the map that takes a function $f = \sum_{n=1}^N \theta_n\varphi_n \in H$ to its natural embedding $\tau f := (\theta_1, \ldots, \theta_N)^\top \in \mathbb{R}^N$. Not only is the map τ an isomorphism between H and $\mathbb{R}^N$, but it is also an isometry, i.e. $\|f\|_H = \|\tau f\|_2$ for all $f \in H$. Further, τ satisfies the relation $\tau^\top = \tau^{-1}$.

Define the "hatted" processes $(\widehat{S}_t)_{t\geq 1}$ and $(\widehat{V}_t)_{t\geq 1}$, which take values in $\mathbb{R}^N$ and $\mathbb{R}^{N\times N}$ respectively as

$$\widehat{S}_t = \sum_{s=1}^t \epsilon_s \tau k(\cdot, X_s) \qquad \text{and} \qquad \widehat{V}_t = \sum_{s=1}^t (\tau k(\cdot, X_s))(\tau k(\cdot, X_s))^\top.$$

It is not hard to see that, by the linearity of τ, that for any $t \geq 1$, we have $\widehat{S}_t = \tau S_t$ and $\widehat{V}_t = \tau V_t \tau^\top$. We observe that (a) $(\widehat{V}_t + \rho I_N)^{-1/2} = \tau(V_t + \rho\mathrm{id}_H)^{-1/2}\tau^\top$ and (b) that the eigenvalues of $\widehat{V}_t$ are exactly those of V_t.

Since the processes $(\widehat{S}_t)_{t\geq 1}$ and $(\widehat{V}_t)_{t\geq 1}$ satisfy the assumptions of Theorem 5, we see that the process $(M_t)_{t\geq 0}$ given by

$$M_t := \frac{1}{\sqrt{\det(I_N + \rho^{-1}\widehat{V}_t)}} \exp\left\{\frac{1}{2}\left\|(\rho I_N + \widehat{V}_t)^{-1/2}\widehat{S}_t\right\|_2^2\right\}$$

is a non-negative supermartingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. From observation (a), the fact τ is an isometry, and the fact $\tau^\top = \tau^{-1}$, it follows that

$$\begin{aligned}
\left\|(\widehat{V}_t + \rho I_N)^{-1/2}\widehat{S}_t\right\|_2 &= \left\|\tau(V_t + \rho\mathrm{id}_H)^{-1/2}\tau^\top\tau S_t\right\|_2 \\
&= \left\|(V_t + \rho\mathrm{id}_H)^{-1/2}\tau^{-1}\tau S_t\right\|_H \\
&= \left\|(V_t + \rho\mathrm{id}_H)^{-1/2}S_t\right\|_H.
\end{aligned}$$

Further, observation (b) implies that

$$\det(I_N + \rho\widehat{V}_t) = \det(\mathrm{id}_H + \rho V_t).$$

Substituting these identities into the definition of $(M_t)_{t\geq 0}$ yields the desired result, i.e. that

$$M_t = \frac{1}{\sqrt{\det(\mathrm{id}_H + \rho^{-1}V_t)}} \exp\left\{\frac{1}{2}\left\|(V_t + \rho I_d)^{-1/2}S_t\right\|_H^2\right\}.$$

is a non-negative supermartingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. The remainder of the result follows from applying Ville's Inequality (Fact 4) and rearranging.

■

We can prove Theorem 1 by truncating the Hilbert space H onto the first N components, applying Lemma 1 to the "truncated" processes $(\pi_N S_t)_{t\geq 0}$ and $(\pi_N V_t \pi_N)_{t\geq 0}$ to construct a relevant, non-negative supermartingale $M_t^{(N)}$, and then show that the error from truncation in this non-negative supermartingale tends towards zero as N grows large. The following two technical lemmas are useful in showing that this latter truncation tends towards zero.

Lemma 2. *For any $t \geq 1$, let V_t be as in the statement of Theorem 1, and let π_N be as in Section 2. Then, we have*

$$\pi_N V_t \pi_N \xrightarrow[N\to\infty]{} V_t,$$

where the above convergence holds under the operator norm on H.

Proof. Fix $\epsilon > 0$, $t \geq 1$, and for $s \in [t]$, let us write $f_s = \sum_{n=1}^{\infty} \theta_n(s)\varphi_n$. Since we have assumed $\|f_t\|_H < \infty$ for all $t \geq 1$, there exists some $N_t < \infty$ such that, for all $s \in [t]$, $\|\pi_{N_t}^{\perp} f_s\|_H^2 = \sum_{n=N_t+1}^{\infty} \theta_n(s)^2 < \frac{\epsilon}{2t}$. We also have, for any $s \in [t]$ and $N \geq 1$, that f_s is an eigenfunction of $f_s f_s^{\top} \pi_N^{\perp} = f_s \langle f_s, \pi_N^{\perp}(\cdot)\rangle_H$ with corresponding (unique) eigenvalue $\|f_s f_s^{\top}\pi_N^{\perp}\|_{op} = \lambda_{\max}(f_s f_s^{\top} \pi_N^{\perp}) = \|\pi_N^{\perp} f_s\|_H^2 = \sum_{n=N+1}^{\infty} \theta_n(s)^2$. Observe that, as an orthogonal projection operator, π_N is self-adjoint, i.e. $\pi_N = \pi_N^{\top}$. With this information, we see that, for $N \geq N_t$, we have

$$\begin{aligned}
\|\pi_N V_t \pi_N - V_t\|_{op} &\leq \sum_{s=1}^{t} \left\|\pi_N f_s f_s^{\top} \pi_N - f_s f_s^{\top}\right\|_{op} \\
&= \sum_{s=1}^{t} \left\|\pi_N f_s f_s^{\top} \pi_N - \pi_N f_s f_s^{\top} + \pi_N f_s f_s^{\top} - f_s f_s^{\top}\right\|_{op} \\
&\leq \sum_{s=1}^{t} \left\|\pi_N f_s f_s^{\top} \pi_N - \pi_N f_s f_s^{\top}\right\|_{op} + \left\|\pi_N f_s f_s^{\top} - f_s f_s^{\top}\right\|_{op} \\
&\leq \sum_{s=1}^{t} \|\pi_N\|_{op} \left\|f_s f_s^{\top} \pi_N - f_s f_s^{\top}\right\|_{op} + \left\|\pi_N f_s f_s^{\top} - f_s f_s^{\top}\right\|_{op} \\
&= \sum_{s=1}^{t} 2\left\|f_s f_s^{\top} \pi_N^{\perp}\right\|_{op} = \sum_{s=1}^{t} 2\|\pi_N^{\perp} f_s\|_H^2 < \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, we have shown the desired result. ■

Lemma 3. *For any $t \geq 1$, let V_t be as in Theorem 1, $\rho > 0$ arbitrary, and π_N as in Section 2. Then, we have*

$$\det(\mathrm{id}_H + \rho^{-1}\pi_N V_t \pi_N) \xrightarrow[N\to\infty]{} \det(\mathrm{id}_H + \rho^{-1} V_t).$$

Proof. We know that the mapping $A \mapsto \det(\mathrm{id}_H + A)$ is continuous under the "trace norm" $\|A\|_1 := \sum_{n=1}^{\infty} |\lambda_n(A)|$ [15]. Thus, to show the desired result, it suffices to show that $\|\pi_N V_t \pi_N - V_t\|_1 \xrightarrow[N\to\infty]{} 0$. Observe that both $\pi_N V_t \pi_N$ and V_t are operators of rank at most t, so so their difference $\pi_N V_t \pi_N - V_t$ has rank at most $2t$. Thus, we know that

$$\|\pi_N V_t \pi_N - V_t\|_1 \leq 2t\|\pi_N V_t \pi_N - V_t\|_{op} \xrightarrow[N\to\infty]{} 0,$$

where the final convergence follows from Lemma 2. Thus, we have shown the desired result. ■

We now tie together all of these technical (but intuitive) results in the proof of Theorem 1 below.

Proof of Theorem 1. Let $(\varphi_n)_{n\geq 1}$ be an orthonormal basis for H, and for $N \geq 1$, let π_N denote the projection operator outlined in Section 2. Recall that $\pi_N = \pi_N^{\top}$. Further $H_N := \mathrm{span}\{\varphi_1, \dots, \varphi_N\} \subset H$ is the image of H under π_N. Since $(S_t)_{t\geq 0}$ is an H-valued martingale with respect to $(\mathcal{F}_t)_{t\geq 1}$, it follows that the projected process $(\pi_N S_t)_{t\geq 1}$ is an H_N-valued martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. Further, note that the projected variance process $(\pi_N V_t \pi_N^{\top})_{t\geq 0}$ satisfies

$$\pi_N V_t \pi_N^{\top} = \sum_{s=1}^{t} (\pi_N f_s)(\pi_N f_s)^{\top}.$$

Since, for any $N \geq 1$, H_N is a finite-dimensional Hilbert space, it follows from Lemma 1 that the process $(M_t^{(N)})_{t\geq 0}$ given by

$$\begin{aligned} M_t^{(N)} &:= \frac{1}{\sqrt{\widetilde{\det}(\mathrm{id}_{H_N} + \rho^{-1}\pi_N V_t \pi_N^\top)}} \exp\left\{\frac{1}{2}\left\|(\rho\mathrm{id}_{H_N} + \pi_N V_t \pi_N^\top)^{-1/2}\pi_N S_t\right\|_{H_N}^2\right\} \\ &= \frac{1}{\sqrt{\det(\mathrm{id}_H + \rho^{-1}\pi_N V_t \pi_N^\top)}} \exp\left\{\frac{1}{2}\left\|(\rho\mathrm{id}_H + \pi_N V_t \pi_N^\top)^{-1/2}\pi_N S_t\right\|_H^2\right\}, \end{aligned}$$

is a non-negative supermartingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. In the above id_{H_N} denotes the identity id_H restricted to $H_N \subset H$ and $\widetilde{\det}$ denotes the determinant restricted to the subspace H_N. The equivalence of the second and third terms above is trivial.

We now argue that for any $t \geq 1$,

$$\lim_{N\to\infty} M_t^{(N)} = M_t. \tag{1}$$

If we show this to be true, then we have, for any $t \geq 1$

$$\begin{aligned} \mathbb{E}\left(M_t \mid \mathcal{F}_{t-1}\right) &= \mathbb{E}\left(\liminf_{N\to\infty} M_t^{(N)} \mid \mathcal{F}_{t-1}\right) \\ &\leq \liminf_{N\to\infty} \mathbb{E}\left(M_t^{(N)} \mid \mathcal{F}_{t-1}\right) \\ &\leq \liminf_{N\to\infty} M_{t-1}^{(N)} \\ &= M_{t-1}, \end{aligned}$$

which implies $(M_t)_{t\geq 0}$ is a non-negative supermartingale with respect to $(\mathcal{F}_t)_{t\geq 0}$ thus proving the result. In the above, the first inequality follows from Fatou's lemma for conditional expectations (see Durrett [8], for instance), and the second inequality follows from the supermartingale property.

Lemma 3 tells us that $\det(\mathrm{id}_H + \rho^{-1}\pi_N V_t \pi_N) \xrightarrow[N\to\infty]{} \det(\mathrm{id}_H + \rho^{-1}V_t)$ for all $t \geq 1$, so to show the desired convergence in (1), it suffices to show that

$$\|(\rho\mathrm{id}_H + \pi_N V_t \pi_N)^{-1/2}\pi_N S_t\|_H \xrightarrow[N\to\infty]{} \|(\rho\mathrm{id}_H + V_t)^{-1/2} S_t\|_H \text{ for any } t.$$

Let $\mathcal{V}_t := \rho\mathrm{id}_H + V_t$ and $\mathcal{V}_t(N) := \rho\mathrm{id}_H + \pi_N V_t \pi_N$ in the following line of reason for simplicity. We trivially have

$$\begin{aligned} \left|\|\mathcal{V}_t(N)^{-1/2}\pi_N S_t\|_H - \|\mathcal{V}_t^{-1/2} S_t\|_H\right| &\leq \left\|\mathcal{V}_t(N)^{-1/2}\pi_N S_t - \mathcal{V}_t^{-1/2} S_t\right\|_H \\ &= \left\|\mathcal{V}_t(N)^{-1/2}\pi_N S_t - \mathcal{V}_t(N)^{-1/2} S_t + \mathcal{V}_t(N)^{-1/2} S_t - \mathcal{V}_t^{-1/2} S_t\right\|_H \\ &\leq \left\|\mathcal{V}_t(N)^{-1/2}\right\|_{op} \left\|\pi_N^\perp S_t\right\|_H + \left\|\mathcal{V}_t(N)^{-1/2} - \mathcal{V}_t^{-1/2}\right\|_{op} \|S_t\|_H \\ &\xrightarrow[N\to\infty]{} 0. \end{aligned}$$

as $\lim_{N\to\infty} \|\pi_N^\perp f\| = 0$ for any $f \in H$ of finite norm, and Lemma 2 tells us that $\|V_t - \pi_N V_t \pi_N\|_{op} \xrightarrow[N\to\infty]{} 0$, which in turn implies that $\|\mathcal{V}_t(N)^{-1/2} - \mathcal{V}_t^{-1/2}\|_{op} = \|(\rho\mathrm{id}_H + \pi_N V_t \pi_N)^{-1/2} - (\rho\mathrm{id}_H + V_t)^{-1/2}\|_H \xrightarrow[N\to\infty]{} 0$. Thus, we have shown the desired result.

The second part of the claim follows from a direct application of Fact 4 and rearranging.

■

As a final result in this appendix, we provide a proof of Corollary 1. This corollary allows for a more direct comparison of our results with those of Chowdhury and Gopalan [4]. Our proof is a simple generalization Lemma 1 in the aforementioned paper to the case of arbitrary regularization parameters.

Proof of Corollary 1. The first result is straightforward, and follows from the identity

$$\det(\mathrm{id}_H + \rho^{-1}V_t) = \det(I_t + \rho^{-1}K_t),$$

which we bring to attention in Section 2.

The second result follows from the following line of reasoning. Before proceeding, recall that $\Phi_t := (k(\cdot, X_1), \dots, k(\cdot, X_t))^\top$, $V_t = \Phi_t^\top \Phi_t$, $K_t = \Phi_t \Phi_t^\top$ and $S_t = \sum_{s=1}^t \epsilon_s k(\cdot, X_s) = \Phi_t^\top \epsilon_{1:t}$.

$$\begin{aligned}
\left\|(\rho\mathrm{id}_H + V_t)^{-1}S_t\right\|_H^2 &= \epsilon_{1:t}^\top \Phi_t(\rho\mathrm{id}_H + \Phi_t^\top\Phi_t)^{-1}\Phi_t^\top \epsilon_{1:t} \\
&= \epsilon_{1:t}^\top(\rho^{-1/2}\Phi_t)\left(\mathrm{id}_H + (\rho^{-1/2}\Phi_t)^\top(\rho^{-1/2}\Phi_t)\right)^{-1}(\rho^{-1/2}\Phi_t)^\top \epsilon_{1:t} \\
&= \epsilon_{1:t}^T \rho^{-1}\Phi_t\Phi_t^\top\left(I_t + \rho^{-1}\Phi_t\Phi_t^\top\right)^{-1}\epsilon_{1:t} \\
&= \epsilon_{1:t}^\top(\rho^{-1}K_t)(I_t + \rho^{-1}K_t)^{-1}\epsilon_{1:t} \\
&= \epsilon_{1:t}^\top(I_t + \rho K_t^{-1})^{-1}\epsilon_{1:t} \\
&= \left\|(I_t + \rho K_t^{-1})^{-1/2}\epsilon_{1:t}\right\|_2^2.
\end{aligned}$$

In the above, the second equality comes from pulling out a multiplicative factor of ρ form the center operator inverse. The third inequality comes from the famed "push through" identity. Lastly, the second to last equality comes from observing that (a) $\rho^{-1}K_t$ and $(I_t + \rho^{-1}K_t)^{-1}$ are simultaneously diagonalizable matrices and (b) for scalars, we have the identity $(1 + a^{-1})^{-1} = a(1 + a)^{-1}$. Thus, we have shown the desired result. ■

C Technical Lemmas for Theorem 2

In this appendix, we provide various technical lemmas needed for the proof of Theorem 2. We then follow these lemmas with a full proof of Theorem 2, which extends the sketch provided in the main body of the paper. Most of the following technical lemmas either already exist in the literature [4] or are extensions of what is known in the case of finite-dimensional, linear bandits [1]. We nonetheless provide self-contained proofs for the sake of completeness.

Lemma 4. *Let $(f_t)_{t\geq 1}$ be the sequence of functions defined in Algorithm 1, and assume Assumption 1 holds. Let $\delta \in (0, 1)$ be an arbitrary confidence parameter. Then, with probability at least $1 - \delta$, simultaneously for all $t \geq 1$, we have*

$$\left\|(V_t + \rho\mathrm{id}_H)^{1/2}(f_t - f^*)\right\|_H \leq \sigma\sqrt{2\log\left(\frac{1}{\delta}\sqrt{\det(\mathrm{id}_H + \rho^{-1}V_t)}\right)} + \rho^{1/2}D,$$

where we recall that the right hand side equals U_t.

Proof. First, observe that we have

$$\begin{aligned}
f_t - f^* &= (\rho\mathrm{id}_H + V_t)^{-1}\Phi_t^\top Y_{1:t} - f^* \\
&= (\rho\mathrm{id}_H + V_t)^{-1}\Phi_t^\top(\Phi_t f^* + \epsilon_{1:t}) - f^* \\
&= (\rho\mathrm{id}_H + V_t)^{-1}\Phi_t^\top(\Phi_t f^* + \epsilon_{1:t}) - f^* \pm \rho(\rho\mathrm{id}_H + V_t)^{-1}f^* \\
&= (\rho\mathrm{id}_H + V_t)^{-1}\Phi_t^\top\epsilon_{1:t} - \rho(\rho\mathrm{id}_H + V_t)^{-1}f^*.
\end{aligned}$$

Applying the triangle inequality to the above, we have

$$\begin{aligned}
\left\|(\rho\mathrm{id}_H + V_t)^{1/2}(f_t - f^*)\right\|_H &\leq \left\|(\rho\mathrm{id}_H + V_t)^{-1/2}\Phi_t^\top\epsilon_{1:t}\right\|_H + \rho\left\|(\rho\mathrm{id}_H + V_t)^{-1/2}f^*\right\|_H \\
&\leq \sigma\sqrt{2\log\left(\frac{1}{\delta}\sqrt{\det(\mathrm{id}_H + \rho^{-1}V_t)}\right)} + \rho^{1/2}D.
\end{aligned}$$

To justify the final inequality, we look at each term separately. For the first term, observe that $V_t = \rho\mathrm{id}_H + \sum_{s=1}^t k(\cdot, X_t)k(\cdot, X_t)^\top$ and $S_t := \Phi_t^\top\epsilon_{1:t} = \sum_{s=1}^t \epsilon_s k(\cdot, X_s)$. Thus, we are in the setting of Theorem 1, and thus have, with probability at least $1 - \delta$, simultaneously for all $t \geq 0$,

$$\left\|(\rho\mathrm{id}_H + V_t)^{-1/2}\Phi_t^\top\epsilon_{1:t}\right\|_H \leq \sigma\sqrt{2\log\left(\frac{1}{\delta}\sqrt{\det(\mathrm{id}_H + \rho^{-1}V_t)}\right)}.$$

For the second term, observe that (a) $\lambda_{\min}(\rho \mathrm{id}_H + V_t) \geq \rho$ and (b) by Assumption 1, we have $\|f^*\|_H \leq D$. Thus applying Holder's inequality, we have, deterministically

$$\rho \left\|(\rho \mathrm{id}_H + V_t)^{-1/2} f^*\right\|_H \leq \rho \left\|(\rho \mathrm{id}_H + V_t)^{-1/2}\right\|_{op} \|f^*\|_H \leq \rho^{1/2}\|f^*\|_H \leq \rho^{1/2} D.$$

These together give us the desired result.

∎

The following "elliptical potential" lemma, abstractly, aims to control the the growth of the squared, self-normalized norm of the selected actions. We more or less port the argument from Abbasi-Yadkori et al. [1], which provides an analogue in the linear stochastic bandit case. We just need to be mildly careful to work around the fact we are using Fredholm determinants.

Lemma 5. *For any $t \geq 1$, let V_t be the covariance operator defined in Algorithm 1, and let $\rho > 0$ be arbitrary. We have the identity*

$$\det(\mathrm{id}_H + \rho^{-1} V_t) = \prod_{s=1}^t \left(1 + \left\|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H^2\right).$$

In particular, if $\rho \geq 1 \vee L$, where L is the bound outlined in Assumption 2, we have

$$\sum_{s=1}^t \left\|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H^2 \leq 2 \log\det(\mathrm{id}_H + \rho^{-1} V_t).$$

Proof. Let $H_t \subset H$ be the finite-dimensional Hilbert space $H_t := \mathrm{span}\{k(\cdot, X_1), \dots, k(\cdot, X_t)\}$. Let $\det_{H_t}$ denote the determinant restricted to H_t, i.e. the map that acts on a (symmetric) operator $A : H_t \to H_t$ by $\det_{H_t}(A) := \prod_{s=1}^t \lambda_s(A)$, where $\lambda_1(A), \dots, \lambda_t(A)$ are the enumerated eigenvalues of A. Observe the identity

$$\det(\mathrm{id}_H + \rho^{-1} V_t) = \det_{H_t}(\mathrm{id}_{H_t} + \rho^{-1} V_t),$$

where we recall the determinant on the lefthand side is the Fredholm determinant, as defined in Section 2. Next, following the same line of reasoning as Abbasi-Yadkori et al. [1], we have

$$\begin{aligned}
&\det_{H_t}(\rho \mathrm{id}_{H_t} + V_t) \\
&= \det_{H_t}(\rho \mathrm{id}_{H_t} + V_{t-1}) \det_{H_t}\left(\mathrm{id}_{H_t} + (\rho \mathrm{id}_{H_t} + V_{t-1})^{-1/2} k(\cdot, X_t) k(\cdot, X_t)^\top (\rho \mathrm{id}_{H_t} + V_{t-1})^{-1/2}\right) \\
&= \det_{H_t}(\rho \mathrm{id}_{H_t} + V_{t-1}) \left(1 + \left\|(\rho \mathrm{id}_{H_t} + V_{t-1})^{-1/2} k(\cdot, X_t)\right\|_H^2\right) \\
&= \cdots \text{ (Iterating } t-1 \text{ more times)} \\
&= \det_{H_t}(\rho \mathrm{id}_H) \prod_{s=1}^t \left(1 + \left\|(\rho \mathrm{id}_{H_t} + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H^2\right) \\
&= \det_{H_t}(\rho \mathrm{id}_H) \prod_{s=1}^t \left(1 + \left\|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H^2\right),
\end{aligned}$$

where the last equality comes from realizing, for all $s \in [t]$, $\|(\rho \mathrm{id}_{H_t} + V_{s-1})^{-1/2} k(\cdot, X_s)\|_H = \|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\|_H$. Thus, rearranging yields

$$\det_{H_t}(\mathrm{id}_{H_t} + \rho^{-1} V_t) = \prod_{s=1}^t \left(1 + \left\|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H^2\right),$$

which yields the first part of the claim.

Now, to see the second part of the claim, observe the bound $x \leq 2\log(1+x), \forall x \in [0,1]$. Observing that, for all $s \in [t]$, $\left\|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H \leq 1$ when $\rho \geq 1 \vee L$, we have

$$
\begin{aligned}
\sum_{s=1}^{t} \left\|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H^2 &\leq 2 \sum_{s=1}^{t} \log\left(1 + \left\|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H^2\right) \\
&= 2\log\left(\prod_{s=1}^{t}\left(1 + \left\|(\rho \mathrm{id}_H + V_{s-1})^{-1/2} k(\cdot, X_s)\right\|_H^2\right)\right) \\
&= 2\log\det(\mathrm{id}_H + \rho^{-1} V_t),
\end{aligned}
$$

proving the second part of the lemma. ■

With the above lemmas, along with the concentration results provided by Theorem 1, we can provide a full proof for Theorem 2.

Proof of Theorem 2. We take the standard approach of (a) first bounding instantaneous regret and then (b) applying the Cauchy-Schwarz inequality to bound the aggregation of terms. To start, for any $t \in [T]$, define the "instantaneous regret" as $r_t := f^*(x^*) - f^*(X_t)$, where we recall $x^* := \arg\max_{x \in \mathcal{X}} f^*(x)$. By applying Lemma 4, we have with probability at least $1 - \delta$ that

$$
\begin{aligned}
r_t &= f^*(x^*) - f^*(X_t) \\
&\leq \widetilde{f}_t(X_t) - f^*(X_t) \\
&= \widetilde{f}_t(X_t) - f_{t-1}(X_t) + f_{t-1}(X_t) - f^*(X_t) \\
&= \langle \widetilde{f}_t - f_{t-1}, k(\cdot, X_t)\rangle_H - \langle f_{t-1} - f^*, k(\cdot, X_t)\rangle_H \\
&\leq \left\|(\rho \mathrm{id}_H + V_{t-1})^{-1/2} k(\cdot, X_t)\right\|_H \left(\left\|(\rho \mathrm{id}_H + V_{t-1})^{1/2}(\widetilde{f}_t - f_{t-1})\right\|_H + \left\|(\rho \mathrm{id}_H + V_{t-1})^{1/2}(f_{t-1} - f^*)\right\|_H\right) \\
&\leq 2U_{t-1} \left\|(\rho \mathrm{id}_H + V_{t-1})^{-1/2} k(\cdot, X_t)\right\|_H,
\end{aligned}
$$

where $\widetilde{f}_t$ and f_{t-1} are as in Algorithm 1. Note that, in the above, we apply Lemma 4 in obtaining the first inequality (which is the "optimism in the face of uncertainty" part of the bound), and additionally in obtaining the last inequality. The second to last inequality follows from applying Cauchy-Schwarz.

With the above bound, we can apply again the Cauchy-Schwarz inequality to see

$$
\begin{aligned}
R_T &= \sum_{t=1}^{T} r_t \leq \sqrt{T \sum_{t=1}^{T} r_t^2} \leq U_T \sqrt{2T \sum_{t=1}^{T} \left\|(\rho \mathrm{id}_H + V_{t-1})^{-1/2} k(\cdot, X_t)\right\|_H^2} \\
&\leq U_T \sqrt{2T \log\det(\mathrm{id}_H + \rho^{-1} V_T)} \\
&= \left(\sigma \sqrt{2\log\left(\frac{1}{\delta}\sqrt{\det(\mathrm{id}_H + \rho^{-1} V_T)}\right)} + \rho^{1/2} D\right) \sqrt{2T \log\det(\mathrm{id}_H + \rho^{-1} V_T)} \\
&\leq \left(\sigma\sqrt{2\log(1/\delta)} + \sigma\sqrt{2\gamma_T(\rho)} + \rho^{1/2} D\right) \sqrt{4T\gamma_T(\rho)} \\
&= \sigma\gamma_T(\rho)\sqrt{8T} + D\sqrt{4\rho\gamma_T(\rho)T} + \sigma\sqrt{8T\log(1/\delta)} \\
&= O\left(\gamma_T(\rho)\sqrt{T} + \sqrt{\rho\gamma_T(\rho)T}\right).
\end{aligned}
$$

In the above, the second inequality follows from the second part of Lemma 5, the following equality follows from substituting in U_T, and the final inequality follows from the definition of the maximum information gain $\gamma_T(\rho)$ and the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for all $a, b \geq 0$. The last, big-Oh bound is straightforward. With this, we have proven the first part of the theorem.

Now, suppose the kernel k experiences (C, β)-polynomial eigendecay. Then, by Fact 2, we know that

$$\begin{aligned}\gamma_T(\rho) &\leq \left(\left(\frac{CB^2T}{\rho}\right)^{1/\beta}\log^{-1/\beta}\left(1+\frac{LT}{\rho}\right)+1\right)\log\left(1+\frac{LT}{\rho}\right)\\ &= \widetilde{O}\left(\left(\frac{T}{\rho}\right)^{1/\beta}\right).\end{aligned}$$

We aim to set $\rho = \left(\frac{T}{\rho}\right)^{1/\beta}$, which occurs when $\rho = T^{\frac{1}{1+\beta}}$. When this happens, we have

$$\left(\frac{T}{\rho}\right)^{1/\beta}\sqrt{T} = T^{\frac{1}{1+\beta}+\frac{1}{2}} = T^{\frac{3+\beta}{2+2\beta}}.$$

Applying this, we have that

$$\begin{aligned}R_T &= O\left(\gamma_T(\rho)\sqrt{T} + \sqrt{\rho\gamma_T(\rho)T}\right)\\ &= \widetilde{O}\left(T^{\frac{3+\beta}{2+2\beta}}\right),\end{aligned}$$

which, in particular, is sublinear for any $\beta > 1$. Thus, we are done.

■