
Directional Complexity of SIGReg with Anisotropic Gaussian Targets

Ahmad Raza
Independent Researcher

ahmadraza.res@gmail.com

Abstract

Sketched Isotropic Gaussian Regularization (SIGReg) aligns learned representations with an isotropic Gaussian target along random one-dimensional projections. LeJEPA establishes finite-direction guarantees because the rotational symmetry of this target makes every direction identical. Structured representations, however, require anisotropic targets that vary across directions. We prove that while target anisotropy preserves the geometric approximation rate, directional complexity depends on the target’s centered spectral moments and spectral endpoints. We show empirically that across targets with matched condition numbers, spectral shape governs direction budgets. These budgets scale with the target’s spherical Dirichlet energy and initial slicing-gradient variance, while its leading angular harmonic couples directly to directional sampling error. To eliminate this quadratic coupling, we introduce Orthogonal Directional Slicing (ODS), which draws complete random orthonormal frames that enforce exact second-moment balance. While ODS reduces gradient variance across all conditions, it does not reliably lower direction budgets, as higher-degree angular harmonics dominate directional complexity. Reducing this complexity therefore requires canceling these higher-order harmonics, rather than simply drawing more directions.

1 Introduction

Projection-based distribution matching replaces a high-dimensional comparison with an average of one-dimensional discrepancies along random directions. In LeJEPA, SIGReg uses this approach to match embeddings to an isotropic Gaussian target $\mathcal{N}(0, I_K)$ (Balestriero & LeCun, 2025). Because an isotropic target has a direction-independent characteristic function, it contributes no angular structure over the sphere. We study anisotropic Gaussian targets $\mathcal{N}(0, \Sigma)$ as a tractable extension: their one-dimensional projections remain available in closed form, while the target varies over the sphere through the Rayleigh quotient $a^\top \Sigma a$.

Within the spherical Sobolev framework used to analyze SIGReg, we show that anisotropy changes the target-dependent factor without changing the geometric approximation exponent. Under $\text{tr}(\Sigma) = K$, the centered quadratic form $q_\Sigma(a) = a^\top (\Sigma - I_K) a$ acts as a degree-two spherical harmonic, while its exponential generates higher even degrees. The resulting H^2 bound depends on centered spectral moments and the minimal eigenvalue (with both spectral endpoints entering through Proposition 4), showing that condition number alone cannot determine this complexity. Because this guarantee holds for deterministic quasi-uniform interpolation, it leaves an empirical question: whether targets with matched condition numbers require different direction budgets under stochastic optimization.

These same spectral quantities govern the target’s integrated tangential-gradient energy and the variance of the initial slicing gradient. Furthermore, we isolate a degree-two finite-sample interaction between the traceless target covariance and the empirical second moment of the sampled directions. This interaction motivates ODS, which forms direction batches from complete random orthonormal frames. Complete frames preserve the population SIGReg objective and cancel the degree-two interaction exactly when $M \equiv 0 \pmod{K}$.

2 Background and related work

2.1 SIGReg and directional distribution matching

Let P and Q be probability distributions on $\mathbb{R}^K$. By the Cramér–Wold theorem, $P = Q$ if and only if $P^{(a)} = Q^{(a)}$ for every direction $a \in \mathbb{S}^{K-1}$ (Cramér & Wold, 1936). Sliced methods use this property to replace an intractable high dimensional comparison with an average of one-dimensional discrepancies (Rabin et al., 2012; Bonneel et al., 2015). For an embedding matrix $Z = \{z_n\}_{n=1}^N \subset \mathbb{R}^K$, SIGReg compares the empirical characteristic function $\widehat{\varphi}_{Z,a}(t) = \frac{1}{N} \sum_{n=1}^N \exp(it a^\top z_n)$ with the target characteristic function $\varphi_{Q,a}(t) = \mathbb{E}_{X \sim Q}[\exp(it a^\top X)]$ through the directional discrepancy

$$\ell_Q(a; Z) = N \int_{\mathbb{R}} |\widehat{\varphi}_{Z,a}(t) - \varphi_{Q,a}(t)|^2 w(t) dt, \quad (1)$$

where $w(t)$ is a symmetric integrable weight. Given directions $\mathcal{A}_M = \{a_m\}_{m=1}^M$, the finite-direction objective is

$$\widehat{\mathcal{L}}_M(Z; Q) = \frac{1}{M} \sum_{m=1}^M \ell_Q(a_m; Z). \quad (2)$$

Standard SIGReg specifies an isotropic target $Q = \mathcal{N}(0, I_K)$, whose characteristic function $\varphi_{Q,a}(t) = e^{-t^2/2}$ is angularly constant (Balestrieri & LeCun, 2025). An anisotropic target $Q_\Sigma = \mathcal{N}(0, \Sigma)$, however, introduces direction-dependent variation into $\ell_Q(a; Z)$ through the target covariance Σ .

2.2 Finite-direction approximation on the sphere

Let $\mathbb{S}^{K-1}$ carry normalized surface measure. At fixed frequency t , define the discrepancy $g_t(a) = \varphi_{P,a}(t) - \varphi_{Q,a}(t)$, and let $I_M g_t$ be its spherical interpolant on M quasi-uniform directions. If $g_t \in H^\alpha(\mathbb{S}^{K-1})$, spherical polynomial interpolation (Balestrieri & LeCun, 2025) with spatial scaling $M \asymp L^{K-1}$ gives

$$\|g_t - I_M g_t\|_{L^2(\mathbb{S}^{K-1})}^2 \leq C(K, \alpha) M^{-2\alpha/(K-1)} \|g_t\|_{H^\alpha(\mathbb{S}^{K-1})}^2. \quad (3)$$

where $C(K, \alpha) > 0$ is the interpolation constant depending only on dimension K and Sobolev order α . Integrating over frequency with weight $w(t)$ yields the total interpolation error $\mathcal{E}_M(P, Q)$ and Sobolev constant $B_\alpha(P, Q)$:

$$\mathcal{E}_M(P, Q) = \int_{\mathbb{R}} w(t) \|g_t - I_M g_t\|_{L^2(\mathbb{S}^{K-1})}^2 dt, \quad (4)$$

$$B_\alpha(P, Q) = \int_{\mathbb{R}} w(t) \|g_t\|_{H^\alpha(\mathbb{S}^{K-1})}^2 dt. \quad (5)$$

Applying equation 3 gives the integrated guarantee

$$\mathcal{E}_M(P, Q) \leq C(K, \alpha) M^{-2\alpha/(K-1)} B_\alpha(P, Q). \quad (6)$$

If the discrepancy vanishes at all sampled directions, the zero interpolant fits the data, and equation 6 bounds the residual error across the unsampled sphere. This guarantee characterizes deterministic interpolation on quasi-uniform point sets. It does not model stochastic optimization where directions are freshly resampled at each step. The sufficient count derived from equation 6 is therefore conceptually distinct from the empirical optimization threshold M^* .

2.3 Finite-direction sampling and structured targets

Prior work has explored structured direction sets for sliced objectives: quasi-Monte Carlo methods improve spherical coverage (Nguyen et al., 2024), orthogonal coupling reduces projection redundancy (Rowland et al., 2019), and Markovian or hierarchical formulations introduce structured dependence (Nguyen et al., 2023a;b; Sun et al., 2024). Adaptive sliced methods instead reweight directions during training (Zhang et al., 2025), altering the population objective, whereas ODS preserves the uniform population objective. While spherical-harmonic control variates reduce finite-direction variance by regressing on zero-mean spherical harmonics (Leluc et al., 2024), ODS eliminates the degree-two component algebraically through exact second-moment matching, requiring no control coefficients. While complete orthogonal frames have been used in other sliced settings (Rowland et al., 2019; Petrovic et al., 2026), our contribution is to derive the explicit degree-two interaction between sampled frames and an anisotropic Gaussian target, and to prove that complete frames cancel this interaction identically. Recent JEPA literature also studies targets beyond the isotropic Gaussian, modifying target geometry or predictive couplings (Alvarez, 2026; Nicollier et al., 2026a;b). These approaches motivate structured targets, but they do not characterize how target anisotropy alters the finite-direction complexity of SIGReg or interacts with sampled directions.

3 Finite-direction approximation for anisotropic targets

We now consider the anisotropic Gaussian target $Q_\Sigma = \mathcal{N}(0, \Sigma)$, where $\Sigma \succ 0$ and $\text{tr}(\Sigma) = K$. The trace constraint fixes the average target variance and separates anisotropy from scale. For a unit direction $a \in \mathbb{S}^{K-1}$, its directional characteristic function is

$$T_{\Sigma,t}(a) := \varphi_{Q_\Sigma,a}(t) = \exp\left(-\frac{t^2}{2} a^\top \Sigma a\right).$$

Let $w_\rho(t) = e^{-\rho^2 t^2}$ with $\rho > 0$, write $\Sigma_0 = \Sigma - I_K$, $S_j = \text{tr}(\Sigma_0^j)$, and set $\gamma = \rho^2 + \lambda_{\min}(\Sigma)$. Using the spectral H^2 convention

$$\|f\|_{H^2}^2 = \|f\|_{L^2}^2 + 2\|\nabla_{\mathbb{S}} f\|_{L^2}^2 + \|\Delta_{\mathbb{S}} f\|_{L^2}^2,$$

define

$$\begin{aligned} B_{2,P} &= \int_{\mathbb{R}} w_\rho(t) \|\varphi_{P,\cdot}(t)\|_{H^2(\mathbb{S}^{K-1})}^2 dt, \\ B_{2,Q}(\Sigma) &= \int_{\mathbb{R}} w_\rho(t) \|T_{\Sigma,t}\|_{H^2(\mathbb{S}^{K-1})}^2 dt. \end{aligned}$$

While P is treated here as a fixed distribution, in optimization it corresponds to the empirical embedding distribution $P_Z = N^{-1} \sum_{n=1}^N \delta_{z_n}$, which evolves with Z . The following result extends equation 6 at Sobolev order two by bounding the target contribution explicitly.

Theorem 1 (Anisotropic Finite-Direction Bound). *Let $\Sigma \succ 0$ satisfy $\text{tr}(\Sigma) = K$, suppose $B_{2,P} < \infty$, and let I_M be constructed from M quasi-uniform directions as in Section 2.2. Then $B_{2,Q}(\Sigma) \leq B_{2,Q}^{\text{base}}(\Sigma)$, where*

$$B_{2,Q}^{\text{base}}(\Sigma) = \frac{\sqrt{\pi}}{\gamma^{1/2}} + \frac{3\sqrt{\pi}(2K+1)}{2(K+2)\gamma^{5/2}} S_2 + \frac{105\sqrt{\pi}}{8K(K+2)\gamma^{9/2}} (S_2^2 + 2S_4). \quad (7)$$

Consequently,

$$\mathcal{E}_M(P, Q_\Sigma) \leq 2C(K, 2)M^{-4/(K-1)} [B_{2,P} + B_{2,Q}^{\text{base}}(\Sigma)]. \quad (8)$$

In particular, $\mathcal{E}_M(P, Q_\Sigma) \leq \varepsilon$ is guaranteed whenever

$$M \geq \left\lceil \frac{2C(K, 2) (B_{2,P} + B_{2,Q}^{\text{base}}(\Sigma))}{\varepsilon} \right\rceil^{(K-1)/4}. \quad (9)$$

Lemma 2 (Empirical Embedding Sobolev Bound). *Let $P_Z = N^{-1} \sum_{n=1}^N \delta_{z_n}$ with $\max_n \|z_n\|_2 \leq R$. Then*

$$B_{2,P_Z} \leq \frac{\sqrt{\pi}}{\rho} + \frac{\sqrt{\pi}[(K-1)^2 + 2]R^2}{2\rho^3} + \frac{2(K-1)R^3}{\rho^4} + \frac{3\sqrt{\pi}R^4}{4\rho^5} < \infty. \quad (10)$$

The finiteness condition $B_{2,P} < \infty$ therefore holds at every finite empirical embedding state; uniform control along an optimization trajectory requires a uniform bound on $\max_n \|z_n\|_2$. The proof is given in Appendix A.2. Theorem 1 shows that anisotropy enters exclusively through the target factor, leaving the geometric rate $M^{-4/(K-1)}$ unchanged. Within $B_{2,Q}^{\text{base}}(\Sigma)$, the second moment S_2 controls tangential derivatives, $S_2^2 + 2S_4$ controls Laplacian curvature, and the minimal eigenvalue $\lambda_{\min}(\Sigma)$ governs frequency damping through $\gamma = \rho^2 + \lambda_{\min}(\Sigma)$. As established in Section 2.2, equation 9 provides a sufficient direction budget for deterministic quasi-uniform interpolation, rather than the stochastic optimization threshold M^* studied in Section 4.

3.1 Angular structure of the target

Because $\text{tr}(\Sigma_0) = 0$, the centered quadratic form $q_\Sigma(a) = a^\top \Sigma_0 a$ isolates the angular variation of the target:

$$T_{\Sigma,t}(a) = e^{-t^2/2} \exp\left(-\frac{t^2}{2} q_\Sigma(a)\right). \quad (11)$$

Proposition 3 (Centered Covariance Harmonic). *The centered quadratic form q_Σ is a spherical harmonic of degree two. In particular,*

$$-\Delta_{\mathbb{S}} q_\Sigma = 2K q_\Sigma, \quad \|q_\Sigma\|_{L^2(\mathbb{S}^{K-1})}^2 = \frac{2 \text{tr}(\Sigma_0^2)}{K(K+2)}. \quad (12)$$

Proof. The quadratic polynomial $x \mapsto x^\top \Sigma_0 x$ satisfies $\Delta(x^\top \Sigma_0 x) = 2 \text{tr}(\Sigma_0) = 0$. Its restriction to the sphere is therefore a harmonic polynomial of degree two, with Laplace–Beltrami eigenvalue $2(2 + K - 2) = 2K$. Integrating $(a^\top \Sigma_0 a)^2$ against normalized surface measure via the spherical fourth-moment identity gives

$$\mathbb{E}_a[(a^\top \Sigma_0 a)^2] = \frac{\text{tr}(\Sigma_0)^2 + 2 \text{tr}(\Sigma_0^2)}{K(K+2)} = \frac{2 \text{tr}(\Sigma_0^2)}{K(K+2)},$$

which proves equation 12. □

Because $q_\Sigma(-a) = q_\Sigma(a)$, the target characteristic function is even in a , so its spherical-harmonic expansion contains only even degrees:

$$T_{\Sigma,t} = T_{\Sigma,t}^{(0)} + T_{\Sigma,t}^{(2)} + T_{\Sigma,t}^{(4)} + \dots \quad (13)$$

The linear term in the Taylor expansion of $\exp(-t^2 q_\Sigma/2)$ produces the leading degree-two harmonic, while higher powers generate higher even degrees. Degree two therefore dominates in the weak-anisotropy regime, but moderate to severe anisotropy excites higher-order angular harmonics.

3.2 Target-dependent Sobolev complexity

The target bound in Theorem 1 is derived directly from the first two tangential derivatives of $T_{\Sigma,t}$. Let $P_a = I_K - aa^\top$ be the tangent-space projector and define $s_\Sigma(a) = a^\top \Sigma_0^2 a - q_\Sigma(a)^2 = \|P_a \Sigma_0 a\|_2^2$. Differentiation yields

$$\nabla_{\mathbb{S}} T_{\Sigma,t}(a) = -t^2 T_{\Sigma,t}(a) P_a \Sigma_0 a, \quad (14)$$

$$\Delta_{\mathbb{S}} T_{\Sigma,t}(a) = T_{\Sigma,t}(a) (t^2 K q_\Sigma(a) + t^4 s_\Sigma(a)). \quad (15)$$

Substituting these identities into the H^2 norm, majorizing $T_{\Sigma,t}(a)^2 \leq e^{-\lambda_{\min}(\Sigma)t^2}$, and integrating over frequency proves equation 7 (detailed in Appendix A.1). The baseline bound $B_{2,Q}^{\text{base}}$ replaces the angular exponential by its spectral maximum $e^{-\nu_- t^2}$ and splits the square in equation 15.

Let $\nu_1, \dots, \nu_K$ be the eigenvalues of Σ_0 with $\nu_- = \min_j \nu_j$ and $\nu_+ = \max_j \nu_j$. We obtain a strictly sharper bound by preserving exponential variation across the Rayleigh-quotient spectrum $[\nu_-, \nu_+]$ and the Laplacian cross term. For $\psi_\Sigma(\theta) = \mathbb{E}_a[e^{-\theta q_\Sigma(a)}]$, Gaussian radial decomposition yields the exact identity

$$\mathbb{E}_a \left[(1 + 2\lambda q_\Sigma(a))^{-K/2} \right] = \mathbb{E}_{z \sim \mathcal{N}(0, I_K)} \left[e^{-\lambda z^\top \Sigma_0 z} \right] = \prod_{j=1}^K (1 + 2\lambda \nu_j)^{-1/2},$$

which upper bounds $\psi_\Sigma(\theta)$ at $\lambda = \theta/K$, alongside the Rayleigh bound $e^{-\nu_- \theta}$.

Proposition 4 (Secant Sobolev Refinement). *Let $B_{2,Q}^{\text{sec}}(\Sigma)$ be the bound obtained by majorizing $q \mapsto e^{-t^2 q}$ by its endpoint secant on $[\nu_-, \nu_+]$ while retaining the Laplacian cross term. Then*

$$B_{2,Q}(\Sigma) \leq B_{2,Q}^{\text{sec}}(\Sigma) \leq B_{2,Q}^{\text{base}}(\Sigma). \quad (16)$$

This bound depends explicitly on both spectral endpoints (ν_-, ν_+) and centered moments $S_2, \dots, S_5$. The second inequality is strict unless $\Sigma = I_K$.

The closed form, secant optimality, and moment algebra of $B_{2,Q}^{\text{sec}}(\Sigma)$ are given in Appendix A.3. While $B_{2,Q}^{\text{sec}}(\Sigma)$ sharpens the target-side bound, any target-only factor is inherently incomplete: bounding embedding and target norms separately discards their interaction.

Lemma 5 (Insufficiency of Component Norms). *In a nontrivial Hilbert space, no function of $\|u\|$ and $\|v\|$ alone can equal $\|u - v\|^2$ for every pair (u, v) .*

Proof. For any nonzero u , the pairs (u, u) and $(u, -u)$ have identical component norms, whereas their squared differences are 0 and $4\|u\|^2$. $\square$

Under the complex H^2 inner product $\langle f, g \rangle_{H^2} = \langle f, g \rangle_{L^2} + 2\langle \nabla_{\mathbb{S}} f, \nabla_{\mathbb{S}} g \rangle_{L^2} + \langle \Delta_{\mathbb{S}} f, \Delta_{\mathbb{S}} g \rangle_{L^2}$, define the joint coupling

$$C_2(P, \Sigma) = \int_{\mathbb{R}} w_\rho(t) \langle \varphi_{P,\cdot}(t), T_{\Sigma,t} \rangle_{H^2} dt.$$

The exact Sobolev factor is then

$$B_2(P, Q_\Sigma) = B_{2,P} + B_{2,Q}(\Sigma) - 2 \operatorname{Re} C_2(P, \Sigma). \quad (17)$$

Target-side factors $B_{2,Q}(\Sigma)$ are therefore sufficient rather than tight: the cross term $C_2(P, \Sigma)$ captures joint embedding-target geometry that no descriptor of the target spectrum alone can reflect.

4 The spectral structure of directional complexity

Section 3 shows that anisotropy enters the finite-direction bound through spectral endpoints and centered moments, allowing covariances with the same condition number to have distinct theoretical bounds. Because this bound does not predict optimization dynamics directly, we test whether targets with matched condition numbers require different direction budgets M^* in practice.

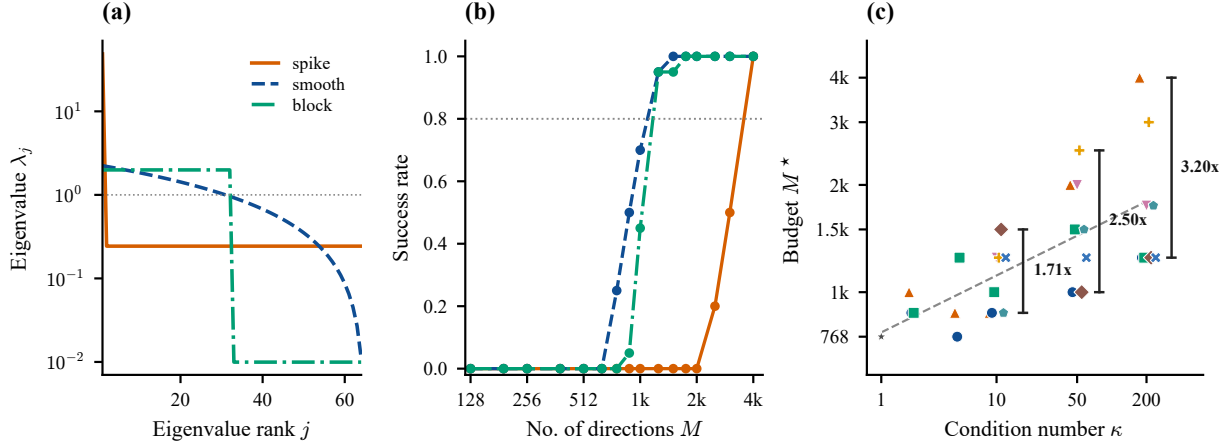


Figure 1: Spectral shape separates finite-direction requirements at matched condition number. (a) Eigenvalue profiles λ_j by rank j for spike, smooth, and block targets at $\kappa = 200$ and $\text{tr}(\Sigma) = 64$; the dotted line marks the isotropic baseline $\lambda_j = 1$. (b) Success rate $p_\Sigma(M)$ across direction budget M on the $\kappa = 200$ triad. The dotted line marks the 80% success threshold defining M^* . (c) Empirical direction requirements M^* across all 31 benchmark targets against condition number κ . Vertical brackets mark the spread between minimum and maximum M^* within matched condition-number groups.

4.1 Experimental setup

We use a direct-embedding experiment to isolate the finite-direction objective from encoder architecture and data augmentation, following the controlled analysis of Balestrierio & LeCun (2025). Each run optimizes $N = 1000$ embeddings in $\mathbb{R}^{64}$ toward $Q_\Sigma = \mathcal{N}(0, \Sigma)$ under trace normalization $\text{tr}(\Sigma) = 64$, drawing a fresh batch of M uniform spherical directions at every step. Across conditions, only the target spectrum changes. The benchmark comprises 31 covariances spanning eight anisotropic spectral families and an isotropic anchor, with shared condition numbers across families to isolate the effect of spectral shape.

Let $E_M^{(s)}(\Sigma)$ denote the discrepancy of seed s on evaluation directions drawn independently of training, and $E_{\text{ref}}(\Sigma)$ be the median reference discrepancy at $M_{\text{ref}} = 16,384$. We define the empirical success frequency $p_\Sigma(M)$ and direction requirement $M^*(\Sigma)$ by

$$p_\Sigma(M) = \frac{1}{n_\Sigma} \sum_{s=1}^{n_\Sigma} \mathbf{1}\left\{E_M^{(s)}(\Sigma) \leq 1.10 E_{\text{ref}}(\Sigma)\right\}, \quad (18)$$

$$M^*(\Sigma) = \min \{M \in \mathcal{M} : p_\Sigma(M) \geq 0.80\},$$

where $\mathcal{M}$ is the discrete grid of direction budgets given in Appendix B.

4.2 Matched-spectrum variation

Targets with the same condition number require markedly different direction budgets. Across the eight spectral families, the ratio between the maximum and minimum M^* widens from 1.71 at $\kappa = 10$ to 3.20 at $\kappa = 200$ (Table 1; Figure 1c).

This divergence is sharpest at $\kappa = 200$ (Figure 1a,b): the spike target requires $M^* = 4096$, while smooth and block require only $M^* = 1280$. For individual seeds, the threshold medians across 20 runs are 3072, 896, and 1280 for spike, smooth, and block, respectively. Every observed seed threshold for spike exceeds those for smooth and block (uncorrected Mann–Whitney $p < 0.001$, Cliff’s $\delta = 1$ for both comparisons); smooth and block also differ significantly ($p = 0.0008$, $\delta = -0.598$). This separation persists under stricter (5%) and

Table 1: Empirical direction requirements across spectral families at matched condition number.

κ	Min M^*	Max M^*	Ratio	Min; Max families
10	896	1536	1.71	smooth/spike/square-root decay; linear
50	1024	2560	2.50	smooth/linear; trispike
200	1280	4096	3.20	smooth/block/linear/cosine; spike

more lenient (20%) success tolerances (Appendix B.5). Condition number therefore fails to determine M^* : spectral shape retains critical information discarded by this scalar ratio.

4.3 Spectral predictors of directional complexity

We next evaluate how closely the spectral descriptors from Section 3 track M^* , using Spearman rank correlations across all 31 targets (Table 2).

Table 2: Rank correlation between spectral descriptors and empirical direction requirements across 31 targets.

Descriptor	Spearman ρ	Fisher- z 95% interval
Condition number κ	0.657	[0.39, 0.82]
$S_2 = \text{tr}(\Sigma_0^2)$	0.791	[0.60, 0.90]
$S_4 = \text{tr}(\Sigma_0^4)$	0.665	[0.40, 0.83]
$B_{2,Q}^{\text{base}}(\Sigma)$	0.877	[0.76, 0.94]
$B_{2,Q}^{\text{sec}}(\Sigma)$	0.834	[0.68, 0.92]

The centered second moment S_2 correlates more strongly with M^* ($\rho = 0.791$) than condition number ($\rho = 0.657$), while the target-side Sobolev bounds achieve rank correlations between 0.834 and 0.877. Because their CIs overlap, the benchmark does not establish a strict empirical ranking among the Sobolev descriptors. In particular, the tighter secant-majorant bound need not yield the highest in-sample rank correlation: analytic sharpness and rank association with an empirical optimization threshold are distinct mathematical properties. These correlations confirm that the spectral moments appearing in the Sobolev bound track finite-direction difficulty. However, consistent with Lemma 5, no target-only descriptor predicts M^* exactly because separated bounds discard joint embedding-target cross-terms. We retain $B_{2,Q}^{\text{sec}}$ for out-of-family generalization because it represents the sharper theoretical construction.

4.4 Generalization to held-out spectra

To test whether the target-side bound generalizes beyond the benchmark families, we fit an OLS log-linear model across the 31 targets, yielding $\log M^* = 0.152 \log B_{2,Q}^{\text{sec}} + 6.563$ ($R^2 = 0.816$), and use it to predict M^* on six held-out spectra (Table 3).

Table 3: Predicted and observed direction requirements for six held-out spectra.

Family	κ	$B_{2,Q}^{\text{sec}}$	Predicted M^*	Observed M^*
Exponential	50	1.251×10^2	1474	896
Step	50	3.169×10^2	1697	1536
Log-normal	50	2.185×10^2	1604	1280
Exponential	200	2.643×10^2	1651	1536
Step	200	4.210×10^2	1772	1792
Log-normal	200	2.386×10^2	1625	1024

The predictions correctly rank the held-out targets within each condition number (Table 3), with observed requirements ranging from 896 to 1536 at $\kappa = 50$ and from 1024 to 1792 at $\kappa = 200$, confirming that

matched-condition variation generalizes to unseen spectral shapes. While the endpoint-secant bound remains informative outside the training families, it is not an exact sample-complexity law. The fitted exponent is sublinear (0.152), and noticeable residuals remain for the exponential target at $\kappa = 50$ and log-normal target at $\kappa = 200$. This behavior is consistent with a sufficient target-side bound that bounds worst-case spherical discrepancy rather than fully characterizing stochastic optimization dynamics.

5 Angular structure and directional sampling

While M^* correlates with target-side Sobolev complexity, this does not specify how target geometry affects optimization. We show that the same spectral structure appears in the target’s angular derivatives, the variance of stochastic directional gradients, and the empirical second-moment error of sampled directions. These quantities provide convergent evidence for a shared mechanism, without claiming any single one as the sole cause of variation in M^* .

5.1 Angular variation of the target

Define the tangential energy density $s_\Sigma(a) = a^\top \Sigma_0^2 a - q_\Sigma(a)^2$. From equation 14, the integrated angular Dirichlet energy of the target is

$$\mathcal{C}_1(\Sigma) = \int_{\mathbb{R}} w(t) \|\nabla_{\mathbb{S}} T_{\Sigma,t}\|_{L^2(\mathbb{S}^{K-1})}^2 dt = \int_{\mathbb{R}} w(t) t^4 \mathbb{E}_a \left[e^{-t^2 a^\top \Sigma a} s_\Sigma(a) \right] dt. \quad (19)$$

This target-only population quantity is nonnegative and, under trace normalization and a nondegenerate weight $w(t)$, satisfies $\mathcal{C}_1(\Sigma) = 0$ if and only if $\Sigma = I_K$. Indeed, $\mathcal{C}_1(\Sigma) = 0$ implies $P_a \Sigma a = 0$ for almost every a ; continuity extends this to every direction, so every vector is an eigenvector of Σ , giving $\Sigma = c I_K$, where trace normalization fixes $c = 1$. Quadrature details are given in Appendix B.3.

The spherical identity $\mathbb{E}_a[s_\Sigma(a)] = S_2/(K+2)$ establishes the leading spectral dependence. For $\Sigma = I_K + A$ with $\text{tr}(A) = 0$ and small $\|A\|_{\text{op}}$,

$$\mathcal{C}_1(I_K + A) = \frac{\text{tr}(A^2)}{K+2} \int_{\mathbb{R}} w(t) t^4 e^{-t^2} dt + O(\|A\|_{\text{op}}^3). \quad (20)$$

Thus S_2 governs weak anisotropy, while the exponential factor introduces higher spectral moments as anisotropy increases. Equivalently, the harmonic expansion equation 13 gives

$$\mathcal{C}_1(\Sigma) = \int_{\mathbb{R}} w(t) \sum_{\ell \in 2\mathbb{N}_0} \ell(\ell + K - 2) \|T_{\Sigma,t}^{(\ell)}\|_{L^2}^2 dt.$$

The isotropic component ($\ell = 0$) has zero angular energy, degree two ($\ell = 2$) provides the first nonzero contribution, and higher even degrees enter at finite anisotropy. Across all benchmark targets, $\mathcal{C}_1(\Sigma)$ has Spearman rank correlation $\rho = 0.818$ with M^* ($p = 1.9 \times 10^{-8}$), comparable to the target-side H^2 bounds in Table 2. This association reflects global target roughness rather than local slope.

Let $v_\Sigma(a) = a^\top \Sigma a$, $D_M(a) = \ell_{Q_\Sigma}(a; Z_M)$, and define relative directional error $R_M(a) = [D_M(a) - D_{\text{ref}}(a)]/D_{\text{ref}}(a)$. At $M/M^* \approx 0.5$, adding $s_\Sigma(a)$ to $v_\Sigma(a)$ increases the explained variance in $R_M(a)$ by at most 0.0003; the increment from local target-gradient energy was zero at reported precision. Global target roughness tracks directional complexity without implying that local angular slope marks the directions retaining optimization error.

5.2 Directional gradient variance

To connect target geometry to the stochastic objective, let $G_\Sigma(a; Z) = \nabla_Z \ell_{Q_\Sigma}(a; Z)$ and define the slicing-gradient variance for an IID batch $\mathcal{A}_M = \{a_m\}_{m=1}^M$ by

Table 4: Spearman correlation between initial slicing-gradient variance ($M = 64$) and M^* .

Target group	Number of targets	Spearman ρ_s	p -value
All targets	31	0.818	1.90×10^{-8}
$\kappa = 10$	8	0.100	0.8135
$\kappa = 50$	8	0.934	0.0007
$\kappa = 200$	8	0.932	0.0007

$$\sigma_{\text{slice}}^2(Z; \Sigma, M) = \mathbb{E}_{\mathcal{A}_M} \left\| \frac{1}{M} \sum_{m=1}^M G_{\Sigma}(a_m; Z) - \mathbb{E}_a[G_{\Sigma}(a; Z)] \right\|_F^2. \quad (21)$$

Unlike $\mathcal{C}_1(\Sigma)$, this quantity depends on the embedding state and the loss gradient. We evaluate it at initialization with $M = 64$ (Appendix B.3). Initial gradient variance correlates strongly with M^* across the full benchmark and within the matched $\kappa = 50$ and $\kappa = 200$ groups, but not at $\kappa = 10$ (Table 4). The association is therefore not an artifact of differences between condition-number groups, nor is it uniform across anisotropy levels.

For the $\kappa = 200$ triad, the estimated variances are 3.861, 0.130, and 0.235 for spike, smooth, and block, matching their direction requirements 4096, 1280, and 1280. While the full-set rank correlation equals that of $\mathcal{C}_1(\Sigma)$ ($\rho_s = 0.818$), the two quantities are distinct: $\mathcal{C}_1(\Sigma)$ is a population property of the target, whereas equation 21 reflects the loss gradient at the embedding state. Their agreement provides convergent evidence that target angular roughness is expressed in stochastic optimization, without establishing gradient variance as the sole driver of M^* (Figure 2).

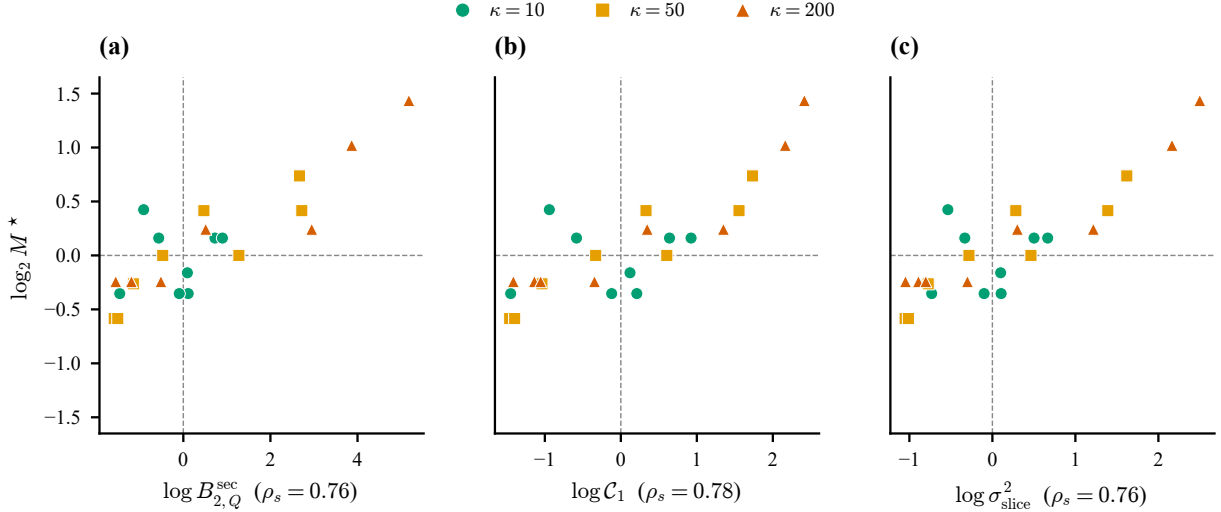


Figure 2: Within- κ associations with the empirical direction requirement. Each panel compares median-centered $\log_2 M^*$ against median-centered descriptors: (a) log secant bound $\log B_{2,Q}^{\text{sec}}$, (b) log Dirichlet energy $\log \mathcal{C}_1$, and (c) log initial gradient variance $\log \sigma_{\text{slice}}^2$ at $M = 64$. ρ_s is the Spearman rank correlation of median-centered residuals pooled across $\kappa \in \{10, 50, 200\}$.

5.3 The degree-two sampling interaction

The target's leading angular harmonic couples directly to finite directional sampling error. For $a_1, \dots, a_M \stackrel{\text{iid}}{\sim} \text{Unif}(\mathbb{S}^{K-1})$, define the empirical second-moment error

$$\Delta_M = \frac{1}{M} \sum_{m=1}^M a_m a_m^\top - \frac{1}{K} I_K. \quad (22)$$

Because $\text{tr}(\Sigma_0) = 0$, the sample mean of the target's degree-two harmonic $q_\Sigma(a) = a^\top \Sigma_0 a$ reduces exactly to

$$\frac{1}{M} \sum_{m=1}^M q_\Sigma(a_m) = \text{tr}(\Sigma_0 \Delta_M). \quad (23)$$

The sampled degree-two error is therefore the projection of the empirical direction-frame error onto the traceless target covariance.

Corollary 6 (IID Degree-Two Interaction Variance). *For IID uniform directions on $\mathbb{S}^{K-1}$,*

$$\mathbb{E}_{\mathcal{A}_M} \left[(\text{tr}(\Sigma_0 \Delta_M))^2 \right] = \frac{2 \text{tr}(\Sigma_0^2)}{MK(K+2)}. \quad (24)$$

Proof. Because $\mathbb{E}_a[q_\Sigma(a)] = \text{tr}(\Sigma_0)/K = 0$, the terms $q_\Sigma(a_m)$ are centered and independent. Their sample mean has variance $M^{-1} \mathbb{E}_a[q_\Sigma(a)^2]$, and equation 12 gives $\mathbb{E}_a[q_\Sigma(a)^2] = 2 \text{tr}(\Sigma_0^2)/[K(K+2)]$. $\square$

This interaction vanishes at isotropy, decays as M^{-1} under IID sampling, and scales with $S_2 = \text{tr}(\Sigma_0^2)$. The same invariant governs the weak-anisotropy expansion equation 20. While this quadratic interaction does not determine M^* , as it omits the embedding distribution and higher target harmonics, it isolates an explicit source of finite-sample error. Controlling this interaction through the dependence structure of sampled directions motivates the design of ODS in Section 6.

6 Orthogonal directional slicing

Since the degree-two interaction $\text{tr}(\Sigma_0 \Delta_M)$ depends only on the empirical second moment of the sampled directions, it can be eliminated by enforcing exact second-moment balance without altering the directional marginals or weighting the SIGReg objective. ODS enforces this balance using complete random orthonormal frames.

6.1 Eliminating the quadratic sampling interaction

Let $M = cK$ for some integer $c \geq 1$, and draw independent Haar-distributed orthogonal matrices $U_1, \dots, U_c \in O(K)$. ODS forms the direction batch from their columns, $\mathcal{A}_M^{\text{ODS}} = \{U_b e_j : b = 1, \dots, c, j = 1, \dots, K\}$, where $\{e_j\}_{j=1}^K$ is the standard basis. In practice, each U_b is generated by a sign-corrected QR factorization of a Gaussian matrix.

Theorem 7 (Second-moment Balance). *For the ODS batch $\mathcal{A}_M^{\text{ODS}}$, every direction is marginally uniform on $\mathbb{S}^{K-1}$ and*

$$\frac{1}{M} \sum_{a \in \mathcal{A}_M^{\text{ODS}}} aa^\top = \frac{1}{K} I_K. \quad (25)$$

Consequently, $\Delta_M = 0$ and, for every trace-normalized covariance $\Sigma = I_K + \Sigma_0$,

$$\frac{1}{M} \sum_{a \in \mathcal{A}_M^{\text{ODS}}} q_{\Sigma}(a) = \text{tr}(\Sigma_0 \Delta_M) = 0 \quad (26)$$

for every realization of the sampled frames. Moreover, for any integrable directional loss,

$$\mathbb{E}_{\text{ODS}} \left[\frac{1}{M} \sum_{a \in \mathcal{A}_M^{\text{ODS}}} \ell_Q(a; Z) \right] = \mathbb{E}_{a \sim \text{Unif}(\mathbb{S}^{K-1})} [\ell_Q(a; Z)]. \quad (27)$$

Proof. For each frame, $\sum_{j=1}^K (U_b e_j)(U_b e_j)^\top = U_b U_b^\top = I_K$. Summing over $b = 1, \dots, c$ and dividing by $M = cK$ gives equation 25, hence $\Delta_M = 0$ and equation 26. Haar invariance makes every column $U_b e_j$ marginally uniform on the sphere. Linearity of expectation then gives equation 27; independence between directions is not required. $\square$

Corollary 8 (Partial-Frame Error). *Let $M = cK + r$ with $0 \leq r < K$, and append the first r columns of an independent Haar orthogonal matrix to c complete frames. If P_r denotes the projector onto their span, then*

$$\Delta_M = \frac{1}{M} \left(P_r - \frac{r}{K} I_K \right), \quad \mathbb{E}[\Delta_M] = 0, \quad (28)$$

and

$$\|\Delta_M\|_F^2 = \frac{r(K-r)}{KM^2}. \quad (29)$$

For every traceless symmetric matrix A ,

$$\mathbb{E}[\text{tr}(A \Delta_M)^2] = \frac{2r(K-r)}{M^2 K(K-1)(K+2)} \text{tr}(A^2). \quad (30)$$

Thus an incomplete frame remains unbiased in expectation, but incurs nonzero variance across realizations. Appendix A.4 proves equation 29–equation 30.

Corollary 9 (Degree-Two Cancellation). *Let h be any spherical harmonic of degree two on $\mathbb{S}^{K-1}$. Then, for every ODS batch,*

$$\frac{1}{M} \sum_{a \in \mathcal{A}_M^{\text{ODS}}} h(a) = \mathbb{E}_a[h(a)] = 0. \quad (31)$$

Consequently, ODS cancels the complete degree-two component $T_{\Sigma, t}^{(2)}$ of the anisotropic Gaussian target for every frequency t .

Proof. Every degree-two spherical harmonic has the form $h(a) = a^\top B a$ for a symmetric matrix B with $\text{tr}(B) = 0$. Therefore,

$$\frac{1}{M} \sum_{a \in \mathcal{A}_M^{\text{ODS}}} h(a) = \text{tr} \left(B \frac{1}{M} \sum_{a \in \mathcal{A}_M^{\text{ODS}}} a a^\top \right) = \frac{1}{K} \text{tr}(B) = 0,$$

where the second equality uses equation 25. Degree-two spherical harmonics have zero spherical mean, which proves equation 31. $\square$

Proposition 10 (Degree-Four Residual). *Let $q_\Sigma(a) = a^\top \Sigma_0 a$ and $S_2 = \text{tr}(\Sigma_0^2)$. The degree-four component of q_Σ^2 is*

$$h_{4,\Sigma}(a) = q_\Sigma(a)^2 - \frac{4}{K+4} a^\top \Sigma_0^2 a + \frac{2S_2}{(K+2)(K+4)}. \quad (32)$$

If a complete orthogonal frame is aligned with an eigenbasis $\{v_j\}_{j=1}^K$ of Σ_0 , then

$$\frac{1}{K} \sum_{j=1}^K h_{4,\Sigma}(v_j) = \frac{S_2}{K+2}. \quad (33)$$

For a Haar frame U ,

$$\mathbb{E}_U \left[\frac{1}{K} \sum_{j=1}^K h_{4,\Sigma}(Ue_j) \right] = 0. \quad (34)$$

The second-order term in the expansion of equation 11 is $e^{-t^2/2}(t^4/8)q_\Sigma^2$, so $h_{4,\Sigma}$ yields an explicit target-side degree-four residual. Complete frames achieve exactness at degree two, but not at degree four. In the terminology of spherical designs, a complete ODS batch is a design of harmonic index 2 (Zhu & Wang, 2015). It eliminates the leading degree-two target interaction, while leaving higher-degree target harmonics and embedding-side directional variation uncanceled.

Because exactness follows directly from the tight-frame identity $\sum_{j=1}^K (Ue_j)(Ue_j)^\top = I_K$, ODS requires neither knowledge of the target covariance nor importance weighting. Generating c orthogonal frames costs $O(cK^3) = O(MK^2)$ operations. Appendix A.5 proves Proposition 10.

6.2 Effect on finite-direction optimization

We evaluate ODS against IID directional sampling across the all benchmark targets using paired initial embeddings and seeds (Appendix B.4). ODS reduces initial slicing-gradient variance in every condition, by 5.5%–28.5% relative to IID sampling (Figure 3a). Because $\Sigma_0 = 0$ at isotropy, the 28.5% reduction there reflects reduced directional variation on the embedding side alone. Under severe anisotropy ($\kappa = 200$), the reduction drops to 5.6%–13.3%, consistent with uncanceled higher-degree angular harmonics.

Across all 31 conditions, the shift in M^* is not statistically significant (Figure 3b). The direction requirement decreases in 9 targets, remains unchanged in 18, and increases in 4. The largest improvement occurs for the $\kappa = 200$ spike target (from 4096 to 3072), while the largest regression occurs for the $\kappa = 200$ linear target (from 896 to 1280 in the paired evaluation). Although ODS eliminates the leading degree-two sampling term and reduces gradient variance in every condition, it does not reliably lower M^* . The persistence of finite-direction difficulty indicates that higher angular degrees and embedding-target interactions dominate optimization.

7 Discussion and limitations

Directional complexity in sliced distribution matching is governed by target angular geometry. Standard SIGReg samples directions uniformly, treating slicing as an uncoupled Monte Carlo step. Under anisotropy, however, probability mass concentrates along low-dimensional subspaces, exciting higher-degree harmonics that demand large direction budgets. Eliminating the leading degree-two error with ODS reduces gradient variance, yet does not lower direction requirements.

On the other hand, our analysis remains limited to trace-normalized Gaussian targets and direct embedding optimization. The analytical bounds evaluate static quasi-uniform interpolation on the sphere. Separating the embedding and target Sobolev norms discards their cross-coupling and does not model stochastic optimization dynamics directly. Non-Gaussian targets introduce odd harmonics that further reshape the

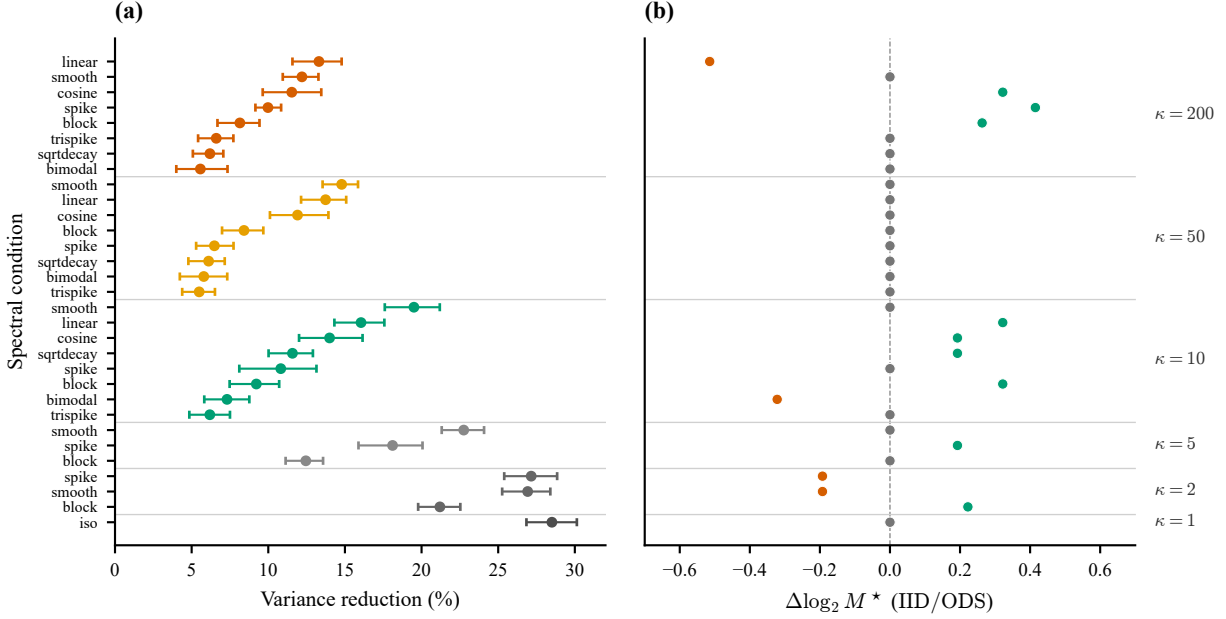


Figure 3: Paired ODS and IID comparison across the 31 benchmark targets. (a) Percentage reduction in initial slicing-gradient variance under ODS at $M = 64$; points are sample means and bars are 95% bootstrap CIs. (b) Paired direction-budget shift $\log_2(M_{\text{IID}}^*/M_{\text{ODS}}^*)$, where positive values favor ODS (paired Wilcoxon $p = 0.11$, sign test $p = 0.13$).

angular spectrum. In representation learning, deep encoders and data augmentations introduce inductive biases that may alter finite-direction requirements. Finally, ODS eliminates only degree-two error, leaving higher angular orders unresolved.

8 Conclusion

In this work, we showed that finite-direction SIGReg depends fundamentally on target angular geometry. Our spherical approximation bounds identify the spectral endpoints and centered moments that govern directional complexity, and our empirical benchmark confirms that matched condition numbers produce widely different direction requirements. The exact second-moment balance achieved by ODS shows that eliminating degree-two directional error reduces gradient variance, but leaves finite-direction requirements dominated by higher angular orders. More broadly, finite-direction guarantees for structured targets must treat target geometry and directional design as a coupled problem.

References

- Robert Jenkinson Alvarez. Beyond isotropy in JEPAs: Hamiltonian geometry and symplectic prediction. *arXiv preprint arXiv:2605.20107*, 2026.
- Randall Balestriero and Yann LeCun. LeJEPA: Provable and scalable self-supervised learning without the heuristics. *arXiv preprint arXiv:2511.08544*, 2025.
- Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015. doi: 10.1007/s10851-014-0506-3.
- Harald Cramér and Herman Wold. Some theorems on distribution functions. *Journal of the London Mathematical Society*, s1-11(4):290–294, 1936. doi: 10.1112/jlms/s1-11.4.290.
- Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.
- Rémi Leluc, Aymeric Dieuleveut, François Portier, Johan Segers, and Aigerim Zhuman. Sliced-Wasserstein estimation with spherical harmonics as control variates. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 27191–27214, 2024.
- Khai Nguyen, Tongzheng Ren, and Nhat Ho. Markovian sliced wasserstein distances: Beyond independent projections. In *Advances in Neural Information Processing Systems*, volume 36, 2023a.
- Khai Nguyen, Tongzheng Ren, Huy Nguyen, Litu Rout, Tan Nguyen, and Nhat Ho. Hierarchical sliced wasserstein distance. In *International Conference on Learning Representations*, 2023b.
- Khai Nguyen, Nicola Barileto, and Nhat Ho. Quasi-monte carlo for 3d sliced wasserstein. In *International Conference on Learning Representations*, 2024.
- Léo Nicollier, Max Dunitz, Marc Pic, Pablo Musé, Enric Meinhardt-Llopis, and Gabriele Facciolo. SPHERE-JEPA: Spherical prediction with homogeneous embeddings. *arXiv preprint arXiv:2605.26900*, 2026a.
- Léo Nicollier, Enric Meinhardt-Llopis, Marc Pic, Pablo Musé, and Gabriele Facciolo. Beyond gaussian worlds: Latent geometry matters for JEPAs. *arXiv preprint arXiv:2609.21656*, 2026b.
- Vladimir Petrovic, Rémi Bardenet, and Agnès Desolneux. Repulsive monte carlo on the sphere for the sliced wasserstein distance. *Transactions on Machine Learning Research*, 2026.
- Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. *Scale Space and Variational Methods in Computer Vision*, pp. 435–446, 2012.
- Mark Rowland, Jiri Hron, Yunhao Tang, Krzysztof Choromanski, Tamas Sarlos, and Adrian Weller. Orthogonal estimation of wasserstein distances. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 186–195, 2019.
- Zhengjie Sun, Wenwu Gao, and Xingping Sun. Spherical quasi-interpolation using scaled zonal kernels. *arXiv preprint arXiv:2408.14803*, 2024.
- Hongliang Zhang, Shuo Chen, Lei Luo, and Jian Yang. Towards better spherical sliced-Wasserstein distance learning with data-adaptive discriminative projection direction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.
- Yuan Zhu and Hui Wang. Harmonic index designs and their applications. *Journal of Complexity*, 31(6): 837–853, 2015.

A Proofs

Throughout this appendix, $a \sim \text{Unif}(\mathbb{S}^{K-1})$ and $\mathbb{E}_a$ denotes expectation with respect to normalized surface measure. We write $P_a = I_K - aa^\top$ for the tangent-space projector and $\Delta_{\mathbb{S}}$ for the Laplace–Beltrami operator.

A.1 Proof of Theorem 1

Proof. Let $q_{\Sigma}(a) = a^\top \Sigma_0 a$ and $s_{\Sigma}(a) = a^\top \Sigma_0^2 a - q_{\Sigma}(a)^2$. Under normalized surface measure on $\mathbb{S}^{K-1}$ and the spectral H^2 convention of Section 3, the spherical gradient equation 14 and Laplacian equation 15 yield

$$\|T_{\Sigma,t}\|_{H^2}^2 = \mathbb{E}_a \left[T_{\Sigma,t}(a)^2 \left\{ 1 + 2t^4 s_{\Sigma}(a) + (t^2 K q_{\Sigma}(a) + t^4 s_{\Sigma}(a))^2 \right\} \right]. \quad (35)$$

Applying the spherical fourth-moment identity gives

$$\mathbb{E}_a[q_{\Sigma}(a)^2] = \frac{2S_2}{K(K+2)}, \quad \mathbb{E}_a[s_{\Sigma}(a)] = \frac{S_2}{K+2}. \quad (36)$$

Since $0 \leq s_{\Sigma}(a) \leq a^\top \Sigma_0^2 a$, we have

$$\mathbb{E}_a[s_{\Sigma}(a)^2] \leq \frac{S_2^2 + 2S_4}{K(K+2)}. \quad (37)$$

Using $(x+y)^2 \leq 2x^2 + 2y^2$ on the squared Laplacian term in equation 35, substituting the spherical expectations equation 36 and equation 37, and applying the uniform bound $T_{\Sigma,t}(a)^2 \leq e^{-\lambda_{\min}(\Sigma)t^2}$ yields

$$\|T_{\Sigma,t}\|_{H^2}^2 \leq e^{-\lambda_{\min}(\Sigma)t^2} \left[1 + \frac{2(2K+1)S_2}{K+2}t^4 + \frac{2(S_2^2 + 2S_4)}{K(K+2)}t^8 \right]. \quad (38)$$

Multiplying equation 38 by $w_{\rho}(t) = e^{-\rho^2 t^2}$, setting $\gamma = \rho^2 + \lambda_{\min}(\Sigma)$, and integrating over $t \in \mathbb{R}$ via the Gaussian moments

$$\int_{\mathbb{R}} e^{-\gamma t^2} dt = \frac{\sqrt{\pi}}{\gamma^{1/2}}, \quad \int_{\mathbb{R}} t^4 e^{-\gamma t^2} dt = \frac{3\sqrt{\pi}}{4\gamma^{5/2}}, \quad \int_{\mathbb{R}} t^8 e^{-\gamma t^2} dt = \frac{105\sqrt{\pi}}{16\gamma^{9/2}},$$

proves equation 7. Finally, convexity of the squared Hilbert norm gives $\|g_t\|_{H^2}^2 \leq 2\|\varphi_{P,\cdot}(t)\|_{H^2}^2 + 2\|T_{\Sigma,t}\|_{H^2}^2$. Integrating against $w_{\rho}(t)$ and applying the spherical interpolation bound equation 6 proves equation 8 and equation 9. $\square$

A.2 Proof of Lemma 2

Proof. For a single embedding vector $z \in \mathbb{R}^K$ with $f_z(a) = e^{ita^\top z}$, direct differentiation gives

$$\nabla_{\mathbb{S}} f_z = it f_z P_a z, \quad \Delta_{\mathbb{S}} f_z = f_z [-t^2 (\|z\|_2^2 - (a^\top z)^2) - it(K-1)a^\top z].$$

Hence, whenever $\|z\|_2 \leq R$,

$$\|f_z\|_{H^2}^2 \leq 1 + [(K-1)^2 + 2]R^2 t^2 + 2(K-1)R^3 |t|^3 + R^4 t^4.$$

By convexity of the squared Hilbert norm,

$$\left\| \frac{1}{N} \sum_{n=1}^N f_{z_n} \right\|_{H^2}^2 \leq \frac{1}{N} \sum_{n=1}^N \|f_{z_n}\|_{H^2}^2.$$

Integrating against $w_\rho(t) = e^{-\rho^2 t^2}$ via

$$\int_{\mathbb{R}} e^{-\rho^2 t^2} dt = \frac{\sqrt{\pi}}{\rho}, \quad \int_{\mathbb{R}} t^2 e^{-\rho^2 t^2} dt = \frac{\sqrt{\pi}}{2\rho^3}, \quad \int_{\mathbb{R}} |t|^3 e^{-\rho^2 t^2} dt = \frac{1}{\rho^4}, \quad \int_{\mathbb{R}} t^4 e^{-\rho^2 t^2} dt = \frac{3\sqrt{\pi}}{4\rho^5},$$

proves equation 10. $\square$

A.3 Proof of Proposition 4

Proof. Let ν_- and ν_+ be the minimal and maximal eigenvalues of Σ_0 , with $\Delta_\nu = \nu_+ - \nu_-$. For $\theta \geq 0$, convexity of $x \mapsto e^{-\theta x}$ implies the endpoint-secant majorant

$$e^{-\theta x} \leq L_\theta(x) := \frac{\nu_+ - x}{\Delta_\nu} e^{-\theta \nu_-} + \frac{x - \nu_-}{\Delta_\nu} e^{-\theta \nu_+}, \quad x \in [\nu_-, \nu_+]. \quad (39)$$

Any secant majorant dominating $e^{-\theta x}$ at the endpoints must satisfy $f(\nu_-) \geq e^{-\theta \nu_-}$ and $f(\nu_+) \geq e^{-\theta \nu_+}$; secant interpolation ensures that L_θ is the pointwise minimal secant majorant on $[\nu_-, \nu_+]$.

Writing $q = q_\Sigma(a)$ and $s = s_\Sigma(a)$ for brevity, define

$$\begin{aligned} H_\theta(q, s) &= 2\theta^2 s + (K\theta q + \theta^2 s)^2, \\ \psi_\Sigma(\theta) &= \mathbb{E}_a[e^{-\theta q}], \\ H_0(\theta) &= \mathbb{E}_a[H_\theta(q, s)], \\ H_1(\theta) &= \mathbb{E}_a[q H_\theta(q, s)]. \end{aligned}$$

The exact integrated target norm equation 35 is

$$B_{2,Q}(\Sigma) = \int_{\mathbb{R}} e^{-(\rho^2+1)t^2} \left[\psi_\Sigma(t^2) + \mathbb{E}_a \left[e^{-t^2 q} H_{t^2}(q, s) \right] \right] dt. \quad (40)$$

Since $H_\theta(q, s) \geq 0$, majorizing $e^{-\theta q}$ via equation 39 yields

$$\mathbb{E}_a[e^{-\theta q} H_\theta(q, s)] \leq e^{-\theta \nu_-} H_0(\theta) + \frac{e^{-\theta \nu_+} - e^{-\theta \nu_-}}{\Delta_\nu} [H_1(\theta) - \nu_- H_0(\theta)]. \quad (41)$$

The required spherical moments follow from Gaussian radial decomposition. Let $A = \Sigma_0$, let $z = Ra \sim \mathcal{N}(0, I_K)$, and let $B_1, \dots, B_m$ be commuting symmetric matrices. Wick's formula and $\mathbb{E}[R^{2m}] = K(K+2) \cdots (K+2m-2)$ give the spherical contraction identity

$$\mathbb{E}_a \left[\prod_{i=1}^m (a^\top B_i a) \right] = \frac{\sum_{\sigma \in \mathfrak{S}_m} 2^{m-c(\sigma)} \prod_{C \in \text{cyc}(\sigma)} \text{tr} \left(\prod_{i \in C} B_i \right)}{K(K+2) \cdots (K+2m-2)}, \quad (42)$$

where $c(\sigma)$ is the number of cycles in permutation σ . Applying equation 42 to powers of A with $\text{tr}(A) = 0$, and setting $r = a^\top A^2 a$ and $s = r - q^2$, yields

$$\mathbb{E}_a[q^3] = \frac{8S_3}{K(K+2)(K+4)}, \quad \mathbb{E}_a[qs] = \frac{2S_3}{(K+2)(K+4)}, \quad (43)$$

$$\mathbb{E}_a[q^2 s] = \frac{2(S_2^2 + 4S_4)}{(K+2)(K+4)(K+6)}, \quad \mathbb{E}_a[s^2] = \frac{(K^2 + 6K + 12)S_2^2 + 2K(K+2)S_4}{K(K+2)(K+4)(K+6)}, \quad (44)$$

$$\mathbb{E}_a[qs^2] = \frac{4[(K^2 + 4K + 8)S_2 S_3 + 2K(K+2)S_5]}{K(K+2)(K+4)(K+6)(K+8)}. \quad (45)$$

Expanding $H_0(\theta) = \sum_{j=2}^4 \beta_j \theta^j$ gives

$$\begin{aligned}\beta_2 &= \frac{2(K+1)S_2}{K+2}, & \beta_3 &= \frac{4KS_3}{(K+2)(K+4)}, \\ \beta_4 &= \frac{(K^2+6K+12)S_2^2 + 2K(K+2)S_4}{K(K+2)(K+4)(K+6)}.\end{aligned}\tag{46}$$

Similarly, $H_1(\theta) - \nu_- H_0(\theta) = \sum_{j=2}^4 \eta_j \theta^j$ with coefficients

$$\begin{aligned}\eta_2 &= \frac{4(2K+1)S_3}{(K+2)(K+4)} - \nu_- \beta_2, \\ \eta_3 &= \frac{4K(S_2^2 + 4S_4)}{(K+2)(K+4)(K+6)} - \nu_- \beta_3, \\ \eta_4 &= \frac{4[(K^2+4K+8)S_2S_3 + 2K(K+2)S_5]}{K(K+2)(K+4)(K+6)(K+8)} - \nu_- \beta_4.\end{aligned}\tag{47}$$

For the zeroth-order term, let $\nu_1, \dots, \nu_K$ be the eigenvalues of Σ_0 , set $\Pi_\Sigma(\lambda) = \prod_{i=1}^K (1 + 2\lambda\nu_i)^{-1/2}$, and define

$$\theta_* = \begin{cases} -K/(2\nu_-), & \nu_- < 0, \\ +\infty, & \nu_- \geq 0, \end{cases} \quad \psi_{\Sigma, \text{ub}}(\theta) = \begin{cases} \min\{\Pi_\Sigma(\theta/K), e^{-\nu_- \theta}\}, & 0 \leq \theta < \theta_*, \\ e^{-\nu_- \theta}, & \theta \geq \theta_*. \end{cases}$$

To verify this bound, condition on a in $z = Ra$. Whenever $1 + 2\lambda\nu_i > 0$ for all i ,

$$\mathbb{E}_a \left[(1 + 2\lambda q_\Sigma(a))^{-K/2} \right] = \mathbb{E}_z [e^{-\lambda z^\top \Sigma_0 z}] = \Pi_\Sigma(\lambda).\tag{48}$$

Since $\log(1+x) \leq x$ implies $(1+u/K)^{-K/2} \geq e^{-u/2}$ for $u > -K$, evaluating at $u = 2\theta q_\Sigma(a)$ and $\lambda = \theta/K$ gives $\mathbb{E}_a [e^{-\theta q_\Sigma(a)}] \leq \Pi_\Sigma(\theta/K)$ for $0 \leq \theta < \theta_*$. Combining this with the Rayleigh bound $q_\Sigma(a) \geq \nu_-$ proves the validity of $\psi_{\Sigma, \text{ub}}$.

Letting $G_j(x) = \Gamma(j + \frac{1}{2})x^{-(j+1/2)}$ and $\gamma = \rho^2 + \lambda_{\min}(\Sigma)$, substituting $\psi_{\Sigma, \text{ub}}$, equation 41, equation 46, and equation 47 into equation 40 yields

$$\begin{aligned}B_{2,Q}^{\text{sec}}(\Sigma) &:= \int_{\mathbb{R}} e^{-(\rho^2+1)t^2} \psi_{\Sigma, \text{ub}}(t^2) dt + \sum_{j=2}^4 \beta_j G_j(\gamma) \\ &+ \frac{1}{\Delta_\nu} \sum_{j=2}^4 \eta_j [G_j(\gamma + \Delta_\nu) - G_j(\gamma)],\end{aligned}\tag{49}$$

evaluating by continuity at $\Delta_\nu = 0$. Because the final sum retains the negative secant correction and the Laplacian cross term is preserved rather than split via $(x+y)^2 \leq 2x^2 + 2y^2$, we obtain $B_{2,Q}(\Sigma) \leq B_{2,Q}^{\text{sec}}(\Sigma) \leq B_{2,Q}^{\text{base}}(\Sigma)$, with strict upper inequality whenever $\Sigma \neq I_K$. $\square$

A.4 Proof of Corollary 8

Proof. The c complete frames contribute cI_K , while the partial frame contributes P_r , so

$$\frac{1}{M} \sum_{m=1}^M a_m a_m^\top = \frac{cI_K + P_r}{M}.$$

Using $M = cK + r$ gives equation 28. Haar invariance yields $\mathbb{E}[P_r] = (r/K)I_K$. Since $P_r^2 = P_r$ and $\text{tr}(P_r) = r$,

$$\left\|P_r - \frac{r}{K}I_K\right\|_F^2 = r - \frac{r^2}{K} = \frac{r(K-r)}{K},$$

which proves equation 29.

Let $X = P_r - (r/K)I_K$. Rotational invariance makes the covariance of X isotropic on the $(K-1)(K+2)/2$ -dimensional space of traceless symmetric matrices. Therefore, for traceless symmetric A ,

$$\mathbb{E}[\text{tr}(AX)^2] = \frac{\mathbb{E}\|X\|_F^2}{(K-1)(K+2)/2} \text{tr}(A^2) = \frac{2r(K-r)}{K(K-1)(K+2)} \text{tr}(A^2).$$

Dividing by M^2 proves equation 30. □

A.5 Proof of Proposition 10

Proof. Let $A = \Sigma_0$ and define the homogeneous polynomial

$$H_4(x) = (x^\top Ax)^2 - \frac{4\|x\|_2^2}{K+4} x^\top A^2 x + \frac{2S_2\|x\|_2^4}{(K+2)(K+4)}.$$

Since $\text{tr}(A) = 0$,

$$\Delta(x^\top Ax)^2 = 8x^\top A^2 x,$$

$$\Delta(\|x\|_2^2 x^\top A^2 x) = 2(K+4)x^\top A^2 x + 2S_2\|x\|_2^2, \quad \Delta\|x\|_2^4 = 4(K+2)\|x\|_2^2.$$

Thus $\Delta H_4 = 0$, and its restriction to the sphere is the degree-four spherical harmonic equation 32.

If v_j is an eigenvector of A with eigenvalue α_j , then $q_\Sigma(v_j)^2 = v_j^\top A^2 v_j = \alpha_j^2$. Averaging equation 32 and using $\sum_j \alpha_j^2 = S_2$ gives equation 33. Finally, each column of a Haar frame is marginally uniform, and every nonconstant spherical harmonic has zero spherical mean, so the frame average of $h_{4,\Sigma}$ has expectation zero. □

B Experimental details

B.1 Finite-direction complexity benchmark

Table 5 gives the shared protocol for the direct-embedding benchmark.

Table 5: Benchmark protocol.

Quantity	Specification
Embedding matrix	$Z \in \mathbb{R}^{1000 \times 64}$
Initialization	iID standard Gaussian entries $\mathcal{N}(0, 1)$
Target	$Q_\Sigma = \mathcal{N}(0, \Sigma)$, $\text{tr}(\Sigma) = 64$
Optimizer	Adam (Kingma & Ba, 2015), 10,000 steps, learning rate 0.01
Adam parameters	$(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$
Training directions	fresh $a = g/\ g\ _2$, $g \sim \mathcal{N}(0, I_{64})$, each step
Frequency integration	$w(t) = e^{-t^2}$ on $[0, 5]$, 17-point Gauss–Legendre quadrature

The target suite contains an isotropic anchor and eight anisotropic spectral families (Figure 4). Spike, smooth, and block are evaluated at condition numbers 2, 5, 10, 50, 200; the remaining families are evaluated at 10, 50, 200. Every covariance is rescaled to trace 64.

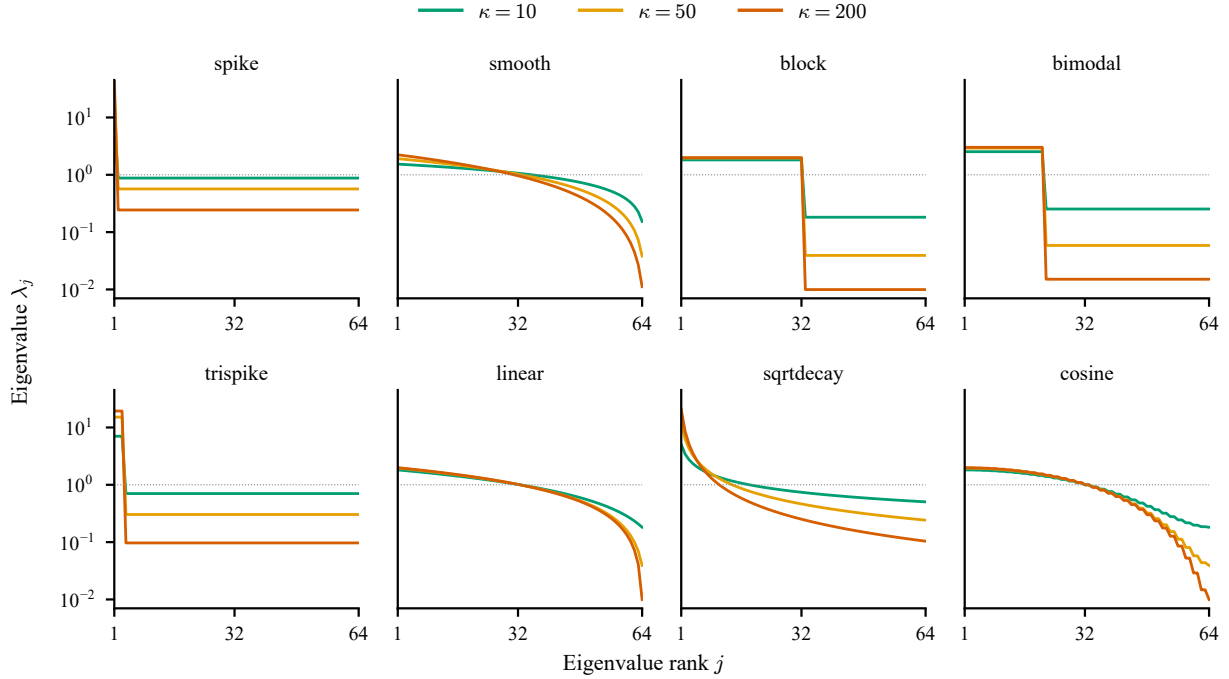


Figure 4: Eigenvalue profiles across the eight target families. Each panel shows eigenvalues λ_j in descending order against rank j on a logarithmic scale for $\kappa \in \{10, 50, 200\}$; dotted horizontal lines mark the isotropic level $\lambda_j = 1$.

We evaluate across 20 direction budgets: $\mathcal{M} = \{2^k\}_{k=3}^{12} \cup \{192, 384, 640, 768, 896, 1280, 1536, 1792, 2560, 3072\}$. The isotropic anchor and the $\kappa = 200$ triad (spike, smooth, block) use 20 training seeds; all remaining conditions use 10.

B.2 Evaluation and reference construction

Each trained embedding is evaluated on four fixed panels of 2048 uniform spherical directions (8192 directions total), independent of training and shared across all conditions and seeds, averaging the discrepancy over panels. Reference runs use $M_{\text{ref}} = 16,384$, with 15 seeds for the isotropic anchor and the three $\kappa = 200$ focal targets, and 5 seeds for each remaining condition. The reference value $E_{\text{ref}}(\Sigma)$ is the median over reference seeds with a 95% nonparametric bootstrap interval. To verify that $M_{\text{ref}} = 16,384$ is sufficient, we tested $M \in \{2048, 4096, 8192\}$ on three difficult conditions using paired validation seeds. A budget was considered adequate if the simultaneous one-sided 95% bootstrap upper bound on the condition-wise median relative excess was at most 2%. The observed bounds were 15.31%, 7.92%, and 4.22%, confirming that smaller budgets $M \leq 8192$ remain under-saturated relative to the $M_{\text{ref}} = 16,384$ reference. Success frequencies in equation 18 use 95% Wilson score intervals.

B.3 Angular mechanism diagnostics

The pointwise directional analysis in Section 5.1 evaluates seven focal targets: the isotropic anchor and the spike, smooth, and block families at $\kappa \in \{50, 200\}$. For each target and direction budget, 20 training seeds are evaluated on a fixed panel of 20,000 uniform spherical directions. Target Dirichlet energy $C_1(\Sigma)$ in equation 19 is computed using 17-point Gauss–Legendre quadrature on $[0, 5]$. Initial slicing-gradient variance is evaluated at $M = 64$ under standard IID sampling across $B = 50$ independent direction batches at shared

initialization Z_0 . Letting $G_b = M^{-1} \sum_{a \in \mathcal{A}_M^{(b)}} \nabla_Z \ell_{Q_\Sigma}(a; Z_0)$ and $\bar{G} = B^{-1} \sum_{b=1}^B G_b$, the estimator reported in Section 5.2 is

$$\hat{\sigma}_{\text{slice}}^2 = \frac{1}{B} \sum_{b=1}^B \|G_b - \bar{G}\|_F^2. \quad (50)$$

B.4 Paired ODS evaluation

Algorithm 1 gives the ODS batch-generation procedure. The ODS evaluation compares paired runs across all 31 benchmark targets over 17 direction budgets: $\mathcal{M} = \{2^k\}_{k=6}^{12} \cup \{192, 384, 640, 768, 896, 1280, 1536, 1792, 2560, 3072\}$. For each condition, seed, and budget, IID and ODS runs share the same initialization and Adam settings. The isotropic anchor and the spike, smooth, and block targets at $\kappa \in \{2, 5, 50, 200\}$ use 15 paired seeds; remaining targets use 10 paired seeds.

Algorithm 1 Orthogonal directional slicing

Require: Embeddings Z , target Q , dimension K , direction budget $M = cK$

Ensure: Finite-direction loss $\hat{\mathcal{L}}_M(Z; Q)$

- 1: $\mathcal{A} \leftarrow \emptyset$
 - 2: **for** $b = 1, \dots, c$ **do**
 - 3: Sample $H_b \in \mathbb{R}^{K \times K}$ with IID $\mathcal{N}(0, 1)$ entries
 - 4: Compute $H_b = U_b R_b$ by QR factorization
 - 5: $U_b \leftarrow U_b \text{diag}(\text{sgn}(\text{diag}(R_b)))$
 - 6: Append the columns of U_b to $\mathcal{A}$
 - 7: **end for**
 - 8: **return** $M^{-1} \sum_{a \in \mathcal{A}} \ell_Q(a; Z)$
-

Post-training evaluation uses the four-panel suite and reference thresholds $E_{\text{ref}}(\Sigma)$ from Appendix B.2. Initial frame error, degree-two coupling, and slicing-gradient variance are measured at $M = 64$ prior to optimization. Across-condition inference on M^* uses a two-sided Wilcoxon signed-rank test ($W = 63.0$, $p = 0.1067$, Hodges–Lehmann median shift of 0) and a one-sided paired sign test (9/13 non-tied wins, $p = 0.1334$).

B.5 Tolerance sensitivity of the matched- κ comparison

To verify that the empirical separation between spectral shapes does not depend on the choice of tolerance $\tau = 0.10$ in equation 18, we re-extract M^* at $\tau = 0.05$ and $\tau = 0.20$ for the matched $\kappa = 200$ targets across 20 training seeds. At $\tau = 0.05$, the spike target fails to reach 80% success within the tested grid ($M^* > 4096$), while smooth and block require $M^* = 3072$ and 4096. At $\tau = 0.10$, the requirements are 4096 (spike), 1280 (smooth), and 1280 (block). At $\tau = 0.20$, they decrease to 1536 (spike), 384 (smooth), and 512 (block). Spike requires much more directions than smooth and block across every tested tolerance.