References

[1] Yasin Abbasi-Yadkori. Online learning for linearly parametrized control problems. 2013.

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

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

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

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

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

[7] 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.

[8] 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.

[9] Victor H de la Peña, Michael J Klass, and Tze Leung Lai. Theory and applications of multivariate self-normalized processes. *Stochastic Processes and their Applications*, 119(12):4210–4227, 2009.

[10] Audrey Durand, Odalric-Ambrym Maillard, and Joelle Pineau. Streaming kernel regression with provably adaptive mean, variance, and regularization. *The Journal of Machine Learning Research*, 19(1):650–683, 2018.

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

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

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

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

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

[16] David Janz. *Sequential decision making with feature-linear models.* PhD thesis, 2022.

[17] 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.

[18] Tor Lattimore. A lower bound for linear and kernel regression with adaptive covariates. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2095–2113. PMLR, 2023.

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

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

[21] 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.

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

[23] 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.

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

[25] 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.

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

[27] 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.

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

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

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

[31] 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.

[32] 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.

[33] 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.

[34] 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.

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

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

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

[38] 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. [31], 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 [5] 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. [31]. Unfortunately, the arguments introduced by Chowdhury and Gopalan [5] 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. Lastly, Janz [16] are able to obtain sublinear regret guarantees for certain parameter settings of the Matérn kernel — in particular in settings where the eigenfunctions of the Hilbert-Schmidt operator associated with the kernel are uniformly bounded independent of scale (Definition 28 in the cited work).

There are many other algorithms that have been created for kernelized bandits. Janz et al. [17] 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 [28] 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 [5], 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. [34], 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. [25] up to logarithmic factors.

Intimately tied to the kernelized bandit problem is the information-theoretic quantity of maximum information gain [6, 31], 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. [32]. 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. [32] can be coupled with the regret analysis in Chowdhury and Gopalan [5] 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 [26, 38] 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. [32].

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. [2], 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. [7], which provide self-normalized concentration for a wide variety of martingale-related processes, several of which have been recently improved [14]. In a work that is largely overlooked in the kernel bandit community, Abbasi-Yadkori [1] extend their concentration result from Abbasi-Yadkori et al. [2] to separable Hilbert spaces by using advanced functional analytic machinery. The bound we present in this work is equivalent to the aforementioned bound in separable Hilbert spaces — we provide an independent, simpler proof that avoids needing advanced tools from functional analysis. Perhaps the best-known result on concentration in Hilbert spaces is that of Chowdhury and Gopalan [5], who extend the results of Abbasi-Yadkori et al. [2] to the kernel setting using a "double mixture" technique, allowing them to construct self-normalized concentration inequalities for infinite-dimensional processes in RKHS's. This bound has historically been used in analyzing kernel bandit algorithms, although as we show in this work the bound of Abbasi-Yadkori [1] (which we independently derive in Theorem 1) is perhaps better suited for online kernelized learning problems.

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 $\varphi_1, \dots, \varphi_N$ such that*

$$H := \mathrm{span}\left\{\varphi_1, \dots, \varphi_N\right\}.$$

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 $\tau f := (\theta_1, \dots, \theta_N)^\top \in \mathbb{R}^N$. 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" $\|A\|_1 := \sum_{n=1}^{\infty} |\lambda_n(A)|$ [20]. 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 [11], 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 Theorem 1 (and thus Corollary 3.5 of Abbasi-Yadkori [1]) with those of Chowdhury and Gopalan [5]. 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 [5] or are extensions of what is known in the case of finite-dimensional, linear bandits [2]. 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. [2], 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 $H_t := \mathrm{span}\{k(\cdot, X_1), \ldots, k(\cdot, X_t)\}$. 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 $\lambda_1(A), \ldots, \lambda_t(A)$ 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. [2], 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 $\rho \asymp \left(\frac{T}{\rho}\right)^{1/\beta}$, which occurs when $\rho = O(T^{\frac{1}{1+\beta}})$. 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.

■