Near-Optimal Regret in Adversarial Kernel Bandits

Yu-Jie Zhang*
University of Washington
yujiez7@cs.washington.edu

Hao Qiu*
National University of Singapore
qiuhaosai@gmail.com

Jonathan Scarlett
National University of Singapore
scarlett@comp.nus.edu.sg

Kevin Jamieson
University of Washington
jamieson@cs.washington.edu

Abstract

We study the adversarial kernel bandit problem, in which the loss at each round is induced by an arbitrary bounded element of a reproducing kernel Hilbert space (RKHS). We propose an exponential-weights algorithm built on a regularized importance-weighted loss estimator, together with an explicit correction term that cancels the bias introduced by the regularization. Our main result bounds the regret by $\widetilde{\mathcal{O}}\big(\sqrt{T\, d_*(\lambda)\, \log|\mathcal{X}|}\big)$, where $d_*(\lambda)$ is a widely-adopted notion of effective dimension that captures the complexity of the kernel. Up to logarithmic factors, this matches the known rate achieved in the related stochastic kernel bandit problem. A notable application is the Matérn(ν, d) kernel with smoothness parameter ν on $\mathbb{R}^d$, for which our bound specializes to $\widetilde{\mathcal{O}}\big(T^{(\nu+d)/(2\nu+d)}\big)$, improving over the best-known prior rate of Chatterji et al. [2019] while simultaneously removing the rank-one adversary assumption required by their analysis. Moreover, this rate is the same as the known optimal rate for stochastic kernel bandits, and also matches a lower bound from concurrent work up to a $\log T$ factor.

1 Introduction

Online decision-making under bandit feedback (e.g., see Cesa-Bianchi and Lugosi [2006], Hazan [2016], Lattimore and Szepesvári [2020]) is a fundamental problem in machine learning, with applications ranging from recommendation systems [Li et al., 2010] to experimental design [Srinivas et al., 2010]. In the classical adversarial bandit setting, each arm has an arbitrary loss sequence, a learner repeatedly selects actions and observes only the loss incurred by the chosen action, with the goal of competing against the best fixed action in hindsight. A well-studied extension of this framework is adversarial linear bandits [Awerbuch and Kleinberg, 2004, Dani et al., 2007, Bubeck et al., 2012], where the loss of an action at each time step is assumed to be a linear function of its features. While this linear structure enables efficient learning and strong regret guarantees, it is often too restrictive to capture complex, nonlinear relationships present in modern applications.

To address this limitation, we study the *adversarial kernel bandits* problem [Chatterji et al., 2019, Takemori and Sato, 2021], in which losses are modeled as functions in a reproducing kernel Hilbert space (RKHS). This framework generalizes linear bandits by allowing nonlinear losses with a potentially infinite-dimensional feature representation. It serves as an adversarially time-varying counterpart to the well-studied stochastic kernel bandit problem, which is typically viewed through the lens of Bayesian optimization via Gaussian process methods [Garnett, 2023]. The added expressivity of RKHS functions (compared to linear functions) can be highly valuable in diverse applications. This is evidenced by the extensive applications of the stochastic formulation, such as hyperparameter

*Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

tuning Chen et al. [2018], molecule design Korovina et al. [2020], environmental monitoring Marchant and Ramos [2012], and many more. The adversarial formulation is of interest in scenarios where we need algorithms robust to unknown, potentially arbitrary fluctuations in the environment, and also has a strong theoretical precedent based on the extensive work on adversarial (non-kernel) bandits.

The additional expressiveness of this setup introduces significant challenges: under bandit feedback, the learner observes only a scalar loss per round, making it difficult to estimate loss functions in a potentially infinite-dimensional RKHS. Despite its practical and theoretical importance, existing regret guarantees for adversarial kernel bandits remain incomplete for some commonly used kernels. The closest prior work, Chatterji et al. [2019], considers kernels whose Mercer eigenvalues exhibit either exponential or polynomial decay. In the former case, which includes the squared-exponential (SE) kernel, they obtain a near-optimal $\widetilde{\mathcal{O}}(\sqrt{T})$ rate. Under polynomial eigendecay, however, their bound is no longer tight. This looseness is particularly evident for the Matérn(ν, d) kernel, a standard choice in Gaussian process modeling and Bayesian optimization. For this kernel, their rate becomes $\widetilde{\mathcal{O}}(T^{(2\nu+d)/(4\nu)})$, which is far from the $\widetilde{\mathcal{O}}(T^{(\nu+d)/(2\nu+d)})$ rate known to be optimal in the analogous stochastic kernel bandit problem. In fact, it fails to be sublinear in T when the smoothness ν is small relative to the ambient dimension d, namely $\nu \leq d/2$. This gap motivates our central question:

Can one design an adversarial kernel bandit algorithm whose regret matches the optimal stochastic kernel bandit rate (within log factors)?

Contributions. This paper makes substantial progress on this question, including an affirmative answer for the Matérn kernel. We propose a simple and principled algorithm for adversarial kernel bandits that combines exponential weights with a regularized importance-weighted loss estimator in the RKHS. The regularization keeps the estimator bounded in an infinite-dimensional feature space, while an explicit correction term compensates for the bias introduced by regularization. This correction is central to the analysis: it cancels the comparator-side bias and leaves only a learner-side residual that can be controlled through a well-known notion of *effective dimension* of the kernel.

Our main result is a regret bound of the form $\widetilde{\mathcal{O}}(\sqrt{T\, d_*(\lambda)\, \log|\mathcal{X}|})$, where $d_*(\lambda)$ is the effective dimension of the kernel and $|\mathcal{X}|$ is the size of the action set. Up to logarithmic factors, this matches the rate achievable in the analogous stochastic kernel bandit problem when expressed via the maximum information gain (as defined in [Srinivas et al., 2010] among many others). In particular, when specialized to the Matérn(ν, d) kernel, our bound yields the regret $\widetilde{\mathcal{O}}(T^{(\nu+d)/(2\nu+d)})$, improving over the best-known prior rate $\widetilde{\mathcal{O}}(T^{(2\nu+d)/(4\nu)})$ of Chatterji et al. [2019] while simultaneously removing their restrictive rank-one adversary assumption. This rate further matches the lower bound appearing in a concurrent work [Iwazaki, 2026] up to logarithmic factors in T. For the SE kernel, our bound recovers the near-optimal $\widetilde{\mathcal{O}}(\sqrt{T})$ rate already established by Chatterji et al. [2019], while again removing the rank-one assumption. Our bound also yields corollaries under suitably-defined polynomial and exponential eigendecay conditions on the kernel spectrum: Under polynomial eigenvalue decay with exponent $\beta > 1$, our algorithm achieves regret $\widetilde{O}\big(T^{(\beta+1)/(2\beta)}\big)$, and under exponential eigenvalue decay, the regret improves to $\widetilde{O}(\sqrt{T})$. Although the polynomial-decay rate has the same expression in the exponent as a lower bound from Chatterji et al. [2019], the two results are proved under different settings and are not directly comparable; see Appendix B.1 for details.

Concurrent work. In independent concurrent work, Iwazaki [2026] closes the same gap with an algorithm and analysis close to ours. Both works use exponential weights with a regularized importance-weighted loss estimator, together with an additive correction term that cancels the bias introduced by the regularization. There are some minor differences in the problem setup and complexity measures used, but these are essentially inconsequential; see Appendix B.2 for the details.

Beyond this shared core, Iwazaki [2026] additionally provides a complete analysis of a Nyström-approximation variant, which we discuss only briefly as a computational simplification, and establishes algorithm-independent lower bounds for the SE and Matérn kernels.

2 Related Work

Stochastic kernel bandits. There have been extensive studies on the stochastic kernel bandit problem, i.e., a fixed underlying function and random observation noise. Here we focus only on the most

fundamental and/or relevant works. Prominent early works studied UCB-style algorithms Srinivas et al. [2010], Chowdhury and Gopalan [2017] and established regret upper bounds of $\widetilde{\mathcal{O}}(\gamma_T\sqrt{T})$, where γ_T is an *information gain* quantity capturing the kernel complexity, and is closely related to the *effective dimension* (e.g., see [Calandriello et al., 2019, Prop. 5]). This was subsequently improved to $\widetilde{\mathcal{O}}(\sqrt{T\gamma_T})$ via the SupKernelUCB algorithm Valko et al. [2013] and subsequently by several other methods Li and Scarlett [2022], Camilleri et al. [2021], Salgia et al. [2021]. For the widely-adopted squared exponential (SE) and Matérn kernels, tight upper bounds on γ_T were derived in Vakili et al. [2021], and these bounds in turn establish that the $\widetilde{\mathcal{O}}(\sqrt{T\gamma_T})$ regret upper bound is tight (to within logarithmic factors) in view of matching lower bounds Scarlett et al. [2017], Cai and Scarlett [2021]. See also the recent work [Janz et al., 2026] for refined information gain bounds, and [Ma and Scarlett, 2026] for a general $\widetilde{\Omega}(\sqrt{T\gamma_T})$ lower bound beyond SE and Matérn. Under polynomial eigenvalue decay with exponent $\beta > 1$, the information gain scales as $\widetilde{\mathcal{O}}(T^{1/\beta})$, yielding a regret bound of order $\widetilde{\mathcal{O}}(T^{\frac{\beta+1}{2\beta}})$, whereas for exponential eigenvalue decay the regret becomes $\widetilde{\mathcal{O}}(\sqrt{T})$. Various works have also studied non-stationary stochastic kernel bandits (see Hong et al. [2023], Cai and Scarlett [2025], Iwazaki and Takeno [2025] and the references therein), but their setup is fundamentally different from adversarial kernel bandits; they seek to minimize *dynamic regret* (i.e., track the maximum at every time step) and they crucially place explicit bounds on how much the function can vary.

Adversarial linear bandits. The adversarial linear bandit problem was introduced to the online learning literature by Awerbuch and Kleinberg [2004]. In this setting, a learner repeatedly selects an action from a decision set and incurs a loss that is linear in the chosen action, while the loss vectors may be chosen adversarially. Subsequent work established algorithms for compact convex action sets with regret scaling as $\widetilde{\mathcal{O}}(d\sqrt{T})$, where d denotes the dimension of the linear bandit problem, using variants of exponential weights combined with additional exploration [Dani et al., 2007, Cesa-Bianchi and Lugosi, 2012, Bubeck et al., 2012]. Abernethy et al. [2008] proposed the first computationally efficient algorithm that achieves $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ regret using the FTRL framework. Other efficient methods achieving $\widetilde{\mathcal{O}}(d\sqrt{T})$ regret include OMD for specific action sets [Bubeck et al., 2012] and volumetric-spanner exploration with membership-oracle access [Hazan and Karnin, 2016].

Adversarial kernel bandits. The closest prior work on adversarial kernel bandits is Chatterji et al. [2019], who study a closely related setting under a rank-one adversary. Their guarantees crucially depend on the eigenvalue decay of the kernel. When the Mercer eigenvalues decay exponentially, they obtain the tight $\widetilde{\mathcal{O}}(\sqrt{T})$ regret. However, the situation is less satisfactory under *polynomial* eigendecay. Under polynomial eigendecay with exponent $\beta > 1$, the $T^{\beta/(2(\beta-1))}$ dependence in their upper bound does not match the $T^{(\beta+1)/(2\beta)}$ dependence in their lower bound (see also Appendix B for some important caveats regarding their lower bound). For the Matérn(ν, d) kernel, the upper bound becomes $\widetilde{\mathcal{O}}(T^{(2\nu+d)/(4\nu)})$, which falls short of the optimal stochastic rate $\widetilde{\mathcal{O}}(T^{(\nu+d)/(2\nu+d)})$, and is only sublinear when $d < 2\nu$. A subsequent work [Takemori and Sato, 2021] removed the rank-one adversary assumption by reducing the problem to a misspecified linear bandit via approximation theory. Their algorithm thus extends to more general adversaries, but the resulting $\widetilde{\mathcal{O}}(T^{(\nu+d)/(2\nu)})$ rate is even further from optimality for the Matérn kernel.

Our effective-dimension bound improves on the upper bounds of Chatterji et al. [2019] in two complementary ways. Applied directly to specific kernels, it recovers the same $\widetilde{\mathcal{O}}(\sqrt{T})$ rate on the SE kernel, and yields an improved $\widetilde{\mathcal{O}}(T^{(\nu+d)/(2\nu+d)})$ rate on the Matérn(ν, d) kernel, sharpening their $\widetilde{\mathcal{O}}(T^{(2\nu+d)/(4\nu)})$ bound. Moreover, under a strengthened version of polynomial eigendecay that holds uniformly over all distributions on $\mathcal{X}$, our bound improves their general rate from $\widetilde{\mathcal{O}}(T^{\beta/(2(\beta-1))})$ to $\widetilde{\mathcal{O}}(T^{(\beta+1)/(2\beta)})$. We re-iterate that, as we detail in Appendix B, their lower bound is not directly comparable to our upper bound despite both having $T^{(\beta+1)/(2\beta)}$ dependence.

Adversarial contextual kernel bandits. Another relevant work is Neu et al. [2024], who study an adversarial kernelized *contextual* bandit with $|\mathcal{X}|$ actions and stochastic contexts, where the kernel is defined over the context space and a separate RKHS element is associated with each action. Although their setup is not the same as ours,[2] their algorithm can potentially be applied here as well, and it attains the optimal T-exponent under polynomial eigendecay. However, because it is built on FTRL

[2]The main difference is that their setup uses kernel modeling for generalizing across contexts, not across actions.

with a *log-barrier* regularizer, it incurs a polynomial dependence on $|\mathcal{X}|$ rather than the $\log|\mathcal{X}|$ dependence of our method. Thus, there is an *exponentially large gap* between the two dependencies, and this is compounded by the fact that a standard covering argument of a continuous domain suggests choosing $|\mathcal{X}| = T^{O(d)}$ [Janz et al., 2020, Ma and Scarlett, 2026], making the higher $|\mathcal{X}|$ dependence prohibitive. We defer further discussion to Section 4.1.

3 Problem Setup

This section introduces the problem setup for adversarial kernel bandits, a generalization of adversarial linear bandits in which the loss functions belong to an RKHS. We begin by stating the kernel loss model and assumptions, and then present the learning protocol and define the notion of regret.

Kernels and RKHS. We focus on a finite (possibly large) action space $\mathcal{X} \subset \mathbb{R}^d$, but our results easily extend to continuous domains via a covering argument. For $\mathcal{X} = [0,1]^d$, taking a grid of size $T^{O(d)}$ leads to negligible discretization error under mild assumptions on the kernel that are satisfied by the SE and Matérn classes [Janz et al., 2020, Ma and Scarlett, 2026].

We are given a symmetric positive definite kernel $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, and denote by $\mathcal{H}$ its associated reproducing kernel Hilbert space (RKHS). The RKHS $\mathcal{H}$ is a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$, equipped with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\|\cdot\|_{\mathcal{H}}$. It satisfies the reproducing property: for every $f \in \mathcal{H}$ and $\mathbf{x} \in \mathcal{X}$, $f(\mathbf{x}) = \langle f, \mathcal{K}(\cdot, \mathbf{x}) \rangle_{\mathcal{H}}$. The kernel also induces a feature map $\Phi : \mathcal{X} \to \mathcal{H}$ such that $\mathcal{K}(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathcal{H}}$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. The feature vector $\Phi(\mathbf{x})$ may be infinite-dimensional. When $\mathcal{K}$ admits a Mercer decomposition under a reference measure, an explicit feature map can be constructed by setting $\Phi_j(\mathbf{x}) = \sqrt{\mu_j}\phi_j(\mathbf{x})$ for each $j \geq 1$, where $\Phi_j(\mathbf{x})$ denotes the j-th coordinate of $\Phi(\mathbf{x})$, and $\{\mu_j\}_{j \geq 1}$ and $\{\phi_j\}_{j \geq 1}$ are the eigenvalues and eigenfunctions of the associated kernel integral operator. We refer the reader to [Chatterji et al., 2019, Section 2.1] and [Kanagawa et al., 2018] for further background on RKHS functions.

We use the following operator notation on $\mathcal{H}$. For $\mathbf{v}, \mathbf{w} \in \mathcal{H}$, we use the matrix-style notation $\mathbf{v}\mathbf{w}^\top$ to denote the rank-one linear operator on $\mathcal{H}$ defined by $(\mathbf{v}\mathbf{w}^\top)\mathbf{u} = \mathbf{v}\langle \mathbf{w}, \mathbf{u} \rangle_{\mathcal{H}}$ for all $\mathbf{u} \in \mathcal{H}$. The identity operator on $\mathcal{H}$ is denoted by $I_{\mathcal{H}}$. For a bounded, self-adjoint, positive semidefinite operator $A : \mathcal{H} \to \mathcal{H}$ and any $\mathbf{w} \in \mathcal{H}$, we define $\|\mathbf{w}\|_A := \sqrt{\langle \mathbf{w}, A\mathbf{w} \rangle_{\mathcal{H}}}$, which is a seminorm in general and a norm when A is positive definite. We write $A \preceq B$ if $B - A$ is positive semidefinite.

Learning protocol. Throughout the paper, we use $\mathbb{Z}_+$ to denote the set of positive integers and $[m]$ to denote the set $\{1, 2, \ldots, m\}$ for $m \in \mathbb{Z}_+$. At each round $t \in [T]$, the learner selects an action $\mathbf{x}_t \in \mathcal{X}$ while the environment simultaneously selects a vector $\mathbf{w}_t \in \mathcal{H}$. The learner then incurs the loss of its decision $\ell_t(\mathbf{x}_t) = \langle \Phi(\mathbf{x}_t), \mathbf{w}_t \rangle_{\mathcal{H}}$. Afterward, the learner observes only this scalar loss $\ell_t(\mathbf{x}_t)$ and then proceeds to the next round. We consider an oblivious adversary, in which the sequence $\{\mathbf{w}_t\}_{t=1}^T$ is fixed before the interaction begins and does not depend on the learner's realized actions or internal randomness. The goal of the learner is to minimize the expected regret:

$$\mathcal{R}_T \;:=\; \mathbb{E}\left[\sum_{t=1}^{T} \langle \Phi(\mathbf{x}_t), \mathbf{w}_t \rangle_{\mathcal{H}} \;-\; \sum_{t=1}^{T} \langle \Phi(\mathbf{x}_*), \mathbf{w}_t \rangle_{\mathcal{H}}\right],$$

where $\mathbf{x}_* = \arg\min_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^{T} \langle \Phi(\mathbf{x}), \mathbf{w}_t \rangle_{\mathcal{H}}$ denotes the best fixed action in hindsight and the expectation is over the randomness in the algorithm. We also impose the following standard regularity assumptions, which are common in the literature [Chatterji et al., 2019, Li and Scarlett, 2022].

Assumption 1 (Bounded Kernel and Adversary)**.** There exists a constant $B > 0$ such that $\|\mathbf{w}_t\|_{\mathcal{H}} \leq B$ for all $t \in [T]$. The kernel is bounded by $\mathcal{K}(\mathbf{x}, \mathbf{x}) \leq 1$ for any $\mathbf{x} \in \mathcal{X}$.[3]

Remark 1. We note that the adversarial kernel bandit model of Chatterji et al. [2019] imposes a *rank-one* restriction on the adversary, which requires $\mathbf{w}_t = \Phi(\mathbf{x}_t)$ for some $\mathbf{x}_t \in \mathcal{X}$. This is quite restrictive, and to our knowledge it has not been considered in any works on stochastic kernel bandits. Our framework removes this restriction, allowing $\mathbf{w}_t$ to be an arbitrary element of the RKHS $\mathcal{H}$ subject only to a bounded norm condition. See Figure 1 for an illustration. The rank-one restriction in Chatterji et al. [2019] is tied to their projection-based loss estimator, whereas our use of a regularized loss estimator frees us from this requirement.

[3]This assumption generalizes to $\mathcal{K}(\mathbf{x}, \mathbf{x}) \leq G^2$ for any fixed G, but we focus on $\mathcal{K}(\mathbf{x}, \mathbf{x}) \leq 1$ to reduce notation and since this normalization is very standard in the stochastic kernel bandit literature (e.g., see Srinivas et al. [2010]).

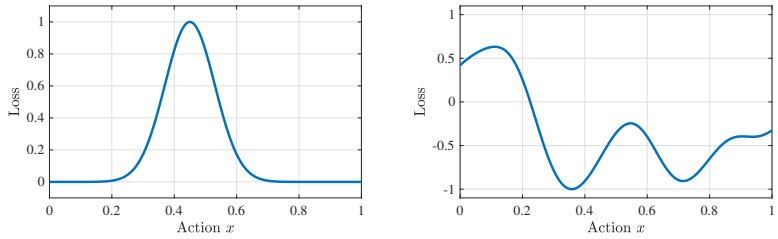

Figure 1: Example of a rank-one function (Left) and a general RKHS-norm bounded function (Right) under the squared exponential kernel, which has exponential eigenvalue decay.

4 Exponential Weights with Regularized Estimator

In this section, we develop a regularized variant of Exp2 for adversarial kernel bandits and derive its regret in terms of the effective dimension. Up to logarithmic factors, this matches the analogous stochastic rate, and specializes to near-optimal rates for commonly used kernels such as the SE and Matérn kernels. We first discuss the challenges of extending Exp2 from linear to kernel losses and the limitations of existing approaches, then present the algorithm and its regret guarantee.

4.1 Exp2 and the Kernel Obstacle

A central challenge in generalizing from linear to kernel bandits is that the parameter $\mathbf{w}_t$ may be infinite-dimensional. This added expressiveness can make the standard importance-weighted loss estimator for linear bandits take values of unbounded magnitude, preventing its direct use here.

To make this concrete, we briefly recall the Exp2 algorithm developed for linear bandits. At each round t, Exp2 maintains a distribution p_t over the action set, samples $\mathbf{x}_t \sim p_t$, and observes the scalar loss $y_t = \langle \mathbf{w}_t, \mathbf{x}_t \rangle$. From this single observation, it constructs the importance-weighted estimator

$$\widehat{\mathbf{w}}_t \;=\; y_t\, \Sigma_t^{-1} \mathbf{x}_t, \qquad \Sigma_t \;=\; \mathbb{E}_{\mathbf{x}\sim p_t}[\mathbf{x}\mathbf{x}^\top].$$

Exp2 then updates p_t by exponential weights on the loss proxy $\widehat{\ell}_t(\mathbf{x}) := \langle \widehat{\mathbf{w}}_t, \mathbf{x} \rangle$. The Exp2 algorithm requires a positive lower bound on the minimum eigenvalue of Σ_t, which is ensured by mixing p_t with a suitable exploration distribution (e.g., John's distribution), and yields the optimal regret $\widetilde{\mathcal{O}}(\sqrt{d\,T \log |\mathcal{X}|})$ [Dani et al., 2007, Bubeck et al., 2012]. Translating this construction to the kernel setting amounts to replacing $\mathbf{x}$ by $\Phi(\mathbf{x})$ throughout, with $\Sigma_t = \mathbb{E}_{\mathbf{x}\sim p_t}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$ now acting on the RKHS $\mathcal{H}$. However, in an infinite-dimensional RKHS, the covariance operator Σ_t may have eigenvalues arbitrarily close to zero, which can make the loss estimator unbounded.

Existing biased estimators and limitations. To circumvent this issue, existing work [Chatterji et al., 2019, Neu et al., 2024] resorts to biased estimators that trade some bias for variance control. Chatterji et al. [2019] construct a finite-dimensional proxy kernel by projecting the Mercer expansion of $\mathcal{K}$ to its top-m eigenfunctions, and run Exp2 in this m-dimensional space. While the truncation restores invertibility of the covariance and makes Exp2 applicable, the bias it introduces is hard to balance tightly against the estimator variance, leaving room for improvement in the resulting bound. When the kernel satisfies a polynomial eigenvalue decay condition with exponent $\beta > 1$, their analysis yields a regret bound of order $\widetilde{\mathcal{O}}(T^{\beta/(2(\beta-1))})$. This gives a suboptimal rate for Matérn kernels and leaves a gap to the lower bound $\Omega(T^{(\beta+1)/(2\beta)})$ established in the same paper. Moreover, to control the bias introduced by their projection-based estimator, Chatterji et al. [2019] restrict the adversary to rank-one losses $\mathbf{w}_t = \Phi(\mathbf{z}_t)$, so that the loss can be written as $\ell_t(\mathbf{x}) = \mathcal{K}(\mathbf{x}, \mathbf{z}_t)$. Our approach removes this restriction and handles arbitrary bounded RKHS losses $\mathbf{w}_t \in \mathcal{H}$ with $\|\mathbf{w}_t\|_{\mathcal{H}} \le B$.

As outlined in the introduction, Neu et al. [2024] consider the kernelized contextual bandit setting and design a different biased estimator via Kernel Geometric Resampling, a finite-step truncation of a Neumann-series approximation to Σ_t^{-1}. Even with this approximation, the small eigenvalues of Σ_t still leave the loss estimator with potentially large magnitude, so they replace exponential weights with a FTRL update under log-barrier regularization, which is more robust to unbounded loss estimators. Their bound matches the optimal T-exponent $T^{(\beta+1)/(2\beta)}$ under polynomial eigenvalue decay. Although their setting is not the same as ours, their estimator can also be applied to our setting,

Algorithm 1 Exponential Weights with Regularized Estimator

Input: Action set $\mathcal{X}$, kernel $\mathcal{K}$, norm bound $B > 0$, learning rate $\eta > 0$, mixing coefficient $\gamma \in (0, 1]$, regularization coefficient $\lambda > 0$.

Initialize: Let q_1 be a uniform distribution over $\mathcal{X}$, and define the exploration distribution

$$\nu_G \in \operatorname*{arg\,min}_{\nu \in \Delta(\mathcal{X})} \max_{\mathbf{x} \in \mathcal{X}} \|\Phi(\mathbf{x})\|^2_{(\Sigma(\nu)+(\lambda/\gamma)I_{\mathcal{H}})^{-1}}.$$

for $t = 1, 2, \ldots, T$ **do**
  Set $p_t = (1-\gamma)q_t + \gamma\nu_G$, draw $\mathbf{x}_t \sim p_t$, and observe $y_t = \langle \Phi(\mathbf{x}_t), \mathbf{w}_t \rangle_{\mathcal{H}}$.
  Compute the regularized estimator $\widehat{\mathbf{w}}_t$ via Eq. (1).
  Construct the corrected loss proxy $\widehat{\ell}_t$ via Eq. (3).
  Update $q_{t+1}(\mathbf{x}) \propto q_t(\mathbf{x}) \exp\big(-\eta\, \widehat{\ell}_t(\mathbf{x})\big)$ for all $\mathbf{x} \in \mathcal{X}$.
end for

but the use of log-barrier inherently incurs at least a $\sqrt{|\mathcal{X}|}$ dependence on the action-set cardinality, which is highly unfavorable in the kernel bandit setting as we discussed at the end of Section 2.

4.2 Our Algorithm

Our method follows the Exp2 template with two key modifications targeting the obstacles identified above: a *regularized estimator* that keeps the loss estimator bounded despite the vanishing eigenvalues of Σ_t, and a *correction term* that compensates for the bias introduced by regularization.

Regularized estimator. We replace the design covariance $\Sigma_t = \mathbb{E}_{\mathbf{x} \sim p_t}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$ with its ridge-regularized version $\Sigma_t^\lambda := \Sigma_t + \lambda I_{\mathcal{H}}$ and form the importance-weighted estimator

$$\widehat{\mathbf{w}}_t \;=\; y_t \cdot (\Sigma_t^\lambda)^{-1}\Phi(\mathbf{x}_t), \qquad y_t = \langle \Phi(\mathbf{x}_t), \mathbf{w}_t \rangle_{\mathcal{H}}. \tag{1}$$

The added $\lambda I_{\mathcal{H}}$ term enforces a positive lower bound on the spectrum of Σ_t^λ, keeping $\widehat{\mathbf{w}}_t$ bounded even when Σ_t has eigenvalues close to zero; the price paid for this is bias. Letting $\mathcal{F}_{t-1}$ denote the filtration of the history before round t, a direct calculation (deferred to Lemma 4 in the appendix) shows that the conditional bias of $\widehat{\mathbf{w}}_t$ takes the explicit form

$$\mathbf{w}_t - \mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] \;=\; \lambda\,(\Sigma_t^\lambda)^{-1}\mathbf{w}_t. \tag{2}$$

This explicit form is what the correction term introduced next will be designed to cancel.

Correction term. To compensate for this regularization bias, we define the loss proxy

$$\widehat{\ell}_t(\mathbf{x}) \;=\; \langle \widehat{\mathbf{w}}_t, \Phi(\mathbf{x}) \rangle_{\mathcal{H}} - c_t(\mathbf{x}), \qquad c_t(\mathbf{x}) \;=\; B\sqrt{\lambda}\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}}. \tag{3}$$

The specific form of the correction term c_t comes directly from the bias formula (2). Applying Cauchy–Schwarz together with $\|\mathbf{w}_t\|_{\mathcal{H}} \le B$ to the right-hand side of (2) yields, for every action $\mathbf{x}$,

$$\big|\langle \Phi(\mathbf{x}), \mathbf{w}_t - \mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] \rangle_{\mathcal{H}}\big| \;\le\; B\sqrt{\lambda}\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}} \;=\; c_t(\mathbf{x}),$$

so $c_t(\mathbf{x})$ is a pointwise upper bound on the bias of $\langle \widehat{\mathbf{w}}_t, \Phi(\mathbf{x}) \rangle_{\mathcal{H}}$, and subtracting it ensures that the estimator $\widehat{\ell}_t(\mathbf{x})$ underestimates the true loss in expectation, thereby encouraging exploration as in UCB methods. Specifically, as detailed in Lemma 2, c_t offsets the comparator-side bias $B_T^{\mathbf{x}_*} := \mathbb{E}\big[\sum_{t=1}^T \langle \Phi(\mathbf{x}_*), \mathbf{w}_t - \widehat{\mathbf{w}}_t \rangle_{\mathcal{H}}\big]$, which could be large when $\mathbf{x}_*$ is poorly explored. This makes the resulting residual non-positive, leaving only the learner-side bias, which is tractable to bound.

We note that the use of a correction term arises in various online optimization settings [Steinhardt and Liang, 2014, Zhao et al., 2024], where it serves to translate a bias (or error) term from the comparator side to the learner side of the regret decomposition. In the context of adversarial linear/kernel bandits and policy optimization, different forms of correction have been designed to compensate for the bias of different loss estimators [Luo et al., 2021, Liu et al., 2023, Neu et al., 2024] and to obtain high-probability bounds with quadratic correction terms [Zimmert and Lattimore, 2022]. Our correction arises naturally from the closed-form bias and integrates seamlessly into the exponential-weights update without inflating the order of the loss-estimator range, which is essential for our analysis.

Overall algorithm. Combining the two ingredients above with a mixing of exploration distribution and exponential-weights update yields Algorithm 1. At each round t, the learner samples $\mathbf{x}_t \sim p_t = (1-\gamma)q_t + \gamma\nu_G$, a mixture of the current distribution q_t and a fixed exploration distribution ν_G given by a regularized G-optimal design (e.g., see [Camilleri et al., 2021]). Upon observing y_t, it forms the corrected loss proxy $\widehat{\ell}_t$ via (1) and (3) and updates q_{t+1} by exponential weights on $\widehat{\ell}_t$. Although defined through $\widehat{\mathbf{w}}_t \in \mathcal{H}$, the proxy $\widehat{\ell}_t(\mathbf{x})$ admits a closed form in terms of kernel evaluations on $\mathcal{X}$, so $\widehat{\mathbf{w}}_t$ is never instantiated explicitly. We next discuss the computational aspects of the algorithm.

Remark 2 (Computational Cost)**.** The computational cost of Algorithm 1 mainly comes from two parts: evaluating the loss proxy $\widehat{\ell}_t$ at each round, and computing the exploration distribution ν_G once at initialization. Both can be implemented entirely via kernel evaluations $\mathcal{K}(\mathbf{x},\mathbf{y}) = \langle\Phi(\mathbf{x}),\Phi(\mathbf{y})\rangle_{\mathcal{H}}$, without ever instantiating the feature map $\Phi(\cdot)$ explicitly. The two terms in the loss proxy $\widehat{\ell}_t(\mathbf{x}) = \langle\widehat{\mathbf{w}}_t,\Phi(\mathbf{x})\rangle_{\mathcal{H}} - c_t(\mathbf{x})$ are both bilinear forms $\Phi(\mathbf{x})^\top(\Sigma_t^\lambda)^{-1}\Phi(\mathbf{y})$ in the regularized covariance. By Camilleri et al. [2021, Lemma 1], any such form admits the closed-form expression $\Phi(\mathbf{x})^\top(\Sigma_t^\lambda)^{-1}\Phi(\mathbf{y}) = \frac{1}{\lambda}\big(\mathcal{K}(\mathbf{x},\mathbf{y}) - k_t(\mathbf{x})^\top(K_t + \lambda I_{|\mathcal{X}|})^{-1}\,k_t(\mathbf{y})\big)$, where $[K_t]_{i,j} = \sqrt{p_t(\mathbf{x}_i)\,p_t(\mathbf{x}_j)}\,\mathcal{K}(\mathbf{x}_i,\mathbf{x}_j)$ and $[k_t(\mathbf{x})]_i = \sqrt{p_t(\mathbf{x}_i)}\,\mathcal{K}(\mathbf{x}_i,\mathbf{x})$ are both expressed only through kernel evaluations on $\mathcal{X}$. The dominant per-round cost is the inversion of $K_t + \lambda I_{|\mathcal{X}|}$ at $\mathcal{O}(|\mathcal{X}|^3)$. Once this is computed, each evaluation of $\widehat{\ell}_t(\mathbf{x})$ takes $\mathcal{O}(|\mathcal{X}|^2)$. This complexity might ideally be brought down to at most linear in $|\mathcal{X}|$ (the cost of iterating over the arms), but we leave this for future work. In practice, since our estimator follows the same structure as kernel ridge regression (KRR), acceleration techniques, such as Nyström [Drineas and Mahoney, 2005] or random Fourier feature [Avron et al., 2017] approximations, can also be applied. One might also adapt the sketching approach of Calandriello et al. [2019] to accelerate our algorithm, but preserving comparator-side bias cancellation under such approximations requires further study. As noted in the introduction, a Nyström-based approach was formally analyzed in concurrent work Iwazaki [2026].

The exploration distribution $\nu_G \in \arg\min_{\nu\in\Delta(\mathcal{X})}\max_{\mathbf{x}\in\mathcal{X}}\|\Phi(\mathbf{x})\|^2_{(\Sigma(\nu)+(\lambda/\gamma)I_{\mathcal{H}})^{-1}}$ depends only on the fixed parameters λ and γ, so it is computed once at initialization. The objective is convex in ν and its gradients are again bilinear forms of the same kind, so ν_G can be obtained by first-order methods using only kernel evaluations [Camilleri et al., 2021]. ◆

4.3 Regret Guarantees

The following theorem states the regret guarantee of Algorithm 1, and will be proved below.

Theorem 1. *Under Assumption 1, let $\eta > 0$ and $\gamma,\lambda \in (0,1]$. Define the effective dimension as $d_*(\rho) = \max_{\nu\in\Delta_{\mathcal{X}}}\mathrm{Tr}\left((\Sigma(\nu)+\rho I_{\mathcal{H}})^{-1}\Sigma(\nu)\right)$, where $\Sigma(\nu) = \mathbb{E}_{\mathbf{x}\sim\nu}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$. If $\lambda \le 1/(2\eta B)$ and $d_*(\lambda/\gamma) \le \gamma/(2\eta B)$, Algorithm 1 with uniform initialization q_1 guarantees*

$$\mathcal{R}_T \le \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta d_*(\lambda)\,T + 4B\sqrt{\lambda T^2 d_*(\lambda)} + B\sqrt{\lambda\gamma T^2\,d_*(\lambda/\gamma)} + 2\gamma BT,$$

With appropriately chosen parameters, our algorithm enjoys the following regret guarantee in terms of the effective dimension. The proof and the kernel-specific rates are deferred to Appendix A.5.

Corollary 1. Under Assumption 1 , with the choices

$$\lambda = \frac{1}{T},\quad \eta = \frac{1}{2B}\sqrt{\frac{\log(e|\mathcal{X}|)}{2(1+\lambda)d_*(\lambda)T}},\quad \gamma = \min\left\{\sqrt{\frac{2d_*(\lambda)\log(e|\mathcal{X}|)}{(1+\lambda)T}},1\right\},$$

Algorithm 1 satisfies $\mathcal{R}_T = \mathcal{O}\left(B\sqrt{d_*(\lambda)T\log(|\mathcal{X}|)}\right)$.

In general, the effective dimension d_* may be difficult to evaluate or bound directly. We then proceed to detail two approaches to doing so and attaining more specific regret bounds.

Bounding d_* via maximum information gain. The effective dimension $d_*(\lambda)$ in Corollary 1 can be related to the maximum information gain widely studied in stochastic kernel bandits. Here, we define

$$\bar{\gamma}_T := \max_{\mathbf{x}_1,\dots,\mathbf{x}_T\in\mathcal{X}}\frac{1}{2}\log\det\left(I_{\mathcal{H}} + \sum_{t=1}^{T}\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)^\top\right),$$

which corresponds to the maximum information gain of Srinivas et al. [2010] at noise level $\sigma^2 = 1$. Lemma 8 in A.1 shows that $d_*(1/T) \leq \mathcal{O}(\bar{\gamma}_T)$, so the regret bound of Corollary 1 translates to

$$\mathcal{R}_T = \widetilde{\mathcal{O}}\Big(B\sqrt{\bar{\gamma}_T\, T\, \log|\mathcal{X}|}\Big)\,.$$

In particular, when B is constant and $\log|\mathcal{X}| = O(\log T)$,[4] this simplifies to $\widetilde{\mathcal{O}}(\sqrt{T\bar{\gamma}_T})$, matching the $\widetilde{\mathcal{O}}(\sqrt{T\bar{\gamma}_T})$ rate achieved by several stochastic kernel bandit algorithms [Valko et al., 2013, Salgia et al., 2021, Li and Scarlett, 2022]. Moreover, bounds on the maximum information gain are known for commonly used kernels on a compact subset of $\mathbb{R}^d$ (e.g., $[0,1]^d$) such as Matérn and SE [Vakili et al., 2021], which lead to the following kernel-specific rates:

- **Matérn**(ν, d) **kernel:** we have $\bar{\gamma}_T = \widetilde{\mathcal{O}}(T^{d/(2\nu+d)})$, giving $\mathcal{R}_T = \widetilde{\mathcal{O}}(T^{(\nu+d)/(2\nu+d)})$.
- **SE kernel:** we have $\bar{\gamma}_T = \mathcal{O}((\log T)^{d+1})$, giving $\mathcal{R}_T = \widetilde{\mathcal{O}}(\sqrt{T})$.

The concurrent work of Iwazaki [2026] establishes lower bounds for both of these kernels, showing that our upper bound is tight up to logarithmic factors in T. Both rates also match the optimal regret rates known for stochastic kernel bandits; see, e.g., Vakili et al. [2021], Li and Scarlett [2022].

Remark 3. While the adversarial kernel bandit problem intuitively seems to be more difficult than the stochastic counterpart, we highlight that neither is a special case of the other. In particular, the additive noise in the stochastic problem can create bottlenecks that are not present in the adversarial problem, and for linear rewards these can actually lead to strictly higher regret; see for example [Lattimore and Szepesvári, 2020, Chapter 29]. ◆

Bounding d_* via eigenvalue decay conditions. While relating the effective dimension to the extensively studied maximum information gain is a natural approach to obtaining results for specific kernels, another route is to place assumptions directly on the kernel eigenvalue decay and upper bound $d_*(\lambda)$ directly based on that decay, with suitable tuning of η. Corollaries 2 and 3 below carry this out for polynomial and exponential eigenvalue decay, respectively, defined as follows.

Definition 1 (Eigenvalue Decay)**.** Let p be any probability measure over $\mathcal{X}$, and let $\mu_1 \geq \mu_2 \geq \cdots \geq 0$ denote the eigenvalues of the covariance operator $\Sigma = \mathbb{E}_{\mathbf{x}\sim p}\big[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top\big]$.

1. $\mathcal{K}$ is said to have (C,β)-polynomial eigenvalue decay, with $\beta > 1$, if there exists a constant $C > 0$ such that $\mu_j \leq Cj^{-\beta}$ for all integers $j \geq 1$, regardless of p.
2. $\mathcal{K}$ is said to have (C,β)-exponential eigenvalue decay if there exist constants $C > 0$ and $\beta > 0$ such that $\mu_j \leq Ce^{-\beta j}$ for all integers $j \geq 1$, regardless of p.

It is worth noting that the uniformity over p makes this definition more restrictive than the assumption of similar decay for a fixed measure (e.g., see [Chatterji et al., 2019, Definition 11]). Nevertheless, under these conditions, we have the following corollaries, whose proofs are deferred to Appendix A.6.

Corollary 2. Suppose $\mathcal{K}$ has (C,β)-polynomial eigenvalue decay defined in Definition 1. Under Assumption 1 , Algorithm 1 initialized with the uniform distribution q_1 over $\mathcal{X}$ and run with

$$m = \Big\lceil T^{1/\beta}\Big\rceil, \quad \lambda = m^{-\beta}, \quad \gamma = m^{-\frac{\beta-1}{2}}, \quad \eta = \frac{1}{2B(1+C/(\beta-1))}\, m^{-\frac{\beta+1}{2}}$$

guarantees

$$\mathcal{R}_T = \mathcal{O}\left(B\left(1 + \left(1+\frac{C}{\beta-1}\right)\log|\mathcal{X}| + \sqrt{1+\frac{C}{\beta-1}}\right) T^{\frac{\beta+1}{2\beta}}\right).$$

Thus, when (B,β,C) are constant and $\log|\mathcal{X}| = O(\log T)$, we have $\mathcal{R}_T = O\big(T^{\frac{\beta+1}{2\beta}}\log T\big)$.

Corollary 3. Suppose $\mathcal{K}$ has (C,β)-exponential eigenvalue decay defined in Definition 1. Under Assumption 1, Algorithm 1 initialized with the uniform distribution q_1 over $\mathcal{X}$ and run with

$$m = \left\lceil \frac{\log T}{\beta}\right\rceil, \quad \lambda = e^{-\beta m}, \quad \gamma = \sqrt{\frac{\Big(m + \frac{C}{e^\beta - 1}\Big)\log(e|\mathcal{X}|)}{T}}, \quad \eta = \frac{\gamma}{2B\Big(m+\frac{C}{e^\beta-1}\Big)},$$

[4]This condition holds when the dimension is fixed and we let $|\mathcal{X}| = T^{O(d)}$ in the covering argument described at Section 3.

provided that T is large enough so that $\gamma \le 1$, guarantees

$$\mathcal{R}_T = \mathcal{O}\left(B\sqrt{T\left(\frac{\log T}{\beta} + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}\right).$$

Thus, when (B, β, C) are constant and $\log|\mathcal{X}| = O(\log T)$, we have $\mathcal{R}_T = O\big(\sqrt{T}\log T\big)$.

Remark 4. For the exponential decay case, the regret bound is $\widetilde{\mathcal{O}}(\sqrt{T})$, which matches the tight rate in Chatterji et al. [2019]. For the polynomial decay case, our regret has the key dependence $\widetilde{O}\big(T^{\frac{\beta+1}{2\beta}}\big)$, which improves over the previous $\widetilde{O}\big(T^{\frac{\beta}{2(\beta-1)}}\big)$ upper bound from Chatterji et al. [2019], albeit under the stronger version of the decay assumption as noted above. Although the same paper also proves a lower bound of order $\Omega\big(T^{\frac{\beta+1}{2\beta}}\big)$, this lower bound does not directly establish the tightness of our result. One obstacle is that their construction uses a T-dependent action set, making the $\log|\mathcal{X}|$ factor in our upper bound no longer lower order. Moreover, the result relies on a different eigendecay condition and adversary model, making it incompatible with our setting. We provide a more detailed discussion of these issues in Appendix B.1.

5 Regret Analysis

In this section, we provide the regret analysis that establishes Theorem 1. The proofs of some auxiliary lemmas are deferred to the appendices.

Proof of Theorem 1. Let $p_* \in \Delta(\mathcal{X})$ denote the probability distribution that puts all its mass on $\mathbf{x}_*$, i.e., $p_*(\mathbf{x}) = \mathbf{1}\{\mathbf{x} = \mathbf{x}_*\}$, where $\mathbf{x}_* \in \arg\min_{\mathbf{x}\in\mathcal{X}} \sum_{t=1}^T \langle\Phi(\mathbf{x}), \mathbf{w}_t\rangle_{\mathcal{H}}$ is a best fixed action in hindsight. We also define $\mathcal{F}_{t-1} = \sigma(\mathbf{x}_1, \mathbf{w}_1, \ldots, \mathbf{x}_{t-1}, \mathbf{w}_{t-1})$ as the history available before round t. Since both $\mathbf{w}_t$ and the learner's prediction distribution p_t are $\mathcal{F}_{t-1}$-measurable, and $\mathbf{x}_t$ is sampled conditionally on $\mathcal{F}_{t-1}$ according to p_t, i.e., $\mathbb{P}(\mathbf{x}_t = \mathbf{x} \mid \mathcal{F}_{t-1}) = p_t(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}$, applying the tower property of conditional expectation yields

$$\begin{aligned}
\mathcal{R}_T &= \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}[\langle\Phi(\mathbf{x}_t), \mathbf{w}_t\rangle_{\mathcal{H}} \mid \mathcal{F}_{t-1}] - \sum_{t=1}^T \langle\Phi(\mathbf{x}_*), \mathbf{w}_t\rangle_{\mathcal{H}}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T \sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\langle\Phi(\mathbf{x}), \mathbf{w}_t\rangle_{\mathcal{H}} - \sum_{t=1}^T \sum_{\mathbf{x}\in\mathcal{X}} p_*(\mathbf{x})\langle\Phi(\mathbf{x}), \mathbf{w}_t\rangle_{\mathcal{H}}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T \sum_{\mathbf{x}\in\mathcal{X}} \big(p_t(\mathbf{x}) - p_*(\mathbf{x})\big)\langle\Phi(\mathbf{x}), \mathbf{w}_t\rangle_{\mathcal{H}}\right]. \qquad (4)
\end{aligned}$$

We next relate the loss function induced by $\mathbf{w}_t$ to the loss estimator used in Algorithm 1. Recall the choice $c_t(\mathbf{x}) := B\sqrt{\lambda}\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}}$, so that $\widehat{\ell}_t(\mathbf{x}) = \langle\Phi(\mathbf{x}), \widehat{\mathbf{w}}_t\rangle_{\mathcal{H}} - c_t(\mathbf{x})$. Then

$$\langle\Phi(\mathbf{x}), \mathbf{w}_t\rangle_{\mathcal{H}} \;=\; \widehat{\ell}_t(\mathbf{x}) + \langle\Phi(\mathbf{x}), \mathbf{w}_t - \widehat{\mathbf{w}}_t\rangle_{\mathcal{H}} + c_t(\mathbf{x}),$$

and separating contributions yields the following *decomposition with correction terms*:

$$\begin{aligned}
\mathcal{R}_T \;=\; &\underbrace{\mathbb{E}\left[\sum_{t=1}^T \sum_{\mathbf{x}\in\mathcal{X}} \big(p_t(\mathbf{x}) - p_*(\mathbf{x})\big)\widehat{\ell}_t(\mathbf{x})\right]}_{\mathcal{R}_T^{\widehat{\ell}}} + \underbrace{\mathbb{E}\left[\sum_{t=1}^T \sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\langle\Phi(\mathbf{x}), \mathbf{w}_t - \widehat{\mathbf{w}}_t\rangle_{\mathcal{H}}\right]}_{B_T^{\mathbf{x}_t}} \\
&+ \underbrace{\mathbb{E}\left[\sum_{t=1}^T \sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\, c_t(\mathbf{x})\right]}_{C_T^{\mathbf{x}_t}} - \underbrace{\mathbb{E}\left[\sum_{t=1}^T \langle\Phi(\mathbf{x}_*), \mathbf{w}_t - \widehat{\mathbf{w}}_t\rangle_{\mathcal{H}}\right]}_{B_T^{\mathbf{x}_*}} - \underbrace{\mathbb{E}\left[\sum_{t=1}^T c_t(\mathbf{x}_*)\right]}_{C_T^{\mathbf{x}_*}}. \qquad (5)
\end{aligned}$$

In the above, $\mathcal{R}_T^{\widehat{\ell}}$ is the regret term defined in terms of the loss estimator. The terms $B_T^{\mathbf{x}_t}$ and $B_T^{\mathbf{x}_*}$ are the regularization-induced bias terms on the learner side and comparator side, respectively. The

learner-side bias $B_T^{\mathbf{x}_t}$ is easier to control because it is averaged under the sampling distribution p_t, whereas $B_T^{\mathbf{x}_*}$ may be harder to bound when the optimal fixed action $\mathbf{x}_*$ is poorly explored. The correction terms $C_T^{\mathbf{x}_t}$ and $C_T^{\mathbf{x}_*}$ are therefore designed so that the comparator-side residual $-(B_T^{\mathbf{x}_*} + C_T^{\mathbf{x}_*})$ is non-positive. The learner-side bias, comparator-side bias, and estimated-loss regret are controlled by Lemma 1, Lemma 2, and Lemma 3, respectively. The detailed proofs of these lemmas are deferred to Appendix A.

Lemma 1 (Learner-side Residual)**.** *Under Assumption 1, the learner-side residual satisfies*

$$B_T^{\mathbf{x}_t} + C_T^{\mathbf{x}_t} \;\le\; 4B\sqrt{\lambda T\,\mathbb{E}\left[\sum_{t=1}^{T} d_{\mathrm{eff}}(\Sigma_t,\lambda)\right]}, \tag{6}$$

for any $\lambda > 0$. Here, for each $t \in [T]$, $d_{\mathrm{eff}}(\Sigma_t,\lambda) := \mathrm{Tr}\big((\Sigma_t^{\lambda})^{-1}\Sigma_t\big)$ denotes the p_t-dependent version of effective dimension, where $\Sigma_t = \mathbb{E}_{\mathbf{x}\sim p_t}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$ and $\Sigma_t^{\lambda} = \Sigma_t + \lambda I_{\mathcal{H}}$.

Lemma 2 (Comparator-side Residual)**.** *Under Assumption 1 and for any $\lambda > 0$, the comparator-side residual is non-positive, i.e.,*

$$-B_T^{\mathbf{x}_*} - C_T^{\mathbf{x}_*} \;\le\; 0. \tag{7}$$

Lemma 3 (Regret Bound via EW)**.** *Under Assumption 1, for any $\lambda > 0$, $\eta > 0$, and $\gamma \in (0,1]$, let*

$$\nu_D \;\in\; \operatorname*{arg\,max}_{\nu\in\Delta(\mathcal{X})} \log\det\Big(\mathbb{E}_{\mathbf{x}\sim\nu}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top] + \tfrac{\lambda}{\gamma} I_{\mathcal{H}}\Big)$$

be the regularized D-optimal exploration distribution, and define $\Sigma(\nu_D) := \mathbb{E}_{\mathbf{x}\sim\nu_D}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$. Under the condition $d_{\mathrm{eff}}(\Sigma(\nu_D),\lambda/\gamma) \;\le\; \gamma/(2\eta B)$, and $\lambda \le 1/(2\eta B)$, Algorithm 1 satisfies

$$\mathcal{R}_T^{\widehat{\ell}} \;\le 2\gamma BT + \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta\cdot\mathbb{E}\left[\sum_{t=1}^{T} d_{\mathrm{eff}}(\Sigma_t,\lambda)\right] + B\sqrt{\lambda\gamma T\sum_{t=1}^{T} d_{\mathrm{eff}}(\Sigma(\nu_D),\lambda/\gamma)}.$$

Combining (5) with Lemma 1, Lemma 2 and Lemma 3, we obtain

$$\begin{aligned}\mathcal{R}_T \le 2\gamma BT + \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta\cdot\mathbb{E}\left[\sum_{t=1}^{T} d_{\mathrm{eff}}(\Sigma_t,\lambda)\right] \\ + B\sqrt{\lambda\gamma T\sum_{t=1}^{T} d_{\mathrm{eff}}(\Sigma(\nu_D),\lambda/\gamma)} + 4B\sqrt{\lambda T\,\mathbb{E}\left[\sum_{t=1}^{T} d_{\mathrm{eff}}(\Sigma_t,\lambda)\right]}.\end{aligned}$$

Let $d_*(\rho) = \max_{\nu\in\Delta_{\mathcal{X}}} \mathrm{Tr}\left((\Sigma(\nu)+\rho I_{\mathcal{H}})^{-1}\Sigma(\nu)\right)$, where $\Sigma(\nu) = \mathbb{E}_{\mathbf{x}\sim\nu}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$. We obtain

$$\mathcal{R}_T \le \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta d_*(\lambda)\,T + B\sqrt{\lambda\gamma T^2\, d_*(\lambda/\gamma)} + 4B\sqrt{\lambda T^2 d_*(\lambda)} + 2\gamma BT,$$

which completes the proof. □

6 Conclusion

For the adversarial kernel bandit problem, we proposed an exponential-weights algorithm with a regularized importance-weighted loss estimator and a correction term that mitigates the regularization bias. We established a regret bound of $\widetilde{\mathcal{O}}\big(B\sqrt{T\,d_*(\lambda)\,\log|\mathcal{X}|}\big)$, which matches the optimal rate of the analogous stochastic kernel bandit problem up to logarithmic factors. As a direct consequence, our result yields a near-optimal $\widetilde{\mathcal{O}}\big(T^{(\nu+d)/(2\nu+d)}\big)$ regret bound for the Matérn(ν,d) kernel, improving over the best-known prior rate of Chatterji et al. [2019] while removing their rank-one adversary assumption. We also derived corollaries under structural assumptions on the kernel spectrum, obtaining $\widetilde{\mathcal{O}}\big(T^{(\beta+1)/(2\beta)}\big)$ regret under polynomial eigenvalue decay with exponent $\beta > 1$ and $\widetilde{\mathcal{O}}(\sqrt{T})$ regret under exponential eigenvalue decay. Possible future directions include obtaining high-probability regret bounds and considering the contextual kernel bandit setting.

Acknowledgment

We thank Shogo Iwazaki for helpful comments on an earlier version of our paper, in particular leading to the addition of Lemma 8. This work is supported by the Singapore National Research Foundation (NRF) and the Ministry of Digital Development and Information (MDDI) under the AI Visiting Professorship programme (Award AIVP-2024-003).

References

Jacob D. Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pages 263–274, 2008.

Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. Random fourier features for kernel ridge regression: Approximation bounds and statistical guarantees. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 253–262, 2017.

Baruch Awerbuch and Robert D Kleinberg. Adaptive routing with end-to-end feedback: Distributed learning and geometric approaches. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC)*, pages 45–53, 2004.

Sébastien Bubeck, Nicolò Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pages 41.1–41.14. JMLR Workshop and Conference Proceedings, 2012.

Xu Cai and Jonathan Scarlett. On lower bounds for standard and robust Gaussian process bandit optimization. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

Xu Cai and Jonathan Scarlett. Lower bounds for time-varying kernelized bandits. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 73–81, 2025.

Daniele Calandriello, Luigi Carratino, Alessandro Lazaric, Michal Valko, and Lorenzo Rosasco. Gaussian process optimization with adaptive sketching: Scalable and no regret. In *Proceedings of the 32nd Conference on Learning Theory (COLT)*, 2019.

Romain Camilleri, Julian Katz-Samuels, and Kevin Jamieson. High-dimensional experimental design and kernel bandits. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pages 1227–1237, 2021.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Nicolò Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

Niladri S. Chatterji, Aldo Pacchiano, and Peter L. Bartlett. Online learning with kernel losses. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 971–980, 2019.

Yutian Chen, Aja Huang, Ziyu Wang, Ioannis Antonoglou, Julian Schrittwieser, David Silver, and Nando de Freitas. Bayesian optimization in AlphaGo. *arXiv preprint arXiv:1812.06855*, 2018.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 345–352, 2007.

Petros Drineas and Michael W. Mahoney. On the Nyström method for approximating a gram matrix for improved kernel-based learning. *Journal of Machine Learning Research*, 6:2153–2175, 2005.

Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2 (3-4):157–325, 2016.

Elad Hazan and Zohar Karnin. Volumetric spanners: An efficient exploration basis for learning. *Journal of Machine Learning Research*, 17(119):1–34, 2016.

Kihyuk Hong, Yuhang Li, and Ambuj Tewari. An optimization-based algorithm for non-stationary kernel bandits without prior knowledge. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 3048–3085, 2023.

Shogo Iwazaki. Nearly-optimal algorithm for adversarial kernelized bandits. *arXiv preprint arXiv:2605.10299*, 2026.

Shogo Iwazaki and Shion Takeno. Near-optimal algorithm for non-stationary kernelized bandits. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 406–414, 2025.

David Janz, David Burt, and Javier González. Bandit optimisation of functions in the Matérn kernel RKHS. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

David Janz, Arya Akhavan, and Alexandre B. Tsybakov. Maximum effective dimension and information gain. *arXiv preprint arXiv:2608.24450*, 2026.

Motonobu Kanagawa, Philipp Hennig, Dino Sejdinovic, and Bharath K Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.

Ksenia Korovina, Sailun Xu, Kirthevasan Kandasamy, Willie Neiswanger, Barnabas Poczos, Jeff Schneider, and Eric Xing. ChemBO: Bayesian optimization of small organic molecules with synthesizable recommendations. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 3393–3403, 2020.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

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 (WWW)*, pages 661–670, 2010.

Zihan Li and Jonathan Scarlett. Gaussian process bandit optimization with few batches. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

Haolin Liu, Chen-Yu Wei, and Julian Zimmert. Bypassing the simulator: Near-optimal adversarial linear contextual bandits. In *Advances in Neural Information Processing Systems 36 (NeurIPS)*, 2023.

Haipeng Luo, Chen-Yu Wei, and Chung-Wei Lee. Policy optimization in adversarial MDPs: Improved exploration via dilated bonuses. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, pages 22931–22942, 2021.

Chenkai Ma and Jonathan Scarlett. A general $\widetilde{\Omega}(\sqrt{T\gamma_T})$ lower bound for kernel bandits. *arXiv preprint arXiv:2610.11082*, 2026.

Roman Marchant and Fabio Ramos. Bayesian optimisation for intelligent environmental monitoring. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2242–2249. IEEE, 2012.

Gergely Neu, Julia Olkhovskaya, and Sattar Vakili. Adversarial contextual bandits go kernelized. In *Proceedings of The 35th International Conference on Algorithmic Learning Theory (ALT)*, pages 907–929, 2024.

Sudeep Salgia, Sattar Vakili, and Qing Zhao. A domain-shrinking based Bayesian optimization algorithm with order-optimal regret performance. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2021.

Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy Gaussian process bandit optimization. In *Proceedings of the 30th Conference on Learning Theory (COLT)*, 2017.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 2010.

Jacob Steinhardt and Percy Liang. Adaptivity and optimism: An improved exponentiated gradient algorithm. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 1593–1601, 2014.

Sho Takemori and Masahiro Sato. Approximation theory based methods for RKHS bandits. In *International Conference on Machine Learning*, pages 10076–10085. PMLR, 2021.

Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in Gaussian process bandits. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

Michal Valko, Nathan Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. Finite-time analysis of kernelised contextual bandits. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2013.

Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization. *Journal of Machine Learning Research*, 25(98):1–52, 2024.

Julian Zimmert and Tor Lattimore. Return of the bias: Almost minimax optimal high probability bounds for adversarial linear bandits. In *Proceedings of the 35th Conference on Learning Theory (COLT)*, pages 3285–3312, 2022.

A Omitted Details from Section 5

A.1 Useful Lemmas

The following lemma characterizes the conditional expectation of the regularized importance-weighted estimator. It shows that the estimator is biased due to regularization: instead of recovering $\mathbf{w}_t$ in expectation, it returns a shrunk version $(\Sigma_t^\lambda)^{-1}\Sigma_t\mathbf{w}_t = (I_\mathcal{H} - \lambda(\Sigma_t^\lambda)^{-1})\mathbf{w}_t$. This explicit form of the bias will be useful in controlling the effect of regularization in the regret analysis.

Lemma 4 (Conditional Mean of Loss Estimator)**.** *Fix any round $t \in [T]$, and let $\Sigma_t := \mathbb{E}_{\mathbf{x}\sim p_t}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$ be the unregularized covariance operator. Then for any $\lambda > 0$, $\Sigma_t^\lambda = \Sigma_t + \lambda I_\mathcal{H}$. Conditioning on the history $\mathcal{F}_{t-1}$, the estimator*

$$\widehat{\mathbf{w}}_t = \langle \Phi(\mathbf{x}_t), \mathbf{w}_t\rangle_\mathcal{H} (\Sigma_t^\lambda)^{-1}\Phi(\mathbf{x}_t), \quad (\mathbf{x}_t \,|\, \mathcal{F}_{t-1}) \sim p_t,$$

satisfies

$$\mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] = (\Sigma_t^\lambda)^{-1}\Sigma_t\,\mathbf{w}_t = \big(I_\mathcal{H} - \lambda(\Sigma_t^\lambda)^{-1}\big)\mathbf{w}_t. \tag{8}$$

Proof. We condition on $\mathcal{F}_{t-1}$, meaning that p_t, $\mathbf{w}_t$, Σ_t, and $(\Sigma_t^\lambda)^{-1}$ are fixed. Thus,

$$\mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] = (\Sigma_t^\lambda)^{-1}\mathbb{E}_{\mathbf{x}\sim p_t}[\langle \Phi(\mathbf{x}), \mathbf{w}_t\rangle_\mathcal{H}\, \Phi(\mathbf{x})]\,.$$

By the definition of Σ_t,

$$\mathbb{E}_{\mathbf{x}\sim p_t}[\langle \Phi(\mathbf{x}), \mathbf{w}_t\rangle_\mathcal{H}\, \Phi(\mathbf{x})] = \mathbb{E}_{\mathbf{x}\sim p_t}\left[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top\right]\mathbf{w}_t = \Sigma_t\mathbf{w}_t,$$

which implies

$$\mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] = (\Sigma_t^\lambda)^{-1}\Sigma_t\mathbf{w}_t.$$

Finally, since $\Sigma_t^\lambda = \Sigma_t + \lambda I_\mathcal{H}$, we have

$$(\Sigma_t^\lambda)^{-1}\Sigma_t = (\Sigma_t^\lambda)^{-1}(\Sigma_t^\lambda - \lambda I_\mathcal{H}) = I_\mathcal{H} - \lambda(\Sigma_t^\lambda)^{-1}.$$

Substituting this identity gives $\mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] = \big(I_\mathcal{H} - \lambda(\Sigma_t^\lambda)^{-1}\big)\mathbf{w}_t$ as claimed. □

The following lemma establishes a uniform bound on the corrected proxy loss. This boundedness condition allows us to invoke the standard exponential-weights regret guarantee despite the estimator being defined in a potentially infinite-dimensional RKHS.

Lemma 5 (Bounded Loss Estimator)**.** *Under Assumption 1, suppose that the effective-dimension condition $d_{\mathrm{eff}}(\Sigma(\nu_D), \lambda/\gamma) \le \gamma/(2\eta B)$ in Lemma 3 holds and that $\lambda \le 1/(2\eta B)$. Then, the loss proxy $\widehat{\ell}_t(\mathbf{x})$ satisfies $|\eta\widehat{\ell}_t(\mathbf{x})| \le 1$.*

Proof. By the definition of the loss proxy, for any $\mathbf{x} \in \mathcal{X}$ we have

$$\begin{aligned}
|\widehat{\ell}_t(\mathbf{x})| &\le |\langle \widehat{\mathbf{w}}_t, \Phi(\mathbf{x})\rangle_\mathcal{H}| + B\sqrt{\lambda}\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}} \\
&\le \|\mathbf{w}_t\|_\mathcal{H}\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}}\|\Phi(\mathbf{x}_t)\|_{(\Sigma_t^\lambda)^{-1}} + B\sqrt{\lambda}\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}} \\
&\le B\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}}\,\|\Phi(\mathbf{x}_t)\|_{(\Sigma_t^\lambda)^{-1}} + B\sqrt{\lambda}\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}},
\end{aligned} \tag{9}$$

where the second inequality holds using the definition $\widehat{\mathbf{w}}_t = \langle \Phi(\mathbf{x}_t), \mathbf{w}_t\rangle(\Sigma_t^\lambda)^{-1}\Phi(\mathbf{x}_t)$, and applying $|\langle \widehat{\mathbf{w}}_t, \Phi(\mathbf{x})\rangle| = |\langle \Phi(\mathbf{x}_t), \mathbf{w}_t\rangle \cdot \langle \Phi(\mathbf{x}), (\Sigma_t^\lambda)^{-1}\Phi(\mathbf{x}_t)\rangle| \le \|\mathbf{w}_t\|_\mathcal{H}\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}}\|\Phi(\mathbf{x}_t)\|_{(\Sigma_t^\lambda)^{-1}}$ by Assumption 1 and two applications of Cauchy-Schwarz. The last inequality uses $\|\mathbf{w}_t\|_\mathcal{H} \le B$, again by Assumption 1.

Next, we derive a uniform upper bound on $\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}}$. Recall that Algorithm 1 uses the exploration distribution

$$\nu_G \in \operatorname*{arg\,min}_{\nu\in\Delta(\mathcal{X})} M_{\lambda/\gamma}(\nu),$$

where $M_{\lambda/\gamma}(\nu) := \max_{\mathbf{x}\in\mathcal{X}}\|\Phi(\mathbf{x})\|^2_{\left(\Sigma(\nu)+\frac{\lambda}{\gamma}I_\mathcal{H}\right)^{-1}}$ and $\Sigma(\nu) := \mathbb{E}_{\mathbf{z}\sim\nu}\left[\Phi(\mathbf{z})\Phi(\mathbf{z})^\top\right]$.

By Camilleri et al. [2021, Lemma 3], we can relate $M_{\lambda/\gamma}$ to the effective dimension as follows:

$$M_{\lambda/\gamma}(\nu_G) \le \mathrm{Tr}\Big((\Sigma(\nu_D) + \frac{\lambda}{\gamma} I_{\mathcal{H}})^{-1}\Sigma(\nu_D)\Big) = d_{\mathrm{eff}}(\Sigma(\nu_D), \lambda/\gamma), \tag{10}$$

where $\nu_D \in \arg\max_{\nu \in \Delta(\mathcal{X})} \log\det\big(\mathbb{E}_{\mathbf{x}\sim\nu}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top] + \frac{\lambda}{\gamma} I_{\mathcal{H}}\big)$. Furthermore, since $p_t = (1-\gamma)q_t + \gamma\nu_G$, we have

$$\Sigma_t^\lambda = \Sigma_t + \lambda I_{\mathcal{H}} \succeq \gamma\Sigma(\nu_G) + \lambda I_{\mathcal{H}} = \gamma\Big(\Sigma(\nu_G) + \frac{\lambda}{\gamma} I_{\mathcal{H}}\Big).$$

By monotonicity of matrix inversion, we have

$$(\Sigma_t^\lambda)^{-1} \preceq \frac{1}{\gamma}\big(\Sigma(\nu_G) + \frac{\lambda}{\gamma} I_{\mathcal{H}}\big)^{-1},$$

which implies for all $\mathbf{x} \in \mathcal{X}$ that

$$\|\Phi(\mathbf{x})\|^2_{(\Sigma_t^\lambda)^{-1}} \le \frac{1}{\gamma} M_{\lambda/\gamma}(\nu_G) \le \frac{1}{\gamma} d_{\mathrm{eff}}(\Sigma(\nu_D), \lambda/\gamma). \tag{11}$$

Finally, substituting (11) into (9), we obtain

$$|\widehat{\ell}_t(\mathbf{x})| \le \frac{B}{\gamma} d_{\mathrm{eff}}(\Sigma(\nu_D), \lambda/\gamma) + B\sqrt{\frac{\lambda\, d_{\mathrm{eff}}(\Sigma(\nu_D), \lambda/\gamma)}{\gamma}} \le \frac{1}{2\eta} + \sqrt{\frac{B\lambda}{2\eta}},$$

where the last inequality follows from the effective-dimension condition $d_{\mathrm{eff}}(\Sigma(\nu_D), \lambda/\gamma) \le \gamma/(2\eta B)$ in Lemma 3. Using $\lambda \le 1/(2\eta B)$, it then follows that $|\widehat{\ell}_t(\mathbf{x})| \le \frac{1}{\eta}$ for all $\mathbf{x} \in \mathcal{X}$, and hence

$$|\eta\widehat{\ell}_t(\mathbf{x})| \le 1.$$

This completes the proof. □

The next lemma translates the eigenvalue-decay condition into explicit upper bounds on the effective dimension under both polynomial and exponential decay regimes. These bounds are then used to specialize the general effective-dimension regret bound to each decay setting.

Lemma 6. *Let $\xi > 0$. Define the effective dimension $d_*(\xi) := \max_{p\in\Delta} d_{\mathrm{eff}}(\Sigma(p), \xi)$, where $d_{\mathrm{eff}}(\Sigma(p), \xi) := \mathrm{Tr}\big((\Sigma^\xi(p))^{-1}\Sigma(p)\big)$, $\Sigma(p) = \mathbb{E}_{\mathbf{x}\sim p}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$, and $\Sigma^\xi(p) = \Sigma(p) + \xi I_{\mathcal{H}}$. Then, under the eigenvalue decay conditions of Definition 1, the following hold.*

1. *If $\mathcal{K}$ has (C, β)-polynomial eigenvalue decay with $\beta > 1$, then for any integer $m \ge 1$,*

$$d_*(\xi) \le m + \frac{\varepsilon_{\mathrm{poly}}}{\xi},$$

where

$$\varepsilon_{\mathrm{poly}} := \frac{C}{\beta - 1} m^{1-\beta}.$$

2. *If $\mathcal{K}$ has (C, β)-exponential eigenvalue decay with $\beta > 0$, then for any integer $m \ge 1$,*

$$d_*(\xi) \le m + \frac{\varepsilon_{\mathrm{exp}}}{\xi},$$

where

$$\varepsilon_{\mathrm{exp}} := \frac{Ce^{-\beta(m+1)}}{1 - e^{-\beta}} = \frac{Ce^{-\beta m}}{e^\beta - 1}.$$

Proof. Fix any $p \in \Delta$ and let $\mu_1 \ge \mu_2 \ge \cdots \ge 0$ denote the eigenvalues of $\Sigma(p)$. Since $\Sigma^\xi(p) = \Sigma(p) + \xi I_{\mathcal{H}}$, the operator $(\Sigma^\xi(p))^{-1}\Sigma(p)$ has eigenvalues

$$\frac{\mu_i}{\mu_i + \xi}, \qquad i \ge 1.$$

Therefore, for any integer $m \geq 1$,

$$\begin{aligned} d_{\text{eff}}(\Sigma(p), \xi) &= \mathrm{Tr}\big((\Sigma^{\xi}(p))^{-1}\Sigma(p)\big) \\ &= \sum_{i=1}^{\infty} \frac{\mu_i}{\mu_i + \xi} \\ &= \sum_{i=1}^{m} \frac{\mu_i}{\mu_i + \xi} + \sum_{i=m+1}^{\infty} \frac{\mu_i}{\mu_i + \xi} \\ &\leq m + \frac{1}{\xi} \sum_{i=m+1}^{\infty} \mu_i, \end{aligned} \tag{12}$$

where we used

$$\frac{\mu_i}{\mu_i + \xi} \leq 1, \qquad \frac{\mu_i}{\mu_i + \xi} \leq \frac{\mu_i}{\xi}.$$

Suppose $\mathcal{K}$ has (C, β)-polynomial eigenvalue decay with $\beta > 1$. By Definition 1, the bound $\mu_i \leq Ci^{-\beta}$ holds regardless of p, hence

$$\sum_{i=m+1}^{\infty} \mu_i \leq \sum_{i=m+1}^{\infty} Ci^{-\beta} \leq \int_{m}^{\infty} Cx^{-\beta}\, dx = \frac{C}{\beta - 1} m^{1-\beta} =: \varepsilon_{\text{poly}}.$$

Substituting this p-independent tail bound into (12) gives

$$d_{\text{eff}}(\Sigma(p), \xi) \leq m + \frac{\varepsilon_{\text{poly}}}{\xi} = m + \frac{C}{(\beta - 1)\xi} m^{1-\beta},$$

for every $p \in \Delta$. Taking the maximum over $p \in \Delta$ on the left-hand side yields

$$d_*(\xi) = \max_{p \in \Delta} d_{\text{eff}}(\Sigma(p), \xi) \leq m + \frac{\varepsilon_{\text{poly}}}{\xi}.$$

Suppose instead $\mathcal{K}$ has (C, β)-exponential eigenvalue decay with $\beta > 0$. By Definition 1, $\mu_i \leq Ce^{-\beta i}$ regardless of p, hence

$$\sum_{i=m+1}^{\infty} \mu_i \leq \sum_{i=m+1}^{\infty} Ce^{-\beta i} = Ce^{-\beta(m+1)} \sum_{k=0}^{\infty} e^{-\beta k} = \frac{Ce^{-\beta(m+1)}}{1 - e^{-\beta}} =: \varepsilon_{\text{exp}},$$

or equivalently,

$$\varepsilon_{\text{exp}} = \frac{Ce^{-\beta m}}{e^{\beta} - 1}.$$

Substituting this p-independent tail bound into (12) and taking the maximum over $p \in \Delta$, we obtain

$$d_*(\xi) \leq m + \frac{\varepsilon_{\text{exp}}}{\xi} = m + \frac{Ce^{-\beta(m+1)}}{(1 - e^{-\beta})\xi}.$$

This completes the proof. □

The next lemma bounds the second moment of the regularized kernel features appearing in the proxy-loss variance term. This provides the key variance control needed in the exponential-weights regret analysis.

Lemma 7. *Fix any round $t \in [T]$ and condition on $\mathcal{F}_{t-1}$. Let*

$$\Sigma_t := \mathbb{E}_{\mathbf{x} \sim p_t}\left[\Phi(\mathbf{x})\Phi(\mathbf{x})^{\top}\right], \qquad \Sigma_t^{\lambda} := \Sigma_t + \lambda I_{\mathcal{H}},$$

and let $A_t := (\Sigma_t^{\lambda})^{-1}$. Then

$$\mathbb{E}_{\mathbf{x}_t \sim p_t}\left[\sum_{\mathbf{x} \in \mathcal{X}} p_t(\mathbf{x}) \left\langle \Phi(\mathbf{x}), A_t \Phi(\mathbf{x}_t) \right\rangle_{\mathcal{H}}^2 \;\middle|\; \mathcal{F}_{t-1}\right] \leq d_{\text{eff}}(\Sigma_t, \lambda),$$

where $d_{\text{eff}}(\Sigma_t, \lambda) := \mathrm{Tr}\big((\Sigma_t + \lambda I_{\mathcal{H}})^{-1}\Sigma_t\big)$.

Proof. We condition on $\mathcal{F}_{t-1}$, meaning that p_t, Σ_t, and $A_t = (\Sigma_t + \lambda I_{\mathcal{H}})^{-1}$ are fixed, and the only remaining randomness is over $\mathbf{x}_t \sim p_t$. For any fixed $\mathbf{z} \in \mathcal{X}$, by the definition of Σ_t, we have

$$\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x}) \langle \Phi(\mathbf{x}), A_t\Phi(\mathbf{z})\rangle_{\mathcal{H}}^2 = \langle A_t\Phi(\mathbf{z}), \Sigma_t A_t \Phi(\mathbf{z})\rangle_{\mathcal{H}} .$$

Taking expectation over $\mathbf{z} = \mathbf{x}_t \sim p_t$ gives

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}_t\sim p_t}\left[\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x}) \langle \Phi(\mathbf{x}), A_t\Phi(\mathbf{x}_t)\rangle_{\mathcal{H}}^2 \;\middle|\; \mathcal{F}_{t-1}\right] \\
&\quad = \mathbb{E}_{\mathbf{z}\sim p_t}\left[\langle A_t\Phi(\mathbf{z}), \Sigma_t A_t \Phi(\mathbf{z})\rangle_{\mathcal{H}}\right] \\
&\quad = \mathrm{Tr}(\Sigma_t A_t \Sigma_t A_t) .
\end{aligned}$$

Since $A_t = (\Sigma_t + \lambda I_{\mathcal{H}})^{-1}$ is a spectral function of Σ_t, the two operators commute. Hence

$$\mathrm{Tr}(\Sigma_t A_t \Sigma_t A_t) = \mathrm{Tr}\big((A_t\Sigma_t)^2\big) .$$

Let $\{\mu_{t,i}\}_{i\ge 1}$ be the eigenvalues of Σ_t. Then the eigenvalues of $A_t\Sigma_t$ are

$$\frac{\mu_{t,i}}{\mu_{t,i}+\lambda}, \qquad i \ge 1.$$

Therefore,

$$\mathrm{Tr}\big((A_t\Sigma_t)^2\big) = \sum_{i\ge 1}\left(\frac{\mu_{t,i}}{\mu_{t,i}+\lambda}\right)^2 \le \sum_{i\ge 1}\frac{\mu_{t,i}}{\mu_{t,i}+\lambda} = d_{\mathrm{eff}}(\Sigma_t,\lambda),$$

where the inequality follows because $\mu_{t,i}/(\mu_{t,i}+\lambda) \in [0,1]$ for every i. This proves the claim. □

The final lemma of this subsection bounds the effective dimension $d_*(\lambda)$ directly by the standard T-point maximum information gain used in the stochastic kernel bandit literature [Srinivas et al., 2010, Valko et al., 2013, Vakili et al., 2021]. This is the technical bridge that lets us reformulate our effective-dimension based regret bound in the conventional information gain form.

Lemma 8 (Effective dimension vs. maximum information gain)**.** *Assume $\mathcal{X}$ is finite, let $\Phi : \mathcal{X} \to \mathcal{H}$ be a feature map with $\|\Phi(\mathbf{x})\|_{\mathcal{H}}^2 \le 1$ for every $\mathbf{x} \in \mathcal{X}$, and define*

$$\Sigma(\nu) := \mathbb{E}_{\mathbf{x}\sim\nu}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top], \qquad d_*(\lambda) := \max_{\nu\in\Delta(\mathcal{X})} \mathrm{Tr}\big((\Sigma(\nu)+\lambda I_{\mathcal{H}})^{-1}\Sigma(\nu)\big).$$

Then, for any $T \ge 1$ and any $\lambda \ge 1/T$, we have

$$d_*(\lambda) \;\le\; 2 \max_{\mathbf{x}_1,\dots,\mathbf{x}_T\in\mathcal{X}} \log\det\left(I_{\mathcal{H}} + \frac{1}{T\lambda}\sum_{t=1}^{T}\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)^\top\right).$$

Proof. By Calandriello et al. [2019, Proposition 5], the effective dimension is bounded by an operator-level log-determinant quantity:

$$d_*(\lambda) \;\le\; \max_{\nu\in\Delta(\mathcal{X})} \log\det\left(I_{\mathcal{H}} + \frac{\Sigma(\nu)}{\lambda}\right). \tag{13}$$

It thus suffices to bound the right-hand side by the T-point information gain in the lemma. To this end, set $\alpha := T\lambda$, so that $\alpha \ge 1$ by the assumption $\lambda \ge 1/T$, and $\Sigma(\nu)/\lambda = T\Sigma(\nu)/\alpha$. Fix any $\nu \in \Delta(\mathcal{X})$, write $M_\nu := T\Sigma(\nu)$, and define $F(A) := \log\det(I_{\mathcal{H}} + A/\alpha)$ on the positive semidefinite cone. We will construct T points $\mathbf{x}_1,\dots,\mathbf{x}_T \in \mathcal{X}$ such that

$$F\left(\sum_{t=1}^{T}\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)^\top\right) \;\ge\; \tfrac{1}{2}\,F(M_\nu), \tag{14}$$

which, combined with (13) and taking the maximum over $\nu \in \Delta(\mathcal{X})$, yields the lemma.

We construct the points greedily: set $V_0 := 0$ and, for $t = 0, 1, \dots, T-1$, choose

$$\mathbf{x}_{t+1} \;\in\; \arg\max_{\mathbf{x}\in\mathcal{X}} \|\Phi(\mathbf{x})\|^2_{(\alpha I_{\mathcal{H}}+V_t)^{-1}}, \qquad V_{t+1} := V_t + \Phi(\mathbf{x}_{t+1})\Phi(\mathbf{x}_{t+1})^\top, \tag{15}$$

where the maximizer exists since $\mathcal{X}$ is finite. With the trace inner product $\langle A, B\rangle_{\mathrm{tr}} = \mathrm{Tr}(AB)$, we have $\nabla F(A) = (\alpha I_{\mathcal{H}} + A)^{-1}$, and by concavity of F,

$$\begin{aligned} F(M_\nu) - F(V_t) &\le \mathrm{Tr}\big((\alpha I_{\mathcal{H}} + V_t)^{-1}(M_\nu - V_t)\big) \\ &\le \mathrm{Tr}\big((\alpha I_{\mathcal{H}} + V_t)^{-1} M_\nu\big) \\ &= T\,\mathbb{E}_{\mathbf{x}\sim\nu}\big[\,\|\Phi(\mathbf{x})\|^2_{(\alpha I_{\mathcal{H}}+V_t)^{-1}}\,\big] \\ &\le T\,\|\Phi(\mathbf{x}_{t+1})\|^2_{(\alpha I_{\mathcal{H}}+V_t)^{-1}}, \end{aligned}$$

where the second inequality drops the non-positive term $-\mathrm{Tr}((\alpha I_{\mathcal{H}} + V_t)^{-1} V_t)$ (note that $V_t \succeq 0$ and $(\alpha I_{\mathcal{H}} + V_t)^{-1} \succeq 0$), the equality uses $M_\nu = T\,\mathbb{E}_{\mathbf{x}\sim\nu}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$, and the last step is the defining property of $\mathbf{x}_{t+1}$ in (15). On the other hand, the matrix determinant lemma gives

$$F(V_{t+1}) - F(V_t) \;=\; \log\big(1 + \|\Phi(\mathbf{x}_{t+1})\|^2_{(\alpha I_{\mathcal{H}}+V_t)^{-1}}\big). \tag{16}$$

Let $z_t := \|\Phi(\mathbf{x}_{t+1})\|^2_{(\alpha I_{\mathcal{H}}+V_t)^{-1}}$. Since $V_t \succeq 0$, $\alpha \ge 1$, and $\|\Phi(\mathbf{x})\|^2_{\mathcal{H}} \le 1$, we have $z_t \le \frac{1}{\alpha}\|\Phi(\mathbf{x}_{t+1})\|^2_{\mathcal{H}} \le 1$. Applying $\log(1+z) \ge (\log 2)\, z$ for $z \in [0,1]$ and combining with (16) gives

$$F(M_\nu) - F(V_t) \;\le\; \frac{T}{\log 2}\,\big(F(V_{t+1}) - F(V_t)\big). \tag{17}$$

We conclude the proof using a telescoping argument. Let $\Delta_t := F(M_\nu) - F(V_t)$. If $\Delta_t \le 0$ for some $t \le T$, then since $F(V_s)$ is non-decreasing in s we have $F(V_T) \ge F(V_t) \ge F(M_\nu)$, and (14) follows. Otherwise $\Delta_t > 0$ for all $t < T$, and (17) rearranges to $\Delta_{t+1} = \Delta_t - (F(V_{t+1}) - F(V_t)) \le (1 - \frac{\log 2}{T})\Delta_t$. Since $V_0 = 0$, we have $\Delta_0 = F(M_\nu)$, and iterating T times yields

$$\Delta_T \;\le\; \Big(1 - \tfrac{\log 2}{T}\Big)^T F(M_\nu) \;\le\; e^{-\log 2}\, F(M_\nu) \;=\; \tfrac{1}{2}\, F(M_\nu),$$

where the second inequality uses $(1-x)^T \le e^{-xT}$. Therefore $F(V_T) \ge \frac{1}{2}\, F(M_\nu)$, which establishes (14) and completes the proof. □

A.2 Proof of Lemma 1 (Learner-side Residual)

The following lemma bounds the learner-side residual terms introduced by the regularized estimator. It shows that these terms can be controlled by the effective dimension induced by the covariance operator Σ_t.

Lemma 1 (Learner-side Residual)**.** *Under Assumption 1, the learner-side residual satisfies*

$$B_T^{\mathbf{x}_t} + C_T^{\mathbf{x}_t} \;\le\; 4B\sqrt{\lambda T\,\mathbb{E}\left[\sum_{t=1}^{T} d_{\mathrm{eff}}(\Sigma_t, \lambda)\right]}, \tag{6}$$

for any $\lambda > 0$. Here, for each $t \in [T]$, $d_{\mathrm{eff}}(\Sigma_t, \lambda) := \mathrm{Tr}\big((\Sigma_t^\lambda)^{-1}\Sigma_t\big)$ denotes the p_t-dependent version of effective dimension, where $\Sigma_t = \mathbb{E}_{\mathbf{x}\sim p_t}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$ and $\Sigma_t^\lambda = \Sigma_t + \lambda I_{\mathcal{H}}$.

Proof. Define the regularization bias operator

$$E_t := I_{\mathcal{H}} - (\Sigma_t^\lambda)^{-1}\Sigma_t = \lambda(\Sigma_t^\lambda)^{-1}.$$

Since $\Sigma_t^\lambda = \Sigma_t + \lambda I_{\mathcal{H}}$, we have $0 \preceq E_t \preceq I_{\mathcal{H}}$. Moreover, recalling that

$$c_t(\mathbf{x}) = B\sqrt{\lambda}\,\|\Phi(\mathbf{x})\|_{(\Sigma_t^\lambda)^{-1}} = B\|\Phi(\mathbf{x})\|_{E_t},$$

we can express both residual terms $B_T^{\mathbf{x}_t}$ and $C_T^{\mathbf{x}_t}$ in terms of the same operator E_t.

As for the term $B_T^{\mathbf{x}_t}$, by definition we have

$$B_T^{\mathbf{x}_t} = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\,\langle\Phi(\mathbf{x}), \mathbf{w}_t - \widehat{\mathbf{w}}_t\rangle_{\mathcal{H}}\right].$$

Conditioning on $\mathcal{F}_{t-1}$, the only remaining randomness is in $\mathbf{x}_t$. Hence, by Lemma 4,

$$\mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] = (\Sigma_t^\lambda)^{-1}\Sigma_t \mathbf{w}_t.$$

Applying the tower property gives

$$\begin{aligned} B_T^{\mathbf{x}_t} &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\,\langle \Phi(\mathbf{x}), \mathbf{w}_t - \mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}]\rangle_{\mathcal{H}}\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\,\langle \Phi(\mathbf{x}), \big(I_{\mathcal{H}} - (\Sigma_t^\lambda)^{-1}\Sigma_t\big)\mathbf{w}_t\rangle_{\mathcal{H}}\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\,\langle \Phi(\mathbf{x}), E_t\mathbf{w}_t\rangle_{\mathcal{H}}\right]. \end{aligned}$$

Since $E_t \succeq 0$, Cauchy–Schwarz in the E_t-seminorm yields

$$\langle \Phi(\mathbf{x}), E_t\mathbf{w}_t\rangle_{\mathcal{H}} = \langle E_t^{1/2}\Phi(\mathbf{x}), E_t^{1/2}\mathbf{w}_t\rangle_{\mathcal{H}} \le \|\Phi(\mathbf{x})\|_{E_t}\|\mathbf{w}_t\|_{E_t}.$$

Using $0 \preceq E_t \preceq I_{\mathcal{H}}$ and Assumption 1, we have $\|\mathbf{w}_t\|_{E_t} \le \|\mathbf{w}_t\|_{\mathcal{H}} \le B$. Therefore, one can further bound the bias term as

$$B_T^{\mathbf{x}_t} \le B\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\|\Phi(\mathbf{x})\|_{E_t}\right].$$

On the other hand, by the definition of $C_T^{\mathbf{x}_t}$, we have

$$C_T^{\mathbf{x}_t} = B\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\|\Phi(\mathbf{x})\|_{E_t}\right].$$

Combining the two displayed equations, we obtain

$$B_T^{\mathbf{x}_t} + C_T^{\mathbf{x}_t} \le 2B\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\|\Phi(\mathbf{x})\|_{E_t}\right]. \tag{18}$$

It remains to relate the right-hand side to the effective dimension. For any $\alpha > 0$, Young's inequality gives

$$B\|\Phi(\mathbf{x})\|_{E_t} \le \alpha\|\Phi(\mathbf{x})\|_{E_t}^2 + \frac{B^2}{\alpha}.$$

Multiplying by $p_t(\mathbf{x})$ and summing over $\mathbf{x}\in\mathcal{X}$, we get

$$\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\,B\|\Phi(\mathbf{x})\|_{E_t} \le \alpha\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\|\Phi(\mathbf{x})\|_{E_t}^2 + \frac{B^2}{\alpha},$$

where we used $\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x}) = 1$. Substituting this into (18) gives

$$\begin{aligned} B_T^{\mathbf{x}_t} + C_T^{\mathbf{x}_t} &\le 2\,\mathbb{E}\left[\sum_{t=1}^{T}\left(\alpha\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\|\Phi(\mathbf{x})\|_{E_t}^2 + \frac{B^2}{\alpha}\right)\right] \\ &= 2\alpha\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\|\Phi(\mathbf{x})\|_{E_t}^2\right] + \frac{2B^2T}{\alpha}. \end{aligned}$$

We now handle the quadratic term. By the definition $\|\Phi(\mathbf{x})\|_{E_t}^2 = \langle\Phi(\mathbf{x}), E_t\Phi(\mathbf{x})\rangle_{\mathcal{H}}$, we have

$$\begin{aligned} \sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\|\Phi(\mathbf{x})\|_{E_t}^2 &= \sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\langle\Phi(\mathbf{x}), E_t\Phi(\mathbf{x})\rangle_{\mathcal{H}} \\ &= \mathrm{Tr}\left(E_t\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\Phi(\mathbf{x})\Phi(\mathbf{x})^\top\right) \\ &= \mathrm{Tr}(E_t\Sigma_t). \end{aligned}$$

Using $E_t = \lambda(\Sigma_t^\lambda)^{-1}$, we obtain

$$\mathrm{Tr}(E_t\Sigma_t) = \lambda\,\mathrm{Tr}\big((\Sigma_t^\lambda)^{-1}\Sigma_t\big).$$

Therefore,

$$\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\|\Phi(\mathbf{x})\|_{E_t}^2 = \lambda\,\mathrm{Tr}\big((\Sigma_t^\lambda)^{-1}\Sigma_t\big).$$

Substituting this identity into the previous bound yields

$$B_T^{\mathbf{x}_t} + C_T^{\mathbf{x}_t} \le 2\alpha\lambda\,\mathbb{E}\left[\sum_{t=1}^T \mathrm{Tr}\big((\Sigma_t^\lambda)^{-1}\Sigma_t\big)\right] + \frac{2B^2T}{\alpha} = 2\alpha\lambda\cdot\mathbb{E}\left[\sum_{t=1}^T d_{\mathrm{eff}}(\Sigma_t,\lambda)\right] + \frac{2B^2T}{\alpha},$$

which proves the claim by choosing $\alpha = B\sqrt{\frac{T}{\lambda\,\mathbb{E}\left[\sum_{t=1}^T d_{\mathrm{eff}}(\Sigma_t,\lambda)\right]}}$. Note that if $\mathbb{E}\left[\sum_t d_{\mathrm{eff}}(\Sigma_t,\lambda)\right] = 0$, the claim is trivial. □

A.3 Proof of Lemma 2 (Comparator-side Residual)

The following lemma shows that the comparator-side residual is non-positive. For any fixed comparator $\mathbf{x}_*$, the correction term dominates the bias induced by regularization in the comparator direction. Consequently, this residual does not contribute positively to the regret. This is a key structural property of the analysis, as it eliminates the need to further bound the comparator-side bias and leaves only the learner-side residual to be controlled.

Lemma 2 (Comparator-side Residual)**.** *Under Assumption 1 and for any $\lambda > 0$, the comparator-side residual is non-positive, i.e.,*

$$-B_T^{\mathbf{x}_*} - C_T^{\mathbf{x}_*} \;\le\; 0. \tag{7}$$

Proof. Define the regularization bias operator

$$E_t := I_{\mathcal{H}} - (\Sigma_t^\lambda)^{-1}\Sigma_t = \lambda(\Sigma_t^\lambda)^{-1}.$$

As in the proof of Lemma 1, we have $0 \preceq E_t \preceq I_{\mathcal{H}}$. By Lemma 4, conditioning on $\mathcal{F}_{t-1}$, we have

$$\mathbb{E}[\widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] = (\Sigma_t^\lambda)^{-1}\Sigma_t\mathbf{w}_t,$$

and hence

$$\mathbb{E}[\mathbf{w}_t - \widehat{\mathbf{w}}_t \mid \mathcal{F}_{t-1}] = E_t\mathbf{w}_t.$$

Since the comparator $\mathbf{x}_*$ is fixed, $\Phi(\mathbf{x}_*)$ is $\mathcal{F}_{t-1}$-measurable. Applying the tower property,

$$\begin{aligned} B_T^{\mathbf{x}_*} &= \mathbb{E}\left[\sum_{t=1}^T \langle\Phi(\mathbf{x}_*), \mathbf{w}_t - \widehat{\mathbf{w}}_t\rangle_{\mathcal{H}}\right] \\ &= \mathbb{E}\left[\sum_{t=1}^T \langle\Phi(\mathbf{x}_*), E_t\mathbf{w}_t\rangle_{\mathcal{H}}\right]. \end{aligned}$$

Recall that

$$C_T^{\mathbf{x}_*} = B\sqrt{\lambda}\mathbb{E}\left[\sum_{t=1}^T \|\Phi(\mathbf{x}_*)\|_{(\Sigma_t^\lambda)^{-1}}\right] = B\,\mathbb{E}\left[\sum_{t=1}^T \|\Phi(\mathbf{x}_*)\|_{E_t}\right].$$

We now combine $B_T^{\mathbf{x}_*}$ and $C_T^{\mathbf{x}_*}$:

$$-B_T^{\mathbf{x}_*} - C_T^{\mathbf{x}_*} = -\mathbb{E}\left[\sum_{t=1}^T \langle\Phi(\mathbf{x}_*), E_t\mathbf{w}_t\rangle_{\mathcal{H}}\right] - B\,\mathbb{E}\left[\sum_{t=1}^T \|\Phi(\mathbf{x}_*)\|_{E_t}\right].$$

Since $E_t \succeq 0$, Cauchy–Schwarz in the E_t-seminorm yields

$$|\langle\Phi(\mathbf{x}_*), E_t\mathbf{w}_t\rangle_{\mathcal{H}}| = |\langle E_t^{1/2}\Phi(\mathbf{x}_*), E_t^{1/2}\mathbf{w}_t\rangle_{\mathcal{H}}| \le \|\Phi(\mathbf{x}_*)\|_{E_t}\,\|\mathbf{w}_t\|_{E_t}.$$

Therefore,

$$\begin{aligned}
-B_T^{\mathbf{x}_*} - C_T^{\mathbf{x}_*} &\le \mathbb{E}\left[\sum_{t=1}^{T} \left(\|\Phi(\mathbf{x}_*)\|_{E_t} \|\mathbf{w}_t\|_{E_t} - B\|\Phi(\mathbf{x}_*)\|_{E_t}\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \|\Phi(\mathbf{x}_*)\|_{E_t}\left(\|\mathbf{w}_t\|_{E_t} - B\right)\right] \\
&\le 0,
\end{aligned}$$

where the last inequality holds due to $\|\mathbf{w}_t\|_{E_t} \le \|\mathbf{w}_t\|_{\mathcal{H}} \le B$ by Assumption 1.

□

A.4 Proof of Lemma 3 (Regret Bound via Exponential Weights)

The following lemma bounds the estimated regret term arising from the exponential-weights update. Under a suitable exploration condition, the proxy losses are uniformly bounded, allowing a standard exponential-weights analysis. The resulting bound separates the contribution of the effective dimension term, the entropy term $\log|\mathcal{X}|/\eta$, and the cost of explicit exploration.

Lemma 3 (Regret Bound via EW)**.** *Under Assumption 1, for any $\lambda > 0$, $\eta > 0$, and $\gamma \in (0, 1]$, let*

$$\nu_D \in \operatorname*{arg\,max}_{\nu \in \Delta(\mathcal{X})} \log\det\left(\mathbb{E}_{\mathbf{x}\sim\nu}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top] + \tfrac{\lambda}{\gamma} I_{\mathcal{H}}\right)$$

be the regularized D-optimal exploration distribution, and define $\Sigma(\nu_D) := \mathbb{E}_{\mathbf{x}\sim\nu_D}[\Phi(\mathbf{x})\Phi(\mathbf{x})^\top]$. Under the condition $d_{\text{eff}}(\Sigma(\nu_D), \lambda/\gamma) \le \gamma/(2\eta B)$, and $\lambda \le 1/(2\eta B)$, Algorithm 1 satisfies

$$\mathcal{R}_T^{\widehat{\ell}} \le 2\gamma BT + \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta \cdot \mathbb{E}\left[\sum_{t=1}^{T} d_{\text{eff}}(\Sigma_t, \lambda)\right] + B\sqrt{\lambda\gamma T \sum_{t=1}^{T} d_{\text{eff}}(\Sigma(\nu_D), \lambda/\gamma)}.$$

Proof. We can decompose $\mathcal{R}_T^{\widehat{\ell}}$ into two terms using $p_t = (1-\gamma)q_t + \gamma\nu_G$:

$$\begin{aligned}
\mathcal{R}_T^{\widehat{\ell}} &= (1-\gamma)\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}}\left(q_t(\mathbf{x}) - p_*(\mathbf{x})\right)\widehat{\ell}_t(\mathbf{x})\right]}_{\texttt{Term (A)}} \\
&\quad + \gamma\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}}\left(\nu_G(\mathbf{x}) - p_*(\mathbf{x})\right)\widehat{\ell}_t(\mathbf{x})\right]}_{\texttt{Term (B)}}.
\end{aligned}$$

Bounding Term (A). When $\gamma = 1$, the coefficient $1-\gamma$ makes Term (A) vanish and the regret reduces to $\gamma \cdot$ `Term (B)`. It therefore suffices to consider $\gamma \in (0, 1)$ in the following. Term (A) is the regret of the distribution q_t generated by the exponential weights algorithm. According to Lemma 5, we have $|\eta\widehat{\ell}_t(\mathbf{x})| \le 1$. Then, the standard analysis for the exponential weights algorithm (e.g., see [Bubeck et al., 2012, Theorem 1]) yields

$$\begin{aligned}
\texttt{Term (A)} &\le \eta\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} q_t(\mathbf{x})\widehat{\ell}_t(\mathbf{x})^2\right] + \frac{\log|\mathcal{X}|}{\eta} \\
&\le \frac{\eta}{1-\gamma}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\widehat{\ell}_t(\mathbf{x})^2\right] + \frac{\log|\mathcal{X}|}{\eta} \\
&= \frac{\eta}{1-\gamma}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_{\mathbf{x}_t\sim p_t}\Big[\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\widehat{\ell}_t(\mathbf{x})^2 \mid \mathcal{F}_{t-1}\Big]\right] + \frac{\log|\mathcal{X}|}{\eta} \qquad (19)
\end{aligned}$$

where the second inequality uses $p_t(\mathbf{x}) = (1-\gamma)q_t(\mathbf{x}) + \gamma\nu_G(\mathbf{x}) \ge (1-\gamma)q_t(\mathbf{x})$.

We can further upper bound the quadratic term on the right-hand side by

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}_t\sim p_t}\left[\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\widehat{\ell}_t(\mathbf{x})^2 \,\middle|\, \mathcal{F}_{t-1}\right] \\
&\le\ 2\,\mathbb{E}_{\mathbf{x}_t\sim p_t}\left[\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\,|\langle\Phi(\mathbf{x}_t),\mathbf{w}_t\rangle_{\mathcal{H}}|^2\cdot\langle\Phi(\mathbf{x}),(\Sigma_t^\lambda)^{-1}\Phi(\mathbf{x}_t)\rangle_{\mathcal{H}}^2 \,\middle|\, \mathcal{F}_{t-1}\right] + 2\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})c_t(\mathbf{x})^2 \\
&\le\ 2B^2\,\mathbb{E}_{\mathbf{x}_t\sim p_t}\left[\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})\,\langle\Phi(\mathbf{x}),(\Sigma_t^\lambda)^{-1}\Phi(\mathbf{x}_t)\rangle_{\mathcal{H}}^2 \,\middle|\, \mathcal{F}_{t-1}\right] + 2\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})c_t(\mathbf{x})^2 \\
&\le\ 2B^2 d_{\mathrm{eff}}(\Sigma_t,\lambda) + 2\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})c_t(\mathbf{x})^2, \qquad (20)
\end{aligned}$$

where:

- the first inequality follows from the definition $\widehat{\ell}_t(\mathbf{x}) = \langle\Phi(\mathbf{x}_t),\mathbf{w}_t\rangle_{\mathcal{H}}\langle\Phi(\mathbf{x}),(\Sigma_t^\lambda)^{-1}\Phi(\mathbf{x}_t)\rangle_{\mathcal{H}} - c_t(\mathbf{x})$ and the fact that $(u-v)^2 \le 2u^2+2v^2$ with $u = \langle\Phi(\mathbf{x}_t),\mathbf{w}_t\rangle_{\mathcal{H}}\langle\Phi(\mathbf{x}),(\Sigma_t^\lambda)^{-1}\Phi(\mathbf{x}_t)\rangle_{\mathcal{H}}$ and $v = c_t(\mathbf{x})$;
- the second inequality uses Assumption 1;
- the last step applies Lemma 7.

For the second term in (20), using the definition of $c_t(\mathbf{x})$, we have

$$\sum_{\mathbf{x}\in\mathcal{X}} p_t(\mathbf{x})c_t(\mathbf{x})^2 \ = B^2\lambda\,\mathrm{Tr}\big((\Sigma_t^\lambda)^{-1}\Sigma_t\big) = B^2\lambda d_{\mathrm{eff}}(\Sigma_t,\lambda). \qquad (21)$$

Substituting (20) and (21) into (19) gives

$$(1-\gamma)\cdot\texttt{Term (A)} \ \le 2(1+\lambda)B^2\eta\cdot\sum_{t=1}^{T}\mathbb{E}\left[d_{\mathrm{eff}}(\Sigma_t,\lambda)\right] + \frac{\log|\mathcal{X}|}{\eta}.$$

Bounding Term (B). Using the tower property together with Lemma 4, we have

$$\texttt{Term (B)} = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{x}\in\mathcal{X}}(\nu_G(\mathbf{x})-p_*(\mathbf{x}))\Big(\langle\Phi(\mathbf{x}),(\Sigma_t^\lambda)^{-1}\Sigma_t\,\mathbf{w}_t\rangle_{\mathcal{H}} - c_t(\mathbf{x})\Big)\right].$$

We bound the two contributions separately. For the first part, since $(\Sigma_t^\lambda)^{-1}\Sigma_t \preceq I_{\mathcal{H}}$ and $\|\mathbf{w}_t\|_{\mathcal{H}} \le B$, we have

$$\big|\langle\Phi(\mathbf{x}),(\Sigma_t^\lambda)^{-1}\Sigma_t\,\mathbf{w}_t\rangle_{\mathcal{H}}\big| \le \|\Phi(\mathbf{x})\|_{\mathcal{H}}\,\|(\Sigma_t^\lambda)^{-1}\Sigma_t\,\mathbf{w}_t\|_{\mathcal{H}} \le B \quad \text{for all } \mathbf{x}\in\mathcal{X}.$$

Moreover, using $\sum_{\mathbf{x}\in\mathcal{X}}|\nu_G(\mathbf{x})-p_*(\mathbf{x})| \le 2$ gives

$$\left|\sum_{\mathbf{x}\in\mathcal{X}}(\nu_G(\mathbf{x})-p_*(\mathbf{x}))\langle\Phi(\mathbf{x}),(\Sigma_t^\lambda)^{-1}\Sigma_t\,\mathbf{w}_t\rangle_{\mathcal{H}}\right| \le 2B.$$

For the correction term,

$$-\sum_{\mathbf{x}\in\mathcal{X}}(\nu_G(\mathbf{x})-p_*(\mathbf{x}))c_t(\mathbf{x}) = c_t(\mathbf{x}_*) - \sum_{\mathbf{x}\in\mathcal{X}}\nu_G(\mathbf{x})\,c_t(\mathbf{x}) \le c_t(\mathbf{x}_*),$$

since $\sum_{\mathbf{x}\in\mathcal{X}}\nu_G(\mathbf{x})\,c_t(\mathbf{x}) \ge 0$.

Combining both contributions and summing over $t=1,\dots,T$ and taking expectation yields

$$\texttt{Term (B)} \le 2BT + C_T^{\mathbf{x}_*}.$$

We can further upper bound the term $C_T^{\mathbf{x}_*}$ by

$$C_T^{\mathbf{x}_*} = B\mathbb{E}\Big[\sum_{t=1}^{T}\|\Phi(\mathbf{x}_*)\|_{E_t}\Big] \le B\mathbb{E}\left[\sqrt{T\sum_{t=1}^{T}\|\Phi(\mathbf{x}_*)\|_{E_t}^2}\right],$$

where $E_t = \lambda(\Sigma_t^\lambda)^{-1}$ and the last step is due to the Cauchy–Schwarz inequality. Now, writing $\|\Phi(\mathbf{x}_*)\|_{E_t}^2 = \lambda\|\Phi(\mathbf{x}_*)\|_{(\Sigma_t^\lambda)^{-1}}^2$, and using $\Sigma_t \succeq \gamma\Sigma(\nu_G)$, we have

$$(\Sigma_t^\lambda)^{-1} \preceq \frac{1}{\gamma}\Big(\Sigma(\nu_G) + \frac{\lambda}{\gamma}I_{\mathcal{H}}\Big)^{-1}.$$

Hence,

$$\|\Phi(\mathbf{x}_*)\|_{E_t}^2 \le \frac{\lambda}{\gamma}\mathrm{Tr}\Big[\big(\Sigma(\nu_D) + \frac{\lambda}{\gamma}I_{\mathcal{H}}\big)^{-1}\Sigma(\nu_D)\Big] = \frac{\lambda}{\gamma}d_{\mathrm{eff}}(\Sigma(\nu_D), \lambda/\gamma),$$

where the inequality holds following the same arguments in deriving (10) in the proof of Lemma 5. Substituting back, we obtain

$$C_T^{\mathbf{x}_*} \le B\sqrt{\frac{\lambda T}{\gamma}\sum_{t=1}^{T} d_{\mathrm{eff}}(\Sigma(\nu_D), \lambda/\gamma)}$$

Finally, combining the bounds for $(1-\gamma)\cdot$ `Term (A)` and $\gamma\cdot$ `Term (B)` completes the proof. □

A.5 Proof of Corollary 1 and Applications to Specific Kernels

Here we provide the omitted proof details for Corollary 1 and the subsequent applications to specific kernels.

Proof of Corollary 1. Recall that Theorem 1 bounds the regret in terms of

$$d_*(\rho) = \max_{\nu\in\Delta(\mathcal{X})} \mathrm{Tr}\left((\Sigma(\nu) + \rho I_{\mathcal{H}})^{-1}\Sigma(\nu)\right),$$

where $\Sigma(\nu) = \mathbb{E}_{\mathbf{x}\sim\nu}\left[\Phi(x)\Phi(x)^\top\right]$. Fixing $\alpha = 1$, we have $\lambda = 1/T$ and Lemma 8 gives

$$d_*(\lambda) \le 2\max_{\mathbf{x}_1,\dots,\mathbf{x}_T\in\mathcal{X}} \log\det\left(I + \sum_{t=1}^{T}\Phi(x_t)\Phi(x_t)^\top\right) =: 4\bar{\gamma}_T.$$

We will also need to control $d_*(\lambda/\gamma)$, where γ denotes the mixing coefficient in Algorithm 1. Since $\gamma \in (0, 1]$, we have

$$\frac{\lambda}{\gamma} = \frac{1}{\gamma T} \ge \frac{1}{T}.$$

Moreover, since $d_*(\rho)$ is non-increasing in ρ, we have

$$d_*(\lambda/\gamma) \le d_*(\lambda) \le 4\bar{\gamma}_T. \tag{22}$$

We choose

$$\eta = \frac{1}{2B}\sqrt{\frac{\log(e|\mathcal{X}|)}{2(1+\lambda)d_*(\lambda)T}}, \qquad \gamma = \min\left\{\sqrt{\frac{2d_*(\lambda)\log(e|\mathcal{X}|)}{(1+\lambda)T}}, 1\right\}. \tag{23}$$

It suffices to consider $T \ge 2d_*(\lambda)\log(e|\mathcal{X}|)$ and $d_*(\lambda) \ge 1/2$, as the remaining cases already satisfy the claimed bound. Indeed, for $T < 2d_*(\lambda)\log(e|\mathcal{X}|)$, we have $\mathcal{R}_T \le 2BT \le 2\sqrt{2}B\sqrt{Td_*(\lambda)\log(e|\mathcal{X}|)}$. For $d_*(\lambda) < 1/2$, fix any $\mathbf{x} \in \mathcal{X}$ and take ν with $\nu(\mathbf{x}) = 1$. Then $\Sigma(\nu) = \Phi(\mathbf{x})\Phi(\mathbf{x})^\top$, so

$$d_*(\lambda) \ge \mathrm{Tr}\big((\Sigma(\nu) + \lambda I_{\mathcal{H}})^{-1}\Sigma(\nu)\big) = \frac{\mathcal{K}(\mathbf{x},\mathbf{x})}{\mathcal{K}(\mathbf{x},\mathbf{x}) + \lambda}.$$

It follows that $\mathcal{K}(\mathbf{x},\mathbf{x}) \le 2\lambda d_*(\lambda)$ for any $\mathbf{x} \in \mathcal{X}$. Thus $\mathcal{R}_T \le 2BT\sqrt{2\lambda d_*(\lambda)} = 2\sqrt{2}B\sqrt{Td_*(\lambda)}$. Under these conditions, we have $\gamma < 1$, and recalling that $\lambda = 1/T$, a direct substitution readily verifies that $\lambda \le 1/(2\eta B)$. Moreover, we observe from (22) that

$$d_*(\lambda/\gamma) \le 2d_*(\lambda) = \frac{\gamma}{2\eta B}.$$

Hence, both conditions of Theorem 1 are satisfied. Applying Theorem 1 (with $\lambda = 1/T$), we obtain

$$\mathcal{R}_T \le \frac{\log|\mathcal{X}|}{\eta} + 2\Big(1+\frac{1}{T}\Big)B^2\eta d_*(\lambda)T + 4B\sqrt{Td_*(\lambda)} + B\sqrt{\gamma T d_*(\lambda)} + 2\gamma BT.$$

The first two terms satisfy

$$\frac{\log|\mathcal{X}|}{\eta} + 2\Big(1+\frac{1}{T}\Big)B^2\eta d_*(\lambda)T \le 4B\sqrt{(1+\lambda)d_*(\lambda)T\log(e|\mathcal{X}|)},$$

and the fourth term satisfies the following since $\gamma \le 1$:

$$B\sqrt{\gamma T d_*(\lambda)} \le B\sqrt{Td_*(\lambda)}.$$

For the final term, we have

$$2\gamma BT = 2BT\sqrt{\frac{2d_*(\lambda)\log(e|\mathcal{X}|)}{(1+\lambda)T}} \le 2B\sqrt{2d_*(\lambda)T\log(e|\mathcal{X}|)}\,.$$

Combining these bounds and using (22) gives

$$\mathcal{R}_T = \mathcal{O}\left(B\sqrt{d_*(\lambda)T\log(e|\mathcal{X}|)}\right) = \mathcal{O}\left(B\sqrt{\bar{\gamma}_T T\log(|\mathcal{X}|)}\right). \tag{24}$$

□

Application to specific kernels. For the Matérn(ν, d) kernel with $\nu > 0$, the standard information-gain bound is given by [Vakili et al., 2021, Janz et al., 2026]

$$\bar{\gamma}_T = \widetilde{\mathcal{O}}\left(T^{d/(2\nu+d)}\right).$$

Substituting this into (24), we obtain

$$\begin{aligned}
\mathcal{R}_T &= \widetilde{\mathcal{O}}\left(B\sqrt{T\cdot T^{d/(2\nu+d)}\log(e|\mathcal{X}|)}\right) \\
&= \widetilde{\mathcal{O}}\left(BT^{\frac{1}{2}+\frac{d}{2(2\nu+d)}}\sqrt{\log(e|\mathcal{X}|)}\right) \\
&= \widetilde{\mathcal{O}}\left(BT^{(\nu+d)/(2\nu+d)}\sqrt{\log(e|\mathcal{X}|)}\right).
\end{aligned}$$

In particular, if $\log|\mathcal{X}| = \mathcal{O}(\log T)$ and B is treated as a constant, then

$$\mathcal{R}_T = \widetilde{\mathcal{O}}\left(T^{(\nu+d)/(2\nu+d)}\right),$$

which proves the desired claim.

For the Squared Exponential kernel, the standard information-gain bound is given by [Vakili et al., 2021]

$$\bar{\gamma}_T = \mathcal{O}\big((\log T)^{d+1}\big)\,.$$

Substituting this into (24), we obtain

$$\begin{aligned}
\mathcal{R}_T &= \mathcal{O}\left(B\sqrt{T\log(e|\mathcal{X}|)(\log T)^{d+1}}\right) \\
&= \mathcal{O}\left(B\sqrt{T\log(e|\mathcal{X}|)}(\log T)^{(d+1)/2}\right).
\end{aligned}$$

In particular, if $\log|\mathcal{X}| = \mathcal{O}(\log T)$ and B is treated as a constant, then

$$\mathcal{R}_T = \mathcal{O}\left(\sqrt{T}(\log T)^{(d+2)/2}\right) = \widetilde{\mathcal{O}}(\sqrt{T}),$$

which proves the desired claim.

A.6 Proofs of Corollary 2 and Corollary 3 (Polynomial and Exponential Eigendecay)

Proof of Corollary 2. Suppose $\mathcal{K}$ has (C, β)-polynomial eigenvalue decay defined in Definition 1. Lemma 6 shows that the effective dimension can be upper bounded by

$$d_*(\lambda) \leq m + \frac{Cm^{1-\beta}}{(\beta-1)\lambda}$$

and

$$d_*(\lambda/\gamma) \leq m + \frac{\gamma C m^{1-\beta}}{(\beta-1)\lambda}.$$

Then, Theorem 1 gives

$$\begin{aligned}\mathcal{R}_T \leq 2\gamma BT + \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta T\left(m + \frac{Cm^{1-\beta}}{(\beta-1)\lambda}\right) \\ + BT\sqrt{\lambda\gamma m + \frac{\gamma^2 C m^{1-\beta}}{\beta-1}} + 4BT\sqrt{\lambda m + \frac{Cm^{1-\beta}}{\beta-1}}.\end{aligned}$$

We now choose

$$m := \left\lceil T^{1/\beta} \right\rceil, \qquad \lambda := m^{-\beta}, \qquad \gamma := m^{-\frac{\beta-1}{2}}, \qquad \eta := \frac{1}{2B(1 + C/(\beta-1))} m^{-\frac{\beta+1}{2}}. \quad (25)$$

By this choice,

$$m + \frac{Cm^{1-\beta}}{(\beta-1)\lambda} = \left(1 + \frac{C}{\beta-1}\right) m$$

and

$$m + \frac{\gamma C m^{1-\beta}}{(\beta-1)\lambda} = \left(1 + \frac{C\gamma}{\beta-1}\right) m \leq \left(1 + \frac{C}{\beta-1}\right) m,$$

where we used $\gamma \leq 1$. Substituting these bounds into the regret bound yields

$$\begin{aligned}\mathcal{R}_T \leq 2\gamma BT + \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta T\left(1 + \frac{C}{\beta-1}\right) m \\ + BT\sqrt{\lambda\gamma m\left(1 + \frac{C\gamma}{\beta-1}\right)} + 4BT\sqrt{\lambda m\left(1 + \frac{C}{\beta-1}\right)}.\end{aligned}$$

Since $m = \lceil T^{1/\beta} \rceil$, we have $m \leq T^{1/\beta} + 1$ and $m^{-a} \leq T^{-a/\beta}$ for any $a > 0$. Therefore,

$$\begin{aligned}\mathcal{R}_T \leq 2B\left(1 + \frac{C}{\beta-1}\right)\log|\mathcal{X}|\ \left(T^{1/\beta} + 1\right)^{\frac{\beta+1}{2}} \\ + BT^{\frac{\beta+1}{2\beta}}\left(3 + T^{-1} + 4\sqrt{1 + \frac{C}{\beta-1}}\right) \\ + BT^{\frac{\beta+3}{4\beta}}\sqrt{1 + \frac{CT^{-\frac{\beta-1}{2\beta}}}{\beta-1}}.\end{aligned}$$

Since the last term is lower order for $\beta > 1$, we obtain

$$\mathcal{R}_T = \mathcal{O}\left(B\left(1 + \left(1 + \frac{C}{\beta-1}\right)\log|\mathcal{X}| + \sqrt{1 + \frac{C}{\beta-1}}\right) T^{\frac{\beta+1}{2\beta}}\right).$$

It remains to verify the conditions of Theorem 1. By the parameter choice,

$$d_{\text{eff}}(\Sigma(\nu_D), \lambda/\gamma) \leq m + \frac{\gamma C m^{1-\beta}}{(\beta-1)\lambda} = \left(1 + \frac{C\gamma}{\beta-1}\right) m \leq \left(1 + \frac{C}{\beta-1}\right) m = \frac{\gamma}{2\eta B}.$$

Moreover,

$$\lambda = m^{-\beta} \leq 1 \leq \left(1 + \frac{C}{\beta-1}\right) m^{\frac{\beta+1}{2}} = \frac{1}{2\eta B}.$$

Thus both conditions of Theorem 1 are satisfied. □

Proof of Corollary 3. Suppose $\mathcal{K}$ has (C, β)-exponential eigenvalue decay defined in Definition 1. Lemma 6 shows that the effective dimension can be upper bounded by

$$d_*(\lambda) \le m + \frac{Ce^{-\beta m}}{(e^\beta - 1)\lambda}, \qquad d_*(\lambda/\gamma) \le m + \frac{\gamma C e^{-\beta m}}{(e^\beta - 1)\lambda}.$$

Then, Theorem 1 gives

$$\begin{aligned}\mathcal{R}_T \le 2\gamma BT + \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta T\left(m + \frac{Ce^{-\beta m}}{(e^\beta-1)\lambda}\right) \\ + BT\sqrt{\lambda\gamma m + \frac{\gamma^2 Ce^{-\beta m}}{e^\beta - 1}} + 4BT\sqrt{\lambda m + \frac{Ce^{-\beta m}}{e^\beta - 1}}.\end{aligned}$$

We now choose

$$m := \left\lceil \frac{\log T}{\beta} \right\rceil, \qquad \lambda := e^{-\beta m}, \tag{26}$$

and

$$\gamma := \sqrt{\frac{\left(m + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}{T}}, \qquad \eta := \frac{\gamma}{2B\left(m + \frac{C}{e^\beta - 1}\right)}.$$

Since $m = \lceil (\log T)/\beta \rceil$, we have

$$\lambda = e^{-\beta m} \le \frac{1}{T}.$$

Moreover,

$$m + \frac{Ce^{-\beta m}}{(e^\beta - 1)\lambda} = m + \frac{C}{e^\beta - 1},$$

and

$$m + \frac{\gamma C e^{-\beta m}}{(e^\beta - 1)\lambda} = m + \frac{\gamma C}{e^\beta - 1} \le m + \frac{C}{e^\beta - 1},$$

where the last inequality uses $\gamma \le 1$.

Substituting these bounds into the regret bound yields

$$\begin{aligned}\mathcal{R}_T \le 2\gamma BT + \frac{\log|\mathcal{X}|}{\eta} + 2(1+\lambda)B^2\eta T\left(m + \frac{C}{e^\beta - 1}\right) \\ + BT\sqrt{\lambda\gamma\left(m + \frac{\gamma C}{e^\beta - 1}\right)} + 4BT\sqrt{\lambda\left(m + \frac{C}{e^\beta - 1}\right)}.\end{aligned}$$

We estimate the terms separately. First, by the choice of γ,

$$2\gamma BT = 2B\sqrt{T\left(m + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}.$$

For the entropy term, since $\log|\mathcal{X}| \le \log(e|\mathcal{X}|)$,

$$\frac{\log|\mathcal{X}|}{\eta} = \frac{2B\left(m + \frac{C}{e^\beta - 1}\right)\log|\mathcal{X}|}{\gamma} \le 2B\sqrt{T\left(m + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}.$$

For the variance term, using $\lambda \le 1$,

$$\begin{aligned}2(1+\lambda)B^2\eta T\left(m + \frac{C}{e^\beta - 1}\right) &\le 4B^2\eta T\left(m + \frac{C}{e^\beta - 1}\right) \\ &= 2B\gamma T \\ &= 2B\sqrt{T\left(m + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}.\end{aligned}$$

For the first square-root term, using $\lambda \le 1/T$ and $\gamma \le 1$,

$$\begin{aligned} BT\sqrt{\lambda\gamma\left(m + \frac{\gamma C}{e^\beta - 1}\right)} &\le BT\sqrt{\frac{\gamma}{T}\left(m + \frac{C}{e^\beta - 1}\right)} \\ &= B\sqrt{T\gamma\left(m + \frac{C}{e^\beta - 1}\right)} \\ &\le B\sqrt{T\left(m + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}, \end{aligned}$$

where the last inequality uses $\gamma \le 1 \le \log(e|\mathcal{X}|)$. Similarly,

$$\begin{aligned} 4BT\sqrt{\lambda\left(m + \frac{C}{e^\beta - 1}\right)} &\le 4BT\sqrt{\frac{1}{T}\left(m + \frac{C}{e^\beta - 1}\right)} \\ &= 4B\sqrt{T\left(m + \frac{C}{e^\beta - 1}\right)} \\ &\le 4B\sqrt{T\left(m + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}. \end{aligned}$$

Combining the above estimates gives

$$\mathcal{R}_T = \mathcal{O}\left(B\sqrt{T\left(m + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}\right).$$

Since $m = \left\lceil \frac{\log T}{\beta} \right\rceil$, we obtain

$$\mathcal{R}_T = \mathcal{O}\left(B\sqrt{T\left(\frac{\log T}{\beta} + \frac{C}{e^\beta - 1}\right)\log(e|\mathcal{X}|)}\right).$$

In particular, when (B, β, C) are constant and $\log|\mathcal{X}| = O(\log T)$, we have $\mathcal{R}_T = O\Big(\sqrt{T}\log T\Big)$.

It remains to verify the conditions of Theorem 1. By the parameter choice,

$$d_{\text{eff}}(\Sigma(\nu_D), \lambda/\gamma) \le m + \frac{\gamma C}{e^\beta - 1} \le m + \frac{C}{e^\beta - 1} = \frac{\gamma}{2\eta B}.$$

Moreover,

$$\lambda = e^{-\beta m} \le 1 \le \frac{m + \frac{C}{e^\beta - 1}}{\gamma} = \frac{1}{2\eta B},$$

where we used $\gamma \le 1$. Thus both conditions of Theorem 1 are satisfied. □

B Further Discussion on Existing and Concurrent Work

B.1 Lower Bound from Chatterji et al. [2019]

In [Chatterji et al., 2019, Theorem 43], a lower bound of $\widetilde{\Omega}(T^{\frac{\beta+1}{2\beta}})$ is given for adversarial kernel bandits in a setting where the kernel has eigenvalues decaying as $O(j^{-\beta})$. While our upper bound also exhibits $T^{\frac{\beta+1}{2\beta}}$ dependence, there are important differences in the underlying assumptions that make these bounds incomparable and/or introduce a gap between them.

Perhaps most importantly, the kernel used in [Chatterji et al., 2019, Theorem 43] is an infinite-dimensional linear kernel, as opposed to a kernel with low-dimensional input such as $[0,1]^d$. Their proof can be adapted to a finite action set of the form $\{-1,1\}^M$ for suitably-chosen M (with all values beyond M implicitly set to zero), but this requires M to grow at rate $T^{\frac{1}{2\beta}}$, thus inducing an

action set $|\mathcal{X}|$ satisfying $\log|\mathcal{X}| = \Omega(T^{\frac{1}{2\beta}})$, rather than the usual $\log|\mathcal{X}| = O(\log T)$. Accordingly, the $\log|\mathcal{X}|$ dependence in our upper bound (and in related upper bounds, including that of Chatterji et al. [2019]) would be significant due to being polynomial in T, and the overall regret scaling would be strictly higher than $T^{\frac{\beta+1}{2\beta}}$.

Two other notable differences are (i) our Definition 1 considers uniform bounds with respect to the measure p, whereas the decay in [Chatterji et al., 2019, Theorem 43] is with respect to a fixed measure, and (ii) we assume an oblivious adversary, whereas the adversary in [Chatterji et al., 2019, Theorem 43] is non-oblivious. These differences may potentially be less of a hurdle to overcome compared to the previous paragraph, but they are nevertheless also important.

B.2 Algorithm and Upper Bound from Iwazaki [2026]

Here we highlight a few differences between our main results and the algorithm and its upper bound from the concurrent work Iwazaki [2026], but note that all of these differences are ultimately minor:

1. We study the expected regret under an oblivious adversary, whereas Iwazaki [2026] studies the pseudoregret under an adaptive adversary. However, this distinction ultimately has almost no impact on the analysis, and it is straightforward to adapt either analysis to the other setting.
2. We focus more on the effective dimension d_*, whereas Iwazaki [2026] focuses more on the maximum information gain γ_T, but the two are well known to be closely related, e.g., see Calandriello et al. [2019] as well as our Lemma 8.
3. We use a G-optimal design for our exploration term in Algorithm 1, whereas Iwazaki [2026] uses maximum variance reduction (MVR). These are essentially equivalent for statistical/regret purposes (though possibly different in terms of computation), since MVR is a greedy Frank-Wolfe like procedure for the log-det (D-optimal design) objective, whose optimality conditions are equivalent to G-optimal design. In fact, it is explicitly stated in [Iwazaki, 2026, Appendix I] that exploration via a G-optimal design yields the same regret guarantee.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The theoretical results are presented in Theorem 1 and Corollaries 2 and 3.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The computation efficiency of the method is discussed in Remark 2 (Section 4.2).

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We explicitly state these assumptions and conditions in Assumption 1 and Definition 1.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [N/A]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [N/A]

Justification: [N/A]

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conforms to the NeurIPS Code of Ethics, and we are not aware of any ethical concerns raised by this work.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The paper is theoretical in nature and focuses on foundational algorithmic analysis, with no direct societal impact.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer:[N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification:[N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification:[N/A]

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.