INTERPRETING SAMPLEWISE FUNCTION PERTURBATIONS: SIGNAL, NOISE, AND NONLINEAR STRUCTURE

Anonymous authors
Paper under double-blind review

ABSTRACT

Samplewise adversarial perturbations have been motivated as stronger regularization, but a larger worst-case penalty does not specify how the learned predictor changes. We study exact samplewise logistic estimation in an RKHS with a unit-diagonal kernel and fixed positive ridge, also realized by per-example perturbations of normalized fixed-feature heads. A common operator governs weak-signal population response and balanced-label limiting covariance. Increasing the radius raises the gain of every observable signal mode and inflates the limiting covariance, although the balanced population minimizer remains zero. These leading responses are not specific to the perturbed objective: ordinary logistic regression with an analytically chosen ridge and output scale matches both exactly. The agreement leaves nonlinear structure undetermined. A general cubic interaction operator identifies the first mechanism-specific difference after matching, with a fifth-order remainder. In a periodic weak-signal model, its third-harmonic projection has opposite local responses: samplewise perturbation attenuates the harmonic while its matched control amplifies it. This attenuation coexists with norm growth, and the Bayes score itself contains the harmonic. Separate $O(n^{-1})$ covariance-error bounds quantify nonlinear fitting and averaged random-design effects. Independent kernel and frozen ResNet/ViT fits test the predictions, while ordinary logistic controls reproduce the fixed-setting task gains. Thus, samplewise perturbation cannot be summarized as uniform shrinkage, and component attenuation alone does not establish predictive benefit.

1 INTRODUCTION

Worst-case local perturbations underlie several widely used learning objectives. Adversarial training chooses a separate perturbation around each input (Madry et al., 2018), whereas Sharpness-Aware Minimization (SAM) perturbs model parameters, typically sharing a perturbation across a batch (Foret et al., 2021). Subsequent work makes the granularity of weight perturbations an explicit design choice: m-sharpness and microbatch SAM associate perturbations with smaller groups of examples (Andriushchenko & Flammarion, 2022; Behdin et al., 2023), while Zhou et al. (2022) formulate per-instance adversarial weight perturbations and develop δ-SAM to approximate this computationally expensive objective. These choices change the order of perturbation and aggregation:

$$\underbrace{\sup_{\|\nu\|\le\delta} \frac{1}{n}\sum_{i=1}^{n} \ell_i(w+\nu)}_{\text{shared perturbation}} \quad \rightsquigarrow \quad \underbrace{\frac{1}{n}\sum_{i=1}^{n} \sup_{\|\nu_i\|\le\delta} \ell_i(w+\nu_i)}_{\text{samplewise perturbations}}.$$

We focus on the finest-grained case, in which each example has its own worst-case perturbation. A normalized fixed-feature model connects this parameter-space objective exactly to function-space perturbations, making the inner problem tractable while preserving the samplewise choice.

Samplewise adversarial perturbations have been motivated as a stronger form of regularization. For example, Zhou et al. (2022) interpret the larger per-instance sharpness penalty as suggesting stronger regularization effects and discuss its potential generalization benefits. Yet this objective-level interpretation is too coarse to explain what changes in the learned predictor. A larger worst-case penalty does not by itself determine how the estimator responds to genuine signal, finite-sample

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.

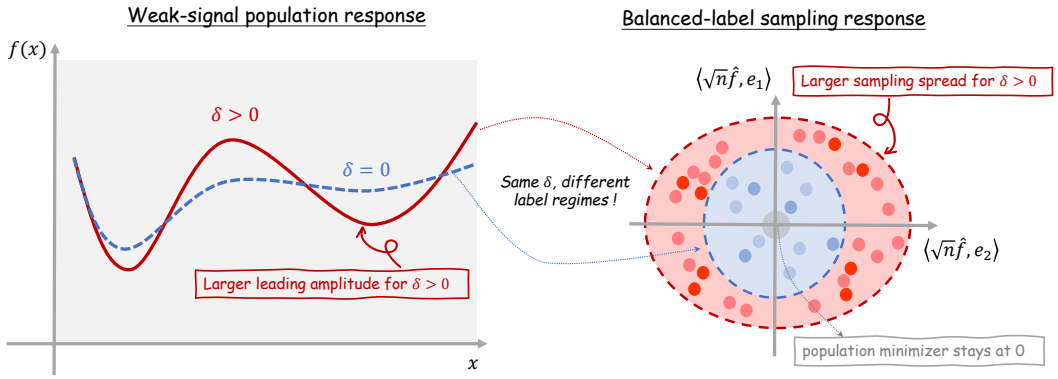

Figure 1: **The same perturbation strength, two label settings.** For normalized kernels and fixed positive ridge, a weak label signal produces a larger leading population response as the radius increases (left). With conditionally balanced labels, the population minimizer remains zero, but the limiting covariance of $\sqrt{n}\hat{f}_{n,\delta}$ grows (right; two RKHS coordinates). Colored guides pair radii across the different label laws, not a map from one panel to the other. Curves and contours are schematic.

label fluctuations, or nonlinear components generated during fitting. Nor does it establish a reduction in function norm or predictive risk. Understanding samplewise perturbation therefore requires separating *what is amplified, what is attenuated, and what is actually beneficial*. Figure 1 illustrates this distinction for population signal and sampling variability under two different label settings. We ask: *How does samplewise perturbation affect population signal and sampling noise? Which leading responses can ordinary ridge logistic regression with output scaling reproduce? What nonlinear changes distinguish the two rules after matching, and what do these changes imply for prediction?* Our goal is to move beyond an aggregate description of "stronger regularization" toward a precise account of signal response, sampling uncertainty, and nonlinear structure, while assessing predictive value separately.

We address these questions in four steps. **First**, an exact margin reduction turns the samplewise inner maximization into a smooth, ridge-regularized logistic objective. Independent per-example parameter perturbations of normalized fixed-feature heads realize the same objective. We can therefore study the minimizers directly, separating estimator response from optimization dynamics. **Second**, we identify a single transfer operator governing weak-signal population response and balanced-label limiting covariance. With normalized kernels, increasing the radius amplifies every nonzero leading signal coefficient in magnitude and every observable covariance mode. In the balanced setting the population minimizer nevertheless stays zero: population response and sampling stability are different questions. **Third**, we test whether these responses are distinctive of the samplewise objective. They are not. Ordinary logistic regression, with an analytically chosen ridge and output scale, matches the entire weak-signal map and balanced-label limiting covariance. We therefore examine the next nonlinear order. A general cubic interaction operator gives the first possible mechanism-specific difference after this matching, with a fifth-order remainder. In a periodic weak-signal model, its $3k$ projection has the opposite local sign for the samplewise estimator and its matched ordinary-logistic control. This attenuation can coexist with growth of the fundamental component and total RKHS norm. Moreover, the Bayes logistic score already contains the same harmonic, so attenuating it is not automatically beneficial. **Finally**, explicit covariance-error bounds and independent kernel and frozen ResNet/ViT fits quantify finite-sample accuracy. Ordinary logistic controls also reproduce the fixed-setting task gains. The central insight is that *matching leading signal and noise responses does not determine how a learning rule changes nonlinear structure, or whether prediction improves.*

Contributions. We make three main contributions.

- **A common signal–noise transfer law.** We characterize weak-signal response and balanced-label limiting covariance by one operator. For normalized kernels, its modal gains grow with radius even though the balanced population minimizer stays zero.
- **First-order matching and general cubic separation.** An analytic ridge-and-scale control matches both leading responses. A general operator expansion identifies the cubic difference between the two fitted rules, while the periodic corollary proves opposite local third-harmonic responses. This separates component attenuation from global norm reduction and predictive benefit.
- **Finite-sample accuracy and structural boundaries.** Separate $O(n^{-1})$ covariance-error bounds control nonlinear fitting and random design. A nonunit-kernel counterexample and kernel and frozen-feature experiments locate the scope of the response laws and test ordinary-logistic controls.

2 Related Work

Perturbation objectives and granularity. Noise injection and robust optimization connect perturbations to regularization, but the correspondence depends on the loss and uncertainty set (Bishop, 1995; Xu et al., 2009). Input-adversarial training uses an examplewise worst case (Goodfellow et al., 2015; Madry et al., 2018), while virtual adversarial training controls changes in predictive distributions (Miyato et al., 2019). Robustness, accuracy, and sample complexity can trade off (Tsipras et al., 2019; Zhang et al., 2019; Schmidt et al., 2018). For weight perturbations, SAM and ASAM specify different local geometries (Foret et al., 2021; Kwon et al., 2021); m-sharpness, microbatch SAM, and δ-SAM vary how examples share a perturbation (Andriushchenko & Flammarion, 2022; Behdin et al., 2023; Zhou et al., 2022). Our exact samplewise objective is the per-instance choice in a fixed-feature geometry, not a new approximation to the shared adversary.

Sharpness, dynamics, and estimator response. Sharpness and norm-based explanations depend on parameterization and on the quantity being measured (Keskar et al., 2017; Dinh et al., 2017; Neyshabur et al., 2017; Jiang et al., 2020). SAM dynamics have been linked to curvature, oscillations, and stochastic gradient noise (Bartlett et al., 2023; Agarwala & Dauphin, 2023; Wen et al., 2023; Luo et al., 2025); practical updates need not converge to the objective's minimizer (Si & Yun, 2023). Feature-rank and depth-dependent amplification give further mechanisms (Andriushchenko et al., 2023; Moon et al., 2026). Functional SAM instead separates Jacobian and logit statistics (Singh et al., 2025). Its affine specialization is the ordinary loss gradient, whereas the samplewise objective can change the minimizer. Statistical analyses of SAM (Behdin & Mazumder, 2023) complement these algorithmic views. We isolate exact minimizers and ask which response statistics can distinguish their learning rules.

Kernel spectra and regularized estimation. Spectral bias and kernel learning curves distinguish signal alignment from estimation error (Rahaman et al., 2019; Bordelon et al., 2020; Canatar et al., 2021; Simon et al., 2023). NTK and random-feature analyses motivate fixed representations (Jacot et al., 2018; Lee et al., 2019; Rudi & Rosasco, 2017), but do not identify different perturbation objectives. Regularized-kernel Gaussian limits, convex-estimator expansions, and self-concordant analysis supply statistical foundations (Hable, 2012; Bellec & Kuchibhotla, 2019; Bach, 2010; Ostrovskii & Bach, 2021). Inverse-Hessian sensitivity (Koh & Liang, 2017) and algorithmic stability (Bousquet & Elisseeff, 2002) address related but distinct responses. Xie & Huo (2025) study shrinking input-adversarial radii; Xie & Huo (2026) study a noise premium under squared loss and kernel-feature perturbations. Here the prediction shift scales with the evaluation norm rather than the fitted function norm. We derive its common signal–noise operator, nonlinear separation, and separate finite-sample covariance errors.

Loss geometry and calibration. Surrogate-risk guarantees depend on the link, target loss, and hypothesis class (Bartlett et al., 2006; Reid & Williamson, 2010; Bao et al., 2020; Awasthi et al., 2021). Output rescaling can change calibration without changing the classification rule (Guo et al., 2017). Recent theory analyzes calibration through SAM and functional gradients (Tan et al., 2026; Futami & Nitanda, 2026). Our analytic control uses ridge and output scale to match leading responses;

its higher-order discrepancy and the Bayes-score expansion show why component attenuation alone does not order predictive risks.

3 An Exact Model of Samplewise Perturbation

Let $\mathcal{H}$ be a real separable RKHS with strongly measurable section map $x \mapsto K_x = K(x, \cdot)$ and $K(x, x) = 1$. The input law P_X is arbitrary. For iid observations $Z_i = (X_i, Y_i)$, $Y_i \in \{-1, 1\}$, write $P_n h = n^{-1} \sum_i h(Z_i)$, $\ell(t) = \log(1 + e^{-t})$, and $\sigma(t) = (1 + e^{-t})^{-1}$. We study

$$\hat{f}_{n,\delta} = \arg\min_{f \in \mathcal{H}} \left\{ P_n \sup_{\|\nu\|_{\mathcal{H}} \le \delta} \ell\big(Y(f + \nu)(X)\big) + \frac{\tau}{2}\|f\|_{\mathcal{H}}^2 \right\}, \qquad \tau > 0. \tag{1}$$

Each loss chooses its own perturbation; ridge penalizes the unperturbed predictor and stays outside the supremum. This placement makes the inner problem depend on a single evaluation.

Proposition 1 (Exact margin reduction). *For every f, x, y and $\delta \ge 0$,*

$$\sup_{\|\nu\|_{\mathcal{H}} \le \delta} \ell\big(y(f + \nu)(x)\big) = \ell\big(yf(x) - \delta\big). \tag{2}$$

The empirical and population objectives are smooth and τ-strongly convex, with unique minimizers.

Indeed, $y\nu(x) \ge -\delta$ by the reproducing property, equality is attained by $\nu = -\delta y K_x$, and ℓ is decreasing. Thus the adversary subtracts a fixed amount from each margin. It does not directly penalize spatial derivatives or suppress particular frequencies. To first order in radius it adds $\delta P_n \sigma(-Yf(X))$, the mean single-observation gradient norm; a shared perturbation instead adds $\delta \|P_n \nabla_f \ell(Yf(X))\|_{\mathcal{H}}$. Both expansions have remainder at most $\delta^2/8$.

Why this model represents an existing objective. For fixed features $\psi(x)$ of unit norm, the head $f_w(x) = \langle w, \psi(x) \rangle$ induces $K(x, z) = \langle \psi(x), \psi(z) \rangle$. Independent perturbations $\|a_i\| \le \delta$ of the head weights have exactly the loss in (2). The RKHS norm is the minimum representing weight norm, so positive ridge gives identical minimizers (Proposition A.3). This is an exact fixed-feature realization of per-instance weight perturbation, rather than a linearization of a changing representation.

Separate signal from input geometry. Fix the kernel and P_X, and vary only the label law through $\mathbb{E}[Y \mid X] = \epsilon q(X)$, where q is bounded and $|\epsilon| \|q\|_\infty < 1$. Write $f_{\epsilon,\delta}$ for the population minimizer. The evaluation map $\mathcal{J} : \mathcal{H} \to L^2(P_X)$ gives $T = \mathcal{J}^* \mathcal{J} = \mathbb{E}[K_X \otimes K_X]$, with $\mathrm{tr}(T) = 1$. Weak-signal expansions keep τ, K, P_X, q and a finite radius ceiling D fixed; sampling limits keep the label law fixed. The case $\epsilon = 0$ isolates sampling variability without population signal.

4 What Leading Signal and Noise Responses Reveal

Set $s_\delta = \sigma(\delta)$, $p_\delta = \sigma'(\delta)$, and define

$$\mathcal{D}_\delta = s_\delta (\tau I + p_\delta T)^{-1} \mathcal{J}^*. \tag{3}$$

This operator combines the strength of the label score, s_δ, with the inverse regularized curvature. We first identify its population meaning, then its sampling meaning, and finally ask whether it identifies the learning rule.

Theorem 2 (Weak-signal response). *Under Section 3's assumptions, for sufficiently small $|\epsilon|$,*

$$\|f_{\epsilon,\cdot} - \epsilon \mathcal{D}_\cdot q\|_{C^1([0,D];\mathcal{H})} \le C|\epsilon|^3.$$

Here the C^1 norm controls both the function and its radius derivative. In a prediction coordinate associated with a positive eigenvalue μ of T, the leading signal multiplier is

$$A_\delta(\mu) = \frac{\mu s_\delta}{\tau + \mu p_\delta}, \qquad A_\delta(0) = 0. \tag{4}$$

The zero-signal score has derivative $\tau I + p_\delta T$; the label perturbation contributes $s_\delta \mathcal{J}^* q$. Inverting the former gives the latter's response. Label-sign symmetry removes even powers of ϵ, and strong convexity converts the remaining score residual into the stated error (Appendix A). The derivative control matters: an error bound on function values alone would not justify signs of radius responses.

Amplification is not uniform rescaling. For $\mu > 0$,

$$\partial_\delta A_\delta(\mu) = \frac{\mu p_\delta(\tau + \mu s_\delta^2)}{(\tau + \mu p_\delta)^2} > 0, \qquad \frac{A_\delta(\mu)}{A_0(\mu)} = \frac{2s_\delta(\tau + \mu/4)}{\tau + \mu p_\delta}. \tag{5}$$

Thus every nonzero leading coefficient grows in magnitude. The gain ratio also increases with μ for $\delta > 0$: large-eigenvalue directions gain more *relatively*, although small-eigenvalue signals also grow *absolutely*. For a general kernel, these directions need not be spatial frequencies.

Theorem 3 (Balanced-label sampling response)**.** *If $\mathbb{E}[Y \mid X] = 0$, then $f_{0,\delta} = 0$ for every $\delta \geq 0$. At each fixed radius, $\sqrt{n}\hat{f}_{n,\delta}$ converges weakly in $\mathcal{H}$ to a centered Gaussian with covariance*

$$C_\delta = \mathcal{D}_\delta \mathcal{D}_\delta^* = s_\delta^2(\tau I + p_\delta T)^{-1} T (\tau I + p_\delta T)^{-1}. \tag{6}$$

Its eigenvalue in a positive T-mode is

$$V_\delta(\mu) = \frac{A_\delta(\mu)^2}{\mu}, \tag{7}$$

which strictly increases with radius. Covariance vanishes on $\ker T$; both C_δ and $\mathcal{J} C_\delta \mathcal{J}^*$ are trace class. For every deterministic nonnegative $\delta_n \to 0$, same-sample estimators satisfy $\|\hat{f}_{n,\delta_n} - \hat{f}_{n,0}\| = O_p(\delta_n n^{-1/2})$.

At balance, the empirical label average $P_n(YK_X)$ fluctuates around zero. The same inverse-curvature operator that transmits a weak population signal transmits this fluctuation. Its covariance has the sandwich form (6), not simply the inverse Hessian. Hence a population predictor can remain fixed while repeated fitted predictors become more variable. Equations (4)–(7) describe two different label regimes linked by one operator, as in Figure 1.

Reading the same mode in function and prediction coordinates. Let $Te = \mu e$ with $\|e\|_{\mathcal{H}} = 1$ and $\mu > 0$, and put $\phi = \mathcal{J}e/\sqrt{\mu}$, so $\|\phi\|_{L^2(P_X)} = 1$. If G_δ is the limiting Gaussian in Theorem 3, then

$$\langle \mathcal{J} f_{\epsilon,\delta}, \phi \rangle = \epsilon A_\delta(\mu) \langle q, \phi \rangle + O(\epsilon^3), \qquad \mathrm{Var}\langle \mathcal{J} G_\delta, \phi \rangle = \mu V_\delta(\mu) = A_\delta(\mu)^2.$$

The factor μ converts an RKHS coordinate to a prediction coordinate. Consequently the common transfer law is not a claim that signal amplitude and variance have the same units: signal is multiplied by A_δ, whereas prediction variance is multiplied through its square. This coordinate distinction also fixes the interpretation of the spectral probes in the experiments.

Why does variance grow at an unchanged population solution? At $f = 0$, the balanced-label score is $-s_\delta Y K_X$ and its covariance is $s_\delta^2 T$, while the regularized Hessian is $\tau I + p_\delta T$. As radius increases, s_δ increases and p_δ decreases from $1/4$. The individual label scores become stronger while their curvature response weakens. Both effects increase $V_\delta(\mu)$, although opposite labels still cancel exactly in the population score. The variance increase therefore comes from how empirical imbalances are transmitted, not from motion of the population optimum. This distinguishes uncertainty across repeated datasets from sensitivity of the population solution itself.

Can ordinary logistic reproduce both responses? Write the ridge as a superscript and define the post-scaled ordinary fit $g_{\epsilon,\delta} = \gamma_\delta f_{\epsilon,0}^{(\lambda_\delta)}$.

Proposition 4 (Analytic response matching)**.** *For a normalized kernel, choose*

$$\lambda_\delta = \frac{\tau}{4p_\delta}, \qquad \gamma_\delta = \frac{s_\delta}{2p_\delta}. \tag{8}$$

Then

$$\mathcal{D}_\delta^{(\tau)} = \gamma_\delta \mathcal{D}_0^{(\lambda_\delta)}, \qquad C_\delta^{(\tau)} = \gamma_\delta^2 C_0^{(\lambda_\delta)}.$$

The second identity is for conditionally balanced labels.

The same two scalars match *all* signal directions and the whole limiting covariance, without fitting individual modes. Indeed, $\lambda_\delta I + T/4 = (\tau I + p_\delta T)/(4p_\delta)$, so the inverse operators differ by a scalar; output scaling supplies the remaining factor. Notably $\lambda_\delta \geq \tau$: the control increases ridge and then rescales its scores. For fixed δ, the complete population fits differ by $O_{\mathcal{H}}(\epsilon^3)$, rather than being identical (Appendix B). This leads to the next question: *what happens at precisely the nonlinear order discarded by matching?*

Proposition 5 (General third-order separation after matching)**.** *Under the general normalized-kernel assumptions, define the cubic response $u_{3,\delta}^{(r)}$ at ridge r as in Appendix A.3, and set*

$$\Delta_\delta[q] = u_{3,\delta}^{(\tau)} - \gamma_\delta u_{3,0}^{(\lambda_\delta)}. \tag{9}$$

Then

$$f_{\epsilon,\delta}^{(\tau)} - \gamma_\delta f_{\epsilon,0}^{(\lambda_\delta)} = \epsilon^3 \Delta_\delta[q] + O_{C^1([0,D];\mathcal{H})}(\epsilon^5). \tag{10}$$

The first-order responses match exactly, while the first possible mechanism-specific difference is the operator $\Delta_\delta[q]$. At a probe $\ell \in \mathcal{H}^*$, a nonzero cubic separation requires $\ell(\Delta_\delta[q]) \neq 0$. The periodic harmonic result below is the Fourier specialization of this statement.

The proposition is an equality of local expansions, not a universal sign or a prediction-risk ordering. In particular, $\Delta_0[q] = 0$ and special kernels, signals, or probes can make a selected cubic projection vanish. For the logistic loss, its local nondegeneracy is explicit:

$$\partial_\delta \Delta_0[q] = \left(\tau I + \tfrac{1}{4}T\right)^{-1} \mathcal{J}^* \left[-\tfrac{1}{16} q h_0^2 + \tfrac{1}{48} h_0^3\right], \qquad h_0 = \mathcal{J}\left[\tfrac{1}{2}\left(\tau I + \tfrac{1}{4}T\right)^{-1} \mathcal{J}^* q\right]. \tag{11}$$

Thus a probe with nonzero projection of (11) separates the two rules linearly in the small radius.

5 What First Order Leaves Undetermined

To expose a generated component, specialize to a population in which input geometry does not mix frequencies. Let X be uniform on $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$, with $m(dx) = dx/(2\pi)$, and

$$K(x, x') = \mu_0 + 2\sum_{j\geq 1} \mu_j \cos(j(x - x')), \qquad \mu_j \geq 0, \quad \mu_0 + 2\sum_{j\geq 1} \mu_j = 1. \tag{12}$$

Assume infinitely many coefficients are positive and take $q(x) = \cos(kx)$, with fixed $k \geq 1$ and $\mu_k, \mu_{3k} > 0$. A cosine coefficient $c_j(f) = 2\int f(x)\cos(jx)\, dm(x)$ contributes $E_j(f) = c_j(f)^2/(2\mu_j)$ to $\|f\|_{\mathcal{H}}^2$ and $c_j(f)^2/2$ to prediction energy. These distinguish a component's amplitude from its norm cost.

This specialization makes the source of each component identifiable within the model. The input law and translation-invariant kernel diagonalize the linear response in the same Fourier basis. Consequently, a frequency absent from the label mean cannot enter through linear mixing. A $3k$ component must instead come from nonlinear terms in the score. The example is useful precisely because it separates these two routes to an apparently new frequency before comparing the fitted rules.

The appendix first gives a mode-free third-order interaction expansion with a fifth-order C^1 remainder. Its interaction source is

$$\zeta_\delta = \frac{\sigma''(\delta)}{2} q h_\delta^2 - \frac{\sigma'''(\delta)}{6} h_\delta^3, \qquad h_\delta = \mathcal{J} u_{1,\delta},$$

under the general normalized-kernel assumptions. The periodic model below then makes this source observable: for $q = \cos(kx)$ it is proportional to $\cos^3(kx)$, so its only cubic modes are k and $3k$.

Theorem 6 (Cubic population response)**.** *Set $a_\delta = A_\delta(\mu_k)$ and*

$$B_\delta = \frac{\sigma''(\delta) a_\delta^2}{8} - \frac{\sigma'''(\delta) a_\delta^3}{24}, \quad b_{1,\delta} = \frac{3\mu_k B_\delta}{\tau + \mu_k p_\delta}, \quad b_{3,\delta} = \frac{\mu_{3k} B_\delta}{\tau + \mu_{3k} p_\delta}. \tag{13}$$

Uniformly for $\delta \in [0, D]$ and sufficiently small $|\epsilon|$,

$$f_{\epsilon,\delta} = \epsilon a_\delta \cos(kx) + \epsilon^3 \{b_{1,\delta}\cos(kx) + b_{3,\delta}\cos(3kx)\} + r_{\epsilon,\delta}, \tag{14}$$

where $\|r_{\epsilon,\cdot}\|_{C^1([0,D];\mathcal{H})} \leq C|\epsilon|^5$. Constants depend on the fixed kernel, τ, D, k, but not on ϵ, δ.

The linear response inherits only frequency k. The cubic score contains $\cos^3(kx) = (3\cos(kx) + \cos(3kx))/4$, creating the $3k$ component. Appendix C.3 constructs a candidate that cancels both orders and transfers its fifth-order residual, including its radius derivative, to the actual minimizer. Thus the harmonic is a consequence of nonlinear fitting, not an extra component placed in the label mean. The general matched-control proposition identifies the corresponding third-order difference before this Fourier projection; the signs below are therefore a geometric corollary rather than a universal frequency law.

A generated component can move against the leading signal. At zero radius,

$$a_0' = \frac{a_0}{2}, \qquad b_{3,0} = \frac{\mu_{3k} a_0^3}{192(\tau + \mu_{3k}/4)} > 0, \qquad b_{3,0}' = \frac{\mu_{3k} a_0^2 (a_0 - 2)}{128(\tau + \mu_{3k}/4)} < 0, \tag{15}$$

since fixed positive ridge gives $0 < a_0 < 2$. The C^1 remainder turns these coefficient signs into statements about component energy.

Corollary 7 (Attenuation without norm reduction)**.** *For sufficiently small nonzero ϵ,*

$$\partial_\delta E_k(f_{\epsilon,\delta})\big|_0 > 0, \qquad \partial_\delta E_{3k}(f_{\epsilon,\delta})\big|_0 < 0, \qquad \partial_\delta \|f_{\epsilon,\delta}\|_{\mathcal{H}}^2\big|_0 > 0.$$

Each sign persists on a sufficiently small positive-radius interval.

The leading positive norm derivative is $\epsilon^2 a_0^2/(2\mu_k)$; the negative third-mode derivative is of order ϵ^6. A smaller generated harmonic can therefore coexist with a larger total function norm. The result concerns a local change, not suppression of all high frequencies at all radii.

Corollary 8 (Matched responses, opposite harmonic changes)**.** *Under the same periodic assumptions, let $g_{\epsilon,\delta} = \gamma_\delta f_{\epsilon,0}^{(\lambda_\delta)}$ use* (8)*. For sufficiently small nonzero ϵ,*

$$\partial_\delta E_{3k}(g_{\epsilon,\delta})\big|_{0^+} = E_{3k}(f_{\epsilon,0}^{(\tau)}) > 0, \qquad \partial_\delta E_{3k}(f_{\epsilon,\delta}^{(\tau)})\Big|_{0^+} < 0.$$

Both paths start at the same ordinary fit. But $\lambda_0' = 0$ and $\gamma_0' = 1/2$, so the control initially scales the entire fit, increasing its third-harmonic energy. Samplewise perturbation instead changes the cubic score and decreases that energy. This is why the matched control is informative: *even after matching the full leading signal and noise responses, the two rules change a nonlinear component in opposite directions.*

Why attenuation is not a prediction guarantee. The Bayes logistic score for this label law is

$$\log \frac{1 + \epsilon\cos(kx)}{1 - \epsilon\cos(kx)} = 2\epsilon\cos(kx) + \frac{\epsilon^3}{2}\cos(kx) + \frac{\epsilon^3}{6}\cos(3kx) + O(\epsilon^5). \tag{16}$$

Thus the optimal unpenalized score already contains the third harmonic. Its attenuation alone orders neither prediction risk nor global smoothness. The periodic construction separates inherited signal from loss-generated structure; under nonuniform inputs, frequency mixing can also occur at linear order. Appendix E gives a local-radius certificate, and Appendix F.6 examines this change of input geometry.

6 Finite-Sample Accuracy and the Role of Normalization

The limiting covariance explains a response but not the accuracy of a finite-sample approximation. Return to general normalized kernels and balanced labels, and condition first on the inputs. Define

$$T_n = P_n(K_X \otimes K_X), \qquad H_{n,\delta} = \tau I + p_\delta T_n,$$
$$\widetilde{f}_{n,\delta} = s_\delta H_{n,\delta}^{-1} P_n(Y K_X), \qquad C_{n,\delta} = s_\delta^2 H_{n,\delta}^{-1} T_n H_{n,\delta}^{-1}.$$

The linearized fit has exact conditional covariance $C_{n,\delta}$ after $\sqrt{n}$ scaling. It separates nonlinear fitting from randomness of the input operator.

Theorem 9 (Two sources of covariance error)**.** *For every $n \geq 1$ and fixed $\delta \geq 0$, the actual and linearized estimators have conditional mean zero. There is an explicit constant $\mathfrak{B}_\delta$, depending only on δ, τ (given explicitly in Appendix D.1), such that $(\mathbb{E}[\|\hat{f}_{n,\delta} - \widetilde{f}_{n,\delta}\|^2 \mid X_{1:n}])^{1/2} \leq \mathfrak{B}_\delta n^{-3/2}$, and, in trace norm,*

$$\|\operatorname{Cov}(\sqrt{n}\hat{f}_{n,\delta}) - \mathbb{E}C_{n,\delta}\|_1 \leq \frac{2 s_\delta \mathfrak{B}_\delta}{\tau n} + \frac{\mathfrak{B}_\delta^2}{n^2}, \tag{17}$$

$$\|\mathbb{E}C_{n,\delta} - C_\delta\|_1 \leq \frac{4 s_\delta^2 p_\delta (1 - \operatorname{tr}(T^2))}{\tau^3 n}. \tag{18}$$

Equation (17) measures fitting nonlinearity at fixed inputs; (18) measures averaging over random designs. The $1/n$ covariance rate comes from two cancellations: balanced-label symmetry controls signed nonlinear terms, and the centered design expansion removes its linear term in expectation. Appendix D.1 gives the constant, proof, and a sharper conditional bound using the observed spectrum. These are absolute errors: a positive limiting variance gap guarantees a finite-sample increase only when it exceeds the errors at both radii. This also motivates the paired linearized references used in the experiments.

An observable decomposition. For a fixed RKHS probe h, write $U = \sqrt{n}\langle \hat{f}_{n,\delta}, h\rangle$, $V = \sqrt{n}\langle \widetilde{f}_{n,\delta}, h\rangle$, $c_n = \mathbb{E}[V^2 \mid X_{1:n}]$, and $c = \langle h, C_\delta h\rangle$. The exact identity

$$U^2 - c = (U^2 - V^2) + (V^2 - c_n) + (c_n - c)$$

separates nonlinear fitting, label Monte Carlo fluctuation, and random design. The middle term has zero conditional expectation. Pairing exact and linearized fits therefore exposes nonlinear error without confusing it with label fluctuation; computing c_n analytically isolates the design term. This is the experimental counterpart of the two trace-norm bounds.

For a bounded nonunit kernel, the exact shift is $\delta\kappa(x)$ with $\kappa(x) = \sqrt{K(x,x)}$. The signal expansion and $C_\delta = \mathcal{D}_\delta \mathcal{D}_\delta^*$ survive with input-dependent weights $\sigma(\delta\kappa(x))$ and $\sigma'(\delta\kappa(x))$. Every observable direction still has positive covariance derivative at zero, but global covariance order can fail: three equally likely features $(-5, 6), (-1, 4), (-6, 0)$ with $\tau = 1/250$ satisfy $v^\top C_2' v < -258510$ for $v = (-5, -3)$, certified by interval arithmetic (Appendix E). Normalization therefore controls global monotonicity, not the existence of a common response operator. For fixed nonzero periodic signal, same-sample drift has remainder $O_p(\delta_n n^{-1/2} + \delta_n^2)$, giving the $n^{-1/2}$ critical radius when the population drift is nonzero. At balance that drift is exactly zero, and Theorem 3's paired bound holds at every shrinking rate. Appendix D.3 develops these complementary sampling limits.

7 Experiments

The experiments follow the theory's comparisons: matched leading response, nonlinear separation, finite-sample covariance error, and predictive controls. Every displayed fit optimizes its own objective. Periodic fits use the infinite-rank kernel $K(t) = (1 + 2\sum_{j\geq 1} j^{-4}\cos jt)/(1 + \pi^4/45)$; its closed polynomial form avoids Fourier truncation. Unless stated otherwise, $\tau = .02$. Appendices F and G give all configurations, solver tolerances, and independent numerical checks. Our code is available at `https://anonymous.4open.science/r/2026sam-B3ED/README.md`

Can matched leading responses conceal opposite structure? Figure 2(a–c) uses independent population fits on a 1,024-point grid at four signal strengths. Fundamental gains nearly coincide, yet the generated third-harmonic coefficient locally decreases under samplewise perturbation and increases under its matched control. At $\epsilon = .1, \delta = .25$, their coefficients are 3.1132×10^{-6} and 4.3919×10^{-6}, despite only $.0367\%$ relative L^2 disagreement between the full predictors. Close fitted functions can therefore conceal opposite component responses. Absolute and relative mismatch slopes of 3.0006 and 1.9989 agree with cubic and quadratic orders; doubling the grid changes third coefficients by at most 1.80×10^{-13}. Full-radius paths show that the sign claim is local: the samplewise coefficient later reverses and grows (Appendix F.5).

Does the transfer law survive beyond periodic geometry? Frozen ImageNet-pretrained ResNet-18 and ViT-B/16 heads exhibit the predicted signal gains and covariance inflation across ranks, weak-signal strengths, and radii; all 18 balanced-label variance-ratio intervals exceed one. This supports the transfer law beyond Fourier geometry while the features remain fixed. The RBF check separates finite-sample covariance errors due to nonlinear fitting and random design (Figure 2(d–f)).

Are task-level gains specific to samplewise perturbation? At the fixed samplewise setting, ordinary controls often achieve lower test log loss across three binary CIFAR-10 tasks and both backbones (Appendix G.4). A gain over basic ERM in that setting is therefore not perturbation-specific evidence. Under five independent splits and expanded grids, samplewise-minus-ERM differences remain mixed; tuned controls lie within 0.006 log-loss on five of six backbone/task combinations.

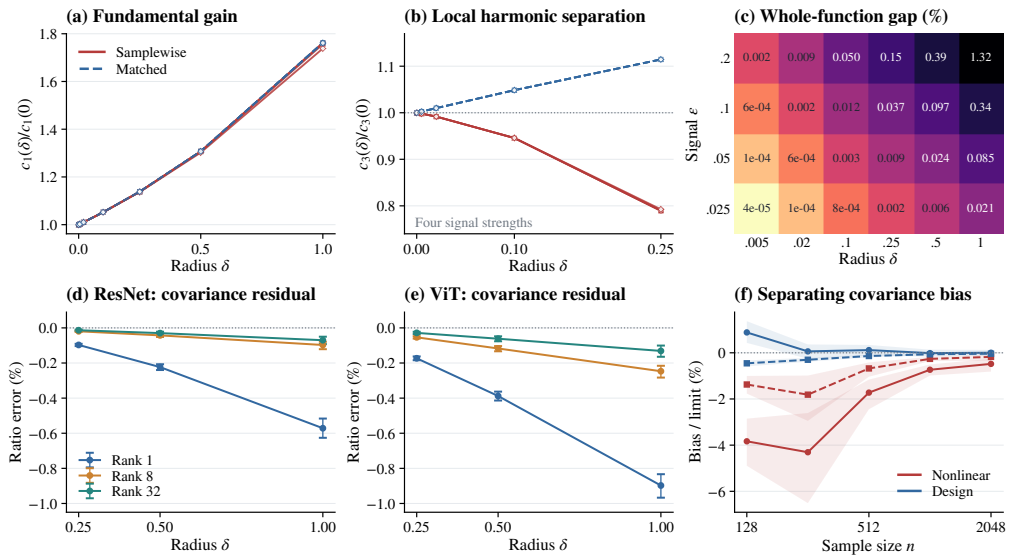

Figure 2: **Leading agreement, nonlinear separation, and finite-sample error.** (a,b) Independent samplewise and matched population fits at $\epsilon \in \{.025, .05, .1, .2\}$; third coefficients are normalized by their common positive zero-radius value. (c) Full-function relative L^2 differences (%; logarithmic color scale), all 24 positive-radius settings. (d,e) Balanced-label variance-ratio error, $100(\text{observed}/\text{predicted} - 1)$; zero denotes the conditional linearized prediction. (f) RBF constant-probe covariance bias; solid/dashed paths use $\tau = .02/.08$. Intervals and shading are pointwise 95% paired-bootstrap intervals over 100 independent replicates; (a–c) are deterministic quadrature, without sampling bands.

Several selected ridges hit the 10^{-5} boundary, so this comparison probes tuning sensitivity rather than establishing a general performance ordering. Validation curves and the centered cubic-residual diagnostic appear in Appendix G.4.

Which conclusions depend on geometry or the evaluation metric? Nonuniform inputs can reverse the third-frequency response, so its attenuation is geometry-dependent. Matched controls change only .12%–1.02% of classifications despite different log loss (Appendix G.4); classification agreement therefore does not settle the log-loss comparison.

8 Discussion and Conclusion

Samplewise function perturbation reduces exactly to a logistic margin shift, but its fitted predictor has a nontrivial response. For normalized kernels and fixed positive ridge, increasing the radius amplifies every nonzero leading signal response and the balanced-label limiting covariance, even though the balanced-label population minimizer remains zero. Finite-sample bounds quantify the effects of nonlinear fitting and random input design; a nonunit-diagonal counterexample shows where global covariance monotonicity can fail. These leading responses do not identify the learning rule. An ordinary logistic fit with adjusted ridge and output scale matches both the weak-signal operator and balanced-label covariance. The first possible difference appears at cubic order, characterized by $\Delta_\delta[q]$ with a fifth-order remainder. Its periodic third-harmonic sign is one consequence of this general result, also tested through independent fits in nonperiodic RBF geometry.

Finally, attenuation of a generated component need not reduce the RKHS norm or predictive risk: in the periodic example, the fundamental and total norm increase, and the Bayes score itself contains the attenuated harmonic. The resulting lesson is to evaluate signal, sampling variability, nonlinear structure, and prediction separately. Our conclusions concern exact minimizers with per-example perturbations and fixed features; shared-perturbation training and changing representations remain open.

AI USE STATEMENT

AI-assisted tools were used for language polishing, grammar checking, and literature retrieval and discovery (e.g., identifying relevant related work). The authors take full responsibility for the content and integrity of the manuscript.

REFERENCES

Atish Agarwala and Yann Dauphin. SAM operates far from home: eigenvalue regularization as a dynamical phenomenon. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 152–168. PMLR, 2023. URL https://proceedings.mlr.press/v202/agarwala23a.html.

Maksym Andriushchenko and Nicolas Flammarion. Towards understanding sharpness-aware minimization. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 639–668. PMLR, 2022. URL https://proceedings.mlr.press/v162/andriushchenko22a.html.

Maksym Andriushchenko, Dara Bahri, Hossein Mobahi, and Nicolas Flammarion. Sharpness-aware minimization leads to low-rank features. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/92dd1adab39f362046f99dfe3c39d90f-Abstract-Conference.html.

Pranjal Awasthi, Natalie Frank, Anqi Mao, Mehryar Mohri, and Yutao Zhong. Calibration and consistency of adversarial surrogate losses. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/514a70448c235ccb8b6842ef5e02ad3b-Abstract.html.

Francis Bach. Self-concordant analysis for logistic regression. *Electronic Journal of Statistics*, 4: 384–414, 2010. doi: 10.1214/09-EJS521. URL https://arxiv.org/abs/0910.4627.

Han Bao, Clay Scott, and Masashi Sugiyama. Calibrated surrogate losses for adversarially robust classification. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 408–451. PMLR, 2020. URL https://proceedings.mlr.press/v125/bao20a.html.

Peter L. Bartlett, Michael I. Jordan, and Jon D. McAuliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156, 2006. doi: 10.1198/016214505000000907. URL https://doi.org/10.1198/016214505000000907.

Peter L. Bartlett, Philip M. Long, and Olivier Bousquet. The dynamics of sharpness-aware minimization: Bouncing across ravines and drifting towards wide minima. *Journal of Machine Learning Research*, 24(316):1–36, 2023. URL https://jmlr.org/papers/v24/23-043.html.

Kayhan Behdin and Rahul Mazumder. On statistical properties of sharpness-aware minimization: Provable guarantees. *arXiv preprint arXiv:2302.11836*, 2023. URL https://arxiv.org/abs/2302.11836.

Kayhan Behdin, Qingquan Song, Aman Gupta, Sathiya Keerthi, Ayan Acharya, Borja Ocejo, Gregory Dexter, Rajiv Khanna, David Durfee, and Rahul Mazumder. mSAM: Micro-batch-averaged sharpness-aware minimization. *arXiv preprint arXiv:2302.09693*, 2023. URL https://arxiv.org/abs/2302.09693.

Pierre Bellec and Arun Kuchibhotla. First order expansion of convex regularized estimators. In *Advances in Neural Information Processing Systems*, volume 32, pp. 3457–3468, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/0609154fa35b3194026346c9cac2a248-Abstract.html.

Patrick Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, second edition, 1999. doi: 10.1002/9780470316962. URL https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316962.

Christopher M. Bishop. Training with noise is equivalent to Tikhonov regularization. *Neural Computation*, 7(1):108–116, 1995. doi: 10.1162/neco.1995.7.1.108. URL https://doi.org/10.1162/neco.1995.7.1.108.

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 of *Proceedings of Machine Learning Research*, pp. 1024–1034. PMLR, 2020. URL https://proceedings.mlr.press/v119/bordelon20a.html.

Olivier Bousquet and André Elisseeff. Stability and generalization. *Journal of Machine Learning Research*, 2:499–526, 2002. URL https://www.jmlr.org/papers/v2/bousquet02a.html.

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*, 12:2914, 2021. doi: 10.1038/s41467-021-23103-1. URL https://arxiv.org/abs/2006.13198.

Laurent Dinh, Razvan Pascanu, Samy Bengio, and Yoshua Bengio. Sharp minima can generalize for deep nets. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1019–1028. PMLR, 2017. URL https://proceedings.mlr.press/v70/dinh17b.html.

Pierre Foret, Ariel Kleiner, Hossein Mobahi, and Behnam Neyshabur. Sharpness-aware minimization for efficiently improving generalization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=6Tm1mposlrM.

Futoshi Futami and Atsushi Nitanda. Smooth calibration error: Uniform convergence and functional gradient analysis. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/00a5213b3479c9733eb3743459badc6b-Abstract-Conference.html.

Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6572.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2017. URL https://proceedings.mlr.press/v70/guo17a.html.

Robert Hable. Asymptotic normality of support vector machine variants and other regularized kernel methods. *Journal of Multivariate Analysis*, 106:92–117, 2012. doi: 10.1016/j.jmva.2011.11.004. URL https://arxiv.org/abs/1010.0535.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/5a4be1fa34e62bb8a6ec6b91d2462f5a-Abstract.html.

Yiding Jiang, Behnam Neyshabur, Hossein Mobahi, Dilip Krishnan, and Samy Bengio. Fantastic generalization measures and where to find them. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/1912.02178.

Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. In *International Conference on Learning Representations*, 2017. URL https://arxiv.org/abs/1609.04836.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1885–1894. PMLR, 2017. URL https://proceedings.mlr.press/v70/koh17a.html.

Jungmin Kwon, Jeongseop Kim, Hyunseo Park, and In Kwon Choi. ASAM: Adaptive sharpness-aware minimization for scale-invariant learning of deep neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5905–5914. PMLR, 2021. URL https://proceedings.mlr.press/v139/kwon21b.html.

Jaehoon Lee, Lechao Xiao, Samuel S. Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/0d1a9651497a38d8b1c3871c84528bd4-Abstract.html.

Haocheng Luo, Mehrtash Harandi, Dinh Phung, and Trung Le. Unveiling m-sharpness through the structure of stochastic gradient noise. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2509.18001.

Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJzIBfZAb.

Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1979–1993, 2019. doi: 10.1109/TPAMI.2018.2858821. URL https://arxiv.org/abs/1704.03976.

Chaewon Moon, Dongkuk Si, and Chulhee Yun. Minor first, major last: A depth-induced implicit bias of sharpness-aware minimization. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/1c492d4247429a075cf2c9de9026271d-Abstract-Conference.html.

Behnam Neyshabur, Srinadh Bhojanapalli, David McAllester, and Nathan Srebro. Exploring generalization in deep learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1706.08947.

Dmitrii M. Ostrovskii and Francis Bach. Finite-sample analysis of M-estimators using self-concordance. *Electronic Journal of Statistics*, 15(1):326–391, 2021. doi: 10.1214/20-EJS1780. URL https://arxiv.org/abs/1810.06838.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5301–5310. PMLR, 2019. URL https://proceedings.mlr.press/v97/rahaman19a.html.

Mark D. Reid and Robert C. Williamson. Composite binary losses. *Journal of Machine Learning Research*, 11(83):2387–2422, 2010. URL https://www.jmlr.org/papers/v11/reid10a.html.

Alessandro Rudi and Lorenzo Rosasco. Generalization properties of learning with random features. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/61b1fb3f59e28c67f3925f3c79be81a1-Abstract.html.

Ludwig Schmidt, Shibani Santurkar, Dimitris Tsipras, Kunal Talwar, and Aleksander Madry. Adversarially robust generalization requires more data. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/f708f064faaf32a43e4d3c784e6af9ea-Abstract.html.

Dongkuk Si and Chulhee Yun. Practical sharpness-aware minimization cannot converge all the way to optima. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/5305b7891e1098dd9773d35cd9333180-Abstract-Conference.html.

James B. Simon, Madeline Dickens, Dhruva Karkada, and Michael R. DeWeese. The eigenlearning framework: A conservation law perspective on kernel ridge regression and wide neural networks. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=FDbQGCAViI.

Sidak Pal Singh, Hossein Mobahi, Atish Agarwala, and Yann Dauphin. Avoiding spurious sharpness minimization broadens applicability of SAM. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 55702–55719. PMLR, 2025. URL https://proceedings.mlr.press/v267/singh25b.html.

Chengli Tan, Yubo Zhou, Haishan Ye, Guang Dai, Junmin Liu, Zengjie Song, Jiangshe Zhang, Zixiang Zhao, Yunda Hao, and Yong Xu. Towards understanding the calibration benefits of sharpness-aware minimization. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/8ce0b141fb61bb847898d0b6258f9acf-Abstract-Conference.html.

Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, and Aleksander Madry. Robustness may be at odds with accuracy. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1805.12152.

Kaiyue Wen, Tengyu Ma, and Zhiyuan Li. How does sharpness-aware minimization minimize sharpness? In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=5spDgWmpY6x.

William Wolberg, Olvi Mangasarian, Nick Street, and W. Street. Breast cancer wisconsin (diagnostic). UCI Machine Learning Repository, 1993. URL https://doi.org/10.24432/C5DW2B. Dataset; CC BY 4.0.

Yiling Xie and Xiaoming Huo. Asymptotic behavior of adversarial training estimator under ℓ_∞-perturbation. *Journal of the American Statistical Association*, 120(552):2672–2683, 2025. doi: 10.1080/01621459.2025.2485346. URL https://arxiv.org/abs/2401.15262.

Yiling Xie and Xiaoming Huo. The noise premium in adversarial training for kernel regression. *arXiv preprint arXiv:2607.27995*, 2026. URL https://arxiv.org/abs/2607.27995.

Huan Xu, Constantine Caramanis, and Shie Mannor. Robustness and regularization of support vector machines. *Journal of Machine Learning Research*, 10(51):1485–1510, 2009. URL https://www.jmlr.org/papers/v10/xu09b.html.

Hongyang Zhang, Yaodong Yu, Jiantao Jiao, Eric Xing, Laurent El Ghaoui, and Michael Jordan. Theoretically principled trade-off between robustness and accuracy. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 7472–7482. PMLR, 2019. URL https://proceedings.mlr.press/v97/zhang19p.html.

Wenxuan Zhou, Fangyu Liu, Huan Zhang, and Muhao Chen. Sharpness-aware minimization with dynamic reweighting. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 5686–5699, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-emnlp.417. URL https://aclanthology.org/2022.findings-emnlp.417/.

Appendix

The appendix follows the main argument: common response, analytic matching, nonlinear separation, and finite-sample accuracy. Structural boundaries precede complete experimental protocols. All entries and proof-guide references are clickable.

Contents: theory

A General signal–noise response and fixed-feature geometry . . . 16
A.1 Assumptions and response theorem . . . 16
A.2 Common transfer proof . . . 16
A.3 General cubic interaction response . . . 18
A.4 Fixed-feature correspondence . . . 19

B Analytic matching and the ordinary-logistic control . . . 20
B.1 First-order matching . . . 20
B.2 General cubic separation after matching . . . 20
B.3 Ridge path and nonlinear separation . . . 21

C Periodic signal and nonlinear harmonic response . . . 22
C.1 Assumptions and coordinates . . . 22
C.2 Smooth paths and residual comparison . . . 23
C.3 Cubic expansion and signed energies . . . 26

D Finite-sample covariance and sampling limits . . . 30
D.1 Finite-sample covariance error . . . 30
D.2 Fixed-radius Hilbert limits . . . 32
D.3 Same-sample shrinking radii . . . 34
D.4 Critical radii and zero drift . . . 34

Main-result proof guide

Main-text result	Proof and supporting ingredients
Proposition 1: exact loss	Proposition C.2; fixed features in Proposition A.3
Theorems 2–3: common response	Theorem A.1, p. 16
Proposition 4: matching	Appendix B.1, p. 20
Proposition 5: cubic separation	Proposition B.1, p. 20
Theorem 6: cubic response	Lemmas C.7–C.8 and Propositions C.9 and C.10
Corollaries 7–8: opposite signs	Proposition C.12; Appendix B.3
Theorem 9: covariance errors	Theorem D.1, p. 30

Contents: boundaries and experiments

E Structural boundaries and local-radius certificates **35**
E.1 Nonunit covariance and counterexample 35
E.2 Certified local harmonic radius 37

F Kernel experiments: protocols and complete diagnostics **38**
F.1 Kernel and solver 38
F.2 Configurations and uncertainty 38
F.3 Sampling diagnostics 39
F.4 Shared and samplewise objectives 40
F.5 Matched periodic fits 40
F.6 Nonuniform inputs and RBF kernels 41
F.7 Paired covariance decomposition 44

G Frozen-feature experiments and predictive controls **46**
G.1 Data and features 46
G.2 Signal and label-noise tests 46
G.3 Fixed-setting task behavior 47
G.4 Matched and validation-selected controls 49
G.5 Real-covariate instance 52

H Objective, geometry, and statistical scope **53**

Reading order. For the main identification argument, read Appendices A–C. For statistical approximation, start with the conditional error decomposition in Appendix D.1, then the sampling limits in Appendix D.2. The nonunit counterexample and local-radius certificate in Appendix E address separate assumptions.

Experimental references. Kernel optima and sampling records are in Appendix F; frozen features and true-label controls are in Appendix G. The accompanying artifact package contains the raw numerical records and the figure-building scripts used for the reported diagnostics. Population quadrature, conditional label repeats, random-input replicates, and training bootstraps are identified separately in each protocol.

A General signal–noise response and fixed-feature geometry

A.1 General-kernel assumptions and response theorem

Let P_X be a probability measure on a measurable input space and let $\mathcal{H}$ be a real separable RKHS whose section map $x \mapsto K_x$ is strongly measurable and satisfies $K(x,x) = 1$ for every input. Fix $\tau > 0$, $0 < D < \infty$, and a measurable real q with $\|q\|_\infty < \infty$. Write $\ell(t) = \log(1 + e^{-t})$ and $\sigma(t) = (1+e^{-t})^{-1}$. Binary labels $Y \in \{-1, 1\}$ satisfy $P_\epsilon(Y = y \mid X = x) = (1+y\epsilon q(x))/2$, with $|\epsilon|\|q\|_\infty < 1$. Observations $Z_i = (X_i, Y_i)$ are iid with this joint law P_ϵ, and $P_n h = n^{-1}\sum_{i=1}^n h(Z_i)$. Define $f_{\epsilon,\delta}$ and $\hat{f}_{n,\delta}$ to be, respectively, the minimizers over $f \in \mathcal{H}$ of $P_\epsilon \ell(Yf(X) - \delta) + \tau\|f\|^2/2$ and $P_n \ell(Yf(X) - \delta) + \tau\|f\|^2/2$. Write $\mathcal{J} : \mathcal{H} \to L^2(P_X)$ for evaluation, $T = \mathcal{J}^*\mathcal{J}$, $s_\delta = \sigma(\delta)$, $p_\delta = \sigma'(\delta)$, and $H_\delta = \tau I + p_\delta T$.

Theorem A.1 (General signal and noise transfer)**.** *Under the preceding conditions, define $u_\delta = s_\delta H_\delta^{-1}\mathcal{J}^* q$. There exist $C, \epsilon_0 > 0$, depending only on the fixed setting, such that*

$$\sup_{0\le\delta\le D}\left(\|f_{\epsilon,\delta} - \epsilon u_\delta\| + \|\partial_\delta f_{\epsilon,\delta} - \epsilon\partial_\delta u_\delta\|\right) \le C|\epsilon|^3 \qquad (|\epsilon| \le \epsilon_0).$$

At balance ($\epsilon = 0$), $f_{0,\delta} = 0$ at every radius, and for each fixed $\delta \ge 0$,

$$\sqrt{n}\hat{f}_{n,\delta} \;\Rightarrow\; N_{\mathcal{H}}(0, C_\delta), \qquad C_\delta = s_\delta^2 H_\delta^{-1} T H_\delta^{-1}.$$

For every positive eigenvalue μ of T, the corresponding eigenvalue of C_δ is

$$V_\delta(\mu) = \frac{\mu s_\delta^2}{(\tau + \mu p_\delta)^2} = \frac{A_\delta(\mu)^2}{\mu}, \qquad A_\delta(\mu) = \frac{\mu s_\delta}{\tau + \mu p_\delta}.$$

It strictly increases with δ. If $\delta_2 > \delta_1 \ge 0$, then $\langle h, (C_{\delta_2} - C_{\delta_1})h\rangle > 0$ for every nonzero $h \in \overline{\operatorname{ran} T}$, and the difference is zero on $\ker T$. This is strict positivity on the observable subspace, not a uniform positive lower bound. Both C_δ and $\mathcal{J}C_\delta\mathcal{J}^*$ are trace class. For every deterministic nonnegative $\delta_n \to 0$, at balance,

$$\|\hat{f}_{n,\delta_n} - \hat{f}_{n,0}\| = O_p(\delta_n n^{-1/2}).$$

All norms in the displayed approximation and rate are RKHS norms. Spatial frequency interpretations are introduced only in the periodic specialization.

A.2 Proof of the common transfer law

Proof. Operators and well-posedness. The evaluation bound is $|f(x)| \le \|f\|$; hence $\|\mathcal{J}\| \le 1$. Bochner integration gives $\mathcal{J}^* v = E[v(X)K_X]$ and $T = E[K_X \otimes K_X]$. For any orthonormal basis (e_j), monotone convergence and Parseval give $\sum_j \langle e_j, Te_j\rangle = E\|K_X\|^2 = 1$. Thus T is positive trace class, with operator norm at most one. The loss is nonnegative and its penalty is coercive. A bounded minimizing sequence has a weakly convergent subsequence; evaluations converge pointwise, so Fatou's lemma and weak lower semicontinuity of the norm give a minimizer. Its score and Hessian, for either population or empirical measure M, are

$$F_{M,\delta}(f) = \tau f - M[Y\sigma(\delta - Yf(X))K_X], \qquad H_{M,\delta}(f) = \tau I + M[\sigma'(\delta - Yf(X))K_X \otimes K_X].$$

All derivatives used below pass under the integral: every fixed-order derivative of σ is bounded, and evaluated directions have absolute value at most their RKHS norms. The Hessian obeys $\tau I \preceq H_{M,\delta}(f) \preceq (\tau + 1/4)I$, so the minimizer is unique. Each Hessian is onto and its inverse has norm at most $1/\tau$: the lower bound gives closed range, and self-adjointness gives dense range. Integrating the Hessian on a line segment proves the residual inequality

$$\|f - f_{M,\delta}\| \le \tau^{-1}\|F_{M,\delta}(f)\|. \tag{A.1}$$

These facts apply also to auxiliary real radii near the interval endpoints. The empirical minimizer is a measurable $\mathcal{H}$-valued random element. Indeed, start at zero and iterate $v_{r+1} = v_r - (\tau + 1/4)^{-1}F_{P_n,\delta}(v_r)$. Each iterate is strongly measurable in the sample: the section map is strongly measurable, Hilbert inner products are continuous, and the score uses finite sums and continuous scalar functions. The mean Hessian on any segment has spectrum in $[\tau, \tau + 1/4]$, so this iteration contracts distances by at most $1 - \tau/(\tau + 1/4) < 1$. It converges in norm to the unique score zero;

the limit is strongly measurable. The same argument holds when radius is a deterministic function of sample size. The implicit function theorem gives smooth radius paths. For completeness, its local inversion condition can also be checked by the contraction map $v \mapsto v - H_{M,\delta_0}(f_0)^{-1}F_{M,\delta}(v)$ in a sufficiently small ball around a zero (f_0, δ_0); continuity of the Hessian makes the derivative in v smaller than one there. Differentiating its zero equation then gives the derivatives used below. Uniqueness identifies all local paths.

Uniform weak-signal expansion. Conditional averaging writes the population score as

$$F_{\epsilon,\delta}(f) = \tau f + \mathcal{J}^*\left[a_\delta(\mathcal{J}f) - \epsilon q\, b_\delta(\mathcal{J}f)\right],$$
$$a_\delta(t) = \frac{\sigma(\delta+t) - \sigma(\delta-t)}{2}, \qquad b_\delta(t) = \frac{\sigma(\delta+t) + \sigma(\delta-t)}{2}.$$

Taylor's theorem, including one derivative in each of t and δ, gives $a_\delta(t) = p_\delta t + O(t^3)$ and $b_\delta(t) = s_\delta + O(t^2)$. The constants are uniform on bounded t and $\delta \in [0, D]$. More explicitly, the t derivatives of these remainders are $O(t^2)$ and $O(|t|)$, while their δ derivatives remain $O(t^3)$ and $O(t^2)$, respectively; bounded derivatives of σ through order four suffice. Since $\|H_\delta^{-1}\| \le \tau^{-1}$, both u_δ and its derivative are uniformly bounded; the derivative follows by differentiating $H_\delta u_\delta = s_\delta \mathcal{J}^* q$. Set $v_\delta = \epsilon u_\delta$. Its linear score terms cancel exactly. The evaluation bound, $\|\mathcal{J}^*\| \le 1$, and the Taylor estimates give

$$r_\delta := F_{\epsilon,\delta}(v_\delta), \qquad \sup_{\delta\in[0,D]} (\|r_\delta\| + \|r'_\delta\|) \le C|\epsilon|^3.$$

The derivative here is total: the additional terms contain $v'_\delta = O(\epsilon)$ multiplying the stated t-derivative remainders. Equation (A.1) gives $\|v_\delta - f_{\epsilon,\delta}\| \le C|\epsilon|^3$ uniformly. Let $B(f) = \partial_\delta F_{\epsilon,\delta}(f)$ and $H(f) = D_f F_{\epsilon,\delta}(f)$ at the current radius. Bounded logistic derivatives imply $\|B(f) - B(v)\| + \|H(f) - H(v)\|_{\mathrm{op}} \le C\|f - v\|$. Subtracting the differentiated score identities gives the exact equation

$$H(f)(v' - f') = r' + [H(f) - H(v)]v' + B(f) - B(v).$$

The inverse bound and $v' = O(\epsilon)$ prove the required $O(\epsilon^3)$ derivative bound. This transfers a residual to the actual minimizer; it does not differentiate an uncontrolled remainder. When $q = 0$ the score zero is exactly zero and the expansion is exact.

Balanced central limit. At balance the conditional mean of Y is zero, so $F_{P,\delta}(0) = 0$ and the unique population minimizer is zero. Put $M_n = n^{-1}\sum_i Y_i K_{X_i}$ and $T_n = n^{-1}\sum_i K_{X_i} \otimes K_{X_i}$. Independence and centering give $E\|M_n\|^2 = 1/n$ and $E\|T_n - T\|_{\mathrm{HS}}^2 \le 1/n$. The residual inequality at zero gives $\|\hat{f}_{n,\delta}\| \le s_\delta\|M_n\|/\tau = O_p(n^{-1/2})$. The empirical score Taylor formula is

$$0 = -s_\delta M_n + (\tau I + p_\delta T_n)\hat{f}_{n,\delta} + R_n, \qquad \|R_n\| \le C\|\hat{f}_{n,\delta}\|^2.$$

Consequently $\hat{f}_{n,\delta} = s_\delta H_\delta^{-1} M_n + O_p(n^{-1})$. The centered iid Hilbert vectors YK_X have covariance T and second moment one. Their Hilbert central limit follows directly from finite-dimensional projection: in an orthonormal basis, each finite projection satisfies the ordinary multivariate central limit theorem, and the expected squared norm of the complementary part of $\sqrt{n}M_n$ is $\sum_{j>m}\langle e_j, Te_j\rangle \to 0$, uniformly in n. The Gaussian with covariance T exists as the mean-square convergent series of independent normal coefficients with variances the eigenvalues of T; its projection tails satisfy the same bound. Finite projection convergence and these uniform tails prove convergence against bounded Lipschitz functions in $\mathcal{H}$. Applying the bounded linear map $s_\delta H_\delta^{-1}$ proves the asserted central limit.

Transfer and strictness. Choose a unit eigenvector e of T with eigenvalue $\mu > 0$ and set $\phi = \mathcal{J}e/\sqrt{\mu}$. Then $\|\phi\|_2 = 1$, $\mathcal{J}^*\phi = \sqrt{\mu}e$, and $\langle \mathcal{J}u_\delta, \phi\rangle = A_\delta(\mu)\langle q, \phi\rangle$. Components orthogonal to $\overline{\operatorname{ran}\mathcal{J}}$ have zero image under $\mathcal{J}^*$. Thus the signal and noise use the same transfer. Direct differentiation, using $p = s(1-s)$, gives

$$A'_\delta(\mu) = \frac{\mu p_\delta(\tau + \mu s_\delta^2)}{(\tau + \mu p_\delta)^2} > 0.$$

Hence $V'_\delta = 2A_\delta A'_\delta/\mu > 0$. The spectral theorem for compact positive self-adjoint T supplies an orthonormal eigenbasis on $\overline{\operatorname{ran} T}$; summing the strictly positive differences times squared coordinates gives the stated strict quadratic inequality. On $\ker T$ all covariances vanish. Trace bounds follow from $V_\delta(\mu) \le \mu/\tau^2$ and $\|\mathcal{J}\| \le 1$. Prediction-coordinate variances are $\mu V_\delta(\mu)$. The fixed ridge supplies the inverse bound even when the spectrum approaches zero.

Every shrinking radius. The deterministic bound $\|\hat{f}_{n,\delta}\| \le \|M_n\|/\tau$ holds simultaneously for all $\delta \ge 0$. At the empirical optimum,

$$\partial_\delta F_{n,\delta}(f) = -p_\delta M_n + E_n(f), \qquad \|E_n(f)\| \le L\|f\|,$$

where $L = \sup_t |\sigma''(t)| < \infty$. Differentiating the optimum and using the Hessian inverse bound proves $\sup_{0\le\delta\le D} \|\partial_\delta \hat{f}_{n,\delta}\| \le \tau^{-1}(1/4 + L/\tau)\|M_n\|$. Integration over $[0, \delta_n]$ gives the paired rate, even if δ_n decreases arbitrarily faster than $n^{-1/2}$. It is identically zero when $\delta_n = 0$. Together with the zero-radius central limit this also proves the common root-n limit for every shrinking sequence. □

A.3 General cubic interaction response

The first-order theorem above does not require a spatial basis. The same conditional-score calculation also identifies the next nonzero term on an arbitrary normalized input geometry.

Proposition A.2 (General cubic interaction response)**.** *Fix $0 < r_- \le r_+ < \infty$. For each $r \in [r_-, r_+]$, let $f^{(r)}_{\epsilon,\delta}$ denote the population minimizer with ridge r, and write $F^{(r)}_{\epsilon,\delta}$ for its population score. Also write*

$$H_{r,\delta} = rI + p_\delta T, \qquad u^{(r)}_{1,\delta} = s_\delta H^{-1}_{r,\delta}\mathcal{J}^* q, \qquad h^{(r)}_\delta = \mathcal{J}u^{(r)}_{1,\delta},$$
$$\zeta^{(r)}_\delta = \frac{\sigma''(\delta)}{2} q (h^{(r)}_\delta)^2 - \frac{\sigma'''(\delta)}{6}(h^{(r)}_\delta)^3, \qquad u^{(r)}_{3,\delta} = H^{-1}_{r,\delta}\mathcal{J}^*\zeta^{(r)}_\delta.$$

There are $C, \epsilon_0 > 0$, depending only on the fixed compact ridge interval and the other fixed quantities, such that

$$\sup_{r\in[r_-,r_+]} \left\| f^{(r)}_{\epsilon,\cdot} - \epsilon u^{(r)}_{1,\cdot} - \epsilon^3 u^{(r)}_{3,\cdot} \right\|_{C^1([0,D];\mathcal{H})} \le C|\epsilon|^5, \qquad |\epsilon| \le \epsilon_0. \tag{A.2}$$

In addition,

$$\sup_{r\in[r_-,r_+]} \left\| \partial_r \left(f^{(r)}_{\epsilon,\cdot} - \epsilon u^{(r)}_{1,\cdot} - \epsilon^3 u^{(r)}_{3,\cdot} \right) \right\|_{C^1([0,D];\mathcal{H})} \le C|\epsilon|^5. \tag{A.3}$$

The parameter-uniform form is what permits the ridge path $r = \lambda_\delta$ in the matched-control theorem below. No spatial frequency or ordering is assumed in this statement. If $Te_j = \mu_j e_j$ with $\mu_j > 0$ and $\varphi_j = \mathcal{J}e_j/\sqrt{\mu_j}$, then

$$\langle \mathcal{J}u^{(r)}_{3,\delta}, \varphi_j\rangle_{L^2(P_X)} = \frac{\mu_j}{r + \mu_j p_\delta}\langle \zeta^{(r)}_\delta, \varphi_j\rangle_{L^2(P_X)}. \tag{A.4}$$

Proof. Put $L_{r,\delta} = H_{r,\delta}$ and first note that $\|\mathcal{J}\| \le 1$ because $K(x,x) = 1$. The common lower bound $\|L^{-1}_{r,\delta}\| \le r_-^{-1}$ makes all estimates below uniform in $r \in [r_-, r_+]$. Hence $u^{(r)}_{1,\delta}$ and $h^{(r)}_\delta$ are uniformly bounded on the rectangle $[r_-, r_+] \times [0, D]$. Differentiating $L_{r,\delta}u^{(r)}_{1,\delta} = s_\delta \mathcal{J}^* q$ gives

$$L_{r,\delta}\partial_\delta u^{(r)}_{1,\delta} = s'_\delta \mathcal{J}^* q - p'_\delta T u^{(r)}_{1,\delta},$$

so $u^{(r)}_1$, $h^{(r)}_\delta$, and their radius derivatives are uniformly bounded. Differentiating in the ridge gives $\partial_r L^{-1}_{r,\delta} = -L^{-2}_{r,\delta}$, so the same is true for the ridge derivatives. The same resolvent bound then gives uniform $L^2(P_X)$ bounds, and bounds for both parameter derivatives, for $\zeta^{(r)}_\delta$ and $u^{(r)}_{3,\delta}$.

For a trial path $v^{(r)}_\delta$ with $\|v^{(r)}\|_{C^1} = O(|\epsilon|)$, conditional averaging of the logistic score and Taylor expansion through degree four give

$$F^{(r)}_{\epsilon,\delta}(v^{(r)}) = H_{r,\delta}v^{(r)} - \epsilon s_\delta \mathcal{J}^* q + \mathcal{J}^*\left[\frac{\sigma'''(\delta)}{6}(\mathcal{J}v^{(r)})^3 - \frac{\epsilon\sigma''(\delta)}{2} q (\mathcal{J}v^{(r)})^2\right] + \mathcal{J}^* R^{(r)}_\delta,$$

where bounded sigmoid derivatives imply $\|\mathcal{J}^* R^{(r)}\|_{C^1([0,D];\mathcal{H})} \le C|\epsilon|^5$, uniformly in r; the same bound holds after one r derivative. This uses only the bounded evaluation map and the fact that P_X is a probability measure; products are pointwise $L^2(P_X)$ products.

Take $v_\delta^{(r)} = \epsilon u_{1,\delta}^{(r)} + \epsilon^3 u_{3,\delta}^{(r)}$. The linear term cancels by the definition of $u_1^{(r)}$. At order ϵ^3, the displayed source is

$$\mathcal{J}^*\left[\frac{\sigma'''(\delta)}{6}(h_\delta^{(r)})^3 - \frac{\sigma''(\delta)}{2}q(h_\delta^{(r)})^2\right],$$

which cancels $L_{r,\delta}u_{3,\delta}^{(r)}$ by the definition of $\zeta_\delta^{(r)}$. The remaining substitutions are

$$(\mathcal{J}v^{(r)})^3 - \epsilon^3(h^{(r)})^3 = O_{C^1}(\epsilon^5), \qquad \epsilon q\big((\mathcal{J}v^{(r)})^2 - \epsilon^2(h^{(r)})^2\big) = O_{C^1}(\epsilon^5),$$

with the bounds interpreted in $L^2(P_X)$ after multiplication by q. Thus the score residual and its total radius derivative are $O(\epsilon^5)$ in $\mathcal{H}$, uniformly in r.

For the additional ridge derivative, differentiate the score equation and subtract the corresponding equation for the trial path. Since $\partial_r F_{\epsilon,\delta}^{(r)}(w) = w$, the residual $e = f^{(r)} - v^{(r)}$ satisfies

$$H_{r,\delta}(f^{(r)})\,\partial_r e = -e - \big[H_{r,\delta}(f^{(r)}) - H_{r,\delta}(v^{(r)})\big]\partial_r v^{(r)} - \partial_r\Big\{F_{\epsilon,\delta}^{(r)}(v^{(r)})\Big\},$$

where $H_{r,\delta}(w)$ denotes the full Hessian at w (the operator $H_{r,\delta}$ above is its zero-signal value). The last derivative is the total derivative of the trial residual, including the path derivative $\partial_r v^{(r)}$. The strong-monotonicity residual inequality first gives $e = O(\epsilon^5)$; the bounded inverse, the global Hessian Lipschitz bound, and the already bounded trial derivatives then give $\partial_r e = O(\epsilon^5)$. The same argument applied to the radius derivative transfers its $O(\epsilon^5)$ residual. This proves (A.2) and the asserted partial-r bound.

Finally, self-adjointness and $\mathcal{J}^*\varphi_j = \sqrt{\mu_j}e_j$ give

$$\langle \mathcal{J}u_3^{(r)}, \varphi_j\rangle = \langle u_3^{(r)}, \mathcal{J}^*\varphi_j\rangle = \frac{\mu_j}{r + \mu_j p_\delta}\langle \zeta_\delta^{(r)}, \varphi_j\rangle,$$

which proves (A.4). □

A.4 Exact fixed-feature correspondence

Proposition A.3 (Exact normalized fixed-feature correspondence)**.** *Let $\psi : \mathcal{X} \to \mathcal{E}$ be a strongly measurable feature map into a real separable Hilbert parameter space with $\|\psi(x)\| = 1$ for every x. Let $K(x,z) = \langle\psi(x), \psi(z)\rangle$ and $\mathcal{E}_0 = \overline{\operatorname{span}\{\psi(x) : x \in \mathcal{X}\}}$. Then $w \mapsto f_w$, $f_w(x) = \langle w, \psi(x)\rangle$, is an isometry from $\mathcal{E}_0$ onto the RKHS of K, and*

$$\|f_w\|_{\mathcal{H}} = \|\Pi_{\mathcal{E}_0} w\| = \min_{v : f_v = f_w} \|v\|.$$

For either an empirical or population average M, minimizing over $w \in \mathcal{E}$ the objective

$$M\left[\sup_{\|a\|_{\mathcal{E}} \le \delta} \ell\big(Y\langle w + a, \psi(X)\rangle\big)\right] + \frac{\tau}{2}\|w\|_{\mathcal{E}}^2$$

is exactly equivalent to the samplewise RKHS objective. Its unique minimizing parameter belongs to $\mathcal{E}_0$ and represents the unique RKHS minimizer. The perturbation is chosen separately for each observation, and the features are fixed independently of these optimization variables.

Proof. Orthogonal projection onto $\mathcal{E}_0$ leaves all predictions unchanged. On $\mathcal{E}_0$, the map $w \mapsto f_w$ is injective: zero evaluations mean orthogonality to the dense span defining $\mathcal{E}_0$. Give its image the transported inner product. It is complete, evaluations are bounded, and the image of $\psi(x)$ represents evaluation at x, with kernel K. Thus this image is the RKHS of K. Every parameter representing the same function has the form $\Pi_{\mathcal{E}_0} w + b$ with $b \in \mathcal{E}_0^\perp$, which proves the minimum norm statement by Pythagoras. For each observation, $Y\langle a, \psi(x)\rangle \ge -\delta$, with equality at $a = -\delta Y\psi(x)$. Since ℓ is decreasing, the inner maximum equals $\ell(Yf_w(x) - \delta)$. The loss depends only on the projected parameter, whereas its ridge penalty adds $\tau\|b\|^2/2$ for a redundant component. The optimum therefore has $b = 0$; the remaining objective is isometric to the RKHS objective. The equivalence concerns fixed features and independent per-example perturbations. □

B ANALYTIC MATCHING AND THE ORDINARY-LOGISTIC CONTROL

The matching calculation is valid on any normalized kernel. Its ridge-path argument then reduces nonlinear separation to the signed periodic coefficients proved in Appendix C.3. The ridge superscript distinguishes an ordinary logistic fit at a changed penalty from the samplewise fit at fixed τ. Throughout this appendix $K(x,x) = 1$, $\tau > 0$, the input law is fixed, and q is bounded as in Section 3.

B.1 MATCHING EVERY LEADING SIGNAL AND NOISE DIRECTION

At radius zero, $s_0 = 1/2$ and $p_0 = 1/4$. With $\lambda_\delta = \tau/(4p_\delta)$,

$$\lambda_\delta I + T/4 = (4p_\delta)^{-1}(\tau I + p_\delta T).$$

Both operators have bounded inverses because their ridges are positive. Multiplying the inverse identity by $\gamma_\delta/2 = s_\delta/(4p_\delta)$ and then by $\mathcal{J}^*$ proves the asserted equality of transfer maps. The covariance identity follows from $C_\delta^{(\tau)} = \mathcal{D}_\delta^{(\tau)}(\mathcal{D}_\delta^{(\tau)})^*$ and the same balanced-label formula at ridge λ_δ. A deterministic output scaling by γ_δ multiplies the limiting covariance by γ_δ^2. For each fixed δ, Theorem 2 applied at the two positive ridge values also gives

$$f_{\epsilon,\delta}^{(\tau)} - \gamma_\delta f_{\epsilon,0}^{(\lambda_\delta)} = O_{\mathcal{H}}(|\epsilon|^3).$$

If $\mathcal{D}_\delta^{(\tau)} q \neq 0$, the denominator $\|f_{\epsilon,\delta}^{(\tau)}\|_{\mathcal{H}}$ is bounded below by a positive multiple of $|\epsilon|$ for sufficiently small nonzero ϵ. The relative RKHS error is therefore $O(\epsilon^2)$. The same argument applies in prediction norm when $\mathcal{J}\mathcal{D}_\delta^{(\tau)} q \neq 0$. These are upper orders; a nonzero cubic difference is needed for an exact leading order.

B.2 GENERAL CUBIC SEPARATION AFTER MATCHING

The preceding identity matches the complete linear response. The first mechanism-specific discrepancy can also be written without a Fourier assumption.

Proposition B.1 (Matched third-order separation)**.** *For any positive ridge r, write*

$$H_{r,\rho} = rI + p_\rho T, \qquad u_{1,\rho}^{(r)} = s_\rho H_{r,\rho}^{-1}\mathcal{J}^* q, \qquad h_\rho^{(r)} = \mathcal{J} u_{1,\rho}^{(r)},$$

and

$$\zeta_\rho^{(r)} = \frac{\sigma''(\rho)}{2} q (h_\rho^{(r)})^2 - \frac{\sigma'''(\rho)}{6}(h_\rho^{(r)})^3, \qquad u_{3,\rho}^{(r)} = H_{r,\rho}^{-1}\mathcal{J}^*\zeta_\rho^{(r)}.$$

Let

$$\lambda_\delta = \frac{\tau}{4p_\delta}, \qquad \gamma_\delta = \frac{s_\delta}{2p_\delta}, \qquad \Delta_\delta[q] = u_{3,\delta}^{(\tau)} - \gamma_\delta u_{3,0}^{(\lambda_\delta)}.$$

Then, uniformly for $\delta \in [0, D]$,

$$\left\| f_{\epsilon,\delta}^{(\tau)} - \gamma_\delta f_{\epsilon,0}^{(\lambda_\delta)} - \epsilon^3 \Delta_\delta[q] \right\|_{C^1([0,D];\mathcal{H})} \leq C|\epsilon|^5. \tag{B.1}$$

Equivalently, if $h_\delta = h_\delta^{(\tau)}$, then

$$\Delta_\delta[q] = H_{\tau,\delta}^{-1}\mathcal{J}^*\left(\zeta_\delta^{(\tau)} - \zeta_\delta^{\mathrm{ctl}}\right), \quad \zeta_\delta^{\mathrm{ctl}} = 4p_\delta\gamma_\delta\left[\frac{\sigma''(0)}{2\gamma_\delta^2} q h_\delta^2 - \frac{\sigma'''(0)}{6\gamma_\delta^3} h_\delta^3\right]. \tag{B.2}$$

In particular $\Delta_0[q] = 0$. For every bounded linear probe $\ell : \mathcal{H} \to \mathbb{R}$,

$$\frac{\ell(f_{\epsilon,\delta}^{(\tau)}) - \gamma_\delta \ell(f_{\epsilon,0}^{(\lambda_\delta)})}{\epsilon^3} \longrightarrow \ell(\Delta_\delta[q]) \quad (\epsilon \to 0).$$

An exact cubic separation for that probe therefore requires the explicit nondegeneracy condition $\ell(\Delta_\delta[q]) \neq 0$; no universal sign is asserted for arbitrary kernels, input laws, or signals.

Proof. The general cubic expansion in Proposition A.2 remains uniform when the ridge r ranges over a compact positive interval. Indeed, all inverse bounds use the common lower ridge bound, and differentiating the resolvent with respect to r gives $\partial_r H_{r,\rho}^{-1} = -H_{r,\rho}^{-2}$; the same Taylor-residual argument therefore controls one (r, ρ) derivative. Since $p_\delta > 0$ on the compact interval, λ_δ lies in such a compact positive interval and its derivative is bounded.

Apply that expansion first at $(r, \rho) = (\tau, \delta)$ and then at $(r, \rho) = (\lambda_\delta, 0)$. The first-order matching identity gives

$$u_{1,\delta}^{(\tau)} = \gamma_\delta u_{1,0}^{(\lambda_\delta)}, \qquad h_0^{(\lambda_\delta)} = h_\delta/\gamma_\delta.$$

Moreover,

$$H_{\lambda_\delta,0} = (4p_\delta)^{-1} H_{\tau,\delta},$$

so the scaled control cubic term equals

$$\gamma_\delta u_{3,0}^{(\lambda_\delta)} = H_{\tau,\delta}^{-1} \mathcal{J}^* \zeta_\delta^{\mathrm{ctl}}.$$

Subtracting the two cubic expansions cancels the order-ϵ terms and leaves exactly $\epsilon^3 \Delta_\delta[q]$. The uniform C^1 remainder is preserved under the smooth paths λ_δ and γ_δ by the additional ridge-derivative bound. This proves (B.1); the probe limit follows by applying ℓ. □

Corollary B.2 (Local logistic separation certificate)**.** *For the logistic loss, let $H_0 = \tau I + T/4$ and $h_0 = \mathcal{J}(\frac{1}{2} H_0^{-1} \mathcal{J}^* q)$. Then*

$$\partial_\delta \Delta_0[q] = H_0^{-1} \mathcal{J}^* \left[-\frac{1}{16} q h_0^2 + \frac{1}{48} h_0^3 \right]. \tag{B.3}$$

Consequently, if a bounded probe ℓ has $\ell(\partial_\delta \Delta_0[q]) \neq 0$, then

$$\ell(\Delta_\delta[q]) = \delta\, \ell(\partial_\delta \Delta_0[q]) + o(\delta) \qquad (\delta \downarrow 0).$$

Proof. At zero radius, $p_0' = 0$, $\gamma_0 = 1$, $\gamma_0' = 1/2$, and $\partial_\delta h_\delta^{(\tau)}|_0 = h_0/2$. Differentiating the samplewise source gives $-q h_0^2/16 + h_0^3/32$. The scaled control source in (B.2) has derivative $h_0^3/96$. Their difference is the bracket in (B.3); the derivative of the resolvent does not contribute because the two sources coincide at zero. The final expansion follows from differentiability of $\Delta_\delta[q]$. □

B.3 Ridge paths and the matched harmonic derivative

Fix an admissible ϵ and write $F_\lambda = f_{\epsilon,0}^{(\lambda)}$. Let $A(f)$ be the population gradient of the unpenalized ordinary logistic risk. Convexity makes A monotone. The optimality equations are $A(F_\lambda) + \lambda F_\lambda = 0$ and $A(F_\tau) + \tau F_\tau = 0$. Taking their difference against $F_\lambda - F_\tau$ yields

$$\lambda \|F_\lambda - F_\tau\|_{\mathcal{H}}^2 \leq |\lambda - \tau| \|F_\tau\|_{\mathcal{H}} \|F_\lambda - F_\tau\|_{\mathcal{H}}.$$

Division, with the zero-difference case immediate, proves

$$\|F_\lambda - F_\tau\|_{\mathcal{H}} \leq \frac{|\lambda - \tau|}{\lambda} \|F_\tau\|_{\mathcal{H}}.$$

Since $p_\delta = 1/(4\cosh^2(\delta/2))$,

$$\lambda_\delta = \tau \cosh^2(\delta/2) = \tau + O(\delta^2), \qquad \gamma_\delta = (1 + e^\delta)/2 = 1 + \delta/2 + O(\delta^2).$$

Consequently $g_{\epsilon,\delta} = F_\tau + \delta F_\tau/2 + O_{\mathcal{H}}(\delta^2)$, proving its right derivative at zero. The Fourier coefficient c_{3k} is a bounded linear functional on $\mathcal{H}$, and $E_{3k} = c_{3k}^2/(2\mu_{3k})$. Thus

$$\partial_\delta E_{3k}(g_{\epsilon,\delta})|_{0^+} = \frac{c_{3k}(F_\tau)^2}{2\mu_{3k}} = E_{3k}(F_\tau).$$

Theorem 6 gives $c_{3k}(F_\tau) = \epsilon^3 b_{3,0} + O(|\epsilon|^5)$ with $b_{3,0} > 0$. The energy is strictly positive for sufficiently small nonzero ϵ. Corollary 7 gives the strictly negative derivative for the samplewise estimator under the same periodic assumptions. This proves the two claimed signs at zero. Their interpretation concerns a particular component of the score, not a prediction-risk ordering.

C PERIODIC SIGNAL AND NONLINEAR HARMONIC RESPONSE

C.1 ASSUMPTIONS AND COORDINATE CONVENTIONS

We study exact minimizers of a samplewise perturbed logistic risk in a real reproducing kernel Hilbert space. The perturbation is chosen separately for each observation, and the quadratic penalty is external to the supremum. The statistical limits keep the signal law and the positive ridge parameter fixed.

Assumption 1 (Fixed normalized Fourier kernel)**.** On the circle $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$ with $m(dx) = dx/(2\pi)$, let

$$K(x, x') = \mu_0 + 2\sum_{j\geq 1} \mu_j \cos(j(x - x')), \qquad \mu_j \geq 0, \qquad \mu_0 + 2\sum_{j\geq 1} \mu_j = 1.$$

Infinitely many coefficients are positive. We use the real separable RKHS $\mathcal{H}$ of this continuous kernel, with $K_x = K(x, \cdot)$.

Assumption 2 (Fixed ridge and bounded radii)**.** The parameters $\tau > 0$ and $0 < D < \infty$ are fixed, and $0 \leq \delta \leq D$. Constants may depend on τ, D, K, q, and on a fixed frequency k when specified. No uniform limit as $\tau \downarrow 0$, $D \uparrow \infty$, or $k \uparrow \infty$ is asserted.

Assumption 3 (Signal law and independent sampling)**.** Let q be a fixed bounded real trigonometric polynomial. Under P_ϵ, $X \sim m$, $Y \in \{-1, 1\}$, and

$$P_\epsilon(Y = 1 \mid X = x) = \frac{1 + \epsilon q(x)}{2}, \qquad |\epsilon|\|q\|_\infty < 1.$$

The observations $Z_i = (X_i, Y_i)$ are independent with this common law, and $P_n = n^{-1}\sum_{i=1}^n \delta_{Z_i}$. Weak-signal results range over a sufficiently small interval of ϵ around zero; statistical limits keep ϵ and q fixed.

Assumption 4 (Single-mode specialization)**.** For the cubic response and signed harmonic conclusions only, $q(x) = \cos(kx)$ for a fixed integer $k \geq 1$ with $\mu_k, \mu_{3k} > 0$.

Assumption 5 (Shrinking-radius specialization)**.** For shrinking-radius conclusions only, (δ_n) is deterministic, belongs to $[0, D]$, and tends to zero. The exponent specialization uses $\delta_n = \eta n^{-\gamma}$ for fixed $\eta > 0$, $\gamma > 0$, once this sequence lies in $[0, D]$.

Put $\ell(t) = \log(1 + e^{-t})$ and $\sigma(t) = (1 + e^{-t})^{-1}$. The population and empirical objectives are

$$Q_{\epsilon,\delta}(f) = P_\epsilon \ell(Y f(X) - \delta) + \frac{\tau}{2}\|f\|_{\mathcal{H}}^2, \qquad Q_{n,\delta}(f) = P_n \ell(Y f(X) - \delta) + \frac{\tau}{2}\|f\|_{\mathcal{H}}^2.$$

Their minimizers are denoted by $f_{\epsilon,\delta}$ and $\hat{f}_{n,\delta}$. Existence, uniqueness, smoothness, and the identity with the exact samplewise supremum will be proved. The following lemmas establish the coordinate, path, and remainder estimates needed for the nonlinear response.

C.1.1 FOURIER AND STATISTICAL COORDINATES

Let $I : \mathcal{H} \to L^2(m)$ be the inclusion, I^* its adjoint, and $T = I^*I$. For $u, v \in \mathcal{H}$, $(u \otimes v)h = \langle v, h\rangle_{\mathcal{H}} u$. We use the real $L^2(m)$ Fourier functions $\varphi_0 = 1$, $\varphi_{j,c} = \sqrt{2}\cos(jx)$, and $\varphi_{j,s} = \sqrt{2}\sin(jx)$. At positive eigenvalues put $e_{j,a} = \sqrt{\mu_j}\varphi_{j,a}$. The constant mode is counted once, and each positive frequency has a cosine and sine coordinate. These coordinates distinguish RKHS covariance from prediction covariance. Missing kernel modes are excluded from the RKHS basis.

Write $s_\delta = \sigma(\delta)$ and $p_\delta = \sigma'(\delta)$. The linear response multiplier and its action on the fixed signal are

$$A_\delta(\mu) = \frac{\mu s_\delta}{\tau + \mu p_\delta}, \quad A_\delta(0) = 0, \qquad \mathcal{A}_\delta q = \sum_{j,a} A_\delta(\mu_j)\langle q, \varphi_{j,a}\rangle_{L^2(m)}\varphi_{j,a}.$$

For the single-mode specialization define

$$a_\delta = A_\delta(\mu_k), \quad B_\delta = \frac{\sigma''(\delta)a_\delta^2}{8} - \frac{\sigma'''(\delta)a_\delta^3}{24}, \quad b_{1,\delta} = \frac{3\mu_k B_\delta}{\tau + \mu_k p_\delta}, \quad b_{3,\delta} = \frac{\mu_{3k} B_\delta}{\tau + \mu_{3k} p_\delta}.$$

These are the coefficients of the response theorem. Its path norm is

$$\|h\|_{C^1([0,D];\mathcal{H})} = \sup_{0\le\delta\le D} (\|h(\delta)\|_{\mathcal{H}} + \|h'(\delta)\|_{\mathcal{H}}).$$

Endpoint derivatives use a smooth extension. For $\mu_j > 0$ define the cosine amplitude $c_j(f) = 2\int f(x)\cos(jx)\,dm(x)$ and its RKHS energy $E_j(f) = c_j(f)^2/(2\mu_j)$.

For statistics put $P = P_\epsilon$, $f_\tau = f_{\epsilon,0}$, and define

$$g(Z) = -Y\sigma(-Yf_\tau(X))K_X, \quad H_\tau = \tau I_{\mathcal{H}} + P[\ell''(Yf_\tau(X))K_X \otimes K_X],$$
$$R_\tau = -P[Y\ell''(Yf_\tau(X))K_X], \quad d = H_\tau^{-1}R_\tau, \quad \Sigma = \mathrm{Cov}(g), \quad C_0 = H_\tau^{-1}\Sigma H_\tau^{-1}.$$

At any fixed radius the corresponding quantities are

$$g_\delta(Z) = -Y\sigma(\delta - Yf_{\epsilon,\delta}(X))K_X,$$
$$H_\delta = \tau I_{\mathcal{H}} + P[\ell''(Yf_{\epsilon,\delta}(X) - \delta)K_X \otimes K_X],$$
$$C_\delta = H_\delta^{-1}\,\mathrm{Cov}(g_\delta)H_\delta^{-1}.$$

The inverses and trace-class Gaussian laws appearing in these formulas are established below. Finally set $V_\delta(\mu) = \mu s_\delta^2/(\tau + \mu p_\delta)^2$, including $V_\delta(0) = 0$; this is the balanced-label RKHS covariance multiplier.

C.2 Exact objectives, global minimizers, and smooth paths

Lemma C.1 (RKHS operator and coordinate identities)**.** *Under Assumption 1, $|f(x)| \le \|f\|_{\mathcal{H}}$ and $\|I\| \le 1$. Moreover,*

$$I^*v = \int v(x)K_x\,dm(x), \qquad T = I^*I = \int K_x \otimes K_x\,dm(x), \qquad \mathrm{tr}\,T = 1.$$

The operator T is positive and self-adjoint. At positive modes the vectors $e_{j,a}$ form a complete orthonormal basis of $\mathcal{H}$ and satisfy

$$Te_{j,a} = \mu_j e_{j,a}, \quad Ie_{j,a} = \sqrt{\mu_j}\varphi_{j,a}, \quad I^*\varphi_{j,a} = \sqrt{\mu_j}e_{j,a}.$$

Absent modes have zero image under I^*. For $j \ge 1$ with $\mu_j > 0$, $\|\cos(jx)\|_{\mathcal{H}}^2 = (2\mu_j)^{-1}$ and $|c_j(f)| \le \sqrt{2\mu_j}\|f\|_{\mathcal{H}}$.

Proof. Summability gives uniform convergence of the Fourier kernel. Thus K is continuous and

$$\|K_x - K_z\|_{\mathcal{H}}^2 = K(x,x) + K(z,z) - 2K(x,z) \longrightarrow 0 \quad (z \to x).$$

The feature map is continuous and strongly measurable in the separable space $\mathcal{H}$. The reproducing identity and $K(x,x) = 1$ prove the evaluation and inclusion bounds. If $v \in L^2(m)$, then $\int |v|\,dm \le \|v\|_2$, so $v(x)K_x$ is Bochner integrable. Pairing its integral with f gives $\int v(x)f(x)\,dm = \langle v, If\rangle_2$, proving the adjoint formula.

Realize the RKHS as Fourier series $f = \sum_{j,a}\alpha_{j,a}\varphi_{j,a}$ over positive modes with $\sum \alpha_{j,a}^2/\mu_j < \infty$. Cauchy–Schwarz bounds the pointwise series by

$$\left(\sum_{j,a}\frac{\alpha_{j,a}^2}{\mu_j}\right)^{1/2}\left(\sum_{j,a}\mu_j\varphi_{j,a}(x)^2\right)^{1/2} = \|f\|_{\mathcal{H}}.$$

These series define functions with reproducing kernel K. The kernel sections have dense span: a vector orthogonal to every section is the zero function, whose Fourier coefficients vanish. This identifies the realization with $\mathcal{H}$ and proves completeness of the displayed basis. Fourier orthogonality proves the identities for I, I^*, and T. If a mode is absent, its integral against every kernel section is zero, hence $I^*\varphi_{j,a} = 0$.

The tensor $K_x \otimes K_x$ has operator and Hilbert–Schmidt norms one. Pairing its integral with two vectors identifies it with I^*I, which is positive and self-adjoint. Nonnegative sums commute with the integral, and completeness gives

$$\sum_{j,a}\langle e_{j,a}, Te_{j,a}\rangle = \int \sum_{j,a}|\langle e_{j,a}, K_x\rangle|^2\,dm = \int \|K_x\|_{\mathcal{H}}^2\,dm = 1.$$

The finite positive diagonal sum proves trace class and fixes the multiplicities. Finally, $\cos(jx) = e_{j,c}/\sqrt{2\mu_j}$ and $c_j(f) = \sqrt{2\mu_j}\langle f, e_{j,c}\rangle_{\mathcal{H}}$, proving both remaining claims. □

Proposition C.2 (Exact samplewise margin shift)**.** *Under Assumptions 1 and 2, for $f \in \mathcal{H}$, $y \in \{-1, 1\}$, $x \in \mathbb{T}$, and $\delta \geq 0$,*

$$\sup_{\|\nu\|_{\mathcal{H}} \leq \delta} \ell(y(f+\nu)(x)) = \ell(yf(x) - \delta).$$

Thus the shifted population and empirical objectives equal the respective samplewise adversarial objectives with external ridge.

Proof. Lemma C.1 gives $y\nu(x) \geq -\delta$. Since $\ell'(t) = -\sigma(-t) < 0$, every admissible loss is at most $\ell(yf(x) - \delta)$. The choice $\nu = -\delta y K_x$ has norm δ and $y\nu(x) = -\delta$, so it attains that bound. This chooses a perturbation separately for each observation. The shifted value is measurable, and its expectation or empirical average gives the asserted objective identities. At radius zero the identity recovers the unperturbed loss. A common perturbation across observations is a different supremum. □

Lemma C.3 (Global well-posedness and residual comparison)**.** *Under Assumptions 1, 2, and 3, let M be either P_ϵ or any realized empirical measure P_n. For any auxiliary real δ, define*

$$F_{M,\delta}(f) = \tau f - M[Y\sigma(\delta - Yf(X))K_X], \quad H_{M,\delta}(f) = \tau I_{\mathcal{H}} + M[\sigma'(\delta - Yf(X))K_X \otimes K_X].$$

The shifted objective has a unique minimizer $f_{M,\delta}$, with $F_{M,\delta}(f_{M,\delta}) = 0$ and $\|f_{M,\delta}\| \leq \tau^{-1}$. For every f, $\tau I_{\mathcal{H}} \preceq H_{M,\delta}(f) \preceq (\tau + 1/4) I_{\mathcal{H}}$; its inverse exists on $\mathcal{H}$ and has norm at most τ^{-1}. For every $u \in \mathcal{H}$,

$$\|u - f_{M,\delta}\|_{\mathcal{H}} \leq \tau^{-1} \|F_{M,\delta}(u)\|_{\mathcal{H}}.$$

Proof. The nonnegative loss gives $Q(f) \geq \tau\|f\|^2/2$, whereas $Q(0) = \ell(-\delta) < \infty$. A minimizing sequence is bounded. Expand it in a countable orthonormal basis and take a diagonal subsequence with every coordinate convergent. Every finite sum of squared limiting coordinates is bounded by the same squared norm bound; the limiting coordinates therefore define a vector $f \in \mathcal{H}$. Approximation of any test vector by its finite projections proves weak convergence, using uniform boundedness of the original sequence. Testing with K_x gives pointwise evaluation convergence. Continuity and nonnegativity of ℓ give the lower bound for the loss by Fatou's inequality; the latter follows here by monotone convergence applied to successive infima of the nonnegative losses. Finite coordinate sums give $\|f\|^2 \leq \liminf_r \|f_r\|^2$. Hence f attains the infimum.

Scalar differentiation gives F and H. It passes through M because $|\ell'| \leq 1$, $|\ell''| \leq 1/4$, and every evaluated direction is bounded by its Hilbert norm. Scalar integral Taylor formulas bound the difference quotients by these same envelopes. Consequently the Hessian is bounded and self-adjoint, with the stated upper and lower bounds. Along the segment from f to u,

$$\langle F(u) - F(f), u - f\rangle = \int_0^1 \langle u - f, H(f + t(u-f))(u-f)\rangle \, dt \geq \tau\|u - f\|^2.$$

This proves strict convexity and uniqueness. At the unconstrained minimizer all directional derivatives vanish, giving $F(f) = 0$. The zero equation yields $\tau\|f\| \leq M\|Y\sigma(\delta - Yf(X))K_X\| \leq 1$. Cauchy–Schwarz in the last displayed inequality proves residual comparison, including the trivial case $u = f$.

To prove surjectivity, $\|Hu\| \geq \tau\|u\|$ gives injectivity and closed range: if Hu_r converges, (u_r) is Cauchy, and its limit maps to the given limit. If v is orthogonal to the range, self-adjointness gives $Hv = 0$, hence $v = 0$. The range is therefore both closed and dense, so it is all of $\mathcal{H}$. The same bound proves $\|H^{-1}\| \leq \tau^{-1}$, in infinite dimension and with missing Fourier modes as well. □

Lemma C.4 (Smooth paths and deterministic radius bounds)**.** *Under Assumptions 1, 2, and 3, the map $(\epsilon, \delta) \mapsto f_{\epsilon,\delta}$ is smooth on the open admissible signal interval times $\mathbb{R}$. For each realized sample, $\delta \mapsto \hat{f}_{n,\delta}$ is smooth on $\mathbb{R}$. Either path satisfies*

$$\|\partial_\delta f\|_{\mathcal{H}} \leq \frac{1}{4\tau}, \qquad \|\partial_\delta^2 f\|_{\mathcal{H}} \leq \frac{1}{4\tau}\left(1 + \frac{1}{4\tau}\right)^2.$$

Population mixed derivatives of each fixed finite order are bounded on any compact admissible signal interval times $[0, D]$, with constants depending only on that interval and τ, D, K, q. Score derivatives of each finite order are bounded multilinear maps. The Hessian and radius score are globally Lipschitz in f, each with constant at most $1/4$.

Proof. The identity $\sigma' = \sigma(1-\sigma)$ gives $0 < \sigma' \le 1/4$ and $|\sigma''| = |\sigma'(1-2\sigma)| \le 1/4$. More generally $\sigma^{(r)} = P_r(\sigma)$, where

$$P_0(z) = z, \qquad P_{r+1}(z) = z(1-z)P_r'(z), \qquad L_r = \max_{0\le z\le 1} |P_r(z)| < \infty.$$

Thus L_r bounds the rth derivative on all of $\mathbb{R}$. Conditioning on X gives the exact population score

$$F_{\epsilon,\delta}(f) = \tau f + I^* \left\{ \frac{\sigma(\delta+f) - \sigma(\delta-f)}{2} - \epsilon q \frac{\sigma(\delta+f) + \sigma(\delta-f)}{2} \right\}.$$

It is affine in ϵ when f is fixed. Each derivative in f produces products of evaluated directions bounded by their Hilbert norms. Radius derivatives replace σ by higher derivatives, and one signal derivative contributes a factor bounded by $\|q\|_\infty$; higher signal derivatives with f fixed vanish. Scalar integral Taylor remainders and the next L_r prove Fréchet differentiability and operator-norm continuity of all these derivatives. Bounded inclusion transfers the estimates to $\mathcal{H}$. A finite average proves the same smoothness for the empirical score.

Here is the zero-map argument explicitly. At a zero (f_0, t_0), with t either δ or (ϵ, δ), fix $A = F_f(f_0, t_0)$. Lemma C.3 supplies its inverse. Define $\Psi_t(f) = f - A^{-1}F(f,t)$. Continuity of F_f gives a product neighborhood with $\|I_{\mathcal{H}} - A^{-1}F_f\| \le 1/2$. Choose a closed f-ball inside that neighborhood and shrink the parameter neighborhood so that $\|\Psi_t(f_0) - f_0\|$ is at most half its radius. The line-integral derivative bound makes Ψ_t a self-map of the ball with Lipschitz constant $1/2$. The successive differences of its iterates are geometrically summable, so completeness gives a unique fixed point. The contraction estimate also gives continuous dependence on t. Subtracting its equations at t and $t+h$, followed by the Fréchet expansion and inverse bound, gives $f_t = -F_f^{-1}F_t$. The resolvent identity $B^{-1} - A^{-1} = B^{-1}(A-B)A^{-1}$ gives continuity and differentiability of the inverse near A. Repeated differentiation of this derivative formula proves smoothness of every finite order. By global uniqueness the local zero is the minimizer, and overlapping neighborhoods agree, proving the full path assertion.

At the zero, put $H = F_f$. With f fixed,

$$F_\delta = -M[Y\sigma'(\delta - Yf(X))K_X], \qquad F_{\delta\delta} = -M[Y\sigma''(\delta - Yf(X))K_X].$$

Both have norm at most $1/4$. Differentiating the first in a direction u gives norm at most $\|u\|/4$. Differentiating H in u and applying to v gives norm at most $\|u\|\|v\|/4$. Thus $\|F_{f\delta}\|, \|F_{ff}\| \le 1/4$. Line integration proves the two global Lipschitz bounds. Differentiating the zero equation twice yields

$$Hf' = -F_\delta, \qquad Hf'' = -F_{\delta\delta} - 2F_{f\delta}f' - F_{ff}[f', f'].$$

The inverse bound therefore gives

$$\|f'\| \le (4\tau)^{-1}, \quad \|f''\| \le \frac{1}{4\tau}(1 + 2\|f'\| + \|f'\|^2) \le \frac{1}{4\tau}(1 + (4\tau)^{-1})^2.$$

These inequalities hold for every sample, every sample size, and every real radius. Continuous population mixed derivatives have finite maxima on any fixed compact admissible parameter rectangle. Their existence follows from the smooth construction just given; alternatively, differentiating the zero equation bounds each order by τ^{-1} times previously bounded derivatives and numerical L_r. This proves the stated dependence and supplies in particular the derivatives through order six used below. Endpoint derivatives are furnished by the real-radius extension. Finally, integration gives the useful deterministic remainder

$$\|f_\delta - f_0 - \delta f_0'\| \le \int_0^\delta (\delta - t)\|f_t''\|\, dt \le \frac{\delta^2}{8\tau}(1 + (4\tau)^{-1})^2.$$

□

Lemma C.5 (Signal and spatial symmetries)**.** *Under Assumptions 1, 2, and 3, for every admissible ϵ and every real auxiliary radius, $f_{-\epsilon,\delta} = -f_{\epsilon,\delta}$ and $f_{0,\delta} = 0$. Under Assumption 4, the minimizer is even and invariant under $x \mapsto x + 2\pi/k$. Its sine coefficients vanish and its nonzero frequencies are multiples of k.*

Proof. Replacing (ϵ, f) by $(-\epsilon, -f)$ negates the conditional score in Lemma C.4, including ridge. Uniqueness in Lemma C.3 proves oddness. At $\epsilon = 0$ the score at the origin is zero at every radius, proving the zero path directly.

Reflection changes signs of sine coordinates and translation rotates each cosine/sine pair by an orthogonal matrix. Both preserve its common weight μ_j, the RKHS norm, and the available-mode subspace, including missing pairs. When $q = \cos(kx)$, reflection and translation by $2\pi/k$ preserve both m and q. A change of integration variable leaves the conditional objective invariant, so uniqueness fixes the minimizer under these operations. Reflection eliminates sines. Translation invariance requires $e^{2\pi ij/k} = 1$ for each nonzero complex Fourier coefficient, hence j is a multiple of k. This does not restrict the exact minimizer to the two modes k and $3k$. □

Proposition C.6 (Affine specialization of the Functional SAM update)**.** *Suppose the prediction map in Singh et al. (2025), equation (11), is $\mathcal{F}(\theta) = B\theta + b$ with bounded linear B between the parameter and prediction inner-product spaces, and the output loss L is differentiable. For every permitted direction e and radius ρ, its direction equals*

$$-D\mathcal{F}(\theta + \rho e)^* \nabla L(\mathcal{F}(\theta)) = -B^* \nabla L(\mathcal{F}(\theta)) = -\nabla_\theta [L(\mathcal{F}(\theta))].$$

Proof. The published rule evaluates the prediction Jacobian at the perturbed parameter and the output-loss gradient at the original parameter, as displayed. For an increment v, $\mathcal{F}(\theta + v) - \mathcal{F}(\theta) = Bv$ exactly. Its derivative is therefore B at both parameters. Substitution proves the first equality. The directional derivative of the composite loss is $\langle \nabla L, Bv \rangle = \langle B^* \nabla L, v \rangle$, proving the second. Thus identical subsequent optimizer transformations and identical external regularization preserve the equality of directions. Proposition C.2 instead changes a samplewise objective by an exact margin shift; the two constructions are distinct. □

C.3 Uniform expansion and signed energy responses

The proof proceeds from a scalar Taylor residual to the optimized function. Lemma C.7 controls the residual and its radius derivative; Lemma C.8 transfers both through strong monotonicity. Propositions C.9–C.12 then prove Theorem 6 and Corollary 7.

Lemma C.7 (Uniform scalar and score remainders)**.** *Under Assumptions 1, 2, and 3, suppose a C^1 trial path $u_\delta \in \mathcal{H}$ satisfies $\sup_{[0,D]}(\|u_\delta\| + \|u'_\delta\|) \le M|\epsilon|$ for fixed finite M, with admissible $|\epsilon| \le 1$. Then*

$$\begin{aligned} F_{\epsilon,\delta}(u) &= (\tau I_{\mathcal{H}} + p_\delta T)u - \epsilon s_\delta I^* q \\ &\quad + I^* \left[\frac{\sigma'''(\delta)}{6} u^3 - \frac{\epsilon q \sigma''(\delta)}{2} u^2 \right] + I^* \rho_\delta, \\ \|I^* \rho\|_{C^1([0,D];\mathcal{H})} &\le C|\epsilon|^5. \end{aligned}$$

Here C depends only on M, $\|q\|_\infty$, and the numerical envelopes L_4, L_5, L_6 from Lemma C.4. Products inside I^* are pointwise L^2 functions.

Proof. Write

$$\begin{aligned} \frac{\sigma(\delta + t) - \sigma(\delta - t)}{2} &= p_\delta t + \frac{\sigma'''(\delta)}{6} t^3 + R_o(\delta, t), \\ \frac{\sigma(\delta + t) + \sigma(\delta - t)}{2} &= s_\delta + \frac{\sigma''(\delta)}{2} t^2 + R_e(\delta, t). \end{aligned}$$

For a smooth scalar h, the degree-r Taylor remainder is $t^{r+1}/r! \int_0^1 (1 - v)^r h^{(r+1)}(\delta + vt)\, dv$, also for negative t. Degree four for the odd combination and degree three for the even one cancel the

unwanted powers by parity. Apply the same calculation to σ' for the partial radius derivatives, and to the differentiated odd and even formulas for partial t derivatives. The resulting bounds are

$$|R_o| \le L_5|t|^5/120, \quad |\partial_\delta R_o| \le L_6|t|^5/120, \quad |\partial_t R_o| \le L_5|t|^4/24,$$
$$|R_e| \le L_4|t|^4/24, \quad |\partial_\delta R_e| \le L_5|t|^4/24, \quad |\partial_t R_e| \le L_4|t|^3/6.$$

These derivative estimates come from Taylor expansions of differentiated formulas, rather than differentiation of an order symbol. Define $\rho_\delta(x) = R_o(\delta, u_\delta(x)) - \epsilon q(x) R_e(\delta, u_\delta(x))$. Bounded evaluation gives $|u|, |u'| \le M|\epsilon|$ pointwise, so

$$\|\rho_\delta\|_\infty \le [L_5 M^5/120 + \|q\|_\infty L_4 M^4/24]|\epsilon|^5,$$
$$\left\| \frac{d\rho_\delta}{d\delta} \right\|_\infty \le [(L_6/120 + L_5/24)M^5 + \|q\|_\infty (L_5/24 + L_4/6)M^4]|\epsilon|^5.$$

The total derivative includes $\partial_t R_o\, u'$ and $-\epsilon q \partial_t R_e\, u'$, with ϵ fixed. Since m is a probability measure and $\|I^*\| \le 1$ by Lemma C.1, the two infinity bounds dominate the respective Hilbert bounds. The bounded derivatives also justify differentiation under I^*. Their sum gives the asserted constant. Only the bounded evaluation map is used for the pointwise products. □

Lemma C.8 (Residual transfer to the actual radius path)**.** *Under Assumptions 1, 2, and 3, for an admissible ϵ and a C^1 trial path u on $[0, D]$, set $v_\delta = F_{\epsilon,\delta}(u_\delta)$ and $r_\delta = f_{\epsilon,\delta} - u_\delta$. Then*

$$\|r_\delta\| \le \tau^{-1}\|v_\delta\|, \qquad \|r'_\delta\| \le \tau^{-1}\left[\|v'_\delta\| + \frac{1 + \|u'_\delta\|}{4}\|r_\delta\|\right].$$

If $\sup \|u'\| \le M$, then $\|r\|_{C^1} \le [\tau^{-1} + (1+M)/(4\tau^2)]\|v\|_{C^1}$.

Proof. The first bound is Lemma C.3, applied to this same population objective. Differentiate $F(f) = 0$ and $v = F(u)$, with $H(f) = F_f(f)$. Subtraction and $f' = r' + u'$ give

$$H(f)r' = -[F_\delta(f) - F_\delta(u)] - [H(f) - H(u)]u' - v'.$$

The two global Lipschitz bounds in Lemma C.4 and the inverse bound give the second inequality. Add it to the first. The coefficient of $\|v\|$ is $\tau^{-1} + (1+M)/(4\tau^2)$ and that of $\|v'\|$ is τ^{-1}. The larger coefficient and the supremum yield the C^1 estimate. □

Proposition C.9 (Norm-controlled linear and cubic responses)**.** *Under Assumptions 1, 2, and 3, there exist $\epsilon_0, C > 0$, depending only on fixed τ, D, K, q, such that*

$$\|f_{\epsilon,\cdot} - \epsilon \mathcal{A}.q\|_{C^1([0,D];\mathcal{H})} \le C|\epsilon|^3 \quad (|\epsilon| \le \epsilon_0).$$

Under Assumption 4, with fixed k, constants of the same permitted dependence give

$$f_{\epsilon,\delta} = \epsilon a_\delta \cos(kx) + \epsilon^3[b_{1,\delta}\cos(kx) + b_{3,\delta}\cos(3kx)] + r_{\epsilon,\delta}, \qquad \|r_{\epsilon,\cdot}\|_{C^1([0,D];\mathcal{H})} \le C|\epsilon|^5.$$

The coefficients are those defined in Section C.1.1, and both statements include $\epsilon = 0$ and all radii in $[0, D]$.

Proof. Let $L_\delta = \tau I_{\mathcal{H}} + p_\delta T$ and $u_{1,\delta} = L_\delta^{-1} s_\delta I^* q$. Positivity of T and the closed-range argument of Lemma C.3 give $\|L_\delta^{-1}\| \le \tau^{-1}$. Fourier identities in Lemma C.1 give $u_1 = \mathcal{A}_\delta q$, with zero coefficients on absent modes. This finite Fourier sum belongs to $\mathcal{H}$. Its denominators are at least τ, so it is smooth through the radius endpoints. In particular,

$$\|u_1\| \le \|q\|_2/\tau, \quad L_\delta u'_1 = p_\delta I^* q - \sigma''(\delta) T u_1, \quad \|u'_1\| \le \|q\|_2[1/(4\tau) + L_2/\tau^2].$$

Choose $\bar\epsilon = [2(1 + \|q\|_\infty)]^{-1}$. Then $|\epsilon| \le \bar\epsilon \le 1/2$ ensures strict admissibility. The trial $u = \epsilon u_1$ satisfies Lemma C.7 with a fixed M. Its linear residual cancels. The remaining cubic terms have C^1 norm at most $C|\epsilon|^3$: pointwise $(u^3)' = 3u^2u'$ and $(u^2)' = 2uu'$, and their coefficient derivatives are bounded by L_4, L_3. The fifth-order error is bounded by third order since $|\epsilon| \le 1$. Lemma C.8 proves the first conclusion. If $q = 0$, the zero-path argument in Lemma C.5 gives it exactly.

For $q = \cos(kx)$ put $c = \cos(kx)$, so $u_1 = a_\delta c$. Cancel the cubic residual by defining

$$L_\delta u_3 = I^*\left[\frac{\sigma''(\delta)}{2} q u_1^2 - \frac{\sigma'''(\delta)}{6} u_1^3\right] = I^*\left[\left(\frac{\sigma''(\delta)a_\delta^2}{2} - \frac{\sigma'''(\delta)a_\delta^3}{6}\right) c^3\right].$$

Using $c^3 = (3\cos(kx) + \cos(3kx))/4$ and $I^* \cos(jx) = \mu_j \cos(jx)$ gives exactly

$$u_3 = b_{1,\delta}\cos(kx) + b_{3,\delta}\cos(3kx), \quad B_\delta = \sigma''(\delta)a_\delta^2/8 - \sigma'''(\delta)a_\delta^3/24,$$
$$b_{1,\delta} = \frac{3\mu_k B_\delta}{\tau + \mu_k p_\delta}, \qquad b_{3,\delta} = \frac{\mu_{3k} B_\delta}{\tau + \mu_{3k} p_\delta}.$$

These coefficients and their first derivatives are bounded on $[0, D]$: their expressions use bounded sigmoid derivatives, the bounds on a, a', and denominators at least τ. The two cosine functions have finite RKHS norm by the positive-mode assumption. Thus $\|u_3\|_{C^1} < \infty$, depending only on fixed τ, D, K, k.

Take $u = \epsilon u_1 + \epsilon^3 u_3$. On the selected signal interval its C^1 norm is bounded by $|\epsilon|(\|u_1\|_{C^1} + \|u_3\|_{C^1})$, which supplies the required M. Its linear and cubic equations cancel orders one and three in the score. The remaining polynomial substitutions are exactly

$$u^3 - \epsilon^3 u_1^3 = 3\epsilon^5 u_1^2 u_3 + 3\epsilon^7 u_1 u_3^2 + \epsilon^9 u_3^3, \quad \epsilon(u^2 - \epsilon^2 u_1^2) = 2\epsilon^5 u_1 u_3 + \epsilon^7 u_3^2.$$

Each term and its total radius derivative are bounded by $C|\epsilon|^5$ using bounded evaluation of u_1, u_3 and their derivatives and $|\epsilon| \le 1$. Application of I^* preserves the bound. Lemma C.7 controls the remaining score error at the same order, and Lemma C.8 gives the claimed C^1 error for the actual minimizer. All higher modes of the minimizer are contained in r. □

Proposition C.10 (Periodic specialization of the general cubic response)**.** *Under Assumptions 1, 2, and 3, define*

$$u_{1,\delta} = s_\delta(\tau I + p_\delta T)^{-1} I^* q, \qquad h_\delta = I u_{1,\delta},$$
$$\zeta_\delta = \frac{\sigma''(\delta)}{2} q h_\delta^2 - \frac{\sigma'''(\delta)}{6} h_\delta^3, \qquad u_{3,\delta} = (\tau I + p_\delta T)^{-1} I^* \zeta_\delta.$$

There exist $\epsilon_0, C > 0$, depending only on the fixed setting, such that

$$\|f_{\epsilon,\cdot} - \epsilon u_{1,\cdot} - \epsilon^3 u_{3,\cdot}\|_{C^1([0,D];\mathcal{H})} \le C|\epsilon|^5 \qquad (|\epsilon| \le \epsilon_0).$$

If $Te_j = \mu_j e_j$ with $\mu_j > 0$ and $Ie_j = \sqrt{\mu_j}\varphi_j$, then

$$\langle I u_{3,\delta}, \varphi_j \rangle_{L^2(m)} = \frac{\mu_j}{\tau + \mu_j p_\delta} \langle \zeta_\delta, \varphi_j \rangle_{L^2(m)}.$$

Proof. The inverse bound in Lemma C.3, together with the bounds for u_1, gives uniform control of the trial-path coefficients. Products of q and h_δ are bounded in $L^2(m)$, and differentiating the resolvent equation for u_1 shows that ζ_δ and $u_{3,\delta}$ also have uniform C^1 bounds. Apply Lemma C.7 to the trial path $u = \epsilon u_1 + \epsilon^3 u_3$. Its linear part cancels because $(\tau I + p_\delta T)u_1 = s_\delta I^* q$. The order-$\epsilon^3$ part is

$$(\tau I + p_\delta T)u_3 + I^*\left[\frac{\sigma'''(\delta)}{6} h_\delta^3 - \frac{\sigma''(\delta)}{2} q h_\delta^2\right] = 0$$

by the definition of u_3. The displayed scalar remainder lemma controls the remaining substitutions $u^3 - \epsilon^3 u_1^3$ and $\epsilon(u^2 - \epsilon^2 u_1^2)$ by $C|\epsilon|^5$; the same calculation after one total radius derivative has the same order. Hence the score residual of the trial path is $O_{C^1}(\epsilon^5)$ in $\mathcal{H}$. Lemma C.8 transfers this residual to the actual minimizer and proves the first claim. Finally, self-adjointness gives

$$\langle I u_3, \varphi_j \rangle = \langle u_3, I^* \varphi_j \rangle = \frac{\mu_j}{\tau + \mu_j p_\delta} \langle \zeta_\delta, \varphi_j \rangle,$$

which proves the mode identity. □

Proposition C.11 (Transfer gains and zero-radius coefficients)**.** *Under Assumptions 1 and 2, for $\mu > 0$ and $\delta \ge 0$,*

$$A'_\delta(\mu) = \frac{\mu p_\delta(\tau + \mu s_\delta^2)}{(\tau + \mu p_\delta)^2} > 0, \qquad \frac{A_\delta(\mu)}{A_0(\mu)} = \frac{2 s_\delta(\tau + \mu/4)}{\tau + \mu p_\delta}.$$

The ratio is strictly increasing in μ when $\delta > 0$, and equals one at $\delta = 0$. Under Assumptions 3 and 4 as well, $0 < a_0 < 2$, $a'_0 = a_0/2$, and

$$b_{3,0} = \frac{\mu_{3k} a_0^3}{192(\tau + \mu_{3k}/4)} > 0, \qquad b'_{3,0} = \frac{\mu_{3k} a_0^2 (a_0 - 2)}{128(\tau + \mu_{3k}/4)} < 0.$$

Proof. Write $s = \sigma(\delta)$, $p = s(1-s)$, and $p' = p(1-2s)$. The quotient derivative numerator is $\mu[p\tau + \mu(p^2 - sp')]$. Since $p^2 - sp' = ps^2$, this gives the first formula and its positive sign. The identities $s_0 = 1/2$ and $p_0 = 1/4$ give the ratio. Its derivative in μ is

$$\frac{2s_\delta \tau(1/4 - p_\delta)}{(\tau + \mu p_\delta)^2} > 0 \qquad (\delta > 0),$$

because $s_\delta > 1/2$ implies $p_\delta < 1/4$. At $\mu = 0$ the separate definition is $A_\delta(0) = 0$; no quotient is asserted there.

At zero, $a_0 = \mu_k/[2(\tau + \mu_k/4)] \in (0, 2)$ and $p_0' = 0$ imply $a_0' = a_0/2$. The derivative recursion gives $\sigma''(0) = 0$, $\sigma'''(0) = -1/8$, and $\sigma''''(0) = 0$. Hence $B_0 = a_0^3/192$, and direct differentiation gives

$$B_0' = \frac{(-1/8)a_0^2}{8} - \frac{(-1/8)3a_0^2(a_0/2)}{24} = -a_0^2/64 + a_0^3/128 = a_0^2(a_0 - 2)/128.$$

The derivative of the denominator of b_3 vanishes at zero. Substituting these two values yields both formulas and strict signs. □

Proposition C.12 (Signed modal and total-energy derivatives)**.** *Under Assumptions 1, 2, 3, and 4, for sufficiently small nonzero ϵ,*

$$\partial_\delta E_k(f_{\epsilon,\delta})\big|_0 = \frac{\epsilon^2 a_0^2}{2\mu_k} + O(\epsilon^4) > 0, \qquad \partial_\delta E_{3k}(f_{\epsilon,\delta})\big|_0 = \frac{\epsilon^6 b_{3,0} b_{3,0}'}{\mu_{3k}} + O(\epsilon^8) < 0,$$

$$\partial_\delta \|f_{\epsilon,\delta}\|_{\mathcal{H}}^2\big|_0 = \frac{\epsilon^2 a_0^2}{2\mu_k} + O(\epsilon^4) > 0.$$

All constants depend only on fixed τ, D, K, k. For each such fixed ϵ the three signs persist on some $[0, d_\epsilon]$ with $0 < d_\epsilon \le D$.

Proof. Proposition C.9 and the bounded coefficient functional in Lemma C.1 imply, in scalar $C^1([0, D])$,

$$c_k(f) = \epsilon a + \epsilon^3 b_1 + O(\epsilon^5), \qquad c_{3k}(f) = \epsilon^3 b_3 + O(\epsilon^5).$$

Bounded linear extraction gives the same errors for their first derivatives. Since $E_j' = c_j c_j'/\mu_j$, the fundamental product is $\epsilon^2 a_0 a_0'/\mu_k + O(\epsilon^4)$; every other product has degree at least four with bounded coefficient. The third-harmonic product is $\epsilon^6 b_{3,0} b_{3,0}'/\mu_{3k} + O(\epsilon^8)$: an order-five error times an order-three term has degree eight, and the product of errors has degree ten, bounded by degree eight for $|\epsilon| \le 1$. Proposition C.11 supplies the values and nonzero signs.

For the total norm use $(\|f\|^2)' = 2\langle f, f'\rangle_{\mathcal{H}}$ directly. The C^1 expansion gives $f = \epsilon a \cos(kx) + O_{\mathcal{H}}(\epsilon^3)$ and $f' = \epsilon a' \cos(kx) + O_{\mathcal{H}}(\epsilon^3)$. Cauchy–Schwarz bounds cross terms by order four and the residual product by order six. The leading term is $2\epsilon^2 a_0 a_0'/(2\mu_k) = \epsilon^2 a_0^2/(2\mu_k)$. The Hilbert-norm bound controls all remaining modes together.

To verify strict signs, let $L_+ = a_0^2/(2\mu_k) > 0$ and $L_- = -b_{3,0} b_{3,0}'/\mu_{3k} > 0$. Let $C_* \ge 1$ be the maximum of the three finite remainder constants just obtained. Intersect the expansion interval with

$$0 < |\epsilon| \le \min\left\{1, \bar{\epsilon}, \sqrt{\frac{\min(L_+, L_-)}{2C_*}}\right\}.$$

Each error is then at most half the magnitude of its leading term, proving the signs for either sign of nonzero ϵ. Lemma C.4 makes all three actual energy derivatives continuous in radius. Each has a strict value at zero, so their three relative neighborhoods intersect in an interval $[0, d_\epsilon]$ with positive $d_\epsilon \le D$. This interval need not be uniform in ϵ. At $\epsilon = 0$ all energies vanish, consistently with excluding that case from strict signs. □

D FINITE-SAMPLE COVARIANCE AND SAMPLING LIMITS

D.1 CONDITIONAL COVARIANCE ACCURACY

Theorem D.1 gives the explicit constant and the conditional refinement of Theorem 9. We retain the general normalized-kernel setting, fixed $\tau > 0$, and balanced independent labels. The result below holds for every $n \geq 1$ and every fixed $\delta \geq 0$. All covariance operators act on the real RKHS, and $\|\cdot\|_1$ denotes trace norm. Put $s = \sigma(\delta)$, $p = \sigma'(\delta)$, $T_n = n^{-1}\sum_i K_{X_i} \otimes K_{X_i}$, $M_n = n^{-1}\sum_i Y_i K_{X_i}$, and

$$H_n = \tau I + pT_n, \qquad \widetilde{f}_n = sH_n^{-1}M_n,$$
$$C_n = s^2 H_n^{-1} T_n H_n^{-1}, \qquad C = s^2(\tau I + pT)^{-1}T(\tau I + pT)^{-1}.$$

Define the explicit constant

$$\mathfrak{B}_\delta = \frac{\sqrt{15}}{\tau}\left(\frac{|\sigma''(\delta)|s^2}{2\tau^2} + \frac{s^3}{24\tau^3}\right).$$

Theorem D.1 (Finite-sample covariance accuracy)**.** *In the preceding setting, conditional on any input sample, both $\hat{f}_{n,\delta}$ and $\widetilde{f}_n$ have mean zero, and*

$$\mathrm{Cov}(\sqrt{n}\widetilde{f}_n \mid X_{1:n}) = C_n, \qquad \left(E[\|\hat{f}_{n,\delta} - \widetilde{f}_n\|^2 \mid X_{1:n}]\right)^{1/2} \leq \mathfrak{B}_\delta n^{-3/2}.$$

Consequently, with $N_n = 2s\mathfrak{B}_\delta/(\tau n) + \mathfrak{B}_\delta^2/n^2$, one has

$$\|\mathrm{Cov}(\sqrt{n}\hat{f}_{n,\delta} \mid X_{1:n}) - C_n\|_1 \leq N_n, \tag{D.1}$$

$$\|EC_n - C\|_1 \leq \frac{4s^2 p}{\tau^3 n}(1 - \mathrm{tr}(T^2)), \tag{D.2}$$

$$\|\mathrm{Cov}(\sqrt{n}\hat{f}_{n,\delta}) - C\|_1 \leq N_n + \frac{4s^2 p}{\tau^3 n}(1 - \mathrm{tr}(T^2)). \tag{D.3}$$

These are absolute errors. They do not imply finite-sample monotonicity or accurate relative errors in directions of small limiting variance. The conditional bounds admit a design-dependent refinement. Set $v_n = \mathrm{tr}(C_n)$ and $\mathfrak{B}_{\delta,n} = \sqrt{15}\tau^{-1}(|\sigma''(\delta)|v_n/2 + v_n^{3/2}/24)$. Then $\mathfrak{B}_\delta$ can be replaced by $\mathfrak{B}_{\delta,n}$ in the conditional estimator bound, and the right-hand side of (D.1) *can be replaced by $2\sqrt{v_n}\mathfrak{B}_{\delta,n}/n + \mathfrak{B}_{\delta,n}^2/n^2$.*

Proof. Conditional Hilbert moments. For deterministic vectors v_i of norm at most one in any real Hilbert space and independent Rademacher signs Y_i,

$$E\left\|n^{-1}\sum_i Y_i v_i\right\|^2 \leq n^{-1}, \qquad E\left\|n^{-1}\sum_i Y_i v_i\right\|^6 \leq 15n^{-3}.$$

For the second assertion, expand the cube of $\|\sum_i Y_i v_i\|^2 = \sum_{ij} Y_i Y_j \langle v_i, v_j\rangle$. An expectation survives only when each index in the six positions occurs an even number of times. Each surviving term has absolute value at most one. Every such six-tuple is compatible with at least one of the fifteen pairings of six positions; each pairing allows n^3 assignments. Summing absolute values proves the bound, including repeated-index tuples. The second-moment identity follows by centering the cross terms.

Apply these facts conditional on the inputs to $v_i = K_{X_i}$ and also to $v_i = K_{X_i}^{\otimes 3}$ in the Hilbert tensor product $\mathcal{H}^{\otimes 3}$. Both have norm one. With $S_n = n^{-1}\sum_i Y_i K_{X_i}^{\otimes 3}$, Hölder's inequality gives

$$E[\|S_n\|^2\|M_n\|^4 \mid X_{1:n}] \leq (E[\|S_n\|^6 \mid X_{1:n}])^{1/3}(E[\|M_n\|^6 \mid X_{1:n}])^{2/3} \leq 15n^{-3}.$$

The Hölder bound applies to the dependent pair (S_n, M_n).

Nonlinear remainder. The empirical score is $F_n(f) = \tau f - n^{-1}\sum_i Y_i \sigma(\delta - Y_i f(X_i))K_{X_i}$. Its Hessian is at least τI, giving $\|f - \hat{f}_{n,\delta}\| \leq \tau^{-1}\|F_n(f)\|$ by integration along the segment between

these two points and Cauchy–Schwarz. The logistic derivatives satisfy $\sigma''' = \sigma'(1 - 6\sigma')$, hence $|\sigma'''| \le 1/4$ globally. Taylor expansion through the quadratic term yields

$$F_n(f) = -sM_n + H_n f - \tfrac{1}{2}\sigma''(\delta)n^{-1}\sum_i Y_i f(X_i)^2 K_{X_i} + r_n(f), \qquad \|r_n(f)\| \le \|f\|^3/24.$$

The quadratic vector is the contraction of S_n with $f \otimes f$, whose norm is at most $\|S_n\|\|f\|^2$ by the Hilbert tensor norm. Since $H_n \succeq \tau I$, $\|\widetilde{f}_n\| \le (s/\tau)\|M_n\|$. The linear terms cancel at $\widetilde{f}_n$. The preceding conditional moment bounds and the triangle inequality in L^2 therefore give

$$\left(E[\|F_n(\widetilde{f}_n)\|^2 \mid X_{1:n}]\right)^{1/2} \le \sqrt{15}\left(\frac{|\sigma''(\delta)|s^2}{2\tau^2} + \frac{s^3}{24\tau^3}\right) n^{-3/2}.$$

The residual inequality proves the claimed estimator bound. All random elements are measurable by the contraction construction in the general-kernel proof; bounded second moments also follow from its residual inequality.

Covariance and symmetry. Replacing every label by its negative sends the objective at f to its value at $-f$. Uniqueness sends its minimizer to its negative. Conditional Rademacher symmetry thus makes its mean zero. The linearized estimator is also odd and centered, with the displayed exact conditional covariance. Let $e = \hat{f}_{n,\delta} - \widetilde{f}_n$. For rank-one operators, $\|a \otimes b\|_1 = \|a\|\|b\|$. Expanding the outer products and using Cauchy–Schwarz bounds their covariance difference by

$$n\left(2\sqrt{E[\|\widetilde{f}_n\|^2 \mid X]}\sqrt{E[\|e\|^2 \mid X]} + E[\|e\|^2 \mid X]\right).$$

Here the first square root is at most $s/(\tau\sqrt{n})$. This proves (D.1), and its expectation has the same bound. Conditional means vanish, so the total covariance is the expected conditional covariance.

For the refinement, the same pairing argument for arbitrary deterministic vectors w_i gives $E\|n^{-1}\sum_i Y_i w_i\|^6 \le 15n^{-6}(\sum_i \|w_i\|^2)^3$: bound each surviving product of inner products by the product of its six vector norms and sum over the fifteen pairings. Take $w_i = sH_n^{-1}K_{X_i}$, so $n^{-1}\sum_i \|w_i\|^2 = v_n$ and $E[\|\widetilde{f}_n\|^6 \mid X] \le 15v_n^3/n^3$. Hölder then gives $E[\|S_n\|^2\|\widetilde{f}_n\|^4 \mid X] \le 15v_n^2/n^3$. Repeating the residual argument proves the refined constant, while the exact identity $E[\|\widetilde{f}_n\|^2 \mid X] = v_n/n$ proves its covariance bound.

Centered design expansion. Write $Q = T_n - T$, $R = (\tau I + pT)^{-1}$ and $R_n = (\tau I + pT_n)^{-1}$. The resolvent identities give

$$\Delta := R_n - R = -pRQR_n = -pR_nQR, \qquad E_n := R_n - R + pRQR = p^2 RQR_nQR.$$

The Hilbert–Schmidt product bound $\|AB\|_1 \le \|A\|_{\mathrm{HS}}\|B\|_{\mathrm{HS}}$ and inverse bounds give

$$\|E_n\|_1 \le p^2\tau^{-3}\|Q\|_{\mathrm{HS}}^2, \qquad \|\Delta\|_{\mathrm{HS}} \le p\tau^{-2}\|Q\|_{\mathrm{HS}}.$$

Writing $L = -pRQR$, the identity $R_n^2 - R^2 - (RL + LR) = RE_n + E_nR + \Delta^2$ yields the trace-norm bound $3p^2\tau^{-4}\|Q\|_{\mathrm{HS}}^2$. Since $p > 0$ for every finite radius,

$$C_n - C = \frac{s^2}{p}\{R_n - R - \tau(R_n^2 - R^2)\}.$$

The individual resolvents need not be trace class, but every difference and every linear or remainder term in this identity is trace class: Q is trace class and the products above satisfy the stated bounds. The linear term is $(s^2/p)\{L - \tau(RL + LR)\}$, whose expectation vanishes because $EQ = 0$. The remaining trace norm is bounded by $4s^2p\tau^{-3}\|Q\|_{\mathrm{HS}}^2$. Finally, independence in the Hilbert space of Hilbert–Schmidt operators gives

$$E\|Q\|_{\mathrm{HS}}^2 = n^{-1}E\|K_X \otimes K_X - T\|_{\mathrm{HS}}^2 = n^{-1}(1 - \mathrm{tr}(T^2)),$$

since $\|K_X \otimes K_X\|_{\mathrm{HS}}^2 = 1$. This proves the design bound; the triangle inequality completes the total bound. □

For an evaluation operator $\mathcal{J}$, $\|\mathcal{J}\| \le 1$ implies the same absolute trace-norm bounds after multiplication by $\mathcal{J}$ and $\mathcal{J}^*$. For a fixed RKHS probe h, the variance error is at most $\|h\|^2$ times the corresponding displayed bound. To compare a radius with zero, add the unconditional total bounds in (D.3) at the two radii. Thus a positive limiting variance increase exceeding that sum certifies a finite-sample increase for this specified probe; a bound that exceeds the limiting increase makes no directional assertion.

D.2 FIXED-RADIUS HILBERT LIMITS

For this part, use the fixed signal law and statistical coordinates in Appendix C.1.1. Localization is established before linearization, and the same-sample radius path is differentiated separately.

Lemma D.2 (Fixed-point vector and operator moments)**.** *Under Assumptions 1, 2, and 3, empirical minimizers at deterministic radii are measurable $\mathcal{H}$-valued variables. For deterministic $f \in \mathcal{H}$ and $\delta \in [0, D]$, put*

$$v = -Y\sigma(\delta - Yf(X))K_X, \quad w = -Y\sigma'(\delta - Yf(X))K_X, \quad B = \sigma'(\delta - Yf(X))K_X \otimes K_X.$$

Then

$$\mathbb{E}\|(P_n - P)v\|^2 \le n^{-1}, \quad \mathbb{E}\|(P_n - P)w\|^2 \le (16n)^{-1}, \quad \mathbb{E}\|(P_n - P)B\|_{\mathrm{HS}}^2 \le (16n)^{-1}.$$

Each fluctuation is $O_p(n^{-1/2})$, the last also in operator norm.

Proof. The feature map and each fixed $f(x)$ are continuous by Lemma C.1. The tensor map is continuous into the separable Hilbert space of Hilbert–Schmidt operators. Thus all three variables are strongly measurable, with norms bounded respectively by $1, 1/4, 1/4$. Independence and centering, in either Hilbert space, give

$$\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^{n}(U_i - \mathbb{E}U)\right\|^2 = \frac{1}{n}\mathbb{E}\|U - \mathbb{E}U\|^2 = \frac{1}{n}(\mathbb{E}\|U\|^2 - \|\mathbb{E}U\|^2).$$

These identities prove the bounds. For envelope b, Markov's inequality bounds the probability of exceeding $t/\sqrt{n}$ by b^2/t^2. Operator norm is no greater than Hilbert–Schmidt norm.

For estimator measurability, take a converging sequence of samples of fixed size in the circle-times-discrete-label product topology. Evaluate their scores at the limiting sample's minimizer u. Continuity of K_x and $u(x)$ gives convergence of these scores in $\mathcal{H}$ to zero. Lemma C.3 bounds the respective minimizers' distances from u by these score norms divided by τ. The minimizer is therefore continuous in the sample, hence measurable. This argument holds at every deterministic radius, including a radius depending on n. □

Lemma D.3 (Global localization and fixed-radius linearization)**.** *Under Assumptions 1, 2, and 3, for each fixed $\delta \in [0, D]$,*

$$\|\hat{f}_{n,\delta} - f_{\epsilon,\delta}\| \le \tau^{-1}\|(P_n - P)g_\delta\| = O_p(n^{-1/2}),$$

$$\hat{f}_{n,\delta} - f_{\epsilon,\delta} = -H_\delta^{-1}(P_n - P)g_\delta + O_p(n^{-1}).$$

For every deterministic $\delta \in [0, D]$, also

$$\|\hat{f}_{n,\delta} - f_\tau\| \le \tau^{-1}(\|(P_n - P)g\| + \delta/4).$$

All norms are Hilbert norms; constants depend only on the fixed setting quantities.

Proof. Population stationarity gives $Pg_\delta = -\tau f_{\epsilon,\delta}$, rather than a zero mean score without ridge. Hence the empirical score at that population minimizer is exactly $(P_n - P)g_\delta$. Residual comparison in Lemma C.3 and the moments in Lemma D.2 prove localization before Taylor expansion. Writing $h = \hat{f}_{n,\delta} - f_{\epsilon,\delta}$, the empirical zero equation gives

$$\begin{aligned} 0 &= (P_n - P)g_\delta + H_\delta h \\ &\quad + (P_n - P)[\sigma'(\delta - Yf_{\epsilon,\delta}(X))K_X \otimes K_X]h + e_n, \qquad \|e_n\| \le \|h\|^2/8. \end{aligned}$$

The error bound integrates the globally $1/4$-Lipschitz Hessian from Lemma C.4 along the segment. The moment lemma and localization bound its last two terms by $O_p(n^{-1})$. Application of the inverse, of norm at most τ^{-1}, gives the linearization. For the final bound, at f_τ the empirical score changes by at most $\delta/4$ from radius zero to δ, since $\sigma' \le 1/4$. The radius-zero score there is $(P_n - P)g$. Another application of global residual comparison proves the result. □

Lemma D.4 (Hilbert Gaussian limit at fixed radius)**.** *Under Assumptions 1, 2, and 3, for each fixed* $\delta \in [0, D]$,

$$\sqrt{n}(\hat{f}_{n,\delta} - f_{\epsilon,\delta}) \Rightarrow N_{\mathcal{H}}(0, C_\delta), \qquad C_\delta = H_\delta^{-1} \operatorname{Cov}(g_\delta) H_\delta^{-1}.$$

The covariance is positive trace class, with trace at most τ^{-2}*; degeneracy is permitted.*

Proof. Let $\xi = g_\delta - Pg_\delta$. Then $\mathbb{E}\xi = 0$ and $\mathbb{E}\|\xi\|^2 \le 1$ by Lemma D.2. Define $Su = \mathbb{E}[\langle \xi, u\rangle \xi]$. For any complete orthonormal basis (e_j),

$$\sum_j \langle e_j, Se_j\rangle = \mathbb{E}\sum_j \langle e_j, \xi\rangle^2 = \mathbb{E}\|\xi\|^2 < \infty.$$

Let Π_m project on its first m vectors. The finite-rank operators $\mathbb{E}[(\Pi_m\xi) \otimes (\Pi_m\xi)]$ approximate S in operator norm, because their difference norm is at most

$$\begin{aligned}\mathbb{E}[\|\xi - \Pi_m\xi\|(\|\xi\| + \|\Pi_m\xi\|)] &\le 2(\mathbb{E}\|\xi - \Pi_m\xi\|^2\, \mathbb{E}\|\xi\|^2)^{1/2} \\ &\longrightarrow 0.\end{aligned}$$

Thus S is compact, positive, and self-adjoint. To recall its spectral construction, maximize the quadratic form on the unit sphere. A maximizing sequence has a weak subsequential limit by the coordinate argument in Lemma C.3; compactness makes its images converge strongly. A positive maximum forces the limiting maximizer to have unit norm. Directional differentiation gives an eigenvector. Iterating on the invariant orthogonal complement gives positive eigenvectors spanning the closure of the range. Compactness forces any infinite list of positive eigenvalues to tend to zero and excludes a nonzero operator on the remaining complement. The diagonal identity above makes their eigenvalues (λ_j) summable.

Independent standard normals (Z_j) therefore define an $L^2(\mathcal{H})$-convergent series $G_\delta = \sum_j \sqrt{\lambda_j} Z_j u_j$. Every linear functional is centered normal, by scalar characteristic functions and L^2 convergence, with covariance S. This constructs an actual Hilbert Gaussian, allowing a finite or empty positive-eigenvalue series.

Put $W_n = n^{-1/2}\sum_{i=1}^n \xi_i$. For fixed m, the finite-dimensional iid central limit theorem gives $\Pi_m W_n \Rightarrow \Pi_m G_\delta$. The projected variables are centered real vectors with finite covariance, including the singular case. One can obtain that theorem by expanding the characteristic function of each scaled summand as $1 - t^\top V t/(2n) + o(n^{-1})$, taking its nth power, and applying the finite-dimensional characteristic-function continuity criterion. We also use the bounded-Lipschitz criterion for weak convergence on a separable metric space; these standard finite-dimensional and weak-convergence tools are described in Billingsley (1999). The infinite-dimensional passage is established here through the exact tail identity

$$\mathbb{E}\|(I_{\mathcal{H}} - \Pi_m)W_n\|^2 = \mathbb{E}\|(I_{\mathcal{H}} - \Pi_m)\xi\|^2 \longrightarrow 0,$$

uniformly in n. The same tail moment holds for G_δ by its covariance. For a function bounded by one and with Lipschitz constant one, replacing W_n by $\Pi_m W_n$ changes its expectation by at most the square root of this tail moment. The Gaussian replacement has the same bound, and the intervening finite-dimensional difference tends to zero. Taking first $n \to \infty$ and then $m \to \infty$ proves bounded-Lipschitz convergence in $\mathcal{H}$, and hence weak convergence.

For an explicit tightness check, fix $\alpha > 0$. Choose increasing m_l so that the tail moment is at most $\alpha 2^{-l} l^{-2}/2$. Markov's inequality and the union bound imply all restrictions $\|(I_{\mathcal{H}} - \Pi_{m_l})W_n\| \le 1/l$ simultaneously except on an event of probability at most $\alpha/2$. Choose R with $\mathbb{E}\|\xi\|^2/R^2 \le \alpha/2$. The closed set

$$\{u : \|u\| \le R, \quad \|(I_{\mathcal{H}} - \Pi_{m_l})u\| \le 1/l \text{ for every } l \ge 1\}$$

is compact: its finite projections are bounded and its tails tend uniformly to zero, giving total boundedness; it is closed in the complete space. It has probability at least $1 - \alpha$ for every n. This establishes Hilbert-space tightness.

Multiplication of the linearization in Lemma D.3 by $\sqrt{n}$ leaves $O_p(n^{-1/2})$. Bounded linear maps preserve the proved weak limit. A vanishing norm perturbation also preserves it: for bounded Lipschitz tests split the error at any positive threshold, and then let the threshold decrease. Hence the estimator limit is $-H_\delta^{-1}G_\delta$. Self-adjointness gives the displayed covariance, and

$$\operatorname{tr} C_\delta = \mathbb{E}\|H_\delta^{-1}G_\delta\|^2 \le \tau^{-2}\mathbb{E}\|\xi\|^2 \le \tau^{-2}.$$

□

D.3 SAME-SAMPLE SHRINKING-RADIUS EXPANSIONS

Proposition D.5 (Marginal and paired shrinking-radius expansions)**.** *Under Assumptions 1, 2, 3, and 5, in $\mathcal{H}$,*

$$\hat{f}_{n,\delta_n} - f_\tau = -H_\tau^{-1}(P_n - P)g - \delta_n d + O_p(n^{-1} + \delta_n n^{-1/2} + \delta_n^2),$$
$$\hat{f}_{n,\delta_n} - \hat{f}_{n,0} = -\delta_n d + O_p(\delta_n n^{-1/2} + \delta_n^2).$$

The constants depend only on fixed K, τ, D, q, ϵ, not on n or the shrinking sequence.

Proof. Write $\delta = \delta_n$ and $h = \hat{f}_{n,\delta} - f_\tau$. Lemma D.3 first gives $h = O_p(n^{-1/2} + \delta)$. At the fixed population point define

$$B_n = (P_n - P)[\sigma'(-Yf_\tau(X))K_X \otimes K_X], \quad w_n = (P_n - P)[-Y\sigma'(-Yf_\tau(X))K_X].$$

Since $\sigma'(-t) = \ell''(t)$, the population score derivatives are precisely H_τ and R_τ. Joint expansion at $(f_\tau, 0)$ gives

$$0 = (P_n - P)g + H_\tau h + \delta R_\tau + B_n h + \delta w_n + e_n, \qquad \|e_n\| \le (\|h\| + \delta)^2/8.$$

Indeed the second derivative along $(f_\tau + th, t\delta)$ is an average of $-Y\sigma''$ times $(\delta - Yh(X))^2 K_X$. Its norm is at most $(\|h\| + \delta)^2/4$, and the Taylor weight integrates to $1/2$. Lemma D.2 gives $\|B_n\|_{\mathrm{op}}, \|w_n\| = O_p(n^{-1/2})$. The three errors are respectively

$$O_p(n^{-1} + \delta n^{-1/2}), \quad O_p(\delta n^{-1/2}), \quad O_p(n^{-1} + 2\delta n^{-1/2} + \delta^2).$$

Applying H_τ^{-1} proves the marginal expansion. Products of these finitely many probability bounds require only a union bound, not independence of the factors.

For the paired expansion, define separately at the actual radius-zero empirical minimizer

$$\widehat{H}_n = H_{P_n,0}(\hat{f}_{n,0}), \quad \widehat{R}_n = -P_n[Y\sigma'(-Y\hat{f}_{n,0}(X))K_X], \quad d_n = \widehat{H}_n^{-1}\widehat{R}_n.$$

Radius-zero localization, fixed-point moments, and the global derivative bounds of Lemma C.4 give

$$\|\widehat{H}_n - H_\tau\|_{\mathrm{op}} \le \|B_n\|_{\mathrm{op}} + \tfrac{1}{4}\|\hat{f}_{n,0} - f_\tau\| = O_p(n^{-1/2}),$$
$$\|\widehat{R}_n - R_\tau\| \le \|w_n\| + \tfrac{1}{4}\|\hat{f}_{n,0} - f_\tau\| = O_p(n^{-1/2}).$$

Both inverse norms are at most τ^{-1} and $\|R_\tau\| \le 1/4$. The resolvent identity gives

$$\|d_n - d\| \le \tau^{-1}\|\widehat{R}_n - R_\tau\| + (4\tau^2)^{-1}\|\widehat{H}_n - H_\tau\|_{\mathrm{op}} = O_p(n^{-1/2}).$$

The same empirical sample defines the smooth path with $\hat{f}'_{n,0} = -d_n$. Integrating the deterministic second derivative bound in Lemma C.4 gives

$$\|\hat{f}_{n,\delta} - \hat{f}_{n,0} + \delta d_n\| \le \frac{\delta^2}{8\tau}(1 + (4\tau)^{-1})^2.$$

Substitution proves the paired expansion with no marginal n^{-1} residue. This remains valid at arbitrarily small positive radii; at radius zero the paired difference and remainder are identically zero. □

D.4 CRITICAL RADII AND THE ZERO-DRIFT CASE

Proposition D.6 (Explicit exponent specialization)**.** *Under Assumptions 1, 2, 3, and 5, let $\delta_n = \eta n^{-\gamma}$ with fixed $\eta > 0$ and $\gamma > 0$. If $G \sim N_{\mathcal{H}}(0, \Sigma)$, then*

$$\sqrt{n}(\hat{f}_{n,\delta_n} - f_\tau) \Rightarrow -H_\tau^{-1}G \quad (\gamma > 1/2),$$
$$\sqrt{n}(\hat{f}_{n,\delta_n} - f_\tau) \Rightarrow -H_\tau^{-1}G - \eta d \quad (\gamma = 1/2),$$
$$n^\gamma(\hat{f}_{n,\delta_n} - f_\tau) \to_p -\eta d \quad (0 < \gamma < 1/2).$$

In the last case, $d \neq 0$ implies $\sqrt{n}\|\hat{f}_{n,\delta_n} - f_\tau\| \to_p \infty$. No universal higher-order conclusion is asserted when $d = 0$.

Proof. Radius admissibility holds once $n \geq \max\{1, \lceil(\eta/D)^{1/\gamma}\rceil\}$. The remainder in Proposition D.5 is $O_p(r_n)$ with $r_n = n^{-1} + \eta n^{-\gamma-1/2} + \eta^2 n^{-2\gamma}$. For $\gamma \geq 1/2$,

$$\sqrt{n} r_n = n^{-1/2} + \eta n^{-\gamma} + \eta^2 n^{1/2-2\gamma} \longrightarrow 0,$$

since $1/2 - 2\gamma \leq -1/2$. The centered score has the Hilbert limit proved in Lemma D.4. The drift multiplier $\sqrt{n}\delta_n$ tends to zero if $\gamma > 1/2$ and equals η if $\gamma = 1/2$. For $0 < \gamma < 1/2$,

$$n^\gamma r_n = n^{\gamma-1} + \eta n^{-1/2} + \eta^2 n^{-\gamma} \longrightarrow 0, \qquad n^\gamma (P_n - P) g = O_p(n^{\gamma-1/2}) \to_p 0.$$

This proves all three limits. To make the probability conversion explicit, if $U_n = O_p(a_n)$ with $a_n \to 0$, choose the probability multiplier threshold first, and then take n large enough that its product with a_n is below the desired error tolerance. Thus $U_n = o_p(1)$, as used for each scaled remainder. The constants remain functions only of the fixed setting and the fixed choices η, γ.

If $d \neq 0$, the norm of the n^γ-scaled error is at least $\eta\|d\|/2$ with probability tending to one. For any $M > 0$, further require $n^{1/2-\gamma}\eta\|d\|/2 > M$. The probability that the root-n norm exceeds M tends to one. When $d = 0$ this positive threshold is absent, so only the already stated n^γ limit follows. □

Corollary D.7 (A nonzero-drift cosine signal)**.** *Under Assumptions 1, 2, and 3, if $q(x) = \cos(kx)$, $\mu_k > 0$, and $\epsilon \neq 0$, then $d \neq 0$. Consequently the subcritical divergence conclusion of Proposition D.6 applies whenever its shrinking-radius condition holds.*

Proof. The derivative σ' is even. Conditioning on X therefore gives $R_\tau = -\epsilon I^*(q\sigma'(f_\tau))$. The available vector $\cos(k\cdot) \in \mathcal{H}$ and adjoint identity in Lemma C.1 imply

$$\langle R_\tau, \cos(k\cdot)\rangle_{\mathcal{H}} = -\epsilon \int \cos^2(kx)\sigma'(f_\tau(x))\, dm(x) \neq 0.$$

The integral is strictly positive because σ' is strictly positive at every finite argument and $\cos^2(kx)$ is positive on a set of positive measure. Hence $R_\tau \neq 0$, and invertibility of H_τ gives $d \neq 0$. This directly verifies the conditional conclusion, without a weak-signal approximation or a condition on μ_{3k}. □

Balanced labels. At balance, the population drift is zero. The simultaneous empirical derivative bound in Appendix A.2 gives $\|\hat{f}_{n,\delta_n} - \hat{f}_{n,0}\| = O_p(\delta_n n^{-1/2})$ for every deterministic $\delta_n \to 0$. Its root-n limit is therefore the radius-zero Gaussian at every shrinking rate. This follows from the general normalized-kernel proof and requires no further periodic specialization.

E Structural boundaries and local-radius certificates

E.1 Transfer without unit diagonal

Let $\mathcal{H}$ be a real separable RKHS with strongly measurable section map, $\kappa(x) = \|K_x\| = \sqrt{K(x,x)} \leq M < \infty$, and fixed $\tau > 0$. All inequalities concerning inputs may hold P_X-almost surely. Write $\mathcal{J} : \mathcal{H} \to L^2(P_X)$ for evaluation and $T = \mathcal{J}^*\mathcal{J}$. Let M_v denote multiplication by a bounded measurable function v on $L^2(P_X)$. For $\delta \geq 0$ define

$$s_\delta(x) = \sigma(\delta\kappa(x)), \quad p_\delta(x) = \sigma'(\delta\kappa(x)), \quad H_\delta = \tau I + \mathcal{J}^* M_{p_\delta} \mathcal{J}, \quad D_\delta = H_\delta^{-1} \mathcal{J}^* M_{s_\delta}.$$

Observations $Z_i = (X_i, Y_i)$ are iid with $Y_i \in \{-1, 1\}$. The radius is fixed when taking statistical limits. The signal law is $E[Y \mid X] = \epsilon q(X)$ with bounded measurable q and $|\epsilon|\|q\|_\infty < 1$.

Proposition E.1 (General diagonal transfer and local covariance growth)**.** *The exact samplewise robust loss is $\ell(Yf(X) - \delta\kappa(X))$, including inputs with $\kappa(X) = 0$. For each fixed $0 < D < \infty$, its population minimizer satisfies*

$$\sup_{\delta \in [0,D]} (\|f_{\epsilon,\delta} - \epsilon D_\delta q\| + \|\partial_\delta f_{\epsilon,\delta} - \epsilon \partial_\delta (D_\delta q)\|) = O(|\epsilon|^3).$$

At balanced labels the population minimizer is zero and

$$\sqrt{n}\hat{f}_{n,\delta} \Rightarrow N_{\mathcal{H}}(0, C_\delta), \qquad C_\delta = D_\delta D_\delta^*.$$

This covariance is trace class. At zero radius,

$$C_0' = \frac{1}{4} H_0^{-1} E[\kappa(X) K_X \otimes K_X] H_0^{-1}, \qquad H_0 = \tau I + T/4. \tag{E.1}$$

Consequently $\langle h, C_0' h \rangle > 0$ *for every nonzero* $h \in \overline{\operatorname{ran} T}$*, and it is zero on* $\ker T$*. For each such fixed* h *this gives a positive-radius interval on which its limiting variance increases; no common interval or uniform lower bound over all observable directions is asserted.*

Proof. For $\kappa(x) > 0$, Cauchy–Schwarz gives $Yg(x) \geq -\delta\kappa(x)$, attained at $g = -\delta Y K_x/\kappa(x)$. If $\kappa(x) = 0$, all evaluations vanish. Since ℓ is decreasing, this proves the exact reduction. Boundedness gives $\|\mathcal{J}\| \leq M$, T trace class with $\operatorname{tr} T = E\kappa(X)^2 \leq M^2$, and $\|H_\delta^{-1}\| \leq \tau^{-1}$.

The score for any population or empirical measure Q is

$$F_{Q,\delta}(f) = \tau f - Q[Y\sigma(\delta\kappa(X) - Yf(X))K_X].$$

Its Hessian is between τI and $(\tau + M^2/4)I$. The objective is coercive and weakly lower semicontinuous, hence has a unique minimizer. The Hessian bound gives $\|f - f_{Q,\delta}\| \leq \tau^{-1}\|F_{Q,\delta}(f)\|$ by integration on the segment joining f to the minimizer. Every fixed-order derivative of the score exists and is continuous, because bounded logistic derivatives and $|f(x)| \leq M\|f\|$ provide integrable bounds. Radius derivatives introduce only bounded factors κ. The inverse function argument therefore gives smooth minimizing paths.

Conditional averaging at scalar prediction t gives

$$\tfrac{1}{2}\{\sigma(\delta\kappa + t) - \sigma(\delta\kappa - t)\} - \tfrac{\epsilon q}{2}\{\sigma(\delta\kappa + t) + \sigma(\delta\kappa - t)\}.$$

Its linear terms are $p_\delta t - \epsilon q s_\delta$; the remainder is $O(|t|^3 + |\epsilon|t^2)$, with t derivative $O(t^2 + |\epsilon t|)$ and partial radius derivative of the same order as the remainder. Constants are uniform in x and $\delta \in [0, D]$. Set $v_\delta = \epsilon D_\delta q$. Both v and v' are $O(\epsilon)$, by differentiating $H_\delta D_\delta q = \mathcal{J}^*(s_\delta q)$. The score residual $r = F_{\epsilon,\delta}(v)$ and its total radius derivative are $O(\epsilon^3)$. The residual inequality controls $v - f$. With $B(f) = \partial_\delta F(f)$ and $H(f) = D_f F(f)$, differentiation gives

$$H(f)(v' - f') = r' + [H(f) - H(v)]v' + B(f) - B(v).$$

The last two operator differences are Lipschitz in $f - v$, uniformly on this bounded setting; applying the inverse bound proves the claimed derivative remainder. This proves the expansion for the actual minimizer.

At balance, zero solves the population score. Put $W_n = n^{-1}\sum_i Y_i s_\delta(X_i) K_{X_i}$ and $T_{p,n} = n^{-1}\sum_i p_\delta(X_i) K_{X_i} \otimes K_{X_i}$. Then $E\|W_n\|^2 \leq M^2/n$ and $E\|T_{p,n} - ET_{p,n}\|_{\mathrm{HS}}^2 \leq M^4/(16n)$. The residual inequality at zero gives $\hat{f} = O_p(n^{-1/2})$. Taylor expansion of its score gives $0 = -W_n + (\tau I + T_{p,n})\hat{f} + O_p(n^{-1})$, whence $\hat{f} = H_\delta^{-1} W_n + O_p(n^{-1})$. The iid centered bounded Hilbert summands in W_n have covariance $\mathcal{J}^* M_{s_\delta^2} \mathcal{J}$. Finite-dimensional central limits and the trace tail bound yield their Hilbert central limit: the expected squared norm beyond the first m basis coordinates is the corresponding covariance trace tail, uniformly in n, and tends to zero. The limit covariance after multiplication by H_δ^{-1} is $D_\delta D_\delta^*$. Its trace is at most M^2/τ^2. Empirical measurability follows by measurable gradient-descent iteration from zero with step $(\tau + M^2/4)^{-1}$, a contraction to the unique minimizer.

At zero radius $s_0 = 1/2$, $s_0' = \kappa/4$, and $p_0' = 0$. Thus $H_0' = 0$, and differentiating $C = H^{-1}\mathcal{J}^* M_{s^2} \mathcal{J} H^{-1}$ proves (E.1). For $g = H_0^{-1} h$, its quadratic form is $E[\kappa(X) g(X)^2]/4$. Its vanishing is equivalent to $E[g(X)^2] = 0$: where $\kappa = 0$ evaluations are zero, and where $\kappa > 0$ the integrand is strictly weighted. This is equivalent to $g \in \ker T$. Since $H_0 = \tau I + T/4$ preserves the kernel and its orthogonal complement, strictness follows on the observable subspace. Operator-norm continuity of C' proves persistence for each fixed direction. □

Proposition E.2 (Nonunit diagonal can destroy global covariance order)**.** *Consider three equally likely inputs with feature vectors* $\psi_1 = (-5, 6)$*,* $\psi_2 = (-1, 4)$*, and* $\psi_3 = (-6, 0)$ *in* $\mathbb{R}^2$*, and kernel* $K(i, j) = \langle \psi_i, \psi_j \rangle$*. At balance and* $\tau = 1/250$*, the limiting covariance of the minimizing parameter has, at radius* $\delta = 2$*, a negative derivative in direction* $v = (-5, -3)$*. Thus the full-radius Loewner monotonicity for normalized kernels does not extend to all bounded nonunit kernels.*

Computer-assisted certificate. The feature span is $\mathbb{R}^2$, so its Euclidean norm is the RKHS norm. Put $k_i = \|\psi_i\| \in \{\sqrt{61}, \sqrt{17}, 6\}$, $s_i = \sigma(2k_i)$, $p_i = s_i(1 - s_i)$ and evaluate

$$H = \tfrac{1}{250} I + \tfrac{1}{3} \sum_i p_i \psi_i \psi_i^\top, \qquad H' = \tfrac{1}{3} \sum_i k_i p_i (1 - 2s_i) \psi_i \psi_i^\top,$$
$$B = \tfrac{1}{3} \sum_i s_i^2 \psi_i \psi_i^\top, \qquad B' = \tfrac{2}{3} \sum_i k_i s_i p_i \psi_i \psi_i^\top.$$

With $R = H^{-1}$, direct differentiation gives $C' = RB'R - RH'RBR - RBRH'R$. Outward interval arithmetic gives

$$C'_{11} \in [590023.31, 590023.32],$$
$$C'_{12} = C'_{21} \in [-900652.29, -900652.28],$$
$$C'_{22} \in [1334497.18, 1334497.19].$$

Consequently $v^\top C' v = 25C'_{11} + 30C'_{12} + 9C'_{22} < -258510$. The interval certificate evaluates only square roots, exponentials and rational matrix operations using 60-digit interval arithmetic and records all outward enclosures. Its tighter enclosure is $[-258511.051867, -258511.051866]$. Continuity gives an interval of decreasing variance in this fixed direction, so two ordered radii in that interval violate covariance order. This gives a specific failure of limiting-covariance order for a bounded nonunit kernel. It also embeds in an infinite-rank bounded kernel. Give the three displayed inputs probability $1/6$ each, give countably many orthogonal unit feature vectors positive probabilities summing to $1/2$, and choose $\tau = 1/500$. On the original two-dimensional block the Hessian is half the displayed H, the score covariance is half the displayed B, and therefore the limiting covariance and its derivative are twice the displayed values. The remaining block is orthogonal and cannot change the negative quadratic derivative in v. This embedding has bounded diagonal and infinite rank. □

E.2 A SIGNAL-INDEPENDENT HARMONIC RADIUS CERTIFICATE

Retain the periodic model and the cubic coefficients in Theorem 6. Write $b(\delta) = b_{3,\delta}$, $b_+ = b(0) > 0$, and $d_+ = -b'(0) > 0$. Choose any fixed finite $D > 0$ and verified finite upper bounds

$$L_0 \geq \sup_{[0,D]} |b'(\delta)|, \qquad L_1 \geq \sup_{[0,D]} |b''(\delta)|.$$

The coefficient formula is a smooth scalar expression whose denominators are at least τ; therefore such bounds can be computed by interval evaluation of the differentiated formula. Define

$$r_* = \min\{D, b_+/(2L_0), d_+/(2L_1)\} > 0, \tag{E.2}$$

where a quotient by a zero L_i is $+\infty$.

Proposition E.3 (Uniform local attenuation interval)**.** *Suppose the uniform cubic remainder bound of Theorem 6 holds with constant C for $|\epsilon| \leq \epsilon_0$ on $[0, D]$. For every nonzero ϵ with $|\epsilon| \leq \epsilon_0$ and*

$$\epsilon^2 \sqrt{2\mu_{3k}} C \leq \tfrac{1}{4} \min\{b_+, d_+\}, \tag{E.3}$$

the actual population minimizer satisfies $\partial_\delta E_{3k}(f_{\epsilon,\delta}) < 0$ for every $\delta \in [0, r_*]$. In particular r_* is independent of signal strength throughout a sufficiently small nonzero signal range.

Proof. The derivative bounds and fundamental theorem of calculus give $b(\delta) \geq b_+/2$ and $b'(\delta) \leq -d_+/2$ on $[0, r_*]$. The cosine coefficient functional has RKHS norm $\sqrt{2\mu_{3k}}$. Writing the actual coefficient as $c_{3k}(f) = \epsilon^3 b(\delta) + e(\delta)$, the uniform expansion gives $|e|, |e'| \leq \sqrt{2\mu_{3k}} C |\epsilon|^5$. After division by the nonzero ϵ^3, the errors in both coefficient and derivative are at most $\sqrt{2\mu_{3k}} C \epsilon^2$. Condition (E.3) therefore gives $c_{3k}(f)/\epsilon^3 \geq b_+/4 > 0$ and $\partial_\delta c_{3k}(f)/\epsilon^3 \leq -d_+/4 < 0$. Their product is negative even when $\epsilon < 0$. Since $\partial_\delta E_{3k} = c_{3k} \partial_\delta c_{3k}/\mu_{3k}$, the asserted strict attenuation follows. □

The radius in (E.2) is checkable from the cubic coefficient and derivative bounds. A numerical signal threshold in (E.3) additionally requires a certified value of the full RKHS remainder constant C and admissible ϵ_0. Without those values the statement certifies a radius for all sufficiently weak

signals, not a numerical finite-signal guarantee. These constants may depend on the entire fixed kernel and must not be inferred from only two Fourier eigenvalues. A dense radius grid alone does not certify either derivative supremum. For the kernel used in Section 7, with $\tau = 0.02$, $k = 1$ and $D = 0.1$, an interval-automatic-differentiation certificate for the cubic coefficient over the entire interval gives $L_0 < 0.011407$, $L_1 < 0.064406$ and $r_* > 0.0116089$; hence 0.0116 is a certified sufficient radius. The saved interval enclosures also cover ridge values $0.005, 0.08$ and frequency $k = 2$. These certificates retain the signal-strength qualification in (E.3).

F KERNEL EXPERIMENTS: PROTOCOLS AND COMPLETE DIAGNOSTICS

F.1 KERNEL, SOLVER, AND NUMERICAL ACCURACY

For $t = (x - x') \bmod 2\pi$, the kernel is evaluated as

$$K(x, x') = \frac{1 + 2\{\pi^4/90 - \pi^2 t^2/12 + \pi t^3/12 - t^4/48\}}{1 + \pi^4/45}.$$

Its coefficients are $\mu_0 = (1 + \pi^4/45)^{-1}$ and $\mu_j = j^{-4}(1 + \pi^4/45)^{-1}$. An independent truncated-series check gave maximum discrepancy 1.4×10^{-14}; direct checks of RKHS norms and Fourier coordinates differed by less than 4.2×10^{-14}.

For the Gram decomposition $K_n = U\Lambda U^\top$, retained whitened features are $\Phi = U\Lambda^{1/2}$. The optimized variable w gives sample predictions Φw, coefficients $\alpha = U\Lambda^{-1/2}w$, and squared RKHS norm $\|w\|_2^2$. The empirical objective is

$$\frac{1}{n}\sum_i \log(1 + e^{\delta - y_i(\Phi w)_i}) + \frac{\tau}{2}\|w\|_2^2.$$

Newton steps use its exact gradient and Hessian, Armijo backtracking with coefficient 10^{-4}, at most 80 iterations, and a retained-coordinate gradient tolerance of 10^{-11}. The largest recorded projected gradient residual was 9.90×10^{-12}. Robust fits may start from the ordinary solution, but optimize their own objective to the same tolerance.

Eigenvalues at most $\max(10^{-13}\lambda_{\max}(K_n), 10^{-13})$ are removed for numerical stability. This approximates the representer problem rather than redefining the population kernel. Since $\mathrm{tr}(K_n) = n$ and logistic scores have absolute value at most one, the omitted gradient has norm at most $\sqrt{10^{-13}} = 3.17 \times 10^{-7}$ in exact positive-semidefinite arithmetic. Adding this to the projected residual and dividing by τ gives a conservative strong-convexity distance bound; the argument does not account for arbitrary floating-point roundoff. Up to 65 numerically small directions were removed; the full-space bound includes their omitted gradient.

The empirical first-order prediction is $\hat{f}_{n,0} - \delta\hat{d}_n$, where $\hat{d}_n$ solves the empirical regularized Hessian equation for the radius derivative at $\hat{f}_{n,0}$. Relative error is the RKHS norm of prediction minus robust fit divided by the paired displacement norm. A finite-difference check confirmed the sign and first-order convergence of this derivative.

F.2 CONFIGURATIONS AND SAMPLING UNCERTAINTY

Population quadrature uses grids of 256, 512 and 1,024 equally spaced points, $\tau \in \{0.005, 0.02, 0.08\}$, signal frequencies $k \in \{1, 3\}$, $\epsilon \in \{0.025, 0.05, 0.1, 0.2, 0.4\}$, and $\delta \in \{0, 0.005, 0.02, 0.1, 0.25, 0.5, 1\}$. At each point both labels enter with their conditional probabilities. These are 630 deterministic population-quadrature fits. The comparison between 512- and 1,024-point Fourier coefficients covers every common configuration.

The shrinking-radius experiment uses 100 independent datasets at each of five sample sizes, six radius schedules per dataset, and 3,000 robust-fit records. NumPy RandomState seeds are $410000 + 1000n + r$, with replicate $r \in \{0, \ldots, 99\}$. Inputs are independent uniform circle points and labels are Bernoulli draws with probability $(1 + 0.5\cos x)/2$. The balanced experiment uses seeds $870000 + 1000n + r$, $r \in \{0, \ldots, 199\}$, at three sample sizes. Four fixed radii $\{0, 0.25, 0.5, 1\}$ and $n^{-1/4}$ give 3,000 records. Each comparison retains identical data within a replicate.

The raw files retain sample size, seed index, radius parameters, relevant distances or coordinates, and projected residuals. Analysis asserts the full record counts and paired replicate identities. The code uses Python with NumPy 1.21.6, SciPy 1.1.0 and Matplotlib 2.2.3; the runner defaults the OpenMP and MKL thread-count environment variables to one. The reference execution took approximately 500 seconds for shrinking radii, 782 seconds for balanced labels, and 133 seconds for population quadrature, with the families overlapping in wall-clock time.

Finite-range slopes and covariance uncertainty. The slope is an ordinary least-squares fit of log mean displacement against log sample size. Its percentile interval uses 1,000 bootstrap resamples, resampling independent datasets separately at each sample size. The predicted slope is a limit statement; the bootstrap interval quantifies sampling uncertainty of the observed finite-range statistic, not asymptotic approximation error.

γ	Predicted slope	Measured slope	Bootstrap 95% interval
0.25	+0.25	+0.187	[+0.175, +0.199]
0.50	+0.00	-0.063	[-0.075, -0.049]
0.75	-0.25	-0.312	[-0.326, -0.298]

Table F.1: Finite-range slopes for $\eta = 0.3$. All intervals exclude the limiting slopes, so the experiment is not evidence of exact exponent agreement on this range.

Variance ratios are calculated from the root-n RKHS coordinates, using the sample variance with denominator $R - 1$. For each radius and mode, 2,000 bootstrap draws resample the same replicate indices for numerator and denominator. This pairing explains their narrow intervals. The bootstrap seed is 20260912. It does not remove finite-n covariance bias. Prediction-coordinate ratios are the same because their additional factor μ_j cancels within a mode; the unnormalized covariances remain different.

F.3 Local drift and balanced sampling diagnostics

For each $n \in \{64, 128, 256, 512, 1024\}$, we draw 100 independent datasets with conditional label mean $0.5 \cos x$. Each dataset is shared between its ordinary and robust fits. We evaluate $\delta_n = \eta n^{-\gamma}$ for $\eta \in \{0.3, 1\}$ and $\gamma \in \{0.25, 0.5, 0.75\}$. Figure F.1(a,b) shows $\eta = 0.3$.

The one-step prediction uses the empirical Hessian and empirical radius derivative evaluated at the independently fitted zero-radius optimum. At $n = 1024$, its mean relative error is 1.69%, 0.30% and 0.053% for increasing γ. This measures the accuracy of the empirical path linearization against the independently optimized estimator.

Finite-range slopes require a separate interpretation. The measured slopes of the mean root-n displacement are 0.187, -0.063 and -0.312, versus the nonzero-drift asymptotic reference slopes 0.25, 0 and -0.25. The cosine calculation above verifies nonzero drift for the experimental signal. Bootstrap intervals in Appendix F exclude the respective limiting values. The common offset of about -0.06 indicates finite-sample variation in the displacement magnitude. These runs support increasing, nearly stable, and decreasing displacement trends, but do not establish exact asymptotic exponents from the observed sample-size range.

Balanced-label covariance. We draw 200 independent balanced-label datasets at each of $n \in \{128, 512, 1024\}$. For $\delta_n = n^{-1/4}$, the mean root-n estimator norm changes from 7.79 to 7.19, while the paired root-n difference falls from 1.05 to 0.61 (Figure F.2(b)). This illustrates the zero-drift specialization instead of imposing subcritical divergence.

For fixed radii, Figure F.1(c) compares coordinate variance ratios with (7). At $n = 1024$ and $\delta = 1$, the constant-mode ratio is 3.076 versus the limiting 3.106, and the second-cosine ratio is 2.321 versus 2.331. Across the three recorded modes and three positive radii, relative discrepancies are approximately 0.064%–1.20%. Paired bootstrap intervals are narrow and exclude the corresponding limits, so finite-sample bias remains detectable. Both empirical and limiting variances increase despite the exactly zero population function.

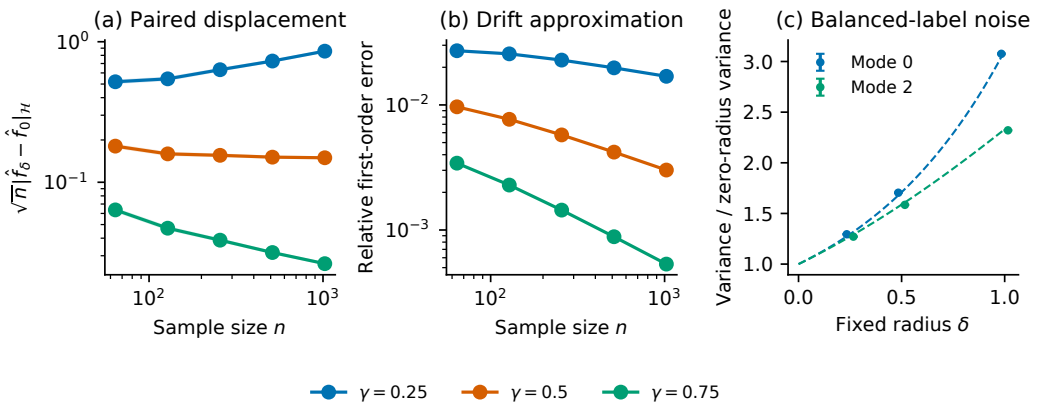

Figure F.1: Sampling response. (a) Paired root-n displacement, mean and one standard error over 100 independent datasets, $\eta = 0.3$. (b) Mean relative error of the empirical first-order path prediction. (c) Balanced-label variance ratios at $n = 1024$, using 200 paired independent datasets: points and bootstrap 95% intervals are empirical, dashed curves are the limiting covariance ratios. Two representative modes are shown; the constant and first cosine share the same kernel eigenvalue.

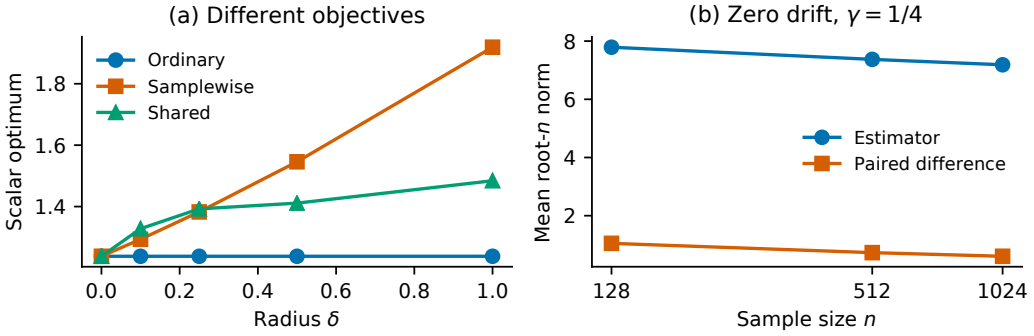

Figure F.2: Objective and zero-drift comparisons. (a) Independently optimized scalar objectives at the same ridge and radius; the ordinary optimum is constant in radius. (b) Balanced-label root-n norms for $\delta_n = n^{-1/4}$, averaged over 200 datasets, distinguish the estimator norm from its paired difference.

F.4 Separating shared and samplewise objectives

For the constant kernel, $P(Y = 1) = 0.8$ and $\tau = 0.02$, we independently minimize three scalar objectives: ordinary ridge logistic risk, samplewise margin-offset risk, and the maximum of ordinary logistic risks at $f - \delta$ and $f + \delta$ with ridge outside. Convexity makes the endpoints exact for the common scalar perturbation. Bounded scalar minimization on $[-10, 10]$ uses tolerance 10^{-12}. The full sweep also includes probabilities 0.5 and 0.65. Figure F.2 displays the scalar comparison and the balanced-label shrinking-radius norms. The latter use the same 200-replicate datasets specified above. Figure F.2(a) shows distinct radius-dependent optimum paths. The shared maximum is computed exactly by comparing its two endpoints. This comparison isolates the effect of moving the supremum across the sample average. The affine Functional SAM identity in Proposition C.6 concerns an update direction; it does not equate either robust objective with ordinary risk.

F.5 Matched periodic fits and resolution checks

We use the closed periodic kernel with $\tau = .02$ and $q(x) = \cos x$, and optimize both conditional labels exactly in expectation on uniform grids of 512 and 1,024 points, with $\epsilon \in \{.025, .05, .1, .2\}$ and $\delta \in \{0, .005, .02, .1, .25, .5, 1\}$. This gives 112 individual objective fits. The matched objective

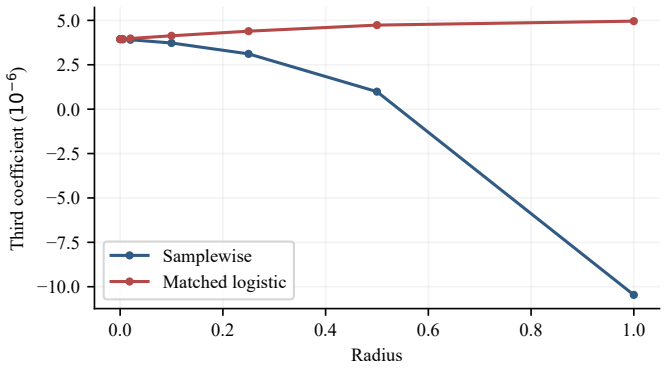

Figure F.3: Full third-coefficient paths at $\epsilon = .1$, $\tau = .02$. The local samplewise attenuation eventually reverses; at radius one the coefficient is negative with larger magnitude than at zero.

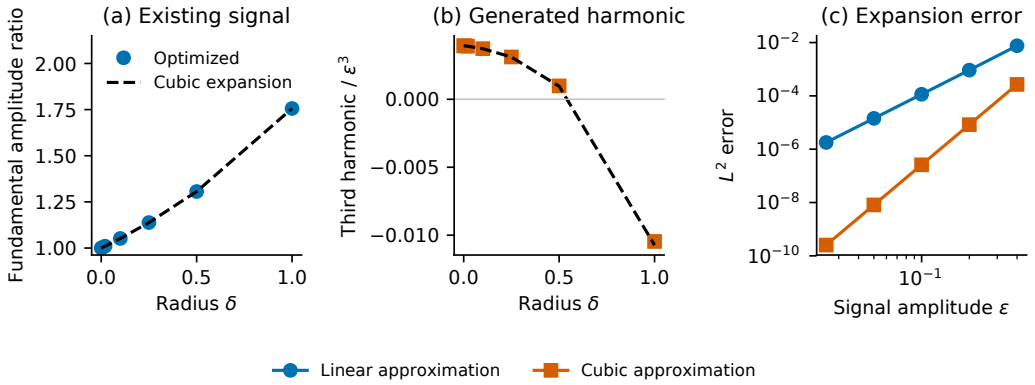

Figure F.4: Independent samplewise fits and cubic approximation. At radius .25, linear and cubic approximation-error slopes over $\epsilon \in \{.1, .2, .4\}$ are 3.024 and 5.008. The matched-control comparison in Figure 2 asks a separate question about the difference between two optimized learning rules.

is ordinary logistic at λ_δ, followed by output scaling γ_δ; no optimized function is synthesized from its expansion. Float64 full Gram coordinates have positive eigenvalues on both grids. Accelerated gradient minimization uses a global curvature bound and a residual tolerance of 10^{-13}. The largest final gradient norm is 1.93×10^{-15}. Four independent Newton comparisons at grid 256 use $\epsilon = .1$, $\delta = .005, .25$ and both objectives; maximum prediction discrepancy is 9.83×10^{-16}. Grid doubling changes samplewise third coefficients by at most 1.80×10^{-13} across the prescribed configurations. These checks measure numerical agreement, not interval-certified total error.

The absolute prediction difference is the uniform-grid L^2 norm; the relative difference divides it by the samplewise prediction norm. At fixed radius .25, their slopes over the four declared amplitudes are 3.0006 and 1.9989. The bounds in Appendix B explain the respective orders but do not assert a nonzero cubic coefficient for every signal. At $\epsilon = .1$, the complete-function relative differences at radii .005, .02, .25 are .000562%, .002286%, .036715%. The local sign is measurable despite the small whole-function difference.

F.6 Nonuniform inputs and nonperiodic kernels

Nonuniform circular inputs and mixed signals. We use the same infinite-rank periodic kernel as Appendix F, but integrate against density $(1 + a\cos x)/(2\pi)$. The full grid crosses $a \in \{0, 0.3, 0.6\}$, $b \in \{0, 0.25, 0.5\}$ in the label mean $0.15(\cos x + b\cos 2x)$, $\tau \in \{0.02, 0.08\}$, radii $0, 0.005, 0.02, 0.1, 0.25, 0.5, 1$ and quadrature sizes 256 and 512, giving 252 fits. Both label

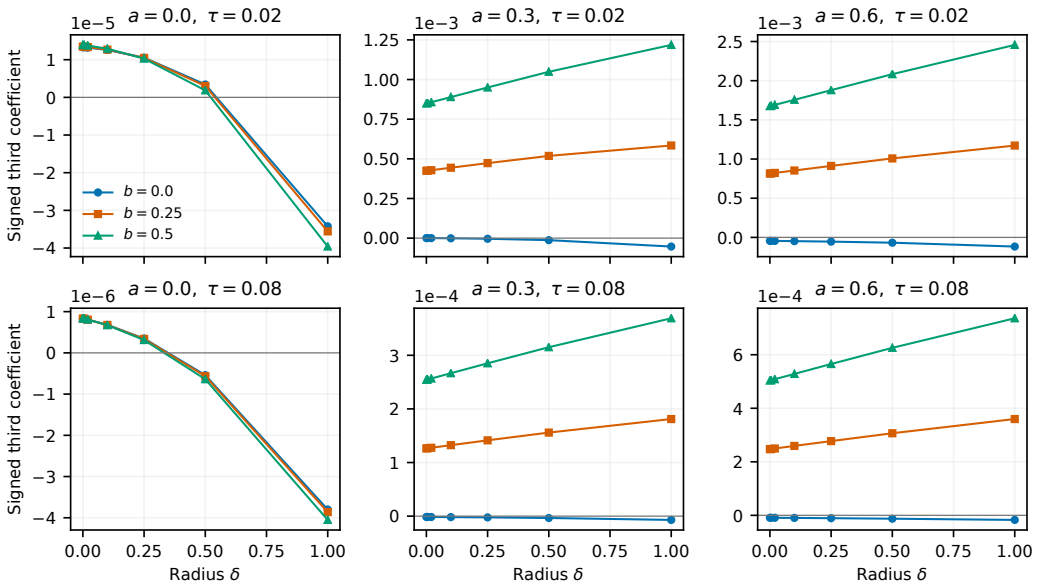

Figure F.5: All signed third-frequency response paths across the density/signal/ridge grid. Each panel fixes (a, τ) and includes the three signal mixtures. A label-absent frequency can arise through linear mixing under nonuniform input density, so these coefficients need not be purely nonlinear harmonics.

outcomes are integrated with their exact conditional probabilities. The input density and label probabilities are valid throughout this grid. All objectives are independently optimized.

Reported Fourier coefficients use uniform Lebesgue measure. We store the constant and cosine coefficients through frequency six for both the optimizer and its leading linear response. For nonuniform inputs, the operator T is not diagonal in this Fourier basis, so frequencies absent from the label function can enter at first order. This separates linear mixing from nonlinear harmonic generation. The maximum absolute difference between corresponding Fourier coefficients on the two grids is 3.089×10^{-10}. The comparison measures quadrature-resolution sensitivity.

At radius 0.02, seven of the eighteen density/signal/ridge settings have a smaller absolute third coefficient and eleven have a larger one than at radius zero. The same partition holds at 0.005; all six uniform-input settings attenuate. At radius one, seven settings have a third coefficient of opposite sign to their zero-radius value. Figure F.5 shows all paths. The particularly small ratio 0.347 at $a = 0.3, b = 0, \tau = 0.02$ reflects near cancellation: the zero-radius coefficient is 3.63425×10^{-7} and the radius-0.02 coefficient is 1.26138×10^{-7}. Its relative size should be read alongside these small absolute values.

Nonperiodic RBF response and sampling covariance. The input space is $[-1, 1]^2$ with density $(1 + 0.6x_1)(1 + 0.4x_2)/4$, and the kernel is $\exp(-\|x - z\|^2/(2 \cdot 0.7^2))$. Balanced labels are independent Rademacher variables. At each $n \in \{128, 256, 512\}$ we generate 100 independent datasets with seeds $1930000 + 1000n + r$, $r = 0, \ldots, 99$. Each is fitted at both ridge values and all radii $0, 0.5, 1$, giving 1,800 fits. Radius and ridge comparisons share data within a replicate.

We project predictions onto $h \in \{1, x_1, x_2\}$ using $\int f(x)h(x)dP_X(x)$ and compute centered sample covariances of their root-n scaled values. These fixed probes are not assumed orthogonal. No empirical eigenvectors are aligned between datasets. The population reference uses the weighted kernel operator on independent Gauss–Legendre product quadrature with 32 nodes per coordinate. If (μ_j, ϕ_j) are its numerical eigenpairs in $L^2(P_X)$, the covariance between probes h_a, h_b is approximated by

$$\sum_j A_\delta(\mu_j)^2 \langle h_a, \phi_j \rangle \langle h_b, \phi_j \rangle.$$

Comparing quadrature orders 24 and 32 yields relative covariance differences below 7.5×10^{-16} in this smooth configuration. This numerical agreement is not an exact-spectrum claim. Both reference orders, all radii and both ridge values are retained, giving 12 reference records.

Each variance-ratio interval uses 2,000 bootstrap resamples of the 100 paired replicate identities within a sample-size/ridge setting. Intervals are pointwise percentile intervals, not simultaneous coverage statements. All 36 positive-radius/probe ratios have intervals above one, while 33 exclude their limiting references. The latter indicate finite-sample discrepancies that remain detectable beyond Monte Carlo uncertainty; they do not assert a separate population-bias theorem. Selected ratios appear in Table F.2; all settings are included in the supplementary data.

Ridge	Probe	Ratio	Pointwise 95% interval	Reference
0.02	1	3.1117	[3.0963, 3.1275]	3.1832
0.02	x_1	2.9383	[2.9148, 2.9639]	2.9910
0.02	x_2	2.9066	[2.8861, 2.9250]	2.9716
0.08	1	2.7720	[2.7649, 2.7791]	2.7923
0.08	x_1	2.6142	[2.5867, 2.6440]	2.6205
0.08	x_2	2.5586	[2.5340, 2.5789]	2.5718

Table F.2: Nonperiodic RBF variance ratios at $n = 512$, radius one, relative to radius zero. Reference values approximate the general-kernel limiting covariance.

For weak signals, $q(x) = (x_1 + 0.5x_2)/1.5$ and $\epsilon \in \{0.05, 0.1, 0.2\}$. Crossing both quadrature orders, all radii and both ridge values gives 36 population-quadrature fits. The largest projection difference between orders is 1.95×10^{-16}. Log–log regressions of $L^2(P_X)$ error against the three signal amplitudes have slopes 2.980–3.026 over the six ridge/radius configurations. These finite-grid slopes test the leading function approximation, without estimating the uniform RKHS radius-derivative remainder.

General cubic separation on the RBF geometry. To test the new matched-control operator away from the periodic setting, we use the same nonuniform RBF population with a 24×24 Gauss–Legendre grid (576 quadrature points), $\tau = 0.02$, and the bounded signal $q(x) = (x_1 + 0.5x_2)/1.5$. For $\delta \in \{0.5, 1\}$ and $\epsilon \in \{0.08, 0.12, 0.16, 0.24\}$, we independently optimize the samplewise and matched population objectives and compare $(f_{\epsilon,\delta} - g_{\epsilon,\delta})/\epsilon^3$ with the general prediction $\Delta_\delta[q]$. The observed normalized $L^2(P_X)$ difference has log–log slopes -0.00569 and -0.02331 for $\delta = 0.5$ and 1, respectively, and the predicted plateau norms are 0.0342376 and 0.152709. Thus the finite-grid curves are consistent with a nonzero third-order difference rather than a vanishing residual. After subtracting $\Delta_\delta[q]$, the centered-error slopes are 1.983 and 1.964. The largest absolute errors of the probe $\langle q, \cdot \rangle_{L^2(P_X)}$ are 9.72×10^{-5} and 1.74×10^{-3}, respectively; these are finite-signal errors, not solver tolerances. Repeating the population calculation on a 32×32 grid changes each predicted norm by less than 2×10^{-15}. This is a deterministic check of the general cubic formula, not a claim of a universal sign or a risk ordering.

Optimization and independent checks. The experiments use whitened Gram coordinates and exact gradients/Hessians of the weighted or empirical shifted logistic objective, with external ridge. For the weighted Gram eigendecomposition $W^{1/2}KW^{1/2} = U\Lambda U^\top$, predictions use the node-value map $W^{-1/2}U\Lambda^{1/2}$; this reconstructs the normalized kernel at the quadrature nodes. In the RBF cubic check, the maximum node-diagonal reconstruction error is 2.1×10^{-10} and $\mathrm{tr}(\Psi^\top W \Psi) = 0.999999999999$. Newton iteration uses Armijo coefficient 10^{-4}, gradient tolerance 10^{-10} and at most 80 iterations. For the covariance, weak-signal, and generalization experiments summarized above, all 2,088 fits met this retained-coordinate tolerance within five iterations; the maximum recorded residual is 9.934×10^{-11}. Eight paired parameter settings, comprising sixteen independently optimized objectives on each quadrature grid, are used for the general cubic check. Their maximum retained gradient residual is 1.6×10^{-14}. Eigenvalues at most $c = \max(10^{-13}\lambda_{\max}, 10^{-13})$ are discarded. For integration weights w_i, the omitted gradient is bounded by $\sqrt{c}\|w\|_2$ in exact positive-semidefinite arithmetic, since each conditional or observed logistic score has magnitude at most one. This term is zero if no mode is discarded. The maximal omitted bound is 2.306×10^{-7}; adding the

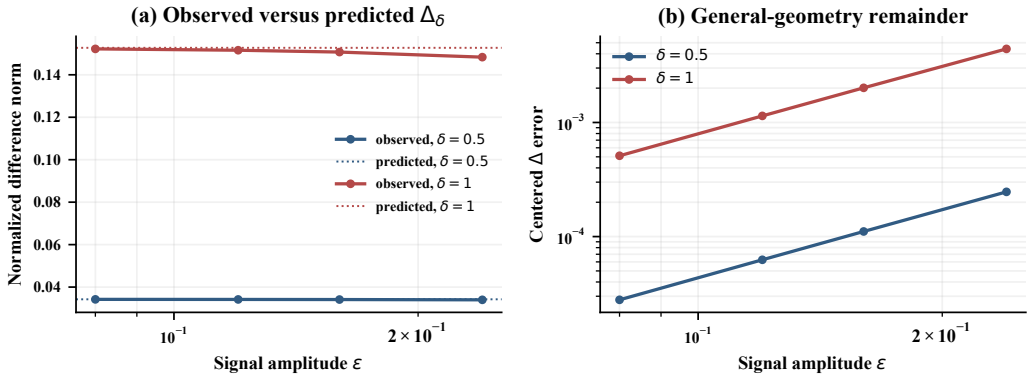

Figure F.6: General-geometry check of the matched cubic operator on the nonuniform RBF population. (a) The normalized samplewise–matched difference approaches the predicted $\|\Delta_\delta[q]\|_{L^2(P_X)}$ (dotted lines). (b) After centering by the predicted operator response, the error follows the expected $O(\epsilon^2)$ order for both radii.

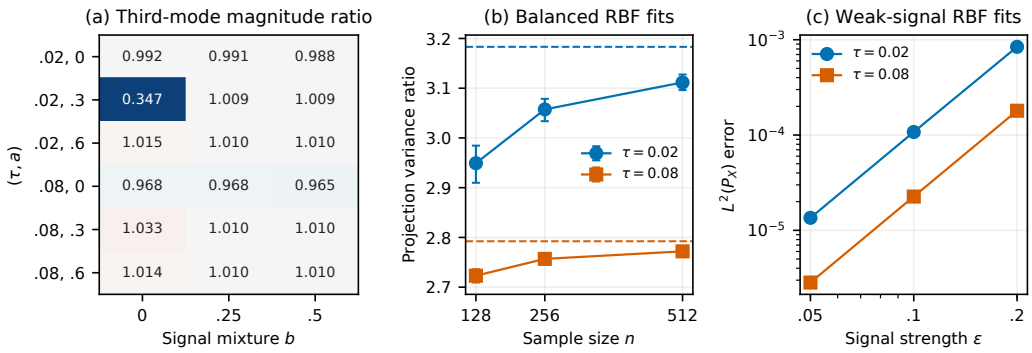

Figure F.7: General transfer and structure-dependent signs. (a) Absolute third-coefficient ratios between radii 0.02 and zero across density strength a, signal mixture b and ridge τ. (b) Balanced RBF constant-projection variance ratios at radius one, relative to zero: pointwise 95% paired-bootstrap intervals use 100 independent datasets per sample size; dashed lines are population-quadrature references. (c) Weak-signal approximation errors at radius one in the RBF model. Panels (a,c) approximate population integrals and have no sampling error bars.

retained residual and dividing by ridge gives a maximal conservative distance bound of 1.153×10^{-5}. It does not account for arbitrary floating-point roundoff.

Independent directional finite-difference checks for empirical and weighted population objectives have errors below 4.7×10^{-12}. Off-grid predictions evaluated through kernel coefficients and through feature coordinates agree within 1.13×10^{-12}. A 200,000-input check of the inverse-CDF sampler gives coordinate-mean standardized errors 1.114 and -0.871 relative to the analytic means. Numerical records, environment metadata, full configuration, and verification results accompany these experiments. The RBF tests concern the general normalized-kernel result; the circular tests probe the boundary of the harmonic specialization. The feature maps are fixed during head fitting.

F.7 Paired nonlinear and random-design covariance errors

The input law and Gaussian kernel match the nonperiodic experiment in Appendix F.6. We use bandwidth $b = 0.7$, $n \in \{128, 256, 512, 1024, 2048\}$, $\tau \in \{0.02, 0.08\}$, $\delta \in \{0, 0.5, 1\}$, and 100 independent input/label replicates at each sample size. Conditions within a replicate share the data, with seed $4930000 + 1000n +$ rep for rep $= 0, \ldots, 99$. This gives 3,000 separately optimized fits. A bandwidth check adds 400 fits at $b \in \{0.4, 1.2\}$, $n \in \{256, 1024\}$, $\tau = 0.02$, and $\delta = 1$. The grids

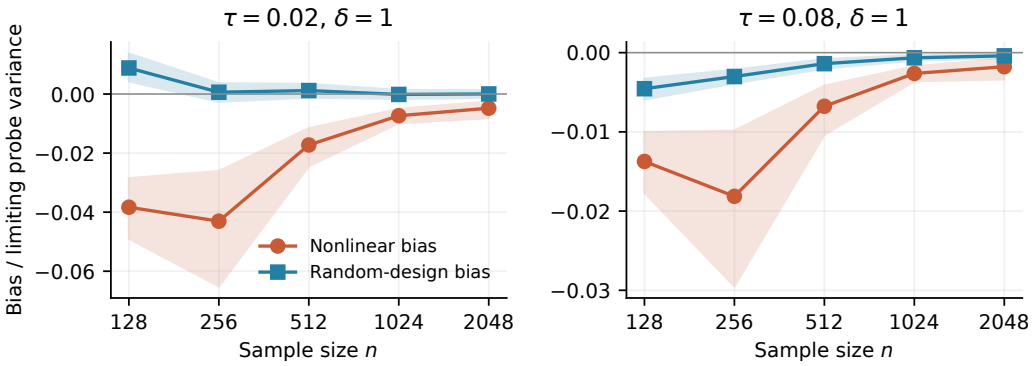

Figure F.8: Finite-sample covariance decomposition for the constant population probe in the nonuniform RBF model, bandwidth 0.7 and radius one. Lines estimate nonlinear and random-design biases divided by limiting variance; shading gives pointwise 95% paired-repeat bootstrap intervals over 100 independent replicates. Analytic conditional covariance removes label Monte Carlo fluctuation from the design estimate. Bias signs are empirical findings.

were recorded before result analysis. Independent seeds separate this study from the covariance-ratio experiment.

The Newton–Armijo representer solver uses the exact Gram kernel. Its relative eigenvalue threshold is 10^{-13}. Across these fits, the largest retained gradient norm is 9.964×10^{-11}; the conservative full-gradient bound is at most 2.717×10^{-7} and the corresponding solution-distance bound at most 1.359×10^{-5}. A separate full Gram solve, without eigenvalue filtering, reproduces the conditional probe covariance at $n = 128$ within 2.50×10^{-14}. No finite-feature approximation is introduced into the Gaussian kernel as n increases.

For each fixed $h \in \{1, x_1, x_2\} \subset L^2(P_X)$, let $U = \sqrt{n}\langle h, \mathcal{J}\hat{f}_{n,\delta}\rangle$ and $V = \sqrt{n}\langle h, \mathcal{J}\widetilde{f}_{n,\delta}\rangle$. Product Gauss–Legendre quadrature of order 32 in each coordinate evaluates these probes. Population covariance references are calculated independently, with orders 24 and 32 compared for each bandwidth. Let $c_n = E[V^2 \mid X_{1:n}]$ and $c = \langle \mathcal{J}^* h, C_\delta \mathcal{J}^* h\rangle$. Balanced-label sign symmetry gives exact conditional zero means, so uncentered second moments estimate covariance. For replicate averages,

$$\overline{U^2} - c = \underbrace{\overline{U^2 - V^2}}_{\text{nonlinear bias estimate}} + \underbrace{\overline{V^2} - \overline{c_n}}_{\text{label Monte Carlo fluctuation}} + \underbrace{\overline{c_n} - c}_{\text{random-design bias estimate}} .$$

The identity is checked in every cell. The middle term has zero expectation and is not a model bias. The control-variate estimate $\overline{c_n + U^2 - V^2}$ has the same expectation as $\overline{U^2}$; it reduces label fluctuation while retaining independently optimized U. Ratios of these estimates need not be unbiased. We retain ordinary centered sample variances as additional diagnostics.

Pointwise 95% intervals use 2,000 bootstrap resamples of the independent replicates, with seed 881133. Within each bootstrap resample, all paired radius conditions stay together; numerator and denominator are recomputed jointly for a variance ratio. These are not simultaneous intervals. The plotted finite-dimensional probe errors examine the decomposition, rather than empirically verifying a full infinite-dimensional trace-norm bound.

Table F.3: Constant-probe covariance bias divided by its limiting variance, at bandwidth 0.7 and radius one. Values are paired estimates.

n	τ	Nonlinear bias	Random-design bias
128	0.02	-3.83%	$+0.882\%$
2048	0.02	-0.480%	$+0.00308\%$
128	0.08	-1.37%	-0.455%
2048	0.08	-0.178%	-0.0402%

Estimated bias magnitude is not monotone at every sampled n: the nonlinear estimate at $n = 256$ has greater magnitude than at $n = 128$. We therefore do not fit or assert an empirical universal $1/n$ law. At $n = 2048$, $\tau = 0.02$, and radius one, the exact second-moment ratio to radius zero is 3.15394 with paired interval $[3.13895, 3.16783]$, compared with the limiting 3.18317. The conditional-linearized ratio is 3.18267, and the control-variate exact ratio is 3.16078. The paired decomposition identifies the error sources within this independent seed set.

At bandwidth 0.4, constant-probe nonlinear biases at $n = 256, 1024$ are -0.32080, -0.05640, versus random-design biases -0.03481, -0.00821 and limiting variance 5.80157. At bandwidth 1.2, the corresponding values are -0.32801, -0.05533, $+0.01284$, $+0.00130$, and 10.41332. The nonlinear term dominates in these displayed cases. All three probes and all conditions are retained in the supplied decomposition and ratio tables.

G FROZEN-FEATURE EXPERIMENTS AND PREDICTIVE CONTROLS

The general normalized-kernel results apply to a linear head on fixed unit-norm features. We test their predictions on two pretrained visual representations, keeping the representation, label mechanism and sampling reference explicit. The protocol fixes all radii, probes, amplitudes and repetition counts before inspecting responses. All 1,592 head fits solve the exact samplewise objective; 72 use expected weak labels, 800 use balanced random labels, and 720 use true-label bootstrap samples.

G.1 IMAGE POOLS, FROZEN FEATURES, AND COMPUTATION

We use CIFAR-10's official training and test split[1]. NumPy seed 9142026 selects 300 training and 100 test images per class without replacement, yielding fixed pools of 3,000 and 1,000 images. Preprocessing, probes, and hyperparameters use training or validation data only. Torchvision 0.22.1 supplies `IMAGENET1K_V1` weights for ResNet-18 and ViT-B/16, with the official resize-to-256, center-crop-to-224 and channel normalization. Evaluation-mode inference replaces the final classifier by the identity. The extracted feature dimensions are 512 and 768, respectively. Each row is normalized in float64 before head fitting; we add no intercept and perform no PCA truncation. Pretraining weights are fixed, and model hashes, image indices and raw feature arrays are supplied with the code. This is a frozen-head transfer study on the declared image pools, rather than full-network fine-tuning or a tuned classification benchmark.

G.2 WEAK-SIGNAL AND BALANCED-LABEL RESPONSE CHECKS

The discrete input distribution is uniform over the 3,000 selected training images. For each backbone, the second-moment feature operator determines three probes at descending eigenvalue ranks 1, 8 and 32. For each probe, divide its feature evaluation by its maximum absolute value on the pool to obtain a bounded signal. Conditional label means multiply this signal by amplitudes .05, .1 and .2. The expected objective sums over both labels with their known conditional probabilities. Its independently optimized minimizer is compared to the leading response in Theorem 2. All radii 0, .25, .5 and 1 use ridge .02. Across the 27 positive-radius signal-gain comparisons per backbone, the maximum relative gain errors are 1.954% and 0.627%; maximum full-vector relative errors over all settings are 1.154% and 0.440%. These checks measure the function-value component of the weak-signal expansion.

Conditional label noise. Seed 9142027 selects 2,048 images from the training pool, without replacement. Inputs stay fixed for 100 independent Rademacher label vectors generated with seed 9142100. The same label vectors pair all radii. Probe ranks are computed from this fixed design's normalized feature operator. The reference is its analytic conditional linearized covariance from Section 6, not an estimate of a new input-design expectation. Balanced-label symmetry gives zero conditional mean, so the mean squared projected coefficient, multiplied by sample size, estimates the corresponding covariance quadratic form. Ordinary centered sample variances are also retained in the CSV files.

[1]Dataset source: `https://www.cs.toronto.edu/~kriz/cifar.html`. The downloaded archive matches the official MD5 `c58f30108f718f92721af3b95e74349a`.

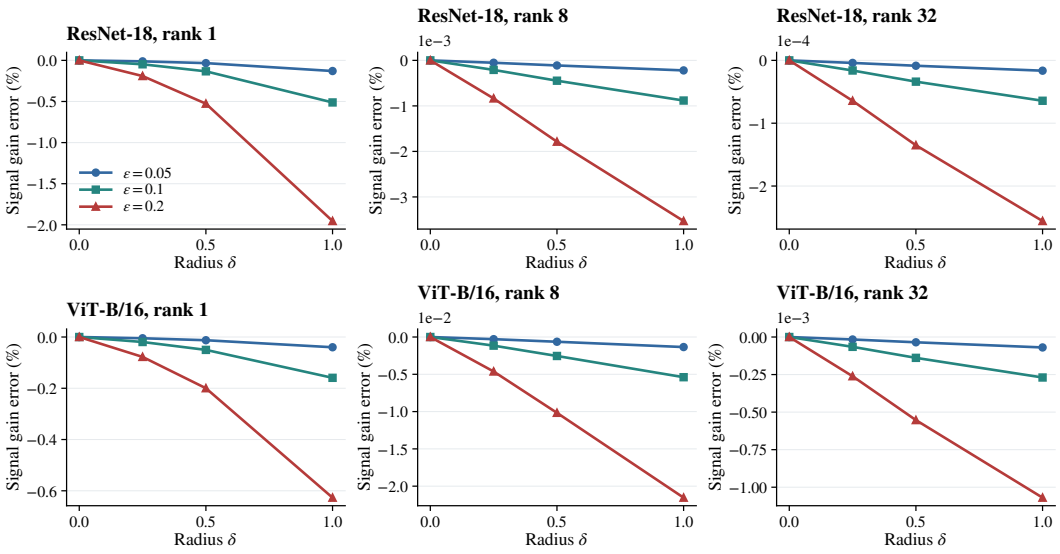

Figure G.1: Complete weak-signal gain errors on frozen features. Each panel fixes a backbone and spectral rank and contains all three signal strengths. Values are $100(\text{observed}/\text{predicted} - 1)$ from independently optimized expected-label objectives. No sampling bands are attached to deterministic population fits.

The nine positive-radius ratios for each backbone are all greater than one, with all 18 pointwise 95% paired-repeat bootstrap intervals above one. Observed ratios range from 1.265 to 3.250. Relative errors against the conditional predictions are at most .571% for ResNet and .898% for ViT. All 18 intervals nevertheless exclude those predictions. Pairing labels strongly reduces uncertainty in a ratio, making small nonlinear deviations detectable. The exact-minus-linearized second-moment difference, normalized by conditional variance, has maximum absolute magnitude .463% and .791% over all declared directions and radii. The realized linearized second moments are saved separately, so label Monte Carlo fluctuation is not identified as nonlinear bias. The 2,000-draw bootstrap uses seed 9142200; intervals are pointwise, and ratios are not unbiased covariance estimators.

G.3 Fixed-setting true-label behavior

The predefined tasks are cat/dog, automobile/truck and airplane/ship. Each uses 600 training-pool images and 200 fixed test images. Thirty bootstrap training samples, drawn with seeds 9142300 plus the task index, are shared across all four radii. Optimization uses the exact bootstrap weights and fixed ridge .02. Table G.1 gives endpoints; the accompanying CSV contains every radius, log-loss and accuracy intervals, and score-variance intervals. Intervals resample the 30 fit replicates using 2,000 draws with seed 9142400 and condition on the fixed training/test pools; they do not express uncertainty across new test datasets. All six endpoint comparisons reduce mean log loss and increase score resampling variance. Accuracy is almost unchanged and does not improve consistently. Fixed-ridge comparison isolates radius response; it does not compare against validation-tuned ERM. True labels are evaluated separately from the balanced-label mechanism, with no claim that their covariance obeys the balanced-label formula.

Computation and independent checks. PyTorch 2.7.1 with CUDA 12.8 runs on an RTX 5060 Laptop GPU (8 GB). Backbones use float32 inference in batches of 32; head optimization uses float64 full-batch accelerated gradient with a valid strongly-convex Lipschitz bound, with no stochastic optimizer noise. Maximum gradient norms are below 10^{-10}. CPU SciPy L-BFGS independently checks weak-signal, balanced-label and true-label fits for both backbones at radius one; parameter differences are below 4.9×10^{-8}, and directional finite-difference gradient errors are below 1.3×10^{-11}. A direct matrix solve also verifies the stored conditional covariance against the eigendecomposition. Feature extraction took approximately 7 seconds for ResNet and 56 seconds for ViT after downloads,

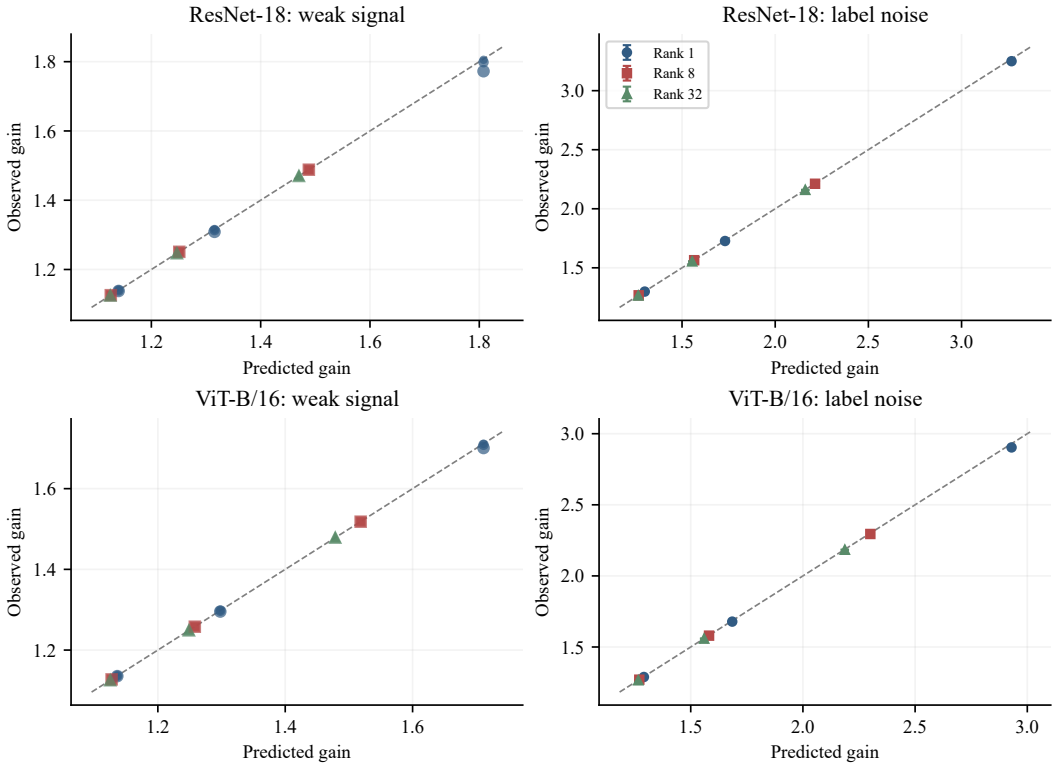

Figure G.2: Backbone-specific views of every declared deep-feature response. Colors and marker shapes denote spectral ranks 1, 8 and 32. Weak-signal panels include all three amplitudes; noise panels use conditional predictions and 100 paired label replicates. Error bars are pointwise 95% bootstrap intervals, mostly smaller than markers. Equality lines are visual references.

Table G.1: Real-label frozen-head results. Each entry compares radii zero and one at fixed ridge .02, using 30 paired bootstrap training samples and the same 200 test images per task. Variance is mean across-test score resampling variance, in units of 10^{-3}. Parentheses give 95% pointwise bootstrap intervals for mean log loss. These intervals condition on the fixed image pools. Accuracy is a percentage.

Backbone / task	Radius	Log loss (95% interval)	Accuracy	Variance
ResNet-18 Cat/dog	0	0.6354 (0.6345, 0.6364)	77.60	7.336
	1	0.6116 (0.6099, 0.6133)	76.95	19.937
ResNet-18 Car/truck	0	0.5843 (0.5834, 0.5851)	90.67	3.945
	1	0.5392 (0.5379, 0.5404)	90.53	9.794
ResNet-18 Airplane/ship	0	0.5742 (0.5732, 0.5752)	87.47	5.407
	1	0.5270 (0.5257, 0.5283)	87.33	13.192
ViT-B/16 Cat/dog	0	0.4633 (0.4621, 0.4645)	90.70	6.022
	1	0.3946 (0.3934, 0.3958)	90.57	12.667
ViT-B/16 Car/truck	0	0.3603 (0.3595, 0.3609)	96.95	3.317
	1	0.2771 (0.2763, 0.2777)	96.98	6.260
ViT-B/16 Airplane/ship	0	0.3804 (0.3794, 0.3814)	96.87	4.278
	1	0.2975 (0.2965, 0.2985)	96.88	8.154

for the 4,000 selected images per backbone. These timings exclude environment installation and model/data transfer.

Table G.2: **Ordinary controls recover the fixed-setting task gains.** Mean test log loss over 30 paired training bootstraps; SW uses $\tau = .02, \delta = 1$. Matched uses analytic ridge/scale; tuned and calibrated controls are validation-selected. The last column is the paired matched-minus-SW difference with a 95% repeat-bootstrap interval, conditional on the fixed test pool and selected settings.

Features	Task	SW	Matched	Tuned	Calibrated	Matched − SW (95% CI)
ResNet-18	Cat/dog	0.6145	0.6141	0.4030	0.4009	$-0.0004\ [-0.0006, -0.0002]$
	Car/truck	0.5417	0.5396	0.2090	0.1800	$-0.0020\ [-0.0023, -0.0017]$
	Plane/ship	0.5293	0.5258	0.2117	0.1826	$-0.0034\ [-0.0036, -0.0032]$
ViT-B/16	Cat/dog	0.3948	0.3780	0.2196	0.2192	$-0.0167\ [-0.0170, -0.0164]$
	Car/truck	0.2774	0.2510	0.0850	0.0797	$-0.0265\ [-0.0268, -0.0261]$
	Plane/ship	0.2998	0.2744	0.0941	0.0790	$-0.0254\ [-0.0256, -0.0252]$

G.4 Matched and validation-selected ordinary controls

Using the same normalized feature arrays, three spectral directions and three amplitudes in Appendix G, we independently fit 72 matched ordinary heads across both backbones and four radii. Relative parameter differences are measured against the saved samplewise optima. Their maxima are 1.93% for ResNet and $.61\%$ for ViT. No Fourier interpretation is assigned to these feature directions. Raw fits retain every direction, amplitude and radius.

Data split and validation protocol. For each class in a binary pair, a seeded permutation of the existing 300 training examples allocates 240 to fitting and 60 to validation. Thus each task has 480 fitting, 120 validation and the fixed 200 independent test examples. Seeds are $9143000 + j$, with pair index $j = 0, 1, 2$ for cat/dog, car/truck and plane/ship. Splits are identical across backbones and methods. Features remain normalized and untruncated; there is no intercept.

Ordinary ridge candidates are $\{.0002, .001, .005, .02, .08, .32\}$. The tuned baseline minimizes validation log loss. For each ridge the calibrated baseline minimizes validation log loss over a positive scalar $c \in [.05, 20]$ by bounded scalar optimization in $\log c$, then selects the ridge with the smallest calibrated validation loss. Ties retain the first candidate. The selected settings below use validation loss only. Fixed ERM uses $\tau = .02$; samplewise uses $\tau = .02, \delta = 1$. The analytic control uses $\lambda_1 = .02543080635$ and $\gamma_1 = 1.859140914$.

Thirty size-480 multinomial training bootstrap samples use seeds $9143100 + j$ and are paired across methods and backbones. Hyperparameters and output scales are held fixed after their initial validation selection. No validation example is added to the fitting pool. Reported variability therefore concerns training resampling with selected parameters held fixed, not the entire hyperparameter-selection procedure. The comparison asks whether ordinary controls reproduce the fixed-setting task gains; it does not rank validation-optimized samplewise methods.

The 900 bootstrap head fits and 36 validation head fits use float64 exact objectives with gradient tolerance 10^{-11}. Independent CPU gradients for the first replicate of all 30 method/task settings are below 2×10^{-10}. CPU L-BFGS checks of one ERM fit per backbone give parameter differences below 6.4×10^{-8}.

Test losses and accuracies average all test examples and 30 repeats. Score and probability variances are centered across repeats with divisor 29, then averaged over the fixed test examples. Probability RMS difference and classification disagreement pair each control with the samplewise fit using the same bootstrap counts. Loss-difference intervals resample the 30 paired repeats 2,000 times with seed 9143200. They condition on the test pool and fixed selected parameters; they do not estimate uncertainty over fresh test populations. All three ordinary controls have negative upper endpoints against samplewise in all six settings. Similar predictive risk is not by itself evidence of identical mechanisms.

Independent split selection and higher-order residuals. The fair-tuning supplement repeats the frozen-feature comparison on five independent stratified splits for each of the three binary tasks and both backbones. Each split uses 240 fitting and 60 validation examples per class; the 200-example

Table G.3: ResNet-18 controls. Variances are multiplied by 10^3; RMS is probability difference against the paired samplewise fit.

Task	Method	Log loss	Brier	Acc.	Score var.	Prob. var.	RMS
3/5	ERM	0.6373	0.2223	0.769	9.412	0.576	0.0266
3/5	Samplewise	0.6145	0.2113	0.761	25.696	1.532	0.0000
3/5	Matched	0.6141	0.2112	0.755	30.302	1.801	0.0035
3/5	Tuned	0.4030	0.1295	0.812	251.815	6.941	0.2595
3/5	Calibrated	0.4009	0.1289	0.812	531.148	10.526	0.3086
1/9	ERM	0.5859	0.1970	0.901	6.637	0.398	0.0350
1/9	Samplewise	0.5417	0.1761	0.897	16.826	0.960	0.0000
1/9	Matched	0.5396	0.1752	0.891	21.879	1.242	0.0048
1/9	Tuned	0.2090	0.0549	0.941	121.261	1.989	0.2957
1/9	Calibrated	0.1800	0.0468	0.941	380.004	3.225	0.3531
0/8	ERM	0.5760	0.1924	0.876	4.704	0.278	0.0384
0/8	Samplewise	0.5293	0.1707	0.875	11.394	0.632	0.0000
0/8	Matched	0.5258	0.1692	0.872	14.809	0.813	0.0050
0/8	Tuned	0.2117	0.0607	0.916	149.315	2.356	0.2824
0/8	Calibrated	0.1826	0.0562	0.916	410.128	3.617	0.3326

Table G.4: ViT-B/16 controls. Variances are multiplied by 10^3; RMS is probability difference against the paired samplewise fit.

Task	Method	Log loss	Brier	Acc.	Score var.	Prob. var.	RMS
3/5	ERM	0.4633	0.1409	0.905	7.626	0.389	0.0599
3/5	Samplewise	0.3948	0.1139	0.905	15.961	0.693	0.0000
3/5	Matched	0.3780	0.1081	0.904	21.987	0.887	0.0165
3/5	Tuned	0.2196	0.0643	0.912	292.668	3.656	0.2200
3/5	Calibrated	0.2192	0.0642	0.912	319.046	3.799	0.2244
1/9	ERM	0.3603	0.0961	0.967	4.386	0.184	0.0683
1/9	Samplewise	0.2774	0.0671	0.968	8.149	0.256	0.0000
1/9	Matched	0.2510	0.0594	0.966	13.414	0.356	0.0239
1/9	Tuned	0.0850	0.0222	0.976	109.105	0.688	0.1888
1/9	Calibrated	0.0797	0.0220	0.976	177.449	0.685	0.2104
0/8	ERM	0.3826	0.1050	0.968	3.975	0.179	0.0669
0/8	Samplewise	0.2998	0.0746	0.968	7.791	0.272	0.0000
0/8	Matched	0.2744	0.0666	0.967	11.215	0.342	0.0223
0/8	Tuned	0.0941	0.0238	0.972	202.786	1.223	0.1972
0/8	Calibrated	0.0790	0.0221	0.972	634.418	1.512	0.2343

Table G.5: Selected ridges and scales. Tuned ridge is .0002 in every setting. Calibrated ridge and scale are jointly validation-selected within the prescribed search.

Features	Pair	Calibrated ridge	Scale
ResNet-18	3/5	0.0002	1.4523
ResNet-18	1/9	0.0002	1.7702
ResNet-18	0/8	0.0002	1.6573
ViT-B/16	3/5	0.0002	1.0441
ViT-B/16	1/9	0.0010	2.2059
ViT-B/16	0/8	0.0002	1.7688

Table G.6: Paired test log-loss differences (control minus samplewise), 95% repeat-bootstrap intervals, and classification disagreement rates. Uncertainty is conditional on the test pool and selected parameters.

Features	Pair	Control	Difference	Interval	Disagreement
ResNet-18	3/5	Matched	-0.0004	[-0.0006, -0.0002]	0.0090
ResNet-18	3/5	Tuned	-0.2115	[-0.2163, -0.2067]	0.1723
ResNet-18	3/5	Calibrated	-0.2136	[-0.2210, -0.2061]	0.1723
ResNet-18	1/9	Matched	-0.0020	[-0.0023, -0.0017]	0.0102
ResNet-18	1/9	Tuned	-0.3327	[-0.3349, -0.3305]	0.0782
ResNet-18	1/9	Calibrated	-0.3617	[-0.3647, -0.3586]	0.0782
ResNet-18	0/8	Matched	-0.0034	[-0.0036, -0.0032]	0.0070
ResNet-18	0/8	Tuned	-0.3176	[-0.3211, -0.3141]	0.0805
ResNet-18	0/8	Calibrated	-0.3466	[-0.3523, -0.3411]	0.0805
ViT-B/16	3/5	Matched	-0.0167	[-0.0170, -0.0164]	0.0042
ViT-B/16	3/5	Tuned	-0.1752	[-0.1792, -0.1716]	0.0412
ViT-B/16	3/5	Calibrated	-0.1756	[-0.1798, -0.1719]	0.0412
ViT-B/16	1/9	Matched	-0.0265	[-0.0268, -0.0261]	0.0022
ViT-B/16	1/9	Tuned	-0.1925	[-0.1937, -0.1912]	0.0138
ViT-B/16	1/9	Calibrated	-0.1978	[-0.1992, -0.1961]	0.0092
ViT-B/16	0/8	Matched	-0.0254	[-0.0256, -0.0252]	0.0012
ViT-B/16	0/8	Tuned	-0.2056	[-0.2074, -0.2039]	0.0112
ViT-B/16	0/8	Calibrated	-0.2208	[-0.2236, -0.2181]	0.0112

Table G.7: Fair validation-selected comparison over five independent stratified splits. Entries are mean test log loss with split-level 95% normal intervals; the fixed test pool is never used for selection. The analytic match is a mechanism control, not a tuned performance baseline.

Features	Task	SW tuned	ERM tuned	ERM + scale	Analytic match
ResNet-18	cat/dog	0.380 ± 0.016	0.389 ± 0.022	0.378 ± 0.014	0.608 ± 0.002
ResNet-18	car/truck	0.173 ± 0.007	0.177 ± 0.007	0.174 ± 0.006	0.535 ± 0.002
ResNet-18	plane/ship	0.175 ± 0.005	0.176 ± 0.011	0.182 ± 0.012	0.522 ± 0.001
ViT-B/16	cat/dog	0.217 ± 0.008	0.217 ± 0.005	0.224 ± 0.009	0.378 ± 0.001
ViT-B/16	car/truck	0.082 ± 0.008	0.076 ± 0.005	0.152 ± 0.136	0.249 ± 0.001
ViT-B/16	plane/ship	0.077 ± 0.005	0.080 ± 0.006	0.140 ± 0.119	0.269 ± 0.001

test pool is held out from selection. Samplewise fits search the Cartesian grid

$$\tau \in \{10^{-5}, 3\times10^{-5}, 10^{-4}, 3\times10^{-4}, 10^{-3}, 3\times10^{-3}, 10^{-2}, \\ 3\times10^{-2}, 10^{-1}, 3\times10^{-1}, 1\}, \qquad \delta \in \{0, .25, .5, 1, 2\}.$$

while ordinary controls use the same expanded ridge grid and, for the calibrated version, a positive validation scale in $[.05, 20]$. Every fit uses the exact convex objective and the same numerical tolerance. The complete split records and selected configurations are included in the artifact package; an independent verification checks split disjointness, test-pool isolation, and solver residuals below 10^{-9}.

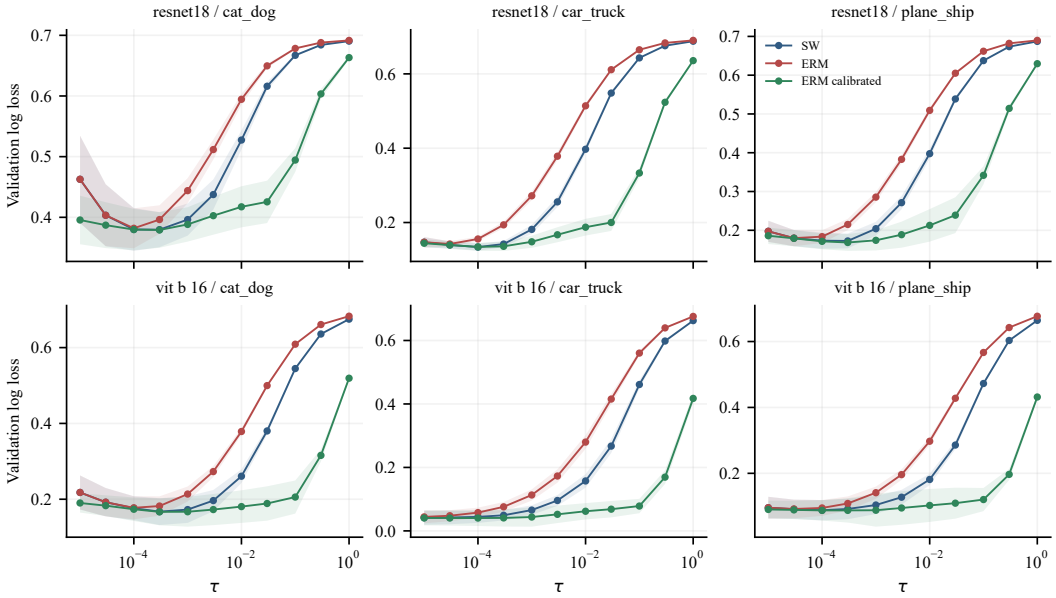

Figure G.3: Validation curves for the expanded ridge sweep. Each panel shows the mean validation log loss and split-level standard-deviation band over the five independent splits. The samplewise curve is the best envelope over the five tested radii, while the ordinary curves use the same ridge grid; the calibrated curve additionally selects a positive output scale. The minima at the smallest ridge values in several panels make the boundary behavior visible rather than hiding it behind a single selected point.

For the higher-order check, we use the independently optimized 1,024-point periodic population fits from Appendix F.5. The raw samplewise and matched remainders are the errors after subtracting the full cubic truncations,

$$R_{\mathrm{sw}}^{(5)} = \|f_{\epsilon,\delta} - \epsilon u_{1,\delta} - \epsilon^3 u_{3,\delta}\|_2, \qquad R_{\mathrm{ctl}}^{(5)} = \|g_{\epsilon,\delta} - \epsilon v_{1,\delta} - \epsilon^3 v_{3,\delta}\|_2,$$

where

$$v_{1,\delta} = \gamma_\delta u_{1,0}^{(\lambda_\delta)}, \qquad v_{3,\delta} = \gamma_\delta u_{3,0}^{(\lambda_\delta)}.$$

and have log–log slopes between 5.001 and 5.003 over radii 0.005–0.25. For the matched comparison, the uncentered quantity

$$M_{\epsilon,\delta} = \frac{f_{\epsilon,\delta} - g_{\epsilon,\delta}}{\epsilon^3}$$

converges to $\Delta_\delta[q]$; it therefore has a nonzero plateau whenever the probe is nondegenerate. We report the centered error

$$E_\Delta(\epsilon,\delta) = \|M_{\epsilon,\delta} - \Delta_\delta[q]\|_2,$$

whose slopes are between 2.00 and 2.03, as predicted by (B.1). At the unplotted radius-one stress test the slope is 1.97 because the finite signal grid is outside the cleanest local regime. These are deterministic population checks; they are not sampling confidence intervals. The numerical records and order summaries are included in the artifact package.

G.5 A REAL-COVARIATE FIXED-FEATURE INSTANCE

We use the packaged Wisconsin Diagnostic Breast Cancer dataset, containing 569 observations and 30 numerical features (Wolberg et al., 1993). Its binary labels directly match the scalar logistic model. One stratified 398/171 training/test split uses seed 927. Standardization is fitted only to the 398 training observations, an intercept is appended, and each 31-dimensional feature row is normalized to unit Euclidean norm. This map is frozen throughout the experiment. The untouched test split is used only for evaluation, with no radius or ridge selection.

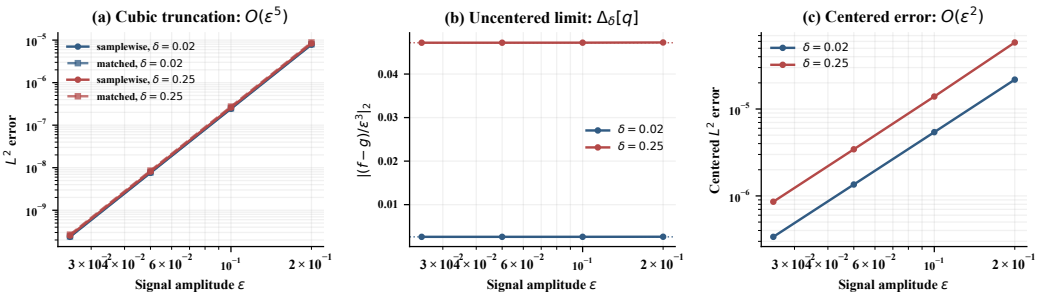

Figure G.4: Numerical order checks for the cubic separation theorem. The raw population-fit remainders subtract the complete cubic truncations and follow the predicted $O(\epsilon^5)$ order. The middle panel shows the uncentered matched difference divided by ϵ^3, which approaches the predicted nonzero $\Delta_\delta[q]$; the final panel centers this quantity by $\Delta_\delta[q]$ and follows the predicted $O(\epsilon^2)$ order. The radius-one stress test is outside the cleanest local regime.

We draw 100 bootstrap training samples of size 398 with seeds $882000+$rep. Each is fitted at $\tau = 0.02$ and $\delta \in \{0, 0.5, 1\}$, using both its true labels and an independent set of balanced randomized labels. Resampled inputs and randomized labels are shared across radii. These 600 fits optimize the exact fixed-feature objective without Gram filtering. A separate L-BFGS implementation reproduces one independently fitted test-score vector within 9.86×10^{-9}. Saved split indices, feature arrays, labels, preprocessing parameters, library version and raw-data hashes permit reconstruction of the instance.

For true labels, Table G.8 reports mean test metrics across the 100 fitted bootstrap samples. Score variance is the mean, over fixed test inputs, of the variance of fitted logits across resamples. The spread measures resampling variability conditional on the fixed split; it is not uncertainty from independent population test samples.

Table G.8: True-label frozen-feature instance, averaged over 100 bootstrap training fits. These describe one fixed split.

Radius	Accuracy	Log loss	Brier score	Score variance
0	0.95018	0.23244	0.05722	0.007493
0.5	0.95199	0.20430	0.05068	0.010407
1	0.95205	0.18278	0.04633	0.013875

For randomized labels, the mean n-scaled test prediction second moments are 9.3857, 15.4958, and 25.7959 at the three radii, while the mean conditional linearized covariances are 9.2401, 15.4046, and 26.0877. The paired nonlinear differences are $+0.06434$, -0.04434, -0.52647; their difference from a direct subtraction of the preceding two lists is the label Monte Carlo term. This is a separate mechanism test on real covariates, conditioning on the frozen preprocessing, training pool and test probes. True labels are not treated as balanced. The example does not establish general task improvement, clinical utility or a claim about changing neural features.

H Objective, geometry, and statistical scope

Curvature and score covariance. At balanced labels and unit kernel diagonal, the random score at the population minimizer zero is $-s_\delta Y K_X$. Its covariance is $s_\delta^2 T$, whereas the population Hessian is $\tau I + p_\delta T$. Thus the sandwich $H_\delta^{-1}(s_\delta^2 T)H_\delta^{-1}$ must retain both operators. At radius zero the data Hessian and score covariance both equal $T/4$, but the regularized Hessian also contains τI. At positive radius, s_δ^2 and p_δ differ. Replacing the sandwich by an inverse Fisher-information operator would discard these differences and generally give the wrong covariance.

Shared perturbations. The comparison in Section 3 concerns a shared supremum of the average loss with the ridge outside. At the balanced population minimizer zero, its mean loss gradient

vanishes, so its first-order perturbation penalty is zero. The samplewise first-order penalty is instead the mean gradient norm, equal to $1/2$ for a normalized kernel. The shared objective therefore requires higher-order analysis at this point; the samplewise transfer and harmonic theorems do not characterize its minimizer. A finite-sample mean gradient need not vanish at zero. These distinctions also prevent an exact-minimizer comparison from being read as a claim about the dynamics of stochastic SAM updates.

Placement of the ridge and vanishing regularization. Putting $\tau\|f+\nu\|^2/2$ inside the supremum couples the perturbation to the norm and direction of the current predictor. The perturbation that minimizes the margin alone no longer resolves that inner optimization in general. Our margin reduction, score equation and covariance decomposition are derived with the ridge outside. Moreover, the inverse-Hessian bound and finite-sample constants depend on fixed positive τ; they do not provide a joint limit with $\tau \downarrow 0$. Such a limit would require additional conditions controlling the spectrum and the signal.

Binary logistic structure. The exact margin and harmonic signs use binary-logistic monotonicity, symmetry, and balanced Rademacher labels. A multiclass extension would require a vector-valued norm and inner-perturbation geometry; the present scalar formulas do not claim that setting.