CLOSING THE REGRET GAP FOR GAUSSIAN PROCESS BANDITS WITH SQUARED EXPONENTIAL KERNEL

Anonymous authors
Paper under double-blind review

ABSTRACT

We study minimax lower bounds for Gaussian process bandits with the squared exponential (SE) kernel over compact domains $\mathcal{X} \subset \mathbb{R}^d$ with non-empty interior, where the reward function is fixed and has a bounded norm in the known reproducing kernel Hilbert space (RKHS). A fundamental open question in this setting has been the dimension-dependent $(\ln T)^{d/4}$ gap between existing upper and lower regret bounds. Although recently resolved on the hypersphere, closing this gap on general compact domains has remained open. In this work, we completely resolve this problem. By establishing a norm embedding from the bandlimited Sinc RKHS into the SE RKHS and utilizing the Prolate Spheroidal Wave Functions (PSWFs), we construct a hypothesis class yielding an $\Omega(\sqrt{T(\ln T)^d(\ln\ln T)^{-d}})$ cumulative regret lower bound and an $\Omega(\epsilon^{-2}(\ln\frac{1}{\epsilon})^d(\ln\ln\frac{1}{\epsilon})^{-d})$ sample complexity lower bound for ϵ-optimal point identification. These results match known upper bounds up to dimension-independent logarithmic factors.

1 INTRODUCTION

Kernelized multi-armed bandits and Gaussian process (GP) optimization serve as primary mathematical frameworks for black-box sequential optimization in machine learning, including robotics (Martinez-Cantin et al., 2007), hyperparameter tuning (Snoek et al., 2012), and experimental design (Lei et al., 2021). In the frequentist setting where the unknown objective function f resides in a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k(\mathcal{X})$ with bounded norm $\|f\|_{\mathcal{H}_k} \leq B$, fundamental research has focused on characterizing the algorithm-independent minimax regret limits.

Among commonly used covariance kernels, the squared exponential (SE) kernel $k(x, \tilde{x}) = \exp(-\|x - \tilde{x}\|_2^2/\theta)$ is ubiquitous. For the Matérn family, state-of-the-art upper and lower bounds match up to dimension-independent logarithmic factors (Camilleri et al., 2021; Li & Scarlett, 2022; Salgia et al., 2021; Scarlett et al., 2017; Vakili et al., 2021a; Valko et al., 2013). However, for the SE kernel, a persistent gap has remained. Specifically, on the Euclidean domain $\mathcal{X} = [0, 1]^d$, existing upper bounds provide $\mathcal{O}^*(\sqrt{T(\ln T)^d})$ cumulative regret and $\mathcal{O}^*(\sqrt{(\ln T)^d/T})$ simple regret (where $\mathcal{O}^*(\cdot)$ suppresses dimension-independent logarithmic terms), whereas the existing lower bounds of Cai & Scarlett (2021); Scarlett et al. (2017) only guarantee $\Omega(\sqrt{T(\ln T)^{d/2}})$ and $\Omega(\sqrt{(\ln T)^{d/2}/T})$, respectively. This discrepancy leaves a dimension-dependent $(\ln T)^{d/4}$ gap, which inflates exponentially with dimension d and remains a significant problem in non-parametric bandit theory.

Recently, Iwazaki (2026) partially tackled this challenge on the hypersphere $\mathbb{S}^d := \{x \in \mathbb{R}^{d+1} \mid \|x\|_2 = 1\}$ by exploiting spherical harmonics. Nevertheless, extending the lower bound to the canonical Euclidean domain $\mathcal{X} = [0, 1]^d$ remained unsolved. As discussed in Iwazaki (2026), the spherical harmonics approach fails on the hypercube because the Mercer eigensystem of the SE kernel under the uniform Lebesgue measure on $[0, 1]^d$ possesses no explicit closed-form expression, while surrogate measures (such as the Gaussian measure with Hermite polynomials) fail to capture the sharpness of the peak on bounded compact sets.

In this work, we resolve this open problem directly on the Euclidean hypercube $\mathcal{X} = [0, 1]^d$, and further demonstrate that our lower bounds extend to arbitrary compact subsets of $\mathbb{R}^d$ with non-empty interior.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Comparison of minimax bounds for the SE kernel on the Euclidean hypercube $\mathcal{X} = [0,1]^d$. For simple regret, the metric represents the time steps required to achieve simple regret at most $\epsilon > 0$.

	Cumulative regret	Simple regret
Upper bound (Iwazaki, 2026)	$\mathcal{O}^*\left(\sqrt{\frac{T(\ln T)^d}{(\ln\ln T)^d}}\right)$	$\mathcal{O}^*\left(\frac{(\ln\frac{1}{\epsilon})^d}{\epsilon^2(\ln\ln\frac{1}{\epsilon})^d}\right)$
Previous Lower bound (Scarlett et al., 2017)	$\Omega(\sqrt{T(\ln T)^{d/2}})$	$\Omega(\frac{1}{\epsilon^2}(\ln\frac{1}{\epsilon})^{d/2})$
Our Lower bound (This work)	$\Omega\left(\sqrt{\frac{T(\ln T)^d}{(\ln\ln T)^d}}\right)$	$\Omega\left(\frac{(\ln\frac{1}{\epsilon})^d}{\epsilon^2(\ln\ln\frac{1}{\epsilon})^d}\right)$

Contributions

- **Tight Regret Lower Bounds on Hypercube:** Under the standard hypercube domain $\mathcal{X} = [0,1]^d$, we prove that any algorithm must suffer a cumulative regret of $\Omega(\sqrt{T(\ln T)^d(\ln\ln T)^{-d}})$. Furthermore, for simple regret, finding an ϵ-optimal point requires at least $\Omega(\epsilon^{-2}(\ln\frac{1}{\epsilon})^d(\ln\ln\frac{1}{\epsilon})^{-d})$ evaluations. And our lower bounds extend to *any compact domain $\mathcal{X} \subset \mathbb{R}^d$ with non-empty interior*.
- **Novel PSWF-Sinc Hard Function Construction:** We overcome the lack of closed-form Mercer eigensystems by establishing a frequency-domain norm embedding inequality between the bandlimited Sinc RKHS and the SE RKHS. We then construct localized hypothesis functions using Prolate Spheroidal Wave Functions (PSWFs) and their Christoffel-Darboux kernels. Crucially, unlike the hyperspherical construction which inherently suffers from antipodal mirror peaks, our Euclidean construction ensures that the ϵ-optimal set is strictly confined to a single localized cell.
- **Complete Minimax Optimality:** Paired with the improved $O((\ln T)^{d+1}(\ln\ln T)^{-d})$ maximum information gain (MIG) bound, our results establish the optimal regret rate for SE kernel bandits on hypercube up to dimension-independent logarithmic terms. These results completely close the $(\ln T)^{d/4}$ gap on Euclidean domains (see Table 1).

2 Related Works

Gaussian Process Bandits Gaussian process (GP) bandits have been investigated across various settings, including Bayesian setting (Russo & Roy, 2014; Srinivas et al., 2010), noiseless feedback (Bull, 2011; Xu et al., 2024), non-stationary environment (Cai & Scarlett, 2025), and adversarially corrupted feedback (Bogunovic et al., 2018). For the frequentist noisy setting—where the unknown reward function resides in an RKHS norm ball and feedback is subject to stochastic noise—early foundational algorithms such as GP-UCB and IGP-UCB (Ray Chowdhury & Gopalan, 2017; Srinivas et al., 2010) established cumulative regret upper bounds scaling with the maximum information gain (MIG) γ_T as $\mathcal{O}^*(\gamma_T\sqrt{T})$. Subsequent works developed non-adaptive batched sampling (Li & Scarlett, 2022), phased exploration (Camilleri et al., 2021), and localized domain-shrinking/tree-search strategies (Salgia et al., 2021; Valko et al., 2013), which successfully reduced the upper bound to the fundamental barrier $\mathcal{O}^*(\sqrt{T\gamma_T})$.

For the squared exponential (SE) kernel on compact subsets of $\mathbb{R}^d$, the standard spectral bound on MIG was long known to be $\gamma_T = O((\ln T)^{d+1})$ (Srinivas et al., 2010; Vakili et al., 2021b). Recently, Iwazaki (2026) sharpened this estimate to $\gamma_T = O((\ln T)^{d+1}(\ln\ln T)^{-d})$. Consequently, combining non-adaptive sampling with this sharpened MIG yields the current state-of-the-art cumulative regret upper bound:

$$R_T \leq \mathcal{O}^*(\sqrt{T\gamma_T}) = \mathcal{O}^*\left(\sqrt{T(\ln T)^d(\ln\ln T)^{-d}}\right), \tag{1}$$

which serves as the benchmark target for lower bound investigations.

Lower Bound Developments. The benchmark worst-case lower bounds for GP bandits were developed by Scarlett et al. (2017) using bump functions derived from the inverse Fourier transform of compactly supported frequency profiles. Subsequent studies extended these bounds to robust (Bogunovic et al., 2018; Cai & Scarlett, 2021), non-stationary (Cai & Scarlett, 2025; Iwazaki & Takeno, 2025), and heavy-tailed environments (Ray Chowdhury & Gopalan, 2019). However, all these works inherited the fundamental $(\ln T)^{d/4}$ looseness for the SE kernel. While Iwazaki (2026) recently managed to close this gap on the hypersphere $\mathbb{S}^d$, extending the result to Euclidean domains was left as an open challenge due to domain-specific analytical limitations.

3 Problem Formulation and Existing Lower Bound Machinery

3.1 Basic Problem Setup

Let $\mathcal{X} = [0,1]^d$ be the d-dimensional unit hypercube. We consider a positive definite SE kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with lengthscale parameter $\theta > 0$:

$$k(x, \tilde{x}) = \exp\left(-\frac{\|x - \tilde{x}\|_2^2}{\theta}\right). \tag{2}$$

The target reward function $f : \mathcal{X} \to \mathbb{R}$ belongs to the known RKHS $\mathcal{H}_k(\mathcal{X})$ and satisfies $\|f\|_{\mathcal{H}_k(\mathcal{X})} \leq B$. We define $\mathcal{H}_k(\mathcal{X}, B) := \{f \in \mathcal{H}_k(\mathcal{X}) \mid \|f\|_{\mathcal{H}_k(\mathcal{X})} \leq B\}$.

At each sequential round $t = 1, \ldots, T$, the learner selects a query $x_t \in \mathcal{X}$ based on prior observations, observing noisy feedback $y_t = f(x_t) + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed Gaussian noises. The performance is assessed via the cumulative regret $R_T := \sum_{t=1}^{T}(\max_{x \in \mathcal{X}} f(x) - f(x_t))$ or the simple regret $r_T := \max_{x \in \mathcal{X}} f(x) - f(\hat{x}_T)$, where $\hat{x}_T \in \mathcal{X}$ is the point recommended after step T.

3.2 Change of Measure Framework

To establish algorithm-independent lower bounds, we follow the information-theoretic change of measure strategy on a finite hypothesis class (Cai & Scarlett, 2021; Scarlett et al., 2017).

Lemma 3.1 (Change of Measure, Lemma 1 in Cai & Scarlett (2021))**.** *Fix $f, \tilde{f} \in \mathcal{H}_k(\mathcal{X}, B)$, $\delta \in (0, 1/3)$, and an arbitrary decision algorithm. Let $(\mathcal{R}_j)_{j \in [M]}$ be a partition of $\mathcal{X}$ into M disjoint regions, and let $\mathcal{A}$ be an event measurable with respect to the T-step history. If $\mathbb{P}_f(\mathcal{A}) \geq 1 - \delta$ and $\mathbb{P}_{\tilde{f}}(\mathcal{A}) \leq \delta$, then*

$$\sum_{j \in [M]} \mathbb{E}_f[N_j(T)] \overline{D}_{f,\tilde{f}}^{(j)} \geq \ln \frac{1}{2.4\delta}, \tag{3}$$

where $N_j(T) := \sum_{t=1}^{T} \mathbb{I}\{x_t \in \mathcal{R}_j\}$ is the number of queries placed in $\mathcal{R}_j$, and $\overline{D}_{f,\tilde{f}}^{(j)} := \sup_{x \in \mathcal{R}_j} \mathrm{KL}(\mathcal{N}(f(x), \sigma^2) \,\|\, \mathcal{N}(\tilde{f}(x), \sigma^2)) = \sup_{x \in \mathcal{R}_j} \frac{(f(x) - \tilde{f}(x))^2}{2\sigma^2}$.

4 Main Results: Tight Lower Bounds on the Hypercube

Our main theoretical contributions are formal lower bounds on the Euclidean hypercube that completely remove the sub-optimality factor.

Theorem 4.1 (Simple Regret Lower Bound on Hypercube)**.** *Fix $\delta \in (0, 1/3)$, $\epsilon \in (0, 1/2)$, $B > 0$, and $T \in \mathbb{N}_+$. Consider the GP bandit problem on $\mathcal{X} = [0,1]^d$ with the SE kernel equation 2 for fixed constants $d \in \mathbb{N}_+$ and $\theta > 0$. If an algorithm guarantees $r_T \leq \epsilon$ with probability at least $1 - \delta$ for any $f \in \mathcal{H}_k(\mathcal{X}, B)$, then for sufficiently small ϵ/B, it is necessary that*

$$T = \Omega\left(\frac{\sigma^2}{\epsilon^2}\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)}\right)^d \ln\frac{1}{\delta}\right). \tag{4}$$

Theorem 4.2 (Cumulative Regret Lower Bound on Hypercube)**.** *Under the same setting as Theorem 4.1, assume that $\sigma^2(\ln(1/\delta))/B^2 = O(T)$ with a sufficiently small implied constant. Then, for*

any algorithm, there exists an objective function $f \in \mathcal{H}_k(\mathcal{X}, B)$ such that with probability at least δ:

$$R_T = \Omega\left(\sqrt{T\sigma^2 \left(\ln \frac{B^2T}{\sigma^2 \ln \frac{1}{\delta}}\right)^d \left(\ln \ln \frac{B^2T}{\sigma^2 \ln \frac{1}{\delta}}\right)^{-d}}\right). \tag{5}$$

Corollary 4.3 (Expected Regret Lower Bounds)**.** *Under the setting of Theorem 4.2, any algorithm must satisfy:*

$$\sup_{f \in \mathcal{H}_k(\mathcal{X},B)} \mathbb{E}_f[r_T] = \Omega\left(\sqrt{\frac{\sigma^2}{T} \left(\ln \frac{B^2T}{\sigma^2}\right)^d \left(\ln \ln \frac{B^2T}{\sigma^2}\right)^{-d}}\right), \tag{6}$$

$$\sup_{f \in \mathcal{H}_k(\mathcal{X},B)} \mathbb{E}_f[R_T] = \Omega\left(\sqrt{T\sigma^2 \left(\ln \frac{B^2T}{\sigma^2}\right)^d \left(\ln \ln \frac{B^2T}{\sigma^2}\right)^{-d}}\right). \tag{7}$$

Moreover, let $\mathcal{X} \subset \mathbb{R}^d$ be any compact domain with non-empty interior. Then, the regret lower bounds in Theorem 4.1, Theorem 4.2, and Corollary 4.3 hold on $\mathcal{X}$ with the exact same dependencies on T, ϵ, and δ, up to constants depending solely on $\mathcal{X}, d$, and θ. The detailed proof is deferred to Appendix E.

5 Proof Overview: The Sinc-PSWF Construction

To achieve Theorems 4.1 and 4.2, we need to construct a bump function that achieves peak width $w_\epsilon = \Theta(\frac{\ln \ln(B/\epsilon)}{\ln(B/\epsilon)})$ while maintaining its SE RKHS norm below B.

A central mathematical obstacle on the Euclidean hypercube $\mathcal{X} = [0,1]^d$ is that the integral operator associated with the squared exponential (SE) kernel under the uniform Lebesgue measure, possesses no known analytical, closed-form eigensystem. Consequently, constructing hard candidate functions through a direct Mercer spectral expansion of k_{SE} on $[0,1]^d$ is analytically intractable.

5.1 Prolate Spheroidal Wave Functions and Sinc RKHS

In this subsection, we review the theory of Prolate Spheroidal Wave Functions (PSWFs) and the associated bandlimited reproducing kernel Hilbert spaces. While spherical harmonics provide an explicit Mercer eigensystem on the hypersphere $\mathbb{S}^d$ (Iwazaki, 2026), they are inapplicable to the Euclidean hypercube. Instead, our Euclidean construction is fundamentally rooted in the spectral theory of the Sinc integral operator on compact intervals, pioneered by Slepian, Landau, and Pollak at Bell Laboratories (Landau & Pollak, 1961; Slepian & Pollak, 1961; Slepian, 1965).

The Energy Concentration Problem and the Sinc Kernel. A classical principle in Fourier analysis is the uncertainty principle: a non-trivial function in $L^2(\mathbb{R})$ cannot be simultaneously strictly time-limited (compactly supported in space) and strictly bandlimited (compactly supported in frequency). To characterize the optimal trade-off, Slepian & Pollak (1961) investigated the variational problem of finding bandlimited functions in the Paley–Wiener space

$$\mathcal{PW}_c := \left\{ f \in L^2(\mathbb{R}) \;\middle|\; \mathrm{supp}(\hat{f}) \subseteq [-c, c] \right\} \tag{8}$$

that maximize the fractional energy concentrated inside the compact spatial interval $[-1, 1]$:

$$\max_{f \in \mathcal{PW}_c \setminus \{0\}} \frac{\int_{-1}^{1} |f(x)|^2 \mathrm{d}x}{\int_{-\infty}^{\infty} |f(x)|^2 \mathrm{d}x}. \tag{9}$$

The reproducing kernel of the Hilbert space $\mathcal{PW}_c$ equipped with the standard $L^2(\mathbb{R})$ inner product is the one-dimensional Sinc kernel:

$$k_c(x, y) := \frac{\sin(c(x-y))}{\pi(x-y)}, \quad x, y \in \mathbb{R}, \tag{10}$$

where $c > 0$ represents the base bandwidth parameter.

It is revealed that the optimal functions are the eigenfunctions of the finite Sinc integral operator $\mathcal{T}_c : L^2([-1,1]) \to L^2([-1,1])$, defined by

$$(\mathcal{T}_c f)(x) := \int_{-1}^{1} \frac{\sin(c(x-y))}{\pi(x-y)} f(y)\,\mathrm{d}y. \tag{11}$$

Mercer Decomposition and Slepian's Super-Exponential Decay. Since the kernel k_c is symmetric, continuous, and strictly positive definite, Mercer's representation theorem ensures that $\mathcal{T}_c$ possesses a countable sequence of positive eigenvalues:

$$1 > \lambda_0(c) > \lambda_1(c) > \lambda_2(c) > \cdots > 0, \quad \lim_{n\to\infty} \lambda_n(c) = 0, \tag{12}$$

and an associated orthonormal basis of eigenfunctions $\{\psi_n(x;c)\}_{n=0}^{\infty}$ in $L^2([-1,1])$, known as the *Prolate Spheroidal Wave Functions* (PSWFs). They satisfy:

$$\int_{-1}^{1} \frac{\sin(c(x-y))}{\pi(x-y)} \psi_n(y;c)\,\mathrm{d}y = \lambda_n(c)\psi_n(x;c), \quad x \in [-1,1], \tag{13}$$

normalized such that $\int_{-1}^{1} \psi_n^2(x;c)\mathrm{d}x = 1$.

The reproducing kernel Hilbert space $\mathcal{H}_{k_c}([-1,1])$ associated with the Sinc kernel on the compact interval $[-1,1]$ admits the explicit spectral representation:

$$\mathcal{H}_{k_c}([-1,1]) = \left\{ f = \sum_{n=0}^{\infty} a_n \sqrt{\lambda_n(c)}\psi_n(\cdot;c) \;\middle|\; \|f\|_{\mathcal{H}_{k_c}}^2 = \sum_{n=0}^{\infty} a_n^2 < \infty \right\}. \tag{14}$$

Remarkably, as discovered by Slepian (1965), the integral operator $\mathcal{T}_c$ share the exact same eigenfunctions $\{\psi_n(x;c)\}_{n=0}^{\infty}$ with a singular second-order differential operator $\mathcal{L}_c$:

$$\mathcal{L}_c\psi(x) := -\frac{\mathrm{d}}{\mathrm{d}x}\left((1-x^2)\frac{\mathrm{d}\psi(x)}{\mathrm{d}x}\right) + c^2x^2\psi(x) = \chi_n(c)\psi(x), \quad x \in [-1,1], \tag{15}$$

subject to the regularity condition that $\psi(x)$ remains bounded at the singular endpoints $x = \pm 1$.

Connection to Legendre Polynomials and Perturbative Viewpoint. A fundamental structural insight is the asymptotic behavior of $\mathcal{L}_c$ as the bandwidth parameter c varies:

- **Unperturbed Base Case ($c = 0$):** When $c = 0$, the potential term c^2x^2 vanishes identically, and $\mathcal{L}_0$ degenerates into the classical Legendre differential operator:
$$-\frac{\mathrm{d}}{\mathrm{d}x}\left((1-x^2)\frac{\mathrm{d}p_n(x)}{\mathrm{d}x}\right) = n(n+1)p_n(x). \tag{16}$$
Its eigenfunctions are precisely the normalized Legendre polynomials $p_n(x) = \sqrt{n+1/2}\,P_n(x)$ on $[-1,1]$, and its spectrum is $\chi_n(0) = n(n+1)$.
- **Perturbed Regime ($c > 0$):** For any fixed bandwidth $c > 0$, the operator $\mathcal{L}_c = \mathcal{L}_0 + c^2x^2$ is an analytic perturbation of the Legendre operator $\mathcal{L}_0$ with bounded symmetric perturbation $V = c^2x^2$. Consequently, each PSWF $\psi_n(x;c)$ admits an orthogonal series expansion in the Legendre basis:
$$\psi_n(x;c) = \sum_{k=0}^{\infty} \beta_k^{(n)}(c)p_k(x), \quad \text{with} \quad \sum_{k=0}^{\infty} |\beta_k^{(n)}(c)|^2 = 1. \tag{17}$$

This connection allows us to transfer the sharp analytical identities of Legendre polynomials to PSWFs via Kato's resolvent perturbation theory.

5.2 Norm Embedding from Bandlimited Sinc RKHS into SE RKHS

To circumvent this fundamental difficulty, our overarching strategy is to identify an analytically tractable subspace of the target hypothesis class $\mathcal{H}_{k_{\mathrm{SE}}}([0,1]^d)$. Specifically, we consider the reproducing kernel Hilbert space associated with the bandlimited tensor-product Sinc kernel, and construct our candidate functions directly within $\mathcal{H}_{k_{\mathrm{Sinc},c}}$. We then establish that controlling the norm in the Sinc RKHS rigorously bounds the function's norm in the SE RKHS.

Multidimensional Sinc RKHS and Paley–Wiener Space Fix a base bandwidth parameter $\Omega_0 = c > 0$. We define the d-dimensional tensor-product Sinc kernel on $\mathbb{R}^d$:

$$k_{\mathrm{Sinc},c}(x,y) := \prod_{i=1}^{d} \frac{\sin(c(x_i - y_i))}{\pi(x_i - y_i)}, \quad x, y \in \mathbb{R}^d. \tag{18}$$

For the multidimensional hypercube $[-1,1]^d$, the tensor-product Sinc kernel $k_{\mathrm{Sinc},c}(x,y) = \prod_{j=1}^{d} k_c(x_j, y_j)$ possesses the multi-index orthonormal basis $\Psi_\alpha(x) = \prod_{j=1}^{d} \psi_{\alpha_j}(x_j; c)$ with eigenvalues $\Lambda_\alpha = \prod_{j=1}^{d} \lambda_{\alpha_j}(c)$ for $\alpha \in \mathbb{N}_0^d$.

On the unbounded domain $\mathbb{R}^d$, the native space $\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)$ coincides isometrically with the Paley–Wiener space $\mathcal{PW}_c^d$ of square-integrable functions whose Fourier transforms are supported within the hypercube $[-c,c]^d$:

$$\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d) = \left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; \mathrm{supp}(\hat{f}) \subseteq [-c,c]^d \right\}, \quad \|f\|_{\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)} = \|f\|_{L^2(\mathbb{R}^d)}. \tag{19}$$

Connecting the Norms: The Triad of Norm Relations. To transport functions constructed in the Sinc RKHS into the target SE RKHS $\mathcal{H}_{k_{\mathrm{SE}}}([0,1]^d)$, we establish three structural norm relations:

Lemma 5.1 (Global Frequency-Domain Embedding)**.** *For any bandlimited function $f \in \mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)$, f also belongs to $\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)$, and satisfies:*

$$\|f\|_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)} \le C_d \|f\|_{\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)}, \tag{20}$$

where $C_d = (\pi\theta)^{-d/4} \exp\left(\frac{d\theta c^2}{8}\right)$.

To relate the RKHS norms on the unbounded domain $\mathbb{R}^d$ to those restricted to compact hypercube subsets, we recall the foundational restriction theorem of reproducing kernel Hilbert spaces and its isometric criterion due to Aronszajn (1950).

Lemma 5.2 (Aronszajn's RKHS Restriction Theorem and Isometric Criterion (Aronszajn, 1950, Part I, Section 5))**.** *Let $k : E \times E \to \mathbb{R}$ be a positive definite reproducing kernel on a set E with associated RKHS $\mathcal{H}_k(E)$, and let $E_1 \subseteq E$ be an arbitrary non-empty subset. Then the restriction of k to $E_1 \times E_1$ is the reproducing kernel of the restricted space $\mathcal{H}_k(E_1) := \{f|_{E_1} : f \in \mathcal{H}_k(E)\}$. Furthermore:*

1. ***Variational Norm Bound:*** *For any $f \in \mathcal{H}_k(E)$, its restriction $f|_{E_1} \in \mathcal{H}_k(E_1)$ satisfies*
$$\|f|_{E_1}\|_{\mathcal{H}_k(E_1)} = \min\left\{ \|g\|_{\mathcal{H}_k(E)} \;\middle|\; g \in \mathcal{H}_k(E),\ g|_{E_1} = f|_{E_1} \right\} \le \|f\|_{\mathcal{H}_k(E)}. \tag{21}$$
2. ***Isometric Equality Criterion:*** *Let $\mathcal{F}_0 := \{g \in \mathcal{H}_k(E) : g|_{E_1} \equiv 0\}$ be the closed subspace of functions vanishing identically on E_1. If $\mathcal{F}_0 = \{0\}$ (i.e., any function in $\mathcal{H}_k(E)$ that vanishes on E_1 must vanish everywhere on E), then the restriction operator $R : f \mapsto f|_{E_1}$ is a linear isometric isomorphism from $\mathcal{H}_k(E)$ onto $\mathcal{H}_k(E_1)$, and for all $f \in \mathcal{H}_k(E)$:*
$$\|f|_{E_1}\|_{\mathcal{H}_k(E_1)} = \|f\|_{\mathcal{H}_k(E)}. \tag{22}$$

We now specialize Lemma 5.2 to the target squared exponential kernel and the bandlimited Sinc kernel. First, for the target squared exponential kernel, applying Lemma 5.2 Item 1 with $E = \mathbb{R}^d$ and $E_1 = \mathcal{X} = [0,1]^d$ guarantees that restricting functions to the unit hypercube cannot increase their SE RKHS norm.

Corollary 5.3 (SE Kernel Norm Non-increase on Hypercube)**.** *For any function $f \in \mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)$, its restriction to the hypercube $\mathcal{X} = [0,1]^d \subset \mathbb{R}^d$ satisfies:*

$$\|f|_{\mathcal{X}}\|_{\mathcal{H}_{k_{\mathrm{SE}}}([0,1]^d)} \le \|f\|_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)}. \tag{23}$$

Next, for the bandlimited Sinc kernel, we exploit the rigid analytic structure of the Paley–Wiener space and apply the isometric criterion in Lemma 5.2 Item 2 to establish that the restriction to any domain with non-empty interior is an exact isometry.

Corollary 5.4 (Isometric Norm Equality for Sinc RKHS)**.** *Let $E_1 \subset \mathbb{R}^d$ be any measurable subset with non-empty interior (in particular, $E_1 = [-1,1]^d$). For any bandlimited function $f \in \mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)$, its restriction $f|_{E_1}$ satisfies the exact norm identity:*

$$\|f|_{E_1}\|_{\mathcal{H}_{k_{\mathrm{Sinc},c}}(E_1)} = \|f\|_{\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)}. \tag{24}$$

The Master Norm Inequality. Combining the global frequency-domain embedding (Lemma 5.1), the SE norm restriction (Corollary 5.3), and the Sinc isometric equality on $[-1,1]^d$ (Corollary 5.4), we obtain the fundamental chain of norm relations for any candidate function $f \in \mathcal{H}_{k_{\mathrm{Sinc},c}}([-1,1]^d)$:

$$\|f\|_{\mathcal{H}_{k_{\mathrm{SE}}}([0,1]^d)} \le \|f\|_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)} \le C_d \|f\|_{\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)} = C_d \|f\|_{\mathcal{H}_{k_{\mathrm{Sinc},c}}([-1,1]^d)}. \tag{25}$$

The significance of inequality equation 25 is profound: to guarantee that a candidate function f satisfies the target constraint $\|f\|_{\mathcal{H}_{k_{\mathrm{SE}}}([0,1]^d)} \le B$, it suffices to construct f in the bandlimited space $\mathcal{H}_{k_{\mathrm{Sinc},c}}([-1,1]^d)$ such that

$$\|f\|_{\mathcal{H}_{k_{\mathrm{Sinc},c}}([-1,1]^d)} \le B' := C_d^{-1} B. \tag{26}$$

This completely shifts our analytical focus from the intractable SE kernel to the Sinc kernel on $[-1,1]^d$, which admits an explicit, celebrated Mercer eigensystem via Prolate Spheroidal Wave Functions, as detailed below.

5.3 Construction via Prolate Spheroidal Wave Functions

The principal advantage of the Sinc kernel is that its integral operator on the compact interval $[-1,1]$ under uniform Lebesgue measure possesses an explicit Mercer eigensystem: the *Prolate Spheroidal Wave Functions* (PSWFs) $\psi_n(x)$ with bandwidth parameter $\Omega_0 = c$. They satisfy:

$$\int_{-1}^{1} \frac{\sin(c(x-y))}{\pi(x-y)} \psi_n(y) \mathrm{d}y = \lambda_n(c)\psi_n(x), \quad x \in [-1,1], \tag{27}$$

normalized such that $\int_{-1}^{1} \psi_n^2(x)\mathrm{d}x = 1$.

For a given truncation order N, we define the multidimensional Christoffel-Darboux type kernel:

$$b_{N,d}(x) := \prod_{i=1}^{d} \left(\sum_{n=0}^{N} \psi_n(0)\psi_n(x_i) \right), \tag{28}$$

and define our prototypical centered bump function:

$$f_{\epsilon,N}(x) := 2\epsilon \frac{b_{N,d}(x)}{b_{N,d}(0)}. \tag{29}$$

5.4 Properties of the Hard Function Class

Lemma 5.5 (Properties of $f_{\epsilon,N}$)**.** *Fix $d \in \mathbb{N}_+$, $\epsilon > 0$, and bandwidth $\Omega_0 = c$. For any integer N sufficiently large:*

1. ***Boundedness:*** *$f_{\epsilon,N}(x)$ satisfies $f_{\epsilon,N}(0) = 2\epsilon$, and $\forall x \in [0,1]^d$, $|f_{\epsilon,N}(x)| < 3\epsilon$.*
2. ***ϵ-Optimal Localization:*** *The ϵ-optimal region is strictly contained within an L_∞-cube of width $w_\epsilon = \Theta(N^{-1})$:*

$$\{x \in [0,1]^d \mid \sup_{z \in [-1,1]^d} f_{\epsilon,N}(z) - f_{\epsilon,N}(x) \le \epsilon\} \subseteq \{x \in [0,1]^d \mid \|x\|_\infty \le w_\epsilon/2\}. \tag{30}$$

3. ***Norm Bound in Sinc RKHS:*** *The Sinc RKHS norm satisfies:*

$$\|f_{\epsilon,N}\|_{\mathcal{H}_{k_c}([-1,1]^d)} \le \frac{2\epsilon}{(\lambda_N(c))^{d/2}}. \tag{31}$$

5.5 Truncation Order via Slepian's Asymptotics

The classical lower bound of Scarlett et al. (2017) constructed a perturbed bump function $g_\epsilon(x) = \frac{2\epsilon}{h(0)} h(\frac{\zeta x}{w_\epsilon})$ via the inverse Fourier transform of a smooth cut-off function, requiring an effective width of $w_\epsilon = \Theta(\ln^{-1/2}(B/\epsilon))$. This led directly to the sub-optimal packing number $M_\epsilon = \Theta(w_\epsilon^{-d}) = \Theta((\ln(B/\epsilon))^{d/2})$.

Crucially for our lower bound construction, by the classical spectral asymptotics of Slepian (1965), fix bandwidth c, as $n \to \infty$, the eigenvalues $\lambda_n(c)$ decay super-exponentially

$$\lambda_n(c) \sim \frac{2\pi}{(n!)^2}\left(\frac{c}{4}\right)^{2n+1} = \exp\left(-2n\ln\left(\frac{4n}{ec}\right) + O(\ln n)\right). \tag{32}$$

By Lemma 5.1, ensuring $\|f_{\epsilon,N}\|_{\mathcal{H}_{k_{\mathrm{SE}}}} \le B$ requires:

$$C_d \frac{2\epsilon}{(\lambda_N(c))^{d/2}} \le B \implies \lambda_N(c) \ge \left(\frac{2C_d\epsilon}{B}\right)^{2/d}. \tag{33}$$

Substituting equation 32:

$$N\ln\left(\frac{4N}{ec}\right) \le \frac{1}{d}\ln\left(\frac{B}{2C_d\epsilon}\right). \tag{34}$$

Solving this asymptotic relation gives the maximal admissible order:

$$\bar{N} = \Theta\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)}\right). \tag{35}$$

Consequently by Lemma 5.5 Property 2, for sufficiently small ϵ, the localization width shrinks at the rate:

$$w_\epsilon = \Theta(\bar{N}^{-1}) = \Theta\left(\frac{\ln\ln(B/\epsilon)}{\ln(B/\epsilon)}\right). \tag{36}$$

5.6 Finite Hypothesis Class and Cross-Interference Control

To avoid boundary distortion near $\partial\mathcal{X}$ and ensure that all candidate peaks operate strictly within the uniform interior decay regime established in Lemma 5.5, we restrict our candidate hypothesis functions to the interior hypercube $\mathcal{X}_0 := [1/4, 3/4]^d \subset \mathcal{X}$, while partitioning the entire domain $\mathcal{X} = [0,1]^d$ into uniform grid cells.

Full-Domain Grid and Interior Packing. For convenience, we set $w_\epsilon := 1/K$ for an integer $K \in \mathbb{N}_+$ satisfying $K = \Theta(\bar{N})$, which guarantees $w_\epsilon = \Theta\left(\frac{\ln\ln(B/\epsilon)}{\ln(B/\epsilon)}\right)$. We define the full multi-index set $\mathcal{I}_{\mathrm{all}} := \{1,\dots,K\}^d$ and partition the entire domain $\mathcal{X}$ into disjoint cubic cells of side length w_ϵ. For each multi-index $\mathbf{u} = (u_1,\dots,u_d) \in \mathcal{I}_{\mathrm{all}}$, the cell center $z_{\mathbf{u}}$ and the corresponding cell $\mathcal{R}_{\mathbf{u}}$ are defined by:

$$z_{\mathbf{u}} := \left(\left(u_1 - \frac{1}{2}\right)w_\epsilon,\ \dots,\ \left(u_d - \frac{1}{2}\right)w_\epsilon\right), \tag{37}$$

$$\mathcal{R}_{\mathbf{u}} := \prod_{j=1}^{d} I_{u_j} \text{ with } I_i := \begin{cases} [(i-1)w_\epsilon,\ iw_\epsilon), & 1 \le i < K, \\ [1 - w_\epsilon,\ 1], & i = K. \end{cases} \tag{38}$$

We then identify the interior index subset whose cells lie entirely within $\mathcal{X}_0$:

$$\mathcal{I}_{\mathrm{int}} := \{\mathbf{u} \in \mathcal{I}_{\mathrm{all}} : \mathcal{R}_{\mathbf{u}} \subseteq \mathcal{X}_0\}, \quad \text{and} \quad \mathcal{I}_{\mathrm{ext}} := \mathcal{I}_{\mathrm{all}} \setminus \mathcal{I}_{\mathrm{int}}. \tag{39}$$

The candidate packing set is given by $\mathcal{N}_\epsilon := \{z_{\mathbf{u}}\}_{\mathbf{u}\in\mathcal{I}_{\mathrm{int}}}$, whose cardinality satisfies:

$$M_\epsilon = |\mathcal{N}_\epsilon| = |\mathcal{I}_{\mathrm{int}}| = \Theta\left(w_\epsilon^{-d}\right) = \Theta\left(\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)}\right)^d\right). \tag{40}$$

For each interior center $z_{\mathbf{u}} \in \mathcal{N}_\epsilon$, we define the candidate hypothesis function via spatial shift $f_{z_{\mathbf{u}};\epsilon}(x) := f_{\epsilon,\bar{N}}(x - z_{\mathbf{u}})$. The exterior partition cells $\{\mathcal{R}_{\mathbf{u}}\}_{\mathbf{u}\in\mathcal{I}_{\mathrm{ext}}}$ serve exclusively as a rejection/error region in the decision rule and do not host any candidate hypotheses.

By the global translation isometry of stationary RKHSs (Lemma F.1, Item 2) and Aronszajn's restriction theorem (Corollary 5.3), we have $\|f_{z_{\mathbf{u}};\epsilon}\|_{\mathcal{H}_{k_{\mathrm{SE}}}([0,1]^d)} \le \|f_{z_{\mathbf{u}};\epsilon}\|_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)} = \|f_{\epsilon,\bar{N}}\|_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)} \le B$, ensuring that all shifted hypotheses are strictly admissible.

Cross-Region Interference Control. To apply the change of measure framework across queries landing anywhere in $\mathcal{X}$, we verify that off-peak oscillations from all candidate hypotheses are summable and uniformly bounded on every cell of the partition.

Lemma 5.6 (Cross-Region Interference Bound)**.** *Under the full partition $\{\mathcal{R}_{\mathbf{u}}\}_{\mathbf{u}\in\mathcal{I}_{\mathrm{all}}}$ of $\mathcal{X}$ and the interior hypothesis class $\{f_{z_{\mathbf{v}};\epsilon}\}_{\mathbf{v}\in\mathcal{I}_{\mathrm{int}}}$, there exists a constant $C_d > 0$ depending solely on d such that for any cell index $\mathbf{u} \in \mathcal{I}_{\mathrm{all}}$:*

$$\sum_{\mathbf{v}\in\mathcal{I}_{\mathrm{int}}} \sup_{x\in\mathcal{R}_{\mathbf{u}}} (f_{z_{\mathbf{v}};\epsilon}(x))^2 \le C_d\epsilon^2. \tag{41}$$

Substituting Lemma 5.6 into the change of measure inequality equation 3 across all alternatives $\mathbf{v} \in \mathcal{I}_{\mathrm{int}} \setminus \{\mathbf{u}\}$ completes the proofs of Theorem 4.1 and Theorem 4.2. The detailed proof is deferred to Appendix C.

6 Discussion: Comparison and Impact on Minimax Rates

Why Previous Euclidean Attempts Failed. The fundamental obstruction is that the Gaussian measure places decaying mass toward the boundary of $\mathbb{R}^d$, artificially dilating the required eigenfunction support on bounded sets. In contrast, our approach relies on the uniform-measure Sinc kernel, whose PSWF eigensystem is intrinsically adapted to hypercube $[0,1]^d$.

Definitive Minimax Optimal Rates. As reviewed in Related Works, state-of-the-art non-adaptive algorithms combined with the sharpened MIG bound (Iwazaki, 2026) establish the upper bound $R_T \le \mathcal{O}^*(\sqrt{T(\ln T)^d(\ln\ln T)^{-d}})$ on hypercube $[0,1]^d$. Our Theorem 4.2 proves a matching lower bound of $\Omega(\sqrt{T(\ln T)^d(\ln\ln T)^{-d}})$. Consequently, the longstanding $(\ln T)^{d/4}$ gap on the Euclidean hypercube is closed, and the minimax cumulative regret is now definitively settled:

$$R_T = \widetilde{\Theta}\left(\sqrt{T\frac{(\ln T)^d}{(\ln\ln T)^d}}\right), \tag{42}$$

where $\widetilde{\Theta}$ hides dimension-independent logarithmic factors.

Extension to General Compact Domains with Non-empty Interior. While formally derived on $[0,1]^d$, our lower bounds seamlessly extend to any compact domain $\mathcal{X} \subset \mathbb{R}^d$ having non-empty interior. By constructing a global affine contraction that maps $\mathcal{X}$ into $[1/4, 3/4]^d$ and combining this with the pullback isometry of the SE kernel, all minimax lower bound rates remain identical (see Appendix E for details).

7 Conclusion

We have proved algorithm-independent lower bounds for Gaussian process bandits with the squared exponential kernel on the Euclidean hypercube $\mathcal{X} = [0,1]^d$ tight up to dimension-independent logarithmic factors. By establishing a norm embedding inequality between Sinc and SE RKHSs and utilizing the spectral decay of Prolate Spheroidal Wave Functions, our construction resolves the longstanding $(\ln T)^{d/4}$ gap on the Euclidean hypercube. And our lower bounds extend to compact domain $\mathcal{X} \subset \mathbb{R}^d$ with non-empty interior.

AI use statement

In this work, we have not used generative AI tools for any tasks with required disclosure. Specifically, generative AI was not used to develop theoretical models, formulate mathematical claims, provide critical ingredients for proofs, or assist in the writing of mathematical derivations; all other tasks with required disclosure (such as dataset generation, experimental design, or method implementation) are not applicable to this theoretical work.

Additionally, we used generative AI tools for tasks with recommended disclosure: specifically, to edit the research paper to improve readability and for literature retrieval and discovery (sourcing information and identifying relevant literature).

We have thoroughly reviewed all AI-assisted work. Specifically, all suggested citations and related references were manually cross-checked against primary sources by the authors to guarantee factual validity and eliminate potential hallucinations. Furthermore, all AI-edited text was meticulously verified by the authors to ensure that the mathematical precision and theoretical claims remain strictly faithful to our original proofs. We take full responsibility for the final content of this work, including all text, mathematical claims, proofs, and references.

REFERENCES

N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950. ISSN 00029947, 10886850. URL http://www.jstor.org/stable/1990404.

Ilija Bogunovic, Jonathan Scarlett, Stefanie Jegelka, and Volkan Cevher. Adversarially robust optimization with Gaussian processes. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Adam D. Bull. Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research (JMLR)*, 12:2879–2904, 2011.

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*, pp. 1216–1226, 2021.

Xu Cai and Jonathan Scarlett. Lower bounds for time-varying kernelized bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 73–81, 2025.

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

Nelson Dunford and Jacob T. Schwartz. *Linear Operators, Part I: General Theory*. Interscience Publishers, New York, 1958.

Shogo Iwazaki. Tighter regret lower bound for Gaussian process bandits with squared exponential kernel in hypersphere. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

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

Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer-Verlag, Berlin, Heidelberg, 1995.

H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty — ii. *The Bell System Technical Journal*, 40(1):65–84, 1961.

Bowen Lei, Trevor Q. Kirk, Anirban Bhattacharya, Debdeep Pati, Xiaofeng Qian, Raymundo Arróyave, and Bani K. Mallick. Bayesian optimization with adaptive surrogate models for automated experimental design. *npj Computational Materials*, 7(1):194, 2021.

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)*, pp. 92–107, 2022.

Ruben Martinez-Cantin, Nando de Freitas, Arnaud Doucet, and José A. Castellanos. Active policy learning for robot planning and exploration under uncertainty. In *Robotics: Science and Systems (RSS)*, 2007.

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

Sayak Ray Chowdhury and Aditya Gopalan. Bayesian optimization under heavy-tailed payoffs. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

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*, volume 34, pp. 28836–28847, 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)*, pp. 1723–1742, 2017.

D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty — i. *The Bell System Technical Journal*, 40(1):43–63, 1961.

David Slepian. Some asymptotic expansions for prolate spheroidal wave functions. *Journal of Mathematics and Physics*, 44(1–4):99–113, 1965.

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, volume 25, 2012.

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

Gábor Szegő. *Orthogonal Polynomials*. American Mathematical Society, Providence, RI, 4th edition, 1975.

Sattar Vakili, Nacime Bouziani, Sepehr Jalali, Alberto Bernacchia, and Da-shan Shiu. Optimal order simple regret for Gaussian process bandits. In *Advances in Neural Information Processing Systems*, volume 34, pp. 25577–25588, 2021a.

Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in gaussian process bandits. In Arindam Banerjee and Kenji Fukumizu (eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 82–90. PMLR, 13–15 Apr 2021b.

Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. Finite-time analysis of kernelised contextual bandits. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 654–663, 2013.

Holger Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2004.

Wenjie Xu, Yuning Jiang, Emilio T. Maddalena, and Colin N. Jones. Lower bounds on the noiseless worst-case complexity of efficient global optimization. *Journal of Optimization Theory and Applications*, 201(2):583–608, 2024.

A Technical Proofs for Hard Function Class Construction

A.1 Proof of Lemma 5.1

Proof. By Lemma F.1 (Item 1), for any stationary kernel on $\mathbb{R}^d$, the RKHS norm is given by:

$$\|f\|^2_{\mathcal{H}_k(\mathbb{R}^d)} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{k}(\omega)} \mathrm{d}\omega. \tag{43}$$

For the squared exponential (SE) kernel $k_{\mathrm{SE}}(\tau) = \exp(-\|\tau\|_2^2/\theta)$, its Fourier transform evaluates explicitly to:

$$\hat{k}_{\mathrm{SE}}(\omega) = (\pi\theta)^{d/2} \exp\left(-\frac{\theta\|\omega\|_2^2}{4}\right). \tag{44}$$

Since $f \in \mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)$ is bandlimited with $\mathrm{supp}(\hat{f}) \subseteq [-c,c]^d$, for all ω in the support we have $\|\omega\|_2^2 \leq dc^2$. Therefore, we have the uniform spectral lower bound on $\mathrm{supp}(\hat{f})$:

$$\hat{k}_{\mathrm{SE}}(\omega) \geq (\pi\theta)^{d/2} \exp\left(-\frac{d\theta c^2}{4}\right), \tag{45}$$

Factoring this uniform upper bound out of the spectral integral yields:

$$\|f\|^2_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)} \leq (\pi\theta)^{-d/2} \exp\left(\frac{d\theta c^2}{4}\right) \left[\frac{1}{(2\pi)^d} \int_{[-c,c]^d} |\hat{f}(\omega)|^2 \mathrm{d}\omega\right]. \tag{46}$$

By Plancherel's theorem, $\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 \mathrm{d}\omega = \|f\|^2_{L^2(\mathbb{R}^d)} = \|f\|^2_{\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)}$. Substituting this equality gives:

$$\|f\|^2_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)} \leq (\pi\theta)^{-d/2} \exp\left(\frac{d\theta c^2}{4}\right) \|f\|^2_{\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)}. \tag{47}$$

Taking the square root on both sides proves equation 20 with $C_d = (\pi\theta)^{-d/4} \exp\left(\frac{d\theta c^2}{8}\right)$. □

A.2 Proof of Corollary 5.4

Proof. Any function $g \in \mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d) = \mathcal{PW}_c^d$ extends to a holomorphic function on $\mathbb{C}^d$. Since E_1 has a non-empty interior in $\mathbb{R}^d$, the identity theorem for multivariate holomorphic functions dictates that if $g|_{E_1} \equiv 0$, then g must vanish identically on all of $\mathbb{R}^d$ ($g \equiv 0$). Hence, the subspace of vanishing extensions is trivial: $\mathcal{F}_0 = \{0\}$. Applying Lemma 5.2 (Item 2) immediately yields equation 24. □

A.3 Proof of Properties 1 and 2 of Lemma 5.5

Lemma A.1 (One-Dimensional Legendre Kernel Properties)**.** *Let $p_k(x) = \sqrt{\frac{2k+1}{2}} P_k(x)$ denote the normalized Legendre polynomials on $[-1,1]$, and let*

$$K_N^p(x,y) = \sum_{k=0}^{N} p_k(x) p_k(y)$$

denote the one-dimensional Christoffel–Darboux kernel of Legendre polynomials on $[-1,1]$. For any integer $N \geq 2$:

1. *$K_N^p(x,0)$ attains its unique global maximum on $[-1,1]$ at $x = 0$, and satisfies $|K_N^p(x,0)| \leq K_N^p(0,0)$ for all $x \in [-1,1]$.*
2. *There exists a universal constant $C_0 > 0$ such that the off-peak envelope decays as*

$$\frac{|K_N^p(x,0)|}{K_N^p(0,0)} \leq \frac{C_0}{N|x|} \quad \text{for all } x \in [-3/4, 3/4] \setminus \{0\}.$$

3. *For sufficiently large N, there exists a universal constant $c_1 > 0$ such that for all $|x| \geq c_1 N^{-1}$ with $x \in [-1,1]$,*

$$|K_N^p(x,0)| \leq \frac{1}{4} K_N^p(0,0).$$

Proof. **Setup: The Christoffel–Darboux Identity at $y = 0$.** Since $P_k(-x) = (-1)^k P_k(x)$, the normalized Legendre polynomials satisfy $p_k(0) = 0$ for all odd integers k, which yields

$$K_{2m+1}^p(x,0) = K_{2m}^p(x,0)$$

Hence, it suffices to prove the assertions for $N = 2m$ ($m \geq 1$).

Applying the Christoffel–Darboux formula from Lemma F.2 at $y = 0$, we have

$$K_N^p(x,0) = a_N \frac{p_{N+1}(x)p_N(0) - p_N(x)p_{N+1}(0)}{x}. \tag{48}$$

evaluating equation 48 yields

$$K_{2m}^p(x,0) = \frac{2m+1}{2} P_{2m}(0) \frac{P_{2m+1}(x)}{x}. \tag{49}$$

Taking the limit $x \to 0$ in equation 49 and using the derivative identity $P'_{2m+1}(0) = (2m+1)P_{2m}(0)$, we obtain the central value

$$K_{2m}^p(0,0) = \frac{2m+1}{2} P_{2m}(0) P'_{2m+1}(0) = \frac{(2m+1)^2}{2} P_{2m}(0)^2 > 0. \tag{50}$$

Dividing equation 49 by equation 50, the normalized kernel reduces to

$$\frac{K_{2m}^p(x,0)}{K_{2m}^p(0,0)} = \frac{P_{2m+1}(x)}{x P'_{2m+1}(0)} = \frac{y(x)}{y(0)}, \tag{51}$$

where $y(x) := \frac{P_{2m+1}(x)}{x}$ and $y(0) := \lim_{x\to 0} y(x) = P'_{2m+1}(0)$.

Proof of Item 1,2,3 of Lemma A.1

By Lemma F.4, the function $|y(x)| = |P_{2m+1}(x)/x|$ attains its unique global maximum on $[-1,1]$ at $x = 0$. In conjunction with equation 51, we have

$$\left| \frac{K_N^p(x,0)}{K_N^p(0,0)} \right| = \frac{|y(x)|}{|y(0)|} \le 1, \quad \forall x \in [-1,1],$$

with equality holding if and only if $x = 0$. Multiplying both sides by $K_N^p(0,0)$ gives $|K_N^p(x,0)| \le K_N^p(0,0)$ on $[-1,1]$.

By Bernstein's inequality for Legendre polynomials (Szegő, 1975, Theorem 7.3.3),

$$|P_{2m+1}(x)| \le \sqrt{\frac{2}{\pi(2m+1)}}(1-x^2)^{-1/4} \le \sqrt{\frac{2}{\pi(2m+1)}} \left(\frac{16}{7}\right)^{1/4}, \quad \forall x \in [-3/4, 3/4].$$

Next, the derivative at the origin is given by $P'_{2m+1}(0) = (-1)^m(2m+1)\frac{(2m-1)!!}{(2m)!!}$. By Gautschi's inequality for the ratio of gamma functions,

$$\frac{(2m-1)!!}{(2m)!!} = \frac{\Gamma(m+1/2)}{\sqrt{\pi}\Gamma(m+1)} > \sqrt{\frac{2}{\pi(2m+1)}}.$$

Thus, we obtain the strict lower bound:

$$|P'_{2m+1}(0)| > (2m+1)\sqrt{\frac{2}{\pi(2m+1)}} = \sqrt{\frac{2(2m+1)}{\pi}}.$$

Substituting these bounds into equation 51 yields, for any $x \in [-3/4, 3/4] \setminus \{0\}$,

$$\frac{|K_N^p(x,0)|}{K_N^p(0,0)} = \frac{|P_{2m+1}(x)|}{|x||P'_{2m+1}(0)|} < \frac{\sqrt{\frac{2}{\pi(2m+1)}}\left(\frac{16}{7}\right)^{1/4}}{|x|\sqrt{\frac{2(2m+1)}{\pi}}} = \frac{(16/7)^{1/4}}{(2m+1)|x|}.$$

Since $2m+1 \ge 2m = N$, setting $C_0 := (16/7)^{1/4}$ immediately implies

$$\frac{|K_N^p(x,0)|}{K_N^p(0,0)} \le \frac{C_0}{N|x|},$$

which proves Item 2.

Let $c_1 := 4C_0$. For all $|x| \ge c_1 N^{-1}$ with $x \in [-3/4, 3/4]$, Item 2 gives directly

$$\frac{|K_N^p(x,0)|}{K_N^p(0,0)} \le \frac{C_0}{N|x|} \le \frac{1}{4}.$$

For $x \in [-1,1] \setminus [-3/4, 3/4]$, note that $\sup_{x\in[-1,1]} |P_n(x)| \le 1 \; \forall n \in \mathbb{N}_0$

$$\frac{|K_N^p(x,0)|}{K_N^p(0,0)} = \frac{|P_{2m+1}(x)|}{|x||P'_{2m+1}(0)|} < \frac{4}{3\sqrt{\frac{2(2m+1)}{\pi}}} \le \frac{1}{4} \quad \forall m \ge 44$$

Multiplying by $K_N^p(0,0)$ establishes $|K_N^p(x,0)| \le \frac{1}{4} K_N^p(0,0)$. This completes the proof of Item 3. □

Recall that the normalized Legendre polynomials $p_k(x) = \sqrt{k+1/2}\, P_k(x)$ form an orthonormal basis of $L^2([-1,1])$. Because the prolate spheroidal wave functions (PSWFs) $\psi_n(x)$ are the eigenfunctions of the Sturm–Liouville differential operator

$$\mathcal{L}_c\psi(x) := -\frac{\mathrm{d}}{\mathrm{d}x}\left((1-x^2)\frac{\mathrm{d}\psi(x)}{\mathrm{d}x}\right) + c^2x^2\psi(x) = \chi_n(c)\psi(x), \tag{52}$$

this admit an orthogonal series expansion in the Legendre polynomial basis:

$$\psi_n(x) = \sum_{k=0}^{\infty} \beta_k^{(n)} p_k(x), \quad \text{with} \quad \sum_{k=0}^{\infty} |\beta_k^{(n)}|^2 = 1, \quad \beta_n^{(n)} > 0. \tag{53}$$

Throughout this section, for any $x \in [-1,1]$, we write $b_N(x) \equiv K_N(x,0) := \sum_{n=0}^{N} \psi_n(0)\psi_n(x)$ and $c_N(x) \equiv K_N^p(x,0) := \sum_{n=0}^{N} p_n(0)p_n(x)$

To understand the fine asymptotic behavior of the truncated kernel $b_N(x)$, we must precisely characterize the perturbation of the expansion coefficients $\beta_k^{(n)}$ away from the unperturbed Legendre basis (where $\beta_k^{(n)}(0) = \delta_{n,k}$). This is formally established in the following lemma.

Lemma A.2 (Boundary-Mode Collapse via Subspace Spectral Projection)**.** *Let $m \in \mathbb{N}$ be any positive integer. For both an even truncation index $N = 2m$ and an odd truncation index $N = 2m+1$, the difference between the prolate spheroidal wave function truncated kernel $b_N(x)$ and the Legendre polynomial truncated kernel $c_N(x)$ satisfies the unified asymptotic formula:*

$$b_N(x) - c_N(x) = \epsilon_{2m+2}^{(2m)}\Big[p_{2m}(0)p_{2m+2}(x) + p_{2m+2}(0)p_{2m}(x)\Big] + E_N(x), \tag{54}$$

where the boundary perturbation coefficient satisfies

$$\epsilon_{2m+2}^{(2m)} = -\frac{c^2}{32m} + O\left(\frac{c^2}{m^2}\right) = -\frac{c^2}{16N} + O\left(\frac{c^2}{N^2}\right), \tag{55}$$

and the error term $E_N(x)$ is represented by an absolutely and uniformly convergent double series on $[-1,1]$, satisfying:

$$\|E_N\|_{L^\infty([-3/4,3/4])} = O\left(\frac{c^4}{N}\right) \quad \|E_N\|_{L^\infty([-1,1])} = O\left(\frac{c^4}{\sqrt{N}}\right) \tag{56}$$

Proof. The proof proceeds by reformulating the truncated kernels directly in terms of the matrix representations of total subspace spectral projections on $\ell^2(\mathbb{N}_0)$, and analyzing their difference via complex contour integration.

Operator Representation on $\ell^2(\mathbb{N}_0)$ and Parity Decoupling. Let $\mathcal{H} = \ell^2(\mathbb{N}_0)$ be the complex separable Hilbert space equipped with standard inner product $\langle \vec{x}, \vec{y} \rangle = \sum_{k=0}^{\infty} \overline{x}_k y_k$ and orthonormal basis $\{\vec{e}_k\}_{k=0}^{\infty}$. The unitary map $U : L^2([-1,1]) \to \ell^2(\mathbb{N}_0)$ defined by $f \mapsto (\langle p_k, f\rangle_{L^2})_{k=0}^{\infty}$ maps the basis element $p_k(x)$ to $\vec{e}_k$.

Under this isomorphism, the unperturbed Legendre operator $\mathcal{L}_0 = -\frac{\mathrm{d}}{\mathrm{d}x}((1-x^2)\frac{\mathrm{d}}{\mathrm{d}x})$ is represented by the diagonal operator:

$$D\vec{e}_k = \lambda_k \vec{e}_k, \quad \lambda_k = k(k+1), \quad \mathcal{D}(D) = \left\{ \vec{x} \in \ell^2(\mathbb{N}_0) : \sum_{k=0}^{\infty} k^2(k+1)^2 |x_k|^2 < \infty \right\}. \tag{57}$$

The perturbation c^2x^2 is represented by $V := c^2J^2$, where J is the symmetric tridiagonal Jacobi operator representing coordinate multiplication $x\cdot$ on $L^2([-1,1])$, with entries $(J)_{k,l} = \langle p_k, xp_l\rangle_{L^2}$ and operator norm $\|J\|_{\mathrm{op}} = \|x\|_{L^\infty([-1,1])} = 1$. By the Kato–Rellich theorem, the total operator

$$A(c) := D + c^2J^2 \tag{58}$$

is self-adjoint on $\mathcal{D}(A(c)) = \mathcal{D}(D)$. The expansion vector $\vec{\beta}^{(n)} = (\beta_k^{(n)})_{k=0}^\infty$ of $\psi_n(x)$ is precisely the normalized eigenvector of $A(c)$ corresponding to $\chi_n(c)$:

$$A(c)\vec{\beta}^{(n)} = \chi_n(c)\vec{\beta}^{(n)}, \quad \|\vec{\beta}^{(n)}\|_{\ell^2} = 1, \quad \beta_n^{(n)} > 0. \tag{59}$$

Because x^2 is an even function, $(J^2)_{k,l} = 0$ whenever $k \not\equiv l \pmod 2$. Consequently, $\mathcal{H}$ decomposes into invariant parity subspaces $\mathcal{H} = \mathcal{H}_{\mathrm{even}} \oplus \mathcal{H}_{\mathrm{odd}}$. Furthermore, since odd modes vanish identically at the origin:

$$\psi_{2k+1}(0) = 0, \qquad p_{2k+1}(0) = 0 \quad (\forall k \in \mathbb{N}_0), \tag{60}$$

both truncated kernels contain only even modes up to degree $2m$ for both $N = 2m$ and $N = 2m+1$:

$$b_N(x) = \sum_{k=0}^{m} \psi_{2k}(0)\psi_{2k}(x), \qquad c_N(x) = \sum_{k=0}^{m} p_{2k}(0)p_{2k}(x). \tag{61}$$

In the invariant even subspace $\mathcal{H}_{\mathrm{even}}$, we define the total orthogonal spectral projections onto the first $m+1$ modes:

$$\Pi_N(c) := \sum_{k=0}^{m} \mathcal{P}_{2k}, \qquad \Pi_N(0) := \sum_{k=0}^{m} P_{2k}, \tag{62}$$

where $\mathcal{P}_{2k} = \vec{\beta}_{2k} \otimes \vec{\beta}^*_{2k}$ and $P_{2k} = \vec{e}_{2k} \otimes \vec{e}^*_{2k}$ are the perturbed and unperturbed one-dimensional spectral projections, respectively.

Series Representation of the Truncated Kernels. Substituting the Legendre series expansions $\psi_{2k}(x) = \sum_{j=0}^\infty \beta_{2j}^{(2k)} p_{2j}(x)$ and $\psi_{2k}(0) = \sum_{i=0}^\infty \beta_{2i}^{(2k)} p_{2i}(0)$ into the definition of $b_N(x)$, we have:

$$b_N(x) = \sum_{k=0}^{m} \left(\sum_{i=0}^{\infty} \beta_{2i}^{(2k)} p_{2i}(0) \right) \left(\sum_{j=0}^{\infty} \beta_{2j}^{(2k)} p_{2j}(x) \right). \tag{63}$$

Noting that the matrix elements of the total projection $\Pi_N(c)$ in the basis $\{\vec{e}_{2j}\}$ are given by:

$$\langle \vec{e}_{2j}, \Pi_N(c)\vec{e}_{2i}\rangle = \sum_{k=0}^{m} \beta_{2j}^{(2k)} \beta_{2i}^{(2k)}, \tag{64}$$

we obtain the exact double series representation:

$$b_N(x) = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \langle \vec{e}_{2j}, \Pi_N(c)\vec{e}_{2i}\rangle p_{2j}(x)p_{2i}(0). \tag{65}$$

Similarly, for the unperturbed Legendre kernel, $\langle \vec{e}_{2j}, \Pi_N(0)\vec{e}_{2i}\rangle = \delta_{j,i}\mathbb{I}\{j \le m\}$, which yields:

$$c_N(x) = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \langle \vec{e}_{2j}, \Pi_N(0)\vec{e}_{2i}\rangle p_{2j}(x)p_{2i}(0). \tag{66}$$

Subtracting the two series term-by-term, the kernel difference is rigorously represented by:

$$b_N(x) - c_N(x) = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \langle \vec{e}_{2j}, \left[\Pi_N(c) - \Pi_N(0)\right]\vec{e}_{2i}\rangle p_{2j}(x)p_{2i}(0). \tag{67}$$

Rigorous justification of the summation interchange: We emphasize that representing the pointwise difference via the double series equation 67 is formally valid provided the series converges absolutely and uniformly on $[-1,1]$ for each fixed N. In the subsequent steps, we rigorously validate this interchange.

Riesz Projection Expansion for the Subspace Cluster. In $\mathcal{H}_{\text{even}}$, the unperturbed eigenvalue cluster is $\sigma_N := \{\lambda_0, \lambda_2, \ldots, \lambda_{2m}\}$ with $\lambda_{2k} = 2k(2k+1)$. For any closed linear operator T and complex parameter $z \in \rho(T)$, we adopt the standard Kato resolvent convention:

$$R(z, T) := (T - zI)^{-1}. \tag{68}$$

For the unperturbed self-adjoint diagonal operator D on $\mathcal{H}_{\text{even}}$, its resolvent admits the explicit spectral resolution:

$$R(z, D) = (D - zI)^{-1} = \sum_{k=0}^{\infty} \frac{1}{\lambda_{2k} - z} \vec{e}_{2k} \otimes \vec{e}_{2k}^{*}, \quad \forall z \in \mathbb{C} \setminus \sigma(D). \tag{69}$$

The spectral gap separating the cluster σ_N from the remaining spectrum is:

$$g_N := \lambda_{2m+2} - \lambda_{2m} = 8m + 6 = \Theta(N). \tag{70}$$

We enclose the entire cluster σ_N with a single closed rectangular contour $\Gamma_{[0,N]} = \partial\Omega_N \subset \mathbb{C}$, where:

$$\Omega_N := \{z \in \mathbb{C} : -h_N \le \mathrm{Re}(z) \le x_N^*,\ |\,\mathrm{Im}(z)| \le h_N\}, \tag{71}$$

with $x_N^* := \frac{\lambda_{2m} + \lambda_{2m+2}}{2} = \lambda_{2m} + 4m + 3$ and $h_N := 4m + 3 = \Theta(N)$. Along $\Gamma_{[0,N]}$, the distance to the spectrum satisfies $\inf_{z \in \Gamma_{[0,N]}} \mathrm{dist}(z, \sigma(D)) \ge 4m + 3$. Since D is self-adjoint, the operator norm of the unperturbed resolvent is given by the reciprocal distance to the spectrum:

$$\sup_{z \in \Gamma_{[0,N]}} \|R(z, D)\|_{\text{op}} = \frac{1}{\inf_{z \in \Gamma_{[0,N]}} \mathrm{dist}(z, \sigma(D))} \le \frac{1}{4m+3} = O\left(\frac{1}{N}\right). \tag{72}$$

For $N \ge 2c^2$, $\|V\|_{\text{op}} \sup_{z \in \Gamma_{[0,N]}} \|R(z, D)\|_{\text{op}} \le \frac{c^2}{4m+3} \le \frac{1}{2}$.

Hence, for any homotopy parameter $t \in [-1, 1]$ and all $z \in \Gamma_{[0,N]}$, the perturbation satisfies:

$$\|tVR(z, D)\|_{\text{op}} \le |t| \|V\|_{\text{op}} \|R(z, D)\|_{\text{op}} \le \frac{c^2}{4m+3} < 1. \tag{73}$$

Consequently, the resolvent $R(z, D + tV) = R(z, D)(I + tVR(z, D))^{-1}$ exists and is bounded for all $z \in \Gamma_{[0,N]}$ and all $t \in [0, 1]$, proving that $\Gamma_{[0,N]}$ lies entirely within the resolvent set $\rho(D + tV)$.

By the Dunford–Riesz holomorphic functional calculus (Dunford & Schwartz, 1958, Chapter VII, §3) and spectral theory of self-adjoint operators (Kato, 1995, Chapter III, §6.4–§6.5), the total Riesz projection along the closed contour $\Gamma_{[0,N]}$ evaluates by Cauchy's residue theorem precisely to the sum of the orthogonal eigenprojections of all eigenvalues enclosed by $\Gamma_{[0,N]}$:

$$P(t) := -\frac{1}{2\pi i} \oint_{\Gamma_{[0,N]}} R(z, D + tV)\, \mathrm{d}z = \sum_{\mu \in \sigma(D+tV) \cap \Omega_N} \mathcal{P}_\mu(t). \tag{74}$$

Furthermore, because $\Gamma_{[0,N]} \subset \rho(D + tV)$ for all $t \in [0, 1]$, the projection family $P(t)$ is norm-continuous on $[0, 1]$. By the projection theorem for Banach spaces (Kato, 1995, Chapter I, §4.6 and Chapter III, §3.4), any two projections with $\|P(t) - P(s)\|_{\text{op}} < 1$ have isomorphic ranges via the invertible transformation $U = P(t)P(s) + (I - P(t))(I - P(s))$. Hence, the dimension is a topological invariant along the homotopy:

$$\mathrm{rank}(P(1)) = \mathrm{rank}(P(0)) = \sum_{k=0}^{m} \mathrm{rank}(P_{2k}) = m + 1. \tag{75}$$

Since the initial spectrum inside Ω_N at $t = 0$ is strictly the first $m + 1$ even Legendre eigenvalues $\{\lambda_0, \ldots, \lambda_{2m}\}$, and the boundary distance satisfies $\mathrm{dist}(\Gamma_{[0,N]}, \sigma(D + tV)) \ge (4m + 3) - c^2 > 0$, no eigenvalues enter or exit Ω_N. Thus, the enclosed spectrum at $t = 1$ consists uniquely of the first $m + 1$ perturbed PSWF eigenvalues $\{\chi_0(c), \ldots, \chi_{2m}(c)\}$, and equation 74 guarantees:

$$P(1) = \sum_{k=0}^{m} \mathcal{P}_{2k} = \Pi_N(c). \tag{76}$$

At $t = 1$, the resolvent expands on $\Gamma_{[0,N]}$ into the uniformly convergent Neumann series:

$$R(z, A(c)) = R(z, D) \sum_{p=0}^{\infty} \big(- VR(z, D)\big)^p. \tag{77}$$

Integrating along $\Gamma_{[0,N]}$, the projection difference rigorously decomposes into:

$$\Pi_N(c) - \Pi_N(0) = -\frac{1}{2\pi i} \oint_{\Gamma_{[0,N]}} \big[R(z, A(c)) - R(z, D)\big] \, \mathrm{d}z = \Pi_N^{(1)} + \mathcal{R}_N. \tag{78}$$

where the first-order operator $\Pi_N^{(1)}$ and the remainder operator $\mathcal{R}_N$ are defined by:

$$\Pi_N^{(1)} := \frac{1}{2\pi i} \oint_{\Gamma_{[0,N]}} R(z, D) V R(z, D) \, \mathrm{d}z, \tag{79}$$

$$\mathcal{R}_N := -\frac{1}{2\pi i} \oint_{\Gamma_{[0,N]}} R(z, D) \sum_{p=2}^{\infty} \big(- VR(z, D)\big)^p \, \mathrm{d}z. \tag{80}$$

Explicit Evaluation of the First-Order Term $\Pi_N^{(1)}$ via Residues. Substituting the spectral resolution equation 69 of $R(z, D)$ into equation 79, the matrix elements are given by:

$$\langle \vec{e}_{2j}, \Pi_N^{(1)} \vec{e}_{2i} \rangle = \langle \vec{e}_{2j}, V \vec{e}_{2i} \rangle \cdot I_{j,i}, \quad \text{where} \quad I_{j,i} := \frac{1}{2\pi i} \oint_{\Gamma_{[0,N]}} \frac{\mathrm{d}z}{(z - \lambda_{2j})(z - \lambda_{2i})}. \tag{81}$$

Evaluating $I_{j,i}$ via Cauchy's residue theorem:

- If both indices are inside the cluster ($2j, 2i \le 2m$): both poles lie inside Ω_N. For $j \neq i$, the sum of residues is $\frac{1}{\lambda_{2j} - \lambda_{2i}} + \frac{1}{\lambda_{2i} - \lambda_{2j}} = 0$; for $j = i$, the pole is double and has residue 0. Hence, $I_{j,i} = 0$.
- If both indices are outside the cluster ($2j, 2i > 2m$): both poles lie outside Ω_N, so $I_{j,i} = 0$.
- If one index is inside and the other outside ($2i \le 2m < 2j$): $\Gamma_{[0,N]}$ encloses only λ_{2i}, yielding $I_{j,i} = \frac{1}{\lambda_{2i} - \lambda_{2j}}$. Similarly, if $2j \le 2m < 2i$, $I_{j,i} = \frac{1}{\lambda_{2j} - \lambda_{2i}}$.

Because $V = c^2 J^2$ is pentadiagonal, $\langle \vec{e}_{2j}, V \vec{e}_{2i} \rangle \neq 0$ only if $|2j - 2i| \in \{0, 2\}$. Under the boundary crossing condition $2i \le 2m < 2j$, the only surviving index pair is $(2i, 2j) = (2m, 2m + 2)$. By symmetry, the only surviving pair for $2j \le 2m < 2i$ is $(2j, 2i) = (2m, 2m + 2)$.

Therefore, all other matrix elements vanish identically, and the double sum for $\Pi_N^{(1)}$ collapses to a finite sum of exactly two terms:

$$\begin{aligned}
&\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \langle \vec{e}_{2j}, \Pi_N^{(1)} \vec{e}_{2i} \rangle p_{2j}(x) p_{2i}(0) \\
&\quad = \langle \vec{e}_{2m+2}, \Pi_N^{(1)} \vec{e}_{2m} \rangle p_{2m+2}(x) p_{2m}(0) + \langle \vec{e}_{2m}, \Pi_N^{(1)} \vec{e}_{2m+2} \rangle p_{2m}(x) p_{2m+2}(0) \\
&\quad = \epsilon_{2m+2}^{(2m)} \Big[p_{2m}(0) p_{2m+2}(x) + p_{2m+2}(0) p_{2m}(x) \Big],
\end{aligned} \tag{82}$$

where the boundary coefficient is given by:

$$\epsilon_{2m+2}^{(2m)} := \frac{c^2 (J^2)_{2m+2,2m}}{\lambda_{2m} - \lambda_{2m+2}}. \tag{83}$$

Using the unperturbed eigenvalue difference $\lambda_{2m} - \lambda_{2m+2} = -(8m + 6)$ and the Jacobi recurrence coefficient $(J^2)_{2m+2,2m} = 1/4 + O(m^{-2})$, we have:

$$\epsilon_{2m+2}^{(2m)} = \frac{c^2 (1/4 + O(m^{-2}))}{-(8m + 6)} = -\frac{c^2}{32m} + O\left(\frac{c^2}{m^2}\right) = -\frac{c^2}{16N} + O\left(\frac{c^2}{N^2}\right), \tag{84}$$

which establishes equation 55.

Uniform Convergence of the Remainder Kernel $E_N(x)$. Substituting equation 78 and equation 82 into equation 67, the remainder kernel is explicitly defined by the double series:

$$E_N(x) := \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\langle \vec{e}_{2j}, \mathcal{R}_N \vec{e}_{2i}\rangle p_{2j}(x)p_{2i}(0). \tag{85}$$

We now prove that this series converges absolutely and uniformly on $[-1,1]$, and bound its supremum norm.

First, by Bernstein's inequality for Legendre polynomials (Szegő, 1975, Theorem 7.3.3), for any $x = 0$ and any $k \in \mathbb{N}_0$:

$$|p_k(0)| \le \sqrt{\frac{2}{\pi}} =: M_0. \tag{86}$$

And $|p_{2j}(x)| \le \sqrt{\frac{4j+1}{2}}$ hold uniformly for all $j \in \mathbb{N}_0$ and $x \in [-1,1]$. Therefore,

$$|E_N(x)| \le M_0 \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}|\langle \vec{e}_{2j}, \mathcal{R}_N \vec{e}_{2i}\rangle| \sqrt{\frac{4j+1}{2}} \tag{87}$$

Consider

$$\mathcal{R}_N = \sum_{p=2}^{\infty} T_p, \quad T_p := -\frac{1}{2\pi i}\oint_{\Gamma_{[0,N]}} R(z,D)\big(-VR(z,D)\big)^p \,\mathrm{d}z$$

Banded Structure and Finite Propagation. The operator $R(z,D)$ is diagonal, while $V = c^2J^2$ is strictly pentadiagonal, satisfying $V_{2j,2k} \neq 0$ if and only if $|j-k| \in \{0,1\}$. Consequently, each multiplication by $(-VR(z,D))$ shifts the basis indices by at most ± 2. By induction, for any $p \ge 1$,

$$\operatorname{supp}\Big(\big(-VR(z,D)\big)^p \vec{e}_{2i}\Big) \subseteq \{2j \in \mathbb{N}_0 : |j-i| \le p\}. \tag{88}$$

Integrating term-by-term along the cluster contour $\Gamma_{[0,N]}$, this immediately implies the support bound.

$$\langle \vec{e}_{2j}, T_p \vec{e}_{2i}\rangle = 0 \quad \text{whenever } p < |i-j|. \tag{89}$$

Furthermore we could show that

Lemma A.3. *For any integer $p \ge 2$ and any indices $i, j, m \in \mathbb{N}_0$,*

$$\langle \vec{e}_{2j}, T_p \vec{e}_{2i}\rangle = 0 \quad \textit{whenever } p < |i-m| + |j-m|. \tag{90}$$

Proof. To evaluate the matrix elements $\langle \vec{e}_{2j}, T_p \vec{e}_{2i}\rangle$, we expand T_p into discrete index paths:

$$\langle \vec{e}_{2j}, T_p \vec{e}_{2i}\rangle = \sum_{k_1=0}^{\infty}\cdots\sum_{k_{p-1}=0}^{\infty} V_{2j,2k_{p-1}}\cdots V_{2k_1,2i}\cdot I_p(j,k_{p-1},\ldots,k_1,i), \tag{91}$$

where the scalar contour integral associated with each discrete path $\gamma = (i = k_0 \to k_1 \to \cdots \to k_p = j)$ is given by

$$I_p(\gamma) := \frac{1}{2\pi i}\oint_{\Gamma_{[0,N]}} \frac{\mathrm{d}z}{\prod_{r=0}^{p}(z-\lambda_{2k_r})}. \tag{92}$$

Recall that the contour $\Gamma_{[0,N]}$ encloses the interior spectral cluster $\sigma_N = \{\lambda_{2k} : 0 \le k \le m\}$, while all exterior eigenvalues $\{\lambda_{2k} : k \ge m+1\}$ lie strictly outside $\Gamma_{[0,N]}$. We first analyze the scalar path integral $I_p(\gamma)$ in equation 92 via Cauchy's residue calculus:

1. *Case A: Strictly Interior Paths ($k_r \le m$ for all $0 \le r \le p$).* All $p+1$ poles $\{\lambda_{2k_r}\}_{r=0}^{p}$ lie entirely within the bounded interior domain Ω_N. Since $p \ge 2$, the rational integrand decays at infinity as $O(|z|^{-(p+1)}) \le O(|z|^{-3})$. By Cauchy's residue theorem, the sum of all interior residues equals the negative of the residue at infinity:

$$I_p(\gamma) = -\operatorname{Res}_{z=\infty}\left(\frac{1}{\prod_{r=0}^{p}(z-\lambda_{2k_r})}\right) = 0. \tag{93}$$

2. *Case B: Strictly Exterior Paths ($k_r \ge m+1$ for all $0 \le r \le p$).* All $p+1$ poles lie strictly in the exterior of $\Gamma_{[0,N]}$. The integrand is thus holomorphic throughout Ω_N, yielding $I_p(\gamma) = 0$ by Cauchy's integral theorem.

Consequently, a path γ yields a non-vanishing contribution $I_p(\gamma) \neq 0$ if and only if it traverses the spectral interface $\{m, m+1\}$, meaning that the index set $\{k_0, \dots, k_p\}$ must contain at least one interior index $k_{\text{in}} \le m$ and at least one exterior index $k_{\text{out}} \ge m+1$.

Next, we exploit the banded structure of the Jacobi operator. In $\mathcal{H}_{\text{even}}$, the perturbation $V = c^2 J^2$ satisfies $V_{2u,2v} \neq 0$ if and only if $|u-v| \le 1$. Hence, every transition in γ has step size at most 1:

$$|k_r - k_{r-1}| \le 1 \quad \text{for all } r \in \{1, \dots, p\}. \tag{94}$$

We now determine the minimum number of steps required to connect i and j while traversing the interface $\{m, m+1\}$. In all cases, any path with $I_p(\gamma) \neq 0$ requires at least $p \ge |i-m| + |j-m|$ transitions. Therefore, if $p < |i-m| + |j-m|$, every term in the expansion equation 91 vanishes identically, establishing equation 90. □

Summability of the Remainder Kernel. By Lemma A.3, for any pair $(i,j) \in \mathbb{N}_0 \times \mathbb{N}_0$, the sum in $\mathcal{R}_N$ truncates from below at $p_{\min}(i,j) := \max\{2, |i-m| + |j-m|\}$. Using the uniform resolvent bound $\|R(z,D)\|_{\text{op}} \le \frac{1}{4m+3} = O(1/N)$ and the perimeter $\text{Length}(\Gamma_{[0,N]}) = O(N^2)$, each operator slice satisfies:

$$\|T_p\|_{\text{op}} \le \frac{1}{2\pi}\,\text{Length}(\Gamma_{[0,N]})\|R(z,D)\|_{\text{op}}^{p+1}\|V\|_{\text{op}}^{p} \le C_0 \frac{c^4}{N}\rho_N^{p-2}, \tag{95}$$

where $\rho_N \le 1/2$. Summing over $p \ge p_{\min}(i,j)$, the matrix elements decay exponentially with distance to m:

$$|\langle \vec{e}_{2j}, \mathcal{R}_N \vec{e}_{2i}\rangle| \le \sum_{p=p_{\min}(i,j)}^{\infty} \|T_p\|_{\text{op}} \le 2C_0 \frac{c^4}{N}\rho_N^{\max\{0,\,|i-m|+|j-m|-2\}}. \tag{96}$$

Summing the geometric tail over $(i,j) \in \mathbb{N}_0 \times \mathbb{N}_0$ yields that on $[-1,1]$:

$$\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\langle \vec{e}_{2j}, \mathcal{R}_N \vec{e}_{2i}\rangle p_{2j}(x)p_{2i}(0) \le \frac{C_1 c^4}{N}\sum_{i=0}^{\infty}\left(\frac{1}{2}\right)^{|i-m|}\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^{|j-m|}\sqrt{\frac{4j+1}{2}} = O\left(\frac{c^4}{\sqrt{N}}\right). \tag{97}$$

And on $[-3/4, 3/4]$

$$\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\langle \vec{e}_{2j}, \mathcal{R}_N \vec{e}_{2i}\rangle p_{2j}(x)p_{2i}(0) \le \frac{C_1' c^4}{N}\sum_{i=0}^{\infty}\left(\frac{1}{2}\right)^{|i-m|}\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^{|j-m|} = O\left(\frac{c^4}{N}\right). \tag{98}$$

This establishes equation 56.

Finally, because the first-order projection $\Pi_N^{(1)}$ consists of strictly two non-zero matrix elements (equation equation 82) and the remainder $\mathcal{R}_N$ generates an absolutely and uniformly convergent double series, the rearrangement and term-by-term summation in equation 67 are rigorously justified by Fubini's theorem. Consequently, the pointwise kernel asymptotic formula equation 54 holds identically for all $x \in [-1,1]$ and completes the proof. □

Corollary A.4 (Uniform Perturbation and Relative Error Bound)**.** *Under the assumptions of Lemma A.2, for both an even truncation index $N = 2m$ and an odd truncation index $N = 2m + 1$ ($m \geq 1$), the kernel difference satisfies the uniform bound:*

$$\|b_N - c_N\|_{L^\infty([-3/4,3/4])} = O(\frac{c^2}{N}) \quad \|b_N - c_N\|_{L^\infty([-1,1])} = O(\frac{c^2}{\sqrt{N}}) \quad \textit{as } N \to \infty. \tag{99}$$

Moreover, comparing with the unperturbed peak height $c_N(0) = \Theta(N)$, the relative error satisfies

$$\frac{\|b_N - c_N\|_{L^\infty([-3/4,3/4])}}{c_N(0)} = O\left(\frac{c^2}{N^2}\right) \quad \textit{as } N \to \infty. \tag{100}$$

Proof. By Lemma A.2, for both $N = 2m$ and $N = 2m + 1$, the difference admits the expansion:

$$b_N(x) - c_N(x) = \epsilon^{(2m)}_{2m+2}\Big[p_{2m}(0)p_{2m+2}(x) + p_{2m+2}(0)p_{2m}(x)\Big] + E_N(x). \tag{101}$$

By the Bernstein inequality (Szegő, 1975, Theorem 7.3.3), the normalized Legendre polynomials are uniformly bounded on $[-3/4, 3/4]$ by a constant M_1:

$$\sup_{k\geq 0}\sup_{x\in[-3/4,3/4]} |p_k(x)| \leq \sqrt{\frac{2}{\pi}}\left(1 - \left(\frac{3}{4}\right)^2\right)^{-1/4} := M_1. \tag{102}$$

In particular, $|p_{2m}(0)| \leq M_1$ and $|p_{2m+2}(0)| \leq M_1$. Moreover, for all $j \in \mathbb{N}_0$

$$\sup_{x\in[-1,1]} |p_k(x)| = \sqrt{\frac{2k+1}{2}}$$

holds uniformly. Since the boundary coefficient satisfies $\epsilon^{(2m)}_{2m+2} = -\frac{c^2}{16N} + O(N^{-2}) = O(1/N)$, the boundary mode in equation 101 is uniformly bounded by:

$$\sup_{x\in[-3/4,3/4]} \left|\epsilon^{(2m)}_{2m+2}\Big[p_{2m}(0)p_{2m+2}(x) + p_{2m+2}(0)p_{2m}(x)\Big]\right| = O\left(\frac{c^2}{N}\right). \tag{103}$$

$$\sup_{x\in[-1,1]} \left|\epsilon^{(2m)}_{2m+2}\Big[p_{2m}(0)p_{2m+2}(x) + p_{2m+2}(0)p_{2m}(x)\Big]\right| = O\left(\frac{c^2}{\sqrt{N}}\right). \tag{104}$$

Combining this with the remainder estimate from Lemma A.2, the triangle inequality yields

$$\|b_N - c_N\|_{L^\infty([-3/4,3/4])} \leq O\left(\frac{c^2}{N}\right) + \|E_N\|_{L^\infty([-3/4,3/4])} = O(\frac{c^2}{N}) \quad \text{as } N \to \infty, \tag{105}$$

$$\|b_N - c_N\|_{L^\infty([-1,1])} \leq O\left(\frac{c^2}{\sqrt{N}}\right) + \|E_N\|_{L^\infty([-1,1])} = O(\frac{c^2}{\sqrt{N}}) \quad \text{as } N \to \infty, \tag{106}$$

which establishes equation 99.

Finally, since Wallis' formula yields

$$c_N(0) = \sum_{k=0}^{m}[p_{2k}(0)]^2 = \Theta(N), \tag{107}$$

Dividing the uniform absolute bound by $c_N(0)$ gives the relative error estimate:

$$\frac{\|b_N - c_N\|_{L^\infty([-3/4,3/4])}}{c_N(0)} = O\left(\frac{c^2}{N^2}\right) \quad \text{as } N \to \infty, \tag{108}$$

which completes the proof. □

Lemma A.5 (One-Dimensional PSWF Kernel Properties)**.** *Recall that $K_N(x, y) = \sum_{n=0}^{N}\psi_n(x)\psi_n(y)$ denote the one-dimensional Christoffel-Darboux kernel of PSWFs on $[-1, 1]$ with fixed bandwidth $\Omega_0 = c$. For integer N sufficiently large:*

1. *The central value $K_N(0,0)$ uniformly dominates across $[-1,1]$ up to a higher-order perturbation:*

$$\frac{\sup_{x\in[-1,1]}|K_N(x,0)|}{K_N(0,0)} = \left(1+O\left(\frac{1}{N^2}\right)\right) \quad \text{as } N\to\infty. \tag{109}$$

2. *There exists a universal constant $C_0' > 0$ such that the off-peak envelope decays as $\frac{|K_N(x,0)|}{K_N(0,0)} \le \frac{C_0'}{N|x|}$ for $x \in [-3/4, 3/4]\setminus\{0\}$.*

3. *In particular, for sufficiently large N, there exists a universal constant $c_1' > 0$ such that for all $|x| \ge c_1' N^{-1}$ with $x \in [-1,1]$, $|K_N(x,0)| \le \frac{1}{3}K_N(0,0)$.*

Proof of Lemma A.5. From Corollary A.4 and Lemma A.1 Item 2, we establish Item 2 that on $[-3/4, 3/4]\setminus\{0\}$

$$\frac{|K_N(x,0)|}{K_N(0,0)} = \frac{|K_N^p(x,0)+O(\frac{1}{N})|}{K_N^p(0,0)+O(\frac{1}{N})} \le \frac{|O(\frac{C_0}{|x|})+O(\frac{1}{N})|}{\Theta(N)+O(\frac{1}{N})} \le \frac{C_0'}{N|x|}$$

Thus for sufficiently large N, there exists a universal constant $c_1' > 0$ such that for all $|x| \ge c_1' N^{-1}$ with $x \in [-3/4, 3/4]$, $|K_N(x,0)| \le \frac{1}{3}K_N(0,0)$, next we focus on $x \notin [-3/4, 3/4]$ to establish Item 3. From Lemma A.1 Item 3, we note that for sufficiently large N.

$$\frac{|K_N(x,0)|}{K_N(0,0)} = \frac{|K_N^p(x,0)+O(\frac{1}{\sqrt{N}})|}{K_N^p(0,0)+O(\frac{1}{\sqrt{N}})} \le \frac{\frac{1}{4}K_N^p(0,0)+O(\frac{1}{\sqrt{N}})}{K_N^p(0,0)+O(\frac{1}{\sqrt{N}})} \le \frac{1}{3}$$

Next note that from Corollary A.4

$$\frac{\sup_{x\in[-3/4,3/4]}|K_N(x,0)|}{K_N(0,0)} = \left(1+O\left(\frac{1}{N^2}\right)\right) \quad \text{as } N\to\infty. \tag{110}$$

From Item 3, we note that as $N\to\infty$, $\sup_{|x|>3/4}|K_N(x,0)| \le K_N(0,0)$ Thus we can establish Item 1. □

Proof of Properties 1 and 2 of Lemma 5.5. By definition equation 28, $b_{N,d}(x) = \prod_{j=1}^d K_N(x_j,0)$, and $b_{N,d}(0) = (K_N(0,0))^d$. The multidimensional function factors into:

$$f_{\epsilon,N}(x) = 2\epsilon \prod_{j=1}^d \frac{K_N(x_j,0)}{K_N(0,0)}. \tag{111}$$

1. **Boundedness:** By Lemma A.5 Item 1, each 1D factor satisfies $\frac{\sup_{x_j\in[-1,1]}|K_N(x_j,0)|}{K_N(0,0)} = 1+O\left(\frac{1}{N^2}\right)$. Taking the product over $j=1,\dots,d$, $f_{\epsilon,N}(0) = 2\epsilon$ and for sufficiently large N, $|f_{\epsilon,N}(x)| \le 3\epsilon$ for all $x\in[-1,1]^d$.

2. **ϵ-Optimal Localization:** By Lemma A.5 Item 3, if at least one coordinate $j^*\in\{1,\dots,d\}$ such that $|x_{j^*}| \ge c_1' N^{-1}$, this coordinate satisfies $|\frac{K_N(x_{j^*},0)}{K_N(0,0)}| \le \frac{1}{3}$. Since all other coordinates satisfy $|\frac{K_N(x_j,0)}{K_N(0,0)}| \le 1+O\left(\frac{1}{N}\right)$, for sufficiently large N, we have:

$$f_{\epsilon,N}(x) \le 2\epsilon\cdot\frac{1}{3}\cdot(1+O(\frac{1}{N}))^{d-1} < \epsilon. \tag{112}$$

Consequently, $f_{\epsilon,N}(0) - f_{\epsilon,N}(x) > 2\epsilon - \epsilon = \epsilon$. This confirms that the ϵ-optimal region $\{x \mid \sup_{z\in[-1,1]^d} f_{\epsilon,N}(z) - f_{\epsilon,N}(x) \le \epsilon\}$ is strictly confined within the cube $\{x \mid \|x\|_\infty \le w_\epsilon/2\}$.

□

A.4 Proof of Property 3 of Lemma 5.5

Proof of Property 3 (Norm Bounds and Admissible Truncation Order). By Mercer's representation theorem on the compact domain $[-1,1]^d$, the Sinc integral operator has orthonormal eigenfunctions $\Psi_\alpha(x) = \prod_{j=1}^d \psi_{\alpha_j}(x_j)$ and corresponding eigenvalues $\Lambda_\alpha = \prod_{j=1}^d \lambda_{\alpha_j}(c)$, where $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$.

The function $b_{N,d}(x) = \prod_{j=1}^d (\sum_{n=0}^N \psi_n(0)\psi_n(x_j))$ can be expanded into the Mercer basis as:

$$b_{N,d}(x) = \sum_{\alpha \in \{0,\ldots,N\}^d} \Psi_\alpha(0)\Psi_\alpha(x) = \sum_{\alpha \in \{0,\ldots,N\}^d} \left(\frac{\Psi_\alpha(0)}{\sqrt{\Lambda_\alpha}}\right) \sqrt{\Lambda_\alpha}\Psi_\alpha(x). \tag{113}$$

Therefore, scaling by $\frac{2\epsilon}{b_{N,d}(0)}$, the squared RKHS norm of $f_{\epsilon,N}$ in the Sinc RKHS $\mathcal{H}_{k_c}([-1,1]^d)$ is:

$$\begin{aligned} \|f_{\epsilon,N}\|^2_{\mathcal{H}_{k_c}([-1,1]^d)} &= \frac{4\epsilon^2}{b_{N,d}(0)^2} \sum_{\alpha_1=0}^N \cdots \sum_{\alpha_d=0}^N \frac{\Psi_\alpha(0)^2}{\Lambda_\alpha} \\ &= \frac{4\epsilon^2}{b_{N,d}(0)^2} \sum_{\alpha_1=0}^N \cdots \sum_{\alpha_d=0}^N \frac{\prod_{j=1}^d \psi_{\alpha_j}(0)^2}{\prod_{j=1}^d \lambda_{\alpha_j}(c)}. \end{aligned} \tag{114}$$

Since the 1D eigenvalues $\lambda_n(c)$ are strictly positive and monotonically decreasing in n ($\lambda_n(c) \geq \lambda_N(c)$ for all $n \leq N$), we have $\prod_{j=1}^d \lambda_{\alpha_j}(c) \geq (\lambda_N(c))^d$. Pulling this worst-case lower bound out of the summation:

$$\begin{aligned} \|f_{\epsilon,N}\|^2_{\mathcal{H}_{k_c}([-1,1]^d)} &\leq \frac{4\epsilon^2}{b_{N,d}(0)^2(\lambda_N(c))^d} \sum_{\alpha_1=0}^N \cdots \sum_{\alpha_d=0}^N \prod_{j=1}^d \psi_{\alpha_j}(0)^2 \\ &= \frac{4\epsilon^2}{b_{N,d}(0)^2(\lambda_N(c))^d} \left(\prod_{j=1}^d \sum_{n=0}^N \psi_n(0)^2\right). \end{aligned} \tag{115}$$

Notice that from the definition of $b_{N,d}(0)$, we have:

$$b_{N,d}(0) = \prod_{j=1}^d \left(\sum_{n=0}^N \psi_n(0)^2\right). \tag{116}$$

Substituting this identity directly simplifies the bound:

$$\|f_{\epsilon,N}\|^2_{\mathcal{H}_{k_c}([-1,1]^d)} \leq \frac{4\epsilon^2}{b_{N,d}(0)^2(\lambda_N(c))^d} \cdot b_{N,d}(0) = \frac{4\epsilon^2}{(\lambda_N(c))^d b_{N,d}(0)}. \tag{117}$$

Since $b_{N,d}(0) = (K_N(0,0))^d = \Theta(N^d) \geq 1$, we conclude that:

$$\|f_{\epsilon,N}\|_{\mathcal{H}_{k_c}([-1,1]^d)} \leq \frac{2\epsilon}{(\lambda_N(c))^{d/2}}. \tag{118}$$

Next, by Corollary 5.4, we have $\|f_{\epsilon,N}\|_{\mathcal{H}_{k_c}([-1,1]^d)} = \|f_{\epsilon,N}\|_{\mathcal{H}_{k_c}(\mathbb{R}^d)}$. Then, invoking the RKHS Norm Embedding Inequality (Lemma 5.1), the norm $\|f_{\epsilon,N}\|_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)}$ satisfies:

$$\|f_{\epsilon,N}\|_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)} \leq C_d \|f_{\epsilon,N}\|_{\mathcal{H}_{k_{\mathrm{Sinc},c}}(\mathbb{R}^d)} \leq C_d \frac{2\epsilon}{(\lambda_N(c))^{d/2}}, \tag{119}$$

To guarantee that $f_{\epsilon,N} \in \mathcal{H}_{k_{\mathrm{SE}}}([0,1]^d, B)$, it suffices to ensure that $\|f_{\epsilon,N}\|_{\mathcal{H}_{k_{\mathrm{SE}}}(\mathbb{R}^d)} \leq B$, which requires:

$$\lambda_N(c) \geq \left(\frac{2C_d\epsilon}{B}\right)^{2/d}. \tag{120}$$

According to Slepian's asymptotic formula equation 32, the eigenvalues satisfy $\lambda_N(c) \sim \exp(-2N\ln(\frac{4N}{ec}))$. Substituting this into equation 120 and taking the logarithm on both sides:

$$-2N\ln\left(\frac{4N}{ec}\right) \geq \frac{2}{d}\ln\left(\frac{2C_d\epsilon}{B}\right) = -\frac{2}{d}\ln\left(\frac{B}{2C_d\epsilon}\right), \tag{121}$$

which is equivalent to:

$$N\ln\left(\frac{4N}{ec}\right) \leq \frac{1}{d}\ln\left(\frac{B}{2C_d\epsilon}\right). \tag{122}$$

Let $Y = \frac{1}{d}\ln\left(\frac{B}{2C_d\epsilon}\right)$. The transcendental equation $N\ln N \leq Y$ has the asymptotic solution $N \leq \frac{Y}{\ln Y}(1+o(1))$ as $Y \to \infty$. Thus, the maximal admissible truncation order is:

$$\bar{N} = \Theta\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)}\right). \tag{123}$$

Under this choice of $\bar{N}$, $\|f_{\epsilon,\bar{N}}\|_{\mathcal{H}_{k_{\mathrm{SE}}}} \leq B$ is guaranteed, completing the proof. □

B Proof of Cross-Region Interference Bound (Lemma 5.6)

Proof. Fix an arbitrary reference cell index $\mathbf{u} \in \mathcal{I}_{\mathrm{all}}$. For any candidate alternative $\mathbf{v} \in \mathcal{I}_{\mathrm{int}}$, the hypothesis function is given by $f_{z_\mathbf{v};\epsilon}(x) = f_{\epsilon,\bar{N}}(x - z_\mathbf{v})$. By the separable tensor-product structure in equation 28 and equation 29, for any $x \in \mathcal{R}_\mathbf{u}$, we have:

$$|f_{z_\mathbf{v};\epsilon}(x)| = 2\epsilon\prod_{j=1}^{d}\left|\frac{K_{\bar{N}}\left(x_j - \left(v_j - \frac{1}{2}\right)w_\epsilon,\, 0\right)}{K_{\bar{N}}(0,0)}\right|. \tag{124}$$

Because the candidate centers are strictly restricted to the interior $\mathcal{X}_0 = [1/4, 3/4]^d$ while $x \in \mathcal{X} = [0,1]^d$, each spatial difference satisfies:

$$x_j - z_{\mathbf{v},j} \in \left[-\frac{3}{4}, \frac{3}{4}\right], \tag{125}$$

which guarantees that the PSWF kernel is evaluated strictly within the uniform interior decay regime (Lemma A.5) everywhere on $\mathcal{X}$.

For each coordinate $j \in \{1,\ldots,d\}$, denote the grid index distance by $k_j := |u_j - v_j|$. We distinguish two cases:

1. If $u_j = v_j$ ($k_j = 0$), x_j and $z_{\mathbf{v},j}$ lie in the same 1D slice of length w_ϵ. By Lemma A.5 (Item 1), for sufficiently large $\bar{N}$, the 1D ratio satisfies:

$$\sup_{x_j \in \mathcal{R}_{\mathbf{u},j}}\left|\frac{K_{\bar{N}}\left(x_j - z_{\mathbf{v},j},\, 0\right)}{K_{\bar{N}}(0,0)}\right| \leq 1 + O\left(\frac{1}{\bar{N}}\right) \leq 2. \tag{126}$$

2. If $u_j \neq v_j$ ($k_j \geq 1$), the coordinate-wise distance between any $x_j \in [(u_j-1)w_\epsilon,\, u_j w_\epsilon]$ and the cell center $z_{\mathbf{v},j} = (v_j - 1/2)w_\epsilon$ satisfies:

$$|x_j - z_{\mathbf{v},j}| \geq \left(k_j - \frac{1}{2}\right)w_\epsilon = \frac{2k_j - 1}{2}w_\epsilon. \tag{127}$$

By the off-peak envelope decay (Lemma A.5, Item 2), there exists a constant $C_0' > 0$ such that:

$$\left|\frac{K_{\bar{N}}(x_j - z_{\mathbf{v},j},\, 0)}{K_{\bar{N}}(0,0)}\right| \leq \frac{C_0'}{\bar{N}|x_j - z_{\mathbf{v},j}|} \leq \frac{C_0'}{\bar{N}\left(\frac{2k_j-1}{2}w_\epsilon\right)} \leq \frac{C_0}{2k_j - 1}, \tag{128}$$

where $C_0 := \frac{2C_0'}{\bar{N}w_\epsilon} = O(1)$, since $\bar{N}w_\epsilon = \Theta(1)$.

Squaring the coordinate ratio, we define the 1D envelope upper-bound function:

$$g(k) := \begin{cases} 4, & k = 0, \\ \dfrac{C_0^2}{(2k-1)^2}, & k \geq 1. \end{cases} \tag{129}$$

Then for any $x \in \mathcal{R}_{\mathbf{u}}$ and $\mathbf{v} \in \mathcal{I}_{\text{int}}$, we obtain $\sup_{x \in \mathcal{R}_{\mathbf{u}}} (f_{z_{\mathbf{v}};\epsilon}(x))^2 \leq 4\epsilon^2 \prod_{j=1}^{d} g(|u_j - v_j|)$.

Summing over all $\mathbf{v} \in \mathcal{I}_{\text{int}}$ and enlarging the index domain to the full grid $\mathcal{I}_{\text{all}}$, the sum cleanly factorizes into d independent 1D series:

$$\begin{aligned} \sum_{\mathbf{v} \in \mathcal{I}_{\text{int}}} \sup_{x \in \mathcal{R}_{\mathbf{u}}} (f_{z_{\mathbf{v}};\epsilon}(x))^2 &\leq 4\epsilon^2 \sum_{\mathbf{v} \in \mathcal{I}_{\text{int}}} \prod_{j=1}^{d} g(|u_j - v_j|) \\ &\leq 4\epsilon^2 \prod_{j=1}^{d} \left(\sum_{v_j=1}^{K} g(|u_j - v_j|) \right) \\ &\leq 4\epsilon^2 \left(g(0) + 2 \sum_{k=1}^{\infty} g(k) \right)^d \\ &= 4\epsilon^2 \left(4 + 2 \sum_{k=1}^{\infty} \frac{C_0^2}{(2k-1)^2} \right)^d \\ &= 4\epsilon^2 \left(4 + \frac{\pi^2 C_0^2}{4} \right)^d =: C_d \epsilon^2, \end{aligned} \tag{130}$$

where $C_d := 4(4 + \pi^2 C_0^2/4)^d < \infty$ depends solely on the dimension d, and the bound holds uniformly for all $\mathbf{u} \in \mathcal{I}_{\text{all}}$. This completes the proof. □

C Proofs of Main Lower Bounds (Theorems 4.1 and 4.2)

C.1 Proof of Theorem 4.1 (Simple Regret Lower Bound)

Proof of Theorem 4.1. Fix $\epsilon > 0$ such that ϵ/B satisfies the smallness conditions of Lemma 5.5 under the norm bound $B/6$. Let $\mathcal{N}_\epsilon = \{z_{\mathbf{u}}\}_{\mathbf{u} \in \mathcal{I}_{\text{int}}}$ and $\{\mathcal{R}_{\mathbf{w}}\}_{\mathbf{w} \in \mathcal{I}_{\text{all}}}$ be the interior packing grid and full-domain partition of $\mathcal{X} = [0,1]^d$ constructed in Section 5.6, with cardinality $M_\epsilon = |\mathcal{I}_{\text{int}}| = \Theta((\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)})^d)$.

Fix an arbitrary reference cell index $\mathbf{u} \in \mathcal{I}_{\text{int}}$. For any alternative index $\mathbf{v} \in \mathcal{I}_{\text{int}} \setminus \{\mathbf{u}\}$, define two competing hypotheses:

$$f := f_{z_{\mathbf{u}};\epsilon}, \quad \tilde{f} := f_{z_{\mathbf{u}};\epsilon} + 5 f_{z_{\mathbf{v}};\epsilon}. \tag{131}$$

By the triangle inequality and and the translation admissibility of shifted hypotheses Lemma F.1, $\|\tilde{f}\|_{\mathcal{H}_k} \leq \|f_{z_{\mathbf{u}};\epsilon}\|_{\mathcal{H}_k} + 5\|f_{z_{\mathbf{v}};\epsilon}\|_{\mathcal{H}_k} \leq B/6 + 5(B/6) = B$.

Let $\mathcal{A}_{\mathbf{u}}$ be the event that the algorithm's recommended point $\hat{x}_T$ lies in the reference cell $\mathcal{R}_{\mathbf{u}}$, i.e., $\mathcal{A}_{\mathbf{u}} := \{\hat{x}_T \in \mathcal{R}_{\mathbf{u}}\}$.

- Under hypothesis f, the global maximum is attained in the interior of $\mathcal{R}_{\mathbf{u}}$ with peak value $\max_{x \in \mathcal{X}} f(x) \geq 2\epsilon$, while for any $x \notin \mathcal{R}_{\mathbf{u}}$, $f(x) < \epsilon$ by the off-peak decay bound. Thus, achieving simple regret $r_T \leq \epsilon$ under f requires $\hat{x}_T \in \mathcal{R}_{\mathbf{u}}$, which implies $\mathbb{P}_f(\mathcal{A}_{\mathbf{u}}) \geq 1 - \delta$.
- Under hypothesis $\tilde{f}$, the global maximum $\max_{x \in \mathcal{X}} \tilde{f}(x) \geq \tilde{f}(z_{\mathbf{v}}) = f_{z_{\mathbf{u}};\epsilon}(z_{\mathbf{v}}) + 5 f_{z_{\mathbf{v}};\epsilon}(z_{\mathbf{v}}) \geq -\epsilon + 10\epsilon = 9\epsilon$. Meanwhile, for any $x \in \mathcal{R}_{\mathbf{u}}$, $\tilde{f}(x) < 3\epsilon + 5\epsilon = 8\epsilon$. Hence, for all $x \in \mathcal{R}_{\mathbf{u}}$, $\max_{\tilde{x}} \tilde{f}(\tilde{x}) - \tilde{f}(x) > 9\epsilon - 8\epsilon = \epsilon$. Therefore, achieving $r_T \leq \epsilon$ under $\tilde{f}$ forces $\hat{x}_T \notin \mathcal{R}_{\mathbf{u}}$, which implies $\mathbb{P}_{\tilde{f}}(\mathcal{A}_{\mathbf{u}}) \leq \delta$.

Applying the change of measure inequality (Lemma 3.1) across the full partition $\{\mathcal{R}_{\mathbf{w}}\}_{\mathbf{w}\in\mathcal{I}_{\text{all}}}$:

$$\sum_{\mathbf{w}\in\mathcal{I}_{\text{all}}} \mathbb{E}_f[N_{\mathbf{w}}(T)]\overline{D}_{f,\tilde{f}}^{(\mathbf{w})} \geq \ln\frac{1}{2.4\delta}, \tag{132}$$

where $N_{\mathbf{w}}(T) := \sum_{t=1}^{T}\mathbb{I}\{x_t \in \mathcal{R}_{\mathbf{w}}\}$, and the maximum pointwise divergence on cell $\mathcal{R}_{\mathbf{w}}$ satisfies:

$$\overline{D}_{f,\tilde{f}}^{(\mathbf{w})} = \sup_{x\in\mathcal{R}_{\mathbf{w}}}\frac{(f(x)-\tilde{f}(x))^2}{2\sigma^2} = \frac{25}{2\sigma^2}\sup_{x\in\mathcal{R}_{\mathbf{w}}}(f_{z_{\mathbf{v}};\epsilon}(x))^2. \tag{133}$$

Substituting this into equation 132, for each alternative $\mathbf{v}\in\mathcal{I}_{\text{int}}\setminus\{\mathbf{u}\}$:

$$\frac{25}{2\sigma^2}\sum_{\mathbf{w}\in\mathcal{I}_{\text{all}}}\mathbb{E}_f[N_{\mathbf{w}}(T)]\sup_{x\in\mathcal{R}_{\mathbf{w}}}(f_{z_{\mathbf{v}};\epsilon}(x))^2 \geq \ln\frac{1}{2.4\delta}. \tag{134}$$

Summing this inequality over all $(M_\epsilon-1)$ possible alternatives $\mathbf{v}\in\mathcal{I}_{\text{int}}\setminus\{\mathbf{u}\}$:

$$\frac{25}{2\sigma^2}\sum_{\mathbf{w}\in\mathcal{I}_{\text{all}}}\mathbb{E}_f[N_{\mathbf{w}}(T)]\left(\sum_{\mathbf{v}\in\mathcal{I}_{\text{int}}\setminus\{\mathbf{u}\}}\sup_{x\in\mathcal{R}_{\mathbf{w}}}(f_{z_{\mathbf{v}};\epsilon}(x))^2\right) \geq (M_\epsilon-1)\ln\frac{1}{2.4\delta}. \tag{135}$$

By Lemma 5.6, the inner sum is upper bounded by $C_d\epsilon^2$. Therefore:

$$\frac{25C_d\epsilon^2}{2\sigma^2}\sum_{\mathbf{w}\in\mathcal{I}_{\text{all}}}\mathbb{E}_f[N_{\mathbf{w}}(T)] \geq (M_\epsilon-1)\ln\frac{1}{2.4\delta}. \tag{136}$$

Since $\sum_{\mathbf{w}\in\mathcal{I}_{\text{all}}}\mathbb{E}_f[N_{\mathbf{w}}(T)] = T$, we obtain:

$$T \geq \frac{2(M_\epsilon-1)\sigma^2}{25C_d\epsilon^2}\ln\frac{1}{2.4\delta}. \tag{137}$$

Finally, substituting $M_\epsilon = \Theta\left(\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)}\right)^d\right)$ establishes equation 4. □

C.2 Proof of Theorem 4.2 (Cumulative Regret Lower Bound)

Proof of Theorem 4.2. Fix $\epsilon > 0$ and an arbitrary reference cell index $\mathbf{u}\in\mathcal{I}_{\text{int}}$. For any alternative index $\mathbf{v}\in\mathcal{I}_{\text{int}}\setminus\{\mathbf{u}\}$, consider the same hypotheses:

$$f := f_{z_{\mathbf{u}};\epsilon}, \quad \tilde{f} := f_{z_{\mathbf{u}};\epsilon} + 5f_{z_{\mathbf{v}};\epsilon}. \tag{138}$$

as in Appendix C.1.

Define the event $\mathcal{A}_{\mathbf{u}} := \{N_{\mathbf{u}}(T) \geq T/2\} = \{\sum_{t=1}^{T}\mathbb{I}\{x_t\in\mathcal{R}_{\mathbf{u}}\} \geq T/2\}$.

- Under f, for any $x_t \notin \mathcal{R}_{\mathbf{u}}$, the instantaneous regret is at least $\max_x f(x) - f(x_t) \geq 2\epsilon - \epsilon = \epsilon$. Thus, if $R_T \leq T\epsilon/2$, the number of rounds where $x_t \notin \mathcal{R}_{\mathbf{u}}$ cannot exceed $T/2$. Hence, $N_{\mathbf{u}}(T) \geq T/2$, which implies $\mathbb{P}_f(R_T \leq T\epsilon/2) \leq \mathbb{P}_f(\mathcal{A}_{\mathbf{u}})$.
- Under $\tilde{f}$, the instantaneous regret for any query inside $\mathcal{R}_{\mathbf{u}}$ is at least $\max_x \tilde{f}(x) - \tilde{f}(x_t) > 9\epsilon - 8\epsilon = \epsilon$. Thus, if $N_{\mathbf{u}}(T) \geq T/2$, the cumulative regret satisfies $R_T > T\epsilon/2$. Consequently, $\mathbb{P}_{\tilde{f}}(\mathcal{A}_{\mathbf{u}}) \leq \mathbb{P}_{\tilde{f}}(R_T > T\epsilon/2)$.

If an algorithm achieves $\mathbb{P}(R_T \leq T\epsilon/2) \geq 1-\delta$ for all functions, we must have $\mathbb{P}_f(\mathcal{A}_{\mathbf{u}}) \geq 1-\delta$ and $\mathbb{P}_{\tilde{f}}(\mathcal{A}_{\mathbf{u}}) \leq \delta$. Following the exact same change of measure steps as in Appendix C.1, this requires:

$$T \geq \frac{2(M_\epsilon-1)\sigma^2}{25C_d\epsilon^2}\ln\frac{1}{2.4\delta}. \tag{139}$$

By contraposition, if ϵ is chosen such that:

$$T < \frac{2(M_\epsilon-1)\sigma^2}{25C_d\epsilon^2}\ln\frac{1}{2.4\delta}, \tag{140}$$

then there must exist some $f \in \mathcal{H}_k(\mathcal{X}, B)$ such that $\mathbb{P}_f(R_T > T\epsilon/2) > \delta$.

We now tune $\epsilon > 0$ to maximize the lower bound $T\epsilon/2$ subject to equation 140. Since $M_\epsilon = \Theta\left(\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)}\right)^d\right)$, condition equation 140 is implied by:

$$\epsilon < c_{d,\theta}\sqrt{\frac{\sigma^2}{T}\left(\ln\frac{B}{\epsilon}\right)^d\left(\ln\ln\frac{B}{\epsilon}\right)^{-d}\ln\frac{1}{\delta}}, \tag{141}$$

for a sufficiently small constant $c_{d,\theta} > 0$. We select ϵ to satisfy the fixed-point equation:

$$\epsilon = \frac{c_{d,\theta}}{2}\sqrt{\frac{\sigma^2}{T}\left(\ln\frac{B}{\epsilon}\right)^d\left(\ln\ln\frac{B}{\epsilon}\right)^{-d}\ln\frac{1}{\delta}}. \tag{142}$$

Dividing both sides by B:

$$\frac{\epsilon}{B} = \frac{c_{d,\theta}}{2}\sqrt{\frac{\sigma^2\ln(1/\delta)}{B^2T}}\left(\ln\frac{B}{\epsilon}\right)^{d/2}\left(\ln\ln\frac{B}{\epsilon}\right)^{-d/2}. \tag{143}$$

Since $\frac{\sigma^2\ln(1/\delta)}{B^2} = O(T)$ with a sufficiently small implied constant, ϵ/B can be made arbitrarily small, ensuring the validity of Lemma 5.5.

Taking logarithms on both sides of equation 142:

$$\ln\frac{B}{\epsilon} = \frac{1}{2}\ln\left(\frac{B^2T}{\sigma^2\ln(1/\delta)}\right) - \frac{d}{2}\ln\left(\left(\frac{c_{d,\theta}}{2}\right)^{2/d}\left(\ln\frac{B}{\epsilon}\right)\left(\ln\ln\frac{B}{\epsilon}\right)^{-1}\right). \tag{144}$$

For sufficiently large B/ϵ, the second term is strictly dominated:

$$\frac{d}{2}\ln\left(\left(\frac{c_{d,\theta}}{2}\right)^{2/d}\left(\ln\frac{B}{\epsilon}\right)\left(\ln\ln\frac{B}{\epsilon}\right)^{-1}\right) \le \frac{1}{4}\ln\frac{B}{\epsilon}. \tag{145}$$

Therefore:

$$\frac{1}{3}\ln\left(\frac{B^2T}{\sigma^2\ln(1/\delta)}\right) \le \ln\frac{B}{\epsilon} \le \ln\left(\frac{B^2T}{\sigma^2\ln(1/\delta)}\right). \tag{146}$$

Since the function $t \mapsto \frac{t}{\ln t}$ is strictly monotonically increasing for $t \ge e$, we can substitute this asymptotic relation back into equation 142 to deduce:

$$\epsilon = \Omega\left(\sqrt{\frac{\sigma^2}{T}\left(\ln\frac{B^2T}{\sigma^2\ln(1/\delta)}\right)^d\left(\ln\ln\frac{B^2T}{\sigma^2\ln(1/\delta)}\right)^{-d}\ln\frac{1}{\delta}}\right). \tag{147}$$

Consequently, the cumulative regret lower bound is:

$$R_T \ge \frac{T\epsilon}{2} = \Omega\left(\sqrt{T\sigma^2\left(\ln\frac{B^2T}{\sigma^2\ln(1/\delta)}\right)^d\left(\ln\ln\frac{B^2T}{\sigma^2\ln(1/\delta)}\right)^{-d}\ln\frac{1}{\delta}}\right). \tag{148}$$

This completes the proof. □

D Proofs of Lower Bounds for Expected Regret

In this section, we provide the formal proofs for Corollary 4.3, which translates our high-probability minimax regret lower bounds into lower bounds on the expected regret. The proof strategy closely follows the reduction established by Iwazaki (2026) via Markov's inequality.

Proof of Expected Simple Regret Bound (Corollary 4.3). Fix an arbitrary algorithm and choose the failure probability $\delta = 1/6$. Suppose for contradiction that the algorithm achieves $\mathbb{E}[r_T] \le \epsilon$ for

all $f \in \mathcal{H}_k(\mathcal{X}, B)$. By Markov's inequality, the probability that the simple regret exceeds 6ϵ must satisfy:

$$\mathbb{P}(r_T > 6\epsilon) \leq \frac{\mathbb{E}[r_T]}{6\epsilon} \leq \frac{\epsilon}{6\epsilon} = \delta. \tag{149}$$

Equivalently, $\mathbb{P}(r_T \leq 6\epsilon) \geq 1 - \delta = 5/6$ holds uniformly for all $f \in \mathcal{H}_k(\mathcal{X}, B)$.

We now invoke Theorem 4.1 with the target accuracy parameter set to $\epsilon' = 6\epsilon$ and the fixed probability parameter $\delta = 1/6$. Theorem 4.1 dictates that if an algorithm achieves simple regret at most 6ϵ with probability at least $5/6$, the number of time steps T must satisfy:

$$T \geq \Omega\left(\frac{\sigma^2}{\epsilon^2}\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)}\right)^d\right), \tag{150}$$

Inverting this monotonic relation with respect to ϵ implies that for any budget T, any algorithm must suffer expected simple regret $\sup_{f\in\mathcal{H}_k(\mathcal{X},B)} \mathbb{E}[r_T] = \Omega\left(\sqrt{\frac{\sigma^2}{T}\left(\ln\frac{B^2T}{\sigma^2}\right)^d\left(\ln\ln\frac{B^2T}{\sigma^2}\right)^{-d}}\right)$, which completes the proof. □

Proof of Expected Cumulative Regret Bound (Corollary 4.3). Under the assumption that $\sigma^2/B^2 = O(T)$ with a sufficiently small implied constant, we set the failure probability to a constant $\delta = 1/6 \in (0, 1/3)$ in Theorem 4.2.

Then, Theorem 4.2 guarantees the existence of a difficult reward function $f^* \in \mathcal{H}_k(\mathcal{X}, B)$ such that the cumulative regret satisfies:

$$\mathbb{P}_{f^*}\left(R_T \geq C_0\sqrt{T\sigma^2\left(\ln\frac{B^2T}{\sigma^2}\right)^d\left(\ln\ln\frac{B^2T}{\sigma^2}\right)^{-d}}\right) \geq \delta = \frac{1}{6}, \tag{151}$$

for some positive constant $C_0 > 0$ depending only on d and θ.

$$\mathbb{E}_{f^*}[R_T] \geq \frac{C_0}{6}\sqrt{T\sigma^2\left(\ln\frac{B^2T}{\sigma^2}\right)^d\left(\ln\ln\frac{B^2T}{\sigma^2}\right)^{-d}}. \tag{152}$$

This immediately implies $\sup_{f\in\mathcal{H}_k(\mathcal{X},B)} \mathbb{E}[R_T] = \Omega\left(\sqrt{T\sigma^2\left(\ln\frac{B^2T}{\sigma^2}\right)^d\left(\ln\ln\frac{B^2T}{\sigma^2}\right)^{-d}}\right)$. □

E Proof of Lower Bounds for General Compact Domains with Non-empty Interior

In this section, we formally establish the extension of our minimax regret lower bounds (Theorems 4.1, 4.2, and Corollary 4.3) from the unit hypercube $[0, 1]^d$ to an arbitrary compact domain $\mathcal{X} \subset \mathbb{R}^d$ with non-empty interior $\mathrm{int}(\mathcal{X}) \neq \emptyset$.

Corollary E.1 (Restatement for General Compact Domains)**.** *Let $\mathcal{X} \subset \mathbb{R}^d$ be any compact subset with non-empty interior. Consider the Gaussian process bandit problem on $\mathcal{X}$ with the squared exponential kernel $k(x, \tilde{x}) = \exp(-\|x - \tilde{x}\|_2^2/\theta)$ and RKHS norm bound $\|f\|_{\mathcal{H}_k(\mathcal{X})} \leq B$. Then:*

1. ***Simple Regret:*** *Any algorithm guaranteeing $r_T \leq \epsilon$ with probability at least $1-\delta$ requires:*

$$T = \Omega\left(\frac{\sigma^2}{\epsilon^2}\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)}\right)^d \ln\frac{1}{\delta}\right). \tag{153}$$

2. ***Cumulative Regret:*** *For any algorithm, there exists $f \in \mathcal{H}_k(\mathcal{X}, B)$ such that with probability at least δ:*

$$R_T = \Omega\left(\sqrt{T\sigma^2\left(\ln\frac{B^2T}{\sigma^2\ln\frac{1}{\delta}}\right)^d\left(\ln\ln\frac{B^2T}{\sigma^2\ln\frac{1}{\delta}}\right)^{-d}}\right). \tag{154}$$

3. ***Expected Regret:*** *The expected simple and cumulative regrets satisfy the exact lower bound rates of Corollary 4.3.*

Here, the asymptotic Ω-notation suppresses constants depending solely on d, θ, and the geometric embedding scale of $\mathcal{X}$.

Proof. The proof extends our Euclidean construction from the canonical unit hypercube to any general compact domain $\mathcal{X} \subset \mathbb{R}^d$ with non-empty interior by establishing a global affine contraction that preserves the interior spectral regime of Appendix A.

Since $\mathrm{int}(\mathcal{X}) \neq \emptyset$, there exists an interior anchor point $x_0 \in \mathrm{int}(\mathcal{X})$ and an open radius $r_0 > 0$ such that the ℓ_∞-ball satisfies $\bar{B}_\infty(x_0, r_0) \subset \mathrm{int}(\mathcal{X})$. Furthermore, because $\mathcal{X}$ is compact, it is bounded with finite radius $R_\infty := \sup_{x \in \mathcal{X}} \|x - x_0\|_\infty < \infty$.

We choose a fixed scaling parameter $s := 4R_\infty > 0$ and denote the center of the unit hypercube by $\mathbf{c}_0 := (1/2, \ldots, 1/2)^\top \in \mathbb{R}^d$. We define the global affine contraction mapping $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ by:

$$u = \Phi(x) := \frac{x - x_0}{s} + \mathbf{c}_0, \quad \text{with inverse} \quad x = \Phi^{-1}(u) = s(u - \mathbf{c}_0) + x_0. \tag{155}$$

By construction, $\Phi(x_0) = \mathbf{c}_0$, and for every $x \in \mathcal{X}$:

$$\|\Phi(x) - \mathbf{c}_0\|_\infty = \frac{\|x - x_0\|_\infty}{s} \leq \frac{R_\infty}{4R_\infty} = \frac{1}{4}. \tag{156}$$

Consequently, the contracted domain $\widetilde{\mathcal{X}} := \Phi(\mathcal{X})$ is strictly embedded inside the central hypercube:

$$\widetilde{\mathcal{X}} \subseteq \left[\frac{1}{4}, \frac{3}{4}\right]^d \subset [0, 1]^d. \tag{157}$$

Moreover, the image of the interior ball contains a fixed central core hypercube:

$$\mathcal{Q}_{\mathrm{core}} := \mathbf{c}_0 + [-\rho_0, \rho_0]^d \subset \Phi\left(\bar{B}_\infty(x_0, r_0)\right) \subset \mathrm{int}(\widetilde{\mathcal{X}}), \tag{158}$$

where $\rho_0 := \frac{r_0}{s} = \frac{r_0}{4R_\infty} \in (0, 1/4)$ is a strictly positive geometric constant that depends solely on $\mathcal{X}$ and is independent of T, ϵ, δ, B, and σ^2.

Under the bijective affine transformation $x = \Phi^{-1}(u)$, optimizing an unknown reward function $f(x)$ over $x \in \mathcal{X}$ with observations $y_t = f(x_t) + \xi_t$ is isometrically equivalent to optimizing $\tilde{f}(u) := f(\Phi^{-1}(u))$ over $u \in \widetilde{\mathcal{X}}$ with queries $u_t = \Phi(x_t)$ and observations $y_t = \tilde{f}(u_t) + \xi_t$.

For the target SE covariance kernel $k(x, \tilde{x}) = \exp(-\|x - \tilde{x}\|_2^2/\theta)$, the pulled-back covariance function on the transformed space is given by:

$$\tilde{k}(u, \tilde{u}) := k(\Phi^{-1}(u), \Phi^{-1}(\tilde{u})) = \exp\left(-\frac{\|s(u - \tilde{u})\|_2^2}{\theta}\right) = \exp\left(-\frac{\|u - \tilde{u}\|_2^2}{\tilde{\theta}}\right), \tag{159}$$

which is an SE kernel on $\widetilde{\mathcal{X}}$ equipped with the scaled lengthscale parameter $\tilde{\theta} := \theta/s^2$. The map $f \mapsto \tilde{f} = f \circ \Phi^{-1}$ is an exact linear isometry between $\mathcal{H}_k(\mathcal{X})$ and $\mathcal{H}_{\tilde{k}}(\widetilde{\mathcal{X}})$:

$$\|\tilde{f}\|_{\mathcal{H}_{\tilde{k}}(\widetilde{\mathcal{X}})} = \|f\|_{\mathcal{H}_k(\mathcal{X})} \leq B. \tag{160}$$

Hence, bounding the minimax regret on $(\mathcal{X}, k)$ is completely equivalent to bounding the minimax regret on $(\widetilde{\mathcal{X}}, \tilde{k})$.

Core Packing and Full-Domain Partition. Let $w_\epsilon = \Theta\left(\frac{\ln \ln(B/\epsilon)}{\ln(B/\epsilon)}\right)$ be the localization peak width determined by the PSWF eigensystem with bandwidth parameter $\Omega_0 = c$ tuned to the kernel parameter $\tilde{\theta}$. We partition the entire bounding unit cube $[0, 1]^d$ into uniform cubic cells of width w_ϵ, indexed by $\mathbf{w} \in \mathcal{I}_{\mathrm{all}}$.

We designate the interior candidate set as those cells whose closures are strictly contained within the core $\mathcal{Q}_{\mathrm{core}}$:

$$\mathcal{I}_{\mathrm{int}} := \{\mathbf{u} \in \mathcal{I}_{\mathrm{all}} : \mathcal{R}_{\mathbf{u}} \subseteq \mathcal{Q}_{\mathrm{core}}\}, \tag{161}$$

with centers $z_{\mathbf{u}} \in \mathcal{Q}_{\text{core}}$. Since $\rho_0 > 0$ is a fixed constant, the packing cardinality preserves the exact asymptotic scaling:

$$M_\epsilon := |\mathcal{I}_{\text{int}}| \geq \left(\left\lfloor \frac{2\rho_0}{w_\epsilon} \right\rfloor \right)^d = \Theta\left(\rho_0^d w_\epsilon^{-d}\right) = \Theta\left(\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)} \right)^d \right). \tag{162}$$

We define the complete partition of the active domain $\widetilde{\mathcal{X}}$ by intersecting the ambient cells with $\widetilde{\mathcal{X}}$:

$$\mathcal{C}_{\mathbf{u}} := \mathcal{R}_{\mathbf{u}} \quad (\text{for } \mathbf{u} \in \mathcal{I}_{\text{int}}), \quad \text{and} \quad \mathcal{C}_{\mathbf{w}} := \mathcal{R}_{\mathbf{w}} \cap \widetilde{\mathcal{X}} \quad (\text{for } \mathbf{w} \in \mathcal{I}_{\text{ext}} := \mathcal{I}_{\text{all}} \setminus \mathcal{I}_{\text{int}}). \tag{163}$$

By construction, $\{\mathcal{C}_{\mathbf{w}}\}_{\mathbf{w} \in \mathcal{I}_{\text{all}}}$ forms a disjoint partition of $\widetilde{\mathcal{X}}$, guaranteeing $\sum_{\mathbf{w} \in \mathcal{I}_{\text{all}}} N_{\mathbf{w}}(T) = T$ for any sequential query sequence $u_1, \ldots, u_T \in \widetilde{\mathcal{X}}$.

For each $\mathbf{u} \in \mathcal{I}_{\text{int}}$, we assign the shifted hypothesis $\tilde{f}_{z_{\mathbf{u}};\epsilon}(u) := f_{\epsilon,\bar{N}}(u - z_{\mathbf{u}})$. By Lemma 5.1 and Aronszajn's restriction theorem (Lemma 5.2), $\|\tilde{f}_{z_{\mathbf{u}};\epsilon}\|_{\mathcal{H}_{\tilde{k}}(\widetilde{\mathcal{X}})} \leq \|\tilde{f}_{z_{\mathbf{u}};\epsilon}\|_{\mathcal{H}_{\tilde{k}}(\mathbb{R}^d)} \leq B/6$, so all candidate hypotheses and their combinations are strictly admissible in $\mathcal{H}_{\tilde{k}}(\widetilde{\mathcal{X}}, B)$.

Crucially, for *any* query point $u \in \widetilde{\mathcal{X}}$ and for any candidate hypothesis center $z_{\mathbf{v}} \in \mathcal{Q}_{\text{core}}$, both points belong to $[1/4, 3/4]^d$ by equation 157. Consequently, the coordinate-wise displacement satisfies:

$$\|u - z_{\mathbf{v}}\|_\infty \leq \frac{1}{2}. \tag{164}$$

Therefore, the spatial off-peak envelope $|f_{\epsilon,\bar{N}}(u - z_{\mathbf{v}})| \leq 2\epsilon \prod_{j=1}^d \frac{C_0'}{\bar{N}|u_j - z_{\mathbf{v},j}|}$ from Lemma 5.5 holds unconditionally for all $u \in \widetilde{\mathcal{X}}$. Following the exact multi-dimensional summation bound in Lemma 5.6, there exists a constant $C_{d,\tilde{\theta}} > 0$ such that for every partition cell $\mathbf{w} \in \mathcal{I}_{\text{all}}$:

$$\sum_{\mathbf{v} \in \mathcal{I}_{\text{int}}} \sup_{u \in \mathcal{C}_{\mathbf{w}}} (\tilde{f}_{z_{\mathbf{v}};\epsilon}(u))^2 \leq C_{d,\tilde{\theta}} \epsilon^2. \tag{165}$$

With the uniform cross-region interference bound equation 165 established across the entire partition of $\widetilde{\mathcal{X}}$, the change-of-measure inequality (Lemma 3.1) across all alternatives $\mathbf{v} \in \mathcal{I}_{\text{int}} \setminus \{\mathbf{u}\}$ yields:

$$\frac{25 C_{d,\tilde{\theta}} \epsilon^2 T}{2\sigma^2} \geq (M_\epsilon - 1) \ln \frac{1}{2.4\delta}. \tag{166}$$

Rearranging terms and substituting $M_\epsilon = \Theta\left(\left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)} \right)^d \right)$ gives:

$$T = \Omega\left(\frac{\sigma^2}{\epsilon^2} \left(\frac{\ln(B/\epsilon)}{\ln\ln(B/\epsilon)} \right)^d \ln \frac{1}{\delta} \right), \tag{167}$$

where the suppressed constants depend only on d, θ, and the scale ratio r_0/R_∞ of the domain $\mathcal{X}$. The cumulative regret lower bound follows from identical parameter tuning ϵ, and the expected regret lower bounds follow directly via Markov's inequality as in Appendix D. This completes the proof. □

F Helper Mathematical Lemmas

Lemma F.1 (Spectral Representation and Translation Isometry in Stationary RKHSs (Wendland, 2004, Theorem 10.12)). *Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a stationary kernel defined by $k(x, y) = \Phi(x - y)$, where $\Phi \in C(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$ is a real-valued symmetric strictly positive definite function. Under the classical Fourier transform convention:*

$$\hat{f}(\omega) := \int_{\mathbb{R}^d} f(x) e^{-i\omega^\top x} \, \mathrm{d}x, \qquad f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\omega) e^{i\omega^\top x} \, \mathrm{d}\omega, \tag{168}$$

The reproducing kernel Hilbert space $\mathcal{H}_k(\mathbb{R}^d)$ associated with k satisfies:

1. ***Explicit Spectral Characterization:*** *The space $\mathcal{H}_k(\mathbb{R}^d)$ is isometrically isomorphic to the weighted L_2 space in the frequency domain:*

$$\mathcal{H}_k(\mathbb{R}^d) = \left\{ f \in L_2(\mathbb{R}^d) \cap C(\mathbb{R}^d) \;\middle|\; \frac{\hat{f}}{\sqrt{\hat{\Phi}}} \in L_2(\mathbb{R}^d) \right\}, \tag{169}$$

equipped with the inner product and RKHS norm:

$$\langle f, g \rangle_{\mathcal{H}_k(\mathbb{R}^d)} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{\hat{f}(\omega)\overline{\hat{g}(\omega)}}{\hat{\Phi}(\omega)} \, \mathrm{d}\omega, \tag{170}$$

$$\|f\|^2_{\mathcal{H}_k(\mathbb{R}^d)} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{\Phi}(\omega)} \, \mathrm{d}\omega. \tag{171}$$

2. ***Global Translation Isometry:*** *For any shift vector $z \in \mathbb{R}^d$ and any $f \in \mathcal{H}_k(\mathbb{R}^d)$, the translated function $f_z(x) := f(x - z)$ also belongs to $\mathcal{H}_k(\mathbb{R}^d)$ and satisfies the exact norm identity:*

$$\|f_z\|_{\mathcal{H}_k(\mathbb{R}^d)} = \|f\|_{\mathcal{H}_k(\mathbb{R}^d)}. \tag{172}$$

Lemma F.2 (Christoffel-Darboux Formula for Orthogonal Polynomials)**.** *Let $\{p_n(x)\}_{n=0}^{\infty}$ be a family of orthonormal polynomials on an interval $[a, b]$ with weight function $w(x)$, satisfying the three-term recurrence $xp_n(x) = a_n p_{n+1}(x) + b_n p_n(x) + a_{n-1}p_{n-1}(x)$. Then for any $x \neq y$:*

$$\sum_{k=0}^{n} p_k(x)p_k(y) = a_n \frac{p_{n+1}(x)p_n(y) - p_n(x)p_{n+1}(y)}{x - y}. \tag{173}$$

Furthermore, by taking the limit as $y \to x$, we obtain the confluent form*:*

$$\sum_{k=0}^{n} p_k(x)^2 = a_n \big(p'_{n+1}(x)p_n(x) - p'_n(x)p_{n+1}(x)\big). \tag{174}$$

Lemma F.3 (Slepian's Super-Exponential Eigenvalue Decay Slepian (1965))**.** *Let $\lambda_n(c)$ denote the eigenvalues of the Sinc kernel integral operator on $[-1, 1]$ with fixed bandwidth $c > 0$. As $n \to \infty$, the eigenvalues satisfy the asymptotic expansion:*

$$\lambda_n(c) \sim \frac{2\pi}{(n!)^2} \left(\frac{c}{4}\right)^{2n+1} = \exp\left(-2n \ln\left(\frac{4n}{ec}\right) + O(\ln n)\right). \tag{175}$$

Lemma F.4 (Extremal Properties of $P_{2m+1}(x)/x$)**.** *Let $P_n(x)$ be the Legendre polynomial of degree $n = 2m + 1 \geq 3$. The function $y(x) = \frac{P_{2m+1}(x)}{x}$ satisfies the self-adjoint differential equation*

$$\frac{d}{dx}\Big(x^2(1 - x^2)y'(x)\Big) + x^2[n(n + 1) - 2]y(x) = 0. \tag{176}$$

Moreover, the successive relative maxima of $|y(x)|$ strictly decrease on $(0, \sqrt{2/3})$ and strictly increase on $(\sqrt{2/3}, 1)$. Consequently, for all $n \geq 3$, the global maximum of $|y(x)|$ on $[-1, 1]$ is uniquely attained at $x = 0$, with $|y(0)| = |P'_{2m+1}(0)| \sim \sqrt{2n/\pi}$.

Proof. Substituting $P_n(x) = xy(x)$ into the standard Legendre differential equation $(1 - x^2)P''_n(x) - 2xP'_n(x) + n(n + 1)P_n(x) = 0$ and multiplying by x yields equation 176.

Writing equation 176 in the Sturm–Liouville form $(p(x)y')' + q(x)y = 0$, we have $p(x) = x^2(1 - x^2)$ and $q(x) = x^2[n(n + 1) - 2]$. By the Sonin–Pólya theorem (Szegő, 1975, Theorem 7.31.1), the successive relative maxima of $|y(x)|$ form an increasing or decreasing sequence according as $p(x)q(x)$ decreases or increases. A straightforward calculation gives

$$\frac{d}{dx}\big[p(x)q(x)\big] = [n(n + 1) - 2] \cdot 2x^3(2 - 3x^2),$$

which is strictly positive on $(0, \sqrt{2/3})$ and strictly negative on $(\sqrt{2/3}, 1)$. Hence, the relative maxima strictly decrease on $(0, \sqrt{2/3})$ and strictly increase on $(\sqrt{2/3}, 1)$.

Finally, for $x \in [\sqrt{2/3}, 1]$, using the universal bound $|P_n(x)| \leq 1$ (Szegő, 1975, Theorem 7.2.1), we have $|y(x)| \leq 1/x \leq \sqrt{3/2}$. On the other hand, at the origin, $|y(0)| = |P'_{2m+1}(0)| \geq 1.5$ for all $n \geq 3$, which strictly dominates the maximum on $[\sqrt{2/3}, 1]$. Thus, the global maximum on $[-1, 1]$ is attained at $x = 0$. □