QUERY LOWER BOUNDS FOR DIFFUSION SAMPLING

Zhiyang Xun
University of Texas at Austin
zxun@cs.utexas.edu

Eric Price
University of Texas at Austin
ecprice@cs.utexas.edu

ABSTRACT

Diffusion models generate samples by iteratively querying learned score estimates. A rapidly growing literature focuses on accelerating sampling by minimizing the number of score evaluations, yet the information-theoretic limits of such acceleration remain unclear.

In this work, we establish the first score query lower bounds for diffusion sampling. We prove that for d-dimensional distributions, given access to score estimates with polynomial accuracy $\varepsilon = d^{-O(1)}$ (in any L^p sense), any sampling algorithm requires $\widetilde{\Omega}(\sqrt{d})$ adaptive score queries. In particular, our proof shows that any sampler must search over $\widetilde{\Omega}(\sqrt{d})$ distinct noise levels, providing a formal explanation for why multiscale noise schedules are necessary in practice.

1 INTRODUCTION

Diffusion models have become a central paradigm in modern generative modeling, enabling major advances in tasks ranging from high-fidelity image synthesis to scientific computing (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021b; Nichol & Dhariwal, 2021; Karras et al., 2022). The key to their success lies in an iterative formulation: rather than generating samples directly in a single shot, these models progressively transform simple noise into structured data by repeatedly evaluating the *smoothed scores* of the distribution (Song & Ermon, 2019; Song et al., 2021b). By reducing the task of sampling from complex high-dimensional multimodal distributions into estimating the scores, this framework has dramatically improved generative modeling capabilities.

A central goal within this framework is improving sampling efficiency. Since each iteration typically requires evaluating a large neural network, the computational cost is dominated by the total number of such evaluations (queries). This has motivated a large body of work on reducing query complexity, including higher-order numerical solvers, optimized discretization schemes, and flow matching techniques that straighten generation trajectories (Lu et al., 2022; 2025; Liu et al., 2022a; Zhang & Chen, 2023; Zhao et al., 2023; Karras et al., 2022; Lipman et al., 2023; Liu et al., 2022b; Fukumizu et al., 2025). These advances have steadily reduced the number of queries required in both theory and practice.

This rapid progress raises a basic theoretical question: as we aim to reduce the number of score queries, what are the intrinsic information-theoretic limits of diffusion sampling in high dimension? When access to the target distribution is limited to querying smoothed score estimates, is there an unavoidable lower bound on the number of such queries? In this work, we ask:

Can we establish a query-complexity lower bound for score-based sampling?

Recent theory has made substantial progress on *upper* bounds. In particular, for targets with bounded second moments, one can sample in $\widetilde{O}(d)$ iterations given access to L^2-accurate smoothed score estimates (Chen et al., 2023c; Lee et al., 2022; Benton et al., 2024). At the same time, practical diffusion samplers often produce high-quality samples in only a handful of steps, far fewer than worst-case theoretical guarantees. This gap makes lower bounds essential: is a polynomial dependence on d truly unavoidable, or could one hope for $\mathrm{polylog}(d)$—or even dimension-independent—query complexity?

Ruling out this possibility turns out to be very non-trivial. Information-theoretically, *smoothed* score estimates are very powerful in sampling, suggesting that such efficiency improvements might indeed be feasible:

1. Each query of a smoothed score returns a full vector in $\mathbb{R}^d$, providing $\Omega(d)$ bits of information. Thus, if each query can extract a near-optimal amount of information, the intrinsic difficulty of sampling might not necessarily grow with dimension.
2. Although scores are by definition local quantities, the smoothed scores can reveal global properties: Gaussian convolution can encode global distributional structure into local gradients, potentially enabling efficient navigation of high-dimensional space.
3. Even when restricted to L^2-accurate score estimates, which only guarantee accuracy in expectation, smoothing can help overcome this limitation. By choosing large smoothing levels, the algorithm can ensure that queries occur within the typical set of the smoothed distribution, where an L^2 guarantee is meaningful.

Furthermore, recent results hint at a gap between limitations of specific discretizations and the potential power of smoothed-score access. For instance, Jiao et al. (2025); Gao & Zhu (2025) show that even for a d-dimensional standard Gaussian, standard discretizations can require $\Omega(\sqrt{d})$ steps due to error accumulation. On the other hand, work on high-accuracy regimes has begun to explore ε-accurate sampling with $\mathrm{polylog}(1/\varepsilon)$ dependence (Gatmiry et al., 2026), raising the possibility of sampling with $\mathrm{polylog}(d)$ steps on product distributions. These considerations leave open whether there is a fundamental, information-theoretic barrier to few-step diffusion sampling in the worst case.

1.1 Our Results

In this work, we establish the first query-complexity lower bounds for diffusion sampling. Our main message is that, under standard distributional assumptions and polynomially accurate score estimates, there is an information-theoretic $\widetilde{\Omega}(\sqrt{d})$ barrier on the number of score queries required to obtain a non-trivial sample.

To state our bounds formally, we first introduce the necessary notation. Let π be a target distribution on $\mathbb{R}^d$. For any noise level $\sigma > 0$ we write

$$\pi_\sigma = \pi * \mathcal{N}(0, \sigma^2 I_d), \qquad s_{\pi,\sigma}(x) = \nabla \log \pi_\sigma(x),$$

for the Gaussian-smoothed distribution and its score. We model diffusion samplers as follows:

Definition 1.1 (Diffusion Sampling)**.** *A diffusion sampling algorithm $\mathcal{A}$ accesses a target distribution π exclusively via adaptive queries to an oracle for smoothed score estimates $\widehat{s}_\sigma(x)$, outputting a sample $\widehat{X}$ whose law aims to approximate π.*

To guarantee convergence, assumptions on the class of target distributions and the quality of the score estimates are needed. We work under two *strong* conditions (i.e., favorable to the algorithm) that are standard in the diffusion-theory literature; therefore, our lower bounds apply to a broad range of algorithms.

First, we assume the target is a Gaussian smoothing of a bounded-support distribution.

Assumption 1 (Bounded plus noise)**.** *For constant $R > 0$ and constant $\gamma \in (0, R/2)$, there exists a distribution π_{pre} supported on $[-R, R]^d$ such that*

$$\pi = \pi_{\mathrm{pre}} * \mathcal{N}(0, \gamma^2 I_d).$$

Bounded-plus-noise is stronger than bounded moments or subgaussian tails, and it automatically implies $M_2 = \mathbb{E}[\|X\|_2^2] = O(d)$ as well as bounds on all higher moments. The additional Gaussian smoothing ensures that π is infinitely differentiable and mirrors the smoothed target distributions typically sampled by diffusion models (e.g., via early stopping). Furthermore, this smoothing enables Total Variation (TV) convergence guarantees for sampling algorithms.

Second, we assume the score estimate is polynomially L^p-accurate. The case $p = 2$ matches the standard assumption in diffusion theory analysis, as it aligns with the standard score matching training objective used in practice (Hyvärinen, 2005; Vincent, 2011; Song & Ermon, 2019; Ho et al., 2020; Song et al., 2021b; Gupta et al., 2024). We state the assumption for any constant $p \geq 2$ to cover a wider range of algorithms (Tang & Yang, 2024; Xun et al., 2025).

Assumption 2 (L^p-accurate score)**.** *There exists a constant $p \geq 2$ and $\varepsilon`_{\mathrm{err}} = 1/\operatorname{poly}(d)$ such that for all $\sigma > 0$,*

$$\mathbb{E}_{X \sim \pi_\sigma}\left[\|\widehat{s}_\sigma(X) - s_{\pi,\sigma}(X)\|_2^p\right] \leq \frac{\varepsilon_{\mathrm{err}}^p}{\sigma^p}.$$

Under Assumptions 1 and 2, the best known upper bound is that $\widetilde{O}(d)$ queries can give us very small TV error. The bound can be achieved by analyzing the discretization error of standard algorithms like DDPM or DDIM (Benton et al., 2024; Li & Yan, 2024; Conforti et al., 2023; Li et al., 2024b). We restate the result of Benton et al. (2024) below, translating their original KL bound to TV distance to facilitate comparison:

Theorem 1.2 (Benton et al. (2024))**.** *Under Assumptions 1 and 2, there exists a diffusion sampling algorithm (DDPM) that achieves TV error 0.01 using $\widetilde{O}(d)$ queries.*

Our main theorem shows that one cannot reduce the query complexity below $\widetilde{\Omega}(\sqrt{d})$ in the worst case, even to achieve a very large constant error.

Theorem 1.3 (Main Theorem)**.** *Any diffusion sampling algorithm under Assumptions 1 and 2 with TV error less than 0.99 requires at least $\widetilde{\Omega}(\sqrt{d})$ queries.*

More recently, researchers have shown that beyond Assumptions 1 and 2, adding additional score Lipschitz assumptions can lead to algorithm acceleration, achieving sublinear iteration complexity in d (Gupta et al., 2025; Li et al., 2024a; Wu et al., 2024; Li & Jiao, 2025). For example, Zhang et al. (2025) showed that if the score estimates are L-Lipschitz, then DDPM only needs $O(L\sqrt{d})$ steps of queries; Jiao & Li (2024) gave an algorithm with $\min(d, L^{1/3}d^{2/3}, Ld^{1/3})$ queries, where L is the Lipschitzness of the true scores.

Therefore, a useful perspective is that our hard instance has globally Lipschitz true smoothed scores and score estimates, with a Lipschitz parameter that scales linearly with dimension.

Remark 1.4 (Lipschitz scale of the smoothed score)**.** *In the hard instance underlying Theorem 1.3, for every $\sigma > 0$, both $s_{\pi,\sigma}$ and $\widehat{s}_\sigma$ are globally $O(d)$-Lipschitz.*

This implies that any algorithm with a query upper bound of $\widetilde{O}(L^a d^b)$ yields a complexity of $\widetilde{O}(d^{a+b})$ when instantiated on our hard instance. Theorem 1.3 therefore rules out query upper bounds of the form $\widetilde{O}(L^a d^b)$ with $a + b < 1/2$.

A more formal version of the theorem, which tracks the dependence on (R, γ), the oracle accuracy, and the exact Lipschitz constant, is given in Theorem A.1.

At the proof level, our lower bound isolates one intrinsic source of hardness: any diffusion sampler must scan through $\widetilde{\Omega}(\sqrt{d})$ distinct noise levels. This matches what diffusion algorithms do in practice, where scores are queried along a multiscale noise schedule, and provides a formal explanation for why such a design is necessary. In fact, we establish something stronger: we construct a null distribution π_{null} and a class of planted distributions $\mathcal{D}$, and show that $\widetilde{\Omega}(\sqrt{d})$ queries at distinct noise levels are needed merely to distinguish whether the score estimates come from π_{null} or from some π drawn from $\mathcal{D}$. Since producing a valid sample from π is strictly harder than extracting this single bit of information, the sampling lower bound follows immediately.

Under Assumptions 1 and 2, a gap remains between our $\widetilde{\Omega}(\sqrt{d})$ lower bound and the best known $\widetilde{O}(d)$ upper bound, and determining the true query complexity is an important open problem. The fact that our proof yields the $\widetilde{\Omega}(\sqrt{d})$ barrier for a task strictly easier than sampling points to a concrete direction for tightening it: going beyond 1-bit distinguishing and directly lower-bounding the difficulty of producing a sample. On the other hand, it is also plausible that scanning noise levels already captures the dominant source of hardness and that the true complexity is $\widetilde{\Theta}(\sqrt{d})$, in which case our hard instance may inform the design of faster diffusion sampling algorithms that close the gap from above.

Subexponential Error Tail. While L^p accuracy is the standard assumption, it is also instructive to consider the impact of stronger error tail conditions. Recent works have shown that stronger error tail requirements might yield accelerated algorithms (Gatmiry et al., 2026).

Assumption 3 (Subexponential Error)**.** *For all $\sigma > 0$ and all $z \geq 0$, the model can query $\widehat{s}_\sigma(x)$ with the guarantee that*

$$\Pr_{X\sim\pi_\sigma}\left[\|\widehat{s}_\sigma(X) - s_{\pi,\sigma}(X)\|_2 \geq z\right] \;\leq\; 2\exp\left(-\frac{z\sigma}{\varepsilon_{\mathrm{err}}}\right).$$

We show that, under a constant sub-exponential error model, $\Omega(d^{1/4})$ queries are still necessary.

Theorem 1.5 (Subexponential Tail)**.** *For any constant $\varepsilon_{\mathrm{err}} > 0$, any algorithm that solves diffusion sampling under Assumptions 1 and 3 with TV error less than 0.99 requires at least $\Omega(d^{1/4})$ queries.*

2 Intuition on Hardness

To build intuition, we consider a simple hard family: mixtures of n well-separated Gaussians. Concretely, pick a uniformly random set of centers $S \subseteq \{-1, +1\}^d$ with $|S| = n$, and define

$$\pi_{\mathrm{pre}} = \mathrm{Unif}(S), \qquad \pi = \pi_{\mathrm{pre}} * \mathcal{N}(0, 0.1I_d).$$

Clearly, any nontrivial sampler must locate at least one center in S using score queries. Our proof shows that this is only possible when the algorithm queries a *narrow window* of smoothing levels determined by n. Since n is hidden and may range over $2^{\Theta(d)}$ possibilities, the sampler is forced to scan many different noise levels.

For conciseness, in this section we slightly abuse notation and write $\pi_\sigma := \pi_{\mathrm{pre}} * \mathcal{N}(0, \sigma^2 I_d)$ and we write $s_{\pi,\sigma}$ for the score of $\pi_{\mathrm{pre}} * \mathcal{N}(0, \sigma^2 I_d)$ throughout (so σ^2 should be read as an effective variance). By Tweedie's formula, suppose we sample a center $Y \sim \pi_{\mathrm{pre}}$ and noise $Z \sim \mathcal{N}(0, \sigma^2 I_d)$, and set $X = Y + Z$, so $X \sim \pi_\sigma$, then the score can be rewritten as

$$s_{\pi,\sigma}(x) = \frac{m(x) - x}{\sigma^2},$$

where $m(x) = \mathbb{E}[Y \mid X = x]$. Since Y is supported on S, $m(x)$ is a posterior average over the modes $y \in S$, with weights proportional to $\exp(-\|x - y\|_2^2/(2\sigma^2))$. Thus a score query is informative only insofar as this posterior average *depends on the planted set S*.

As a reference, let s_σ^{null} denote the score one would obtain if the adversary pretended $S = \{-1, +1\}^d$ (so the answer carries no information about S). We will see that for most σ, an adversary can answer essentially with s_σ^{null} while still satisfying the polynomial $L^p(\pi_\sigma)$ accuracy guarantee, forcing the sampler to scan across σ.

Warmup: Gaussian as a sphere. A Gaussian in $\mathbb{R}^d$ concentrates on a thin shell: $\|Z\|_2 \approx \sigma\sqrt{d}$. Imagine that it were *exactly uniform* on the sphere of radius $\sigma\sqrt{d}$. Then for fixed (x, σ), $m(x)$ would simply average the sampled centers lying in the corresponding sphere around x. Now we count how many points in S fall into this sphere, and let $N(x, \sigma)$ denote this expected number over the random draw of S.

This yields two uninformative extremes:

- If $N(x, \sigma) \gg \mathrm{poly}(d)$, with high probability the averaging washes out the instance: by concentration, $m(x)$ becomes $1/\mathrm{poly}(d)$-close to the null answer, so an adversary can answer with s_σ^{null}.
- If $N(x, \sigma) \ll 1/\mathrm{poly}(d)$, then with high probability over S there are no points in the sphere. This implies that $\pi_\sigma(x)$ has tiny mass, and the $L^p(\pi_\sigma)$ guarantee does not constrain the oracle on these points. The adversary can again answer with s_σ^{null}.

Either way, a query at this σ reveals essentially nothing except that the chosen smoothing level is "wrong."

To get more information, σ must be tuned so that $N(x, \sigma)$ lies within an intermediate window of a $\mathrm{poly}(d)$ range. Note that $N(x, \sigma)$ scales linearly with the underlying n. Therefore, each query would only rule out the range of n up to a $\mathrm{poly}(d)$ factor, and there are

$$\log_{\mathrm{poly}(d)}(2^{\Theta(d)}) = \Theta(d/\log d)$$

such ranges to check. Therefore, we need to query $\Theta(d/\log d)$ distinct smoothing levels to get one "not null" answer that actually helps sampling.

The real Gaussian gives $\sqrt{d}$. The analysis above fails in one place: a Gaussian is *not* uniform even on its thin shell. With $1 - 1/\mathrm{poly}(d)$ probability,

$$\|Z\|_2^2 = \sigma^2 d \pm \widetilde{O}(\sigma^2\sqrt{d}),$$

so the posterior is influenced by a whole band of squared distances of width $\Theta(\sigma^2\sqrt{d})$, and the mass outside is negligible.

Pick two squared radii $r_\pm^2 = \sigma^2 d \pm \Theta(\sigma^2\sqrt{d})$ within this typical band. The per-center likelihood weights satisfy

$$\frac{\exp(-r_-^2/(2\sigma^2))}{\exp(-r_+^2/(2\sigma^2))} = \exp\left(\Theta\left(\frac{r_+^2 - r_-^2}{\sigma^2}\right)\right) = \exp(\Theta(\sqrt{d})).$$

At the same time, the size of the corresponding Hamming shell at radius r_+ versus r_- differs by $\exp(\Theta(\sqrt{d}))$ as well.

When S is a random subset of size n, this compensation creates a gap between the two thresholds in the counting story: as n grows, one first exits the "empty" regime by hitting the high-surface-area part of the band, but only after an additional $\exp(\widetilde{\Theta}(\sqrt{d}))$ factor in n does one reach the true "averaging" regime, where even the low-surface-area (but high-weight) part is well populated. Equivalently, for fixed σ the informative window can pin down n only up to an $\exp(\widetilde{\Theta}(\sqrt{d}))$ multiplicative factor.

Therefore we must scan

$$\log_{\exp(\widetilde{\Theta}(\sqrt{d}))}\big(\exp(\Theta(d))\big) = \widetilde{\Theta}(\sqrt{d})$$

distinct smoothing levels to reliably hit an informative scale, yielding the $\widetilde{\Omega}(\sqrt{d})$ lower bound.

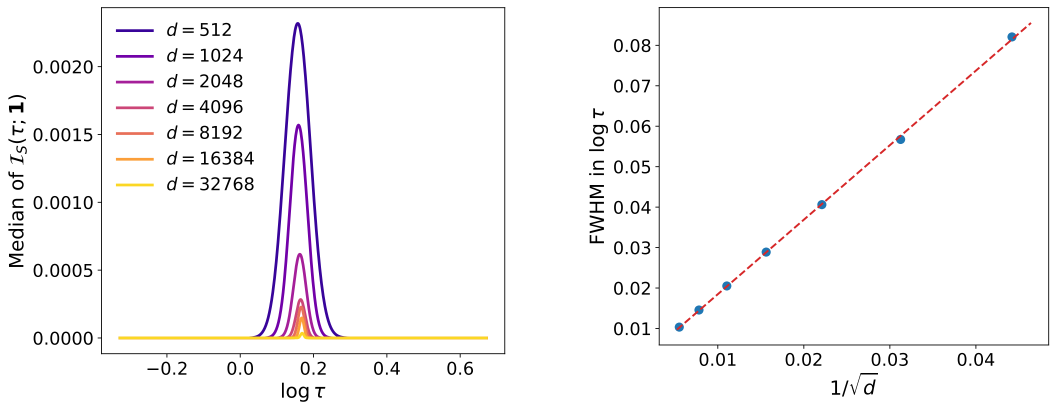
We note that our actual construction replaces the binary hypercube with a higher-resolution product constellation. This enables the $\log(R/\gamma)$ dependence in the parameterized theorem. See Appendix A for formal description of the hard instances.

Synthetic illustration. Figure 1 visualizes the informative-window phenomenon on the hypercube warm-up family described above. For each dimension d, we draw a random codebook S from a Poissonized ensemble on $\{-1, +1\}^d$ in which each vertex is included independently with probability $e^{\rho d - d\log 2}$ (so the expected codebook size is $e^{\rho d}$), and consider the local signal at the fixed query point $x = \mathbf{1}$:

$$\mathcal{I}_S(\tau;\, \mathbf{1}) \;:=\; \frac{\pi_\tau(\mathbf{1})}{\pi_\tau^{\mathrm{null}}(\mathbf{1})} \;\cdot\; \frac{1}{d}\left\|s_{\pi,\tau}(\mathbf{1}) - s_\tau^{\mathrm{null}}(\mathbf{1})\right\|_2^2,$$

where π_τ is the smoothed density of the mixture defined by S and π_τ^{null} is the smoothed null density (corresponding to $S = \{-1, +1\}^d$), with $s_{\pi,\tau}$ and s_τ^{null} their respective scores. This is the pointwise contribution at $x = \mathbf{1}$ to the relative Fisher divergence between planted and null distributions at noise level τ. When $\mathcal{I}_S(\tau;\, \mathbf{1})$ is negligible, a query at noise level τ reveals no information about S at this point.

The left panel plots the median of $\mathcal{I}_S(\tau;\, \mathbf{1})$ over randomly drawn instances as a function of $\log\tau$ for several values of d. The signal concentrates around a single matching scale and becomes progressively narrower as d grows. The right panel confirms that the full width at half maximum scales as $1/\sqrt{d}$, consistent with the $d^{-1/2}$ prediction from the shell fluctuation argument above. Equivalently, the informative noise levels occupy a window of width $\Theta(d^{-1/2})$ in $\log\tau$, so covering an $O(1)$ range of candidate smoothing levels requires $\Theta(\sqrt{d})$ queries.



(a) Median of $\mathcal{I}_S(\tau;\, \mathbf{1})$ for $d = 512$ to 32768.

(b) FWHM versus $1/\sqrt{d}$.

Figure 1: **Informative window on the hypercube warm-up family** ($\rho = 0.2$). The signal concentrates near a single smoothing scale and narrows as d grows; the width scales as $1/\sqrt{d}$, confirming the predicted $d^{-1/2}$ rate.

3 Related Works

While a rapidly growing literature studies the iteration complexity of diffusion sampling under various assumptions Chen et al. (2023c); Bortoli (2022); Conforti et al. (2023); Lee et al. (2022); Chen et al. (2023b); Lee et al. (2023); Chen et al. (2023a;d), lower bounds remain much less developed than convergence analyses. We discuss two directions that are closest in spirit: iteration lower bounds for specific diffusion algorithms and query lower bounds in the classical sampling literature.

3.1 Iteration Lower Bounds for Diffusion Models

Existing algorithmic lower bounds typically fix a discretization or a noise schedule. Jiao et al. (2025) show that under a standard DDPM discretization schedule, even with exact scores and a Gaussian target, the KL error of the terminal marginal satisfies $\mathrm{KL} \gtrsim d/T^2$, which forces $T = \Omega(\sqrt{d}/\varepsilon)$ steps to achieve $\mathrm{KL} \leq \varepsilon^2$. Related Gaussian lower bounds in Wasserstein metrics appear in Gao & Zhu (2025). These results identify barriers for specific solvers, but their analyses are tied to particular discretizations of specific stochastic processes and therefore cannot be generalized to the algorithm-agnostic lower bounds we seek.

It is also worth mentioning a route toward dimension-dependent query lower bounds for small KL error via tensorization. Let p be a one-dimensional distribution, and consider sampling from its product $p^{\otimes d}$. Since each dimension is independent, the best an algorithm can do is to sample d copies of p and output a sample following $\hat{p}^{\otimes d}$. Since KL is additive:

$$D_{\mathrm{KL}}(p^{\otimes d} \| q^{\otimes d}) = d \cdot D_{\mathrm{KL}}(p \| q).$$

In this setting, achieving ε-KL accuracy on $p^{\otimes d}$ would require achieving (ε/d)-KL accuracy on p (Benton et al., 2024).

This suggests that any query lower bound that in terms of $(1/\epsilon)$ in one dimension could be converted into a dimension-dependent lower bound even for product targets. For instance, if achieving ε-KL accuracy on $\mathbb{R}$ required $\Omega(1/\varepsilon^{0.1})$ score queries, then achieving ε-KL on $\mathbb{R}^d$ for a d-fold product target would require $\Omega((d/\varepsilon)^{0.1})$ queries. However, this approach relies on query lower bounds for small-ε KL, and to our knowledge no such bound is currently known.

3.2 Query Lower Bounds for Sampling

A classical oracle model writes the target as $\pi(x) \propto e^{-V(x)}$ and grants pointwise access to $V(x)$ and possibly its derivatives (Chewi et al., 2022). With first-order access, querying $\nabla V(x)$ is equivalent to querying the exact unsmoothed score

$$s_0(x) := \nabla \log \pi(x) = -\nabla V(x).$$

In the strongly log-concave and smooth setting, Chewi et al. (2024) prove dimension-dependent lower bounds that already appear for Gaussian targets. In particular, for a centered Gaussian with condition number κ, any algorithm with first-order oracle access needs

$$\widetilde{\Omega}\big(\min\{\sqrt{\kappa}\log d,\ d\}\big)$$

queries to achieve constant accuracy, matching block-Krylov type methods up to logarithmic factors. Sharp lower bounds are also known on one dimension (Chewi et al., 2022).

For general non-log-concavity, it is known that $\exp(\Omega(d))$ queries are needed. A typical construction hides probability mass in one of exponentially many well-separated regions of $\mathbb{R}^d$, so that local values and gradients are essentially indistinguishable unless the algorithm queries near the right region (Lee et al., 2018; He & Zhang, 2025; Chewi et al., 2023).

Diffusion-style access changes the picture because the oracle is smoothed. Querying $s_{\pi,\sigma} = \nabla \log(\pi * \mathcal{N}(0, \sigma^2 I))$ aggregates information from an $\mathcal{O}(\sigma)$ neighborhood. Large σ can reveal coarse global structure and rule out many candidate regions at once, while small σ refines local details. Therefore, our lower bounds require a different approach, and we show that the bottleneck is identifying an informative smoothing scale σ.

3.3 Applicability to Practical Algorithms

Our oracle model is intended to capture the inference-time information available to a broad class of practical diffusion and diffusion-adjacent samplers: repeated evaluations of a learned network at user-chosen states and noise levels, without any additional access to the target distribution π.

Common network parameterizations are score-equivalent. In standard diffusion setups, a network may be parameterized as a denoiser/x_0-predictor, a noise predictor, or directly as a score/drift model. All of these are information-equivalent to a smoothed-score oracle, up to known, noise-dependent affine transformations. Concretely, under additive Gaussian corruption, the denoiser (posterior mean) $D_\sigma(x)$ and the noise predictor $\varepsilon_\sigma(x)$ satisfy the standard identities

$$s_{\pi,\sigma}(x) \;=\; \frac{D_\sigma(x) - x}{\sigma^2}, \qquad \varepsilon_\sigma(x) \;=\; -\sigma\, s_{\pi,\sigma}(x),$$

and hence any learned approximation of D_σ or ε_σ can be converted into a score estimate $\widehat{s}_\sigma$ by the same formulas. These conversions are routinely used in diffusion models such as DDPM/DDIM and score-SDE formulations (up to conventional rescalings that depend only on the noise/time parameterization) (Ho et al., 2020; Song et al., 2021b; Karras et al., 2022; Robbins, 1956; Efron, 2011; Song et al., 2021a). Therefore, up to constant factors, each neural network evaluation can be counted as one oracle query. (Methods that use higher-order updates or corrector steps may spend multiple network evaluations per macro-step, but the relevant complexity measure for our results is the *total* number of evaluations.)

Consequently, Theorems 1.3 and 1.5 can be interpreted as lower bounds on the number of neural network evaluations required by any inference procedure whose only access to π is through repeated evaluation of such score-equivalent networks, regardless of how the updates are organized or whether the method is deterministic or randomized.

SDE/ODE-based diffusion samplers. This scope includes both classic reverse-time SDE discretizations and ODE-based solvers. Concretely, it covers DDPM and its variants, predictor–corrector samplers, the probability flow ODE, and generic integrators such as DPM-Solver, PNDM, DEIS, UniPC, and EDM-style solvers (Song et al., 2021b; Lu et al., 2022; 2025; Liu et al., 2022a; Zhang & Chen, 2023; Zhao et al., 2023; Karras et al., 2022). These methods differ in discretization order, step-size selection, and whether the path is formulated as an SDE or an ODE, but they all obtain information about π through the same primitive: evaluating a learned score (or an equivalent drift derived from it) at adaptively chosen states and noise levels.

Flow matching and rectified flow in the linear Gaussian interpolation setting. Our oracle model also captures the standard rectified-flow / flow-matching setup with a Gaussian prior and linear interpolation. Specifically, let $X_1 \sim \pi$ and $X_0 \sim \mathcal{N}(0, I_d)$ be independent and define

$$X_t = tX_1 + (1-t)X_0, \qquad t \in (0,1).$$

Rectified flow defines the velocity field

$$v_t(x) := \mathbb{E}[X_1 - X_0 \mid X_t = x].$$

A direct calculation using Tweedie's formula (Robbins, 1956; Efron, 2011) shows that v_t is an explicit affine transform of the Gaussian-smoothed score:

$$v_t(x) = \frac{x}{t} + \frac{1-t}{t^2}\, s_{\pi,\sigma_t}(x/t), \qquad \sigma_t = \frac{1-t}{t}.$$

Equivalently, for each fixed t, querying $v_t(x)$ is invertibly equivalent to querying s_{π,σ_t} at the rescaled input x/t. Therefore, any sampler whose inference-time access to π is only through evaluations of such velocity fields inherits the same query lower bounds (Lipman et al., 2023; Liu et al., 2022b; Benton et al., 2023).

What is not covered. Our lower bounds apply to any algorithm whose only source of information about π at inference time is score-equivalent evaluations at adaptively chosen states and noise levels, regardless of what computation the algorithm performs between or after queries. In contrast, approaches such as progressive distillation (Salimans & Ho, 2022) or consistency models (Song et al., 2023) may fall outside this model: their training procedures can encode additional global information about π into a new mapping, so that each inference-time evaluation is no longer a score-equivalent query. For hybrid systems that combine a partially distilled backbone with residual score-correction steps, applicability depends on whether the correction steps are the sole source of information about π at inference time; if so, the lower bound applies to the number of such correction queries.

4 CONCLUSION

In this work, we establish the first information-theoretic lower bounds for diffusion sampling. We prove that for d-dimensional distributions, acquiring a non-trivial sample using L^p-accurate score estimates requires $\widetilde{\Omega}(\sqrt{d})$ adaptive score queries, even under favorable assumptions of smoothness and polynomial estimation accuracy. These results characterize the computational cost's dependence on dimension as a fundamental barrier intrinsic to the score-based sampling paradigm, rather than a limitation of current solvers.

At a technical level, our analysis reveals that multiscale scanning is a worst-case necessity driven by high-dimensional concentration. Because the score signal is localized within specific noise scales and forms a narrow "informative window", algorithms that aggressively reduce step counts risk missing the signal entirely. This perspective complements discretization-based analyses by identifying information loss, rather than discretization error, as the primary bottleneck in the aggressive acceleration regime.

Our work highlights several open frontiers for future research:

1. **Tightening the Bound.** There remains a gap between our $\widetilde{\Omega}(\sqrt{d})$ lower bound and the best known $\widetilde{O}(d)$ upper bounds. Determining the precise query complexity is a major open question that may require constructing sharper hard instances.
2. **Beyond the Smoothed-Score Oracle.** Our lower bounds apply when inference-time access to the target is only through repeated evaluations of a Gaussian-smoothed score, or an explicitly equivalent field as in rectified flow. A natural next step is to extend lower bounds to paradigms that change this primitive, including flow matching with non-Gaussian priors or alternative couplings, and distilled generators that produce samples in very few steps.
3. **Real-World Structure.** The empirical success of few-step sampling suggests that real-world data possess structures stronger than generic smoothness. Identifying specific geometric properties, such as low intrinsic dimension, that enable practical algorithms to break the worst-case $\sqrt{d}$ barrier is a promising direction for bridging theory and practice.

REFERENCES

Joe Benton, George Deligiannidis, and Arnaud Doucet. Error bounds for flow matching methods. *arXiv preprint arXiv:2305.16860*, 2023.

Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *International Conference on Learning Representations*, 2024. arXiv:2308.03686.

Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL `https://openreview.net/forum?id=MhK5aXo3gB`. Expert Certification.

Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *International Conference on Machine Learning*, pp. 4735–4763. PMLR, 2023a.

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ode is provably fast. *Advances in Neural Information Processing Systems*, 36:68552–68575, 2023b.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. In *International Conference on Learning Representations*, 2023c. arXiv:2209.11215.

Sitan Chen, Giannis Daras, and Alex Dimakis. Restoration-degradation beyond linear diffusions: A non-asymptotic analysis for ddim-type samplers. In *International Conference on Machine Learning*, pp. 4462–4484. PMLR, 2023d.

Sinho Chewi, Patrik Gerber, Chen Lu, Thibaut Le Gouic, and Philippe Rigollet. The query complexity of sampling from strongly log-concave distributions in one dimension. In *Conference on Learning Theory*, 2022. arXiv:2105.14163.

Sinho Chewi, Patrik Gerber, Holden Lee, and Chen Lu. Fisher information lower bounds for sampling. In *Algorithmic Learning Theory*, 2023. arXiv:2210.02482.

Sinho Chewi, Jaume de Dios Pont, Jerry Li, Chen Lu, and Shyam Narayanan. Query lower bounds for log-concave sampling. *Journal of the ACM*, 71(4):29:1–29:42, 2024. doi: 10.1145/3673651. URL `https://doi.org/10.1145/3673651`. Preliminary version in FOCS 2023; arXiv:2304.02599.

Giovanni Conforti, Alain Durmus, and Marta Gentiloni Silveri. Kl convergence guarantees for score diffusion models under minimal data assumptions. *arXiv preprint arXiv:2308.12240*, 2023.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. doi: 10.1198/jasa.2011.tm11181.

Kenji Fukumizu, Taiji Suzuki, Noboru Isobe, Kazusato Oko, and Masanori Koyama. Flow matching achieves almost minimax optimal convergence. In *The Thirteenth International Conference on Learning Representations*, 2025. URL `https://openreview.net/forum?id=2OMyAFjiJJ`.

Xuefeng Gao and Lingjiong Zhu. Convergence analysis for general probability flow ODEs of diffusion models in wasserstein distances. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 258 of *Proceedings of Machine Learning Research*, pp. 1009–1017, 2025. arXiv:2401.17958.

Khashayar Gatmiry, Sitan Chen, and Adil Salim. High-accuracy and dimension-free sampling with diffusions. *arXiv preprint arXiv:2601.10708*, 2026.

Shivam Gupta, Aditya Parulekar, Eric Price, and Zhiyang Xun. Improved sample complexity bounds for diffusion model training. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9798331314385.

Shivam Gupta, Linda Cai, and Sitan Chen. Faster diffusion sampling with randomized midpoints: Sequential and parallel. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/f30307ac840b88f86f4ab5761b2d6595-Abstract-Conference.html. arXiv:2406.00924.

Yuchen He and Chihao Zhang. On the query complexity of sampling from non-log-concave distributions (extended abstract). In Nika Haghtalab and Ankur Moitra (eds.), *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 2786–2787. PMLR, 2025. URL https://proceedings.mlr.press/v291/he25a.html. Full version: arXiv:2502.06200.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.

Yuchen Jiao and Gen Li. Instance-dependent convergence theory for diffusion models. *arXiv preprint arXiv:2410.13738*, 2024.

Yuchen Jiao, Yuchen Zhou, and Gen Li. Optimal convergence analysis of DDPM for general distributions. *arXiv preprint arXiv:2510.27562*, 2025.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, 2022.

Holden Lee, Andrej Risteski, and Rong Ge. Beyond log-concavity: Provable guarantees for sampling multi-modal distributions using simulated tempering langevin monte carlo. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/c6ede20e6f597abf4b3f6bb30cee16c7-Paper.pdf.

Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence for score-based generative modeling with polynomial complexity. In *Advances in Neural Information Processing Systems*, 2022. arXiv:2206.06227.

Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *International Conference on Algorithmic Learning Theory*, pp. 946–985. PMLR, 2023.

Gen Li and Yuchen Jiao. Improved convergence rate for diffusion probabilistic models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=SOd07Qxkw4.

Gen Li and Yuling Yan. O(d/t) convergence theory for diffusion probabilistic models under minimal assumptions. *arXiv preprint arXiv:2409.18959*, 2024.

Gen Li, Yu Huang, Timofey Efimov, Yuting Wei, Yuejie Chi, and Yuxin Chen. Accelerating convergence of score-based diffusion models, provably. In *Proceedings of the 41st International Conference on Machine Learning*, 2024a. arXiv:2403.03852.

Gen Li, Yuting Wei, Yuejie Chi, and Yuxin Chen. A sharp convergence theory for the probability flow odes of diffusion models. *arXiv preprint arXiv:2408.02320*, 2024b.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t. arXiv:2210.02747.

Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. In *International Conference on Learning Representations*, 2022a. URL https://openreview.net/forum?id=PlKWVd2yBkY. arXiv:2202.09778.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022b.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: a fast ode solver for diffusion probabilistic model sampling in around 10 steps. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, 22:730–751, 2025. doi: 10.1007/s11633-025-1562-4. URL `https://doi.org/10.1007/s11633-025-1562-4`. arXiv:2211.01095.

Alex Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, 2021.

Herbert Robbins. An empirical bayes approach to statistics. In Jerzy Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pp. 157–163, Berkeley, CA, 1956. University of California Press.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022. arXiv:2202.00512.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL `https://openreview.net/forum?id=St1giarCHLP`.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b. arXiv:2011.13456.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252, 2023. URL `https://proceedings.mlr.press/v202/song23a.html`. arXiv:2303.01469.

Rong Tang and Yun Yang. Adaptivity of diffusion models to manifold structures. In Sanjoy Dasgupta, Stephan Mandt, and Yingzhen Li (eds.), *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1648–1656. PMLR, 02–04 May 2024. URL `https://proceedings.mlr.press/v238/tang24a.html`.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Yuchen Wu, Yuxin Chen, and Yuting Wei. Stochastic runge–kutta methods: Provable acceleration of diffusion models. *arXiv preprint arXiv:2410.04760*, 2024.

Zhiyang Xun, Shivam Gupta, and Eric Price. Posterior sampling by combining diffusion models with annealed langevin dynamics. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL `https://openreview.net/forum?id=ARZiMmb619`.

Matthew S. Zhang, Stephen Huan, Jerry Huang, Nicholas M. Boffi, Sitan Chen, and Sinho Chewi. Sublinear iterations can suffice even for ddpms. *arXiv preprint arXiv:2511.04844*, 2025.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. In *International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=Loek7hfb46P`. arXiv:2204.13902.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. In *Advances in Neural Information Processing Systems*, 2023. arXiv:2302.04867.

APPENDIX

A FORMAL CONSTRUCTION

We begin by stating a parameterized version of Theorem 1.3 to make the dependence on (R, γ) and the oracle parameters explicit.

Theorem A.1. *Fix constants $p > 0$ and $\rho \in (0, 1/4)$. Then there exists a constant $c = c(p, \rho) > 0$ such that the following holds for every exponent $A \geq 0$, every $R > 0$, every $\gamma \in (0, R/2)$, every sufficiently large dimension d, and every $\varepsilon_{\mathrm{err}} \in [d^{-A}, 1]$. There exists a distribution $\mathcal{D}$ over pairs $(\pi, \{\widehat{s}_\sigma\}_{\sigma>0})$ on $\mathbb{R}^d$ such that:*

(1) ***(bounded-plus-noise)*** *$\pi = \pi_{\mathrm{pre}} * \mathcal{N}(0, \gamma^2 I_d)$ for some π_{pre} supported on $[-R, R]^d$;*

(2) ***(L^p-accurate score oracle)*** *for every $\sigma > 0$,*

$$\mathbb{E}_{X\sim\pi_\sigma}\left[\|\widehat{s}_\sigma(X) - s_{\pi,\sigma}(X)\|_2^p\right] \leq \frac{\varepsilon_{\mathrm{err}}^p}{\sigma^p};$$

(3) ***(global Lipschitzness)*** *for every $\sigma > 0$, both $s_{\pi,\sigma}$ and $\widehat{s}_\sigma$ are globally L_σ-Lipschitz with*

$$L_\sigma \leq \frac{3}{\gamma^2 + \sigma^2} + \frac{7R^2 d}{(\gamma^2 + \sigma^2)^2};$$

(4) ***(query lower bound)*** *every adaptive algorithm making at most*

$$Q \leq c \cdot \min\left\{d,\ \frac{d\log(R/\gamma)}{\sqrt{dH_p} + H_p}\right\}, \qquad H_p := \log(d/\varepsilon_{\mathrm{err}}) + \log(R/\gamma),$$

score queries and outputting $\widehat{X}$ satisfies

$$\mathbb{P}_{(\pi,\{\widehat{s}_\sigma\})\sim\mathcal{D}}\Big[\mathrm{d_{TV}}\left(\mathcal{L}(\widehat{X}), \pi\right) \geq 1-\rho\Big] \geq 1-\rho.$$

Now we formally introduce the hard family and notation used throughout the proofs. We then prove Theorem A.1 and its ψ_1-counterpart in parameterized form; Theorem 1.5 follows by specialization.

Fix $R > 0$ and $\gamma \in (0, R/2)$, and assume throughout this appendix that d is even, writing $d = 2m$. Let

$$M := \left\lceil \pi\frac{R}{\gamma} \right\rceil,$$

and define the 2D constellation

$$a_j := R\big(\cos(2\pi j/M),\, \sin(2\pi j/M)\big) \in \mathbb{R}^2, \qquad j \in \{0, 1, \dots, M-1\}.$$

Write

$$\mathcal{A} := \{a_0, \dots, a_{M-1}\} \subset \mathbb{R}^2, \qquad \mathcal{V} := \mathcal{A}^m \subset \mathbb{R}^d, \qquad U := \mathrm{Unif}(\mathcal{V}).$$

Thus

$$|\mathcal{V}| = M^{d/2}.$$

If $S = (Y^{(1)}, \dots, Y^{(n)})$ is a codebook of size n sampled from U with replacement, let

$$\nu_S := \frac{1}{n}\sum_{i=1}^n \delta_{Y^{(i)}}$$

denote its empirical measure. Duplicates are allowed, and only the empirical measure matters.

For every total noise level $\tau > 0$, define

$$\nu_{S,\tau} := \nu_S * \mathcal{N}(0, \tau^2 I_d), \qquad \nu_{U,\tau} := U * \mathcal{N}(0, \tau^2 I_d).$$

When needed, we write

$$s_{S,\tau} := \nabla \log \nu_{S,\tau}, \qquad s_{U,\tau} := \nabla \log \nu_{U,\tau}$$

for the corresponding score fields. In particular,

$$\pi_{S,\gamma} := \nu_{S,\gamma}, \qquad \pi_{U,\gamma} := \nu_{U,\gamma}.$$

We also write

$$\tau(\sigma) := \sqrt{\gamma^2 + \sigma^2}, \qquad \kappa(n) := \frac{1}{d} \log n.$$

The quantity $\kappa(n)$ is the rate parameter used throughout the appendix. Its natural ambient scale is $\frac{1}{2} \log M$, since $|\mathcal{V}| = M^{d/2}$.

Throughout, we interpret $[a, b] = \emptyset$ when $a > b$, and we write $|[a, b]| := (b - a)_+$ for its length.

We first establish a lower bound for a total-noise oracle indexed by $\tau \geq \gamma$. The fixed-noise theorem reduces this step to agreement with the null score outside a rate interval, a packing of the admissible rate set, and separation at the base noise level γ. The small-sample and large-sample regimes determine the two endpoints of this interval, and the analytic estimates below are used only in the large-sample regime. We then specialize the resulting fixed-noise lower bound to the L^p and ψ_1 settings and transfer it back to the original σ-indexed model via $\tau(\sigma) = \sqrt{\gamma^2 + \sigma^2}$.

B A FIXED-NOISE LOWER BOUND IN RATE SPACE

We begin with a lower bound for a total-noise oracle indexed by $\tau \geq \gamma$.

B.1 THE FIXED-NOISE THEOREM

Fix a query budget $Q \geq 1$, a target error level $\rho \in (0, 1/4)$, and set

$$\delta := \frac{\rho^2}{80Q}.$$

For each codebook S and total-noise level $\tau \geq \gamma$, let $\widehat{s}_\tau^{(S)}$ denote the oracle response at total noise τ, and let $s_{U,\tau}$ denote the corresponding null score. Unless stated otherwise, whenever $\mathbb{P}_S$ or $\mathbb{E}_S$ appears below, the codebook S is sampled i.i.d. from U with the value of n prescribed there.

Theorem B.1 (A fixed-noise lower bound in rate space)**.** *Let $1 \leq n_{\min} \leq n_{\max}$ be integers, and define the discrete admissible rate set*

$$K_d := \{\kappa(n) : \ n \in [n_{\min}, n_{\max}] \cap \mathbb{N}\}.$$

Assume that for every $\tau \geq \gamma$ we are given an interval

$$\mathcal{J}(\tau) \subseteq \mathbb{R},$$

and let

$$w := \sup_{\tau \geq \gamma} |\mathcal{J}(\tau)|,$$

with the convention that $|\mathcal{J}(\tau)| = 0$ when $\mathcal{J}(\tau) = \emptyset$. Suppose that the following hold.

(i) ***(Agreement with the null score outside the interval)*** *For every fixed query point $x \in \mathbb{R}^d$, every $\tau \geq \gamma$, and every integer $n \in [n_{\min}, n_{\max}]$,*

$$\kappa(n) \notin \mathcal{J}(\tau) \quad \Longrightarrow \quad \mathbb{P}_S\big[\widehat{s}_\tau^{(S)}(x) \neq s_{U,\tau}(x)\big] \leq \delta,$$

where S is sampled i.i.d. from U with size n.

(ii) ***(Packing on the rate axis)*** *The set K_d contains a subset G of cardinality*

$$|G| \geq \frac{80Q}{\rho^2}$$

whose points are pairwise separated by more than w.

(iii) ***(Separation at the base noise level)*** *There exists a measurable map $S \mapsto A(S) \subseteq \mathbb{R}^d$ such that*

$$\pi_{S,\gamma}(A(S)) \geq 1 - \rho/2 \qquad \text{for every codebook } S \text{ of size } n \in [n_{\min}, n_{\max}] \cap \mathbb{N},$$

and, if J is uniform on G, if $n(J)$ denotes the unique integer with $\kappa(n(J)) = J$, and if S_J is an i.i.d. codebook of size $n(J)$ drawn from U, then

$$\sup_{x \in \mathbb{R}^d} \mathbb{P}_{J,S_J}\big[x \in A(S_J)\big] \leq \frac{\rho^2}{8}.$$

Let $\mathcal{D}$ denote the law of $(n(J), S_J)$. Then every adaptive algorithm $\mathcal{A}$ making at most Q score queries, with output distribution $Q^{\mathcal{A}}_{n,S}$ on instance (n, S), satisfies

$$\mathbb{P}_{(n,S)\sim\mathcal{D}}\Big[\, \mathrm{d_{TV}}\left(Q^{\mathcal{A}}_{n,S}, \pi_{S,\gamma}\right) \geq 1 - \rho\Big] \geq 1 - \rho.$$

Proof. Fix an arbitrary adaptive algorithm $\mathcal{A}$ making at most Q score queries. Let $Q^{\mathcal{A}}_{n,S}$ be its output distribution on the hard instance (n, S), and let $Q^{\mathcal{A}}_0$ be its output distribution when the oracle is replaced by the null family $\{s_{U,\tau}\}_{\tau\geq\gamma}$.

Condition on the internal randomness ω of $\mathcal{A}$ in the null run. The resulting null transcript determines a deterministic sequence of queries

$$\left(\tau^0_t, x^0_t\right)_{t=1}^{Q}.$$

Let

$$E := \Big\{\kappa(n(J)) \notin \bigcup_{t=1}^{Q} \mathcal{J}(\tau^0_t)\Big\}.$$

Because the points of G are pairwise separated by more than w while each interval $\mathcal{J}(\tau^0_t)$ has length at most w, every queried interval contains at most one point of G. Hence, conditional on ω,

$$\mathbb{P}_J(E^c \mid \omega) \leq \frac{Q}{|G|} \leq \frac{\rho^2}{80}.$$

Averaging over ω yields

$$\mathbb{P}_{J,\omega}(E^c) \leq \frac{\rho^2}{80}.$$

We now couple the run of $\mathcal{A}$ on $(n(J), S_J)$ with its null run by using the same internal randomness ω. Let Diff be the event that the two transcripts ever differ, and let

$$T := \min\{t \leq Q : \text{ the two transcripts differ at time } t\},$$

with the convention $T = \infty$ when they never differ. On the event $\{T = t\}$ the two runs agree up to time $t - 1$, hence issue the same tth query (τ^0_t, x^0_t). Therefore

$$\{T = t\} \subseteq \big\{\widehat{s}^{(S_J)}_{\tau^0_t}(x^0_t) \neq s_{U,\tau^0_t}(x^0_t)\big\}.$$

If E holds, then $\kappa(n(J)) \notin \mathcal{J}(\tau^0_t)$ for every $t \leq Q$, so assumption (i) gives

$$\mathbb{P}_{J,S_J,\omega}(T = t \mid E) \leq \delta \qquad \text{for every } t \leq Q.$$

Summing over t yields

$$\mathbb{P}_{J,S_J,\omega}(\mathsf{Diff} \mid E) \leq Q\delta = \frac{\rho^2}{80}.$$

Consequently

$$\mathbb{P}_{J,S_J,\omega}(\mathsf{Diff}) \leq \mathbb{P}_{J,\omega}(E^c) + \mathbb{P}_{J,S_J,\omega}(\mathsf{Diff} \mid E) \leq \frac{\rho^2}{40}.$$

Outside Diff the coupled outputs coincide, and therefore

$$\mathbb{E}_{(n,S)\sim\mathcal{D}}\big[\,\mathrm{d_{TV}}(Q^{\mathcal{A}}_{n,S}, Q^{\mathcal{A}}_0)\big] \leq \mathbb{P}_{J,S_J,\omega}(\mathsf{Diff}) \leq \frac{\rho^2}{40}.$$

By Markov's inequality,

$$\mathbb{P}_{(n,S)\sim\mathcal{D}}\big[\,\mathrm{d}_{\mathrm{TV}}(Q^{\mathcal{A}}_{n,S}, Q^{\mathcal{A}}_0) \geq \rho/4\big] \leq \frac{\rho}{10}.$$

Let $X_0 \sim Q^{\mathcal{A}}_0$ be independent of (J, S_J). Fubini's theorem and assumption (iii) give

$$\mathbb{E}_{(n,S)\sim\mathcal{D}}\big[Q^{\mathcal{A}}_0(A(S))\big] = \mathbb{E}_{X_0\sim Q^{\mathcal{A}}_0}\Big[\mathbb{P}_{J,S_J}\big[X_0 \in A(S_J)\big]\Big] \leq \frac{\rho^2}{8}.$$

A second application of Markov's inequality yields

$$\mathbb{P}_{(n,S)\sim\mathcal{D}}\big[Q^{\mathcal{A}}_0(A(S)) \geq \rho/4\big] \leq \frac{\rho}{2}.$$

On the complement of the union of the last two events we have

$$Q^{\mathcal{A}}_{n,S}(A(S)) \leq Q^{\mathcal{A}}_0(A(S)) + \mathrm{d}_{\mathrm{TV}}(Q^{\mathcal{A}}_{n,S}, Q^{\mathcal{A}}_0) \leq \frac{\rho}{2},$$

whereas assumption (iii) gives $\pi_{S,\gamma}(A(S)) \geq 1 - \rho/2$. Therefore

$$\mathrm{d}_{\mathrm{TV}}(Q^{\mathcal{A}}_{n,S}, \pi_{S,\gamma}) \geq \pi_{S,\gamma}(A(S)) - Q^{\mathcal{A}}_{n,S}(A(S)) \geq 1 - \rho.$$

Hence

$$\mathbb{P}_{(n,S)\sim\mathcal{D}}\Big[\,\mathrm{d}_{\mathrm{TV}}(Q^{\mathcal{A}}_{n,S}, \pi_{S,\gamma}) < 1 - \rho\Big] \leq \frac{\rho}{10} + \frac{\rho}{2} < \rho.$$

Since $\mathcal{A}$ was arbitrary, the theorem follows. □

B.2 Oracle construction and the two threshold bounds

We now specialize the fixed-noise argument to the product-circle hard family from Section A. We write the argument for even d, so that the reference law U factorizes into $d/2$ identical planar blocks. The odd-dimensional case follows from the standard one-coordinate padding/projection reduction. The small-sample and large-sample regimes below yield the two threshold bounds used in Theorem B.1.

For $\tau > 0$ let

$$\nu_{U,\tau} := U * \mathcal{N}(0, \tau^2 I_d), \qquad \nu_{S,\tau} := \nu_S * \mathcal{N}(0, \tau^2 I_d),$$

and let u_τ denote the density of $\nu_{U,\tau}$, and let φ_{τ^2} denote the density of $\mathcal{N}(0, \tau^2 I_d)$. We also write

$$s_{S,\tau} := s_{\nu_{S,\tau}}, \qquad s_{U,\tau} := s_{\nu_{U,\tau}}.$$

For $y \in \mathcal{V}$ and $x \in \mathbb{R}^d$ define

$$L_\tau(y, x) := \frac{\varphi_{\tau^2}(x - y)}{u_\tau(x)}, \qquad \ell_\tau(y, x) := \log L_\tau(y, x).$$

If $Y \sim U$ and $X = Y + Z$ with $Z \sim \mathcal{N}(0, \tau^2 I_d)$ independent, set

$$\ell_\tau := \ell_\tau(Y, X), \qquad I_\tau := \frac{1}{d}\,\mathbb{E}[\ell_\tau].$$

For a codebook S write

$$J_\tau(S) := \mathbb{E}_{X\sim\nu_{S,\tau}}\big[\|s_{S,\tau}(X) - s_{U,\tau}(X)\|_2^2\big].$$

For $\zeta \in (0, 1/2]$, define

$$\Lambda_\tau(\zeta) := \inf\Big\{\lambda > 0 :\ \mathbb{P}\big[\ell_\tau \leq \log\lambda\big] \geq \zeta\Big\}.$$

We will repeatedly use the elementary consequences

$$\mathbb{E}_{Y\sim U}[L_\tau(Y, x)] = 1, \qquad \mathbb{P}\big[\ell_\tau < \log\Lambda_\tau(\zeta)\big] \leq \zeta, \qquad \mathbb{P}\big[\ell_\tau \leq \log\Lambda_\tau(\zeta)\big] \geq \zeta.$$

Fix auxiliary functions

$$\zeta(\tau) \in (0, 1/2], \qquad \theta(\tau) > 0, \qquad \tau \geq \gamma,$$

and abbreviate $\Lambda_\tau := \Lambda_\tau(\zeta(\tau))$. Define

$$\ell^{\max}_{\tau,S}(x) := \max_{y\in S} \ell_\tau(y,x), \qquad G_\tau(S) := \{x \in \mathbb{R}^d : \ \ell^{\max}_{\tau,S}(x) \ge \log \Lambda_\tau\},$$

and let

$$m_{\tau,S}(x) := \psi\big(\ell^{\max}_{\tau,S}(x) - \log \Lambda_\tau + 1\big),$$

where $\psi : \mathbb{R} \to [0,1]$ is the standard 1-Lipschitz cutoff with $\psi(t) = 0$ for $t \le 0$ and $\psi(t) = 1$ for $t \ge 1$. Given such a choice, we define the oracle by

$$\widehat{s}^{(S)}_\tau(x) := \begin{cases} s_{U,\tau}(x), & J_\tau(S) \le \theta(\tau), \\ s_{U,\tau}(x) + m_{\tau,S}(x)\big(s_{S,\tau}(x) - s_{U,\tau}(x)\big), & J_\tau(S) > \theta(\tau). \end{cases}$$

By symmetry of the product-circle support and isotropy of the Gaussian noise, if S is any codebook supported on $\mathcal{V}$, $Y_S \sim \nu_S$, and $X = Y_S + Z$ with $Z \sim \mathcal{N}(0, \tau^2 I_d)$ independent, then $\ell_\tau(Y_S, X) \stackrel{d}{=} \ell_\tau$. Since

$$G_\tau(S)^c \subseteq \{\ell_\tau(Y_S, X) < \log \Lambda_\tau\},$$

the elementary quantile fact above gives the mass-coverage bound

$$\nu_{S,\tau}(G_\tau(S)) \ge 1 - \zeta(\tau)$$

for every codebook S. We will use this repeatedly.

Proposition B.2 (Regularity of the oracle construction)**.** *For every $\tau \ge \gamma$ and every codebook S,*

$$\mathrm{Lip}(\ell^{\max}_{\tau,S}) \le \frac{2R\sqrt{d}}{\tau^2}, \qquad \mathrm{Lip}(m_{\tau,S}) \le \frac{2R\sqrt{d}}{\tau^2}.$$

Consequently,

$$\mathrm{Lip}\left(\widehat{s}^{(S)}_\tau\right) \le \frac{3}{\tau^2} + \frac{7R^2 d}{\tau^4}.$$

Proposition B.3 (Small-sample agreement with the null score)**.** *Define*

$$n_-(\tau) := \delta e^{-1} \Lambda_\tau, \qquad \kappa_-(\tau) := \frac{1}{d} \log n_-(\tau).$$

Then for every fixed $\tau \ge \gamma$, every fixed query point $x \in \mathbb{R}^d$, and every integer n with

$$\kappa(n) \le \kappa_-(\tau),$$

one has

$$\mathbb{P}_S\big[\widehat{s}^{(S)}_\tau(x) \ne s_{U,\tau}(x)\big] \le \delta.$$

Proof. If $\kappa(n) \le \kappa_-(\tau)$, then

$$n \le e^{d\kappa_-(\tau)} = \delta e^{-1} \Lambda_\tau.$$

If $\widehat{s}^{(S)}_\tau(x) \ne s_{U,\tau}(x)$, then necessarily $m_{\tau,S}(x) > 0$, which implies

$$\ell^{\max}_{\tau,S}(x) > \log \Lambda_\tau - 1.$$

Equivalently, there exists $y \in S$ with

$$L_\tau(y, x) \ge e^{-1} \Lambda_\tau.$$

Hence, by a union bound,

$$\mathbb{P}_S\big[\widehat{s}^{(S)}_\tau(x) \ne s_{U,\tau}(x)\big] \le n\, \mathbb{P}_{Y\sim U}\big[L_\tau(Y, x) \ge e^{-1}\Lambda_\tau\big].$$

Using $\mathbb{E}_{Y\sim U}[L_\tau(Y,x)] = 1$ and Markov's inequality,

$$\mathbb{P}_{Y\sim U}\big[L_\tau(Y,x) \ge e^{-1}\Lambda_\tau\big] \le \frac{e}{\Lambda_\tau}.$$

Therefore

$$\mathbb{P}_S\big[\widehat{s}^{(S)}_\tau(x) \ne s_{U,\tau}(x)\big] \le \frac{en}{\Lambda_\tau} \le \delta,$$

as claimed. □

Theorem B.4 (Large-sample agreement with the null score). *Fix $\tau \geq \gamma$ and define*

$$H(\tau) := \max\left\{1,\ \log\left(\frac{\log(1+|\mathcal{V}|)}{\delta}\cdot\frac{CR^2d^2}{\gamma^2\tau^2\theta(\tau)}\right)\right\},$$

where $C > 0$ is a sufficiently large universal constant. Let $\alpha := 1 - \frac{\gamma^2}{R^2d^2}$ and $\widetilde{\tau} := \sqrt{\alpha}\,\tau$, and define

$$E_{\mathrm{med}}(\tau) := C\sqrt{dH(\tau)} + CH(\tau), \qquad E_{\mathrm{big}}(\tau) := C\frac{R}{\widetilde{\tau}}\sqrt{dH(\tau)} + CH(\tau).$$

Set

$$\kappa_+(\tau) := I_{\widetilde{\tau}} + \frac{1}{d}\min\{E_{\mathrm{med}}(\tau), E_{\mathrm{big}}(\tau)\}.$$

Then, for all sufficiently large d, every fixed query point $x \in \mathbb{R}^d$ and every integer n with

$$\kappa(n) \geq \kappa_+(\tau),$$

one has

$$\mathbb{P}_S\big[\widehat{s}_\tau^{(S)}(x) \neq s_{U,\tau}(x)\big] \leq \delta.$$

Theorem B.5 (Fixed-noise interval theorem). *There exists a universal constant $C > 0$ such that the following holds. Define*

$$\mathcal{J}(\tau) := [\kappa_-(\tau), \kappa_+(\tau)], \qquad H_{\mathrm{win}}(\tau) := \max\big\{\log(1/\zeta(\tau)),\ H(\tau),\ \log(1/\delta)\big\}.$$

Then, for all sufficiently large d and every $\tau \geq \gamma$,

$$|\mathcal{J}(\tau)| \leq C\sqrt{\frac{H_{\mathrm{win}}(\tau)}{d}} + C\frac{H_{\mathrm{win}}(\tau)}{d}.$$

Moreover, for every fixed query point $x \in \mathbb{R}^d$ and every integer $n \geq 1$,

$$\kappa(n) \notin \mathcal{J}(\tau) \quad\Longrightarrow\quad \mathbb{P}_S\big[\widehat{s}_\tau^{(S)}(x) \neq s_{U,\tau}(x)\big] \leq \delta.$$

B.3 Analytic input for the large-sample regime

The large-sample theorem is the only point in the fixed-noise argument that needs additional analytic input. The smoothing identities and KL/Fisher estimate control the Fisher term in Theorem B.4; the likelihood-ratio estimates are then converted into expected-KL bounds for random codebooks.

Gaussian smoothing and KL/Fisher control.

Proposition B.6 (Gaussian smoothing identities and consequences). *Let ν be any probability measure supported on $[-R,R]^d$, and let $\nu_\tau := \nu * \mathcal{N}(0,\tau^2 I_d)$ for $\tau > 0$. Then the following hold.*

(a) For every $x \in \mathbb{R}^d$,

$$s_{\nu,\tau}(x) = \tau^{-2}\big(\mathbb{E}[Y \mid Y+Z=x] - x\big),$$

where $Y \sim \nu$ and $Z \sim \mathcal{N}(0,\tau^2 I_d)$ are independent.

(b) The Jacobian of the posterior mean satisfies the standard Gaussian smoothing identity, and hence

$$\nabla s_{\nu,\tau}(x) = \tau^{-4}\operatorname{Cov}(Y \mid Y+Z=x) - \tau^{-2}I_d.$$

(c) If ν_1, ν_2 are both supported on $[-R,R]^d$, then for every $x \in \mathbb{R}^d$,

$$\|s_{\nu_1,\tau}(x) - s_{\nu_2,\tau}(x)\|_2 \leq \frac{2R\sqrt{d}}{\tau^2}.$$

(d) For every $x \in \mathbb{R}^d$,

$$\mathrm{Lip}(s_{\nu,\tau}) \leq \frac{1}{\tau^2} + \frac{R^2 d}{\tau^4}.$$

Proof. Let $Y \sim \nu$ and $Z \sim \mathcal{N}(0, \tau^2 I_d)$ be independent, and write

$$\nu_\tau(x) = \int \varphi_{\tau^2}(x-y)\,\nu(dy).$$

Proof of (a). Differentiating under the integral sign gives

$$\nabla \nu_\tau(x) = \frac{1}{\tau^2}\int (y-x)\varphi_{\tau^2}(x-y)\,\nu(dy).$$

Dividing by $\nu_\tau(x)$ and recognizing the posterior weights of Y given $Y+Z=x$ yields

$$s_{\nu,\tau}(x) = \nabla \log \nu_\tau(x) = \frac{\mathbb{E}[Y \mid Y+Z=x] - x}{\tau^2}.$$

Proof of (b). Since ν is compactly supported, ν_τ is C^∞ and strictly positive. Let

$$m(x) := \mathbb{E}[Y \mid Y+Z=x], \qquad \Sigma(x) := \mathrm{Cov}(Y \mid Y+Z=x).$$

A standard differentiation of the posterior weights yields

$$\nabla m(x) = \frac{1}{\tau^2}\Sigma(x).$$

Differentiating the identity from part (a),

$$s_{\nu,\tau}(x) = \frac{m(x)-x}{\tau^2},$$

therefore gives

$$\nabla s_{\nu,\tau}(x) = -\frac{1}{\tau^2}I_d + \frac{1}{\tau^4}\Sigma(x).$$

Proof of (c). Apply part (a) to ν_1 and ν_2:

$$s_{\nu_i,\tau}(x) = \frac{m_i(x)-x}{\tau^2}, \qquad m_i(x) := \mathbb{E}_i[Y \mid Y+Z=x].$$

Because each posterior mean $m_i(x)$ lies in the convex hull of $[-R,R]^d$, it has Euclidean norm at most $R\sqrt{d}$. Hence

$$\|m_1(x) - m_2(x)\|_2 \le 2R\sqrt{d},$$

and therefore

$$\|s_{\nu_1,\tau}(x) - s_{\nu_2,\tau}(x)\|_2 \le \frac{2R\sqrt{d}}{\tau^2}.$$

Proof of (d). By part (b),

$$\nabla s_{\nu,\tau}(x) = -\frac{1}{\tau^2}I_d + \frac{1}{\tau^4}\Sigma(x).$$

Since $\Sigma(x) \preceq \mathbb{E}[YY^\top \mid Y+Z=x]$, we have

$$\|\Sigma(x)\|_{\mathrm{op}} \le \mathbb{E}[\|Y\|_2^2 \mid Y+Z=x] \le R^2 d.$$

Thus

$$\|\nabla s_{\nu,\tau}(x)\|_{\mathrm{op}} \le \frac{1}{\tau^2} + \frac{R^2 d}{\tau^4}.$$

Taking the supremum over x proves the Lipschitz bound. □

Proposition B.7 (KL/Fisher control under Gaussian smoothing)**.** *Let S be any codebook and define*

$$D(t) := D_{\mathrm{KL}}(\nu_{S,\sqrt{t}} \,\|\, \nu_{U,\sqrt{t}}), \qquad J(t) := \int \nu_{S,\sqrt{t}}(x)\left\|\nabla \log \frac{d\nu_{S,\sqrt{t}}}{d\nu_{U,\sqrt{t}}}(x)\right\|_2^2 dx.$$

There exists a universal constant $C > 0$ such that the following hold.

(a) *For every $t > 0$,*

$$D'(t) = -\frac{1}{2} J(t),$$

and if

$$\rho(t) := \sup_{x \in \mathbb{R}^d} \lambda_{\max}\big(\nabla^2 \log u_{\sqrt{t}}(x)\big), \qquad \rho_+(t) := \max\{\rho(t), 0\},$$

then

$$J'(t) \le 2\rho_+(t) J(t)$$

for almost every $t > 0$.

(b) *For every $\alpha \in (0,1)$ and every $\tau > 0$,*

$$J(\tau^2) \le \frac{2}{(1-\alpha)\tau^2} \exp\left(2 \int_{\alpha\tau^2}^{\tau^2} \rho_+(s)\, ds\right) D(\alpha\tau^2).$$

(c) *For the null heat flow generated by U,*

$$\rho_+(t) \le \frac{R^2}{t^2} \qquad \forall t > 0.$$

(d) *In particular, with $\alpha := 1 - \frac{\gamma^2}{R^2 d^2}$ and $\widetilde{\tau} := \sqrt{\alpha}\,\tau$,*

$$J_\tau(S) \le \frac{C R^2 d^2}{\gamma^2 \tau^2}\, D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}} \,\|\, \nu_{U,\widetilde{\tau}})$$

for all sufficiently large d and every $\tau \ge \gamma$.

Proof. For the proof it is convenient to work with the variance parameter $t = \tau^2$. For a fixed codebook S, define

$$D(t) := D_{\mathrm{KL}}(\nu_{S,\sqrt{t}} \,\|\, \nu_{U,\sqrt{t}}),$$

and

$$J(t) := \int \nu_{S,\sqrt{t}}(x) \left\| \nabla \log \frac{d\nu_{S,\sqrt{t}}}{d\nu_{U,\sqrt{t}}}(x) \right\|_2^2 dx.$$

Thus $J(\tau^2) = J_\tau(S)$.

Proof of (a). Because both $\nu_{S,\sqrt{t}}$ and $\nu_{U,\sqrt{t}}$ are finite Gaussian mixtures, they are smooth and strictly positive, and all integrations by parts are justified. Differentiating

$$D(t) = \int \nu_{S,\sqrt{t}} \log \frac{\nu_{S,\sqrt{t}}}{\nu_{U,\sqrt{t}}}$$

along the heat flow yields the relative de Bruijn identity

$$D'(t) = -\frac{1}{2} J(t).$$

Writing

$$h_t := \log \frac{\nu_{S,\sqrt{t}}}{\nu_{U,\sqrt{t}}}, \qquad \rho(t) := \sup_x \lambda_{\max}\big(\nabla^2 \log u_{\sqrt{t}}(x)\big), \qquad \rho_+(t) := \max\{\rho(t), 0\},$$

the standard evolution formula for the relative Fisher information gives

$$J'(t) = -\int \nu_{S,\sqrt{t}}\, \|\nabla^2 h_t\|_F^2 + 2 \int \nu_{S,\sqrt{t}}\, \langle \nabla^2 \log u_{\sqrt{t}},\, \nabla h_t \otimes \nabla h_t \rangle.$$

Dropping the nonpositive first term and using

$$\langle A, v \otimes v \rangle \le \lambda_{\max}(A) \|v\|_2^2$$

yields

$$J'(t) \le 2\rho_+(t) J(t)$$

for almost every $t > 0$.

Proof of (b). Fix $\alpha \in (0, 1)$. If $J(\tau^2) = 0$ there is nothing to prove. Otherwise, integrating the differential inequality from part (a) over $[\alpha\tau^2, \tau^2]$ gives

$$J(s) \ge J(\tau^2) \exp\left(-2 \int_{\alpha\tau^2}^{\tau^2} \rho_+(r)\, dr\right) \qquad \text{for all } s \in [\alpha\tau^2, \tau^2].$$

Using $D'(s) = -J(s)/2$ and $D(\tau^2) \ge 0$, we obtain

$$\begin{aligned} D(\alpha\tau^2) \ge D(\alpha\tau^2) - D(\tau^2) &= \frac{1}{2} \int_{\alpha\tau^2}^{\tau^2} J(s)\, ds \\ &\ge \frac{1}{2}(1-\alpha)\tau^2\, J(\tau^2)\, \exp\left(-2 \int_{\alpha\tau^2}^{\tau^2} \rho_+(r)\, dr\right). \end{aligned}$$

Rearranging proves part (b).

Proof of (c) and (d). Because U is a product of $d/2$ two-dimensional circle laws, the density $u_{\sqrt{t}}$ factorizes blockwise. By Proposition B.6(b), each two-dimensional block satisfies

$$\nabla^2 \log u^{(2)}_{\sqrt{t}}(x) = -\frac{1}{t} I_2 + \frac{1}{t^2} \operatorname{Cov}(Y \mid Y + Z = x),$$

and since $\|Y\|_2 = R$ almost surely in each block,

$$\lambda_{\max}\big(\operatorname{Cov}(Y \mid Y + Z = x)\big) \le R^2.$$

Therefore

$$\rho_+(t) \le \frac{R^2}{t^2} \qquad \text{for all } t > 0.$$

Now fix

$$\alpha := 1 - \frac{\gamma^2}{R^2 d^2}, \qquad \widetilde{\tau} := \sqrt{\alpha}\, \tau.$$

Then

$$\int_{\widetilde{\tau}^2}^{\tau^2} \rho_+(s)\, ds \le R^2 \left(\frac{1}{\widetilde{\tau}^2} - \frac{1}{\tau^2}\right) = \frac{R^2(1-\alpha)}{\alpha\tau^2} = \frac{\gamma^2}{\alpha d^2 \tau^2} \le \frac{1}{\alpha d^2},$$

since $\tau \ge \gamma$. For all sufficiently large d, the exponential factor in part (b) is therefore bounded by an absolute constant. Using

$$\frac{1}{1-\alpha} = \frac{R^2 d^2}{\gamma^2},$$

part (b) yields

$$J_\tau(S) = J(\tau^2) \le \frac{C R^2 d^2}{\gamma^2 \tau^2}\, D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}} \,\|\, \nu_{U,\widetilde{\tau}}),$$

for a sufficiently large universal constant C. □

Log-likelihood ratio concentration.

Proposition B.8 (Log-likelihood ratio concentration)**.** *There exists a universal constant $c > 0$ such that the following hold for every $\tau > 0$.*

(a) For every $t \ge 0$,

$$\mathbb{P}\big[|\ell_\tau - dI_\tau| \ge t\big] \le 2 \exp\big(-c \min\{t^2/d,\ t\}\big).$$

(b) For every $t \ge 0$,

$$\mathbb{P}\big[\ell_\tau - dI_\tau \ge t\big] \le \exp\left(-c \frac{\tau^2 t^2}{R^2 d}\right).$$

Proof. We prove part (a) by showing that the centered block log-likelihood ratio has a universal ψ_1 norm. Since the full d-dimensional log-likelihood ratio is a sum of $d/2$ independent block contributions, Bernstein's inequality then yields the desired two-sided sub-exponential tail. Part (b) is a separate Gaussian-Lipschitz estimate.

Because U and the Gaussian noise factorize across the $m = d/2$ two-dimensional blocks,

$$\ell_\tau(Y, Y+Z) = \sum_{j=1}^{m} \ell_\tau^{(2)}(Y^{(j)}, Y^{(j)} + Z^{(j)}),$$

where $\ell_\tau^{(2)}$ denotes the two-dimensional log-likelihood ratio for a single block. We first prove a universal ψ_1 bound for one block; the two-sided sub-exponential tail bound in part (a) then follows from Bernstein's inequality.

Fix one block. By rotational symmetry of the M reference points on the circle and isotropy of the noise, the distribution of $\ell_\tau^{(2)}(Y^{(1)}, Y^{(1)} + Z^{(1)})$ does not depend on the underlying block point. We may therefore assume that this point is the reference point $a_0 = (R, 0)$ and write $Z^{(1)} = \tau N$ with $N \sim \mathcal{N}(0, I_2)$. Set

$$\lambda := \frac{\tau}{\gamma}, \qquad b_k := \frac{a_k}{\gamma} \qquad (0 \le k \le M-1),$$

so that the points b_k lie on the circle of radius R/γ and

$$M = \left\lceil \frac{\pi R}{\gamma} \right\rceil.$$

Define

$$S_\lambda(N) := \sum_{k=0}^{M-1} \exp\!\left(-\frac{\|b_0 - b_k\|_2^2 + 2\lambda \langle N, b_0 - b_k \rangle}{2\lambda^2} \right).$$

Then

$$\ell_\tau^{(2)}(a_0, a_0 + \tau N) = \log M - \log S_\lambda(N).$$

It therefore suffices to show that $\log S_\lambda(N) - \mathbb{E} \log S_\lambda(N)$ has a universal ψ_1 norm.

Regime 1: $\lambda \ge M/10$. Set $f_\lambda(N) := \log S_\lambda(N)$. Differentiating gives

$$\nabla f_\lambda(N) = -\frac{1}{\lambda} \sum_{k=0}^{M-1} w_k(N)(b_0 - b_k),$$

where the weights $w_k(N)$ form a convex combination. Hence

$$\|\nabla f_\lambda(N)\|_2 \le \frac{2(R/\gamma)}{\lambda}.$$

Since $M = \lceil \pi R/\gamma \rceil$, the condition $\lambda \ge M/10$ implies $2(R/\gamma)/\lambda \le 20/\pi < 7$. Thus f_λ is uniformly Lipschitz, and Gaussian concentration gives a universal sub-Gaussian bound on $f_\lambda(N) - \mathbb{E} f_\lambda(N)$, hence also a universal ψ_1 bound.

Regime 2: $\lambda < M/10$. Write $N = (N_1, N_2)$. We claim that

$$\big| \log S_\lambda(N) - \log(1+\lambda) \big| \le C\big(1 + \|N\|_2^2\big) \tag{1}$$

for a universal constant C. Taking expectations in equation 1 and using $\mathbb{E}\|N\|_2^2 = 2$ gives

$$\big| \mathbb{E} \log S_\lambda(N) - \log(1+\lambda) \big| \le C,$$

so

$$\big| \log S_\lambda(N) - \mathbb{E} \log S_\lambda(N) \big| \le C\big(1 + \|N\|_2^2\big).$$

Since $\|N\|_2^2$ has a universal ψ_1 norm, this implies the desired block-level ψ_1 bound.

To prove equation 1, let $k' := \min\{k, M-k\}$. Using

$$\|b_0 - b_k\|_2 = 2\frac{R}{\gamma} \sin(\pi k'/M)$$

and the inequalities $(2/\pi)x \le \sin x \le x$ on $[0,\pi/2]$, we have

$$\frac{4R}{\gamma M}k' \le \|b_0 - b_k\|_2 \le \frac{2\pi R}{\gamma M}k'.$$

Hence

$$-\frac{\|b_0 - b_k\|_2^2 + 2\lambda\langle N, b_0 - b_k\rangle}{2\lambda^2} \le -aj^2 + u\|N\|_2\, j$$

with

$$a := \frac{(4R/(\gamma M))^2}{2\lambda^2}, \qquad u := \frac{2\pi R/(\gamma M)}{\lambda}, \qquad j = k'.$$

Since each value of j occurs for at most two indices k,

$$S_\lambda(N) \le 1 + 2\sum_{j=1}^{\lfloor M/2\rfloor} e^{-aj^2 + u\|N\|_2 j}.$$

Completing the square,

$$-aj^2 + u\|N\|_2 j = \frac{u^2}{4a}\|N\|_2^2 - a\Big(j - \frac{u}{2a}\|N\|_2\Big)^2, \qquad \frac{u^2}{4a} = \frac{\pi^2}{8}.$$

A comparison with the Gaussian integral gives

$$\sum_{j\in\mathbb{Z}} e^{-a(j-\mu)^2} \le C\Big(1 + \frac{1}{\sqrt{a}}\Big)$$

uniformly in $\mu \in \mathbb{R}$. Since $M = \lceil \pi R/\gamma \rceil$, the ratio $M/(R/\gamma)$ is bounded above and below by absolute constants, so $1/\sqrt{a} \asymp \lambda$. Therefore

$$S_\lambda(N) \le C(1+\lambda)\exp\big(C\|N\|_2^2\big),$$

and hence

$$\log S_\lambda(N) \le \log(1+\lambda) + C + C\|N\|_2^2.$$

For the lower bound, the $k = 0$ term already contributes 1 to $S_\lambda(N)$. Thus, if $\lambda < 1$ then

$$S_\lambda(N) \ge 1 \ge c(1+\lambda)\exp\big(-C(1 + \|N\|_2^2)\big)$$

for suitable universal constants c, C, since $1 + \lambda \le 2$.

Assume now that $\lambda \ge 1$. Pair the terms k and $M - k$. Writing $\theta_k := 2\pi k/M$, one checks that

$$e^{-A_k - B_k} + e^{-A_{M-k} - B_{M-k}} = 2\exp\Big(-A_k - \frac{R}{\gamma}\frac{1-\cos\theta_k}{\lambda}N_1\Big)\cosh\Big(\frac{R}{\gamma}\frac{\sin\theta_k}{\lambda}N_2\Big),$$

with $A_k = \|b_0 - b_k\|_2^2/(2\lambda^2)$ and $B_k = \langle N, b_0 - b_k\rangle/\lambda$. Since $\cosh(\cdot) \ge 1$ and

$$\frac{R}{\gamma}\frac{1-\cos\theta_k}{\lambda}N_1^+ = \frac{\lambda N_1^+}{R/\gamma}\,A_k,$$

we obtain

$$e^{-A_k - B_k} + e^{-A_{M-k} - B_{M-k}} \ge 2\exp\Big(-\Big(1 + \frac{\lambda N_1^+}{R/\gamma}\Big)A_k\Big).$$

Let $L := \lfloor \lambda \rfloor$. Because $\lambda < M/10$, we have $L \le \lfloor (M-1)/2 \rfloor$. For every $1 \le k \le L$,

$$A_k \le \frac{1}{2}\Big(\frac{2\pi R}{\gamma M}\Big)^2 \le C,$$

and therefore

$$\frac{\lambda N_1^+}{R/\gamma}\,A_k \le \frac{\lambda N_1^+}{R/\gamma}\cdot\frac{1}{2}\Big(\frac{2\pi R}{\gamma M}\Big)^2 \le C\frac{\lambda}{M}N_1^+ \le CN_1^+,$$

using $M \ge \pi R/\gamma$ and $\lambda < M/10$. Hence each pair with $1 \le k \le L$ contributes at least

$$c\exp(-CN_1^+)$$

for universal constants c, C. Summing these L pairs together with the $k = 0$ term gives

$$S_\lambda(N) \geq 1 + cL\exp(-CN_1^+) \geq c(1+\lambda)\exp\big(-C(1+N_1^+)\big).$$

Since $N_1^+ \leq (1 + \|N\|_2^2)/2$, this yields

$$S_\lambda(N) \geq c(1+\lambda)\exp\big(-C(1+\|N\|_2^2)\big),$$

which is the matching lower bound in equation 1. This proves the claim.

Summing the $m = d/2$ independent centered block contributions and applying Bernstein's inequality gives

$$\mathbb{P}\big[|\ell_\tau - dI_\tau| \geq t\big] \leq 2\exp\big(-c\min\{t^2/d,\ t\}\big),$$

which is part (a).

For part (b), fix $y \in \mathcal{V}$ and define

$$f_y(z) := \ell_\tau(y, y+z).$$

Differentiating the identity

$$\ell_\tau(y, y+z) = -\frac{\|z\|_2^2}{2\tau^2} - \log u_\tau(y+z) - \frac{d}{2}\log(2\pi\tau^2)$$

with respect to z gives

$$\nabla_z f_y(z) = -\frac{m_U(y+z) - y}{\tau^2},$$

where $m_U(\cdot)$ is the posterior mean under the null law U. Hence

$$\|\nabla_z f_y(z)\|_2 \leq \frac{2R\sqrt{d}}{\tau^2}.$$

Gaussian concentration for $Z \sim \mathcal{N}(0, \tau^2 I_d)$ therefore gives

$$\mathbb{P}\big[f_y(Z) - \mathbb{E}f_y(Z) \geq t\big] \leq \exp\left(-\frac{\tau^2 t^2}{8R^2 d}\right).$$

By rotational symmetry, the law of $\ell_\tau(y, y+Z)$ does not depend on y, so $\mathbb{E}f_y(Z) = dI_\tau$. Averaging over $Y \sim U$ proves part (b). □

Lemma B.9 (A quantile lower bound from the rate gap)**.** *There exists a universal constant $C > 0$ such that for every $\tau > 0$ and every $\zeta \in (0, 1/2]$,*

$$\log \Lambda_\tau(\zeta) \geq dI_\tau - C\sqrt{d\log(1/\zeta)} - C\log(1/\zeta).$$

Proof. By definition, any threshold $t \in \mathbb{R}$ with

$$\mathbb{P}\big[\ell_\tau \leq t\big] < \zeta$$

must satisfy $t < \log \Lambda_\tau(\zeta)$. Let

$$s := C\Big(\sqrt{d\log(1/\zeta)} + \log(1/\zeta)\Big)$$

with $C > 0$ sufficiently large. Proposition B.8(a) then gives

$$\mathbb{P}\big[\ell_\tau \leq dI_\tau - s\big] \leq 2\exp\Big(-c\min\{s^2/d,\ s\}\Big) \leq \frac{\zeta}{2} < \zeta.$$

Hence $dI_\tau - s < \log \Lambda_\tau(\zeta)$, and therefore $\log \Lambda_\tau(\zeta) \geq dI_\tau - s$. □

Random-codebook KL bounds.

Lemma B.10 (KL bound from a likelihood-ratio tail)**.** *Let U be the uniform law on a finite set $\mathcal{V}$, and for each $y \in \mathcal{V}$ let $W(\cdot \mid y)$ be a probability distribution on $\mathcal{X}$. Define*

$$Q := \frac{1}{|\mathcal{V}|} \sum_{y \in \mathcal{V}} W(\cdot \mid y).$$

If $Y_1, \dots, Y_n$ are i.i.d. from U and

$$P_n := \frac{1}{n} \sum_{i=1}^{n} W(\cdot \mid Y_i),$$

then for every $a \geq 0$,

$$\mathbb{E}\big[D_{\mathrm{KL}}(P_n \| Q)\big] \leq e^{-a} + \log(1 + |\mathcal{V}|)\, \mathbb{P}\big[\imath(Y, X) \geq \log n - a\big],$$

where $Y \sim U$, $X \sim W(\cdot \mid Y)$, and

$$\imath(y, x) := \log \frac{dW(\cdot \mid y)}{dQ}(x)$$

denotes the corresponding pointwise log-likelihood ratio with respect to Q.

Proof. Let I be uniform on $\{1, \dots, n\}$, independent of everything else, and let

$$X \sim W(\cdot \mid Y_I)$$

conditionally on $(Y_1, \dots, Y_n, I)$. Then, conditional on $(Y_1, \dots, Y_n)$, the law of X is exactly P_n, and therefore

$$D_{\mathrm{KL}}(P_n \| Q) = \mathbb{E}\left[\log\Big(\frac{1}{n} \sum_{i=1}^{n} e^{\imath(Y_i, X)}\Big) \,\Big|\, Y_1, \dots, Y_n\right].$$

Taking expectations and using exchangeability, we may replace the random index I by 1 and obtain

$$\mathbb{E}\big[D_{\mathrm{KL}}(P_n \| Q)\big] = \mathbb{E}\left[\log\Big(\frac{1}{n} \sum_{i=1}^{n} e^{\imath(Y_i, X)}\Big)\right],$$

where now $Y_1, \dots, Y_n$ are i.i.d. from U and $X \sim W(\cdot \mid Y_1)$.

Condition on (Y_1, X). For every $i \geq 2$, the random variable Y_i is independent of X and

$$\mathbb{E}\big[e^{\imath(Y_i, X)} \mid X\big] = \sum_{y \in \mathcal{V}} U(y) \frac{dW(\cdot \mid y)}{dQ}(X) = 1.$$

By Jensen's inequality for the concave logarithm,

$$\begin{aligned} \mathbb{E}\left[\log\Big(\frac{1}{n} \sum_{i=1}^{n} e^{\imath(Y_i, X)}\Big) \,\Big|\, Y_1, X\right] &\leq \log\Big(\frac{e^{\imath(Y_1, X)} + (n-1)}{n}\Big) \\ &\leq \log\Big(1 + \frac{e^{\imath(Y_1, X)}}{n}\Big). \end{aligned}$$

Since U is uniform on $\mathcal{V}$, we also have the pointwise bound

$$e^{\imath(y, x)} = \frac{dW(\cdot \mid y)}{dQ}(x) \leq |\mathcal{V}|$$

for all $y \in \mathcal{V}$ and all $x \in \mathcal{X}$. Hence, on the event

$$\big\{\imath(Y_1, X) \leq \log n - a\big\}$$

we have

$$\log\Big(1 + \frac{e^{\imath(Y_1, X)}}{n}\Big) \leq e^{-a},$$

while on its complement

$$\log\Big(1 + \frac{e^{\imath(Y_1, X)}}{n}\Big) \leq \log(1 + |\mathcal{V}|).$$

Averaging over (Y_1, X) proves the claim. $\square$

Lemma B.11 (Explicit random-codebook KL bound for the hard family)**.** *There exists a universal constant $c > 0$ such that the following holds. For every integer $n \ge 1$, every $\tau > 0$, and every $a \ge 0$,*

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\tau}\|\nu_{U,\tau})\big] \le e^{-a}+2\log(1+|\mathcal{V}|)\cdot\exp\Big(-c\min\big\{(d(\kappa(n)-I_\tau)-a)_+^2/d,\ (d(\kappa(n)-I_\tau)-a)_+\big\}\Big).$$

Moreover,

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\tau}\|\nu_{U,\tau})\big] \le e^{-a}+\log(1+|\mathcal{V}|)\cdot\exp\bigg(-c\frac{\tau^2}{R^2 d}\big(d(\kappa(n)-I_\tau)-a\big)_+^2\bigg).$$

Proof. Apply Lemma B.10 to the Gaussian observation distributions

$$W(\cdot\mid y) = \mathcal{N}(y,\tau^2 I_d).$$

Then $Q = \nu_{U,\tau}$, the empirical average over the sampled codebook is $\nu_{S,\tau}$, and the log-likelihood ratio is

$$\imath(y,x) = \ell_\tau(y,x).$$

Hence, for every $a \ge 0$,

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\tau}\|\nu_{U,\tau})\big] \le e^{-a}+\log(1+|\mathcal{V}|)\,\mathbb{P}\big(\ell_\tau \ge \log n - a\big).$$

Set

$$b := d(\kappa(n)-I_\tau)-a, \qquad t := b_+ = (d(\kappa(n)-I_\tau)-a)_+.$$

If $b \le 0$, then the probability above is at most 1, and therefore

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\tau}\|\nu_{U,\tau})\big] \le e^{-a}+\log(1+|\mathcal{V}|),$$

which is stronger than both stated bounds because $t = 0$ in this case.

Assume now that $b > 0$, so that $t = b$. Since $\log n = d\kappa(n)$ and $\mathbb{E}[\ell_\tau] = dI_\tau$,

$$\mathbb{P}\big(\ell_\tau \ge \log n - a\big) = \mathbb{P}\big(\ell_\tau - dI_\tau \ge t\big).$$

Applying Proposition B.8(a) yields

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\tau}\|\nu_{U,\tau})\big] \le e^{-a}+2\log(1+|\mathcal{V}|)\cdot\exp\Big(-c\min\big\{t^2/d,\ t\big\}\Big)$$

This proves the first estimate. Applying Proposition B.8(b) instead gives

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\tau}\|\nu_{U,\tau})\big] \le e^{-a}+\log(1+|\mathcal{V}|)\cdot\exp\bigg(-c\frac{\tau^2}{R^2 d}t^2\bigg)$$

This proves the second estimate. ∎

B.4 Proof of the large-sample and interval theorems

We now return to the statements from Section B.2.

Proof of Proposition B.2. For $y \in \mathcal{V}$, set

$$f_y(x) := \log L_\tau(y,x) = \log\varphi_{\tau^2}(x-y) - \log u_\tau(x).$$

Differentiating and using Proposition B.6(a) for the null law U,

$$\nabla f_y(x) = \frac{y-x}{\tau^2} - s_{U,\tau}(x) = \frac{y-m_U(x)}{\tau^2},$$

where $m_U(x) := \mathbb{E}[Y \mid Y+Z=x]$ under $Y \sim U$ and $Z \sim \mathcal{N}(0,\tau^2 I_d)$. Both y and $m_U(x)$ belong to the convex hull of $[-R,R]^d$, so

$$\|\nabla f_y(x)\|_2 \le \frac{2R\sqrt{d}}{\tau^2}.$$

Thus every f_y is $(2R\sqrt{d}/\tau^2)$-Lipschitz, and the same bound passes to

$$\ell_{\tau,S}^{\max} = \max_{y \in S} f_y \qquad \text{and} \qquad m_{\tau,S} = \psi\big(\ell_{\tau,S}^{\max} - \log \Lambda_\tau + 1\big),$$

since maxima preserve Lipschitz constants and ψ is 1-Lipschitz.

If $J_\tau(S) \le \theta(\tau)$, then $\widehat{s}_\tau^{(S)} = s_{U,\tau}$, so the desired Lipschitz bound follows directly from Proposition B.6(d). Otherwise write

$$\Delta_\tau := s_{S,\tau} - s_{U,\tau}, \qquad \widehat{s}_\tau^{(S)} = s_{U,\tau} + m_{\tau,S}\Delta_\tau.$$

By Proposition B.6(c) and (d),

$$\sup_x \|\Delta_\tau(x)\|_2 \le \frac{2R\sqrt{d}}{\tau^2}, \qquad \mathrm{Lip}(\Delta_\tau) \le 2\Big(\frac{1}{\tau^2} + \frac{R^2 d}{\tau^4}\Big).$$

Applying the product rule for Lipschitz functions,

$$\begin{aligned}
\mathrm{Lip}(\widehat{s}_\tau^{(S)}) &\le \mathrm{Lip}(s_{U,\tau}) + \mathrm{Lip}(m_{\tau,S}) \sup_x \|\Delta_\tau(x)\|_2 + \mathrm{Lip}(\Delta_\tau) \\
&\le \Big(\frac{1}{\tau^2} + \frac{R^2 d}{\tau^4}\Big) + \frac{2R\sqrt{d}}{\tau^2} \cdot \frac{2R\sqrt{d}}{\tau^2} + 2\Big(\frac{1}{\tau^2} + \frac{R^2 d}{\tau^4}\Big) \\
&\le \frac{3}{\tau^2} + \frac{7R^2 d}{\tau^4}.
\end{aligned}$$

□

Proof of Theorem B.4. By Proposition B.7(d),

$$J_\tau(S) \le \frac{C_1 R^2 d^2}{\gamma^2 \tau^2} D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}} \| \nu_{U,\widetilde{\tau}})$$

for a universal constant $C_1 > 0$. Hence Markov's inequality gives

$$\mathbb{P}_S\big[J_\tau(S) > \theta(\tau)\big] \le \frac{C_1 R^2 d^2}{\gamma^2 \tau^2 \theta(\tau)} \,\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}} \| \nu_{U,\widetilde{\tau}})\big].$$

It therefore suffices to show that the expectation on the right is at most

$$\frac{\delta\, \tau^2 \theta(\tau) \gamma^2}{C_1 R^2 d^2}.$$

Let

$$a := \frac{C}{2} H(\tau),$$

where C is the large constant appearing in the statement. Since

$$\kappa(n) \ge I_{\widetilde{\tau}} + \frac{1}{d} \min\{E_{\mathrm{med}}(\tau), E_{\mathrm{big}}(\tau)\},$$

we have

$$d\big(\kappa(n) - I_{\widetilde{\tau}}\big) - a \ge \min\{E_{\mathrm{med}}(\tau), E_{\mathrm{big}}(\tau)\} - a.$$

Set

$$t := d\big(\kappa(n) - I_{\widetilde{\tau}}\big) - a.$$

Then either

$$t \ge C\sqrt{dH(\tau)} + \frac{C}{2} H(\tau)$$

or

$$t \ge C\frac{R}{\widetilde{\tau}}\sqrt{dH(\tau)} + \frac{C}{2} H(\tau).$$

If the first alternative holds, then the first estimate in Lemma B.11 at noise $\widetilde{\tau}$ gives

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}} \| \nu_{U,\widetilde{\tau}})\big] \le e^{-a} + 2\log(1 + |\mathcal{V}|)\, \exp\left(-c \min\left\{\frac{t^2}{d}, t\right\}\right).$$

Since

$$t \geq C\sqrt{dH(\tau)} + \frac{C}{2}H(\tau),$$

choosing the constant C in the present theorem sufficiently large yields

$$\min\left\{\frac{t^2}{d},\, t\right\} \geq c'H(\tau),$$

and therefore

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}}\|\nu_{U,\widetilde{\tau}})\big] \leq e^{-a} + 2\log(1+|\mathcal{V}|)\, e^{-c'H(\tau)}.$$

If the second alternative holds, then the second estimate in Lemma B.11 at noise $\widetilde{\tau}$ gives

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}}\|\nu_{U,\widetilde{\tau}})\big] \leq e^{-a} + \log(1+|\mathcal{V}|)\, \exp\!\left(-c\frac{\widetilde{\tau}^2 t^2}{R^2 d}\right).$$

Since

$$t \geq C\frac{R}{\widetilde{\tau}}\sqrt{dH(\tau)} + \frac{C}{2}H(\tau),$$

choosing the constant C in the present theorem sufficiently large again yields

$$\frac{\widetilde{\tau}^2 t^2}{R^2 d} \geq c'H(\tau),$$

and therefore

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}}\|\nu_{U,\widetilde{\tau}})\big] \leq e^{-a} + \log(1+|\mathcal{V}|)\, e^{-c'H(\tau)}.$$

In either case,

$$\mathbb{E}_S\big[D_{\mathrm{KL}}(\nu_{S,\widetilde{\tau}}\|\nu_{U,\widetilde{\tau}})\big] \leq e^{-a} + 2\log(1+|\mathcal{V}|)\, e^{-c'H(\tau)}.$$

Finally, by the definition of $H(\tau)$ and by increasing the constant C in that definition if necessary, we may ensure that

$$\frac{C_1 R^2 d^2}{\gamma^2\tau^2\theta(\tau)}\big(1 + 2\log(1+|\mathcal{V}|)\big)e^{-c'H(\tau)} \leq \delta.$$

Hence

$$\mathbb{P}_S\big[J_\tau(S) > \theta(\tau)\big] \leq \delta.$$

Whenever $J_\tau(S) \leq \theta(\tau)$, the oracle is everywhere equal to the null score, and therefore equals $s_{U,\tau}$ pointwise. This proves the stated null-coupling conclusion. □

Proof of Theorem B.5. By definition,

$$\kappa_-(\tau) = \frac{1}{d}\log\Lambda_\tau + \frac{\log\delta - 1}{d}.$$

Lemma B.9 therefore gives

$$\kappa_-(\tau) \geq I_\tau - C\sqrt{\frac{\log(1/\zeta(\tau))}{d}} - C\frac{\log(1/\zeta(\tau)) + \log(1/\delta)}{d}.$$

On the other hand, Theorem B.4 gives

$$\kappa_+(\tau) \leq I_{\widetilde{\tau}} + C\sqrt{\frac{H(\tau)}{d}} + C\frac{H(\tau)}{d},$$

where $\widetilde{\tau} = \sqrt{1-\gamma^2/(R^2d^2)}\,\tau$.

We next compare $I_{\widetilde{\tau}}$ and I_τ. Let

$$\mathcal{I}(t) := I(Y; Y + \sqrt{t}\,Z) = dI_{\sqrt{t}},$$

where $Y \sim U$ and $Z \sim \mathcal{N}(0, I_d)$. By the standard de Bruijn identity,

$$\frac{d}{dt}h(\nu_{U,\sqrt{t}}) = \frac{1}{2}J(\nu_{U,\sqrt{t}}),$$

so

$$\mathcal{I}'(t) = \frac{1}{2} J(\nu_{U,\sqrt{t}}) - \frac{d}{2t}.$$

The Fisher information of a Gaussian-smoothed law satisfies the standard Stam bound

$$J(\nu_{U,\sqrt{t}}) \le \frac{d}{t},$$

whence

$$-\frac{d}{2t} \le \mathcal{I}'(t) \le 0.$$

Integrating from $\widetilde{\tau}^2$ to τ^2 yields

$$0 \le dI_{\widetilde{\tau}} - dI_\tau = \mathcal{I}(\widetilde{\tau}^2) - \mathcal{I}(\tau^2) \le \frac{d}{2}\log\Big(\frac{\tau^2}{\widetilde{\tau}^2}\Big) = \frac{d}{2}\log\Big(\frac{1}{1-\gamma^2/(R^2d^2)}\Big) \le \frac{C\gamma^2}{R^2 d}.$$

Hence

$$0 \le I_{\widetilde{\tau}} - I_\tau \le \frac{C\gamma^2}{R^2 d^2}.$$

For all sufficiently large d, this term is dominated by

$$C\sqrt{\frac{H_{\rm win}(\tau)}{d}} + C\frac{H_{\rm win}(\tau)}{d},$$

since $H_{\rm win}(\tau) \ge 1$.

Combining the two bounds yields

$$\kappa_+(\tau) - \kappa_-(\tau) \le C\sqrt{\frac{H_{\rm win}(\tau)}{d}} + C\frac{H_{\rm win}(\tau)}{d},$$

because

$$H_{\rm win}(\tau) = \max\{\log(1/\zeta(\tau)),\, H(\tau),\, \log(1/\delta)\}.$$

By the standing interval convention, this is equivalent to the asserted bound on $|\mathcal{J}(\tau)|$.

For the null-coupling statement, fix $x \in \mathbb{R}^d$ and an integer $n \ge 1$. If $\kappa(n) < \kappa_-(\tau)$, then Proposition B.3 applies. If $\kappa(n) > \kappa_+(\tau)$, then Theorem B.4 applies. Thus

$$\kappa(n) \notin \mathcal{J}(\tau) \Longrightarrow \mathbb{P}_S\big[\widehat{s}_\tau^{(S)}(x) \ne s_{U,\tau}(x)\big] \le \delta,$$

as required. □

B.5 Small-noise information and packing

Each parameterized section chooses its own admissible range of sample sizes. The next two lemmas are the only ingredients shared by the two parameterized arguments: a uniform small-noise mutual-information estimate and a packing lemma on a discrete rate set.

Lemma B.12 (Small-noise mutual-information lower bound)**.** *There exist absolute constants $c_{\rm sm} > 1$ and $c_I > 1/16$ such that, for every $\tau \in [\gamma, c_{\rm sm}\gamma]$,*

$$I_\tau \ge c_I \log M.$$

Proof. Since

$$M = \left\lceil \frac{\pi R}{\gamma} \right\rceil$$

and $\gamma < R/2$, we have $M \ge 7$. Let $d_{\min}$ be the nearest-neighbor spacing in one planar block of the product-circle support. Then

$$d_{\min} = 2R\sin(\pi/M).$$

For a fixed block point, the nearest-point estimator can err only if the Gaussian perturbation crosses one of the two Voronoi bisectors adjacent to that point. Each such bisector lies at distance $d_{\min}/2$

from the point. Writing $\overline{\Phi}(t) := \mathbb{P}[N \geq t]$ for $N \sim \mathcal{N}(0,1)$, the block error probability is therefore at most

$$q(\tau) := 2\overline{\Phi}\left(\frac{d_{\min}}{2\tau}\right),$$

which is a continuous increasing function of τ. Moreover,

$$\frac{d_{\min}}{2\gamma} = \frac{R}{\gamma}\sin(\pi/M) > \frac{M-1}{\pi}\sin(\pi/M) = \left(1 - \frac{1}{M}\right)\frac{\sin(\pi/M)}{\pi/M}.$$

Since $M \geq 7$ and $x \mapsto \sin x / x$ is decreasing on $(0, \pi)$,

$$\left(1 - \frac{1}{M}\right)\frac{\sin(\pi/M)}{\pi/M} \geq \frac{6}{7} \cdot \frac{\sin(\pi/7)}{\pi/7} > 0.8.$$

Fix

$$q_\star := \frac{7}{16}.$$

Since $2\overline{\Phi}(0.8) < 7/16$, we may choose an absolute constant $c_{\mathrm{sm}} > 1$ sufficiently close to 1 so that

$$2\overline{\Phi}\left(\frac{0.8}{c_{\mathrm{sm}}}\right) \leq q_\star.$$

Then for every $\tau \in [\gamma, c_{\mathrm{sm}}\gamma]$,

$$q(\tau) = 2\overline{\Phi}\left(\frac{d_{\min}}{2\tau}\right) \leq 2\overline{\Phi}\left(\frac{0.8}{c_{\mathrm{sm}}}\right) \leq q_\star.$$

Fano's inequality therefore gives a lower bound of

$$F_M(q_\star) := \log M - h_2(q_\star) - q_\star \log(M-1)$$

on the mutual information contributed by a single two-dimensional block. Since

$$F_M(q_\star) \geq (1 - q_\star)\log M - h_2(q_\star)$$

and, for $M \geq 7$,

$$\left(1 - q_\star - \frac{1}{5}\right)\log M \geq \left(1 - q_\star - \frac{1}{5}\right)\log 7 > h_2(q_\star),$$

we obtain

$$F_M(q_\star) > \frac{1}{5}\log M.$$

Because the $d/2$ block observations are independent and identically distributed, the total mutual information is at least $(d/2)F_M(q_\star)$. Hence

$$I_\tau \geq \frac{1}{2}F_M(q_\star) > \frac{1}{10}\log M.$$

The claim follows with $c_I := 1/10$. □

Lemma B.13 (Packing on the admissible rate axis)**.** *Let $1 \leq n_{\min} \leq n_{\max}$ be integers, and write*

$$K_d := \{\kappa(n) : \ n \in [n_{\min}, n_{\max}] \cap \mathbb{N}\}.$$

Then for every $w > 0$, the set K_d contains a subset G whose points are pairwise separated by more than w and whose cardinality satisfies

$$|G| \geq \frac{\log(n_{\max}/(2n_{\min}))}{\log 2 + 2dw}.$$

Proof. If $n_{\max} < 2n_{\min}$, then

$$\log\left(\frac{n_{\max}}{2n_{\min}}\right) < 0,$$

so the asserted lower bound is nonpositive and the claim is trivial. We may therefore assume

$$n_{\max} \geq 2n_{\min}.$$

Set

$$\Gamma := \exp(2dw),$$

and recursively define

$$n_0 := n_{\min}, \qquad n_{j+1} := \lceil \Gamma n_j \rceil$$

for as long as $n_j \le n_{\max}$. Let m be the number of generated points not exceeding $n_{\max}$, and define

$$G := \{\kappa(n_j) : \; 0 \le j \le m-1\} \subseteq K_d.$$

Since $n_{j+1} \ge \Gamma n_j$, we have for every $0 \le j < k \le m-1$,

$$\kappa(n_k) - \kappa(n_j) \ge \frac{k-j}{d} \log \Gamma \ge 2w > w,$$

so the points of G are pairwise separated by more than w.

Because $\Gamma \ge 1$ and $n_j \ge 1$,

$$n_{j+1} = \lceil \Gamma n_j \rceil \le \Gamma n_j + 1 \le 2\Gamma n_j.$$

By induction,

$$n_j \le (2\Gamma)^j n_{\min} \qquad \text{for every } j \ge 0.$$

Set

$$j_\star := \left\lfloor \frac{\log(n_{\max}/(2n_{\min}))}{\log(2\Gamma)} \right\rfloor.$$

Then

$$(2\Gamma)^{j_\star} n_{\min} \le \frac{n_{\max}}{2} < n_{\max},$$

so $n_{j_\star} \le n_{\max}$ and therefore

$$m \ge j_\star + 1 \ge \frac{\log(n_{\max}/(2n_{\min}))}{\log(2\Gamma)}.$$

Since $\log(2\Gamma) = \log 2 + 2dw$, the claimed bound follows. □

C Proof of the parameterized L^p lower bound

Fix $p > 0$, $\rho \in (0, 1/4)$, an exponent $A \ge 0$, and parameters $R > 0$, $\gamma \in (0, R/2)$. For all sufficiently large d, assume

$$\varepsilon_{\text{err}} \in [d^{-A}, 1].$$

Set

$$H_{L^p} := \log(d/\varepsilon_{\text{err}}) + \log(R/\gamma),$$

let $c_0 = c_0(p, \rho) > 0$ be a sufficiently small constant, and define

$$Q_\star := \left\lfloor c_0 \min\left\{ d, \; \frac{d \, \log(R/\gamma)}{\sqrt{dH_{L^p}} + H_{L^p}} \right\} \right\rfloor.$$

We work in the nontrivial regime $Q_\star \ge 1$ and set $\delta := \rho^2/(80Q_\star)$. Choose an auxiliary constant $C_{\min} = C_{\min}(p, \rho) \ge 1$ sufficiently large, and define

$$n_{\min}^{(L^p)} := \left\lceil e^{C_{\min} H_{L^p}} \right\rceil, \qquad n_{\max}^{(L^p)} := \left\lfloor M^{d/32} \right\rfloor, \qquad \kappa_{\min}^{(L^p)} := \kappa(n_{\min}^{(L^p)}), \qquad \kappa_{\max}^{(L^p)} := \kappa(n_{\max}^{(L^p)}),$$

and

$$K_d^{(L^p)} := \{\kappa(n) : \; n \in [n_{\min}^{(L^p)}, n_{\max}^{(L^p)}] \cap \mathbb{N}\}.$$

Since $Q_\star \ge 1$, after taking d sufficiently large we have

$$\frac{d \log(R/\gamma)}{\sqrt{dH_{L^p}} + H_{L^p}} \ge \frac{1}{c_0},$$

and therefore, because $\log M \asymp \log(R/\gamma)$,

$$\sqrt{dH_{L^p}} + H_{L^p} \le C c_0 \, d \log M,$$

where $C > 0$ is an absolute constant.

C.1 The fixed-noise L^p oracle

For $\tau \geq \gamma$ define

$$\zeta^{(p)}(\tau) := \min\left\{\frac{1}{2},\ \left(\frac{\varepsilon_{\mathrm{err}}\tau}{4R\sqrt{d}}\right)^p\right\},$$

and

$$\theta^{(p)}(\tau) := \begin{cases} \varepsilon_{\mathrm{err}}^2/\tau^2, & 0 < p \leq 2, \\ \dfrac{\varepsilon_{\mathrm{err}}^p}{\tau^p}\left(\dfrac{\tau^2}{4R\sqrt{d}}\right)^{p-2}, & p > 2. \end{cases}$$

Let $\widehat{s}_\tau^{(S,p)}$ denote the oracle defined in Section B.2 with $(\zeta, \theta) = (\zeta^{(p)}, \theta^{(p)})$.

Lemma C.1 (Fixed-noise L^p accuracy)**.** *For every $\tau \geq \gamma$, the oracle $\widehat{s}_\tau^{(S,p)}$ is L^p-accurate at total noise τ in the sense that*

$$\mathbb{E}_{X\sim\nu_{S,\tau}}\left[\|\widehat{s}_\tau^{(S,p)}(X) - s_{S,\tau}(X)\|_2^p\right] \leq \frac{\varepsilon_{\mathrm{err}}^p}{\tau^p}$$

for every codebook S. Moreover,

$$\mathrm{Lip}\left(\widehat{s}_\tau^{(S,p)}\right) \leq \frac{3}{\tau^2} + \frac{7R^2 d}{\tau^4}.$$

Proof. Fix S and $\tau \geq \gamma$, and write

$$\Delta_\tau(x) := s_{S,\tau}(x) - s_{U,\tau}(x).$$

If $J_\tau(S) \leq \theta^{(p)}(\tau)$, then $\widehat{s}_\tau^{(S,p)} = s_{U,\tau}$. Hence

$$\widehat{s}_\tau^{(S,p)}(X) - s_{S,\tau}(X) = -\Delta_\tau(X), \qquad \mathbb{E}\|\Delta_\tau(X)\|_2^2 = J_\tau(S).$$

For $0 < p \leq 2$, Jensen gives

$$\mathbb{E}\|\Delta_\tau(X)\|_2^p \leq J_\tau(S)^{p/2} \leq \theta^{(p)}(\tau)^{p/2} = \frac{\varepsilon_{\mathrm{err}}^p}{\tau^p}.$$

For $p > 2$, Proposition B.6(c) gives

$$\|\Delta_\tau(x)\|_2 \leq \frac{2R\sqrt{d}}{\tau^2} \qquad \forall x,$$

so

$$\|\Delta_\tau(X)\|_2^p \leq \left(\frac{2R\sqrt{d}}{\tau^2}\right)^{p-2}\|\Delta_\tau(X)\|_2^2.$$

Using the definition of $\theta^{(p)}(\tau)$,

$$\mathbb{E}\|\Delta_\tau(X)\|_2^p \leq \left(\frac{2R\sqrt{d}}{\tau^2}\right)^{p-2}\theta^{(p)}(\tau) \leq \frac{\varepsilon_{\mathrm{err}}^p}{\tau^p}.$$

Assume next that $J_\tau(S) > \theta^{(p)}(\tau)$. Then

$$\widehat{s}_\tau^{(S,p)}(x) - s_{S,\tau}(x) = -(1 - m_{\tau,S}(x))\Delta_\tau(x),$$

and $1 - m_{\tau,S}(x)$ vanishes on $G_\tau(S)$ and is bounded by $\mathbf{1}\{x \notin G_\tau(S)\}$. Therefore

$$\|\widehat{s}_\tau^{(S,p)}(X) - s_{S,\tau}(X)\|_2^p \leq \|\Delta_\tau(X)\|_2^p \mathbf{1}\{X \notin G_\tau(S)\}.$$

Using again Proposition B.6(c) and the mass-coverage bound

$$\nu_{S,\tau}(G_\tau(S)) \geq 1 - \zeta^{(p)}(\tau),$$

we obtain

$$\mathbb{E}\|\widehat{s}_\tau^{(S,p)}(X) - s_{S,\tau}(X)\|_2^p \leq \left(\frac{2R\sqrt{d}}{\tau^2}\right)^p \zeta^{(p)}(\tau) \leq \frac{\varepsilon_{\mathrm{err}}^p}{\tau^p}.$$

The Lipschitz bound is Proposition B.2. □

Proposition C.2 (Extremal-noise collapse in the L^p case)**.** *There exist constants $c_{\rm sm}, C_{\rm lg} > 0$, depending only on p and ρ, such that the following hold for all sufficiently large d.*

(a) If $\tau \in [\gamma, c_{\rm sm}\gamma]$, then

$$\kappa_-^{(p)}(\tau) > \kappa_{\max}^{(L^p)},$$

where $\kappa_-^{(p)}(\tau) := \frac{1}{d}\log\big(\delta e^{-1}\Lambda_\tau(\zeta^{(p)}(\tau))\big)$.

(b) If $\tau \ge C_{\rm lg}R\sqrt{d}$, then

$$\kappa_+^{(p)}(\tau) < \kappa_{\min}^{(L^p)},$$

where $\kappa_+^{(p)}(\tau)$ is the upper threshold from Theorem B.4 specialized to the choice $(\zeta^{(p)}, \theta^{(p)})$.

Proof. We show that, at the two ends of the total-noise range, the admissible rate interval collapses outside $[\kappa_{\min}^{(L^p)}, \kappa_{\max}^{(L^p)}]$.

Small noise. Fix $\tau \in [\gamma, c_{\rm sm}\gamma]$. By Lemma B.12,

$$I_\tau \ge c_I \log M$$

with $c_I > 1/16$. Lemma B.9 with $\zeta = \zeta^{(p)}(\tau)$ yields

$$\log\Lambda_\tau\big(\zeta^{(p)}(\tau)\big) \ge dI_\tau - C\Big(\sqrt{d\log(1/\zeta^{(p)}(\tau))} + \log(1/\zeta^{(p)}(\tau))\Big).$$

Since $\tau \ge \gamma$,

$$\log(1/\zeta^{(p)}(\tau)) \le C\big(\log(d/\varepsilon_{\rm err}) + \log(R/\gamma)\big) = CH_{L^p}.$$

Therefore Lemma B.9 and the quantitative bound

$$\sqrt{dH_{L^p}} + H_{L^p} \le Cc_0\, d\log M$$

give

$$\log\Lambda_\tau\big(\zeta^{(p)}(\tau)\big) \ge c_I d\log M - C\big(\sqrt{dH_{L^p}} + H_{L^p}\big) \ge \big(c_I - Cc_0\big)d\log M.$$

Recalling the definition of $\kappa_-^{(p)}(\tau)$ and using $\log\delta^{-1} = O(\log d)$, we obtain

$$\kappa_-^{(p)}(\tau) \ge \big(c_I - Cc_0\big)\log M - o(1).$$

Since

$$\kappa_{\max}^{(L^p)} = \frac{1}{d}\log n_{\max}^{(L^p)} = \frac{1}{32}\log M + o(1)$$

and $c_I > 1/16$, choosing $c_0 = c_0(p,\rho)$ sufficiently small gives

$$c_I - Cc_0 > \frac{1}{32}.$$

Hence

$$\kappa_-^{(p)}(\tau) > \kappa_{\max}^{(L^p)}$$

for all sufficiently large d.

Large noise. Fix $\tau \ge C_{\rm lg}R\sqrt{d}$ with $C_{\rm lg}$ sufficiently large. By Theorem B.4,

$$\kappa_+^{(p)}(\tau) = I_{\widetilde{\tau}} + \frac{1}{d}\min\{E_{\rm med}(\tau), E_{\rm big}(\tau)\}.$$

Since $\widetilde{\tau} \asymp \tau$ and $\tau \ge C_{\rm lg}R\sqrt{d}$, the elementary Gaussian-observation bound (using $\mathrm{Cov}(Y) \preceq (R^2/2)I_d$) gives

$$I_{\widetilde{\tau}} \le \frac{1}{2}\log\bigg(1 + \frac{R^2}{2\widetilde{\tau}^2}\bigg) \le \frac{C}{d}.$$

For the L^p choice of (ζ,θ),

$$H(\tau) = \max\bigg\{1, \log\bigg(\frac{\log(1+|\mathcal{V}|)}{\delta}\cdot\frac{CR^2d^2}{\gamma^2\tau^2\theta^{(p)}(\tau)}\bigg)\bigg\} \le CH_{L^p}$$

in this range of τ: indeed $\tau \geq \gamma$, and the explicit form of $\theta^{(p)}(\tau)$ contributes at worst polynomial dependence on d, R/γ, and $\varepsilon_{\mathrm{err}}^{-1}$. Hence

$$\kappa_+^{(p)}(\tau) \leq \frac{CH_{L^p}}{d}.$$

Since

$$\kappa_{\min}^{(L^p)} = \frac{1}{d}\log n_{\min}^{(L^p)} \geq \frac{C_{\min}H_{L^p}}{d},$$

choosing $C_{\min} = C_{\min}(p, \rho)$ sufficiently large ensures

$$\kappa_{\min}^{(L^p)} \geq \frac{2CH_{L^p}}{d}.$$

Therefore

$$\kappa_+^{(p)}(\tau) < \kappa_{\min}^{(L^p)}.$$

□

Corollary C.3 (Rate-interval width in the L^p case)**.** *Under the standing assumptions of this section, there exists a family of intervals $\{\mathcal{J}^{(L^p)}(\tau)\}_{\tau\geq\gamma}$ such that*

$$\sup_{\tau\geq\gamma} |\mathcal{J}^{(L^p)}(\tau)| \leq C\sqrt{\frac{H_{L^p}}{d}} + C\frac{H_{L^p}}{d},$$

and for every $\tau \geq \gamma$, every fixed query point $x \in \mathbb{R}^d$, and every integer $n \in [n_{\min}^{(L^p)}, n_{\max}^{(L^p)}]$,

$$\kappa(n) \notin \mathcal{J}^{(L^p)}(\tau) \quad \Longrightarrow \quad \mathbb{P}_S\big[\widehat{s}_\tau^{(S,p)}(x) \neq s_{U,\tau}(x)\big] \leq \delta.$$

Proof. For

$$\tau \in [\gamma, c_{\mathrm{sm}}\gamma] \cup [C_{\mathrm{lg}}R\sqrt{d}, \infty),$$

Proposition C.2 shows that every admissible rate already lies outside the interval supplied by Theorem B.5. Hence in these regimes we may set

$$\mathcal{J}^{(L^p)}(\tau) := \emptyset$$

without changing the null-coupling conclusion.

On the active range

$$c_{\mathrm{sm}}\gamma \leq \tau \leq C_{\mathrm{lg}}R\sqrt{d},$$

apply Theorem B.5 with

$$\zeta(\tau) = \zeta^{(p)}(\tau), \qquad \theta(\tau) = \theta^{(p)}(\tau),$$

and define

$$\mathcal{J}^{(L^p)}(\tau) := [\kappa_-^{(p)}(\tau), \kappa_+^{(p)}(\tau)].$$

Here

$$\log(1/\zeta^{(p)}(\tau)) \leq CH_{L^p}$$

because $\tau \geq c_{\mathrm{sm}}\gamma$, while the explicit form of $\theta^{(p)}(\tau)$ gives

$$\log\left(\frac{R^2d^2}{\gamma^2\tau^2\theta^{(p)}(\tau)}\right) \leq C_pH_{L^p}$$

uniformly on the active range. Moreover,

$$\log\log(1 + |\mathcal{V}|) = \log(d\log M + O(1)) \leq C(\log d + \log\log M) \leq CH_{L^p},$$

and

$$\log(1/\delta) \leq C\log d \leq CH_{L^p}$$

because $Q_\star \leq d$ and $H_{L^p} \geq \log d$. Thus

$$H_{\mathrm{win}}(\tau) \leq CH_{L^p}$$

uniformly on this range, and Theorem B.5 yields

$$|\mathcal{J}^{(L^p)}(\tau)| \leq C\sqrt{\frac{H_{L^p}}{d}} + C\frac{H_{L^p}}{d}$$

together with the required null-coupling implication. Combining the active and inactive regimes proves the corollary. □

C.2 Proof of Theorem A.1

Proof. Interval and packing. For each $\tau \geq \gamma$, set

$$\widehat{s}_{\tau}^{(S)} := \widehat{s}_{\tau}^{(S,p)},$$

and let $\{\mathcal{J}^{(L^p)}(\tau)\}_{\tau \geq \gamma}$ be the interval family from Corollary C.3. Then assumption (i) of Theorem B.1 holds, and

$$w_{L^p} := \sup_{\tau \geq \gamma} |\mathcal{J}^{(L^p)}(\tau)| \leq C\sqrt{\frac{H_{L^p}}{d}} + C\frac{H_{L^p}}{d}.$$

By Lemma B.13, the admissible rate set $K_d^{(L^p)}$ contains a subset G_p whose points are pairwise separated by more than w_{L^p} and whose cardinality satisfies

$$|G_p| \geq \frac{\log(n_{\max}^{(L^p)}/(2n_{\min}^{(L^p)}))}{\log 2 + 2dw_{L^p}}.$$

Now

$$\log \frac{n_{\max}^{(L^p)}}{2n_{\min}^{(L^p)}} \geq \frac{d}{32}\log M - C_{\min}H_{L^p} - O(1).$$

Using the quantitative bound

$$\sqrt{dH_{L^p}} + H_{L^p} \leq Cc_0\, d\log M,$$

we obtain

$$\log \frac{n_{\max}^{(L^p)}}{2n_{\min}^{(L^p)}} \geq \Big(\frac{1}{32} - CC_{\min}c_0\Big) d\log M - O(1).$$

Choosing $c_0 = c_0(p, \rho)$ sufficiently small gives

$$\log \frac{n_{\max}^{(L^p)}}{2n_{\min}^{(L^p)}} \geq c\, d\log(R/\gamma)$$

for all sufficiently large d, because $\log M \asymp \log(R/\gamma)$. Moreover,

$$\log 2 + 2dw_{L^p} \leq C\big(\sqrt{dH_{L^p}} + H_{L^p}\big).$$

Hence

$$|G_p| \geq c'\, \frac{d\log(R/\gamma)}{\sqrt{dH_{L^p}} + H_{L^p}}.$$

Since

$$Q_\star \leq c_0\, \frac{d\log(R/\gamma)}{\sqrt{dH_{L^p}} + H_{L^p}},$$

choosing $c_0 = c_0(p, \rho)$ sufficiently small ensures

$$|G_p| \geq \frac{80Q_\star}{\rho^2}.$$

Base-noise separation. Set

$$A^{(p)}(S) := G_\gamma(S),$$

where the good set is defined using the threshold $\Lambda_\gamma(\zeta^{(p)}(\gamma))$. By the mass-coverage property of $G_\gamma(S)$,

$$\pi_{S,\gamma}(A^{(p)}(S)) \geq 1 - \zeta^{(p)}(\gamma).$$

Since

$$\zeta^{(p)}(\gamma) \leq \Big(\frac{\varepsilon_{\mathrm{err}}\gamma}{4R\sqrt{d}}\Big)^p \to 0,$$

we have $\zeta^{(p)}(\gamma) \le \rho/2$ for all sufficiently large d, and therefore

$$\pi_{S,\gamma}(A^{(p)}(S)) \ge 1 - \rho/2.$$

For the pointwise overlap bound, fix $x \in \mathbb{R}^d$ and an admissible size n. If $x \in A^{(p)}(S)$, then some $y \in S$ satisfies

$$L_\gamma(y, x) \ge \Lambda_\gamma(\zeta^{(p)}(\gamma)).$$

A union bound and Markov's inequality therefore give

$$\mathbb{P}_S[x \in A^{(p)}(S)] \le \frac{n}{\Lambda_\gamma(\zeta^{(p)}(\gamma))} \le \frac{n_{\max}^{(L^p)}}{\Lambda_\gamma(\zeta^{(p)}(\gamma))}.$$

Moreover,

$$\log(1/\zeta^{(p)}(\gamma)) \le C H_{L^p}.$$

Hence Lemma B.9 at $\tau = \gamma$, Lemma B.12, and the quantitative bound

$$\sqrt{d H_{L^p}} + H_{L^p} \le C c_0\, d \log M$$

give

$$\log \Lambda_\gamma\big(\zeta^{(p)}(\gamma)\big) \ge c_I d \log M - C\big(\sqrt{d H_{L^p}} + H_{L^p}\big) \ge \big(c_I - C c_0\big) d \log M.$$

Since

$$\log n_{\max}^{(L^p)} = \frac{d}{32} \log M + O(1)$$

and $c_I > 1/16$, choosing $c_0 = c_0(p, \rho)$ sufficiently small yields

$$c_I - C c_0 > \frac{1}{32}.$$

It follows that

$$\frac{n_{\max}^{(L^p)}}{\Lambda_\gamma(\zeta^{(p)}(\gamma))} \le \frac{\rho^2}{8}$$

for all sufficiently large d. Thus the separating-set hypothesis in Theorem B.1 holds for the map $A^{(p)}$ and the packing set G_p.

Passage to the original σ-indexed model. Let $\mathcal{D}_0$ be the distribution over instances (n, S) from Theorem B.1. For each such instance define

$$\pi^{(S,\gamma)} := \pi_{S,\gamma} = \nu_S * \mathcal{N}(0, \gamma^2 I_d), \qquad \widehat{s}_\sigma^{(S,\gamma)}(x) := \widehat{s}_{\tau(\sigma)}^{(S)}(x), \qquad \tau(\sigma) := \sqrt{\gamma^2 + \sigma^2}.$$

Because ν_S is supported on $\mathcal{V} \subset [-R, R]^d$, the target $\pi^{(S,\gamma)}$ is of the bounded-plus-noise form considered in Theorem A.1. Moreover,

$$(\pi^{(S,\gamma)})_\sigma = \nu_{S,\tau(\sigma)}, \qquad s_{\pi^{(S,\gamma)},\sigma} = s_{S,\tau(\sigma)}.$$

Hence Lemma C.1 gives, for every $\sigma > 0$,

$$\mathbb{E}_{X \sim (\pi^{(S,\gamma)})_\sigma}\big[\|\widehat{s}_\sigma^{(S,\gamma)}(X) - s_{\pi^{(S,\gamma)},\sigma}(X)\|_2^p\big] \le \frac{\varepsilon_{\mathrm{err}}^p}{\tau(\sigma)^p} \le \frac{\varepsilon_{\mathrm{err}}^p}{\sigma^p},$$

and

$$\mathrm{Lip}\left(\widehat{s}_\sigma^{(S,\gamma)}\right) \le \frac{3}{\tau(\sigma)^2} + \frac{7R^2 d}{\tau(\sigma)^4}.$$

By Proposition B.6(d), the true score obeys the same Lipschitz bound up to smaller constants. Thus the family

$$\big(\pi^{(S,\gamma)}, \{\widehat{s}_\sigma^{(S,\gamma)}\}_{\sigma > 0}\big)$$

has the properties asserted in items (1)–(3) of Theorem A.1.

Now fix an adaptive algorithm $\mathcal{A}$ obeying the query budget in Theorem A.1, and construct $\mathcal{A}^\sharp$ by replacing each query (σ, x) with $(\tau(\sigma), x)$. For every instance (n, S), the two runs induce the same joint law of queries, oracle answers, and final output. Let $\mathcal{D}$ be the induced law of

$$\big(\pi^{(S,\gamma)}, \{\widehat{s}_\sigma^{(S,\gamma)}\}_{\sigma > 0}\big)$$

when $(n, S) \sim \mathcal{D}_0$. Taking the constant in Theorem A.1 to be c_0, the algorithm $\mathcal{A}^\sharp$ makes at most $Q_\star$ queries. Theorem B.1 therefore gives

$$\mathbb{P}_{(\pi, \{\widehat{s}_\sigma\}) \sim \mathcal{D}}\Big[\mathrm{d}_{\mathrm{TV}}\left(\mathcal{L}(\widehat{X}), \pi\right) \ge 1 - \rho\Big] \ge 1 - \rho.$$

This proves Theorem A.1. ☐

Deriving Theorem 1.3. Under Assumptions 1 and 2, the ratio R/γ is fixed and $\varepsilon_{\mathrm{err}} = d^{-O(1)}$. Hence there exists $A \ge 0$ such that $\varepsilon_{\mathrm{err}} \in [d^{-A}, 1]$ for all sufficiently large d, and

$$H_{L^p} = \log(d/\varepsilon_{\mathrm{err}}) + \log(R/\gamma) = O(\log d).$$

Applying Theorem A.1 with this exponent A, we obtain

$$Q = \Omega\left(\frac{d}{\sqrt{d\log d} + \log d}\right) = \Omega\left(\sqrt{\frac{d}{\log d}}\right).$$

This yields Theorem 1.3.

D PROOF OF THE PARAMETERIZED ψ_1 LOWER BOUND

Theorem D.1 (Parameterized ψ_1 lower bound)**.** *Fix $\rho \in (0, 1/4)$. Then there exists a constant $c = c(\rho) > 0$ such that the following holds for every $R > 0$, every $\gamma \in (0, R/2)$, every fixed $\varepsilon_{\mathrm{err}} \in (0, 1]$, and all sufficiently large d. There exists a distribution $\mathcal{D}$ over pairs $(\pi, \{\widehat{s}_\sigma\}_{\sigma>0})$ on $\mathbb{R}^d$ such that:*

(1) ***(bounded-plus-noise)*** *$\pi = \pi_{\mathrm{pre}} * \mathcal{N}(0, \gamma^2 I_d)$ for some π_{pre} supported on $[-R, R]^d$;*

(2) ***(ψ_1-accurate score oracle)*** *for every $\sigma > 0$ and every $z \ge 0$,*

$$\mathbb{P}_{X\sim\pi_\sigma}\left[\|\widehat{s}_\sigma(X) - s_{\pi,\sigma}(X)\|_2 \ge z\right] \le 2\exp\left(-\frac{z\sigma}{\varepsilon_{\mathrm{err}}}\right);$$

(3) ***(global Lipschitzness)*** *for every $\sigma > 0$, both $s_{\pi,\sigma}$ and $\widehat{s}_\sigma$ are globally L_σ-Lipschitz with*

$$L_\sigma \le \frac{3}{\gamma^2 + \sigma^2} + \frac{7R^2 d}{(\gamma^2 + \sigma^2)^2};$$

(4) ***(query lower bound)*** *every adaptive algorithm making at most*

$$Q \le c \cdot \min\left\{d,\ \frac{d\,\log(R/\gamma)}{\sqrt{dH_{\psi_1}} + H_{\psi_1}}\right\}, \qquad H_{\psi_1} := \log d + \log(R/\gamma) + \frac{R}{\gamma}\frac{\sqrt{d}}{\varepsilon_{\mathrm{err}}},$$

score queries and outputting $\widehat{X}$ satisfies

$$\mathbb{P}_{(\pi,\{\widehat{s}_\sigma\})\sim\mathcal{D}}\Big[\mathrm{d_{TV}}(\mathcal{L}(\widehat{X}), \pi) \ge 1 - \rho\Big] \ge 1 - \rho.$$

Fix $\rho \in (0, 1/4)$ and parameters $R > 0$, $\gamma \in (0, R/2)$, and $\varepsilon_{\mathrm{err}} \in (0, 1]$. For all sufficiently large d, set

$$H_{\psi_1} := \log d + \log(R/\gamma) + \frac{R}{\gamma}\frac{\sqrt{d}}{\varepsilon_{\mathrm{err}}},$$

let $c_0 = c_0(\rho) > 0$ be a sufficiently small constant, and define

$$Q_\star := \left\lfloor c_0 \min\left\{d,\ \frac{d\,\log(R/\gamma)}{\sqrt{dH_{\psi_1}} + H_{\psi_1}}\right\}\right\rfloor.$$

We work in the nontrivial regime $Q_\star \ge 1$ and set $\delta := \rho^2/(80Q_\star)$. Fix an auxiliary exponent $C_{\min} \ge 1$, and define

$$n_{\min}^{(\psi_1)} := \lceil d^{C_{\min}} \rceil, \qquad n_{\max}^{(\psi_1)} := \left\lfloor M^{d/32} \right\rfloor, \qquad \kappa_{\min}^{(\psi_1)} := \kappa(n_{\min}^{(\psi_1)}), \qquad \kappa_{\max}^{(\psi_1)} := \kappa(n_{\max}^{(\psi_1)}),$$

and

$$K_d^{(\psi_1)} := \{\kappa(n) :\ n \in [n_{\min}^{(\psi_1)}, n_{\max}^{(\psi_1)}] \cap \mathbb{N}\}.$$

Since $Q_\star \ge 1$, after taking d sufficiently large we have

$$\frac{d\log(R/\gamma)}{\sqrt{dH_{\psi_1}} + H_{\psi_1}} \ge \frac{1}{c_0},$$

and therefore, because $\log M \asymp \log(R/\gamma)$,

$$\sqrt{dH_{\psi_1}} + H_{\psi_1} \le Cc_0\, d\log M,$$

where $C > 0$ is an absolute constant.

D.1 THE FIXED-NOISE ψ_1 ORACLE

For $\tau \geq \gamma$ define

$$\zeta^{(\psi)}(\tau) := \min\left\{\frac{1}{4},\ 2\exp\left(-\frac{2R\sqrt{d}}{\tau\varepsilon_{\text{err}}}\right)\right\}, \qquad \theta^{(\psi)}(\tau) := \frac{R^2 d}{4\tau^4}\exp\left(-\frac{2R\sqrt{d}}{\tau\varepsilon_{\text{err}}}\right),$$

and introduce the high-noise cutoff

$$\tau_\star := \frac{4R\sqrt{d}}{\varepsilon_{\text{err}}}.$$

For $\tau \in [\gamma, \tau_\star)$, let $\widehat{s}_\tau^{(S,\psi)}$ denote the oracle defined in Section B.2 with $(\zeta, \theta) = (\zeta^{(\psi)}, \theta^{(\psi)})$. For $\tau \geq \tau_\star$, define

$$\widehat{s}_\tau^{(S,\psi)} := s_{U,\tau}.$$

Proposition D.2 (Fixed-noise ψ_1 properties)**.** *For every $\tau \geq \gamma$, the oracle $\widehat{s}_\tau^{(S,\psi)}$ satisfies*

$$\mathbb{P}_{X\sim\nu_{S,\tau}}\left[\|\widehat{s}_\tau^{(S,\psi)}(X) - s_{S,\tau}(X)\|_2 \geq z\right] \leq 2\exp\left(-\frac{z\tau}{\varepsilon_{\text{err}}}\right) \qquad \textit{for all } z \geq 0,$$

and

$$\operatorname{Lip}\left(\widehat{s}_\tau^{(S,\psi)}\right) \leq \frac{3}{\tau^2} + \frac{7R^2 d}{\tau^4}.$$

Proof. Set

$$\Delta_\tau(x) := s_{S,\tau}(x) - s_{U,\tau}(x), \qquad B_\tau := \frac{2R\sqrt{d}}{\tau^2}, \qquad \lambda := \frac{\tau}{\varepsilon_{\text{err}}}.$$

By Proposition B.6(c),

$$\|\Delta_\tau(x)\|_2 \leq B_\tau \qquad \forall x \in \mathbb{R}^d.$$

We first verify the ψ_1 tail bound.

High-noise regime. If $\tau \geq \tau_\star$, then by definition

$$\widehat{s}_\tau^{(S,\psi)} = s_{U,\tau}.$$

Thus the oracle error is just $\Delta_\tau(X)$. Since

$$\lambda B_\tau = \frac{\tau}{\varepsilon_{\text{err}}} \cdot \frac{2R\sqrt{d}}{\tau^2} = \frac{2R\sqrt{d}}{\tau\varepsilon_{\text{err}}} \leq \frac{1}{2},$$

we have: if $z > B_\tau$, then the left-hand side is zero; if $0 \leq z \leq B_\tau$, then

$$2e^{-\lambda z} \geq 2e^{-\lambda B_\tau} \geq 2e^{-1/2} > 1.$$

Hence

$$\mathbb{P}\left[\|\widehat{s}_\tau^{(S,\psi)}(X) - s_{S,\tau}(X)\|_2 \geq z\right] \leq 2e^{-\lambda z}$$

for all $z \geq 0$.

Low-noise null regime. Assume $\tau < \tau_\star$ and $J_\tau(S) \leq \theta^{(\psi)}(\tau)$. Then the oracle is again $s_{U,\tau}$, so the error is $\Delta_\tau(X)$. We use the following elementary claim.

Claim. If W is nonnegative, $W \leq B$ almost surely, and

$$\mathbb{E}[W^2] \leq \frac{B^2}{16}e^{-\lambda B},$$

then

$$\mathbb{P}[W \geq z] \leq 2e^{-\lambda z} \qquad \forall z \geq 0.$$

Proof of claim. For $t \in [0, B]$,

$$e^{\lambda t} \leq 1 + \lambda t + \frac{e^{\lambda B}}{B^2}t^2$$

by convexity. Therefore

$$\mathbb{E}[e^{\lambda W}] \le 1 + \lambda \mathbb{E}[W] + \frac{e^{\lambda B}}{B^2}\mathbb{E}[W^2] \le 1 + \lambda\sqrt{\mathbb{E}[W^2]} + \frac{1}{16}.$$

Using the assumed bound on $\mathbb{E}[W^2]$ and the elementary inequality $ue^{-u/2} \le 2e^{-1} < 1$ for $u \ge 0$, we obtain

$$\lambda\sqrt{\mathbb{E}[W^2]} \le \frac{\lambda B}{4} e^{-\lambda B/2} \le \frac{1}{2}.$$

Hence $\mathbb{E}[e^{\lambda W}] \le 2$, and Markov's inequality proves the claim.

Applying the claim to

$$W := \|\Delta_\tau(X)\|_2, \qquad B := B_\tau,$$

and using

$$\mathbb{E}[W^2] = J_\tau(S) \le \theta^{(\psi)}(\tau) = \frac{B_\tau^2}{16} e^{-\lambda B_\tau},$$

gives the required tail bound.

Masked regime. Assume finally that $\tau < \tau_\star$ and $J_\tau(S) > \theta^{(\psi)}(\tau)$. Then

$$\widehat{s}_\tau^{(S,\psi)}(x) - s_{S,\tau}(x) = -(1 - m_{\tau,S}(x))\Delta_\tau(x).$$

If $z = 0$, then

$$\mathbb{P}\big[\|\widehat{s}_\tau^{(S,\psi)}(X) - s_{S,\tau}(X)\|_2 \ge 0\big] = 1 \le 2.$$

Assume therefore that $z > 0$. Since $m_{\tau,S}(x) = 1$ for every $x \in G_\tau(S)$, the event

$$\big\{\|\widehat{s}_\tau^{(S,\psi)}(X) - s_{S,\tau}(X)\|_2 \ge z\big\}$$

can occur only when $X \notin G_\tau(S)$. Hence

$$\mathbb{P}\big[\|\widehat{s}_\tau^{(S,\psi)}(X) - s_{S,\tau}(X)\|_2 \ge z\big] \le \mathbb{P}[X \notin G_\tau(S)] \le \zeta^{(\psi)}(\tau).$$

If $z > B_\tau$ then the left-hand side is zero. If $0 < z \le B_\tau$, then

$$\zeta^{(\psi)}(\tau) \le 2e^{-\lambda B_\tau} \le 2e^{-\lambda z}$$

by the definition of $\zeta^{(\psi)}(\tau)$. This proves the ψ_1-accuracy statement.

The Lipschitz bound is immediate from Proposition B.2 when $\tau < \tau_\star$, and from Proposition B.6(d) when $\tau \ge \tau_\star$. □

Corollary D.3 (Rate intervals in the ψ_1 case)**.** *Under the standing assumptions of this section, there exists a family of intervals $\{\mathcal{J}^{(\psi_1)}(\tau)\}_{\tau \ge \gamma}$ such that*

$$\sup_{\tau \ge \gamma} |\mathcal{J}^{(\psi_1)}(\tau)| \le C\sqrt{\frac{H_{\psi_1}}{d}} + C\frac{H_{\psi_1}}{d},$$

and for every $\tau \ge \gamma$, every fixed query point $x \in \mathbb{R}^d$, and every integer $n \in [n_{\min}^{(\psi_1)}, n_{\max}^{(\psi_1)}]$,

$$\kappa(n) \notin \mathcal{J}^{(\psi_1)}(\tau) \quad \Longrightarrow \quad \mathbb{P}_S\big[\widehat{s}_\tau^{(S,\psi)}(x) \ne s_{U,\tau}(x)\big] \le \delta.$$

Proof. For $\tau \ge \tau_\star$, the oracle is identically $s_{U,\tau}$ by definition, so in this regime we may take

$$\mathcal{J}^{(\psi_1)}(\tau) := \emptyset.$$

On the active range

$$\gamma \le \tau < \tau_\star,$$

apply Theorem B.5 with

$$\zeta(\tau) = \zeta^{(\psi)}(\tau), \qquad \theta(\tau) = \theta^{(\psi)}(\tau),$$

and define

$$\mathcal{J}^{(\psi_1)}(\tau) := [\kappa_-^{(\psi)}(\tau), \kappa_+^{(\psi)}(\tau)].$$

Here

$$\log(1/\zeta^{(\psi)}(\tau)) \le \log 4 + \frac{2R\sqrt{d}}{\tau\varepsilon_{\mathrm{err}}} \le CH_{\psi_1}.$$

Also,

$$\frac{R^2 d^2}{\gamma^2\tau^2\theta^{(\psi)}(\tau)} = \frac{4d\tau^2}{\gamma^2}\exp\left(\frac{2R\sqrt{d}}{\tau\varepsilon_{\mathrm{err}}}\right),$$

so, because $\gamma \le \tau < \tau_\star = 4R\sqrt{d}/\varepsilon_{\mathrm{err}}$,

$$\log\left(\frac{R^2 d^2}{\gamma^2\tau^2\theta^{(\psi)}(\tau)}\right) \le CH_{\psi_1}.$$

Finally,

$$\log\log(1+|\mathcal{V}|) = \log(d\log M + O(1)) \le C(\log d + \log\log M) \le CH_{\psi_1},$$

and

$$\log(1/\delta) \le C\log d \le CH_{\psi_1}$$

because $Q_\star \le d$. Thus

$$H_{\mathrm{win}}(\tau) \le CH_{\psi_1}$$

uniformly on the active range, and Theorem B.5 gives

$$|\mathcal{J}^{(\psi_1)}(\tau)| \le C\sqrt{\frac{H_{\psi_1}}{d}} + C\frac{H_{\psi_1}}{d}$$

together with the required null-coupling implication. Combining the active and inactive regimes proves the corollary. □

D.2 PROOF OF THEOREM D.1

Proof. Interval and packing. For each $\tau \ge \gamma$, set

$$\widehat{s}_\tau^{(S)} := \widehat{s}_\tau^{(S,\psi)},$$

and let $\{\mathcal{J}^{(\psi_1)}(\tau)\}_{\tau\ge\gamma}$ be the interval family from Corollary D.3. Then assumption (i) of Theorem B.1 holds, and

$$w_{\psi_1} := \sup_{\tau\ge\gamma}|\mathcal{J}^{(\psi_1)}(\tau)| \le C\sqrt{\frac{H_{\psi_1}}{d}} + C\frac{H_{\psi_1}}{d}.$$

By Lemma B.13, the admissible rate set $K_d^{(\psi_1)}$ contains a subset G_ψ whose points are pairwise separated by more than w_{ψ_1} and whose cardinality satisfies

$$|G_\psi| \ge \frac{\log(n_{\max}^{(\psi_1)}/(2n_{\min}^{(\psi_1)}))}{\log 2 + 2dw_{\psi_1}}.$$

As above,

$$\log\frac{n_{\max}^{(\psi_1)}}{2n_{\min}^{(\psi_1)}} \ge c\,d\log(R/\gamma)$$

for all sufficiently large d, while

$$\log 2 + 2dw_{\psi_1} \le C\big(\sqrt{dH_{\psi_1}} + H_{\psi_1}\big).$$

Hence

$$|G_\psi| \ge c'\,\frac{d\log(R/\gamma)}{\sqrt{dH_{\psi_1}} + H_{\psi_1}}.$$

Since

$$Q_\star \le c_0\,\frac{d\log(R/\gamma)}{\sqrt{dH_{\psi_1}} + H_{\psi_1}},$$

choosing $c_0 = c_0(\rho)$ sufficiently small ensures

$$|G_\psi| \geq \frac{80Q_\star}{\rho^2}.$$

Base-noise separation. Set

$$A^{(\psi)}(S) := G_\gamma(S),$$

where the threshold is $\Lambda_\gamma(\zeta^{(\psi)}(\gamma))$. Exactly as in the L^p proof,

$$\pi_{S,\gamma}(A^{(\psi)}(S)) \geq 1 - \zeta^{(\psi)}(\gamma), \qquad \mathbb{P}_S[x \in A^{(\psi)}(S)] \leq \frac{n_{\max}^{(\psi_1)}}{\Lambda_\gamma(\zeta^{(\psi)}(\gamma))}$$

for every fixed $x \in \mathbb{R}^d$. Since

$$\zeta^{(\psi)}(\gamma) \leq 2\exp\left(-\frac{2R\sqrt{d}}{\gamma\varepsilon_{\mathrm{err}}}\right) \to 0,$$

we have $\zeta^{(\psi)}(\gamma) \leq \rho/2$ for all sufficiently large d. Moreover, by Lemma B.9 at $\tau = \gamma$, Lemma B.12, and the quantitative bound

$$\sqrt{dH_{\psi_1}} + H_{\psi_1} \leq Cc_0\, d\log M,$$

we obtain

$$\log\Lambda_\gamma\big(\zeta^{(\psi)}(\gamma)\big) \geq c_I d\log M - C\big(\sqrt{dH_{\psi_1}} + H_{\psi_1}\big) \geq \big(c_I - Cc_0\big)d\log M.$$

Since

$$\log n_{\max}^{(\psi_1)} = \frac{d}{32}\log M + O(1)$$

and $c_I > 1/16$, choosing $c_0 = c_0(\rho)$ sufficiently small yields

$$c_I - Cc_0 > \frac{1}{32}.$$

It follows that

$$\frac{n_{\max}^{(\psi_1)}}{\Lambda_\gamma(\zeta^{(\psi)}(\gamma))} \leq \frac{\rho^2}{8}$$

for all sufficiently large d. Thus the separating-set hypothesis in Theorem B.1 holds for the map $A^{(\psi)}$ and the packing set G_ψ.

Passage to the original σ-indexed model. The transfer from the τ-indexed oracle to the original σ-indexed model is identical to the L^p proof, so we record only the ψ_1-specific oracle bound. Let $\mathcal{D}_0$ be the distribution over instances (n, S) from Theorem B.1, and for each such instance define

$$\pi^{(S,\gamma)} := \pi_{S,\gamma} = \nu_S * \mathcal{N}(0, \gamma^2 I_d), \qquad \widehat{s}_\sigma^{(S,\gamma)}(x) := \widehat{s}_{\tau(\sigma)}^{(S)}(x), \qquad \tau(\sigma) := \sqrt{\gamma^2 + \sigma^2}.$$

Then

$$(\pi^{(S,\gamma)})_\sigma = \nu_{S,\tau(\sigma)}, \qquad s_{\pi^{(S,\gamma)},\sigma} = s_{S,\tau(\sigma)}.$$

Hence Proposition D.2 gives, for every $\sigma > 0$ and every $z \geq 0$,

$$\mathbb{P}_{X\sim(\pi^{(S,\gamma)})_\sigma}\big[\|\widehat{s}_\sigma^{(S,\gamma)}(X) - s_{\pi^{(S,\gamma)},\sigma}(X)\|_2 \geq z\big] \leq 2\exp\left(-\frac{z\tau(\sigma)}{\varepsilon_{\mathrm{err}}}\right) \leq 2\exp\left(-\frac{z\sigma}{\varepsilon_{\mathrm{err}}}\right),$$

and both $\widehat{s}_\sigma^{(S,\gamma)}$ and $s_{\pi^{(S,\gamma)},\sigma}$ satisfy the Lipschitz bound stated in Theorem D.1. Thus the family

$$\big(\pi^{(S,\gamma)}, \{\widehat{s}_\sigma^{(S,\gamma)}\}_{\sigma>0}\big)$$

has the properties asserted in items (1)–(3) of Theorem D.1.

Fix an adaptive algorithm $\mathcal{A}$ obeying the query budget in Theorem D.1, and define $\mathcal{A}^\sharp$ exactly as in the proof of Theorem A.1. The same transfer argument shows that the two runs induce the same joint law of queries, oracle answers, and final output on every instance. Let $\mathcal{D}$ be the induced law of

$$\big(\pi^{(S,\gamma)}, \{\widehat{s}_\sigma^{(S,\gamma)}\}_{\sigma>0}\big)$$

when $(n, S) \sim \mathcal{D}_0$. Taking the constant in Theorem D.1 to be c_0, the algorithm $\mathcal{A}^\sharp$ makes at most $Q_\star$ queries. Theorem B.1 therefore gives

$$\mathbb{P}_{(\pi,\{\widehat{s}_\sigma\})\sim\mathcal{D}}\Big[\mathrm{d_{TV}}\left(\mathcal{L}(\widehat{X}), \pi\right) \geq 1 - \rho\Big] \geq 1 - \rho.$$

This proves Theorem D.1. $\square$

Deriving Theorem 1.5. Under Assumptions 1 and 3, the ratio R/γ and the parameter $\varepsilon_{\rm err}$ are fixed constants. Consequently,

$$H_{\psi_1} = \log d + \log(R/\gamma) + \frac{R}{\gamma}\frac{\sqrt{d}}{\varepsilon_{\rm err}} = \Theta(\sqrt{d}).$$

Theorem D.1 therefore gives

$$Q = \Omega\left(\frac{d}{\sqrt{d\sqrt{d}} + \sqrt{d}}\right) = \Omega(d^{1/4}).$$

This yields Theorem 1.5.