# LEARNING BETWEEN THE PEAKS: SHARP ASYMPTOTICS FOR KERNEL RIDGE REGRESSION UNDER POWER-LAW ANISOTROPY

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

We study Kernel Ridge Regression for polynomial inner-product kernels under anisotropic Gaussian data, where the eigenvalues of the input covariance follow a power-law decay with exponent $\alpha \geq 0$. In the polynomial high-dimensional regime $n = \Theta(d^\kappa)$, we derive asymptotically sharp expressions for the kernel spectrum and the generalization error, revealing how anisotropy reshapes the learning curves. For weak anisotropy ($0 < \alpha < 1$), the problem remains effectively high-dimensional and retains some features of the isotropic case, while departing from it in others: the variance still peaks at integer sample complexities $\kappa \in \mathbb{N}$, though these peaks are progressively flattened as $\alpha \to 1^-$. Meanwhile, for targets strongly aligned with the data's principal directions, the bias drops at fractional sample complexities $\kappa = m(1-\alpha), m \in \mathbb{N}$. This early transition effectively decouples the bias from the interpolation peaks. For strong anisotropy ($\alpha > 1$), the effective dimension is constant. The variance stops depending on sample size, plateauing under ridgeless interpolation or vanishing at an explicit rate under fixed ridge penalty. The bias undergoes a sharp transition governed by the target's decay rate: below a threshold, learning is abrupt rather than gradual; above it, the bias decays as a power law, recovering classical source and capacity rates. We finally specialize these results to single-index targets, showing how the alignment of the index with the data's principal directions determines the effect of anisotropy on learning. Together, our results clarify how the input geometry shapes the kernel features and fundamentally impacts its generalization properties.

## 1 INTRODUCTION AND SETTING

Over the last decade, fundamental connections between kernel methods and deep neural networks have been discovered, establishing kernels as analytically tractable proxies for understanding deep learning (Belkin et al., 2018). A prime example of this connection is the Neural Tangent Kernel (NTK) (Jacot et al., 2018; Chizat et al., 2019), which rigorously characterizes infinitely wide networks and underpins much of our modern theoretical understanding of deep learning (Bartlett et al., 2020; Ghorbani et al., 2020; Mallinar et al., 2022). More recently, this connection has gained further relevance in the context of *neural scaling laws*, the empirical observation that the performance of large-scale networks improves in a predictable, power-law fashion (Kaplan et al., 2020; Bahri et al., 2024; Atanasov et al., 2026). A similar scaling behavior emerges in kernel regression as well (Bahri et al., 2024; Maloney et al., 2022; Atanasov et al., 2026). However, while classical kernel theory can perfectly capture these power-law learning curves, it typically relies on abstract *capacity conditions* (Caponnetto & De Vito, 2007). Understanding how the input data geometry explicitly shapes this scaling behavior is a question that only very recently started to be addressed (Wortsman & Loureiro, 2025; Karkada et al., 2026; Paquette et al., 2026), with the consequences for learning remaining a largely open question.

In this work, we build on this recent progress to provide a sharp description of the generalization properties of Kernel Ridge Regression under anisotropic data. More precisely, our **main contributions** are:



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

- **Sharp spectrum characterization.** For polynomial inner-product kernels under power-law anisotropic data, we provide an exact asymptotic characterization of the associated kernel integral operator's spectrum.
- **Weak Anisotropy ($0 < \alpha < 1$).** We derive exact asymptotic formulas for bias and variance deterministic equivalents. We discover fractional learning thresholds: for polynomial targets well aligned with principal data directions, a degree-$m$ component becomes learnable at *fractional* complexities $n = \Theta(d^{m(1-\alpha)})$, bypassing the classical $n = \Theta(d^m)$ isotropic threshold. Strikingly, these bias transitions decouple from the variance interpolation peaks, which remain at integer powers of $d$ but progressively dampen as $\alpha \in (0, 1)$ increases.
- **Strong Anisotropy ($\alpha > 1$).** Here the effective dimension of the problem is finite. The variance becomes independent of $n$, either vanishing or plateauing to a constant depending on the explicit ridge regularization. The bias undergoes a sharp transition governed by the target's decay: below a threshold, a given target degree is learned only abruptly (or, under regularization, gradually but at a delayed sample complexity); above it, the bias decays smoothly as a power-law, recovering the source & capacity rates in the literature.

### Further related work

**Source and capacity conditions.** Classical KRR guarantees and their extensions (e.g., random features, structured prediction) typically rely on abstract, distribution-free source and capacity conditions (Caponnetto & De Vito, 2007; Cui et al., 2022; Rudi & Rosasco, 2017; Rudi et al., 2015; Cabannes et al., 2021). Our work departs from this literature by explicitly deriving these exponents from the input covariance structure, rather than treating them as free parameters.

**High-dimensional asymptotics.** A vast literature leverages random matrix theory and Gaussian equivalence to derive exact asymptotic and dimension-free non-asymptotic risks for KRR and random features (Ghorbani et al., 2020; Canatar et al., 2021; Bordelon et al., 2020; Xiao et al., 2022; Goldt et al., 2022; Schröder et al., 2024; Lu & Yau, 2025; Cheng & Montanari, 2024; Misiakiewicz & Saeed, 2024; Defilippis et al., 2024). We build on these tools to provide an in-depth study of KRR learning curves under anisotropic inputs.

**Data geometry and the kernel spectrum.** The interplay between input density, data geometry, and spectral decay is well-established for Sobolev kernels on manifolds (Bach, 2017; Rosa, 2025) and translation-invariant kernels (Widom, 1963; Williams & Seeger, 2000; Bach, 2024a). We extend this density-kernel interplay to polynomial kernels in the high-dimensional $n = \Theta(d^\kappa)$ regime.

**Learning under power-law anisotropy.** Closest to our setting, three recent works investigate learning under anisotropic data. Wortsman & Loureiro (2025) provide spectral and risk upper bounds for weak anisotropy; we tighten their spectral bounds to an asymptotic equality, yielding the first complete, closed-form phase diagram across both weak and strong anisotropy regimes. Concurrently, Karkada et al. (2026) introduced a Hermite ansatz for orthogonally invariant kernels, while Paquette et al. (2026) studied random feature spectra focusing on strong anisotropy.

### 1.1 Setting: KRR on Anisotropic Data

We consider a supervised regression task with a dataset $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ of i.i.d. samples drawn from a joint distribution $p_X \cdot p_{Y|X}$ on $\mathbb{R}^d \times \mathbb{R}$. We assume a teacher-student setup with target function $f_\star(\boldsymbol{x}) = \mathbb{E}[y|\boldsymbol{x}]$ and additive noise, such that $y_i = f_\star(\boldsymbol{x}_i) + \varepsilon_i$ with $\mathbb{E}[\varepsilon_i] = 0$. We focus on Kernel Ridge Regression (KRR), where the estimator $\hat{f}_\lambda$ minimizes the unnormalized empirical risk $\sum_{i=1}^n (y_i - f(\boldsymbol{x}_i))^2 + \lambda \|f\|_{\mathcal{H}}^2$ with $\mathcal{H}$ the RKHS. We work under the following assumptions:

**Assumption 1.1** (*Kernel and Target*). (i) The kernel is a polynomial inner-product kernel: $k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{m=0}^D h_m \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^m$ with $h_m \geq 0$ and finite (but large) degree $D$. (ii) The target $f_\star \in L^2(p_X)$ is a polynomial of degree up to $D$. (iii) The label noise $\varepsilon = y - f_\star(\boldsymbol{x})$ is mean-zero, independent, and $\sigma_\varepsilon^2$-sub-Gaussian.

Our goal in this work is to characterize the excess population risk of the kernel predictor in the high-dimensional limit $d \to \infty$. Because the input distribution $p_X$ defines the geometry of the learning problem, it also plays a crucial role in generalization. Hence, we ask: *How does the data geometry $p_X$ dictate the performance of the kernel ridge estimator $\hat{f}_\lambda$?* Answering this requires diagonalizing

the underlying kernel operator $T_k$, defined as (Cucker & Smale, 2001):

$$(T_k f)(\boldsymbol{x}) = \int_{\mathbb{R}^d} k(\boldsymbol{x}, \boldsymbol{x}') f(\boldsymbol{x}') p_X(\mathrm{d}\boldsymbol{x}'). \tag{1}$$

This diagonalization is a challenging mathematical problem that has confined most previous work to isotropic distributions (Ghorbani et al., 2021; Mei et al., 2022; Misiakiewicz, 2022). To capture the anisotropic nature of real-world data (Mallat, 1989; Simoncelli & Olshausen, 2001; Cagnetta et al., 2026), we consider **anisotropic Gaussian inputs** $p_X = \mathcal{N}(0, \boldsymbol{\Sigma})$, where the eigenvalues $\{\sigma_j\}_{j\in[d]}$ of $\boldsymbol{\Sigma}$ decay as a power-law:

$$\sigma_j = r_\alpha(d)^{-1} j^{-\alpha}, \quad j \in [d], \tag{2}$$

where $r_\alpha(d) = \sum_{j=1}^d j^{-\alpha}$ ensures $\operatorname{Tr} \boldsymbol{\Sigma} = \mathbb{E}[\|\boldsymbol{x}\|^2] = 1$. The *data anisotropy* exponent $\alpha \geq 0$ interpolates between isotropy ($\alpha = 0$) and effectively low-dimensional data ($\alpha \gg 1$). We analyze the polynomial high-dimensional regime where $d, n \to \infty$ with $n/d^\kappa \to \psi > 0$ for $\kappa > 0$. While our primary focus is on integer values $\kappa \in \mathbb{N}$, we will occasionally explore cases where $\kappa$ is fractional.

## 2 High-dimensional limit of the kernel spectrum

The statistical properties of the KRR estimator heavily depend on the spectrum of the kernel operator in Eq. (1). Under Assumption 1.1 and for Gaussian data with generic covariance $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_d)$, Wortsman & Loureiro (2025) recently showed that the kernel eigenvalues $\lambda_{\boldsymbol{\beta}}$ (indexed by multi-indices $\boldsymbol{\beta} \in \mathbb{Z}^d_{\geq 0}$) can be bounded by

$$C_1 h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \, \sigma_1^{\beta_1} \cdots \sigma_d^{\beta_d} \leq \lambda_{\boldsymbol{\beta}} \leq C_2 h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \, \sigma_1^{\beta_1} \cdots \sigma_d^{\beta_d}, \tag{3}$$

where $|\boldsymbol{\beta}| = \beta_1 + \cdots + \beta_d$ and $C_1, C_2 > 0$ are constants independent of $d$. Our first result is to show that the above bounds can be made sharper in the high-dimensional limit $d \to \infty$:

**Proposition 1** (*Asymptotic kernel spectrum*)**.** Assume $\operatorname{Tr}(\boldsymbol{\Sigma}^2) = o_d(1)$ and Assumption 1.1. Then, the spectrum of the inner-product kernel indexed by multi-indices $\boldsymbol{\beta} \in \mathbb{Z}^d_{\geq 0}$, $|\boldsymbol{\beta}| \leq D$ satisfies:

$$\lambda_{\boldsymbol{\beta}} = h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \, \sigma_1^{\beta_1} \dots \sigma_d^{\beta_d} (1 + o_d(1)). \tag{4}$$

In particular, this holds for $\alpha \in [0, 1]$ under power-law data in Eq. (2).

We prove Proposition 1 and other properties of the kernel operator in Appendix A. Under the power-law setting in Eq. (2) for $\alpha \leq 1$, the eigenvalues explicitly become at leading order $\lambda_{\boldsymbol{\beta}} = h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \, r_\alpha(d)^{-|\boldsymbol{\beta}|} \prod_{j=1}^d j^{-\alpha\beta_j}$. This spectral structure is fundamentally governed by the normalization constant $r_\alpha(d)$, which acts as the problem's *effective dimension* and whose asymptotic scaling with $\alpha$ partitions the problem into distinct regimes. In particular, in the isotropic limit ($\alpha = 0$), Eq. (4) shows that the spectrum is highly degenerate, with $\Theta(d^m)$ eigenvalues of magnitude $\Theta(d^{-m})$ per $m \in \mathbb{N}$. Crucially, under weak anisotropy ($0 < \alpha < 1$), the effective dimension scales as $r_\alpha(d) = \Theta(d^{1-\alpha})$. In this regime, the spectrum shatters into gaps and disconnected sectors. Because simple capacity conditions fail completely here, the exact spectral characterization established in Proposition 1 is strictly necessary. Finally, for strong anisotropy ($\alpha > 1$), the effective dimension is finite ($r_\alpha = \zeta(\alpha)$); here, Wortsman & Loureiro (2025) showed that the kernel spectrum inherits a capacity condition $\lambda_k = \tilde{\Theta}(k^{-\alpha})$ from the data.

## 3 Generalization properties

To sharply compute the excess risk of the KRR predictor under the anisotropic setting (Section 1.1), we rely on recent progress in the non-asymptotic analysis of kernel methods, pioneered by Cheng & Montanari (2024); Misiakiewicz & Saeed (2024); Defilippis et al. (2024), which we briefly review. Consider the canonical bias-variance decomposition of the noise-averaged excess risk:

$$\mathcal{R}(f_\star, \boldsymbol{X}, \sigma_\varepsilon^2, \lambda) := \mathbb{E}_\varepsilon[\mathcal{R}(f_\star, \boldsymbol{X}, \sigma_\varepsilon^2, \lambda, \boldsymbol{\varepsilon})] = \mathcal{B}(f_\star, \boldsymbol{X}, \lambda) + \mathcal{V}(\boldsymbol{X}, \sigma_\varepsilon^2, \lambda) \tag{5}$$

where $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]^\top \in \mathbb{R}^{n \times d}$ is the data matrix, $\boldsymbol{y} = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top \in \mathbb{R}^n$ are the label and noise vectors and $\mathcal{V}(\boldsymbol{X}, \sigma_\varepsilon^2, \lambda), \mathcal{B}(f_\star, \boldsymbol{X}, \lambda)$ are the variance and bias. Under specific assumptions (see Appendix B for a detailed discussion), the following deterministic approximations of the bias and variance hold with high probability:

$$\begin{aligned} \mathcal{B}(f_\star, \boldsymbol{X}, \lambda) &= \left(1 + \tilde{O}(n^{-A})\right) \mathsf{B}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \lambda), \\ \mathcal{V}(\boldsymbol{X}, \sigma_\varepsilon^2, \lambda) &= \left(1 + \tilde{O}(n^{-B})\right) \mathsf{V}_n(\sigma_\varepsilon^2, \boldsymbol{\Lambda}, \lambda), \end{aligned} \tag{6}$$

where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \dots) \in \mathbb{R}^{\infty \times \infty}$ is the diagonal matrix of the kernel eigenvalues, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \mathbb{R}^\infty$ are the coefficients of the target function expanded in the $L^2(p_X)$ basis of kernel eigenfunctions, and $A, B > 0$. The functions $\mathsf{B}_n, \mathsf{V}_n$, which only depend on these deterministic quantities, are known as *deterministic equivalents* for the bias and variance, and are explicitly given by:

$$\mathsf{V}_n(\sigma_\varepsilon^2, \boldsymbol{\Lambda}, \lambda) = \sigma_\varepsilon^2 \frac{\operatorname{Tr}(\boldsymbol{\Lambda}^2(\boldsymbol{\Lambda} + \nu_\star \boldsymbol{I})^{-2})}{n - \operatorname{Tr}(\boldsymbol{\Lambda}^2(\boldsymbol{\Lambda} + \nu_\star \boldsymbol{I})^{-2})}, \quad \mathsf{B}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \lambda) = \frac{\nu_\star^2 \langle \boldsymbol{\theta}, (\boldsymbol{\Lambda} + \nu_\star \boldsymbol{I})^{-2} \boldsymbol{\theta} \rangle}{1 - \frac{1}{n} \operatorname{Tr}(\boldsymbol{\Lambda}^2(\boldsymbol{\Lambda} + \nu_\star \boldsymbol{I})^{-2})}, \tag{7}$$

where $\boldsymbol{I} \in \mathbb{R}^{\infty \times \infty}$ is the infinite-dimensional identity matrix and the scalar $\nu_\star(\lambda) > 0$ is the unique solution to the following self-consistency equation:

$$n - \frac{\lambda}{\nu} = \operatorname{Tr}(\boldsymbol{\Lambda}(\boldsymbol{\Lambda} + \nu \boldsymbol{I})^{-1}). \tag{8}$$

**Remark 1.** Note that the quantity $\nu_\star(\lambda) \geq 0$ plays a similar role to the $\ell_2$ regularization in the random expressions for the bias and variance. Remarkably, Eq. (8) implies that $\nu_\star(0) > 0$, meaning that the interpolator $\hat{f}_{\lambda=0}$ will not necessarily overfit. For this reason, it is known as the *effective regularization* (Bartlett et al., 2020; Bach, 2024b).

Checking the conditions for the deterministic equivalents in Eq. (7) to hold requires, in principle, diagonalizing the kernel operator, which is a non-trivial task. For this reason, these conditions have only been proven for isotropic distributions in the kernel setting (Misiakiewicz, 2022; Misiakiewicz & Saeed, 2024). Our first result shows that these deterministic equivalents are valid also in the anisotropic power-law setting introduced in Section 1.1.

**Proposition 2** (*Validity of deterministic equivalents*)**.** Consider the KRR setting introduced in Section 1.1. Define the deterministic equivalent for the excess risk:

$$\mathsf{R}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \lambda, \sigma_\varepsilon^2) := \mathsf{B}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \lambda) + \mathsf{V}_n(\sigma_\varepsilon^2, \boldsymbol{\Lambda}, \lambda),$$

where $\mathsf{B}_n, \mathsf{V}_n$ are given by Eq. (7). Under Assumption 1.1, with probability going to 1 as $n$ grows, we have

$$\left| \mathcal{R}(f_\star, \boldsymbol{X}, \sigma_\varepsilon^2, \lambda) - \mathsf{R}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \lambda, \sigma_\varepsilon^2) \right| \leq C_{x,\varepsilon,D,K}\, \mathcal{E}_{R,n}(m)\, \mathsf{R}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \lambda, \sigma_\varepsilon^2),$$

where $\mathcal{E}_{R,n}(m)$ is a relative approximation error defined in Appendix B that satisfies $\mathcal{E}_{R,n}(m) \to 0$ with $n$.

With this result in hand, we can move to our core contribution, which is the study of how anisotropy impacts the generalization properties of the kernel predictor. We start with the weakly anisotropic regime, where $0 < \alpha < 1$. Proofs for the following theorems are deferred to Appendix D.

### 3.1 Weakly anisotropic regime

Our analysis in this regime begins by characterizing the effective regularization $\nu_\star$ through the self-consistency Eq. (8). In particular, we have:

**Lemma 3.1** (*Kernel state equation*)**.** Consider the KRR setting introduced in Section 1.1 for $\alpha \in [0, 1)$. In the limit $d \to \infty$ with $n/d^\kappa \to \psi > 0$ and $\kappa \in \mathbb{N}$, the solution $\nu_\star > 0$ of Eq. (8) takes the form $\nu_\star = \xi_\star d^{-\kappa}$, where $\xi_\star > 0$ is the unique positive root of the following equation:

$$\psi = \frac{\lambda_{\text{eff}}}{\xi} + \frac{1}{\kappa!(\kappa - 1)!} \int_0^\infty e^{-t} t^{\kappa - 1} \frac{h_\kappa \kappa!}{h_\kappa \kappa! + (1 - \alpha)^{-\kappa} \xi e^{-\alpha t}} \mathrm{d}t \tag{9}$$

where $\lambda_{\text{eff}} := \lambda + \sum_{m > \kappa} h_m$ acts as an *effective ridge regularization*.

**Remark 2.** This result mathematically demonstrates why the effective regularization remains strictly positive ($\nu_\star(0) > 0$) even in the ridgeless limit. Specifically, the high-order tails of the kernel Taylor expansion $\sum_{m>\kappa} h_m$ act as an implicit regularization mechanism. This high-degree component of the kernel prevents the interpolator from overfitting and provides a self-regularizing effect exactly as first described for the isotropic case by Misiakiewicz (2022).

Given a solution $\xi_\star \in \mathbb{R}$ of Eq. (9), one can compute $\mathsf{V}_n(\sigma_\varepsilon^2, \mathbf{\Lambda}, \lambda)$ and $\mathsf{B}_n(\boldsymbol{\theta}, \mathbf{\Lambda}, \lambda)$. We start by discussing the variance term.

### 3.1.1 Variance

Plugging the solution $\xi_\star$ into the formula for the deterministic variance $\mathsf{V}_n$, we obtain the following:

**Theorem 1** (*Variance in the weakly anisotropic regime*)**.** Consider KRR in the setting described in Section 1.1 with $0 \le \alpha < 1$. Let $\xi_\star$ be the unique root of the Kernel State Equation defined in Lemma 3.1. Under the polynomial scaling $n = \psi d^\kappa$ (with $\kappa \in \mathbb{N}, \psi = \Theta_d(1)$), the deterministic equivalent for the variance $\mathsf{V}_n$ in the high-dimensional limit is given by:

$$\mathsf{V}_n = \sigma_\varepsilon^2 \frac{\tau}{1-\tau} \tag{10}$$

where $\tau = \tau(\psi, \kappa, \alpha, \lambda, \{h_m\})$ is given by the integral:

$$\tau = \frac{1}{\psi} \frac{1}{\kappa!(\kappa-1)!} \int_0^\infty e^{-t} t^{\kappa-1} \frac{(h_\kappa \kappa!)^2}{(h_\kappa \kappa! + (1-\alpha)^{-\kappa} \xi_\star e^{-\alpha t})^2} \mathrm{d}t \tag{11}$$

For $\kappa \notin \mathbb{N}$, the variance asymptotically vanishes to zero as $\mathsf{V}_n = \Theta(d^{-\gamma_V})$ where:

$$\gamma_V = \begin{cases} \min\left(\kappa - \lfloor \kappa \rfloor,\ \lceil \kappa \rceil - \kappa\right) & \text{if } 0 \le \alpha \le \frac{1}{2} \\ \min\left(\kappa - \lfloor \kappa \rfloor,\ \frac{1-\alpha}{\alpha}(\lceil \kappa \rceil - \kappa)\right) & \text{if } \frac{1}{2} < \alpha < 1 \end{cases} \tag{12}$$

**Remark 3.** The quantity $\tau$ in Eq. (11) is closely tied to the effective dimensionality of the kernel operator. Formally, its value corresponds to

$$\tau = \frac{\mathrm{df}_2(\nu_\star)}{\mathrm{df}_1(\nu_\star) + \frac{\lambda}{\nu_\star}}$$

where $\mathrm{df}_1(\nu)$ and $\mathrm{df}_2(\nu)$ denote respectively the *first and second degrees of freedom* of the kernel eigenvalue matrix (Zhang, 2005). Specifically, $\tau$ quantifies the dissimilarity between these alternative capacity metrics, reducing to $\mathrm{df}_2(\nu_\star)/\mathrm{df}_1(\nu_\star)$ in the ridgeless limit. This ratio ultimately dictates the variance behavior, which explodes when $\mathrm{df}_1 \approx \mathrm{df}_2$ and collapses to zero when $\mathrm{df}_2 \ll \mathrm{df}_1$. Interestingly, as $\alpha \to 1^-$, $\tau$ vanishes independently of $\psi$, implying that the second degree of freedom becomes negligible when the data covariance decays as $\sigma_j \propto j^{-1}$.

Given $\tau$, the behavior of the variance $\mathsf{V}_n$ as a function of $\psi$ at fixed scaling $\kappa \in \mathbb{N}$ can be easily evaluated. Fig. 1 (Left) shows that the variance exhibits a distinct peak at a finite value $\psi^\star > 0$, and systematically vanishes at both extremes of the domain ($\psi \to 0, \psi \to \infty$). This behavior is consistent with the variance vanishing for fractional values of $\kappa$. By horizontally concatenating these curves across different scales $\kappa$, one obtains a global picture given by a *multiple descent* curve. Each variance peak emerges at a sample size $n$ matching the number of degrees of freedom in the next polynomial feature space (i.e., whenever $n \asymp d^\kappa$ for $\kappa \in \mathbb{N}$), while the variance vanishes for intermediate values. This multiple descent pattern is akin to the one shown by Misiakiewicz (2022); Canatar et al. (2021); Liang et al. (2020) for isotropic data, and the fractional decay rates in Eq. (12) are consistent with the results of Liang et al. (2020) in the isotropic limit $\alpha \to 0$.

A striking departure from the isotropic setting is the severe dampening effect of the anisotropy, which ultimately mitigates completely the multiple descent profile. As $\alpha \in [0, 1)$ increases, the variance curve progressively flattens out, a phenomenon driven by two concurrent mechanisms. On one hand, at integer $\kappa$, Fig. 1 (Left) shows that the interpolation peaks are progressively quenched towards zero. On the other hand, for fractional $\kappa$, the second argument of the minimum in Eq. (12) vanishes, yielding $\gamma_V \to 0$; as a consequence, the variance ceases to decay. Together, these two effects close the gap between the integer and fractional regime: the peaks collapse to zero and the deep intermediate valleys disappear, reducing the entire variance curve to a flat line. Indeed, in the strongly anisotropic limit $\alpha \to 1^-$, Eq. (11) implies that $\mathsf{V}_n \to 0$ identically, since $(1-\alpha)^{-\kappa} \to \infty$. This is also a consequence of the fact that, as $\alpha \to 1^-$, $\tau \to 0$ and $\mathrm{df}_2(\nu)$ becomes negligible compared to the first degree of freedom $\mathrm{df}_1(\nu)$.

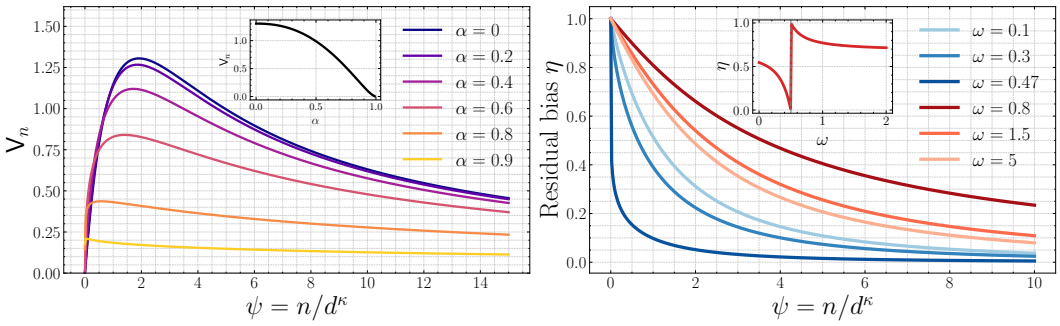


Figure 1: **(Left)** Variance $\mathsf{V}_n$ vs. $\psi = n/d^\kappa$ for varying $\alpha$ ($\kappa = 2$). Inset: $\mathsf{V}_n$ vs. $\alpha$ for fixed $\psi = 2$, showing how $\mathsf{V}_n \to 0$ as $\alpha \to 1^-$. **(Right)** Residual bias $\eta$ vs. $\psi$ for varying $\omega$ ($\kappa_{\text{eff}} = 1, \alpha = 0.5$), quantifying the fraction of the resonant shell unlearned. Inset: Dependence of $\eta$ on $\omega$ at fixed $\kappa_{\text{eff}} = 1, \psi = 1, \alpha = 0.5$. *Setup:* $\sigma_\varepsilon = 4, \lambda = 0$, degree-4 exponential kernel.

### 3.1.2 Bias

We now focus on the bias, whose deterministic equivalent in Eq. (7) depends on the coefficients of the target function $f_\star$ in the $L^2(p_X)$ basis that diagonalises the kernel operator. For polynomial inner-product kernels, this defines a basis of degree $D$ polynomials in $\mathbb{R}^d$, which we denote $(e_{\boldsymbol{\beta}})_{\boldsymbol{\beta} \in \mathbb{Z}^d_{\geq 0}, |\boldsymbol{\beta}| \leq D}$. While one can naturally write $f_\star(x) = \sum_{\boldsymbol{\beta}} \theta_{\boldsymbol{\beta}} e_{\boldsymbol{\beta}}(x)$, working directly with the multi-index $\boldsymbol{\beta}$ is analytically challenging. As detailed in Appendix C, we adopt a more intuitive parametrization: we group the components by their polynomial degree $m = |\boldsymbol{\beta}|$ (the *shell*) and map $\boldsymbol{\beta}$ to a sequence of indices $1 \leq i_1 \leq \cdots \leq i_m \leq d$ where $\beta_j = \sum_{k=1}^m \delta_{i_k, j}$, intuitively listing the active input dimensions forming the corresponding basis tensor. Assuming $f_\star$ is not sparse, we can then extend its coefficients into a completely symmetric tensor $\theta_{i_1,\ldots,i_m}$. In the high-dimensional limit $d \to \infty$, this symmetric extension allows us to rewrite the target function decomposition as:

$$f_\star(\boldsymbol{x}) = \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{i_1,\ldots,i_m=1}^{d} \theta_{i_1,\ldots,i_m} e_{i_1,\ldots,i_m}(\boldsymbol{x}). \tag{13}$$

As a consequence, the final formula for the bias will depend on the specific shape of the completely symmetric tensor $\theta_{i_1,\ldots,i_m}$. In this work, we focus on the following case:

$$\theta_{i_1,\ldots,i_m} = \theta_m (i_1 \ldots i_m)^{-\omega}, \qquad \omega \in \mathbb{R}_{\geq 0}, \tag{14}$$

where $\omega$ represents the *target anisotropy*, and the scalar $\theta_m$ depends only on $m$ and is chosen by imposing the normalization of the target function $||f_\star||_{L^2(p_X)} := \mathbb{E}[f_\star(\boldsymbol{x})^2] = 1$. Notice that, while Eq. (14) resembles a standard *source condition* ($\theta_k = \Theta(k^{-g})$), mapping this multi-dimensional tensor to an exact univariate sequence is non-trivial (Simon, 2026). Denote by $E(m)$ the *spectral energy* of the target function contained in the $m$-th shell, that is:

$$E(m) = \sum_{|\boldsymbol{\beta}|=m} \theta_{\boldsymbol{\beta}}^2 = \frac{1}{m!} \sum_{i_1,\ldots,i_m=1}^{d} \left(\theta_{i_1,\ldots,i_m}\right)^2 = \|\mathrm{P}_m f_\star\|_{L^2}^2$$

where $\mathrm{P}_m$ is the projection onto the subspace spanned by the polynomials of degree $m$ in the kernel orthonormal basis. Under the assumption of Eq. (14), we have the following result for the bias.

**Theorem 2** (*Bias in the weakly anisotropic regime*)**.** Consider the setting from Section 1.1 with target $f_\star \in L^2(p_X)$ satisfying Eq. (14). Let $\xi_\star$ solve Eq. (9) and $\tau$ be given by Eq. (11). Defining the effective degree as $\kappa_{\text{eff}} = \kappa$ if $0 \leq \omega < 1/2$, and $\kappa_{\text{eff}} = \frac{\kappa}{1-\alpha}$ if $\omega > 1/2$, the bias reads:

$$\mathsf{B}_n = \frac{1}{1-\tau} \left( \mathbf{1}_{\{\kappa_{\text{eff}} \in \mathbb{N}\}} \eta \, E(\kappa_{\text{eff}}) + \sum_{m > \kappa_{\text{eff}}} E(m) \right), \tag{15}$$

where the *residual bias fraction* $\eta \in [0, 1]$ is defined for integer $\kappa_{\text{eff}}$ as:

$$\eta = \frac{|1 - 2\omega|^{\kappa_{\text{eff}}}}{(\kappa_{\text{eff}} - 1)!} \int_0^\infty \frac{e^{-|1-2\omega|t} t^{\kappa_{\text{eff}} - 1}}{\left(1 + h_{\kappa_{\text{eff}}} \kappa_{\text{eff}}! \xi_\star^{-1} (1-\alpha)^{\kappa_{\text{eff}}} e^{\operatorname{sgn}(1-2\omega)\alpha t}\right)^2} \, \mathrm{d}t. \tag{16}$$

This theorem provides a detailed picture of how the kernel learns different polynomial degrees of the target function in the weakly anisotropic regime. For small target anisotropies $0 \leq \omega < \frac{1}{2}$, $\kappa_{\text{eff}} = \kappa$ and the kernel can learn at best the degree-$\kappa$ components of the target function with $n = \Theta(d^\kappa)$ samples, regardless of the data anisotropy $\alpha$. The exact fraction of the unlearned degree-$\kappa$ component is quantified by $\eta \in [0, 1]$, which depends on $\psi$. As shown in Fig. 1 (Right, blue lines), as long as $\omega < \frac{1}{2}$, a larger target decay $\omega$ accelerates the learning of this shell. Naturally, in the data-scarce ($\psi \to 0$) and data-rich ($\psi \to \infty$) limits, we recover $\eta \to 1$ and $\eta \to 0^+$, respectively. Concatenating this behavior across integer scales yields a staircase bias curve where successive polynomial degrees are resolved precisely at sample complexities $n = \Theta(d^m)$ for $m \in \mathbb{N}$, akin to the isotropic case (Misiakiewicz, 2022), with the distinction that a larger $\omega$ leads to sharper transitions between consecutive plateaus.

Conversely, for strongly decaying targets ($\omega > 1/2$), the learning dynamics shift dramatically. The kernel is now able to learn features up to the effective order $\kappa_{\text{eff}} = \lfloor \frac{\kappa}{1-\alpha} \rfloor$ with only $n = \Theta(d^\kappa)$ samples. In other words, learning degree-$m$ target features strictly requires $n = \Theta(d^{m(1-\alpha)})$ samples in this regime, drastically improving upon the $n = \Theta(d^m)$ scaling required when $\omega < 1/2$ or $\alpha = 0$. Ultimately, as we approach the strongly anisotropic regime ($\alpha \to 1^-$), this scaling suggests the predictor can resolve all degrees independently of $\kappa$. Zooming in on the transition shell itself, when $\kappa_{\text{eff}} \in \mathbb{N}$, the unlearned fraction is still quantified by $\eta$ in Eq. (16). Specifically, just above the $\omega = 1/2$ threshold, the kernel does access this $\kappa_{\text{eff}}$-th shell, but the actual learned fraction is initially vanishing ($\eta \to 1$). As $\omega$ grows, however, the residual bias strictly decreases (Fig. 1, Right, red lines), and the model learns a progressively larger portion of the resonant shell. Fig. 2 summarizes this discussion on the bias.

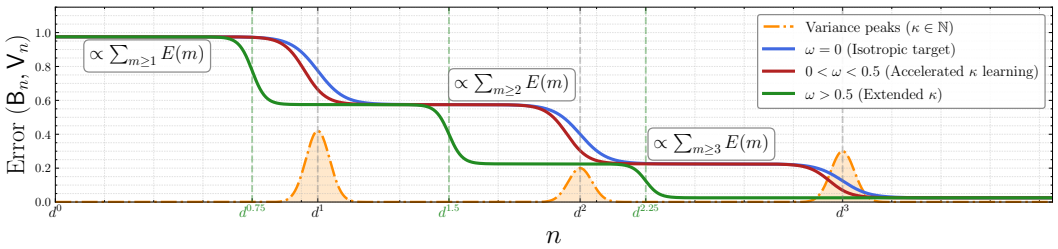


Figure 2: Schematic bias $\mathsf{B}_n$ (solid lines) and variance $\mathsf{V}_n$ (shaded orange region) under weak anisotropy ($\alpha = 0.25$) for varying $\omega$. For $\omega < 1/2$, degree-$m$ target features are resolved at $n \asymp d^m$, with larger $\omega$ accelerating the transition. For $\omega > 1/2$, learning the same features occurs at fractional scales $n \asymp d^{m(1-\alpha)} = d^{0.75m}$, decoupling the bias drops from the variance interpolation peaks and allowing the model to *learn between the peaks*.

The results above highlight a fundamental phenomenon: for strongly aligned targets ($\omega > 1/2$), the bias drops at fractional scales while the variance still peaks at integer scales. This effectively *decouples* signal acquisition from variance peaks. As the variance is by definition independent of the target, this decoupling admits a natural reading in terms of overfitting. From Eq. (7), a direction $\boldsymbol{\beta}$ shifts from contributing to the bias to contributing to the variance as its eigenvalue $\lambda_{\boldsymbol{\beta}}$ crosses the effective regularization $\nu_\star$. In the isotropic case, spectral degeneracy forces an entire degree-$m$ shell through this transition simultaneously, locking signal acquisition and variance costs at $n = \Theta(d^m)$. Anisotropy, instead, spreads the shell's eigenvalues and lifts the degeneracy. If the target aligns with the top principal directions, its components cross the threshold early, resolving the bias at minimal variance cost. The shell's bulk, largely orthogonal to the target, crosses much later at $n = \Theta(d^m)$, producing a massive variance peak uncoupled from any matching gain in bias.

### 3.2 Strongly anisotropic regime

We now turn our attention to the strongly anisotropic regime ($\alpha > 1$), where the effective dimension $r_\alpha(d)$ converges to a finite constant $r_\alpha \to \zeta(\alpha)$ as $d \to \infty$. This reflects the fact that, even though the data is supported in $\mathbb{R}^d$, it is effectively finite dimensional. Proofs are deferred to Appendix E.

### 3.2.1 VARIANCE

**Theorem 3** (*Variance in the strongly anisotropic regime*)**.** Consider the KRR setting of Section 1.1 under strong anisotropy ($\alpha > 1$). In the high-dimensional limit ($d \to \infty$), the deterministic equivalent of the variance exhibits two distinct behaviors depending on the ridge penalty:

$$\mathsf{V}_n(\lambda, \sigma_\varepsilon^2) = \begin{cases} C\sigma_\varepsilon^2(\alpha - 1) & \text{if } \lambda = 0^+ \text{ (ridgeless)}, \\ \tilde{\Theta}\left(d^{-\kappa\left(1-\frac{1}{\alpha}\right)}\right) = \tilde{\Theta}\left(n^{-1+\frac{1}{\alpha}}\right) & \text{if } \lambda \asymp 1 \text{ (regularized)}, \end{cases} \tag{17}$$

where $C = \Theta_d(1)$ is a factor independent of $\psi$ and $\kappa$ (and thus independent of $n$).

Here, we retrieve two expected limits from the source & capacity literature on kernels (Cui et al., 2022; Defilippis et al., 2024), but from a different angle. Indeed, for $\alpha > 1$ the effective dimension of the data $r_\alpha = \Theta_d(1)$ is constant, meaning that the polynomial growth of samples ($n = \psi d^\kappa$) effectively results in an infinitely large dataset compared to the input dimension, and thus we recover the classical large-$n$ asymptotics in the statistical analysis of kernel methods.

### 3.2.2 BIAS

Following the setup of Section 3.1.2, we characterize the target function through the decay rate $\omega \geq 0$ of its eigen-coefficients (Eqs. (13) and (14)). Under strong anisotropy ($\alpha > 1$), the deterministic bias behaves as follows.

**Theorem 4** (*Bias in the strongly anisotropic regime*)**.** Let the target anisotropy be $\omega \geq 0$. Depending on $\lambda$, the bias $\mathsf{B}_n$ exhibits the asymptotic behavior summarized in Table 1 in the limit $d \to \infty$, where $C, C_1, C_2$ are constants and, for integer $\kappa_{\text{eff}} = \kappa/\alpha$, the multiplier $\eta' \in (0, 1)$ is given by:

$$\eta' = \frac{(1-2\omega)^{\kappa_{\text{eff}}}}{(\kappa_{\text{eff}} - 1)!} \int_0^\infty e^{t(2\omega-1)} t^{\kappa_{\text{eff}}-1} \frac{1}{\left(1 + h_{\kappa_{\text{eff}}} \kappa_{\text{eff}}! \zeta(\alpha)^{-\kappa_{\text{eff}}} \psi \lambda^{-1} e^{\alpha t}\right)^2} \, \mathrm{d}t. \tag{18}$$

Table 1: Asymptotic behavior of the deterministic bias in the strongly anisotropic regime.

| **Target anisotropy** ($\omega$) | **Ridgeless** ($\lambda = 0^+$) | **Regularized** ($\lambda \asymp 1$) |
|---|---|---|
| $0 \leq \omega < \frac{1}{2}$ | $C\alpha \sum_{m \geq \kappa} E(m)$ | $C_1 \mathbf{1}_{\left\{\frac{\kappa}{\alpha} \in \mathbb{N}\right\}} \eta' E\left(\frac{\kappa}{\alpha}\right) + C_2 \sum_{m > \frac{\kappa}{\alpha}} E(m)$ |
| $\frac{1}{2} < \omega \leq \alpha + \frac{1}{2}$ | $\tilde{\Theta}\left(n^{-(2\omega-1)}\right)$ | $\Theta\left(n^{-\frac{2\omega-1}{\alpha}}\right)$ |
| $\omega > \alpha + \frac{1}{2}$ | $\tilde{\Theta}\left(n^{-2\alpha}\right)$ | $\Theta\left(n^{-2}\right)$ |

Theorem 4 provides a fine analysis of which target components are resolved by the kernel as the sample complexity increases. For slowly decaying targets ($0 \leq \omega < 1/2$), a finite portion of the target remains unlearned. In the ridgeless limit, the model fails to resolve features of degree $m > \kappa$ but also $m = \kappa$, marking a sharp learning transition compared to the weakly anisotropic case (where the $\kappa$-th degree was partially learned). Indeed, learning the $\kappa$-th degree now requires $n = \Theta(d^{\kappa+\epsilon})$ for an arbitrary $\epsilon > 0$. Physically, the kernel's fast spectral decay limits its effective capacity, preventing it from fitting high-dimensional, slow-decaying targets. Finite regularization ($\lambda \asymp 1$) further restricts learning to degrees $m < \kappa/\alpha$. However, the ridge penalty restores a smooth transition via the fractional multiplier $\eta'$. This delays the learning of degree-$m$ features to a sample complexity $n = \Theta(d^{m\alpha})$, perfectly illustrating the bias-variance trade-off: the explicit penalty suppresses the variance (Theorem 3) at the cost of increased bias.

Conversely, for rapidly decaying targets ($\omega > 1/2$), the target coefficients strongly align with the kernel's spectrum. In this regime, the bias asymptotically vanishes regardless of the regularization $\lambda$, with a convergence rate dictated by $\omega$ that shifts at the crossover $\omega = \alpha + 1/2$. These high-dimensional rates perfectly recover the classical source & capacity bounds established by Cui et al. (2022); Defilippis et al. (2024). A comprehensive visual summary of the various analysed regimes, illustrating the different behaviours for $\alpha \lesseqgtr 1$ and $\omega \lesseqgtr 1/2$ for $\lambda = 0$, is presented in Fig. 3.

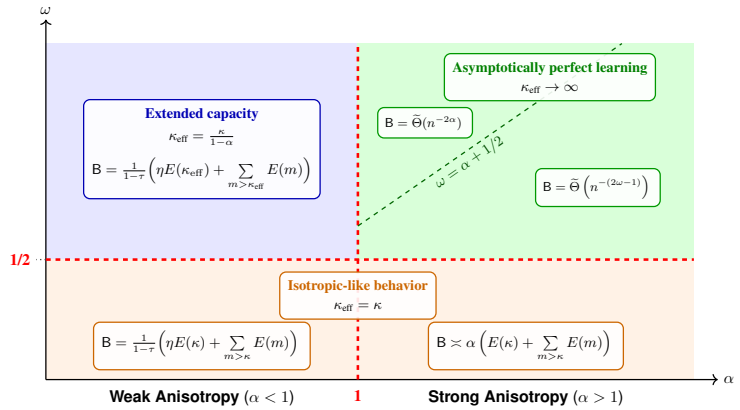


Figure 3: Phase diagram summarizing the asymptotic bias $\mathsf{B}_n$ in the ridgeless limit ($\lambda = 0^+$), highlighting the various regimes as $\alpha$ (data anisotropy) and $\omega$ (target decay anisotropy) are varied.

**Remark 4** (*Continuity at the critical boundary*)**.** The weakly and strongly anisotropic regimes continuously match at $\alpha = 1$. As $\alpha \to 1^{\pm}$, the variance vanishes from both sides (Eqs. (11) and (17)). This continuity naturally extends to the bias, notably since taking $\alpha \to 1^-$ in Eq. (16) yields $\eta \to 1$.

## 4 Case Study: Single-Index Models

In Sections 3.1.2 and 3.2.2, we characterized the deterministic bias starting from assumptions on the target function's spectral decomposition (see Eq. (14)). As a case study, we now discuss *single-index models* (SIMs), a class of functions that has drawn significant recent interest. While SIMs are defined directly in the input space, they allow us to link the abstract spectral decay $\omega$ to concrete geometric properties of the target. A SIM is a low-rank function defined as $f_\star(\boldsymbol{x}) = g(\boldsymbol{v}^\top \boldsymbol{x})$, where $\boldsymbol{v} \in \mathbb{R}^d$ is the *index* (normalized as $\mathbb{E}[(\boldsymbol{v}^\top \boldsymbol{x})^2] = 1$) and $g \in L^2(\mathcal{N}(0,1))$ is the link function. To analyze SIMs within our framework, we rely on the *Hermite ansatz* (Karkada et al., 2026), which postulates that, for $d \to \infty$, the kernel eigenfunctions $e_{\boldsymbol{\beta}}$ are well-approximated by the multivariate Hermite polynomials $H_{\boldsymbol{\beta}}$ (see Appendix F for a deeper discussion and empirical validation of this hypothesis). Under this Hermite ansatz, the tensor coefficients $\theta_{i_1,\ldots,i_m}$ of the target function can be explicitly characterized (see Appendix F.1 for the proof). For a dense index vector $\boldsymbol{v}$, as $d \to \infty$:

$$\theta_{i_1,\ldots,i_m} \simeq g_m \sqrt{m!}\, r_\alpha(d)^{-m/2} (i_1 \ldots i_m)^{-\alpha/2} \prod_{k=1}^{m} v_{i_k}, \tag{19}$$

where $\{g_m\}$ are the 1D Hermite coefficients of $g(t)$. Eq. (19) can be used to map the structural properties of the index $\boldsymbol{v}$ directly to the effective decay rate $\omega$ of our power-law framework in Eq. (14), revealing three distinct scenarios:

- **Power-law target:** If the index components decay as $v_j \propto j^{-\gamma/2}$, Eq. (19) yields an effective spectral decay $\omega = \frac{1}{2}(\alpha + \gamma)$, tying generalization directly to the interplay of $\alpha$ and $\gamma$.
- **Delocalized target:** For a random isotropic index $\boldsymbol{v} \sim \mathcal{N}(0, \boldsymbol{I}_d)$, the average effective decay is $\omega = \alpha/2$, mathematically equivalent to a power-law target with no intrinsic decay ($\gamma = 0$).
- **Localized target:** If the index is strictly aligned with a leading principal direction (e.g., $\boldsymbol{v} = \boldsymbol{e}^{(1)}$), the target becomes highly sparse, yielding an effective decay rate $\omega \to \infty$.

As a consequence, depending on the geometrical properties of the index $\boldsymbol{v}$, SIMs inherently undergo the exact same geometric phase transitions, fractional sample complexities, and bias-variance decoupling characterized in our main results.

### AI use statement

In this work, we used generative AI tools for various research and manuscript preparation subtasks. Specifically, we used them to aid and polish writing, improve overall readability, and assist with

translation. Furthermore, AI was utilized to create scientific figures (e.g., Fig. 2) and to assist with generating Python code for the empirical Kernel Ridge Regression (KRR) tests and the empirical diagonalization procedure in Appendix F via inline programming assistants. Regarding the theoretical content, AI tools were employed to assist in the writing of proofs and to brainstorm ideas for tackling specific theoretical challenges, such as the conceptualization of the Irwin-Hall trick proposed in Lemma D.1 to transform $m$-dimensional integrals into one-dimensional ones. Finally, we made limited use of AI to summarize and analyze existing scientific literature to better pinpoint and understand relevant related works. We have rigorously reviewed all AI-assisted work. All LLM-generated text, translations, and figures were reviewed by the authors, while the generated code was verified and empirically tested for correctness. For the theoretical claims, the authors carefully read and analyzed the mathematical hints provided by the AI, strictly evaluating their formal soundness and validity before integrating them into the mathematical process and final proofs. Any literature insights were also cross-referenced directly with the original published papers. We take full responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

### Reproducibility statement

The work presented in this paper is predominantly theoretical in nature. Formal proofs for the theoretical claims made in the main text are provided in the Appendices. Specifically, Appendix A for the fundamental Proposition 1, Appendix B for the validation of the deterministic equivalent formulas in our setting, Appendix D for the weakly anisotropic regime, Appendix E for the strongly anisotropic regime, and finally Appendix F for the SIM formula in Eq. (19). All underlying assumptions are clearly stated and elaborated where necessary, and conjectures (such as the Hermite Ansatz) are explicitly acknowledged. For the Hermite Ansatz, we also provide numerical evidence in Appendix F.

Furthermore, to validate our asymptotic predictions, we also perform numerical simulations. These results are illustrated in Figs. 6 to 8 for the variance, and in Appendix F for the bias and SIM-related formulas. The simulations are implemented in Python by training Kernel Ridge Regression (KRR) models on synthetic datasets drawn from a multivariate Gaussian distribution with a power-law covariance decay. Model hyperparameters are detailed in the respective sections and correspond to the scenarios investigated throughout the main text.

## References


Alexander Atanasov, Jacob A Zavatone-Veth, and Cengiz Pehlevan. Scaling and renormalization in high-dimensional regression. *Journal of Statistical Mechanics: Theory and Experiment*, 2026(4): 043404, 2026.

Francis Bach. On the equivalence between kernel quadrature rules and random feature expansions. *Journal of Machine Learning Research*, 18(21):1–38, 2017.

Francis Bach. Unraveling spectral properties of kernel matrices – i. Machine Learning Research Blog, 2024a.

Francis Bach. High-dimensional analysis of double descent for linear regression with random projections. *SIAM Journal on Mathematics of Data Science*, 6(1):26–50, 2024b.

Yasaman Bahri, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma. Explaining neural scaling laws. *Proceedings of the National Academy of Sciences*, 121(27):e2311878121, 2024.

Peter L. Bartlett, Philip M. Long, Gábor Lugosi, and Alexander Tsigler. Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences*, 117(48):30063–30070, 2020.

Mikhail Belkin, Siyuan Ma, and Soumik Mandal. To understand deep learning we need to understand kernel learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 541–549. PMLR, 2018.

Blake Bordelon, Abdulkadir Canatar, and Cengiz Pehlevan. Spectrum dependent learning curves in kernel regression and wide neural networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 1024–1034. PMLR, 2020.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Vivien A Cabannes, Francis Bach, and Alessandro Rudi. Fast rates for structured prediction. In *Conference on Learning Theory*, pp. 823–865. PMLR, 2021.

Francesco Cagnetta, Allan Raventós, Surya Ganguli, and Matthieu Wyart. Deriving neural scaling laws from the statistics of natural language. *arXiv preprint arXiv:2602.07488*, 2026.

Abdulkadir Canatar, Blake Bordelon, and Cengiz Pehlevan. Spectral bias and task-model alignment explain generalization in kernel regression and infinitely wide neural networks. *Nature Communications*, 12(1):2914, 2021.

A. Caponnetto and E. De Vito. Optimal rates for the regularized least-squares algorithm. *Foundations of Computational Mathematics*, 7(3):331–368, 2007.

Chen Cheng and Andrea Montanari. Dimension free ridge regression. *The Annals of Statistics*, 52 (6):2879–2912, 2024.

Lenaic Chizat, Edouard Oyallon, and Francis Bach. On lazy training in differentiable programming. *Advances in Neural Information Processing Systems*, 32:2933–2943, 2019.

Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39(1):1–49, 2001.

Hugo Cui, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. Generalization error rates in kernel regression: the crossover from the noiseless to noisy regime*. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(11):114004, 2022.

Leonardo Defilippis, Bruno Loureiro, and Theodor Misiakiewicz. Dimension-free deterministic equivalents and scaling laws for random feature regression. *Advances in Neural Information Processing Systems*, 37:104630–104693, 2024.

Behrooz Ghorbani, Song Mei, Theodor Misiakiewicz, and Andrea Montanari. When do neural networks outperform kernel methods? *Advances in Neural Information Processing Systems*, 33: 14820–14830, 2020.

Behrooz Ghorbani, Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Linearized two-layers neural networks in high dimension. *The Annals of Statistics*, 49(2):1029–1054, 2021.

Sebastian Goldt, Bruno Loureiro, Galen Reeves, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. The gaussian equivalence of generative models for learning with shallow neural networks. In *Mathematical and Scientific Machine Learning*, pp. 426–471. PMLR, 2022.

Arthur Jacot, Franck Gabriel, and Clement Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in Neural Information Processing Systems*, 31:8571–8580, 2018.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

Dhruva Karkada, Joseph Turnbull, Yuxi Liu, and James B Simon. Predicting kernel regression learning curves from only raw data statistics. In *The Fourteenth International Conference on Learning Representations*, 2026.

T. H. Koornwinder, R. S. C. Wong, R. Koekoek, and R. F. Swarttouw. Orthogonal polynomials. In *NIST Handbook of Mathematical Functions*, pp. 435–484. Cambridge University Press, 2010.

Tengyuan Liang, Alexander Rakhlin, and Xiyu Zhai. On the multiple descent of minimum-norm interpolants and restricted lower isometry of kernels. In *Proceedings of the 33rd Conference on Learning Theory*, volume 125, pp. 2683–2711. PMLR, 2020.

Yue M Lu and Horng-Tzer Yau. An equivalence principle for the spectrum of random inner-product kernel matrices with polynomial scalings. *The Annals of Applied Probability*, 35(4):2411–2470, 2025.

S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

Neil Mallinar, James Simon, Amirhesam Abedsoltan, Parthe Pandit, Misha Belkin, and Preetum Nakkiran. Benign, tempered, or catastrophic: Toward a refined taxonomy of overfitting. *Advances in Neural Information Processing Systems*, 35:1182–1195, 2022.

Alexander Maloney, Daniel A Roberts, and James Sully. A solvable model of neural scaling laws. *arXiv preprint arXiv:2210.16859*, 2022.

Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Generalization error of random feature and kernel methods: Hypercontractivity and kernel matrix concentration. *Applied and Computational Harmonic Analysis*, 59:3–84, 2022.

Theodor Misiakiewicz. Spectrum of inner-product kernel matrices in the polynomial regime and multiple descent phenomenon in kernel ridge regression. *arXiv preprint arXiv:2204.10425*, 2022.

Theodor Misiakiewicz and Basil Saeed. A non-asymptotic theory of kernel ridge regression: deterministic equivalents, test error, and gcv estimator. *arXiv preprint arXiv:2403.08938*, 2024.

Elliot Paquette, Ke Liang Xiao, and Yizhe Zhu. Power-law spectrum of the random feature model. *arXiv preprint arXiv:2603.14578*, 2026.

Paul Rosa. $l^2-$ posterior contraction rates for gaussian process and random series priors in bayesian nonparametric regression models. *arXiv preprint arXiv:2512.20503*, 2025.

Alessandro Rudi and Lorenzo Rosasco. Generalization properties of learning with random features. *Advances in Neural Information Processing Systems*, 30:3215–3225, 2017.

Alessandro Rudi, Raffaello Camoriano, and Lorenzo Rosasco. Less is more: Nyström computational regularization. *Advances in Neural Information Processing Systems*, 28:1657–1665, 2015.

Dominik Schröder, Daniil Dmitriev, Hugo Cui, and Bruno Loureiro. Asymptotics of learning with deep structured (Random) features. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 43862–43894. PMLR, 21–27 Jul 2024.

Jamie Simon. Predicting kernel spectra powerlaw exponents from raw data statistics. Personal website, 2026.

Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1):1193–1216, 2001.

Harold Widom. Asymptotic behavior of the eigenvalues of certain integral equations. *Transactions of the American Mathematical Society*, 109(2):278–295, 1963.

Christopher KI Williams and Matthias Seeger. The effect of the input density distribution on kernel-based classifiers. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 1159–1166, 2000.

Arie Wortsman and Bruno Loureiro. Kernel ridge regression under power-law data: spectrum and generalization. *arXiv preprint arXiv:2510.04780*, 2025.

Lechao Xiao, Hong Hu, Theodor Misiakiewicz, Yue Lu, and Jeffrey Pennington. Precise learning curves and higher-order scalings for dot-product kernel regression. *Advances in Neural Information Processing Systems*, 35:4558–4570, 2022.

Tong Zhang. Learning bounds for kernel regression using effective data dimensionality. *Neural Computation*, 17(9):2077–2098, 2005.

## A Properties of the Kernel Operator

In this section, we prove Proposition 1 from the main manuscript.

**Proposition 3** (Proposition 1 in the main manuscript)**.** Assume $\mathrm{Tr}(\boldsymbol{\Sigma}^2) = o_d(1)$, and $h_k > 0$ for all $k \in [D]$ in Assumption 1.1. Then, the spectrum of an inner-product kernel indexed by multi-indices $\boldsymbol{\beta} \in \mathbb{Z}_{\geq 0}^d$, $|\boldsymbol{\beta}| \leq D$ satisfies:

$$\lambda_{\boldsymbol{\beta}} = h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \, \sigma_1^{\beta_1} \dots \sigma_d^{\beta_d} (1 + o_d(1)) \,. \tag{20}$$

*Proof.* Denote $\boldsymbol{z} = \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{x} \sim \mathcal{N}(0, \boldsymbol{I}_d)$. We will follow the arguments of Appendix 1 in Wortsman & Loureiro (2025). Given $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$, we can write our kernel in the following form by expanding the inner product:

$$\begin{aligned} k(\boldsymbol{x}, \boldsymbol{x}') &= \sum_{k=0}^{D} h_k \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^k \\ &= \sum_{k=0}^{D} h_k \sum_{\boldsymbol{\beta} \in \mathbb{Z}_{\geq 0}^d : |\boldsymbol{\beta}| = k} \binom{k}{\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\beta}} \boldsymbol{x}'^{\boldsymbol{\beta}} \\ &= \sum_{k=0}^{D} h_k \sum_{\boldsymbol{\beta} \in \mathbb{Z}_{\geq 0}^d : |\boldsymbol{\beta}| = k} \binom{k}{\boldsymbol{\beta}} \boldsymbol{\sigma}^{\boldsymbol{\beta}} \boldsymbol{z}^{\boldsymbol{\beta}} \boldsymbol{z}'^{\boldsymbol{\beta}} \,, \end{aligned} \tag{21}$$

where we denote $\boldsymbol{\sigma}^{\boldsymbol{\beta}} = \prod_{i=1}^d \sigma_i^{\beta_i}$. Let

$$\Phi_{\boldsymbol{\beta}}(\boldsymbol{x}) = \sqrt{h_{|\boldsymbol{\beta}|} \binom{|\boldsymbol{\beta}|}{\boldsymbol{\beta}} \boldsymbol{\sigma}^{\boldsymbol{\beta}}} \boldsymbol{z}^{\boldsymbol{\beta}} \,, \tag{22}$$

and denote by $\boldsymbol{\Phi}(\boldsymbol{x}) \in \mathbb{R}^{\binom{D+d}{D}}$ the vector with these coordinates. We can expand each coordinate into Hermite polynomials via the transformation detailed in Wortsman & Loureiro (2025):

$$\Phi_{\boldsymbol{\beta}}(\boldsymbol{x}) = \sqrt{h_{|\boldsymbol{\beta}|} \binom{|\boldsymbol{\beta}|}{\boldsymbol{\beta}} \boldsymbol{\sigma}^{\boldsymbol{\beta}}} \boldsymbol{z}^{\boldsymbol{\beta}} = \sqrt{h_{|\boldsymbol{\beta}|} \binom{|\boldsymbol{\beta}|}{\boldsymbol{\beta}} \boldsymbol{\sigma}^{\boldsymbol{\beta}}} \sum_{\boldsymbol{\gamma} \leq \boldsymbol{\beta} : \boldsymbol{\gamma} \equiv_2 \boldsymbol{\beta}} \frac{\boldsymbol{\beta}!}{2^{\frac{|\boldsymbol{\beta}-\boldsymbol{\gamma}|}{2}} \left(\frac{|\boldsymbol{\beta}-\boldsymbol{\gamma}|}{2}\right)! \sqrt{\boldsymbol{\gamma}!}} He_{\boldsymbol{\gamma}}(\boldsymbol{z}) \,. \tag{23}$$

Let $d_{\boldsymbol{\beta}} = h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \boldsymbol{\sigma}^{\boldsymbol{\beta}}$, and define $\boldsymbol{D} = \mathrm{diag}\left((d_{\boldsymbol{\beta}})_{|\boldsymbol{\beta}| \leq D}\right)$. We also define the matrix $\boldsymbol{B} \in \mathbb{R}^{\binom{d+D}{D} \times \binom{d+D}{D}}$ as:

$$B_{\boldsymbol{\beta},\boldsymbol{\gamma}} := \begin{cases} \sqrt{\frac{h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \boldsymbol{\beta}!}{h_{|\boldsymbol{\gamma}|} |\boldsymbol{\gamma}|! \boldsymbol{\gamma}!}} \frac{\boldsymbol{\sigma}^{\frac{\boldsymbol{\beta}-\boldsymbol{\gamma}}{2}}}{2^{\frac{|\boldsymbol{\beta}-\boldsymbol{\gamma}|}{2}} \left(\frac{|\boldsymbol{\beta}-\boldsymbol{\gamma}|}{2}\right)!} & \text{if } \boldsymbol{\gamma} \leq \boldsymbol{\beta} \text{ and } \boldsymbol{\gamma} \equiv_2 \boldsymbol{\beta} \,, \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

In particular, note that $B_{\boldsymbol{\beta},\boldsymbol{\beta}} = 1$. Then, by Eq. (23) we have that:

$$\boldsymbol{\Phi}(\boldsymbol{x}) = \boldsymbol{B} \boldsymbol{D}^{\frac{1}{2}} \,. \tag{25}$$

Now, assume that

$$\|\boldsymbol{B} - \boldsymbol{I}\|_{\mathrm{op}} = o_d(1) \,. \tag{26}$$

Then we would have that:

$$(1 - o_d(1))^2 \boldsymbol{I} \preceq \boldsymbol{B}^{\boldsymbol{T}} \boldsymbol{B} \preceq (1 + o_d(1)) \boldsymbol{I} \,, \tag{27}$$

and therefore, multiplying by $\boldsymbol{D}^{\frac{1}{2}}$:

$$(1 - o_d(1))^2 \boldsymbol{D} \preceq \boldsymbol{D}^{\frac{1}{2}} \boldsymbol{B}^{\boldsymbol{T}} \boldsymbol{B} \boldsymbol{D}^{\frac{1}{2}} \preceq (1 + o_d(1)) \boldsymbol{D} \,, \tag{28}$$

and since the eigenvalues of the kernel are the same as the eigenvalues of $\boldsymbol{D}^{\frac{1}{2}} \boldsymbol{B}^{\boldsymbol{T}} \boldsymbol{B} \boldsymbol{D}^{\frac{1}{2}}$ by Eq. A.26 in Wortsman & Loureiro (2025), we would conclude the result. Therefore, we now focus on proving Eq. (26).

Denote by $\mathcal{I}_k := \{\boldsymbol{\beta} \in \mathbb{Z}^d_{\geq 0} : |\boldsymbol{\beta}| = k\}$ for $k \in [D]$. Given $\ell, k \in [d]$, define

$$B_{k,\ell} = (B_{\boldsymbol{\beta},\boldsymbol{\gamma}})_{|\boldsymbol{\beta}|=k,|\boldsymbol{\gamma}|=\ell}\,.$$

Note that if $k - \ell$ is not even, then $B_{k,\ell} = 0$. Also, note that $B_{k,k} = \boldsymbol{I}$. Let $\ell > 0$, and let $\boldsymbol{u} \in \mathbb{R}^{\binom{d+k-1}{k}}$, so the coordinates of $\boldsymbol{u}$ can be indexed by $(u_{\boldsymbol{\beta}})_{|\boldsymbol{\beta}|=k}$. Then we have:

$$(B_{k+2\ell,k}\boldsymbol{u})_{\boldsymbol{\beta}} = \sqrt{\frac{h_{k+2\ell}(k+2\ell)!}{h_k k!}} \sum_{\boldsymbol{\gamma}\in\mathbb{Z}^d_{\geq 0}:|\boldsymbol{\gamma}|=\ell,2\boldsymbol{\gamma}\leq\boldsymbol{\beta}} \frac{\boldsymbol{\sigma}^{\boldsymbol{\gamma}}}{2^\ell \boldsymbol{\gamma}!} \sqrt{\frac{\boldsymbol{\beta}!}{(\boldsymbol{\beta}-2\boldsymbol{\gamma})!}} u_{\boldsymbol{\beta}-2\boldsymbol{\gamma}}\,. \tag{29}$$

Since we can bound most of the coefficients of the sum by constants that do not depend on the dimension:

$$\|B_{k+2\ell,k}\boldsymbol{u}\|_2^2 \leq C_D \sum_{|\boldsymbol{\beta}|=k+2\ell} \sum_{\boldsymbol{\gamma}\in\mathbb{Z}^d_{\geq 0}:|\boldsymbol{\gamma}|=\ell,2\boldsymbol{\gamma}\leq\boldsymbol{\beta}} \boldsymbol{\sigma}^{2\boldsymbol{\gamma}} u^2_{\boldsymbol{\beta}-2\boldsymbol{\gamma}}\,. \tag{30}$$

Applying the Cauchy-Schwarz inequality, we get:

$$\|B_{k+2\ell,k}\boldsymbol{u}\|_2^2 \leq C\|\boldsymbol{u}\|_2^2 \sum_{|\boldsymbol{\gamma}|=\ell} \boldsymbol{\sigma}^{2\boldsymbol{\gamma}} \leq C\|\boldsymbol{u}\|_2^2 \left(\mathrm{Tr}(\boldsymbol{\Sigma}^2)\right)^\ell\,. \tag{31}$$

Hence, we conclude:

$$\|B_{k+2\ell,k}\|_{\mathrm{op}} \leq \left(\mathrm{Tr}(\boldsymbol{\Sigma}^2)\right)^{\frac{\ell}{2}}\,. \tag{32}$$

Then, by applying the triangle inequality:

$$\|\boldsymbol{B} - \boldsymbol{I}\|_{\mathrm{op}} \leq C_D \sum_{k=0}^{D} \sum_{\substack{\ell=1 \\ k+2\ell\leq D}}^{k} \left(\mathrm{Tr}(\boldsymbol{\Sigma}^2)\right)^{\frac{\ell}{2}}\,. \tag{33}$$

Since we have a constant number of sums, we conclude:

$$\|\boldsymbol{B} - \boldsymbol{I}\|_{\mathrm{op}} \leq C_D (\mathrm{Tr}(\boldsymbol{\Sigma}^2))^{\frac{1}{2}}\,, \tag{34}$$

and since $\mathrm{Tr}(\boldsymbol{\Sigma}^2) = o_d(1)$ by assumption, this concludes the proof. □

## B Deterministic Equivalents

The bias-variance decomposition in Eq. (6) consists on two random terms. As discussed in Section 3, under specific assumptions, Cheng & Montanari (2024); Misiakiewicz & Saeed (2024); Defilippis et al. (2024) proved that this quantities concentrate around deterministic quantities, called *deterministic equivalents*, and presented in Eq. (7). In particular, Misiakiewicz & Saeed (2024) provides an abstract assumption to derive deterministic equivalents for Kernel Ridge Regression. We will first briefly state these assumptions and then we will provide a proof of their validity in our setting.

The idea behind the following assumptions is to separate the kernel eigenfunctions into lower are higher order. Given the $m$ largest eigenvalues, we denote by $\Lambda_{\leq m}$ the covariance of the associated eigenfunctions (which we will sometimes call the lower-degree part) and $\Lambda_{>m}$ the covariance associated to the higher eigenvalues (which we will also call the higher-degree part). The exposition of the different notions and assumptions in this section follows Misiakiewicz & Saeed (2024).

**Definition 1** (Ranks)**.** Let $p \in \mathbb{N} \cup \{\infty\}$ be the rank of the kernel $k$. We introduce the following notions of rank for our covariance $\Lambda$:

1. **Regularized tail Rank:** For any integer $m \leq p$ and regularization parameter $\lambda \geq 0$, the regularized tail rank $r_\lambda(m)$ is defined by

$$r_\lambda(m) = \frac{\lambda + \sum_{j=m+1}^p \lambda_j}{\lambda_{m+1}}.$$

We denote $\lambda_{>m} := \lambda + \sum_{j=m+1}^p \lambda_j$.

2. **Effective Rank:** For any integers $n$ and $m \leq p$, the effective rank $r_{\text{eff},m}(n)$ of $(n, \Lambda_{\leq m})$ is defined as the smallest scalar such that $r_{\text{eff},m} \geq n$ and

$$r_{\text{eff},m}(n) \geq \frac{\sum_{j=k+1}^{m} \xi_j}{\xi_{k+1}}, \text{ for all } 0 \leq k \leq \min(n, m) - 1.$$

3. **Relative Approximation bound**: For any $n, m \in \mathbb{N}$ and ridge regularization parameter $\lambda \geq 0$, we define:

$$\nu_{\lambda,m}(n) = 1 + \frac{\lambda_{\lfloor \eta n \rfloor, m} r_{\text{eff},m}(n) \sqrt{\log(r_{\text{eff},m}(n))}}{\lambda + \text{Tr}(K_{>m})}. \tag{35}$$

for $\eta \in (0, \frac{1}{2})$, and $\lambda_{\lfloor \eta n \rfloor, m} = \lambda_{\lfloor \eta n \rfloor}$ if $\lfloor \eta x \rfloor \leq m$, and 0 otherwise.

The first assumption we will need concerns the behaviour of the lower and higher order eigenfunctions.

**Assumption B.1** (Concentration at $n \in \mathbb{N}$)**.** There exists constants $C_x, c_x, \beta_x > 0$ and $m \in \mathbb{N} \cup \{\infty\}$, $m \leq p$, such that the regularized tail rank at $m$ satisfies

$$r_\lambda(m) \geq 2n, \tag{36}$$

and the following hold:

1. **Low-degree Features:** There exists $\varphi_1(m) > 0$ such that for any deterministic vector $v \in \mathbb{R}^m$, with $\|\Lambda^{\frac{1}{2}} v\|_2 < \infty$ and p.s.d matrix $A \in \mathbb{R}^{m \times m}$ with $\text{Tr}(\Lambda_{\leq m} A) < \infty$ we have

$$\mathbb{P}\left(|\langle v, \phi_{\leq m} \rangle| \geq t v^t \Lambda_{\leq m} v\right) \leq C_x \exp\left(-c_x t^{\frac{2}{\beta_x}}\right)$$
$$\mathbb{P}\left(\left|\phi_{\leq m}^T A \phi_{\leq m} - \text{Tr}(\Lambda_{\leq m} A)\right| \geq t\varphi_1(m) \|\Lambda_{\leq m}^{\frac{1}{2}} A \Lambda_{\leq m}^{\frac{1}{2}}\right) \leq C_x \exp\left(-c_x t^{\frac{1}{\beta_x}}\right)$$

2. **High-degree Features:** There exists $p_{2,n}(m) \in (0, 1)$ and $\varphi_{2,n}(m) \geq 1$ such that with probability at least $1 - p_{2,n}(m)$, we have

$$\|\Phi_{>m} \Phi_{>m}^T - \text{Tr}(\Lambda_{>m}) I_n\|_{\text{op}} \leq \varphi_{2,n}(m) \sqrt{\frac{n}{r_\lambda(m)}} (\lambda + \text{Tr}(\Lambda_{\geq m}))$$

3. **Target Function:** The high-degree part of the target function satisfies the tail bound

$$\mathbb{P}\left(|f_{\star,>m}(x)| \geq t \|f_{\star,>m}\|_{L^2}\right) \leq C_x \exp(-c_x t^{\frac{2}{\beta}}).$$

At last, we need the following assumptions on the target function $f^\star(x)$.

**Assumption B.2** (Label noise)**.** The label noises $\varepsilon_i, i \in [n]$ are independent, mean-zero, $\sigma_\varepsilon^2$-sub-Gassian random variables.

We can now state the main theorem that allows us to rigorously state the quantities in Equation (7) as deterministic equivalents for the generalization error in our setting.

**Theorem 5** (Theorem 1 in (Misiakiewicz & Saeed, 2024))**.** Consider $D, K > 0$, an integer $n$, a regularization parameter $\lambda \geq 0$, and a target function $f_* \in L^2(\mathcal{X})$ with parameters

$$\|\theta_*\|_2 = \|f_*\|_{L^2} < \infty.$$

Suppose Assumption B.1 is satisfied for some

$$m := m(n) \in \mathbb{N} \cup \{\infty\},$$

and that $\{\varepsilon_i\}_{i \in [n]}$ satisfy Assumption B.2. There exist constants

$$\eta := \eta_x \in \left(0, \frac{1}{2}\right), \qquad C_{D,K} > 0, \qquad C_{x,\varepsilon,D,K} > 0,$$

such that, for all $n \geq C_{D,K}$ and $\lambda_{>m} > 0$, if

$$\lambda_{>m}\, \nu_{\lambda,m}(n) \geq n^{-K}, \qquad \varphi_{2,n}(m)\sqrt{\frac{n}{r_\lambda(m)}} \leq \frac{1}{2}, \qquad \varphi_1(m)\nu_{\lambda,m}(n)^8 \log^{3\beta+\frac{1}{2}}(n) \leq K\sqrt{n},$$

then, with probability at least $1 - n^{-D} - p_{2,n}(m)$, we have

$$\left|\mathcal{R}\big(\theta_*; X, \varepsilon, \lambda\big) - \mathsf{R}_n(\theta_*, \lambda)\right| \leq C_{x,\varepsilon,D,K}\, \mathcal{E}_{R,n}(m)\, R_n(\theta_*, \lambda),$$

where the relative approximation rate is given by

$$\mathcal{E}_{R,n}(m) := \frac{\varphi_1(m)\nu_{\lambda,m}(n)^6 \log^{3\beta+\frac{1}{2}}(n)}{\sqrt{n}} + \nu_{\lambda,m}(n)\varphi_{2,n}(m)\sqrt{\frac{n}{r_\lambda(m)}}.$$

Theorem 5 roughly tells us that, if we can prove both Assumption B.1 and Assumption B.2, then we can apply the deterministic equivalents from Equation (7). Hence, before manipulating the quantities in Equation (7) by applying Eq. (3), we need to prove that Assumption B.1 and Assumption B.2 are satisfied in our setting.

In the following sections, we will prove the Assumptions of Theorem 5. We will separate the spectrum into the weakly anisotropic setting $\alpha < 1$ and the strongly anisotropic setting $\alpha > 1$.

Before beginning, we note that the rank of our kernel is $p := O(d^D)$.

### B.1 The Weakly Anisotropic Setting

We begin by proving Assumption B.1. We begin by defining our threshold $m = m(n)$. We will only do the case $0 < \alpha < 1$. The case $\alpha = 0$ follows by the same arguments.

**Lemma B.1.** Let $\alpha \in (0, 1)$, and let $n = \psi d^\kappa$, for $\psi > 0$ independent of $d$ and $\kappa > 0$ also independent of $d$, with $\lfloor \kappa \rfloor \in [1, D]$. Define:

$$m := \left|\{k \in \mathbb{N} : k \leq p, \lambda_k > d^{-(\kappa+\rho)}\}\right|, \tag{37}$$

with $\rho > 0$ such that $0 < \rho < (1-\alpha)(\lfloor \kappa \rfloor + 1 - \kappa)$. Then, if $\kappa \notin \mathbb{N}$,

$$\exists \delta_0 > 0 \text{ s. t } m \leq n^{1-\delta_0}.$$

If $\kappa \in \mathbb{N}$, then

$$m \leq \Theta(n).$$

*Proof.* Define:

$$m := \left|\{k \in \mathbb{N} : k \leq p, \lambda_k > d^{-(\kappa+\rho)}\}\right|. \tag{38}$$

Note that $\lambda_{m+1} = O(d^{-(\kappa+\rho)})$ by construction. By Lemma C.10 from Wortsman & Loureiro (2025), we conclude the first equation. For the second case, when $\kappa \in \mathbb{N}$, the set that defines $m$ necessarily contains the whole degree $\kappa$ shell, and the result follows. □

**Lemma B.2.** Let $m = m(n)$ as defined in Lemma B.1. Then, for any ridge regularization parameter $\lambda > 0$,

$$r_\lambda(m) \geq 2n.$$

Moreover, for $n = \Theta(d^\kappa)$ the relative approximation bound $\nu_{\lambda,m} = 1$ when $\kappa \notin \mathbb{N}$, and $\nu_{\lambda,m} = O(\text{poly}\log(n))$ when $\kappa \in \mathbb{N}$.

*Proof.* By definition, we have:

$$r_\lambda(m) = \frac{1}{\lambda_{m+1}}\left(\lambda + \sum_{j \geq m+1} \lambda_j\right). \tag{39}$$

Assuming the kernel has a shell after degree $\lfloor \kappa \rfloor$, and using the fact that $\lambda_{m+1} \leq Cd^{-(\kappa+\rho)}$ we get:

$$r_\lambda(m) \geq Cd^{\kappa+\rho}.$$

Therefore:

$$\frac{r_\lambda(m)}{n} \geq Cd^\rho = \omega_d(1). \tag{40}$$

Hence, for $d$ large enough,

$$r_\lambda(m) \geq 2n. \tag{41}$$

We now focus on proving the results for $\nu_{\lambda,m}$. Recall:

$$\nu_{\lambda,m}(n) = 1 + \frac{\lambda_{\lfloor \eta n \rfloor, m} r_{\text{eff},m}(n) \sqrt{\log(r_{\text{eff},m}(n))}}{\lambda + \text{Tr}(K_{>m})}. \tag{42}$$

If $\kappa \notin \mathbb{N}$, by Lemma B.1 we have $m \leq n^{1-\delta_0}$. Therefore, $\lambda_{\lfloor \eta n \rfloor, m} = 0$ and we conclude $\nu_{\lambda,m}(n) = 1$.

Now assume $\kappa \in \mathbb{N}$. Since by Lemma B.1 we have $m = \Theta(n)$, for any constant $\eta$, the index $\lfloor \eta n \rfloor$ lies in the $\kappa$-th shell. Then, by Corollary 2 in (Wortsman & Loureiro, 2025):

$$\lambda_{\lfloor \eta n \rfloor} = O(d^{(1-\alpha)\kappa} \lfloor \eta n \rfloor^{-\alpha}) = O\left(d^{-(1-\alpha)\kappa} d^{-\kappa\alpha}\right) = O\left(d^{-\kappa}\right). \tag{43}$$

On the other hand, to compute $r_{\text{eff},m}(n)$ we have:

$$r_{\text{eff},m}(n) = \max_{0 \leq k \leq \min(m,n)-1} \left\{ \frac{\sum_{j=k+1}^m \lambda_j}{\lambda_{k+1}} \right\} \tag{44}$$

$$\leq \max_{0 \leq k \leq \min(m,n)-1} m \frac{\lambda_{k+1}}{\lambda_{k+1}} \leq m = \Theta(n). \tag{45}$$

Finally, since $\lambda > 0$, we can replace Equation (43) and Equation (45) in the definition of $\nu_{\lambda,m}(n)$ to obtain:

$$\nu_{\lambda,m}(n) = O(\text{poly}\log(n)). \tag{46}$$

□

So far, we have only proved the regularized effective dimension condition from Assumption B.1. We now proceed to prove the rest of the conditions. We begin by Lower-degree conditions.

**Lemma B.3** (Lower-Order Conditions in the Weakly Anisotropic Setting)**.** Let $\alpha \in [0, 1)$, and assume $n = \psi d^\kappa$ for $\psi, \kappa > 0$ constants independent of the dimension. Then the lower-degree conditions of Assumption B.1 are satisfied with

$$\varphi_1(m) = O\left(\sqrt{m}\text{poly}\log(n)\right)$$

*Proof.* The condition on the lower-order eigenfunctions says that we need

$$\mathbb{P}\left(|\langle v, \phi_{\leq m}\rangle| \geq t v^t \Lambda_{\leq m} v\right) \leq C_x \exp\left(-c_x t^{\frac{2}{\beta_x}}\right) \tag{47}$$

$$\mathbb{P}\left(\left|\phi_{\leq m}^T A \phi_{\leq m} - \text{Tr}(\Lambda_{\leq m} A)\right| \geq t\varphi_1(m)\|\Lambda_{\leq m}^{\frac{1}{2}} A \Lambda_{\leq m}^{\frac{1}{2}}\right) \leq C_x \exp\left(-c_x t^{\frac{1}{\beta_x}}\right). \tag{48}$$

(Misiakiewicz & Saeed, 2024) notes that, in order to proof this condition it suffices to prove that the top eigenspaces are hypercontractive, that is, there exists constants $C, c$ such that for all $g \in \text{span}\{\phi_j : j \in [m]\}$, and for any $q \geq 2$:

$$\|g\|_{L^q}^2 \leq (C)^c \|g\|_{L^2}^2. \tag{49}$$

For our cases, we are working with polynomial inner product kernels. Its not hard to prove that, under this assumption, the eigenfunctions of the kernel are polynomials (although not Hermite polynomials), and hence, satisfy Equation (49) with $C = 1$ and $c = D$. This implies Equation (47) is satisfied by Markov's Inequality.

On the other hand, as noted after Equation 25 in (Misiakiewicz & Saeed, 2024), Equation (49) and the fact that we have Gaussian polynomials also implies Equation (48) is satisfied with

$$\varphi_1(m) = O\left(\sqrt{m}\log(n)\right). \tag{50}$$

Therefore, Lower-order Conditions are satisfied in our setting. □

We can actually improve the bound in Equation (50).

**Lemma B.4.** Let $e(x) = (e_j(x))_{j\in[m]}$ be the vector with the top $m$ eigenfunctions of the kernel. Then, there exists $\delta > 0$ such that, for any matrix $A \in \mathbb{R}^{m\times m}$

$$\mathbb{P}\left(\left|e(x)^T Ae(x) - \mathrm{Tr}(A)\right| \geq t\mathrm{poly}\log(n)\sqrt{m\mathrm{Tr}(\Sigma^2)}\|A\|_F\right) \leq C\exp\left(-cmt^m\right). \tag{51}$$

Therefore, we can take

$$\varphi_1(m) = \tilde{O}\left(\sqrt{m\mathrm{Tr}(\Sigma^2)}\log(n)\right) \tag{52}$$

*Proof.* Indeed, note that:

$$e^T A\varphi(x) - \mathrm{Tr}(A) = \phi(x)^T B_m^T AB_m\phi(x) - \mathrm{Tr}(A), \tag{53}$$

where $\phi(x) = (phi_j(x))_{j\in[m]}$ is a vector with $m$ Hermite polynomials with matching multi-indices, and $B_m$ is a change of basis matrix. By the arguments in Appendix A, we have:

$$\|B_m B_m^T - I\|_{\mathrm{op}} \leq C\sqrt{\mathrm{Tr}(\Sigma^2)}. \tag{54}$$

for $0 \leq \alpha \leq 1$. Therefore:

$$\begin{aligned} e(x)^T Ae(x) - \mathrm{Tr}(A) &= \phi(x)^T B_m^T AB_m\phi(x) - \mathrm{Tr}(A) && (55)\\ &= \phi(x)^T B_m^T AB_m\phi(x) - \mathrm{Tr}(B_m AB_m) + \mathrm{Tr}((B_m B_m^T - I_m)A). && (56)\end{aligned}$$

As discussed after Eq. 25 in (Misiakiewicz & Saeed, 2024), for Hermite polynomials, for any p.s.d matrix $C \in \mathbb{R}^{m\times m}$

$$\mathbb{P}\left(\left|\phi(x)^T C\phi(x) - \mathrm{Tr}(C)\right| \geq t\mathrm{poly}\log(n)\sqrt{\frac{m}{d}}\|C\|_F\right) \leq C\exp\left(-cmt^m\right) \tag{57}$$

Since $\|B_m^T AB_m\|_{\mathrm{F}} \leq \|B_m^T B\|_{\mathrm{op}}\|A\|_F \leq C(D)A$, we already get the desired bound for this quantity. On the other hand, by Equation (54):

$$\mathrm{Tr}((B_m B_m^T - I_m)A) \leq C\sqrt{m\mathrm{Tr}(\Sigma^2)}\|A\|_F. \tag{58}$$

Putting Equation (57) and Equation (58) together, we conclude. □

For higher order conditions the fact that eigenfunctions are polynomials also makes things easier, as hypercontractivity is a key property allowing to prove the higher order condition in Assumption B.1. More precisely, following (Misiakiewicz & Saeed, 2024), we need to prove that **a)** The higher-order eigenspaces are hypercontractive, and **b)** The higher-order diagonal of the kernel matrix concentrates. The latter can be written as:

$$\exists\alpha_1 > 0, \text{ such that } \mathbb{E}\left[\max_{i\in[n]}|K_{>m}(x_i,x_i) - \mathbb{E}\left[K_{>m}(x_i,x_i)\right]|\right] \leq \alpha_1\sqrt{\frac{n\lambda_{m+1}}{\mathrm{Tr}(\Lambda_{>m})}}\mathbb{E}\left[K_{>m}(x_i,x_i)\right]. \tag{59}$$

Since we already have hypercontractivity, we are left with proving Equation (59).

**Lemma B.5** (Concentration of Diagonal Features)**.** Let $\alpha \in [0,1)$. Let

$$F_m(x) := K_{>m}(x_i,x_i),$$

Then

$$\mathbb{E}\left[(F_m - \mathrm{Tr}(K_{>m}))^2\right] \leq C\|\Sigma\|_{\mathrm{op}}\mathrm{Tr}(K_{>m}).$$

*Proof.* We first note that

$$\mathbb{E}\left[(F_m - \mathrm{Tr}(K_{>m}))^2\right] = \mathrm{Var}(F_m). \tag{60}$$

Therefore, we want to bound a variance. The Gaussian Poincaré Inequality Lemma G.3 gives:

$$\mathrm{Var}(F_m) \leq \|\Sigma\|_{\mathrm{op}}\mathbb{E}\left[\|\nabla F_m(x)\|^2\right]. \tag{61}$$

Now, let $P_{>m}$ be the orthogonal projector into the higher degree features. We will now compute the gradient on the RHS. Let $j \in [d]$. We have:

$$\partial_j F_m(x) = \partial_j \|P_{>m}\Phi\|^2 = 2\langle P\Phi(x), \partial_j \Phi(x)\rangle, \tag{62}$$

where $\Phi(x)$ is the vector of polynomial features, and the inner product is taken in $\mathbb{R}^p$, for $p$ the range of the kernel. Let $\beta \in \mathbb{Z}^d_{\geq 0}$, with $|\beta| \leq D$. Then:

$$\partial_j \Phi_\beta(x) = \partial_j \sqrt{h_{|\beta|}\binom{|\beta|}{\beta}} x^\beta = \mathbf{1}_{\beta_j > 0}\sqrt{h_{|\beta|}\binom{|\beta|}{\beta}}\beta_j x^{\beta - e_j}. \tag{63}$$

Replacing in Equation (62):

$$\partial_j F_m(x) = 2\sum_{\ell=1}^{D} \sum_{\beta \in \mathbb{Z}^d_{\geq 0}:|\beta|=\ell,\beta_j>0} (P\Phi(x))_\beta \sqrt{h_{|\beta|}}\beta_j x^{\beta - e_j}. \tag{64}$$

By taking the norm and repeatidly using Cauchy-Schwarz, we have:

$$\|\nabla F_m(x)\|^2 \leq \sum_{j=1}^{d}\left(2\sum_{\ell=1}^{D} \sum_{\beta \in \mathbb{Z}^d_{\geq 0}:|\beta|=\ell,\beta_j>0} (P\Phi(x))_\beta \sqrt{h_{|\beta|}}\beta_j x^{\beta - e_j}\right)^2 \tag{65}$$

$$\leq 4\sum_{j=1}^{d}\sum_{\ell=1}^{D} h_\ell \left(\sum_{\beta \in \mathbb{Z}^d_{\geq 0}:|\beta|=\ell} \beta_j (P\Phi)_\beta^2\right)\left(\sum_{\beta \in \mathbb{Z}^d_{\geq 0}:|\beta|=\ell,\beta_j>0} \binom{\ell}{\beta}\beta_j x^{2\beta - 2e_j}\right) \tag{66}$$

$$= 4\sum_{\ell=1}^{D} h_\ell \sum_{j=1}^{d}\left(\sum_{\beta \in \mathbb{Z}^d_{\geq 0}:|\beta|=\ell} \beta_j (P\Phi)_\beta^2\right)\left(\ell \sum_{\beta \in \mathbb{Z}^d_{\geq 0}:|\beta|=\ell-1} \binom{\ell-1}{\beta} x^{2\beta}\right) \tag{67}$$

$$= 4\sum_{\ell=1}^{D} \ell \|x\|^{2(\ell-1)} \sum_{j=1}^{d}\left(\sum_{\beta \in \mathbb{Z}^d_{\geq 0}:|\beta|=\ell} \beta_j (P\Phi)_\beta^2\right) \tag{68}$$

$$\leq 4\sum_{\ell=1}^{D} h_\ell \ell^2 \|x\|^{2(\ell-1)} \|P\Phi_\ell\|_2^2 \tag{69}$$

$$\leq \left(\sum_{\ell=1}^{D} h_\ell \ell^2 \|x\|^{2(\ell-1)}\right)\left(\sum_{\ell=1}^{D} h_\ell \|P\Phi_\ell\|_2^2\right) \leq \left(\sum_{\ell=1}^{D} h_\ell \ell^2 \|x\|^{2(\ell-1)}\right) F_m(x). \tag{70}$$

The, taking expectation:

$$\mathbb{E}\left[\|\nabla F_m(x)\|^2\right] \leq \mathbb{E}\left[\left(\sum_{\ell=1}^{D} h_\ell \ell^2 \|x\|^{2(\ell-1)}\right) F_m(x)\right]. \tag{71}$$

Applying Lemma G.1 and Cauchy-Schwarz, we conclude:

$$\mathbb{E}\left[\|\nabla F_m(x)\|^2\right] \leq C_{D,h}\mathbb{E}\left[F_m^2\right]^{\frac{1}{2}} = C_{D,h}\operatorname{Tr}(K_{>m}). \tag{72}$$

Replacing this in Equation (61):

$$\operatorname{Var}(F_m) \leq C_{D,h}\|\Sigma\|_{\text{op}} \operatorname{Tr}(K_{>m}), \tag{73}$$

and we conclude. □

We can now prove Equation (59).

**Lemma B.6** (Diagonal Concentration)**.** For $m$ as chosen in Lemma B.1, Equation (59) holds with

$$\alpha_1 = \frac{\mathbb{E}\left[\max_{i\in[n]} |F_m(x_i) - \operatorname{Tr}(K_{>m})|\right]}{\sqrt{n\lambda_{m+1}\operatorname{Tr}(K_{>m})}}.$$

*Proof.* By Lemma B.5 and Lemma G.2, we have that:

$$\mathbb{E}\left[(F_m - \operatorname{Tr}(K_{>m})^2\right]^{\frac{1}{2}} \leq C_{D,h} \min\{\operatorname{Tr}(K_{>m}), \sqrt{\|\Sigma\|_{\text{op}} \operatorname{Tr}(K_{>m})}\}. \tag{74}$$

Then, by hypercontractivity:

$$\mathbb{E}\left[(F_m - \operatorname{Tr}(K_{>m})^q\right]^{\frac{1}{q}} \leq C_{D,q,h} \min\{\operatorname{Tr}(K_{>m}), \sqrt{\|\Sigma\|_{\text{op}} \operatorname{Tr}(K_{>m})}\}. \tag{75}$$

Then we have:

$$\begin{aligned}
\mathbb{E}\left[\max_{i\in[n]} |F_m(x_i) - \operatorname{Tr}(K_{>m})|\right] &\leq \mathbb{E}\left[\max_{i\in[n]} |F_m(x_i) - \operatorname{Tr}(K_{>m})|^q\right]^{\frac{1}{q}} && (76)\\
&\leq \mathbb{E}\left[\sum_{i=1}^{n} |F_m(x_i) - \operatorname{Tr}(K_{>m})|^q\right]^{\frac{1}{q}} && (77)\\
&\leq n^{\frac{1}{q}} \mathbb{E}\left[|F_m(x_i) - \operatorname{Tr}(K_{>m})|^q\right]^{\frac{1}{q}} && (78)\\
&\leq C_{D,q,h} n^{\frac{1}{q}} \min\{\operatorname{Tr}(K_{>m}), \sqrt{\|\Sigma\|_{\text{op}} \operatorname{Tr}(K_{>m})}\}. && (79)
\end{aligned}$$

Taking $q$ large enough, we conclude:

$$\mathbb{E}\left[\max_{i\in[n]} |F_m(x_i) - \operatorname{Tr}(K_{>m})|\right] \leq C_{D,h} \text{poly}\log(n) \min\{\operatorname{Tr}(K_{>m}), \sqrt{\|\Sigma\|_{\text{op}} \operatorname{Tr}(K_{>m})}\}. \tag{80}$$

Then, defining $\alpha_1 := \frac{\mathbb{E}\left[\max_{i\in[n]} |F_m(x_i) - \operatorname{Tr}(K_{>m})|\right]}{\sqrt{n\lambda_{m+1} \operatorname{Tr}(K_{>m})}}$, we have:

$$\alpha = O\left(\sqrt{\|\Sigma\|_{\text{op}} \frac{1}{n\lambda_{m+1}}}\right). \tag{81}$$

Note that if $\kappa \in \mathbb{N}$, then we get $\alpha = o_d(1)$ since $\|\Sigma\|_{\text{op}} = o_d(1)$. On the other hand, if $\kappa \notin \mathbb{N}$, since $\rho \leq 1 - \alpha$, we can also conclude that $\alpha = O_n(1)$. Therefore, we conclude the Lemma. □

By Proposition 1 in Misiakiewicz & Saeed (2024) Lemma B.6 implies the concentration for the higher order part of the diagonal of the kernel matrix with a function

$$\varphi_{2,n}(m) \leq \frac{C_D}{p_{2,n}}(\alpha_1 + C_D \text{poly}\log(n)), \tag{82}$$

for any chosen $p_{2,n}$.

Note that Assumption B.2 already assumes Part 3 of Assumption B.1. Therefore, we are only left with proving the spectral assumptions in Theorem 5.

**Lemma B.7.** Let $\alpha \in (0, 1)$, and assume $n = \psi d^\kappa$ for $\psi, \kappa > 0$. Let $m = m(n)$ as defined in Lemma B.1. Then:

1. $\lambda + \operatorname{Tr}(K_{>m})\nu_{\lambda,m}(n) \geq n^{-K}$.
2. $\varphi_{2,m}(m)\sqrt{\frac{n}{r_\lambda(m)}} \leq \frac{1}{2}$.
3. $\varphi_1(m)\nu_{\lambda,m}^8(n) \log^{3\beta+\frac{1}{2}}(n) \leq K\sqrt{n}$.

*Proof.* By Lemma B.2, $\nu_{\lambda,m}(n) = \tilde{O}(1)$ regardless of $\kappa \in \mathbb{N}$ or not. Therefore, the first condition gives:

$$\lambda + \operatorname{Tr}(K_{>m})\nu_{\lambda,m}(n) = \Theta(1). \tag{83}$$

Therefore, for any $n$ large enough the first condition holds.

By Equation (82), the second condition holds for $d$ large enough, as we have:

$$\varphi_{2mn}(m)\sqrt{\frac{n}{r_\lambda(m)}} = o_d(1). \tag{84}$$

For the third condition, using again the fact that $\nu_{\lambda,m}(n) = \tilde{O}(1)$:

$$\varphi_1(m)\nu_{\lambda,m}(n)^8 \log^{3\beta+\frac{1}{2}}(n) = O(\varphi_1(m)\log^{3\beta+\frac{1}{2}}(n)). \tag{85}$$

By Equation (50), we conclude:

$$\varphi_1(m)\nu_{\lambda,m}(n)^8 \log^{3\beta+\frac{1}{2}}(n) = O\left(\sqrt{m}\text{poly}\log(n)\right). \tag{86}$$

Then, for $\kappa \notin \mathbb{N}$, Lemma B.1 tells us that $m \leq n^{1-\delta_0}$, so we can conclude. For $\kappa \in \mathbb{N}$, we use the estimate from Lemma B.4 to conclude for $d$ large enough. □

Putting Lemma B.1, Lemma B.2, Lemma B.3, Lemma B.6 and Lemma B.7, all the assumptions of Theorem 5 hold, and therefore conclude Proposition 2, that is, we can apply the deterministic equivalents in the weakly anisotropic setting.

### B.2 The Strongly Anisotropic Setting

The strongly anisotropic setting is different than the weakly anisotropic setting only by some subtle technicalities. We sketch the validity of the deterministic equivalents in this setting below.

First, we can choose $m(n)$ as in Lemma B.1, however, we have to restrict $\rho \in (0, (\alpha-1)\kappa)$. We get $m \leq n^{1-\delta_0}$ for some $\delta_0 > 0$ regardless of $\kappa \in \mathbb{N}$ or not.

The conclusions for the effective dimensions in Lemma B.2 still hold. Indeed, by Corollary 2 in (Wortsman & Loureiro, 2025), in this case, each shell almost follows a capacity condition:

$$\lambda_k = \Theta(\frac{1}{r_0(\Sigma)}k^{-\alpha}\text{poly}\log(d)) = \Theta(k^{-\alpha}\text{poly}\log(d)), \tag{87}$$

because the normalizing constant $r_0(\Sigma) = \Theta_d(1)$ in this setting. Therefore, in this case:

$$\text{Tr}(K_{>m}) = \sum_{k\geq m}\lambda_m \leq C\text{poly}\log(m)\int_m^\infty x^{-\alpha}dx = C\text{poly}\log(m)m^{1-\alpha}, \tag{88}$$

since $\alpha > 1$. Since $m$ grows with $n$, we have that $\text{Tr}(K_{>m}) \to 0$ with $n$, and therefore we need the ridge regularization parameter $\lambda > 0$ in order to have $\frac{r_m(\lambda)}{n} \not\to 0$ in Lemma B.2.

The lower and higher order conditions, and the conditions in Lemma B.7 still hold in this setting. Therefore, we the deterministic equivalents are still valid as long as we have the ridge regularization $\lambda > c$ for some constant $c$ independent of $n, d$.

## C Reparametrization of the Kernel Eigenvalues

By Proposition 1, we know that in the high-dimensional limit, the kernel eigenvalues can be indexed at leading order by a multi-index $\boldsymbol{\beta} \in \mathbb{Z}^d_{\geq 0}$:

$$\lambda_{\boldsymbol{\beta}} = h_{|\boldsymbol{\beta}|}|\boldsymbol{\beta}|!\, r_\alpha(d)^{-|\boldsymbol{\beta}|}\prod_{j=1}^d j^{-\alpha\beta_j}$$

assuming $\sigma_j = r_\alpha(d)^{-1}j^{-\alpha}$. To evaluate the deterministic equivalents, it is convenient to reparameterize these eigenvalues using a new set of variables over which we can sum more naturally. Specifically, we can group the eigenvalues using $m = |\boldsymbol{\beta}|$ that can be seen as a single *shell*. Within each shell of constant $m$, we can map $\boldsymbol{\beta}$ to a set of $m$ indices $i_1, \dots, i_m \in \{1, \dots, d\}$ with $1 \leq i_1 \leq \cdots \leq i_m \leq d$ defined such that:

$$\beta_j = \sum_{k=1}^m \delta_{i_k,j} \tag{89}$$

Equivalently, the index $i_k$ simply tracks the active dimensions composing the multi-index $\boldsymbol{\beta}$. Using this representation, the eigenvalues can be rewritten as:

$$\lambda^{(m)}_{i_1,\dots,i_m} = h_m m!\, r_\alpha(d)^{-m} (i_1 \cdots i_m)^{-\alpha}$$

where $m \in \mathbb{N}$ and the indices are ordered as $1 \le i_1 \le \cdots \le i_m \le d$. Consequently, summations over the entire spectrum can be recast as:

$$\sum_{\boldsymbol{\beta}\in\mathbb{Z}^d_{\ge 0}} F(\lambda_{\boldsymbol{\beta}}) = \sum_{m=0}^{\infty} \sum_{1\le i_1\le\cdots\le i_m\le d} F\big(\lambda^{(m)}_{i_1,\dots,i_m}\big)$$

for a generic function $F$. While conceptually useful, this formulation still presents analytical difficulties due to the constrained ordered summation. To circumvent this issue, we can relax the strict ordering and extend the sum to a free summation over all possible index permutations:

$$\sum_{1\le i_1\le\cdots\le i_m\le d} F\big(\lambda^{(m)}_{i_1,\dots,i_m}\big) = \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} F\big(\lambda^{(m)}_{i_1,\dots,i_m}\big) \prod_{j=1}^{d} \beta_j! \tag{90}$$

The combinatorial factor $1/m!$ accounts for the permutations of the $m$ free indices, while the correction term $\prod_{j=1}^{d} \beta_j!$ (with $\beta_j$ determined by Eq. (89)) accurately compensates for the multinomial overcounting of "diagonal" elements, where two or more indices coincide. However, in the high-dimensional limit ($d \to \infty$), the number of tuples composed of at least one repeated index becomes strictly negligible. The sum is overwhelmingly dominated by bulk configurations where $\beta_j \in \{0, 1\}$, for which the correction factor reduces identically to 1. Since the contribution of the diagonal elements is asymptotically subleading, we can drop the factorial terms and approximate the trace with a pure free summation:

$$\sum_{1\le i_1\le\cdots\le i_m\le d} F\big(\lambda^{(m)}_{i_1,\dots,i_m}\big) \simeq \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} F\big(\lambda^{(m)}_{i_1,\dots,i_m}\big)$$

A pictorial representation of this reparameterization is provided in Fig. 4.

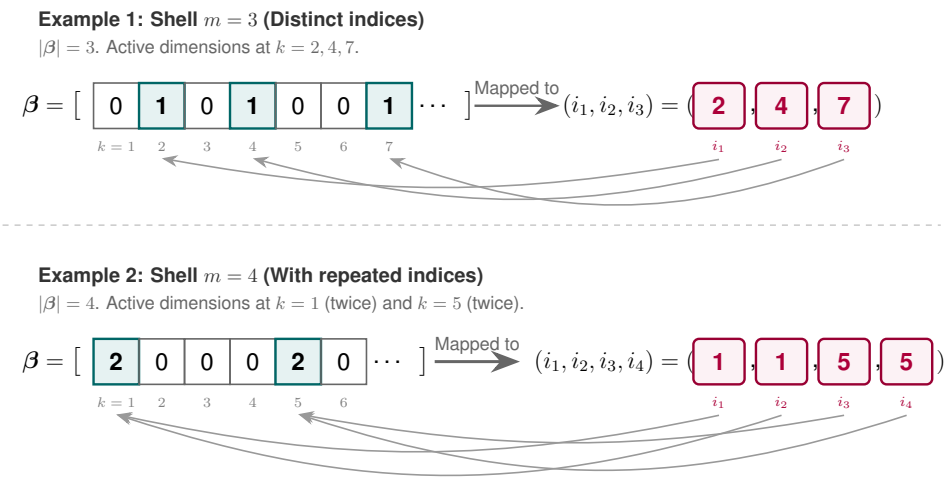


Figure 4: Representation of the reparameterization from the multi-index $\boldsymbol{\beta}$ to the ordered sequence $(i_1, \dots, i_m)$. The top panel shows a standard "bulk" configuration without index collisions ($\beta_k \in \{0, 1\}$). The bottom panel illustrates a "diagonal" configuration where index collisions result in values of $\beta_k > 1$.

A similar approach can be employed to represent the target function $f_\star$ in the kernel eigenbasis. As seen in Section 3.1.2, we can re-write $f_\star$ in the polynomial basis that diagonalises the kernel operator $e_{\boldsymbol{\beta}}(x)$:

$$f_\star(\boldsymbol{x}) = \sum_{\boldsymbol{\beta}\in\mathbb{Z}^d_{\ge 0}} \theta_{\boldsymbol{\beta}} e_{\boldsymbol{\beta}}(\boldsymbol{x})$$

Grouping these functions by their polynomial shell $m = |\boldsymbol{\beta}|$, we can rewrite the target function as:

$$f_\star(x) = \sum_{m=0}^{\infty} \sum_{|\boldsymbol{\beta}|=m} \theta_{\boldsymbol{\beta}} e_{\boldsymbol{\beta}}(\boldsymbol{x}) = \sum_{m=0}^{\infty} \sum_{1 \le i_1 \le \cdots \le i_m \le d} \theta_{i_1,\ldots,i_m} e_{i_1,\ldots,i_m}(\boldsymbol{x}) \tag{91}$$

To convert this strictly ordered sum into a free summation over $\{1, \ldots, d\}^m$, the sequence of scalar weights $\theta_{i_1,\ldots,i_m}$, originally defined only for ordered index tuples, must be symmetrically extended. By defining the extended weights to be invariant under any permutation of the indices $(i_1, \ldots, i_m)$, we can safely drop the ordering constraint. Neglecting again the multinomial corrections from the diagonal elements for dense targets, we obtain the free summation form:

$$f_\star(x) \simeq \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{i_1,\ldots,i_m=1}^{d} \theta_{i_1,\ldots,i_m} e_{i_1,\ldots,i_m}(\boldsymbol{x}) \tag{92}$$

**Remark 5.** The free summation in Eq. (92) relies on the assumption that the bulk terms ($\beta_j \in \{0, 1\}$) overwhelmingly dominate the trace. This holds strictly true for "delocalized" or "dense" target functions, where the spectral energy is distributed across a vast number of tuples. However, if $f_\star$ is highly concentrated on a few specific diagonal features (e.g., purely dependent on a single coordinate raised to the $m$-th power, implying $\beta_j = m$), the diagonal correction factor $\prod_{j=1}^{d} \beta_j!$ from Eq. (90) can no longer be ignored and we must take care of this correction.

# D Weakly anisotropic regime

In this appendix, our goal is to prove mathematically the formulas shown in Sec. 3.1 for the weakly anisotropic regime.

## D.1 Kernel State Equation

Consider Eq. (8) that reads:

$$n - \frac{\lambda}{\nu} = \mathrm{Tr}(\boldsymbol{\Lambda}(\boldsymbol{\Lambda} + \nu \boldsymbol{I})^{-1})$$

Under the polynomial scaling $n = \psi d^\kappa$ and using the results of Appendix C, we can unroll the Tr operator and rewrite the self-consistency equation for $\nu$ as

$$\begin{aligned}
\psi &= \frac{\lambda}{\nu} d^{-\kappa} + d^{-\kappa} \sum_{\boldsymbol{\beta}} \frac{\lambda_{\boldsymbol{\beta}}}{\lambda_{\boldsymbol{\beta}} + \nu} = \\
&= \frac{\lambda}{\nu} d^{-\kappa} + \sum_{m=0}^{\infty} \frac{d^{-\kappa}}{m!} \sum_{i_1,\ldots,i_m=1}^{d} \frac{1}{1 + \nu (i_1 \ldots i_m)^\alpha (h_m m!)^{-1} r_\alpha(d)^m} = \\
&= \frac{\lambda}{\nu} d^{-\kappa} + \sum_{m=0}^{\infty} \psi^{(m)}
\end{aligned} \tag{93}$$

where we define:

$$\psi^{(m)} = \frac{1}{m!} \sum_{i_1,\ldots,i_m=1}^{d} \frac{d^{-\kappa}}{1 + \nu (i_1 \ldots i_m)^\alpha (h_m m!)^{-1} r_\alpha(d)^m}$$

In the regime where $\alpha < 1$, then $r_\alpha(d) = (1 - \alpha)^{-1} d^{1-\alpha}$ at leading order. Let us denote by $\delta_m = h_m m!$, so that:

$$\psi^{(m)} = \frac{1}{m!} \sum_{i_1,\ldots,i_m=1}^{d} \frac{d^{-\kappa}}{1 + \nu (i_1 \ldots i_m)^\alpha d^{m(1-\alpha)} \delta_m^{-1} (1 - \alpha)^{-m}} \tag{94}$$

To solve the self-consistency equation $\psi = \frac{\lambda}{\nu} d^{-\kappa} + \sum_m \psi^{(m)}$, we need to formulate an ansatz on the form of $\nu$. In particolar, we will assume:

$$\nu = \xi d^{-\kappa}$$

where $\xi$ is of order 1 with respect to $d$, $\xi \asymp 1$, and should only depend on $\alpha, \psi, \lambda, \{\delta_m\}$. Under this ansatz, the generic term $\psi^{(m)}$ becomes:

$$\psi^{(m)} = \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{d^{-\kappa}}{1 + (i_1 \dots i_m)^{\alpha} d^{m(1-\alpha)-\kappa} \xi \delta_m^{-1} (1-\alpha)^{-m}} \tag{95}$$

The behavior of $\psi^{(m)}$ depends on $m$. In particular, we can distinguish different regimes depending on the relative value of $m$ with respect to $\kappa \in \mathbb{N}$:

**Case a: $m < \kappa$** Since $i_k \geq 1$, then $(i_1 \cdots i_m)^{\alpha} \geq 1$. This means that we can bound Eq. (95):

$$\begin{aligned} \psi^{(m)} &\leq \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{d^{-\kappa}}{1 + d^{m(1-\alpha)-\kappa} \xi \delta_m^{-1} (1-\alpha)^{-m}} = \\ &= \frac{1}{m!} \frac{d^{m-\kappa}}{1 + d^{m(1-\alpha)-\kappa} \xi \delta_m^{-1} (1-\alpha)^{-m}} \end{aligned} \tag{96}$$

Since $m < \kappa$, then $m(1-\alpha) - \kappa < 0$. This means that when $d \to \infty$, this upper bound simplifies to:

$$\psi^{(m)} \leq \frac{1}{m!} \frac{d^{m-\kappa}}{1 + d^{m(1-\alpha)-\kappa} \xi \delta_m^{-1} (1-\alpha)^{-m}} \sim \frac{1}{m!} d^{m-\kappa} \implies \psi^{(m)} = O(d^{m-\kappa})$$

Similarly, we have that $(i_1 \dots i_m)^{\alpha} \leq d^{m\alpha}$ as $i_k \leq d$ so that:

$$\begin{aligned} \psi^{(m)} &\geq \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{d^{-\kappa}}{1 + d^{m-\kappa} \xi \delta_m^{-1} (1-\alpha)^{-m}} = \\ &= \frac{1}{m!} \frac{d^{m-\kappa}}{1 + d^{m-\kappa} \xi \delta_m^{-1} (1-\alpha)^{-m}} \,. \end{aligned}$$

Again, since $m < \kappa$, the term $d^{m-\kappa} \to 0$ as $d \to \infty$. Thus, the denominator converges to 1:

$$\psi^{(m)} \geq \frac{1}{m!} d^{m-\kappa} \implies \psi^{(m)} = \Omega(d^{m-\kappa}) \,.$$

As a consequence, $\psi^{(m)} = \Theta(d^{m-\kappa})$ and vanishes when $d \to \infty$ for $m < \kappa$. The contribution on the full summation of the under-resonating $m < \kappa$ terms is therefore negligible.

**The integral regime and the Irwin-Hall trick** For values of $m \geq \kappa$, the direct bounding argument employed above is no longer sufficiently tight. However, as $d \to \infty$, the discrete indices $i_q$ can be treated as continuous variables $x_q \in [1, d]$. Since the summand varies slowly over the integer grid, we assume we can approximate the discrete summation with an $m$-dimensional Riemann integral:

$$\psi^{(m)} \simeq \frac{1}{m!} \int_1^d \cdots \int_1^d \mathrm{d}x_1 \cdots \mathrm{d}x_m \, \frac{d^{-\kappa}}{1 + (x_1 \cdots x_m)^{\alpha} d^{m(1-\alpha)} \nu C_m} \,. \tag{97}$$

with $C_m = \delta_m^{-1} (1-\alpha)^{-m}$. Upon introducing the change of variables $y_i = x_i \, C_m^{\frac{1}{m\alpha}} d^{\frac{1-\alpha}{\alpha}} \nu^{\frac{1}{m\alpha}}$ and recalling the ansatz $\nu = \xi d^{-\kappa}$, we obtain:

$$\psi^{(m)} = \frac{1}{m!} C_m^{-1/\alpha} \xi^{-1/\alpha} d^{(m-\kappa)(1-1/\alpha)} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{1 + (y_1 \cdots y_m)^{\alpha}} \,, \tag{98}$$

where the integration boundaries explicitly read:

$$y_{\min} = C_m^{\frac{1}{m\alpha}} d^{\frac{1-\alpha}{\alpha}} \nu^{\frac{1}{m\alpha}} = C_m^{\frac{1}{m\alpha}} \xi^{\frac{1}{m\alpha}} d^{\frac{1}{\alpha}\left(1 - \frac{\kappa}{m} - \alpha\right)} \,, \tag{99a}$$

$$y_{\max} = C_m^{\frac{1}{m\alpha}} d^{\frac{1-\alpha}{\alpha}} \nu^{\frac{1}{m\alpha}} d = C_m^{\frac{1}{m\alpha}} \xi^{\frac{1}{m\alpha}} d^{\frac{1}{\alpha}\left(1 - \frac{\kappa}{m}\right)} \,. \tag{99b}$$

The asymptotic behavior of $y_{\min}$ and $y_{\max}$ as $d \to \infty$ is illustrated in Fig. 5.

Having established that the $m < \kappa$ terms vanish, we restrict our focus to the remaining shells. Before evaluating these specific branches, we establish a key mathematical identity that will repeatedly emerge:

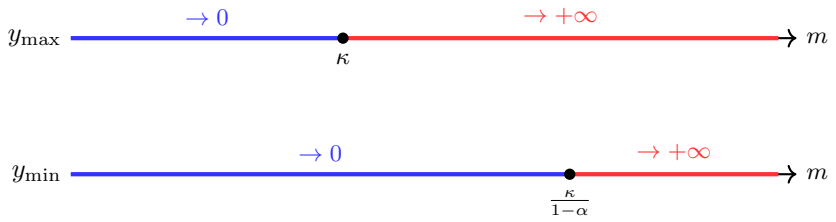


Figure 5: *Asymptotic behavior of the integration boundaries $y_{\min}$ and $y_{\max}$ derived in Eq. (99). Notably, when $m = \kappa$ or $m = \frac{\kappa}{1-\alpha} \in \mathbb{N}$, the respective boundary converges to a strictly positive, macroscopic constant $\Theta_d(1)$.*

**Lemma D.1** (*Irwin-Hall Trick*)**.** Let $y_{\min}, y_{\max} \in \mathbb{R}^+$ such that $y_{\max} = y_{\min} d$. Consider the $m$-dimensional integral defined as:

$$\mathcal{I} = \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} f(y_1 \cdots y_m)\, \mathrm{d}y_1 \cdots \mathrm{d}y_m \,. \tag{100}$$

for some regular enough $f : \mathbb{R} \to \mathbb{R}$. Then, denoting $l = \log(y_{\min})$ and $L = \log(y_{\max})$, this integral can be exactly rewritten through a 1-dimensional integral as:

$$\mathcal{I} = \log(d)^{m-1} \int_{ml}^{mL} \mathrm{d}S\, e^S f(e^S) f_{IH}\left(\frac{S - ml}{\log(d)}; m\right) ,$$

or, equivalently via variable substitution:

$$\mathcal{I} = \log(d)^m \int_0^m \mathrm{d}x\, e^{x\log(d)+ml} f(e^{x\log(d)+ml}) f_{IH}(x; m) \,,$$

where $f_{IH}$ is the Irwin-Hall probability density function:

$$f_{IH}(x; m) = \frac{1}{(m-1)!} \sum_{k=0}^{\lfloor x \rfloor} (-1)^k \binom{m}{k} (x-k)^{m-1} \,.$$

*Proof.* Let us apply the change of variables $y_i = e^{u_i}$ to Eq. (100):

$$\mathcal{I} = \int_l^L \cdots \int_l^L e^{u_1+\cdots+u_m} f(e^{u_1+\cdots+u_m})\, \mathrm{d}u_1 \cdots \mathrm{d}u_m \,, \tag{101}$$

where $l = \log(y_{\min})$ and $L = \log(y_{\max})$. The core insight is to treat $u_1, \ldots, u_m$ as independent random variables uniformly distributed in the interval $[l, L]$, allowing us to interpret Eq. (101) as an expectation value:

$$\mathcal{I} = (L-l)^m\, \mathbb{E}_{u_i \sim \mathcal{U}[l,L]}\left[e^{u_1+\cdots+u_m} f(e^{u_1+\cdots+u_m})\right] .$$

By defining the sum variable $S = \sum_i u_i$, which represents the sum of $m$ independent and uniformly distributed random variables, and noticing that $L - l = \log(y_{\max}/y_{\min}) = \log(d)$, we can rewrite this expectation as:

$$\mathcal{I} = \log(d)^m\, \mathbb{E}_S\left[e^S f(e^S)\right] .$$

The probability distribution for the sum of $m$ independent standard uniform variables $z_i \sim \mathcal{U}[0,1]$ is a well-known result and corresponds to the Irwin-Hall distribution. From this base distribution, it is straightforward to deduce the probability density function for our shifted and scaled sum $S$:

$$p_S(S) = \log(d)^{-1} f_{IH}\left(\frac{S - ml}{\log(d)}; m\right) ,$$

where the standard Irwin-Hall density is defined as:

$$f_{IH}(x; m) = \frac{1}{(m-1)!} \sum_{k=0}^{\lfloor x \rfloor} (-1)^k \binom{m}{k} (x-k)^{m-1} \,.$$

Integrating over the support of $S$ yields the first form of the integral:

$$\mathcal{I} = \log(d)^{m-1} \int_{ml}^{mL} dS\, e^S f(e^S) f_{IH}\left(\frac{S - ml}{\log(d)}; m\right).$$

Alternatively, introducing the substitution $x = \frac{S-ml}{\log(d)}$, we directly obtain the equivalent formulation:

$$\mathcal{I} = \log(d)^m \int_0^m dx\, e^{x\log(d)+ml} f(e^{x\log(d)+ml}) f_{IH}(x; m).$$

□

Equipped with Lemma D.1, we can now systematically analyze the specific branches for $m \geq \kappa$ by turning $m$-dimensional integrals into simpler one-dimensional integrals.

**Case b: $m = \kappa$** When $m = \kappa$, we have from Eq. (99) that $y_{\min} \to 0$ and $y_{\max} = (C_\kappa \xi)^{1/(\kappa\alpha)} = \Theta_d(1)$. From Eq. (98), the resonant term reads:

$$\psi^{(\kappa)} = \frac{1}{\kappa!}(C_\kappa\xi)^{-1/\alpha} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{1 + (y_1 \cdots y_m)^\alpha} = \frac{1}{\kappa!}(C_\kappa\xi)^{-1/\alpha}\mathcal{I}.$$

Applying Lemma D.1, we can exactly recast the integral $\mathcal{I}$ as:

$$\mathcal{I} = \log(d)^\kappa \int_0^\kappa dx\, \frac{e^{x\log(d)+\kappa\log(y_{\min})}}{1 + e^{\alpha x\log(d)+\alpha\kappa\log(y_{\min})}} f_{IH}(x; \kappa). \tag{102}$$

Exploiting the relation $y_{\max} = y_{\min} d$, which implies $\log(y_{\min}) = \log(y_{\max}) - \log(d)$, we obtain:

$$\begin{aligned}\mathcal{I} &= \log(d)^\kappa y_{\max}^\kappa \int_0^\kappa \mathrm{d}x\, \frac{e^{\log(d)(x-\kappa)}}{1 + y_{\max}^{\alpha\kappa} e^{\log(d)\alpha(x-\kappa)}} f_{IH}(x; \kappa) \\ &= \log(d)^\kappa (C_\kappa\xi)^{1/\alpha} \int_0^\kappa \mathrm{d}x\, \frac{e^{\log(d)(x-\kappa)}}{1 + C_\kappa\xi e^{\log(d)\alpha(x-\kappa)}} f_{IH}(x; \kappa).\end{aligned}$$

Finally, introducing the change of variables $t = \log(d)(\kappa - x)$ yields:

$$\mathcal{I} = \log(d)^{\kappa-1}(C_\kappa\xi)^{1/\alpha} \int_0^{\kappa\log(d)} \mathrm{d}t\, \frac{e^{-t}}{1 + C_\kappa\xi e^{-\alpha t}} f_{IH}\left(\kappa - \frac{t}{\log(d)}; \kappa\right). \tag{103}$$

We now compute the asymptotic limit of this integral as $d \to \infty$ using the Dominated Convergence Theorem (DCT). To this end, we recast $\mathcal{I}$ as:

$$\mathcal{I} = (C_\kappa\xi)^{1/\alpha} \int_{\mathbb{R}^+} g_d(t)\, \mathrm{d}t,$$

where:

$$g_d(t) = \mathbf{1}_{[0,\kappa\log(d)]}(t)\, \log(d)^{\kappa-1} \frac{e^{-t}}{1 + C_\kappa\xi e^{-\alpha t}} f_{IH}\left(\kappa - \frac{t}{\log(d)}; \kappa\right),$$

and $\mathbf{1}_{[0,\kappa\log(d)]}(t)$ denotes the indicator function of the domain $[0, \kappa\log(d)]$.

To evaluate the pointwise limit of $g_d(t)$, we exploit the asymptotic expansion of the Irwin-Hall distribution near its rightmost boundary:

$$f_{IH}(\kappa - \epsilon; \kappa) = \frac{\epsilon^{\kappa-1}}{(\kappa-1)!} \quad \text{as } \epsilon \to 0.$$

This can be easily seen from its definition and from the fact that $f_{IH}(\epsilon; \kappa) = f_{IH}(\kappa - \epsilon; \kappa)$. Setting $\epsilon = \frac{t}{\log(d)}$, the diverging logarithmic factor cancels out as $d \to \infty$, yielding:

$$\lim_{d\to\infty} g_d(t) = \lim_{d\to\infty} \log(d)^{\kappa-1} \frac{e^{-t}}{1 + C_\kappa\xi e^{-\alpha t}} \left(\frac{t}{\log(d)}\right)^{\kappa-1} \frac{1}{(\kappa-1)!} = \frac{1}{(\kappa-1)!} \frac{e^{-t}t^{\kappa-1}}{1 + C_\kappa\xi e^{-\alpha t}}.$$

To apply the DCT, we must find a dominating function $G(t) \geq |g_d(t)|$ independent of $d$. Leveraging the global inequality for the Irwin-Hall distribution $f_{IH}(\kappa - x; \kappa) \leq x^{\kappa-1}$ (valid for all $x \geq 0$), we bound $g_d(t)$ as:

$$|g_d(t)| \leq \log(d)^{\kappa-1} \frac{e^{-t}}{1 + C_\kappa \xi e^{-\alpha t}} \left( \frac{t}{\log(d)} \right)^{\kappa-1} = \frac{e^{-t} t^{\kappa-1}}{1 + C_\kappa \xi e^{-\alpha t}} \equiv G(t)\,,$$

which is strictly integrable over $\mathbb{R}^+$ since $\alpha < 1$. Since the hypotheses of the DCT are satisfied, we can exchange the limit and the integral:

$$\lim_{d\to\infty} \mathcal{I} = (C_\kappa \xi)^{1/\alpha} \int_0^\infty \left( \lim_{d\to\infty} g_d(t) \right) dt = (C_\kappa \xi)^{1/\alpha} \frac{1}{(\kappa-1)!} \int_0^\infty \frac{e^{-t} t^{\kappa-1}}{1 + C_\kappa \xi e^{-\alpha t}}\, dt\,.$$

Substituting this back into the expression for $\psi^{(\kappa)}$, we obtain:

$$\psi^{(\kappa)} = \frac{1}{\kappa!} (C_\kappa \xi)^{-1/\alpha} \mathcal{I} = \frac{1}{\kappa!(\kappa-1)!} \int_0^\infty \frac{e^{-t} t^{\kappa-1}}{1 + C_\kappa \xi e^{-\alpha t}}\, dt\,.$$

Finally, expanding the constant $C_\kappa = \delta_\kappa^{-1} (1-\alpha)^{-\kappa} = \frac{(1-\alpha)^{-\kappa}}{h_\kappa \kappa!}$, we arrive at the final form:

$$\psi^{(\kappa)} = \frac{1}{\kappa!(\kappa-1)!} \int_0^\infty dt\, e^{-t} t^{\kappa-1} \frac{1}{1 + \frac{(1-\alpha)^{-\kappa}}{h_\kappa \kappa!} \xi e^{-\alpha t}}\,.$$

This quantity is rigorously well-defined, and the integral securely converges to a finite macroscopic value. This aligns with the physical constraint dictated by the state equation, which requires $\psi^{(m)} \leq \psi = \Theta_d(1)$.

**Case c: $m > \frac{\kappa}{1-\alpha}$** In this regime, both integration boundaries diverge to infinity ($y_{\min} \to \infty$ and $y_{\max} \to \infty$). However, since $y_{\max} = y_{\min} d$, the upper limit scales significantly faster than the lower one. As $d \to \infty$, the unity term in the denominator becomes negligible compared to the polynomial growth of $(y_1 \cdots y_m)^\alpha$. We can therefore asymptotically approximate the integral in Eq. (98) as follows:

$$\begin{aligned} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{1 + (y_1 \cdots y_m)^\alpha} &\simeq \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} (y_1 \cdots y_m)^{-\alpha}\, \mathrm{d}y_1 \cdots \mathrm{d}y_m \\ &= (1-\alpha)^{-m} \left( y_{\max}^{1-\alpha} - y_{\min}^{1-\alpha} \right)^m \\ &\simeq (1-\alpha)^{-m} y_{\max}^{m(1-\alpha)} \\ &= (1-\alpha)^{-m} C_m^{\frac{1-\alpha}{\alpha}} \xi^{\frac{1-\alpha}{\alpha}} d^{\frac{1-\alpha}{\alpha}(m-\kappa)}\,. \end{aligned} \tag{104}$$

Substituting this approximation back into Eq. (98), the fractional scalings in $d$ perfectly cancel out, yielding a dimension-free expression:

$$\psi^{(m)} = \frac{1}{m!} C_m^{-1} \xi^{-1} (1-\alpha)^{-m}\,.$$

Recalling the definition of the constant $C_m = \delta_m^{-1} (1-\alpha)^{-m} = (h_m m!)^{-1} (1-\alpha)^{-m}$, we finally isolate the self-regularizing contribution of these higher-degree modes:

$$\psi^{(m)} = \frac{h_m}{\xi} \quad \text{for} \quad m > \frac{\kappa}{1-\alpha}\,.$$

**Case d: $\kappa < m \leq \frac{\kappa}{1-\alpha}$** In this intermediate band, the integration limits exhibit a mixed asymptotic behavior: the upper limit diverges ($y_{\max} \to \infty$), while the lower limit $y_{\min}$ either vanishes to 0 or converges to a strictly positive constant (when $\frac{\kappa}{1-\alpha} \in \mathbb{N}$). Nevertheless, the integral is heavily ultraviolet-dominated (see Eq. (104), since $y_{\min}$ is raised to the power of $1 - \alpha > 0$). The leading-order asymptotic contribution is overwhelmingly driven by the diverging upper boundary, so that the precise behavior near the origin is asymptotically negligible. Therefore, the evaluation rigorously follows the exact same asymptotic expansion as the super-critical Case c, yielding:

$$\psi^{(m)} = \frac{h_m}{\xi}\,.$$

**Synthesis: The Macroscopic Kernel State Equation** We can finally assemble the self-consistency equation. By substituting the results from the various regimes analyzed above into Eq. (93), we obtain

$$\begin{aligned}\psi &= \frac{\lambda}{\xi} + 0 + \psi^{(\kappa)} + \sum_{m>\kappa} \frac{h_m}{\xi} \\ &= \frac{1}{\xi}\Big(\lambda + \sum_{m>\kappa} h_m\Big) + \psi^{(\kappa)} .\end{aligned} \tag{105}$$

We identify the term in parentheses as the *effective ridge regularization* generated by the higher-order tails, $\lambda_{\text{eff}} = \lambda + \sum_{m>\kappa} h_m$. Substituting the explicit integral evaluation for $\psi^{(\kappa)}$, we obtain the kernel state equation that implicitly determines the macroscopic parameter $\xi$:

$$\psi = \frac{\lambda_{\text{eff}}}{\xi} + \frac{1}{\kappa!(\kappa-1)!}\int_0^\infty dt\, e^{-t} t^{\kappa-1} \frac{1}{1 + \frac{(1-\alpha)^{-\kappa}}{h_\kappa \kappa!}\xi e^{-\alpha t}} . \tag{106}$$

□

### D.2 Variance

The variance V is given by Eq. (7) and can be rewritten as $\mathsf{V} = \sigma_\varepsilon^2 \frac{\tau}{1-\tau}$ where:

$$\tau = \frac{1}{n}\operatorname{Tr}(\mathbf{\Lambda}^2(\mathbf{\Lambda} + \nu_\star \boldsymbol{I})^{-2})$$

with $\nu_\star = \xi_\star d^{-\kappa}$ and $\xi_\star$ is the solution of Eq. (106). Under the polynomial scaling $n = \psi d^\kappa$ and using the findings in Appendix C, we can write $\tau$ as a sum over the shells:

$$\tau = \sum_{m=0}^{\infty} \tau^{(m)}$$

where, using the same notation as in Appendix D.1:

$$\tau^{(m)} = \frac{d^{-\kappa}}{\psi m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{1}{(1 + C_m d^{m(1-\alpha)}(i_1 \dots i_m)^\alpha \nu)^2} \tag{107}$$

**Case a: $m < \kappa$** For the lower-degree shells, we can employ the exact same bounding argument used in Appendix D.1 to conclude that $\tau^{(m)} = \Theta(d^{m-\kappa})$. Consequently, since $m < \kappa$, these terms strictly vanish in the thermodynamic limit: $\tau^{(m)} \to 0$.

**The integral regime** As established previously, for $m \geq \kappa$ we can approximate the discrete summation in Eq. (107) with a Riemann integral. Introducing the coordinate transformation $y_i = x_i\, C_m^{\frac{1}{m\alpha}}\, d^{\frac{1-\alpha}{\alpha}}\, \nu^{\frac{1}{m\alpha}}$ and substituting the ansatz $\nu = \xi d^{-\kappa}$, we obtain:

$$\begin{aligned}\tau^{(m)} &\simeq \frac{d^{-\kappa}}{\psi m!}\int_1^d \cdots \int_1^d \mathrm{d}x_1 \cdots \mathrm{d}x_m \frac{1}{\big(1 + C_m d^{m(1-\alpha)}(x_1 \cdots x_m)^\alpha \nu_\star\big)^2} \\ &= \frac{1}{\psi m!} C_m^{-1/\alpha} \xi^{-1/\alpha} d^{(m-\kappa)(1-1/\alpha)} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{\big(1 + (y_1 \cdots y_m)^\alpha\big)^2} .\end{aligned} \tag{108}$$

The integration boundaries $y_{\min}$ and $y_{\max}$ are identical to those defined in Eqs. equation 99. According to the asymptotic behavior sketched in Fig. 5, we will again distinguish the following regimes:

**Case b: $m > \frac{\kappa}{1-\alpha}$** In this regime, both boundaries diverge ($y_{\min}, y_{\max} \to \infty$). Thus, the unity term in the denominator becomes negligible, allowing us to asymptotically approximate the integral:

$$\begin{aligned}\int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{\big(1 + (y_1 \cdots y_m)^\alpha\big)^2} &\simeq \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} (y_1 \cdots y_m)^{-2\alpha}\, \mathrm{d}y_1 \cdots \mathrm{d}y_m \\ &= (1-2\alpha)^{-m}\big(y_{\max}^{1-2\alpha} - y_{\min}^{1-2\alpha}\big)^m .\end{aligned} \tag{109}$$

The dominant contribution to this expression depends crucially on the sign of $1 - 2\alpha$:

- When $\alpha < \frac{1}{2}$, the integral is asymptotically dominated by the upper limit $y_{\max}$. Substituting this back into Eq. (108) yields:

$$\tau^{(m)} \asymp d^{(m-\kappa)(1-1/\alpha)} y_{\max}^{m(1-2\alpha)} \asymp d^{(m-\kappa)(1-1/\alpha)} d^{\frac{1-2\alpha}{\alpha}(m-\kappa)} = d^{\kappa-m} \to 0\,.$$

  Even when evaluating the infinite sum $\sum d^{\kappa-m}$, the series is bounded by its largest geometric term, behaving as $\Theta(d^{\kappa-\lceil \frac{\kappa}{1-\alpha} \rceil})$, which is asymptotically negligible.
- When $\alpha = \frac{1}{2}$, the integral in Eq. (109) scales logarithmically as $\log(y_{\max}/y_{\min})^m \asymp \log(d)^m$. Overall, we obtain $\tau^{(m)} \asymp d^{\kappa-m}\log(d)^m \to 0$, as the polynomial decay suppresses the poly-logarithmic factor.
- Finally, when $\alpha > \frac{1}{2}$, the dominating term is the lower boundary $y_{\min}$, causing the integral to scale as $y_{\min}^{m(1-2\alpha)} \asymp d^{\frac{1-2\alpha}{\alpha}(m-\kappa-m\alpha)}$. Putting it all together:

$$\tau^{(m)} \asymp d^{\frac{1-2\alpha}{\alpha}(m-\kappa-m\alpha)} d^{(m-\kappa)(1-1/\alpha)} = d^{\kappa-2m(1-\alpha)}\,.$$

  By definition of this regime, $m > \frac{\kappa}{1-\alpha}$, meaning the exponent $\kappa - 2m(1-\alpha)$ is strictly negative, leading to $\tau^{(m)} \to 0$.

We can then conclude that the entire super-critical tail is negligible: $\sum_{m > \frac{\kappa}{1-\alpha}} \tau^{(m)} \to 0$.

**Case c: $\boldsymbol{\kappa < m \leq \frac{\kappa}{1-\alpha}}$** In this intermediate band, $y_{\min} \to 0$ while $y_{\max} \to \infty$. For $\alpha < \frac{1}{2}$, as shown in the previous case, the integral is strictly ultraviolet-dominated by the diverging upper limit $y_{\max}$. Thus, its asymptotic behavior is independent of the origin, directly yielding $\tau^{(m)} \asymp d^{\kappa-m} \to 0$.

Conversely, for $\alpha \geq \frac{1}{2}$, it becomes crucial to account for the integrand's behavior near the origin. We rely once again on the Irwin-Hall representation (Lemma D.1). Rewriting the integral over the sum variable $S$, we obtain:

$$\int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{\left(1 + (y_1 \cdots y_m)^\alpha\right)^2} = \log(d)^{m-1} \int_{ml}^{mL} dS\, \frac{e^S}{(1+e^{\alpha S})^2} f_{IH}\left(\frac{S - ml}{\log d}; m\right)\,. \tag{110}$$

For $\alpha > \frac{1}{2}$, the function $g(S) = \frac{e^S}{(1+e^{\alpha S})^2}$ is integrable over the entire real line. Its exponential tails rapidly suppress the integral away from the origin. Following the same procedure leveraging the DCT as we did in the state equation, we obtain that the overall term scales as:

$$\tau^{(m)} \asymp d^{(m-\kappa)(1-1/\alpha)} \log(d)^{m-1} = \widetilde{\Theta}\left(d^{(m-\kappa)(1-1/\alpha)}\right)\,. \tag{111}$$

Since the exponent $(m-\kappa)(1-1/\alpha)$ is strictly negative, this power-law decay collapses the term to zero.

When $\alpha = \frac{1}{2}$, the integrand exhibits a constant asymptote $g(S) \sim 1$ as $S \to \infty$. Consequently, integrating up to $mL$ yields an additional dimension-dependent factor $\mathcal{O}(\log d)$. Plugging this into Eq. (110), the total term scales as $\tau^{(m)} \asymp d^{-(m-\kappa)} \log(d)^m$. Since $m > \kappa$, the polynomial suppression effortlessly overcomes the logarithmic growth, ensuring $\tau^{(m)} \to 0$.

Hence, throughout the entire intermediate regime, $\tau^{(m)} \to 0$. Depending on the value of $\alpha$, the decay rate is bounded by either $\mathcal{O}(d^{\kappa-m})$ or $\widetilde{\mathcal{O}}(d^{(\kappa-m)(1/\alpha-1)})$.

**Case d: $\boldsymbol{m = \kappa}$** Here $y_{\min} \to 0$ and $y_{\max} = (C_\kappa \xi)^{1/\alpha} = \Theta_d(1)$. The integral reads:

$$\tau^{(\kappa)} = \frac{1}{\kappa!\psi}(C_\kappa \xi)^{-1/\alpha} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{\left(1 + (y_1 \cdots y_m)^\alpha\right)^2}\,.$$

Applying Lemma D.1, we transform this $m$-dimensional integral into a 1-dimensional one:

$$\int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{\left(1 + (y_1 \cdots y_m)^\alpha\right)^2} = \log(d)^\kappa \int_0^\kappa dx\, \frac{e^{x\log(d) + \kappa \log(y_{\min})}}{\left(1 + e^{\alpha x \log(d) + \alpha\kappa \log(y_{\min})}\right)^2} f_{IH}(x;\kappa)\,.$$

The asymptotic evaluation from this point precisely mirrors the derivations in Appendix D.1 for the $m = \kappa$ case (see Eq. (102)). Executing the high-dimensional limit $d \to \infty$, the logarithmic prefactors exactly cancel out, yielding:

$$\tau^{(\kappa)} = \frac{1}{\psi} \frac{1}{\kappa!(\kappa-1)!} \int_0^\infty dt\, e^{-t} t^{\kappa-1} \frac{1}{\left(1 + \frac{(1-\alpha)^{-\kappa}}{h_\kappa \kappa!} \xi e^{-\alpha t}\right)^2}. \tag{112}$$

**Synthesis: Final formula for the variance** Combining the results together, we have that the only contribution to Eq. (107) that survives in the high-dimensional limit $d \to \infty$ is the $m = \kappa$, so that:

$$\tau = \frac{1}{\psi} \frac{1}{\kappa!(\kappa-1)!} \int_0^\infty dt\, e^{-t} t^{\kappa-1} \frac{1}{(1 + \frac{(1-\alpha)^{-\kappa}}{h_\kappa \kappa!} \xi e^{-t\alpha})^2} \tag{113}$$

□

### D.2.1 Fractional sample complexity $\kappa$

If $\kappa \notin \mathbb{N}$, the asymptotic formulas simplify significantly. In this case, the kernel state equation lacks the exact "resonating" shell $m = \kappa$. Consequently, the threshold integral term $\psi^{(\kappa)}$ vanishes, and the self-consistency equation straightforwardly reduces to:

$$\psi = \frac{\lambda_{\text{eff}}}{\xi}$$

The same reasoning rigorously applies to the variance. As demonstrated previously, the only term yielding a macroscopic non-zero contribution $\tau^{(m)} = \Theta_d(1)$ occurs when $m = \kappa$. If $\kappa \notin \mathbb{N}$, then $\tau$ will necessarily go to 0. In this regime, the overall asymptotic decay of the variance as $d \to \infty$ is bounded by the slowest decaying adjacent integer shells, which is $\tau^{(\lfloor \kappa \rfloor)}$ or $\tau^{(\lceil \kappa \rceil)}$.

In particular, as analyzed for the sub-critical terms ($m < \kappa$):

$$\tau^{(\lfloor \kappa \rfloor)} = \Theta(d^{\lfloor \kappa \rfloor - \kappa})$$

For $\tau^{(\lceil \kappa \rceil)}$ we have to distinguish between $\alpha < \frac{1}{2}$ and $\alpha > \frac{1}{2}$. When $\alpha < \frac{1}{2}$, the scaling for $\kappa < m < \frac{\kappa}{1-\alpha}$ follows the same regime as $m > \frac{\kappa}{1-\alpha}$, yielding for $\tau^{(\lceil \kappa \rceil)} = \Theta(d^{\kappa - \lceil \kappa \rceil})$. When $\alpha > \frac{1}{2}$, the integral localizes and the variance term scales as $\Theta(d^{(\lceil \kappa \rceil - \kappa)(1-1/\alpha)})$ as we proved in Eq. 111.

Defining the variance decay exponent $\gamma_V > 0$ such that $\mathsf{V} = \Theta(d^{-\gamma_V})$, the effective exponent is given by the minimum absolute decay rate among the two bounding shells:

$$\gamma_V = \begin{cases} \min\left(\kappa - \lfloor \kappa \rfloor,\ \lceil \kappa \rceil - \kappa\right) & \text{if } 0 \le \alpha \le \frac{1}{2} \\ \min\left(\kappa - \lfloor \kappa \rfloor,\ \frac{1-\alpha}{\alpha}(\lceil \kappa \rceil - \kappa)\right) & \text{if } \frac{1}{2} < \alpha < 1 \end{cases} \tag{114}$$

To validate our theoretical analysis, we performed numerical experiments verifying that an empirically trained KRR on anisotropic data aligns with our asymptotic predictions in the high-dimensional limit. Fig. 6 displays the finite-size scaling of the variance learning curve as a function of $\psi = n/d^\kappa$, while Fig. 7 illustrates the convergence of the variance for fixed $\psi$ and $\kappa$ as the dimension $d$ increases, once again demonstrating agreement with our theory. To isolate the pure variance, we train our KRRs with a null target function $f_\star(\boldsymbol{x}) = 0$, meaning the labels are pure noise.

## D.3 Bias

The deterministic equivalent for the bias $\mathsf{B}$ in Eq. (7) can be recast as:

$$\mathsf{B} = \frac{\nu^2 \langle \boldsymbol{\theta}, (\boldsymbol{\Lambda} + \nu_\star \boldsymbol{I})^{-2} \boldsymbol{\theta} \rangle}{1 - \frac{1}{n} \operatorname{Tr}(\boldsymbol{\Lambda}^2 (\boldsymbol{\Lambda} + \nu_\star \boldsymbol{I})^{-2})} = \frac{U}{1 - \tau} \tag{115}$$

where $U \equiv \nu_\star^2 \langle \boldsymbol{\theta}, (\boldsymbol{\Lambda} + \nu_\star \boldsymbol{I})^{-2} \theta \rangle$ captures the explicit dependence on the target function and can be interpreted as the raw unlearned energy of the target function, and $\tau$ is the term previously evaluated

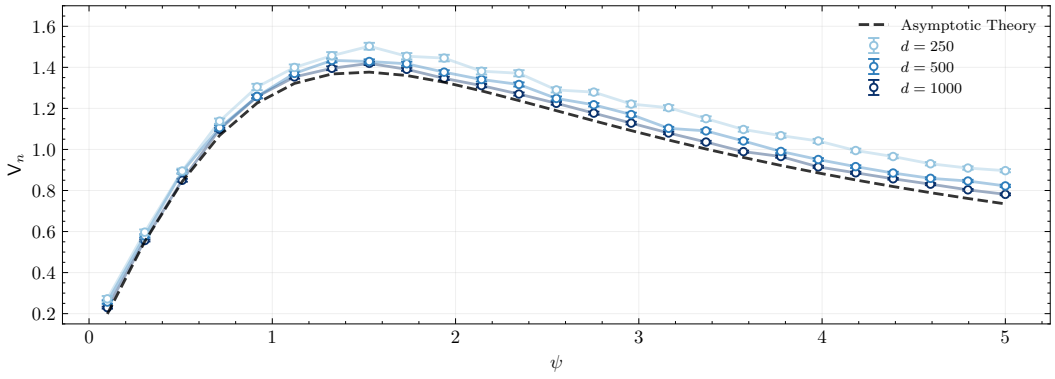


Figure 6: Finite-size scaling of the variance learning curve. The empirical variance $\mathsf{V}_n$ is plotted as a function of the load parameter $\psi = n/d^\kappa$ for varying system dimensions $d$ (blue gradients). As the dimension increases, the empirical realizations progressively collapse onto the asymptotic theoretical prediction (black dashed line) derived from our state equations in Theorem 1. Fixed parameters: $\kappa = 1$, $\alpha = 0.3$, and $\sigma_\varepsilon = 2$, $h(t) = 1 + 8t + 3t^2 + 0.1t^3 + t^4$, $\lambda = 0$.

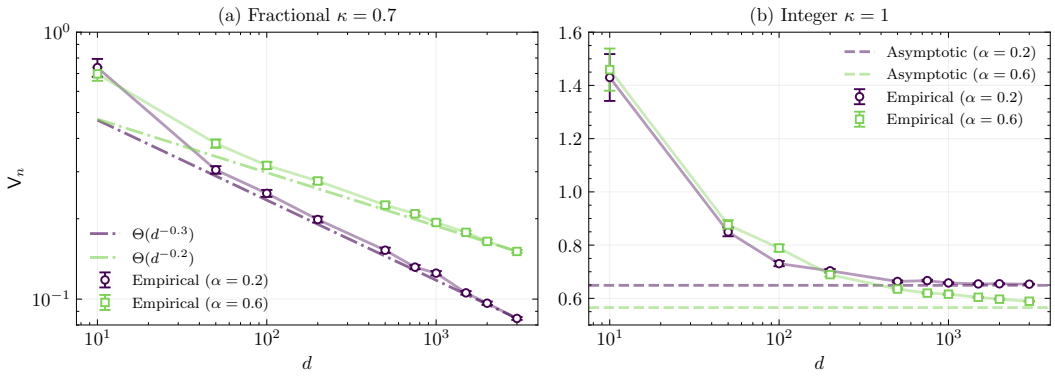


Figure 7: **(Left)** Convergence of the empirical variance as $d \to \infty$ for fractional $\kappa$. The variance decays to zero, perfectly following the asymptotic power law predicted by Eq. (12). **(Right)** Integer scaling: the empirical variance matches the asymptotic limit predicted by our theory (dotted lines). Error bars represent one standard deviation over 25 independent trials. Fixed parameters: $\sigma_\varepsilon = 1$, $\lambda = 0$, $\psi = 0.8$, with an exponential kernel truncated at the fourth Taylor order.

in Appendix D.2 which doesn't depend on the specific target function. The vector $\boldsymbol{\theta}$ contains the projection coefficients of the target function $f_\star$ into the kernel eigenbasis. Following the reparameterization introduced in Appendix C, we can expand this target contribution across the polynomial shells as

$$f_\star(\boldsymbol{x}) = \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{1\le i_1\le\cdots\le i_m\le d} \theta_{i_1,\ldots,i_m} e_{i_1,\ldots,i_m}(\boldsymbol{x})$$

And the numerator of Eq. (115) becomes:

$$\begin{aligned} U \equiv \nu_\star^2 \langle \boldsymbol{\theta}, (\boldsymbol{\Lambda} + \nu_\star \boldsymbol{I})^{-2} \boldsymbol{\theta} \rangle &= \sum_{m=0} \sum_{1\le i_1\le\cdots\le i_m\le d} \frac{\nu^2}{(\lambda_{\boldsymbol{\beta}} + \nu_\star)^2} (\theta_{i_1,\ldots,i_m})^2 \\ &= \sum_{m=0} \sum_{1\le i_1\le\cdots\le i_m\le d} \frac{1}{(1 + \lambda_{\boldsymbol{\beta}}/\nu_\star)^2} (\theta_{i_1,\ldots,i_m})^2 = \\ &= \sum_{m=0}^{\infty} U^{(m)} \end{aligned} \tag{116}$$

Where

$$U^{(m)} = \sum_{1\le i_1\le\cdots\le i_m\le d} \frac{1}{(1 + \lambda^{(m)}_{i_1,\ldots i_m}/\nu_\star)^2} (\theta_{i_1,\ldots,i_m})^2 \tag{117}$$

can be interpreted as the *unlearned fraction of the $m$-th shell* as $0 \le U^{(m)} \le E(m)$. In the limit $d \to \infty$, assuming a "dense" target function and extending $\theta_{i_1,\ldots,i_m}$ so that it is perfectly symmetric we have:

$$U^{(m)} = \frac{1}{m!} \sum_{i_1\ldots i_m=1}^{d} \frac{(\theta_{i_1,\ldots,i_m})^2}{(1 + \nu_\star^{-1} C_m^{-1} d^{m(\alpha-1)} (i_1 \ldots i_m)^{-\alpha})^2} \tag{118}$$

To proceed, we need to assume something about the specific shape of the tensor $\theta_{i_1,\ldots,i_m}$. In particular, we will assume:

$$\theta_{i_1,\ldots,i_m} = \theta_m (i_1 \ldots i_m)^{-\omega}$$

where $\omega \ge 0$ is the target anisotropy parameter governing an equivalent source condition for the function, and $\theta_m$ is a shell-dependent scalar coefficient.

The total spectral energy contained in the $m$-th shell, denoted as $E(m)$, is given by the sum of the squared weights:

$$E(m) = \frac{1}{m!} \sum_{i_1,\ldots,i_m=1}^{d} (\theta_{i_1,\ldots,i_m})^2 = \frac{\theta_m^2}{m!} \sum_{i_1,\ldots,i_m=1}^{d} (i_1 \ldots i_m)^{-2\omega} = \frac{\theta_m^2}{m!} \left( \sum_{i=1}^{d} i^{-2\omega} \right)^m \tag{119}$$

We want to ensure the target function is square-integrable (i.e., its total energy is finite) as $d \to \infty$. Hence, depending on $\omega$ we will have two different regimes:

**Slow decay target $\boldsymbol{\omega < \frac{1}{2}}$** In this regime, the unnormalized energy per shell grows to infinity:

$$E(m) \sim \frac{\theta_m^2}{m!} (1 - 2\omega)^{-m} d^{m(1-2\omega)}$$

so that $\theta_m^2$ needs to scale with $d$ in order to keep the full sum $\sum_m E(m) = 1$, i.e. $\theta_m^2 = \Theta(d^{m(2\omega-1)})$. We can asymptotically approximate the series in Eq. (118) with the corresponding Riemann integral:

$$U^{(m)} = \frac{\theta_m^2}{m!} \int_1^d \cdots \int_1^d \mathrm{d}x_1 \cdots \mathrm{d}x_m \frac{(x_1 \cdots x_m)^{-2\omega}}{(1 + \nu^{-1} C_m^{-1} d^{m(\alpha-1)} (x_1 \cdots x_m)^{-\alpha})^2} \,. \tag{120}$$

Upon the substitution:

$$y_i = x_i \nu^{\frac{1}{m\alpha}} C_m^{\frac{1}{m\alpha}} d^{\frac{1-\alpha}{\alpha}}$$

and using the scaling relation $\nu = \xi d^{-\kappa}$ we get:

$$U^{(m)} = \frac{\theta_m^2}{m!} C_m^{\frac{2\omega-1}{\alpha}} \xi^{\frac{2\omega-1}{\alpha}} d^{\frac{1}{\alpha}(\kappa-m+m\alpha)(1-2\omega)} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \mathrm{d}y_1 \cdots \mathrm{d}y_m \frac{(y_1 \cdots y_m)^{-2\omega}}{(1+(y_1 \cdots y_m)^{-\alpha})^2} = \\ = \frac{\theta_m^2}{m!} C_m^{\frac{2\omega-1}{\alpha}} \xi^{\frac{2\omega-1}{\alpha}} d^{\frac{1}{\alpha}(\kappa-m+m\alpha)(1-2\omega)} \mathcal{I}^{(m)}(\alpha,\omega)\,. \tag{121}$$

The asymptotic behavior of the upper and lower limits of the integral is the same as the one in Fig. 5. Again, we will divide the analysis according to the different values of $m$:

**Case a: $m < \kappa$** Here both $y_{\min}, y_{\max} \to 0$. The integral $\mathcal{I}^{(m)}(\alpha,\omega)$ can thus be computed by re-writing:

$$\mathcal{I}^{(m)}(\alpha,\omega) = \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \mathrm{d}y_1 \cdots \mathrm{d}y_m \frac{(y_1 \cdots y_m)^{2(\alpha-\omega)}}{(1+(y_1 \cdots y_m)^{\alpha})^2}$$

and asymptotically approximating:

$$\mathcal{I}^{(m)}(\alpha,\omega) \sim \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \mathrm{d}y_1 \cdots \mathrm{d}y_m (y_1 \cdots y_m)^{2(\alpha-\omega)} = \\ = \left( \int_{y_{\min}}^{y_{\max}} y^{2(\alpha-\omega)} dy \right)^m = \\ = (1+2(\alpha-\omega))^{-m} (y_{\max}^{1+2(\alpha-\omega)} - y_{\min}^{1+2(\alpha-\omega)})^m\,.$$

Now, the leading term in this last expression is determined by the sign of $1+2(\alpha-\omega)$. Since we are considering $\omega < \frac{1}{2}$, then $1+2(\alpha-\omega) > 0$, and thus the leading term is the one with $y_{\max}$:

$$(1+2(\alpha-\omega))^{-m} (y_{\max}^{1+2(\alpha-\omega)} - y_{\min}^{1+2(\alpha-\omega)})^m \sim (1+2(\alpha-\omega))^{-m} y_{\max}^{m+2m(\alpha-\omega)}$$

which scales as $d^{\frac{1+2(\alpha-\omega)}{\alpha}(m-\kappa)}$. The full term $U^{(m)}$ will then scale like:

$$d^{\frac{1+2(\alpha-\omega)}{\alpha}(m-\kappa)} d^{\frac{\kappa-m+m\alpha}{\alpha}(1-2\omega)} = d^{3m-2\kappa-2m\omega}\,.$$

At a first glance, the sign of this exponent might not be immediately obvious. However, if we rewrite it as:

$$3m - 2\kappa - 2m\omega = 2(m-\kappa) + m(1-2\omega)$$

and consider this expression together with $\theta_m^2$ (which, in the case $\omega < \frac{1}{2}$, must scale with $d$ as well):

$$U^{(m)} \asymp \theta_m^2 d^{2(m-\kappa)+m(1-2\omega)} = \theta_m^2 d^{m(1-2\omega)} d^{2(m-\kappa)} = E(m) d^{2(m-\kappa)}$$

we can see that $U^{(m)} \to 0$ when $m < \kappa$. These under-resonant shells are thus perfectly learnt.

**Case b: $m > \frac{\kappa}{1-\alpha}$** Now $y_{\min}, y_{\max} \to \infty$. The integral $\mathcal{I}^{(m)}(\alpha,\omega)$ in Eq. (121) becomes:

$$\mathcal{I}^{(m)}(\alpha,\omega) \sim \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \mathrm{d}y_1 \cdots \mathrm{d}y_m (y_1 \cdots y_m)^{-2\omega} = \\ = \left( \int_{y_{\min}}^{y_{\max}} y^{-2\omega} dy \right)^m = (1-2\omega)^{-m} (y_{\max}^{1-2\omega} - y_{\min}^{1-2\omega})^m\,.$$

Since $\omega < \frac{1}{2}$, then:

$$(1-2\omega)^{-m} (y_{\max}^{1-2\omega} - y_{\min}^{1-2\omega})^m \sim (1-2\omega)^{-m} y_{\max}^{m-2m\omega}$$

And, asymptotically:

$$\int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \mathrm{d}y_1 \cdots \mathrm{d}y_m \frac{(y_1 \cdots y_m)^{2(\alpha-\omega)}}{(1+(y_1 \cdots y_m)^{\alpha})^2} \sim (1-2\omega)^{-m} y_{\max}^{m-2m\omega} = \\ = (1-2\omega)^{-m} \xi^{\frac{1-2\omega}{\alpha}} C_m^{\frac{1-2\omega}{\alpha}} d^{\frac{1-2\omega}{\alpha}(m-\kappa)}\,.$$

Overall, the quantity $U^{(m)}$ becomes at a leading order:

$$U^{(m)} = \frac{\theta_m^2}{m!} C_m^{\frac{2\omega-1}{\alpha}} \xi^{\frac{2\omega-1}{\alpha}} d^{\frac{1}{\alpha}(\kappa-m+m\alpha)(1-2\omega)} (1-2\omega)^{-m} \xi^{\frac{1-2\omega}{\alpha}} C_m^{\frac{1-2\omega}{\alpha}} d^{\frac{1-2\omega}{\alpha}(m-\kappa)} =$$
$$= \frac{\theta_m^2}{m!} (1-2\omega)^{-m} d^{m(1-2\omega)} .$$

And since we are in the $\omega < \frac{1}{2}$ regime, then the latter can be simply recognized as the spectral energy, yielding:

$$U^{(m)} = E(m) .$$

**Case c: $\kappa < m \leq \frac{\kappa}{1-\alpha}$** In this intermediate regime, the integration boundaries behave as $y_{\min} \to 0$ and $y_{\max} \to \infty$. As established in the previous case $m > \frac{\kappa}{1-\alpha}$, the divergence of the integral is strictly driven by the large-$y$ behavior near the upper boundary $y_{\max}$ (UV-divergence). Consequently, the contribution from the lower limit ($y_{\min} \to 0$) is asymptotically sub-leading and can be safely neglected. Evaluating the integral at leading order thus yields the exact same dimensional scaling derived in the previous case. Therefore, even in this intermediate regime, the bias saturates to the full shell energy:

$$U^{(m)} = E(m) . \tag{122}$$

**Case d: $m = \kappa$** In this case, $y_{\min} \to 0$ and $y_{\max} = C_\kappa^{\frac{1}{\kappa\alpha}} \xi^{\frac{1}{\kappa\alpha}} = \Theta_d(1)$. We are not integrating over the pathological region $y \to \infty$ so the final integral should be of order 1. Indeed, we can transform the integral $\mathcal{I}^{(m)}(\alpha, \omega)$ in Eq. (121) using the Irwin-Hall technique in Lemma D.1 to get:

$$\mathcal{I}^{(m)}(\alpha,\omega) = \log(d)^\kappa \int_0^\kappa dx f_{IH}(x;\kappa) \frac{e^{(1-2\omega)(x\log(d)+m\log(y_{\min}))}}{(1+e^{-\alpha(x\log(d)+m\log(y_{\min}))})^2} =$$
$$= \log(d)^\kappa (\xi C_\kappa)^{(1-2\omega)/\alpha} \int_0^\kappa dx f_{IH}(x;\kappa) \frac{e^{(1-2\omega)(x-\kappa)\log(d)}}{(1+(C_\kappa\xi)^{-1}e^{-\alpha\log(d)(x-\kappa)})^2} . \tag{123}$$

where we used the fact that $\log(y_{\min}) = \log(y_{\max}) - \log(d) = \frac{1}{\kappa\alpha}\log(C_\kappa\xi) - \log(d)$. Under the substitution $(\kappa - x)\log(d) = t$, we obtain:

$$\mathcal{I}^{(m)}(\alpha,\omega) = \log(d)^{\kappa-1} (\xi C_\kappa)^{(1-2\omega)/\alpha} \int_0^{\kappa\log(d)} f_{IH}\left(\kappa - \frac{t}{\log(d)};\kappa\right) \frac{e^{-(1-2\omega)t}}{(1+(C_\kappa\xi)^{-1}e^{\alpha t})^2} dt$$

As $d \to \infty$, we can follow the same procedure as the one we performed in Sec. D.1 with Eq. (103). In particular, to use the DCT, it will suffice to have:

$$g_d(t) = \mathbf{1}_{t\in[0,\kappa\log(d)]} \log(d)^{\kappa-1} f_{IH}\left(\kappa - \frac{t}{\log(d)};\kappa\right) \frac{e^{-(1-2\omega)t}}{(1+(C_\kappa\xi)^{-1}e^{\alpha t})^2}$$

$$G(t) = t^{\kappa-1} \frac{e^{-(1-2\omega)t}}{(1+(C_\kappa\xi)^{-1}e^{\alpha t})^2}$$

and since:

$$g_d(t) \to \frac{t^{\kappa-1}}{(\kappa-1)!} \frac{e^{-(1-2\omega)t}}{(1+(C_\kappa\xi)^{-1}e^{\alpha t})^2}$$

we finally have:

$$\lim_{d\to\infty} \mathcal{I}^{(m)}(\alpha,\omega) = \frac{1}{(\kappa-1)!} (\xi C_\kappa)^{(1-2\omega)/\alpha} \int_0^\infty t^{\kappa-1} \frac{e^{-(1-2\omega)t}}{(1+(C_\kappa\xi)^{-1}e^{\alpha t})^2} dt .$$

Getting back to the formula for $U^{(\kappa)}$ from Eq. (121), we finally have:

$$U^{(\kappa)} = \frac{\theta_\kappa^2}{\kappa!} d^{\kappa(1-2\omega)} \frac{1}{(\kappa-1)!} \int_0^\infty e^{-t} t^{\kappa-1} \frac{e^{2\omega t}}{(1+(C_\kappa\xi)^{-1}e^{\alpha t})^2} dt .$$

Considering $E(\kappa) = \frac{\theta_\kappa^2}{\kappa!}(1-2\omega)^{-\kappa} d^{\kappa(1-2\omega)}$, then finally:

$$U^{(\kappa)} = E(\kappa) \frac{(1-2\omega)^\kappa}{(\kappa-1)!} \int_0^\infty e^{t(2\omega-1)} t^{\kappa-1} \frac{1}{(1+(C_\kappa\xi)^{-1}e^{\alpha t})^2} dt . \tag{124}$$

**Fast decay target ($\omega > 1/2$).** In this regime, the target coefficients decay rapidly enough to guarantee the convergence of the infinite series in the thermodynamic limit $d \to \infty$. As established in Eq. (119), the spectral energy evaluates to:

$$E(m) = \frac{\theta_m^2}{m!}\zeta(2\omega)^m \,, \tag{125}$$

implying that the prefactor $\theta_m$ scales as an intensive quantity, $\Theta_d(1)$.

Substituting the critical scaling ansatz $\nu = \xi d^{-\kappa}$ and the source condition into Eq. (118), we obtain:

$$U^{(m)} = \frac{\theta_m^2}{m!}\sum_{i_1,\dots,i_m=1}^{d} \frac{(i_1\cdots i_m)^{-2\omega}}{\left(1+\xi^{-1}C_m^{-1}d^{\kappa+m(\alpha-1)}(i_1\cdots i_m)^{-\alpha}\right)^2} \,. \tag{126}$$

The asymptotic limit of this expression is governed by the exponent of $d$, namely $\kappa+m(\alpha-1)$. Since $\omega > 1/2$, the bounding series $\sum_{i_1,\dots,i_m}(i_1\cdots i_m)^{-2\omega}$ intrinsically converges. Thus, we can invoke the Dominated Convergence Theorem (DCT) to exchange the limit $d \to \infty$ with the summation sign, using the bound:

$$\frac{(i_1\cdots i_m)^{-2\omega}}{\left(1+\xi^{-1}C_m^{-1}d^{\kappa+m(\alpha-1)}(i_1\cdots i_m)^{-\alpha}\right)^2} \le (i_1\cdots i_m)^{-2\omega} \,. \tag{127}$$

Assuming we are not exactly at a transition boundary (i.e., $\frac{\kappa}{1-\alpha} \neq m$ for any $m \in \mathbb{N}$), we can evaluate the limit term by term:

- **Unlearned features** $\left(m > \frac{\kappa}{1-\alpha}\right)$**:** The exponent of $d$ is strictly negative, causing the denominator to approach 1. The series thus converges to the full spectral energy:

$$U^{(m)} \xrightarrow{d\to\infty} \frac{\theta_m^2}{m!}\sum_{i_1,\dots,i_m=1}^{\infty}(i_1\cdots i_m)^{-2\omega} = \frac{\theta_m^2}{m!}\zeta(2\omega)^m = E(m) \,. \tag{128}$$

- **Learned features** $\left(m < \frac{\kappa}{1-\alpha}\right)$**:** The exponent of $d$ is strictly positive. Consequently, the denominator diverges, yielding $U^{(m)} \xrightarrow{d\to\infty} 0$.

Next, we take care of the case $\kappa_{\text{eff}} = \frac{\kappa}{1-\alpha} \in \mathbb{N}$. This is formally equivalent to the case $m = \kappa$ analyzed in the slow-decay regime $\omega < \frac{1}{2}$. Indeed, from Eq. (121), we have:

$$U^{(\kappa_{\text{eff}})} = \frac{\theta_{\kappa_{\text{eff}}}^2}{\kappa_{\text{eff}}!}C_{\kappa_{\text{eff}}}^{\frac{2\omega-1}{\alpha}}\xi^{\frac{2\omega-1}{\alpha}}\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) \,.$$

with

$$\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \int_{y_{\min}}^{y_{\max}}\cdots\int_{y_{\min}}^{y_{\max}} \mathrm{d}y_1\cdots\mathrm{d}y_{\kappa_{\text{eff}}}\frac{(y_1\cdots y_{\kappa_{\text{eff}}})^{-2\omega}}{(1+(y_1\cdots y_{\kappa_{\text{eff}}})^{-\alpha})^2}$$

We use again the Irwin-Hall technique in Lemma D.1 to manipulate the integral and get:

$$\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \log(d)^{\kappa_{\text{eff}}}\int_0^{\kappa_{\text{eff}}} dx\, f_{IH}(x;\kappa_{\text{eff}})\frac{e^{(1-2\omega)(x\log(d)+\kappa_{\text{eff}}\log(y_{\min}))}}{(1+e^{-\alpha(x\log(d)+\kappa_{\text{eff}}\log(y_{\min}))})^2} \,. \tag{129}$$

When $m = \kappa_{\text{eff}} = \frac{\kappa}{1-\alpha}$, then $y_{\min} = (C_{\kappa_{\text{eff}}}\xi)^{\frac{1}{\kappa_{\text{eff}}\alpha}} = \Theta_d(1)$ (whereas in the $m = \kappa$ previous case $y_{\min}$ was vanishing and $y_{\max}$ was finite), so that:

$$\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \log(d)^{\kappa_{\text{eff}}}(C_{\kappa_{\text{eff}}}\xi)^{\frac{1-2\omega}{\alpha}}\int_0^{\kappa_{\text{eff}}} dx\, f_{IH}(x;\kappa_{\text{eff}})\frac{e^{(1-2\omega)(x\log(d))}}{(1+C_{\kappa_{\text{eff}}}^{-1}\xi^{-1}e^{-\alpha x\log(d)})^2} \,. \tag{130}$$

Under the substitution $x\log(d) = t$, we obtain:

$$\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \log(d)^{\kappa_{\text{eff}}-1}(\xi C_{\kappa_{\text{eff}}})^{(1-2\omega)/\alpha}\int_0^{\kappa_{\text{eff}}\log(d)} f_{IH}\left(\frac{t}{\log(d)};\kappa_{\text{eff}}\right)\frac{e^{(1-2\omega)t}}{(1+(C_{\kappa_{\text{eff}}}\xi)^{-1}e^{-\alpha t})^2}dt$$

As $d \to \infty$, we can perform the usual procedure and use the Dominated Convergence Theorem. We will thus omit the full derivation and provide the final results:

$$\lim_{d\to\infty} \mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \frac{1}{(\kappa_{\text{eff}}-1)!}(\xi C_{\kappa_{\text{eff}}})^{(1-2\omega)/\alpha} \int_0^\infty t^{\kappa_{\text{eff}}-1} \frac{e^{(1-2\omega)t}}{(1+(C_{\kappa_{\text{eff}}}\xi)^{-1}e^{-\alpha t})^2} dt\,.$$

Getting back to the formula for $U^{(\kappa_{\text{eff}})}$ from Eq. (121), we have:

$$U^{(\kappa_{\text{eff}})} = \frac{\theta^2_{\kappa_{\text{eff}}}}{\kappa_{\text{eff}}!\,(\kappa_{\text{eff}}-1)!} \int_0^\infty t^{\kappa_{\text{eff}}-1} \frac{e^{(1-2\omega)t}}{(1+(C_{\kappa_{\text{eff}}}\xi)^{-1}e^{-\alpha t})^2} dt\,.$$

To interpret this result in terms of the initial spectral energy and draw a direct parallel with the $\omega < 1/2$ regime, we must note that the transition to the continuous integral formulation inherently replaces the discrete series with a continuous measure. Under this same continuum approximation ($\sum \to \int$), the baseline spectral energy is evaluated as $E_{\text{cont}}(m) = \frac{\theta^2_m}{m!}\left(\int_1^\infty x^{-2\omega}dx\right)^m = \frac{\theta^2_m}{m!}(2\omega-1)^{-m}$, which acts as the continuous counterpart to the exact discrete energy $E(m) \propto \zeta(2\omega)^m$.

Substituting the discrete energy with the continuous energy baseline approximation, we finally obtain an elegant closed-form expression:

$$U^{(\kappa_{\text{eff}})} \simeq E_{(\kappa_{\text{eff}})} \frac{(2\omega-1)^{\kappa_{\text{eff}}}}{(\kappa_{\text{eff}}-1)!} \int_0^\infty t^{\kappa_{\text{eff}}-1} \frac{e^{(1-2\omega)t}}{(1+(C_{\kappa_{\text{eff}}}\xi)^{-1}e^{-\alpha t})^2} dt\,. \tag{131}$$

□

# E STRONGLY ANISOTROPIC REGIME

In this section we aim at proving the formulas valid in the $\alpha > 1$ regime. In the strongly anisotropic regime, we cannot express the kernel eigenvalues as we did in Proposition 1, and we must instead rely on the bounds provided by Wortsman & Loureiro (2025):

$$c_{1,\boldsymbol{\beta}} h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \sigma_1^{\beta_1} \cdots \sigma_d^{\beta_d} \le \lambda_{\boldsymbol{\beta}} \le c_{2,\boldsymbol{\beta}} h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \sigma_1^{\beta_1} \cdots \sigma_d^{\beta_d}\,. \tag{132}$$

First, let us define the following quantities:

$$c_-^{(m)} = \inf_{|\boldsymbol{\beta}|=m} c_{1,\boldsymbol{\beta}}$$

$$c_+^{(m)} = \sup_{|\boldsymbol{\beta}|=m} c_{2,\boldsymbol{\beta}}$$

Thus, for every shell $m$, we can write:

$$c_-^{(m)} h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \sigma_1^{\beta_1} \cdots \sigma_d^{\beta_d} \le \lambda_{\boldsymbol{\beta}} \le c_+^{(m)} h_{|\boldsymbol{\beta}|} |\boldsymbol{\beta}|! \sigma_1^{\beta_1} \cdots \sigma_d^{\beta_d} \quad \forall \boldsymbol{\beta} : |\boldsymbol{\beta}| = m\,.$$

or, using the reparametrization in Appendix C and using $\sigma_j = r_\alpha(d)^{-1} j^{-\alpha}$:

$$c_-^{(m)} h_m m! r_\alpha(d)^{-m} (i_1 \cdots i_m)^{-\alpha} \le \lambda^{(m)}_{i_1,\dots,i_m} \le c_+^{(m)} h_m m! r_\alpha(d)^{-m} (i_1 \cdots i_m)^{-\alpha}\,.$$

## E.1 KERNEL STATE EQUATION AND VARIANCE

Following the same procedure adopted in Appendix D.1, we can write the self-consistency equation Eq. (8) as $\psi = \sum_m \psi^{(m)} + \frac{\lambda}{\nu} d^{-\kappa}$ where:

$$\psi^{(m)} = \sum_{\boldsymbol{\beta}\in\mathbb{Z}^d_{\ge 0}:|\boldsymbol{\beta}|=m} \frac{d^{-\kappa}}{1+\frac{\nu}{\lambda_{\boldsymbol{\beta}}}} = \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{d^{-\kappa}}{1+\frac{\nu}{\lambda^{(m)}_{i_1,\dots,i_m}}} \tag{133}$$

It is straightforward to see that, by using the bounds in Eq. (132) for $\lambda^{(m)}_{i_1,\dots,i_m}$, we can write:

$$\psi_-^{(m)} \le \psi^{(m)} \le \psi_+^{(m)}$$

where:

$$\psi_{\pm}^{(m)} = \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{d^{-\kappa}}{1 + (c_{\pm}^{(m)})^{-1} \nu (i_1 \dots i_m)^{\alpha} (h_m m!)^{-1} r_{\alpha}(d)^m} \,. \tag{134}$$

In what follows, we will mainly focus on the behavior of $\psi_{\pm}^{(m)}$.

When $\alpha > 1$, we know that $r_\alpha(d) = \zeta(\alpha)$ at leading order and, labeling $C_{m,\pm} = (c_{\pm}^{(m)})^{-1}(h_m m!)^{-1}\zeta(\alpha)^m$, we get:

$$\psi_{\pm}^{(m)} = \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{d^{-\kappa}}{1 + \nu (i_1 \dots i_m)^{\alpha} C_{m,\pm}} \simeq \frac{1}{m!} \int_1^d \cdots \int_1^d \frac{d^{-\kappa}}{1 + \nu (x_1 \dots x_m)^{\alpha} C_{m,\pm}} dx_1 \dots dx_m$$

using the integral asymptotic approximation. Changing variables $x_i = y_i (C_{m,\pm}\nu)^{-\frac{1}{m\alpha}}$, we finally have:

$$\psi_{\pm}^{(m)} = \frac{1}{m!} C_{m,\pm}^{-\frac{1}{\alpha}} d^{-\kappa} \nu^{-\frac{1}{\alpha}} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} dy_1 \dots dy_m \frac{1}{1 + (y_1 \dots y_m)^{\alpha}} \,. \tag{135}$$

We analyze the ridgeless $\lambda = 0$ case and $\lambda \asymp 1$ separately. But before, we claim the following lemma that will come in handy later:

**Lemma E.1** (*Asymptotic Expansion of the Shell Integral*)**.** Let $\alpha > 1$ and consider the following $m$-dimensional integral:

$$\mathcal{I}^{(m)} = \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{\mathrm{d}y_1 \cdots \mathrm{d}y_m}{1 + (y_1 \cdots y_m)^{\alpha}} \,, \tag{136}$$

where the integration bounds are strictly positive and obey the following scaling behavior in the high-dimensional limit ($d \to \infty$):

$$y_{\min} = \Theta\Big(d^{-\frac{\kappa}{m}} \hat{f}(d)\Big) \to 0 \,, \tag{137a}$$

$$y_{\max} = d\, y_{\min} = \Theta\Big(d^{1-\frac{\kappa}{m}} \hat{f}(d)\Big) \to \infty \,. \tag{137b}$$

Here, $m \in \mathbb{N}$, $\kappa \in \mathbb{R}, 0 < \kappa \leq m$, and $\hat{f}(d)$ is a strictly sub-polynomial function (i.e. $f(d) = o(d^{\epsilon})$ for any $\epsilon > 0$).

Then, as $d \to \infty$, the leading-order asymptotic behavior of the integral resolves into two distinct regimes:

- **Super-critical shells** ($m > \kappa$): The upper bound $y_{\max}$ strictly diverges as a power of $d$. The integral exhibits a poly-logarithmic divergence governed by the Irwin-Hall distribution:

$$\mathcal{I}^{(m)} \simeq \log(d)^{m-1} f_{IH}(\kappa; m) \int_{-\infty}^{\infty} dz \frac{e^z}{1 + e^{\alpha z}} \tag{138}$$

  where $f_{IH}(\kappa; m)$ is the probability density function of the Irwin-Hall distribution evaluated at the critical threshold $\kappa$, explicitly given by:

$$f_{IH}(\kappa; m) = \frac{1}{(m-1)!} \sum_{j=0}^{\lfloor \kappa \rfloor} (-1)^j \binom{m}{j} (\kappa - j)^{m-1} \tag{139}$$

- **Resonant shell** ($m = \kappa$): The integral diverges sub-logarithmically as:

$$\mathcal{I}^{(\kappa)} \simeq \big(\log \hat{f}(d)\big)^{\kappa - 1} \frac{\kappa^{\kappa - 1}}{(\kappa - 1)!} \int_{-\infty}^{\infty} dz \frac{e^z}{1 + e^{\alpha z}} \,. \tag{140}$$

*Proof.* To manipulate the integral $\mathcal{I}^{(m)}$, we will resort again to the Irwin-Hall trick in Lemma D.1. We will then obtain a one-dimensional integral:

$$\mathcal{I}^{(m)} = \log(d)^m \int_0^m \mathrm{d}x \frac{e^{x \log(d) + m \log(y_{\min})}}{1 + e^{x\alpha \log(d) + m\alpha \log(y_{\min})}} f_{IH}(x; m) \,.$$

Since $y_{\min} = \Theta(d^{-\kappa/m}\hat{f}(d))$, then $e^{m\log(y_{\min})} = e^{\Delta-\kappa\log(d)}$ where $\Delta = \Theta(m\log\hat{f}(d)) = o(\log(d))$ is a quantity that scales sub-logarithmically with $d$. At the end, we have:

$$\mathcal{I}^{(m)} = \log(d)^m \int_0^m dx \frac{e^{\Delta} e^{(x-\kappa)\log(d)}}{1 + e^{\alpha\Delta} e^{\alpha\log(d)(x-\kappa)}} f_{IH}(x; m)$$

Let us change variables $t = (x-\kappa)\log(d) + \Delta$:

$$\mathcal{I}^{(m)} = \log(d)^{m-1} \int_{-\kappa\log(d)+\Delta}^{(m-\kappa)\log(d)+\Delta} dt \frac{e^t}{1+e^{\alpha t}} f_{IH}\left(\kappa + \frac{t-\Delta}{\log(d)}; m\right) \tag{141}$$

We now analyze the behavior of this expression as $d \to \infty$ by leveraging the Dominated Convergence Theorem (DCT). Specifically, let us rewrite the integral using the indicator function $\mathbf{1}_{[\cdot]}(t)$:

$$\begin{aligned}\mathcal{I}^{(m)} &= \log(d)^{m-1} \int_{\mathbb{R}} dt\, \mathbf{1}_{[-\kappa\log(d)+\Delta,(m-\kappa)\log(d)+\Delta]}(t) \frac{e^t}{1+e^{\alpha t}} f_{IH}\left(\kappa + \frac{t-\Delta}{\log(d)}; m\right) \\ &= \log(d)^{m-1} \int_{\mathbb{R}} g_d(t)\, dt\,,\end{aligned} \tag{142}$$

To apply the theorem, we first compute the pointwise convergence of $g_d(t)$ for any fixed $t \in \mathbb{R}$:

$$\lim_{d\to\infty} g_d(t) = \frac{e^t}{1+e^{\alpha t}} f_{IH}(\kappa; m)\,, \tag{143}$$

as the upper and lower limit of the indicator function go to $\pm\infty$. This is true since $\Delta$ is sub-logarithmic, hence the leading order behavior is given by $-\kappa\log(d) \to \infty$ and $(m-\kappa)\log(d) \to \infty$. Next, we construct an integrable bounding function $g(t)$ independent of $d$. Since the Irwin-Hall distribution $f_{IH}(z; m)$ is a continuous probability density function defined on a compact support ($[0, m]$), it is surely globally bounded by a finite constant $M_m$. Consequently, we can bound $g_d(t)$ as:

$$g_d(t) \le M_m \frac{e^t}{1+e^{\alpha t}} \equiv g(t)\,. \tag{144}$$

The bounding function $g(t)$ is strictly integrable over $\mathbb{R}$ for $\alpha > 1$. Hence, by the Dominated Convergence Theorem, we can safely take the limit $d \to \infty$ inside the integral:

$$\mathcal{I}^{(m)} = f_{IH}(\kappa; m)\log(d)^{m-1} \int_{-\infty}^{\infty} dt \frac{e^t}{1+e^{\alpha t}}\,. \tag{145}$$

The remaining integral is easy to compute and, overall, the integral simplifies to:

$$\mathcal{I}^{(m)} = f_{IH}(\kappa; m)\log(d)^{m-1} \frac{\pi}{\alpha\sin(\pi/\alpha)} \tag{146}$$

Now let us consider the case $m = \kappa$. In this scenario, Eq. (141) becomes:

$$\mathcal{I}^{(\kappa)} = \log(d)^{\kappa-1} \int_{-\kappa\log(d)+\Delta}^{\Delta} dt \frac{e^t}{1+e^{\alpha t}} f_{IH}\left(\kappa - \frac{\Delta - t}{\log(d)}; \kappa\right)\,. \tag{147}$$

We can apply a reasoning similar to the previous case, but we must be more careful. If we were to apply the Dominated Convergence Theorem (DCT) directly, the integral alone would trivially collapse to 0. While mathematically correct, this is not very instructive, as the diverging prefactor $\log(d)^{\kappa-1}$ would lead to an indeterminate form of the type $\infty \times 0$. To extract the correct scaling behavior of the integral as $d \to \infty$, let us recast the expression by factoring out $\Delta^{\kappa-1}$:

$$\begin{aligned}\mathcal{I}^{(\kappa)} &= \Delta^{\kappa-1} \int_{\mathbb{R}} dt\, \mathbf{1}_{[-\kappa\log(d)+\Delta,\Delta]}(t) \left[\log(d)^{\kappa-1}\Delta^{1-\kappa}\right] \frac{e^t}{1+e^{\alpha t}} f_{IH}\left(\kappa - \frac{\Delta-t}{\log(d)}; \kappa\right) \\ &\equiv \Delta^{\kappa-1} \int_{\mathbb{R}} g_d(t)\, dt\,.\end{aligned} \tag{148}$$

Now we have to evaluate the pointwise limit of $g_d(t)$. Setting $\epsilon = \frac{\Delta-t}{\log(d)}$, when $\epsilon \to 0$ we can expand the Irwin-Hall function near its right edge:

$$f_{IH}(\kappa - \epsilon; \kappa) = \frac{\epsilon^{\kappa-1}}{(\kappa-1)!} = \frac{(\Delta-t)^{\kappa-1}}{\log(d)^{\kappa-1}(\kappa-1)!}\,. \tag{149}$$

Substituting this expansion into $g_d(t)$, the diverging logarithmic terms exactly cancel out:

$$\begin{aligned}\lim_{d\to\infty} g_d(t) &= \lim_{d\to\infty} \Delta^{1-\kappa}\frac{e^t}{1+e^{\alpha t}}(\Delta-t)^{\kappa-1}\frac{1}{(\kappa-1)!} \\ &= \lim_{d\to\infty}\left(1-\frac{t}{\Delta}\right)^{\kappa-1}\frac{1}{(\kappa-1)!}\frac{e^t}{1+e^{\alpha t}} = \frac{1}{(\kappa-1)!}\frac{e^t}{1+e^{\alpha t}}\,.\end{aligned} \tag{150}$$

To rigorously apply the DCT, we can bound the sequence using the global upper limit for the Irwin-Hall distribution $f_{IH}(\kappa - x;\kappa) \le x^{\kappa-1}$ (valid for all $x \ge 0$). Substituting this into the absolute value of $g_d(t)$ yields:

$$|g_d(t)| \le \left[\log(d)^{\kappa-1}\Delta^{1-\kappa}\right]\frac{e^t}{1+e^{\alpha t}}\left(\frac{\Delta-t}{\log(d)}\right)^{\kappa-1} = \left(1-\frac{t}{\Delta}\right)^{\kappa-1}\frac{e^t}{1+e^{\alpha t}}\,. \tag{151}$$

Since $t \le \Delta$ on the integration domain, the term $(1 - t/\Delta)$ is less than or equal to 1 for $t \ge 0$, and bounded by $1 + |t|$ for $t < 0$. Therefore, we can uniformly bound $|g_d(t)|$ by a function $G(t)$ independent of $d$:

$$|g_d(t)| \le (1+|t|)^{\kappa-1}\frac{e^t}{1+e^{\alpha t}} \equiv G(t)\,. \tag{152}$$

Because $\alpha > 1$, $G(t)$ is absolutely integrable over $\mathbb{R}$. By the Dominated Convergence Theorem, we finally obtain the leading-order behavior:

$$\mathcal{I}^{(\kappa)} \simeq \Delta^{\kappa-1}\frac{1}{(\kappa-1)!}\int_{\mathbb{R}}\frac{e^t}{1+e^{\alpha t}}dt = \Delta^{\kappa-1}\frac{1}{(\kappa-1)!}\frac{\pi}{\alpha\sin(\pi/\alpha)}\,. \tag{153}$$

Recalling that $\Delta \simeq \kappa\log(\hat{f}(d))$ at leading order, we get:

$$\mathcal{I}^{(\kappa)} \simeq \frac{\kappa^{\kappa-1}}{(\kappa-1)!}\big(\log\hat{f}(d)\big)^{\kappa-1}\frac{\pi}{\alpha\sin(\pi/\alpha)}\,. \tag{154}$$

This demonstrates that, if $\hat{f}(d) \to \infty$, the integral in the case $m = \kappa$ diverges much slower ($\big(\log\hat{f}(d)\big)^{\kappa-1}$ instead of $\log(d)^{\kappa-1}$). □

### E.1.1 Ridgeless variance

In this case, the self-consistency equation reduces to $\psi = \sum_m \psi^{(m)}$. We formulate the following ansatz for $\nu$:

$$\nu = \xi f(d) d^{-\kappa\alpha} \tag{155}$$

where $\xi \asymp 1$ with respect to $d$ and $f(d)$ is a super-poly-logarithmic function in $d$, meaning that:

$$\lim_{d\to\infty}\frac{f(d)}{d^\eta} = 0 \quad \forall\eta > 0 \quad \text{and} \quad \lim_{d\to\infty}\frac{f(d)}{\log(d)^\pi} = \infty \quad \forall\pi > 0$$

The reasons for this specific scaling will become clear later on. The upper and lower limit in Eq. (135) become:

$$y_{\min} = (\xi C_{m,\pm})^{1/m\alpha} d^{-\kappa/m} f(d)^{1/m\alpha}, \qquad y_{\max} = (\xi C_{m,\pm})^{1/m\alpha} d^{1-\kappa/m} f(d)^{1/m\alpha} \tag{156}$$

**$m < \kappa$** We can use similar bounding arguments as the ones used in Appendix D.1 to conclude that the terms $\psi_\pm^{(m)}$ are negligible in the full summation (more precisely, $\psi_\pm^{(m)} = \Theta(d^{m-\kappa})$). Since $\psi_-^{(m)} \le \psi^{(m)} \le \psi_+^{(m)}$, then also $\psi^{(m)} = \Theta(d^{m-\kappa}) \to 0$.

**$m \ge \kappa$** Let us consider the integral in Eq. (135). We can use the Lemma E.1 setting $\hat{f}(d) = f(d)$ (scaling sub-polynomially as required). Overall, Eq. (135) simplifies to:

$$\psi_\pm^{(m)} \sim \frac{1}{m!}C_{m,\pm}^{-1/\alpha}\xi^{-1/\alpha}f(d)^{-1/\alpha}f_{IH}(\kappa;m)\log(d)^{m-1}\int_{-\infty}^{\infty}dz\frac{e^z}{1+e^{\alpha z}} \tag{157}$$

or, when $m = \kappa$:

$$\psi_\pm^{(m)} \sim \frac{1}{m!}C_{m,\pm}^{-1/\alpha}\xi^{-1/\alpha}f(d)^{-1/\alpha}\frac{\kappa^{\kappa-1}}{(\kappa-1)!}\big(\log f(d)\big)^{\kappa-1}\int_{-\infty}^{\infty}dz\frac{e^z}{1+e^{\alpha z}} \tag{158}$$

**Remark 6.** Now it should be clearer why we introduced the scaling $f(d)$. Without it, the generic $\psi_-^{(m)}$ would diverge logarithmically, which would not make sense as $\psi_-^{(m)} \leq \psi^{(m)} < \psi \asymp 1$. To extract the specific shape of $f(d)$, one should solve $\sum_m \psi_\pm^{(m)} \asymp 1$ and $f(d)$ should balance the infinite summation of the logarithmic powers. However, this is not strictly necessary, as we prove now.

Now consider the formula for the variance $\mathsf{V} = \sigma_\varepsilon^2 \frac{\tau}{1-\tau}$ with $\tau$ given by:

$$\tau = \sum_{m=0}^{\infty} \tau^{(m)} = \sum_{m=0}^{\infty} \frac{d^{-\kappa}}{\psi m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{1}{\left(1 + \frac{\nu}{\lambda^{(m)}_{i_1,\dots,i_m}}\right)^2} .$$

Just as we did above, using the bounding equations Eq. (132), we obtain:

$$\tau_-^{(m)} \leq \tau^{(m)} \leq \tau_+^{(m)} \tag{159}$$

where we define:

$$\begin{aligned} \tau_\pm^{(m)} &= \frac{d^{-\kappa}}{\psi m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{1}{(1 + C_{m,\pm}(i_1 \dots i_m)^\alpha \nu)^2} \\ &\sim \frac{d^{-\kappa}}{\psi m!} \int_1^d \cdots \int_1^d dx_1 \dots dx_m \frac{1}{(1 + C_{m,\pm}(x_1 \dots x_m)^\alpha \nu)^2} . \end{aligned}$$

By manipulating this integral as detailed above, we eventually arrive at an integral analogous to the one in Lemma E.1, but featuring a squared denominator. It is straightforward to show that the same asymptotic characterization holds for this modified integral, requiring only the substitution:

$$\int_{-\infty}^{\infty} \frac{e^z}{1 + e^{\alpha z}} \, \mathrm{d}z \quad \to \quad \int_{-\infty}^{\infty} \frac{e^z}{(1 + e^{\alpha z})^2} \, \mathrm{d}z .$$

Consequently, we obtain:

$$\tau_\pm^{(m)} \sim \frac{1}{\psi m!} C_{m,\pm}^{-1/\alpha} \xi^{-1/\alpha} f(d)^{-1/\alpha} f_{IH}(\kappa; m) \log(d)^{m-1} \int_{-\infty}^{\infty} dz \frac{e^z}{(1 + e^{\alpha z})^2} \tag{160}$$

But since:

$$\int_{-\infty}^{\infty} dz \frac{e^z}{(1 + e^{\alpha z})^2} = \left(1 - \frac{1}{\alpha}\right) \int_{-\infty}^{\infty} dz \frac{e^z}{1 + e^{\alpha z}}$$

Then, comparing Eq. (160) with Eq. (157):

$$\tau_\pm^{(m)} \sim \left(1 - \frac{1}{\alpha}\right) \frac{\psi_\pm^{(m)}}{\psi}$$

Therefore, summing over $m$ in Eq. (159), we obtain:

$$\left(1 - \frac{1}{\alpha}\right) \frac{1}{\psi} \sum_{m=0}^{\infty} \psi_-^{(m)} \leq \tau \leq \left(1 - \frac{1}{\alpha}\right) \frac{1}{\psi} \sum_{m=0}^{\infty} \psi_+^{(m)} .$$

By definition, $\psi$ satisfies $\sum_m \psi_-^{(m)} \leq \psi \leq \sum_m \psi_+^{(m)}$. Thus, the macroscopic ratios bounding $\tau$ obey:

$$0 < \frac{1}{\psi} \sum_{m=0}^{\infty} \psi_-^{(m)} \leq 1 \leq \frac{1}{\psi} \sum_{m=0}^{\infty} \psi_+^{(m)} .$$

In other words, we have:

$$k_1 \left(1 - \frac{1}{\alpha}\right) \leq \tau \leq \left(1 - \frac{1}{\alpha}\right) k_2 .$$

with $k_{1,2}$ finite constant such that $k_1 < 1, k_2 > 1$. This is equivalent to the asymptotic equivalence:

$$\tau \asymp 1 - \frac{1}{\alpha} .$$

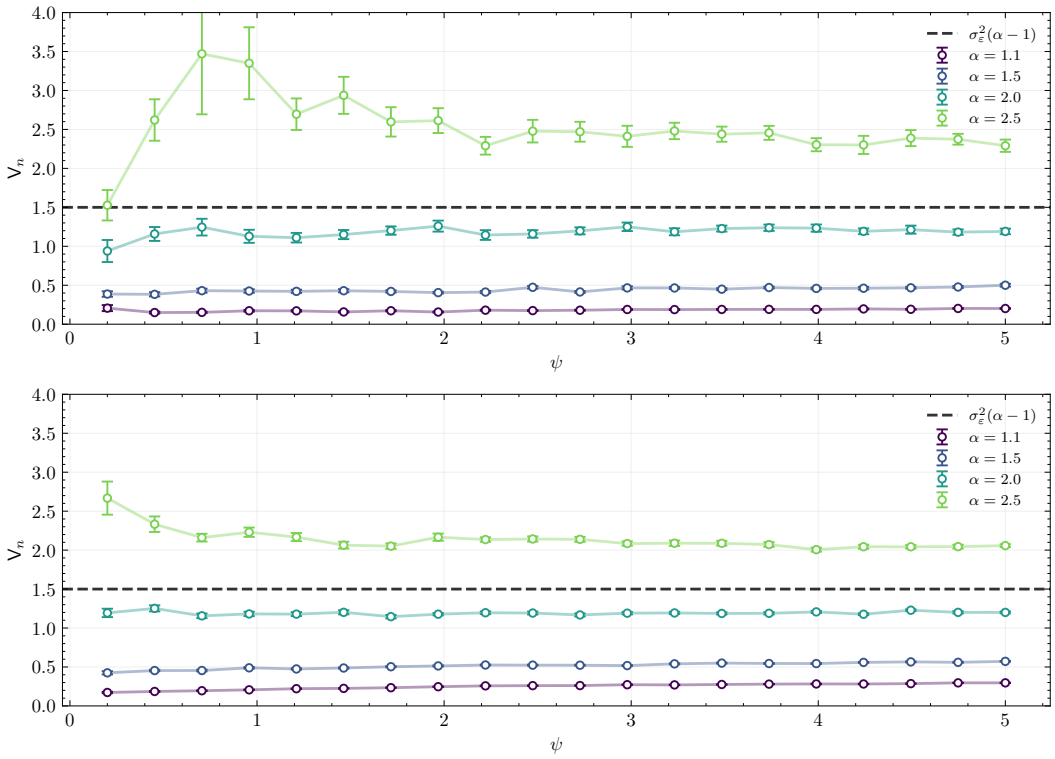


Figure 8: Empirical learning curves for the variance $\mathsf{V}_n$ for $\kappa = 0.5$ (top) and $\kappa = 0.8$ (bottom), across different values of $\alpha$. Results are averaged over 50 independent runs. Experimental setup: $\sigma_\varepsilon^2 = 1$, $\lambda = 0$, $d = 1000$. The models were trained setting the target function to $f_\star = 0$ in order to perfectly isolate the variance component.

**Final formula for the variance** Considering the variance $\mathsf{V} = \sigma_\varepsilon^2 \frac{\tau}{1-\tau}$, and since $\tau \asymp 1 - 1/\alpha$, we finally obtain:

$$k_1' \sigma_\varepsilon^2 (\alpha - 1) \leq \mathsf{V}_n \leq k_2' \sigma_\varepsilon^2 (\alpha - 1)$$

for $d$-independent constant. Equivalently, we have:

$$\mathsf{V}_n = C \sigma_\varepsilon^2 (\alpha - 1)\,,$$

where $C = \Theta_d(1)$ is a strictly positive, order-one factor.

**Remark 7.** Note that for $\alpha \to 1^+$, the variance vanishes identically, matching the $\alpha \to 1^-$ limit derived for the weakly anisotropic regime in Appendix D.2.

Once again, we benchmark our asymptotic findings against empirical KRR simulations in Fig. 8 for two values of $\kappa$, showing that the variance learning curve effectively reaches a plateau, becoming entirely independent of $\psi$ and $\kappa$ and on the order of $\sigma_\varepsilon^2(\alpha - 1)$

### E.1.2 Regularized case ($\lambda \asymp 1$)

In the presence of explicit ridge regularization, the self-consistency equation reads:

$$\psi = \sum_m \psi^{(m)} + \frac{\lambda}{\nu} d^{-\kappa}$$

Assuming $\lambda \asymp 1$ we must modify our scaling ansatz. We postulate:

$$\nu = \xi d^{-\kappa} \tag{161}$$

Under this scaling, the generic term $\psi_\pm^{(m)}$ evaluated through the integral representation in Eq. (135) will scale as:

$$\psi_\pm^{(m)} \sim \frac{1}{m!} C_{m,\pm}^{-\frac{1}{\alpha}} d^{-\kappa} \nu^{-\frac{1}{\alpha}} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{dy_1 \dots dy_m}{1 + (y_1 \dots y_m)^\alpha} \asymp d^{-\kappa\left(1-\frac{1}{\alpha}\right)} \log(d)^{m-1}$$

Since $\alpha > 1$, the exponent $(1 - 1/\alpha)$ is strictly positive. Consequently, $\psi_{\pm}^{(m)}$ gets power-law suppressed by $d^{-\kappa(1-1/\alpha)}$. This implies that also $\psi^{(m)}$ will scale as $\log(d)^{m-1} d^{-\kappa(1-1/\alpha)}$. Even when summing over the macroscopic shells $m$, the accumulated logarithmic factors $\log(d)^{m-1}$ yield at most a sub-polynomial growth $f(d)$. Hence, the entire kernel contribution vanishes in the thermodynamic limit ($\sum_m \psi^{(m)} = \tilde{\Theta}(d^{-\kappa(1-1/\alpha)}) \to 0$). The state equation thus trivially reduces to the regularization term:

$$\psi = \frac{\lambda}{\xi} \tag{162}$$

This implies $\xi = \lambda/\psi$, which is an order one macroscopic constant, definitively validating our scaling ansatz.

Following the same analysis, the terms $\tau_{\pm}^{(m)}$ (Eq. (160)) undergo the exact same algebraic suppression:

$$\tau_{\pm}^{(m)} \asymp d^{-\kappa} \nu^{-1/\alpha} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{dy_1 \dots dy_m}{(1 + (y_1 \dots y_m)^{\alpha})^2} \asymp d^{-\kappa\left(1-\frac{1}{\alpha}\right)} \log(d)^{m-1}$$

Summing all target shell contributions, the total $\tau_{\pm} = \sum_m \tau_{\pm}^{(m)}$ asymptotically vanishes ($\tau_{\pm} = \tilde{\Theta}(d^{-\kappa\left(1-\frac{1}{\alpha}\right)}) \to 0$). Since $\tau_- \le \tau \le \tau_+$, we also have $\tau = \tilde{\Theta}(d^{-\kappa\left(1-\frac{1}{\alpha}\right)}) \to 0$. The variance therefore scales as:

$$\mathsf{V} = \tilde{\Theta}\left(d^{-\kappa\left(1-\frac{1}{\alpha}\right)}\right)$$

where the $\tilde{\Theta}$ notation absorbs the residual sub-polynomial factor $f(d)$.

### E.2 Bias

We now turn our attention to the bias component. Recall that the deterministic equivalent $\mathsf{B}$ can be computed by:

$$\mathsf{B} = \frac{U}{1 - \tau}$$

where the unlearned energy $U$ explicitly depends on the target function:

$$U = \nu^2 \langle \boldsymbol{\theta}, (\boldsymbol{\Lambda} + \nu \boldsymbol{I})^{-2} \boldsymbol{\theta} \rangle \tag{163}$$

In the previous section, we proved that the $\tau$ term evaluates to $\tau \asymp 1 - \frac{1}{\alpha}$ in the ridgeless limit ($\lambda = 0$), whereas it vanishes ($\tau \to 0$) in the regularized regime ($\lambda = \Theta_d(1)$). We now proceed to compute the unlearned energy $U$ across both regularization regimes.

#### E.2.1 Ridgeless limit

The scaling of $\nu$ is given by $\nu = \xi d^{-\kappa\alpha} f(d)$. By definition, the macroscopic unlearned energy $U$ in Eq. (163) decomposes shell-wise as $U = \sum_m U^{(m)}$, where:

$$U^{(m)} = \frac{1}{m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{(\theta_{i_1,\dots,i_m})^2}{(1 + \nu^{-1} \lambda_{i_1,\dots,i_m}^{(m)})^2} \,. \tag{164}$$

Since we cannot evaluate the exact eigenvalues in the strongly anisotropic regime, we once again exploit the spectral inequalities defined in Eq. (132). This yields macroscopic upper and lower bounds $U_-^{(m)} \le U^{(m)} \le U_+^{(m)}$, defined by:

$$U_{\pm}^{(m)} = \frac{\theta_m^2}{m!} \sum_{i_1,\dots,i_m=1}^{d} \frac{(i_1 \dots i_m)^{-2\omega}}{(1 + (i_1 \dots i_m)^{-\alpha} \nu^{-1} C_{m,\pm}^{-1})^2} \,, \tag{165}$$

where $C_{m,\pm} = (c_{\pm}^{(m)})^{-1} (h_m m!)^{-1} \zeta(\alpha)^m$. In what follows, we will compute the explicit scaling of $U_{\pm}^{(m)}$ to establish the asymptotic behavior of the true unlearned energy $U^{(m)}$.

The mathematical evaluation of Eq. (165) naturally splits into two primary categories with the first one requiring a further subdivision based on the magnitude of $\omega$ with respect to $\alpha$:

- **The fast decay regime** ($\omega > \frac{1}{2}$). This range can be further partitioned into two distinct sub-categories:
  - *Ultra-fast decay* ($\omega > \alpha + \frac{1}{2}$);
  - *Moderately fast decay* ($\frac{1}{2} < \omega \leq \alpha + \frac{1}{2}$).
- **The slow decay regime** ($\omega < \frac{1}{2}$).

**Ultra-fast decay $\boldsymbol{\omega > \alpha + \frac{1}{2}}$ :** Let us rewrite the bounds for the unlearned energy $U_{\pm}^{(m)}$ as:

$$U_{\pm}^{(m)} = \frac{\theta_m^2}{m!}\nu^2 \sum_{i_1,\dots,i_m=1}^{d} \frac{(i_1 \dots i_m)^{-2\omega}}{(\nu + (i_1 \dots i_m)^{-\alpha} C_{m,\pm}^{-1})^2} \,.$$

We haven't yet made the scaling of $\nu$ explicit. As it turns out, it's not strictly required; we will only need to assume $\nu \to 0$ as $d \to \infty$. When performing this high-dimensional limit, we can apply the Dominated Convergence Theorem, bounding each term using the absolutely summable sequence $\sum_{i_1,\dots,i_m} (i_1 \dots i_m)^{-2(\omega-\alpha)}$ (converging since $2(\omega - \alpha) > 1$). Taking the limit $\nu \to 0$, then at leading order:

$$U_{\pm}^{(m)} \sim \frac{\theta_m^2}{m!}\nu^2 C_{m,\pm}^2 \sum_{i_1,\dots,i_m=1}^{\infty} (i_1 \dots i_m)^{-2(\omega-\alpha)} \,.$$

Now we have to consider the scaling behavior of each term appearing in the above equation. The infinite summation is converging to $\zeta(2(\omega - \alpha))$ and, since $\omega > \frac{1}{2}$, then $\theta_m$ is of order-one with respect to $d$. Furthermore, the bounding prefactors $C_{m,\pm}^2$ are macroscopic $\Theta_d(1)$ constants. Hence the only asymptotic dependence is on $\nu$:

$$U_{\pm}^{(m)} \asymp \nu^2 \,.$$

Since the true unlearned energy of the shell is strictly sandwiched by these bounds ($U_{-}^{(m)} \leq U^{(m)} \leq U_{+}^{(m)}$), it must obey the exact same asymptotic scaling. We know the scaling of $\nu$ up to a sub-polynomial correction, so that we finally obtain:

$$U^{(m)} = \widetilde{\Theta}(d^{-2\kappa\alpha}) \,,$$

meaning that the kernel has completely learned every polynomial feature. The final bias will therefore scale similarly:

$$\mathsf{B} = \tilde{\Theta}(d^{-2\kappa\alpha}) \,,$$

or, remembering $n \asymp d^{\kappa}$, then:

$$\mathsf{B} = \tilde{\Theta}(n^{-2\alpha}) \,.$$

**Moderately-fast decay $\boldsymbol{\frac{1}{2} < \omega \leq \alpha + \frac{1}{2}}$ :** We can asymptotically approximate the summation defining the bounds $U_{\pm}^{(m)}$ in Eq. (165) as the corresponding Riemann integral:

$$U_{\pm}^{(m)} \sim \frac{\theta_m^2}{m!} \int_1^d \cdots \int_1^d \frac{dx_1 \dots dx_m}{((x_1 \dots x_m)^{\omega} + \nu^{-1} C_{m,\pm}^{-1} (x_1 \dots x_m)^{\omega-\alpha})^2}$$

and performing the change of variables:

$$y_i = x_i \nu^{1/m\alpha} C_{m,\pm}^{1/m\alpha}$$

so that now:

$$U_{\pm}^{(m)} = \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{dy_1 \cdots dy_m}{((y_1 \dots y_m)^{\omega} + (y_1 \cdots y_m)^{\omega-\alpha})^2}$$

or

$$\begin{aligned} U_{\pm}^{(m)} &= \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{(y_1 \dots y_m)^{2(\alpha-\omega)}}{((y_1 \dots y_m)^{\alpha} + 1)^2} dy_1 \cdots dy_m = \\ &= \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} \mathcal{I}^{(m)}(\alpha, \omega) \end{aligned} \tag{166}$$

where the limits of the integral $\mathcal{I}^{(m)}(\alpha,\omega)$ are implicitly dependent on the specific bound chosen:

$$y_{\min} = \nu^{\frac{1}{m\alpha}} C_{m,\pm}^{\frac{1}{m\alpha}} = C_{m,\pm}^{\frac{1}{m\alpha}} \xi^{\frac{1}{m\alpha}} f(d)^{\frac{1}{m\alpha}} d^{-\frac{\kappa}{m}} \tag{167a}$$

$$y_{\max} = d\nu^{\frac{1}{m\alpha}} C_{m,\pm}^{\frac{1}{m\alpha}} = C_{m,\pm}^{\frac{1}{m\alpha}} \xi^{\frac{1}{m\alpha}} f(d)^{\frac{1}{m\alpha}} d^{1-\frac{\kappa}{m}} \,. \tag{167b}$$

These limits are asymptotically represented in Fig. 9. The lower limit $y_{\min} = \widetilde{\Theta}(d^{-\frac{\kappa}{m}})$ always collapses to zero. The upper limit $y_{\max} = \widetilde{\Theta}(d^{1-\frac{\kappa}{m}})$ converges to 0 when $m < \kappa$ and diverges to $\infty$ for $m > \kappa$. Note that when $m = \kappa$, the scaling $f(d)$ makes $y_{\max}$ diverge to $\infty$ instead of converging to an order-one constant. As usual, we will solve Eq. (166) according to the different values of $m$:

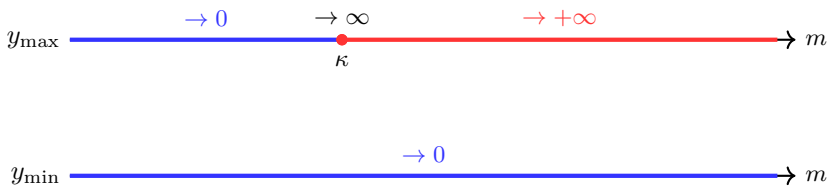


Figure 9: *Asymptotic behavior of the integration boundaries $y_{\min}$ and $y_{\max}$ derived in Eqs. equation 167. At variance with Fig. 5, when $m = \kappa$ the respective boundary does not converge to a constant $\Theta_d(1)$ due to the diverging factor $f(d)$.*

**Case a: $m < \kappa$** Here both $y_{\min}, y_{\max} \to 0$. The integral in Eq. (166) can therefore be asymptotically computed by:

$$\mathcal{I}^{(m)}(\alpha,\omega) = \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{(y_1 \dots y_m)^{2(\alpha-\omega)}}{((y_1 \dots y_m)^{\alpha}+1)^2} dy_1 \dots dy_m \sim \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} (y_1 \dots y_m)^{2(\alpha-\omega)} \tag{168}$$

When $\omega = \alpha + \frac{1}{2}$, then the integral in Eq. (168) becomes logarithmic:

$$\int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} (y_1 \dots y_m)^{-1} \, dy_1 \dots dy_m = \log^m \left( \frac{y_{\max}}{y_{\min}} \right) = \log(d)^m$$

and, finally, $U_{\pm}^{(m)} \asymp \frac{\theta_m^2}{m!} \nu^2 \log(d)^m \asymp \frac{\theta_m^2}{m!} d^{-2\kappa\alpha} f(d)^2 \log(d)^m = \tilde{\Theta}(d^{-2\kappa\alpha}) \to 0$.

If $\frac{1}{2} < \omega < \alpha + \frac{1}{2}$, the integral $\mathcal{I}^{(m)}(\alpha,\omega)$ becomes:

$$\int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} (y_1 \dots y_m)^{2(\alpha-\omega)} dy_1 \dots dy_m \asymp (y_{\max}^{1+2(\alpha-\omega)} - y_{\min}^{1+2(\alpha-\omega)})^m$$

since $1 + 2(\alpha - \omega) > 0$, the leading contribution is the one associated to $y_{\max}$, hence:

$$U_{\pm}^{(m)} \asymp \frac{\theta_m^2}{m!} y_{\max}^{m(1+2(\alpha-\omega))} \nu^{\frac{2\omega-1}{\alpha}} = \frac{\theta_m^2}{m!} d^{m+2m(\alpha-\omega)} \nu^{\frac{1+2(\alpha-\omega)}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} = \frac{\theta_m^2}{m!} d^{m+2m(\alpha-\omega)} \nu^2$$

As $\frac{1}{2} < \omega < \frac{1}{2} + \alpha$, the target coefficient $\theta_m$ is of order 1 and:

$$U_{\pm}^{(m)} \asymp \frac{\theta_m^2}{m!} d^{m+2m(\alpha-\omega)} \nu^2 \asymp d^{m+2m(\alpha-\omega)} d^{-2\kappa\alpha} f(d)^2 = d^{m(1-2\omega)} d^{2\alpha(m-\kappa)} f(d)^2 \to 0 \,,$$

since $m < \kappa$ and $1 - 2\omega < 0$. Since the bounds $U_{\pm}^{(m)}$ converge asymptotically to 0 with the same scaling, the exact energy term $U^{(m)}$ must follow the exact same rate. Therefore, for $m < \kappa$:

$$U^{(m)} = \widetilde{\Theta}(d^{m(1-2\omega)+2\alpha(m-\kappa)}) \to 0 \tag{169}$$

and $\sum_{m<\kappa} U^{(m)}$ is negligible

**Case b: $m > \kappa$ :** Using the technique in Lemma D.1, we can rewrite the integral $\mathcal{I}^{(m)}(\alpha,\omega)$ in a one-dimensional form. Recalling that $m\log(y_{\min}) = -\kappa\log(d) + \Delta$ and $m\log(y_{\max}) = (m-\kappa)\log(d) + \Delta$, where $\Delta = \Theta(\alpha^{-1}\log(f(d))) = o(\log d)$, we obtain:

$$\mathcal{I}^{(m)}(\alpha,\omega) = \log(d)^{m-1}\int_{-\kappa\log(d)+\Delta}^{(m-\kappa)\log(d)+\Delta} dt \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2} f_{IH}\left(\kappa + \frac{t-\Delta}{\log(d)}; m\right). \tag{170}$$

We now analyze the behavior of this expression as $d \to \infty$. This is perfectly equivalent to the integral we analyzed in the proof of Lemma E.1 since the integrand weight function $\eta(t) = \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2}$ is absolutely integrable on the real line for $\frac{1}{2} < \omega < \frac{1}{2} + \alpha$. By dominated convergence, we can therefore evaluate the Irwin-Hall density strictly at the critical threshold $\kappa$ and factor it out of the integral:

$$\mathcal{I}^{(m)}(\alpha,\omega) \simeq \log(d)^{m-1} f_{IH}(\kappa; m)\int_{-\infty}^{\infty} dt \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2}. \tag{171}$$

Finally, substituting this back into the expression for $U_{\pm}^{(m)}$, its value at leading order becomes:

$$U_{\pm}^{(m)} \simeq \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} \log(d)^{m-1} f_{IH}(\kappa; m)\int_{-\infty}^{\infty} dt \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2}.$$

Since the remaining one-dimensional integral over $t$ converges to an order-one constant, and recalling the scalings $\nu = \tilde{\Theta}(d^{-\kappa\alpha})$, $\theta_m = \Theta(1)$, and $C_{m,\pm} = \Theta_d(1)$, we readily obtain:

$$U_{\pm}^{(m)} = \tilde{\Theta}(d^{-\kappa(2\omega-1)}).$$

Since the upper and lower bounds exhibit the same asymptotic behavior, we can drop the $\pm$ indices for the final evaluation. Consequently, summing over the super-critical shells yields:

$$\sum_{m>\kappa}^{\infty} U^{(m)} = \tilde{\Theta}(d^{-\kappa(2\omega-1)}).$$

**Case c: $m = \kappa$ :** We use again Lemma D.1. Recalling the boundary scalings, we have $\kappa\log(y_{\max}) = (m-\kappa)\log(d) + \Delta = \Delta$ and $\kappa\log(y_{\min}) = -\kappa\log(d) + \Delta$, where $\Delta = \Theta(\alpha^{-1}\log(f(d))) = o(\log(d)) \to \infty$. The integral becomes:

$$\mathcal{I}^{(\kappa)}(\alpha,\omega) = \log(d)^{\kappa-1}\int_{-\kappa\log(d)+\Delta}^{\Delta} dt \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2} f_{IH}\left(\kappa + \frac{t-\Delta}{\log(d)}; \kappa\right). \tag{172}$$

This integral is formally equivalent to the one analyzed in the proof of Lemma E.1 in the case $m = \kappa$, the only difference being in the absolutely integrable integrand weight function. Leveraging the DCT and expanding the Irwin-Hall function near its rightmost boundary, we obtain:

$$\mathcal{I}^{(\kappa)}(\alpha,\omega) \simeq \frac{\Delta^{\kappa-1}}{(\kappa-1)!}\int_{-\infty}^{\infty} dt \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2} \asymp \Delta^{\kappa-1} \asymp \big(\log f(d)\big)^{\kappa-1}. \tag{173}$$

Finally, substituting this sub-logarithmic growth back into the macroscopic unlearned energy evaluated in Eq. (166), and recalling that $\nu = \tilde{\Theta}(d^{-\kappa\alpha})$, we obtain:

$$U_{\pm}^{(\kappa)} \asymp \nu^{\frac{2\omega-1}{\alpha}} \mathcal{I}^{(\kappa)}(\alpha,\omega) \asymp d^{-\kappa(2\omega-1)}\big(\log \hat{f}(d)\big)^{\kappa-1} = \widetilde{\Theta}\Big(d^{-\kappa(2\omega-1)}\Big). \tag{174}$$

Exploiting once again the sandwich theorem for the exact shell energy, we conclude that the true resonant shell $m = \kappa$ decays at the exact same polynomial rate as the super-critical shells $m > \kappa$, being immune to the $\log(d)^{\kappa-1}$ explosion:

$$U^{(\kappa)} = \widetilde{\Theta}\Big(d^{-\kappa(2\omega-1)}\Big).$$

Combining all the individual regimes together, we finally arrive at the complete asymptotic picture showing that the unlearned energy, and consequently the macroscopic bias, vanishes in the high-dimensional limit. In particular, the overall convergence rate is dictated by the contributions of the deep shells ($m > \kappa$), yielding:

$$\mathsf{B} = \widetilde{\Theta}\left(d^{-\kappa(2\omega-1)}\right) = \widetilde{\Theta}\left(n^{-(2\omega-1)}\right). \tag{175}$$

**Slow decay $\omega < \frac{1}{2}$ :** Similarly to what we did in the moderately-fast regime, we will split our analysis into three distinct parts. The bounding quantities we want to compute are:

$$U_{\pm}^{(m)} = \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} \frac{dy_1 \cdots dy_m}{((y_1 \dots y_m)^{\omega} + (y_1 \cdots y_m)^{\omega-\alpha})^2} = \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} \mathcal{I}^{(m)}(\alpha,\omega)\,.$$

**Case a: $m < \kappa$:** Following the same procedure as the one in the moderately-fast regime, we approximate the summation as a Riemann integral. Since both limits vanish ($y_{\min}, y_{\max} \to 0$), the integral scales as:

$$\begin{aligned} \mathcal{I}^{(m)}(\alpha,\omega) &\simeq \int_{y_{\min}}^{y_{\max}} \cdots \int_{y_{\min}}^{y_{\max}} (y_1 \cdots y_m)^{2(\alpha-\omega)} dy_1 \cdots dy_m \\ &\asymp \left(y_{\max}^{1+2\alpha-2\omega} - y_{\min}^{1+2\alpha-2\omega}\right)^m \asymp y_{\max}^{m(1+2\alpha-2\omega)}\,. \end{aligned} \tag{176}$$

Consequently, we obtain the scaling for the unlearned energy bounds:

$$U_{\pm}^{(m)} \asymp \frac{\theta_m^2}{m!} y_{\max}^{m(1+2\alpha-2\omega)} \nu^{\frac{2\omega-1}{\alpha}} \asymp \frac{\theta_m^2}{m!} d^{m+2m(\alpha-\omega)} \nu^2\,. \tag{177}$$

The fundamental difference here is that in the slow decay regime ($\omega < 1/2$), the target coefficients $\theta_m$ must scale with $d$ to maintain a finite total energy for the target function $f_\star$. Specifically, we established that their squared norm scales as $\theta_m^2 = \Theta(d^{-m(1-2\omega)})$. Therefore, replacing this into our expression alongside $\nu^2 = \widetilde{\Theta}(d^{-2\kappa\alpha})$, the unlearned energy bounds for the $m < \kappa$ shells elegantly reduce to:

$$U_{\pm}^{(m)} \asymp d^{-m(1-2\omega)} d^{m+2m(\alpha-\omega)} \nu^2 = \widetilde{\Theta}\left(d^{2\alpha(m-\kappa)}\right) \to 0\,. \tag{178}$$

Since $m < \kappa$, the exponent is strictly negative. Because both bounds vanish, the true exact energy $U^{(m)} \to 0$, meaning that the kernel perfectly resolves the sub-critical shells.

**Case b: $m > \kappa$** We will change the variables in Eq. (170) with $t = m\log(y_{\max}) - S = mL - S$:

$$\begin{aligned} \mathcal{I}^{(m)}(\alpha,\omega) &= \log(d)^{m-1} \int_0^{m\log(d)} dt\, \frac{e^{(mL-t)(1+2\alpha-2\omega)}}{(1+e^{\alpha(mL-t)})^2} f_{IH}\left(m - \frac{t}{\log(d)}; m\right) \\ &= \log(d)^{m-1} e^{mL(1+2\alpha-2\omega)} \int_0^{m\log(d)} dt\, \frac{e^{-t(1+2\alpha-2\omega)}}{(1+e^{m\alpha L} e^{-\alpha t})^2} f_{IH}\left(m - \frac{t}{\log(d)}; m\right) \\ &= \log(d)^{m-1} y_{\max}^{m(1+2\alpha-2\omega)} \int_0^{m\log(d)} dt\, \frac{e^{-t(1+2\alpha-2\omega)}}{(1+y_{\max}^{m\alpha} e^{-\alpha t})^2} f_{IH}\left(m - \frac{t}{\log(d)}; m\right) \\ &= y_{\max}^{m(1-2\omega)} \int_0^{\infty} dt\, \mathbf{1}_{[0,m\log(d)]}(t) \log(d)^{m-1} \frac{e^{-t(1+2\alpha-2\omega)}}{(y_{\max}^{-m\alpha} + e^{-\alpha t})^2} f_{IH}\left(m - \frac{t}{\log(d)}; m\right)\,. \end{aligned} \tag{179}$$

Now we can use the Dominated Convergence Theorem (DCT) on the sequence of functions:

$$g_d(t) = \mathbf{1}_{[0,m\log(d)]}(t) \log(d)^{m-1} \frac{e^{-t(1+2\alpha-2\omega)}}{(y_{\max}^{-m\alpha} + e^{-\alpha t})^2} f_{IH}\left(m - \frac{t}{\log(d)}; m\right)\,.$$

Similarly to what we have already shown (expanding the Irwin-Hall function near its boundary in the limit $d \to \infty$), this pointwise converges to:

$$\lim_{d\to\infty} g_d(t) = \frac{1}{(m-1)!} e^{-t(1-2\omega)} t^{m-1}\,.$$

To rigorously bound the sequence, we note that $(y_{\max}^{-m\alpha} + e^{-\alpha t})^2 > e^{-2\alpha t}$. Combined with the global bound $f_{IH}(m-x;m) \le x^{m-1}$, we obtain:

$$|g_d(t)| \le t^{m-1} \frac{e^{-t(1+2\alpha-2\omega)}}{e^{-2\alpha t}} = t^{m-1} e^{-t(1-2\omega)} \equiv G(t)\,,$$

which is absolutely integrable on $\mathbb{R}^+$ since $\omega < \frac{1}{2}$. By the DCT, we get at leading order:

$$\begin{aligned} \mathcal{I}^{(m)}(\alpha,\omega) &\simeq \frac{1}{(m-1)!} y_{\max}^{m(1-2\omega)} \int_0^{\infty} dt\, t^{m-1} e^{-t(1-2\omega)} \\ &= y_{\max}^{m(1-2\omega)} (1-2\omega)^{-m}\,. \end{aligned} \tag{180}$$

Finally, substituting this back into $U_{\pm}^{(m)}$, we obtain:

$$U_{\pm}^{(m)} \sim \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} y_{\max}^{m(1-2\omega)} (1-2\omega)^{-m} .$$

Crucially, recalling from Eq. (167b) that $y_{\max}^{m(1-2\omega)} \propto C_{m,\pm}^{\frac{1-2\omega}{\alpha}}$, the bounding constants structurally cancel out: $C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \times C_{m,\pm}^{\frac{1-2\omega}{\alpha}} = C_{m,\pm}^{0} = 1$. This implies that the asymptotic dependence on the specific eigenvalue fluctuations completely vanishes, yielding identically for both bounds:

$$U_{\pm}^{(m)} \sim \frac{\theta_m^2}{m!} d^{m(1-2\omega)} (1-2\omega)^{-m} ,$$

and since $\omega < \frac{1}{2}$, this last expression is precisely the leading term of the target energy for the $m$-th shell, $E(m)$. Because the upper and lower bounds rigorously squeeze to the identical deterministic limit, the exact unlearned energy perfectly converges:

$$U^{(m)} \to E(m) .$$

Hence, we find that when $m > \kappa$, the unlearned energy saturates to the full spectral energy.

**Case c: $m = \kappa$** Due to the sub-polynomial divergence of $f(d)$ as $d \to \infty$ in Eq. (167), the asymptotic evaluation of the resonant shell $m = \kappa$ identically follows the deep-shell regime ($m > \kappa$). The macroscopic bounding constants $C_{\kappa,\pm}$ undergo the exact same structural cancellation, yielding a perfectly sharp limit $U^{(\kappa)} = E(\kappa)$. This property is entirely driven by the presence of $f(d)$; in its absence, the $\kappa$-th shell would exhibit a different behavior. From a learning perspective, the condition $U^{(\kappa)} = E(\kappa)$ demonstrates that the resonant shell remains completely unlearned, retaining its full initial target energy.

Synthesizing the results across all polynomial shells, we obtain the definitive asymptotic limit for the total exact unlearned energy in the slow decay regime ($\omega < 1/2$):

$$U = \sum_{m=\kappa}^{\infty} E(m) . \tag{181}$$

**Summary for unregularized regime** Putting everything together, since the true unlearned energy $U$ scales identically to its analytical bounds, we have:

- When $\omega > \alpha + \frac{1}{2}$:

$$U = \sum_{m<\kappa} U^{(m)} + U^{(\kappa)} + \sum_{m>\kappa} U^{(m)} = \tilde{\Theta}(d^{-2\kappa\alpha}) ,$$

- When $\frac{1}{2} < \omega \leq \alpha + \frac{1}{2}$:

$$U = \sum_{m<\kappa} U^{(m)} + U^{(\kappa)} + \sum_{m>\kappa} U^{(m)} = \tilde{\Theta}(d^{-\kappa(2\omega-1)}) ,$$

- When $\omega < \frac{1}{2}$:

$$U = \sum_{m<\kappa} U^{(m)} + U^{(\kappa)} + \sum_{m>\kappa} U^{(m)} = \sum_{m\geq\kappa} E(m) .$$

and the full bias is given by $\mathsf{B} = \frac{1}{1-\tau} U \asymp \alpha U$.

### E.2.2 Regularized regime

We now turn to the computation of the unlearned energy of the $m$-th shell in the dense setting. Employing the spectral bounds, we write:

$$U_{\pm}^{(m)} = \frac{\theta_m^2}{m!} \sum_{i_1,\ldots,i_m=1}^{d} \frac{(i_1 \cdots i_m)^{-2\omega}}{\left(1 + C_{m,\pm}^{-1} \nu^{-1} (i_1 \cdots i_m)^{-\alpha}\right)^2} , \tag{182}$$

where we introduce the macroscopic regularization $\lambda = \Theta_d(1)$. As established in Eq. (161), this assumption imposes a strict scaling on the parameter $\nu$:

$$\nu = \xi d^{-\kappa} = \frac{\lambda}{\psi} d^{-\kappa} . \tag{183}$$

Aside from this modified scaling, the functional structure of $U_{\pm}^{(m)}$ remains unchanged, allowing us to adapt the analytical computations performed for the ridgeless setting ($\lambda = 0$). To systematically evaluate the asymptotic behavior of these sums, we transition to the multi-dimensional integral representation of Eq. (166):

$$U_{\pm}^{(m)} \simeq \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \nu^{\frac{2\omega-1}{\alpha}} \mathcal{I}^{(m)}(\alpha, \omega) = \frac{\theta_m^2}{m!} C_{m,\pm}^{\frac{2\omega-1}{\alpha}} \xi^{\frac{2\omega-1}{\alpha}} d^{-\kappa\left(\frac{2\omega-1}{\alpha}\right)} \mathcal{I}^{(m)}(\alpha, \omega) , \tag{184}$$

where the upper and lower integration boundaries, previously defined in Eqs. equation 167, are now updated to reflect the macroscopic ridge scaling of $\nu$ alongside the bounds:

$$y_{\min} = (\nu C_{m,\pm})^{\frac{1}{m\alpha}} = C_{m,\pm}^{\frac{1}{m\alpha}} \xi^{\frac{1}{m\alpha}} d^{-\frac{\kappa}{m\alpha}} , \tag{185a}$$

$$y_{\max} = d y_{\min} = C_{m,\pm}^{\frac{1}{m\alpha}} \xi^{\frac{1}{m\alpha}} d^{1-\frac{\kappa}{m\alpha}} . \tag{185b}$$

By introducing the effective capacity threshold $\kappa_{\text{eff}} = \frac{\kappa}{\alpha} < \kappa$, these boundaries can be elegantly rewritten as:

$$y_{\min} = \Theta\left(d^{-\frac{\kappa_{\text{eff}}}{m}}\right) , \tag{186a}$$

$$y_{\max} = \Theta\left(d^{1-\frac{\kappa_{\text{eff}}}{m}}\right) . \tag{186b}$$

When $m < \kappa_{\text{eff}}$, the exponent $1 - \kappa_{\text{eff}}/m$ becomes strictly negative. Consequently, both the upper and lower integration limits asymptotically converge to zero. We can thus safely leverage the analytical machinery developed for the ridgeless setting: the sub-critical regime $m < \kappa_{\text{eff}}$ under a macroscopic ridge penalty is formally equivalent to the regime $m < \kappa$ in the unregularized case. In both scenarios, the integral vanishes, confirming that the sub-critical shells ($m < \kappa_{\text{eff}}$) are perfectly learned by the kernel, yielding $U^{(m)} \to 0$.

A parallel reasoning applies to the super-critical shells ($m > \kappa_{\text{eff}}$), which directly map to the unregularized $m > \kappa$ scenario. While the ridgeless setting required the modulating sub-polynomial factor $f(d)$ to regularize the summations, here the strong polynomial decay inherently dominates the asymptotic behavior, rendering the two problems formally equivalent. By directly evaluating the integral, we thus obtain the definitive asymptotic behavior for the unlearned exact energy of the deep shells ($m > \kappa_{\text{eff}}$):

$$U^{(m)} \asymp \begin{cases} E(m) & \text{for } \omega < \frac{1}{2} , \\ \log(d)^{m-1} \nu^{\frac{2\omega-1}{\alpha}} = \widetilde{\Theta}\left(n^{-\frac{2\omega-1}{\alpha}}\right) & \text{for } \frac{1}{2} < \omega \leq \alpha + \frac{1}{2} , \\ \nu^2 = \Theta(n^{-2}) & \text{for } \omega > \alpha + \frac{1}{2} . \end{cases} \tag{187}$$

The only main departure from the ridgeless setting comes out when we analyze the resonant shell $m = \kappa_{\text{eff}}$ (equivalent to the ridgeless $m = \kappa$). Assuming there exists an integer such that $m = \kappa_{\text{eff}}$, the integration boundaries defined in Eq. (186) behave differently. In the unregularized case, the presence of the modulating factor $f(d)$ forced the upper limit $y_{\max}$ to slowly diverge to infinity for the resonant shell. Here, however, the strict absence of $f(d)$ implies that the term $1 - \kappa_{\text{eff}}/m$ is exactly zero and, as a direct consequence, the upper bound $y_{\max}$ does not diverge, but instead rigorously settles on a finite, macroscopic $\Theta_d(1)$ constant as $d \to \infty$.

For fast-decaying targets ($\omega > 1/2$), the mathematical treatment maps almost identically to the ridgeless scenario, with the crucial distinction that the multi-dimensional integral now evaluates to a pure, order-one constant, completely devoid of any sub-logarithmic modulations. The resonant shell thus seamlessly joins the power-law scaling of the super-critical modes, yielding an exact unlearned energy that vanishes as:

$$U^{(\kappa_{\text{eff}})} = \Theta\left(d^{-\kappa \min\left(\frac{2\omega-1}{\alpha}, 2\right)}\right) . \tag{188}$$

It remains to evaluate the critical shell $m = \kappa_{\text{eff}}$ in the slow-decay regime ($\omega < 1/2$), under the assumption that the data anisotropy $\alpha$ yields an integer effective threshold, $\kappa_{\text{eff}} \in \mathbb{N}$. Unlike the unregularized limit, the absence of the diverging sub-polynomial factor $f(d)$ stabilizes the upper integration boundary $y_{\max}$ to an $\Theta_d(1)$ constant.

The analytical procedure closely mirrors the techniques developed in the previous sections. We aim to compute the core integral $\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha, \omega)$. Applying Lemma D.1 and evaluating the integration limits, we obtain:

$$\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \log(d)^{\kappa_{\text{eff}}-1} \int_{-\kappa_{\text{eff}}\log(d)+\Delta_\pm}^{\Delta_\pm} dt \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2} f_{IH}\left(\kappa_{\text{eff}} + \frac{t-\Delta_\pm}{\log(d)}; \kappa_{\text{eff}}\right), \quad (189)$$

where the constant integration bound is defined as $\Delta_\pm = \frac{1}{\alpha}\log(C_{\kappa_{\text{eff}},\pm}\xi)$ or, equivalently, $e^{\Delta_\pm/\kappa_{\text{eff}}} = C_{\kappa_{\text{eff}},\pm}^{\frac{1}{\kappa_{\text{eff}}\alpha}} \xi^{\frac{1}{\kappa_{\text{eff}}\alpha}}$. To rigorously evaluate the high-dimensional limit ($d \to \infty$), we apply the Dominated Convergence Theorem (DCT). We define the sequence of integrands:

$$g_d(t) = \mathbf{1}_{[-\kappa_{\text{eff}}\log(d)+\Delta_\pm, \Delta_\pm]}(t) \log(d)^{\kappa_{\text{eff}}-1} \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2} f_{IH}\left(\kappa_{\text{eff}} + \frac{t-\Delta_\pm}{\log(d)}; \kappa_{\text{eff}}\right).$$

By asymptotically expanding the Irwin-Hall distribution near its rightmost boundary, the polynomial $\log(d)$ divergence is perfectly canceled, yielding the point-wise limit:

$$\lim_{d\to\infty} g_d(t) = \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2} \frac{(\Delta_\pm - t)^{\kappa_{\text{eff}}-1}}{(\kappa_{\text{eff}}-1)!},$$

which does not depend on $d$ (recall that $\Delta_\pm$ is now a constant). Leveraging the usual property of the Irwin-Hall distribution, this sequence is globally uniformly bounded by the dominating function:

$$G(t) = \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2} (\Delta_\pm - t)^{\kappa_{\text{eff}}-1},$$

which is strictly integrable over $(-\infty, \Delta_\pm]$. Thus, passing the limit inside the integral yields:

$$\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \frac{1}{(\kappa_{\text{eff}}-1)!} \int_{-\infty}^{\Delta_\pm} dt \frac{e^{t(1+2\alpha-2\omega)}}{(1+e^{\alpha t})^2} (\Delta_\pm - t)^{\kappa_{\text{eff}}-1}.$$

To evaluate this, we introduce the change of variables $z = \Delta_\pm - t$, obtaining:

$$\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \frac{e^{\Delta_\pm(1+2\alpha-2\omega)}}{(\kappa_{\text{eff}}-1)!} \int_0^\infty dz \frac{e^{-z(1+2\alpha-2\omega)}}{(1+e^{\alpha\Delta_\pm}e^{-\alpha z})^2} z^{\kappa_{\text{eff}}-1}.$$

Substituting $e^{\alpha\Delta_\pm} = C_{\kappa_{\text{eff}},\pm}\xi$, the integral assumes its final explicit form:

$$\mathcal{I}^{(\kappa_{\text{eff}})}(\alpha,\omega) = \frac{1}{(\kappa_{\text{eff}}-1)!} (C_{\kappa_{\text{eff}},\pm}\xi)^{\frac{1+2\alpha-2\omega}{\alpha}} \int_0^\infty dz \frac{e^{-z(1+2\alpha-2\omega)}}{(1+C_{\kappa_{\text{eff}},\pm}\xi e^{-\alpha z})^2} z^{\kappa_{\text{eff}}-1}.$$

We can now reinsert this result into the general expression for $U_\pm^{(\kappa_{\text{eff}})}$. Renaming the dummy integration variable to $t$, we obtain:

$$U_\pm^{(\kappa_{\text{eff}})} = \frac{\theta_{\kappa_{\text{eff}}}^2}{\kappa_{\text{eff}}!} d^{-\kappa_{\text{eff}}(2\omega-1)} \frac{(C_{\kappa_{\text{eff}},\pm}\xi)^2}{(\kappa_{\text{eff}}-1)!} \int_0^\infty dt \frac{e^{-t(1+2\alpha-2\omega)}}{(1+C_{\kappa_{\text{eff}},\pm}\xi e^{-\alpha t})^2} t^{\kappa_{\text{eff}}-1}.$$

We can further restructure the integral by factoring out $e^{-2\alpha t}$ from the denominator, which simplifies the exponential terms and explicitly recovers the initial unlearned energy definition $E(\kappa_{\text{eff}})$:

$$U_\pm^{(\kappa_{\text{eff}})} = E(\kappa_{\text{eff}}) \frac{(1-2\omega)^{\kappa_{\text{eff}}}}{(\kappa_{\text{eff}}-1)!} \int_0^\infty dt\, e^{-t} t^{\kappa_{\text{eff}}-1} \frac{e^{2t\omega}}{(1+C_{\kappa_{\text{eff}},\pm}^{-1}\xi^{-1}e^{\alpha t})^2}.$$

Notice that the functional structure of this integral perfectly mirrors the one derived for the residual fraction $\eta$ in the weakly anisotropic regime. By substituting the explicit value of $C_{\kappa_{\text{eff}},\pm}^{-1} = c_\pm^{(\kappa_{\text{eff}})} h_{\kappa_{\text{eff}}} \kappa_{\text{eff}}! \zeta(\alpha)^{-\kappa_{\text{eff}}}$, and recalling that $\xi^{-1} = \frac{\psi}{\lambda}$ under macroscopic regularization, we obtain the final exact form for the critical shell bounds:

$$U_\pm^{(\kappa_{\text{eff}})} = E(\kappa_{\text{eff}}) \frac{(1-2\omega)^{\kappa_{\text{eff}}}}{(\kappa_{\text{eff}}-1)!} \int_0^\infty dt\, e^{-t} t^{\kappa_{\text{eff}}-1} \frac{e^{2t\omega}}{\left(1 + c_\pm^{(\kappa_{\text{eff}})} h_{\kappa_{\text{eff}}} \kappa_{\text{eff}}! \zeta(\alpha)^{-\kappa_{\text{eff}}} \frac{\psi}{\lambda} e^{\alpha t}\right)^2} = \eta'_\pm E(\kappa_{\text{eff}}).$$

By the sandwich theorem, the exact unlearned energy for the critical shell is therefore tightly bounded within a macroscopic $\Theta_d(1)$ interval determined by these structural fractions: $\eta'_- E(\kappa_{\text{eff}}) \leq U^{(\kappa_{\text{eff}})} \leq \eta'_+ E(\kappa_{\text{eff}})$. Alternatively, we can write:

$$U^{(\kappa_{\text{eff}})} = C\eta' E(\kappa_{\text{eff}})$$

with

$$\eta' = \frac{(1-2\omega)^{\kappa_{\text{eff}}}}{(\kappa_{\text{eff}}-1)!} \int_0^\infty dt\, e^{-t} t^{\kappa_{\text{eff}}-1} \frac{e^{2t\omega}}{\left(1 + h_{\kappa_{\text{eff}}} \kappa_{\text{eff}}! \zeta(\alpha)^{-\kappa_{\text{eff}}} \frac{\psi}{\lambda} e^{\alpha t}\right)^2}$$

□

# F Hermite Ansatz Conjecture and Single-Index Models

In this section, we provide a more detailed discussion regarding the Hermite ansatz introduced in Section 4. From Eq. (7), we observe that the deterministic bias is essentially determined by the decomposition of the target function onto the exact kernel eigenbasis $\{e_{\boldsymbol{\beta}}\}$. In general, obtaining these true eigenfunctions is a highly non-trivial task, as it requires the explicit diagonalization of the kernel integral operator. The *Hermite ansatz* circumvents this difficulty by postulating that, as far as the generalization risk is concerned, the kernel behaves effectively as if its eigenfunctions were exactly the Hermite tensors $H_{\boldsymbol{\beta}}$, with corresponding eigenvalues given by the leading-order asymptotic terms identified in Proposition 1. In particular:

$$H_{\boldsymbol{\beta}}(\boldsymbol{x}) = \frac{1}{\sqrt{\boldsymbol{\beta}!}} \prod_{j=1}^{d} \mathrm{He}_{\beta_j}\left(\frac{x_j}{\sqrt{\sigma_j}}\right)$$

with $\mathrm{He}_\beta(z)$ the probabilist's Hermite polynomial of order $\beta$. To be more precise in the formulation of the ansatz, let $c_{\boldsymbol{\beta}} := \langle f_\star, H_{\boldsymbol{\beta}} \rangle_{L^2(p_X)}$ denote the Hermite coefficients of the target function $f_\star$, and let $\theta_{\boldsymbol{\beta}} = \langle f_\star, e_{\boldsymbol{\beta}} \rangle_{L^2(p_X)}$ represent the true projection coefficients onto the exact kernel eigenfunctions, corresponding to the eigenpairs $(\lambda_{\boldsymbol{\beta}}, e_{\boldsymbol{\beta}})$ characterized in Proposition 1. The deterministic equivalents in Eq. (7) depend on $\theta_{\boldsymbol{\beta}}$ only through the cumulative target energy below a given eigenvalue: $A(u) := \sum_{\boldsymbol{\beta}: \lambda_{\boldsymbol{\beta}} \leq u} \theta_{\boldsymbol{\beta}}^2$. Denote by $\widehat{A}$ the same function built from $c_{\boldsymbol{\beta}}$.

**Conjecture 1** (Hermite ansatz)**.** Let $\alpha \geq 0$ and let the kernel $k$ satisfy Assumption 1.1. Then there exists $\varepsilon(d)$, such that $\varepsilon(d) \to 0$ as $d \to \infty$ such that, for every $f_\star \in L^2(p_X)$ with $\|f_\star\|_{L^2(p_X)} = 1$,

$$A(u) = (1 + o_d)\widehat{A}(u) \qquad \text{for all } u > 0, \tag{190}$$

In words: the distribution of squared target coefficients is asymptotically the same in the true kernel eigenbasis and in the Hermite tensor basis. Under Conjecture 1, the deterministic equivalents in Eq. (7) may be evaluated on $(c_{\boldsymbol{\beta}})$ in place of $(\theta_{\boldsymbol{\beta}})$, with $o_d(1)$ relative error.

## F.1 Single-Index Models

Having formulated Conjecture 1, we can now derive the analytical formulas presented in Section 4. The core idea is to compute the exact decomposition of an index function onto the Hermite tensor basis $\{H_{\boldsymbol{\beta}}\}$. We will start by considering a target function whose form is:

$$f_\star(\boldsymbol{x}) = g(\boldsymbol{v}^\top \boldsymbol{x})$$

where $\boldsymbol{v} \in \mathbb{R}^d$ is the main defining property of the single-index function and we assume the link function $g \in L^2(\mathcal{N}(0,1))$, meaning that:

$$\|g\|^2 = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} g(t)^2 dt < \infty\,.$$

Since $\boldsymbol{x} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$ is the usual diagonal covariance matrix power-law distributed, then it will be convenient to redefine:

$$\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{x}$$

so that $\boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I}_d)$. To ensure that $f_\star \in L^2(p_X) = L^2(\mathcal{N}(0, \Sigma))$, one can easily prove we need to assume:

$$\mathbb{E}[(v^\top x)^2] = 1 \iff \sum_{k=1}^{d} v_k^2 \sigma_k = 1$$

which allows us to write $\|g\|^2_{\mathcal{N}(0,1)} = \|f\|^2_{L^2(p_X)}$. Hence, we can decompose $f_\star$ in the Hermite multi-variate orthonormal basis $H_{\boldsymbol{\beta}}$

$$f_\star(x) = \sum_{\boldsymbol{\beta} \in \mathbb{Z}^d_{\geq 0}} C_{\boldsymbol{\beta}} \, H_{\boldsymbol{\beta}}(z) = \sum_{m=0}^{\infty} \sum_{1 \leq i_1 \leq \cdots \leq i_m \leq d} \theta_{i_1,\dots,i_m} H_{i_1,\dots,i_m}(z) \tag{191}$$

where we reparameterized the $\boldsymbol{\beta}$-summation as explained in Appendix C. Let us now consider for a moment the link function $g(t)$. Since we assumed $g \in L^2(\mathcal{N}(0,1))$, we can decompose $g$ in the 1-dimensional orthonormal Hermite polynomials basis obtaining:

$$g(t) = \sum_{m=0}^{\infty} g_m h_m(t)$$

where $h_m(t) = \frac{1}{\sqrt{m!}} \mathrm{He}_m(t)$. At this point, the target function becomes:

$$f_\star(x) = \sum_{m=0}^{\infty} g_m h_m(\boldsymbol{v}^\top \boldsymbol{x})$$

or, using $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{x}$:

$$f_\star(\boldsymbol{x}) = \sum_{m=0}^{\infty} g_m h_m(\boldsymbol{w}^\top \boldsymbol{z})$$

having defined $\boldsymbol{w} = \boldsymbol{\Sigma}^{1/2} \boldsymbol{v}$ as the transformation of the model weights due to the data anisotropy. To proceed, we will use a well-known addition identity of the Hermite polynomials (Koornwinder et al., 2010, Sec. 18.18(ii)). Since by design $\|\boldsymbol{w}\|^2 = 1$, we have:

$$h_m(\boldsymbol{w}^\top \boldsymbol{z}) = \sum_{|\boldsymbol{\beta}|=m} \sqrt{\binom{m}{\boldsymbol{\beta}}} w_1^{\boldsymbol{\beta}_1} \dots w_d^{\boldsymbol{\beta}_d} H_{\boldsymbol{\beta}}(z)$$

where, for convenience, we will define $\binom{m}{\boldsymbol{\beta}} = \binom{m}{\beta_1 \dots \beta_d} = \frac{m!}{\beta_1! \dots \beta_d!}$. Remembering now that $\boldsymbol{w} = \boldsymbol{\Sigma}^{1/2} \boldsymbol{v}$ and thus $w_j = \sigma_j^{1/2} v_j$, we will have:

$$h_m(\boldsymbol{w}^\top \boldsymbol{z}) = \sum_{|\boldsymbol{\beta}|=m} \sqrt{\binom{m}{\boldsymbol{\beta}}} (\sigma_1^{\frac{1}{2}\beta_1} \dots \sigma_d^{\frac{1}{2}\beta_d})(v_1^{\beta_1} \dots v_d^{\beta_d}) H_{\boldsymbol{\beta}}(\boldsymbol{z})$$

It is now convenient to move to the reparametrization $\boldsymbol{\beta} \to (i_1, \dots, i_m), \beta_j = \sum_k \delta_{i_k, j}$, since we are only considering values of $\boldsymbol{\beta}$ whose total modulus is fixed. We can therefore write:

$$h_m(\boldsymbol{w}^\top \boldsymbol{z}) = \sum_{1 \leq i_1 \leq \cdots \leq i_m \leq d} \sqrt{\binom{m}{\boldsymbol{\beta}}} (\sigma_{i_1}^{1/2} \dots \sigma_{i_m}^{1/2})(v_{i_1} \dots v_{i_m}) H_{i_1,\dots,i_m}(\boldsymbol{z})$$

where $\boldsymbol{\beta}$ must be identified once the parameters $m, i_1, \dots, i_m$ are known. At this point, since, by definition, $\sigma_j = r_\alpha^{-1} j^{-\alpha}$, then we will have:

$$h_m(\boldsymbol{w}^\top \boldsymbol{z}) = \sum_{1 \leq i_1 \leq \cdots \leq i_m \leq d} \sqrt{\binom{m}{\boldsymbol{\beta}}} \, r_\alpha^{-m/2} (i_1 \dots i_m)^{-\alpha/2} (v_{i_1} \dots v_{i_m}) H_{i_1,\dots,i_m}(\boldsymbol{z})$$

Putting everything together, we obtain:

$$f_\star(\boldsymbol{x}) = \sum_{m=0}^{\infty} \sum_{1 \leq i_1 \leq \cdots \leq i_m \leq d} g_m \sqrt{\binom{m}{\boldsymbol{\beta}}} \, r_\alpha^{-m/2} (i_1 \dots i_m)^{-\alpha/2} (v_{i_1} \dots v_{i_m}) H_{i_1,\dots,i_m}(\boldsymbol{z})$$

Comparing with Eq. (191), it is easy to obtain the parameters we are interested in:

$$\theta_{i_1,\dots,i_m} = g_m \sqrt{\binom{m}{\boldsymbol{\beta}}}\, r_\alpha^{-m/2} (i_1 \dots i_m)^{-\alpha/2} (v_{i_1} \dots v_{i_m})$$

In general, if the function is dense enough, then we can neglect the contribution of the diagonal terms approximating asymptotically:

$$\theta_{i_1,\dots,i_m} = g_m \sqrt{m!}\, r_\alpha^{-m/2} (i_1 \dots i_m)^{-\alpha/2} (v_{i_1} \dots v_{i_m}) \tag{192}$$

□

### F.2 Numerical Evidence

Next, we provide numerical evidence supporting the Hermite Ansatz introduced in Conjecture 1. To this end, we proceed in two distinct ways.

**Cumulative Target Function Comparison** First, we numerically compare the cumulative target energies $A(u)$ and $\widehat{A}(u)$, defined as:

$$A(u) = \sum_{\boldsymbol{\beta}:\lambda_{\boldsymbol{\beta}} \le u} \theta_{\boldsymbol{\beta}}^2,$$

$$\widehat{A}(u) = \sum_{\boldsymbol{\beta}:\lambda_{\boldsymbol{\beta}} \le u} c_{\boldsymbol{\beta}}^2,$$

where $\theta_{\boldsymbol{\beta}}$ denotes the projection coefficient of the target function $f_\star$ onto the kernel eigenbasis, and $c_{\boldsymbol{\beta}}$ denotes its coefficient when decomposed along the Hermite orthonormal basis. We focus on Single-Index Models as target functions, as this class allows us to leverage the exact Hermite decomposition given in Eq. (192) (which holds even for finite $d$). In particular, to numerically compute $\widehat{A}(u)$:

- We define a Single-Index Model target function using a random direction $\boldsymbol{v}$ and a fixed Hermite expansion for the link function $g$.
- Using Eq. equation 192, we compute the coefficients $c_{\boldsymbol{\beta}}$ indexed by $\boldsymbol{\beta}$. We then sort these target weights according to the kernel eigenvalues.
- To obtain the kernel eigenvalues, we rely on the asymptotic spectrum expression in Proposition 1. While Eq. (192) is valid for finite $d$, Proposition 1 technically applies in the large-$d$ limit; nonetheless, we demonstrate that it provides a remarkably good approximation to the finite-$d$ spectrum.

To build $A(u)$, we rely on the *empirical diagonalization* procedure. This Monte Carlo-inspired technique allows us to obtain an estimate of both the kernel spectrum and the target coefficients $\theta_{\boldsymbol{\beta}}$ on the kernel eigenbasis. The comparison is illustrated in Fig. 10 for different values of $\alpha$, providing strong numerical evidence for the validity of our conjecture in both the weakly and strongly anisotropic regime.

**Learning SIMs** As a second test, we can leverage the Hermite Ansatz to evaluate the theoretical predictions from Sec. 4, which are valid if we assume that the kernel eigenbasis coincides with the Hermite orthonormal basis in which the single-index target function is naturally decomposed. To do so, we empirically consider a KRR learning task trained over an anisotropic dataset and extract the empirical bias (setting $\sigma_\varepsilon = 0$ to eliminate the variance component).

We first consider the case of a sparse single-index model with $j = 1$. The empirical versus asymptotic results are illustrated in Fig. 11 for both the weakly anisotropic regime ($\alpha = 0.35 < 1$) and the strongly anisotropic regime ($\alpha = 1.5 > 1$). The asymptotic predictions are obtained considering the formal limit $\omega \to \infty$ and are in good agreement with the numerical findings.

We then consider the Single-Index power-law decaying case. In Fig. 12, we examine the weakly anisotropic regime ($\alpha < 1$). Since the effective parameter is defined as $\omega = \frac{1}{2}(\alpha + \gamma)$, depending on

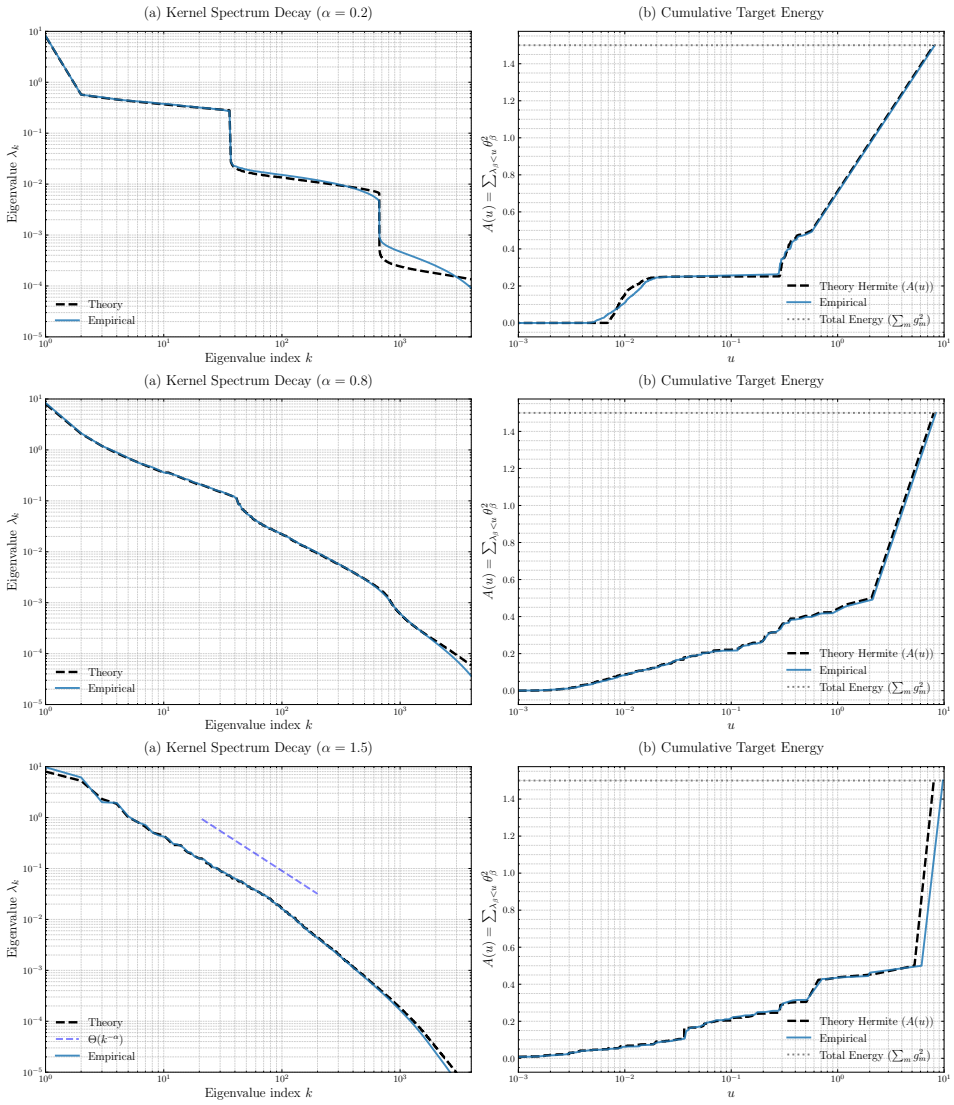


Figure 10: *Kernel spectrum (left panel) and cumulative target function comparison (right panel) for different values of $\alpha$. In the left panel, we observe that the asymptotic theoretical spectrum obtained from Proposition 1 closely matches the spectrum empirically extracted via empirical diagonalization. In the right panel, we see that the cumulative curves $A(u)$ computed on the Hermite basis and on the kernel eigenbasis are in close agreement. Simulation parameter: $d = 35$, polynomial kernel $k(\boldsymbol{x}, \boldsymbol{x}') = (2 + \langle \boldsymbol{x}, \boldsymbol{x}' \rangle)^3$, random index $\boldsymbol{v}$. The empirical diagonalization was performed using $M = 10000$ Monte-Carlo samples*

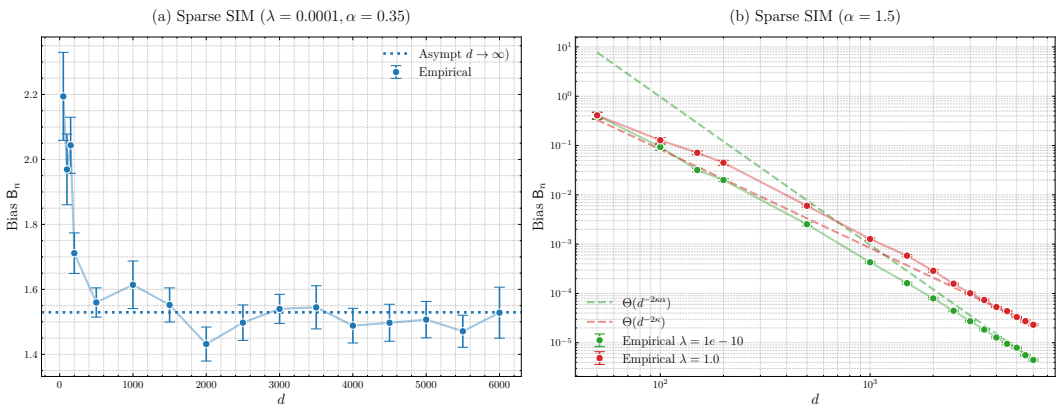


Figure 11: ***Left:*** *Asymptotic versus empirical bias for a sparse SIM in the weakly anisotropic regime.* ***Right:*** *Asymptotic versus empirical bias for a sparse SIM in the strongly anisotropic regime. Simulation parameters: exponential kernel truncated at the fourth order,* $\kappa = 1$, $\psi = 0.8$. *Each empirical KRR run is repeated* 20 *times and averaged, with errors representing the standard deviations.*

$\gamma$ we can fall into either the $\omega < \frac{1}{2}$ or $\omega > \frac{1}{2}$ regime. When $\gamma = 0.2$ (left plot), we have $\omega < \frac{1}{2}$, and the kernel learns up to the $\kappa$-th feature of the target function. When $\gamma = 1$ or $\gamma = 2$ (right plot), we have $\omega > \frac{1}{2}$. In this case, we expect the kernel to learn up to the $\frac{\kappa}{1-\alpha} = \frac{1}{1-0.8} = 5$-th order features of the target function. In our simulation, the link function was set to $g(t) = \sum_{m=0}^{3} \frac{1}{m!} g_m h_m(t)$ (meaning $g_5 = 0$), so we expect the kernel to learn the entire function, and the bias to decay to zero. Indeed, we observe that for both values of $\gamma$, the bias drops to zero (note that we cannot observe the power-law decay here, as it corresponds to a sub-leading correction, whereas we are only tracking the leading term).

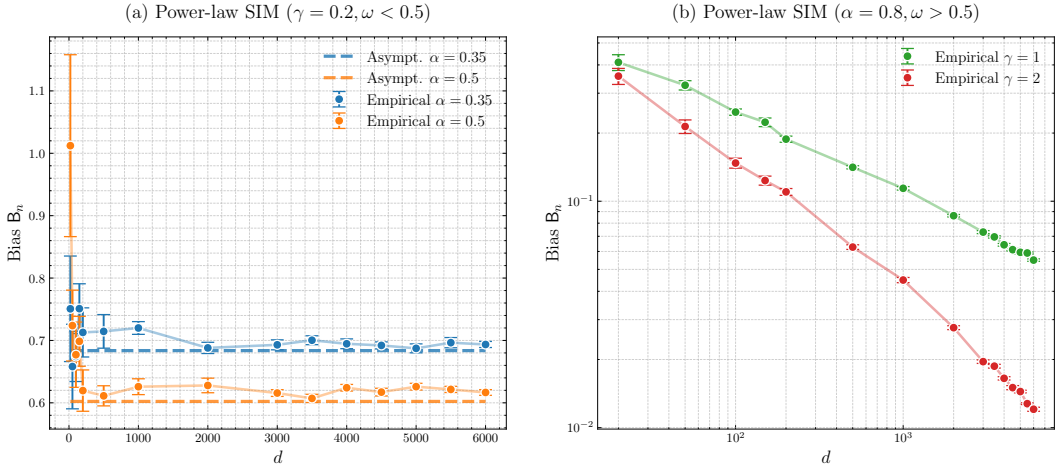


Figure 12: ***Left:*** *Asymptotic versus empirical bias for a power-law SIM in the weakly anisotropic regime with an effective* $\omega < 0.5$. ***Right:*** *Asymptotic versus empirical bias for a power-law SIM in the weakly anisotropic regime for an effective* $\omega > 0.5$. *The link function* $g$ *is chosen such that its expansion is truncated at degree* 3 *(*$g(t) = \sum_{m=0}^{3} g_m h_m(t)$*), allowing the kernel to fully learn it in the* $\omega > \frac{1}{2}$ *regime. All other parameters match those of Fig. 11.*

Finally, we analyze the strongly anisotropic regime (Fig. 13). Here, provided that $\gamma > 0$, we always operate in the fast-decay $\omega > \frac{1}{2}$ regime, meaning we should observe a scaling behavior of the bias, converging to zero. We consider two distinct scenarios: one where $\omega > \alpha + \frac{1}{2}$ and another where

$\omega < \alpha + \frac{1}{2}$, allowing us to verify the different scaling behaviors predicted by Theorem 4, for $\lambda \to 0$ and $\lambda = 1$. Once again, we obtain visually compelling evidence supporting our conjecture.

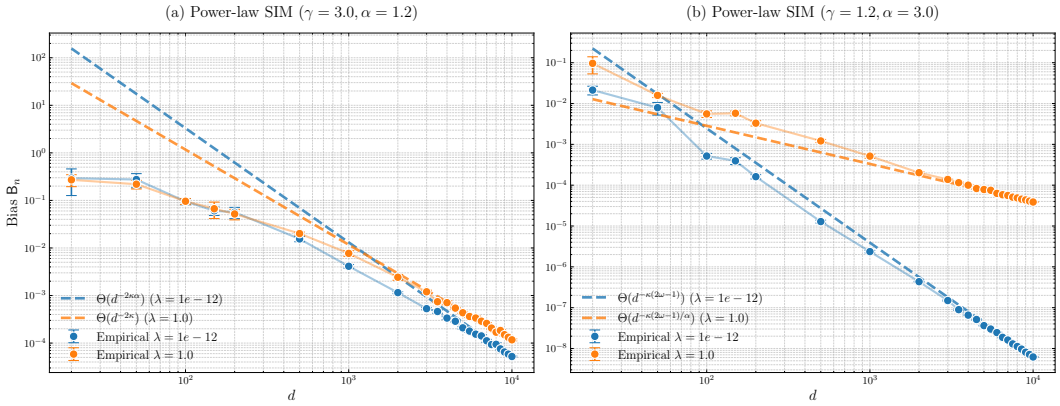


Figure 13: ***Left:*** *Asymptotic versus empirical bias for a power-law SIM in the strongly anisotropic regime when $\omega > \alpha + \frac{1}{2}$.* ***Right:*** *Asymptotic versus empirical bias for a power-law SIM in the strongly anisotropic regime when $\omega < \alpha + \frac{1}{2}$. All other parameters match those of Fig. 11.*

## G Technical and Auxilliary Lemmas

**Lemma G.1.** Let $\ell \in \mathbb{N}$. Then there exists a constant $C_\ell$, independent of $d$, such that

$$\left\| \|x\|^2 \right\|_{L^\ell} = \mathbb{E}_{x \sim \mathcal{N}(0,\Sigma)} \left[ \left( \|x\|^{2\ell} \right)^{\frac{1}{\ell}} \right] \leq C_\ell.$$

*Proof.* Indeed, we have:

$$\mathbb{E}_{x \sim \mathcal{N}(0,\Sigma)} \left[ \|x\|^{2\ell} \right]^{\frac{1}{\ell}} = \mathbb{E}_{z \sim \mathcal{N}(0,I_d)} \left[ \left( \sum_{j=1}^{d} \sigma_j z_j^2 \right)^{\ell} \right]^{\frac{1}{\ell}} \tag{193}$$

$$\leq \sum_{j=1}^{d} \sigma_j \mathbb{E}_{z \sim \mathcal{N}(0,1)} \left[ z_j^{2\ell} \right]^{\frac{1}{\ell}} \leq C_\ell \operatorname{Tr}(\Sigma) = C_\ell, \tag{194}$$

where in the second line we use the triangular inequality for the $L^\ell$- norm. □

**Lemma G.2.** Let $m \in \mathbb{N}$, and denote:

$$F_m(x) = K_{>m}(x,x), \quad \text{and } \operatorname{Tr}(K_{>m}) = \sum_{j \geq m+1} \lambda_j.$$

Then

$$\mathbb{E}\left[ F_m(x)^2 \right]^{\frac{1}{2}} \leq C_D \operatorname{Tr}(K_{>m}), \quad \mathbb{E}\left[ (F_m - \operatorname{Tr}(K_{>m}))^2 \right]^{\frac{1}{2}} \leq (C_D + 1) \operatorname{Tr}\left( K_{>m} \right). \tag{195}$$

*Proof.* For the first inequality, we compute directly:

$$\mathbb{E}\left[F_m(x)^2\right] = \mathbb{E}\left[\left(\sum_{j>m} \lambda_j \Phi_j(x)^2\right)^2\right] \tag{196}$$

$$= \sum_{j,\ell \geq m} \lambda_j \lambda_\ell \mathbb{E}\left[\Phi_j(x)^2 \Phi_\ell(x)^2\right] \tag{197}$$

$$\leq \sum_{j,\ell \geq m} \lambda_j \lambda_\ell \mathbb{E}\left[\Phi_j(x)^4\right]^{\frac{1}{2}} \mathbb{E}\left[\Phi_\ell(x)^4\right]^{\frac{1}{2}}, \tag{198}$$

by Cauchy-Schwarz. Since the eigenfunctions are polynomials, hypercontractivity gives:

$$\mathbb{E}\left[F_m(x)^2\right] \leq C_D(\sum_{j,\ell \geq m} \lambda_j)^2 = C_D \operatorname{Tr}(K_{>m})^2. \tag{199}$$

Taking the square-root, we conclude the first inequality. For the second inequality, we have:

$$\mathbb{E}\left[(F_m - \operatorname{Tr}(K_{>m}))^2\right]^{\frac{1}{2}} \leq \mathbb{E}\left[F_m^2\right]^{\frac{1}{2}} + \operatorname{Tr}(K_{>m}), \tag{200}$$

and applying the first inequality we conclude:

$$\mathbb{E}\left[(F_m - \operatorname{Tr}(K_{>m}))^2\right]^{\frac{1}{2}} \leq (C_D + 1) \operatorname{Tr}(K_{>m}). \tag{201}$$

□

**Lemma G.3** (Gaussian Poincaré Inequality, (Boucheron et al. (2013), Theorem 3.20)**.** Let $x \sim \mathcal{N}(0, \Sigma)$, and let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. Then:

$$\operatorname{Var}(f(x)) \leq \|\Sigma\|_{\mathrm{op}} \mathbb{E}\|\nabla f(x)\|_2^2.$$