BEYOND ESTIMATOR MOMENTS: MINIMAL INFORMATION FOR PREDICTING ZEROTH-ORDER ADAM'S RESPONSE

Anonymous authors
Paper under double-blind review

ABSTRACT

Zeroth-order (ZO) query laws are conventionally compared through estimator mean, covariance, and mean-squared error. We show that these statistics need not determine Adam's response: in a local separable exponential-loss model with an annealed stabilizer, equal-cost laws can match the zero-radius mean, full covariance, and MSE yet induce distinct mean normalized updates at their respective slow equilibria. The mechanism is *root adjustment*: equilibrium motion absorbs part of the direct query-law effect, leaving a generally oblique-projected response. For a sphere-supported class with prescribed data-contracted moments, we characterize exactly how much additional information is required for uniform next-order prediction. The minimum number of exact scalar expectation readings equals the rank of the surviving response operator and, under nondegeneracy and full realizability, is $d - r$, where d is the parameter dimension and r the active-row rank. Indistinguishable admissible laws prove necessity, while explicit polynomial readings attain the bound. On a matched angular-mixture family, a response-aligned reading estimated from only 16 i.i.d. directions reduces held-out prediction MSE to about 1.4% of the reference-only baseline across two disjoint phases, whereas a same-count orthogonal reading misses the response. Finite-radius bounds and a local initialized weak-limit result further delimit the regime in which the response characterization applies.

1 INTRODUCTION

Zeroth-order (ZO) query laws are commonly designed and compared through query cost and gradient-estimator bias, covariance, and mean-squared error (Nesterov & Spokoiny, 2017; Ghadimi & Lan, 2013; Ma & Huang, 2025). These statistics describe the quality of the stochastic gradient supplied to the optimizer. Adam then combines first- and second-moment filters driven by the same stochastic history through a nonlinear coordinatewise ratio (Chen et al., 2019). Matching the input statistics does not, by itself, determine the mean of this ratio.

A simple quadratic example exhibits the difference. Two rotated octagon query laws match estimator mean, full covariance and MSE at every positive smoothing radius, yet give different expected first normalized updates for standard Adam with a fixed stabilizer (Appendix F.1). This first-step difference does not settle what happens after the optimizer's state has adapted: changing the query law also changes the state at which its response is evaluated.

We examine this interaction through the mean normalized update at a query law's own local slow-variable equilibrium. In a separable exponential-loss model with an annealed stabilizer, we show that equilibrium motion absorbs part of the direct law forcing and leaves a generally oblique-projected component. We call this mechanism *root adjustment*. Equal-cost laws can still match the zero-radius estimator mean, MSE, and full covariance while having different own-root responses (Figure 1). The difference is determined by the part of the law effect that survives equilibrium adjustment.

This leads to the question *which additional statistics of the query law suffice to recover its adapted response?* Root adjustment suggests that only the surviving component needs to be identified. To make this precise, we represent each additional statistic as a preassigned exact expectation of a bounded single-direction function, which we call a reading. These readings depend on the law

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.

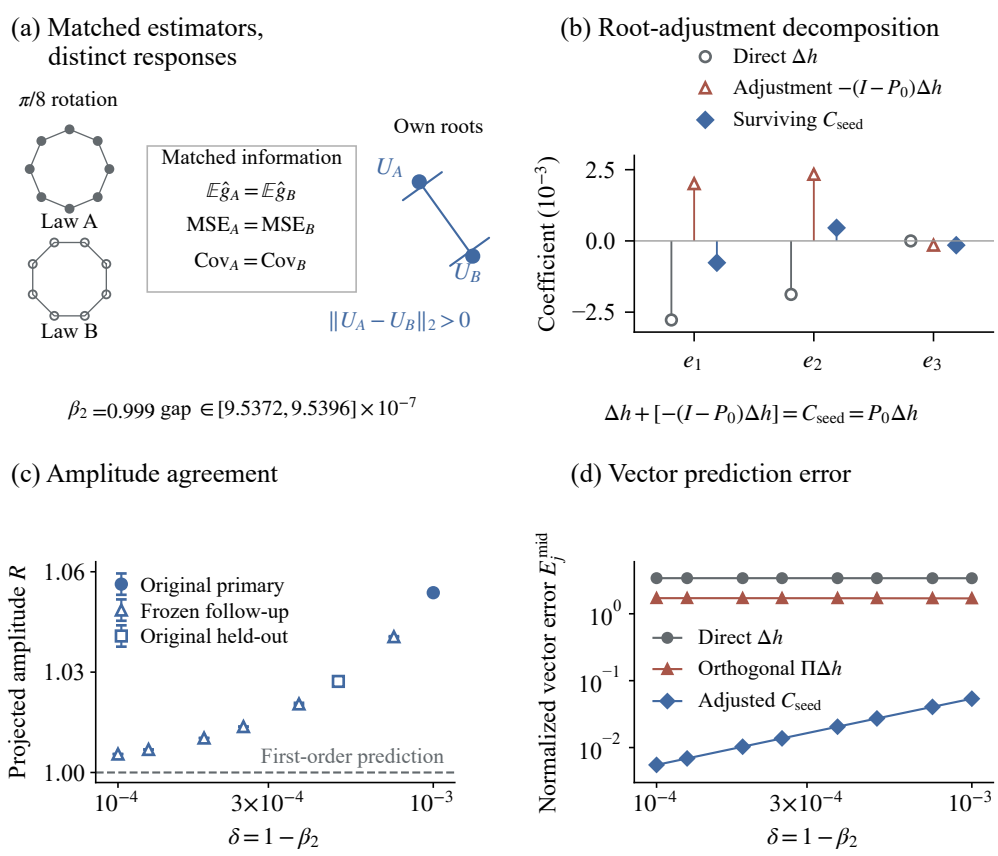


Figure 1: **Matched estimator summaries leave a response difference after equilibrium adjustment.** (a) Planar projections of two sixteen-atom query laws with an independent sign coordinate. At a common state, they match zero-radius estimator mean, MSE and full covariance, yet have distinct own-root responses (glyphs not to scale). (b) Root motion removes part of the direct forcing, leaving the oblique-projected coefficient. (c) Projected amplitude approaches its first-order prediction across eight tested memories; smaller δ means longer memory. (d) The root-adjusted predictor has smaller midpoint vector error than both alternatives at every displayed memory. Panel (c) shows stationary-response enclosures. Construction and non-cancellation: Appendix A.7–A.8. Normalization: Section 5.1; full endpoints: Appendix D.3.1.

without using its unknown own-root response or equilibrium. If compatible laws have the same readings but different responses, no predictor based on that information can distinguish them.

On a sphere-supported contracted-moment class, the first unmatched law information enters the adapted response through

$$U_{D,q} = U_q^{\circ} + \delta^{m-1} \mathsf{B}_{m,q} \theta(D) + O(\delta^m), \qquad \delta = 1 - \beta_2,$$

where $\theta(D)$ collects the first unmatched input moments and $\mathsf{B}_{m,q}$ is the response operator after equilibrium adjustment. Our main result gives the sharp information requirement

$$k_{\min} = \operatorname{rank} \mathsf{B}_{m,q}$$

for uniform next-order prediction. Under coordinate nondegeneracy and full realizability, this count is $d - r$, where d is the parameter dimension and r is the active-row rank. The lower bound is realized by legal query laws that remain indistinguishable under too few readings, while explicit polynomial readings attain the bound. The predictor thus needs only the component of the hidden moment vector that survives root adjustment.

Contributions. Minimal information after root adjustment. We prove that the rank of the surviving response operator is the minimum number of additional exact readings for uniform next-order prediction. Legal indistinguishable laws establish necessity, and explicit polynomial readings attain the bound.

Resolution and estimation of the required information. On separately specified three-dimensional classes, we characterize the next-order residual and the precision floor from inaccurate moments and readings. A sampled-reading identity separates deterministic residual error from projected moment variance: information dimension and estimation cost are distinct. On a matched angular mixture, response-aligned readings yield replicated held-out recovery from 16 directions.

Scope beyond the stationary characterization. A finite-radius bound preserves the leading response gap on a local branch. Under local contraction and coupled high-memory, long-time, small-step limits, the same operator governs initialized unconditional expected displacement. Independent computations test root adjustment, aligned recovery, and response dimension.

2 Related Work

ZO estimation and adaptivity. Classical ZO theory analyzes estimator error and convergence (Nesterov & Spokoiny, 2017; Ghadimi & Lan, 2013). ZO-AdaMM studies adaptive convergence and Mahalanobis geometry (Chen et al., 2019), ZO-signSGD normalized updates (Liu et al., 2019), and MeZO forward-only fine-tuning (Malladi et al., 2023). Perturbation design targets variance and directional alignment (Ma & Huang, 2025), while coordinate adaptivity affects step-size robustness (Dbouk et al., 2026). Our target is the sufficiency of supplied law summaries for an adapted response.

Adaptive dynamics. SDEs, modified equations and vector fields describe Adam dynamics (Malladi et al., 2022; Cattaneo et al., 2024; Dereich et al., 2025b); related analyses address mini-batch noise (Cattaneo & Shigida, 2026), attracting zeros and fluctuations (Dereich & Jentzen, 2024; 2026).

Adaptive geometry and implicit bias. Adaptive geometries (Duchi et al., 2011; Kingma & Ba, 2015) and memory (Reddi et al., 2018) interact with separable-problem implicit bias (Soudry et al., 2018; Gunasekar et al., 2018; Nacson et al., 2019; Ji & Telgarsky, 2021). Adam-specific results cover full-batch maximum-ℓ_∞ margins (Zhang et al., 2024), coordinatewise loss geometry (Xie et al., 2025), structured per-sample bias and a high-memory proxy (Baek et al., 2026), and quadratic-model symmetry (Dereich et al., 2025a). Here the optimizer is fixed while query laws change the shared-history response and its equilibrium.

Information recovery. Optimal recovery and moment-constrained inference reconstruct targets from linear measurements (Donoho, 1994; Bertsimas & Popescu, 2005; Foucart & Liao, 2024). Our measurements are bounded single-direction expectations and our target is the own-root response. Root motion changes the relevant operator, so legal-law lower bounds and attaining readings must recover its surviving image. The proof uses harmonic realizability, finite moment representations and singular-subspace tools at their supplementary points of use.

3 Problem Setup

Our primary target is the mean response after the normalized slow state has adapted to the query law. We first define its stochastic input and shared-history normalization, then the equilibrium and the information available to a predictor.

Model and recursion. Let $Z \in \mathbb{R}^{r\times d}$ have r independent active rows $z_i^\top$ and positive weights r_i. Set

$$L(w) = \sum_{i=1}^{r} r_i e^{-z_i^\top w}, \qquad p_i(w) = r_i e^{-z_i^\top w}/L(w), \qquad a(p) = Z^\top p. \tag{1}$$

Here p_i is row i's share of the loss, so $a(p)$ is the loss-weighted average row and $-\nabla L = La(p)$. The masses preserve relative row contributions as the overall loss decreases. The two-point oracle is $\widehat{g}_\mu(w,v) = v[L(w+\mu v) - L(w-\mu v)]/(2\mu)$, with zero-radius extension $\widehat{g}_0 = vv^\top \nabla L$. Each

step averages q IID directions into $\overline{g}_t$, using $2q$ off-center evaluations. Comparisons fix the loss, optimizer parameters and schedules, changing only the query law D. The shared-history Adam recursion uses the same batch average in both filters:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\overline{g}_t, \qquad v_t = \beta_2 v_{t-1} + (1-\beta_2)\overline{g}_t^{\odot 2}, \tag{2}$$

$$w_{t+1} = w_t - \eta_t \widehat{m}_t/(\sqrt{\widehat{v}_t} + \epsilon_t), \qquad \epsilon_{t+1} = \epsilon_t e^{-\kappa\eta_t}. \tag{3}$$

Operations are coordinatewise; hats are the usual bias corrections. Both filters use the same stochastic history, so their ratio is sensitive to the joint behavior of numerator and denominator. The stabilizer has fixed annealing rate $\kappa > 0$ and is outside the square root, and the second filter squares the batch average. Write $\delta = 1-\beta_2$ and $\lambda = \epsilon/L$; full corrections and coordinates are in Appendix A.1.

Response at a law's own root. Let $X_j(p,v) = v_j v^\top a(p)$ and let $G_{q,j}$ be its IID batch average. In slow coordinates $x = (\log(p_1/p_r), \dots, \log(p_{r-1}/p_r), \log\lambda)$, use $X_j(x,v) = X_j(p(x),v)$ and $a(x) = a(p(x))$. The same stationary batch history drives

$$A_j = (1-\beta_1)\sum_{k\geq 0}\beta_1^k G_{q,j,-k}, \quad V_j = \delta\sum_{k\geq 0}(1-\delta)^k G_{q,j,-k}^2,$$
$$H_{D,\delta,q}(x) = \mathbb{E}_D[A/(\sqrt{V}+\lambda)]. \tag{4}$$

The expectation is taken after forming the correlated-filter ratio; replacing it by a ratio of separate expectations would define a different response. Since $X = -\widehat{g}_0/L$, the averaged descent dynamics have sign $\dot{w} = H$. The slow drift has components $(z_r - z_i)^\top H$ for $i < r$ and $a(p)^\top H - \kappa$ for the last coordinate. An interior attracting root therefore defines

$$U_{D,q} = H_{D,\delta,q}(x_{D,\delta,q}), \qquad ZU_{D,q} = \kappa\mathbf{1}. \tag{5}$$

The first $r-1$ drift equations equate active-margin growth rates; the last matches loss decay to stabilizer decay and fixes their common rate to κ. The slow variables can equilibrate while the parameters continue to move; the own-root response describes this continuing mean motion. The row equation confines all own-root responses to an affine space of dimension $d-r$. Root branches are local, with masses and λ bounded away from zero; we suppress the δ dependence of U. Changing D moves the root, so fixed-state differences need not equal own-root differences.

Available information. A predictor given the same summaries must assign the same estimate to every compatible law. For admissible laws $\mathcal{C}$ sharing the supplied summaries, the smallest worst-case Euclidean error is

$$R(\mathcal{C}) = \inf_{u\in\mathbb{R}^d}\sup_{D\in\mathcal{C}}\|u - U_{D,q}\|_2. \tag{6}$$

A compatible pair gives the lower bound $\|U_a - U_b\|_2/2$. An additional reading is $\mathbb{E}_D\psi(v)$ for a preassigned bounded function of one direction. Equal readings define a fiber. We seek uniform prediction order over the stated class and root branch; reading counts measure exact scalar expectations, not bits, queries or statistical sample complexity.

4 What Information Survives Equilibrium Adjustment?

We first identify the law effect that survives equilibrium motion and the minimal readings that recover it. We then separate the remaining deterministic ambiguity from the error of estimating those readings, before connecting the stationary operator to initialized dynamics.

4.1 Root adjustment

Let D, D_{ref} have interior attracting roots in a common convex differentiable local domain, with the connecting segment inside it. Suppress δ, q and set $E_D = \int_0^1 D_x H_D(x_{\mathrm{ref}} + t(x_D - x_{\mathrm{ref}}))\,dt$. The response difference separates into the direct forcing $\Delta H = H_D(x_{\mathrm{ref}}) - H_{\mathrm{ref}}(x_{\mathrm{ref}})$ and the state contribution $E_D(x_D - x_{\mathrm{ref}})$. Both roots obey the same row equation, so their difference satisfies $Z\Delta H + ZE_D(x_D - x_{\mathrm{ref}}) = 0$. When ZE_D is invertible, eliminating the state displacement gives the exact identity

$$U_D - U_{\mathrm{ref}} = P_D[H_D(x_{\mathrm{ref}}) - H_{\mathrm{ref}}(x_{\mathrm{ref}})], \qquad P_D = I - E_D(ZE_D)^{-1}Z. \tag{7}$$

Let H_0 be the common population high-memory field determined by the matched information (written $H_{0,q}$ for general fixed q below). At its root x_0, $E_0 = D_x H_0(x_0)$ gives $P_0 = I - E_0(ZE_0)^{-1}Z$. Its range is $\ker Z$ and kernel is $\operatorname{range} E_0$: root motion absorbs state-sensitive directions, so the projection is generally oblique. Figure 1(b) depicts this leading P_0, distinct from finite-displacement P_D (Appendix A.1, Appendix A.6).

For the $q = 1$, $m = 2$ matched pair below, fixed β_1 gives $H_D(x, \delta) = H_0(x) + \delta h_D(x) + O_{C^1}(\delta^2)$ (Theorem 9). Thus $\Delta h = h_{D_A}(x_0) - h_{D_B}(x_0)$ is a coefficient, with $U_{D_A} - U_{D_B} = \delta P_0 \Delta h + O(\delta^2)$. Its decomposition is

$$\underbrace{\Delta h}_{\text{direct forcing}} \;=\; \underbrace{(I - P_0)\Delta h}_{\text{absorbed by root motion}} \;+\; \underbrace{P_0 \Delta h}_{\text{surviving response}} \quad .$$

A Euclidean projection onto $\ker Z$ enforces the same row constraint but generally removes a different component: the absorbed directions are set by E_0, not by the row space of Z alone. A law change therefore acts through two channels: it changes the update at the old state and moves the state itself. Forcing along $\operatorname{range} E_0$ can be offset by this motion; the surviving component is what distinguishes the adapted responses.

For a concrete witness, take $v = (\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, R)$, with uniform $\theta \in \{\phi + k\pi/4 : 0 \le k < 8\}$ and an independent Rademacher sign R. Octagon quadrature makes planar moments through degree seven independent of ϕ. The rotations 0 and $\pi/8$ give sixteen-atom equal-cost laws with matching zero-radius estimator mean, MSE and full covariance. At the explicit stable two-row seed, their unmatched moments give $P_0 \Delta h \neq 0$. This existence witness is established in Appendix A.7–A.8; the reading theorem below identifies the information that survives more generally.

4.2 Root-adjusted minimal information

The witness leaves open which parts of the unmatched law information matter. We now specify an information class on which the answer is sharp, including an explicit predictor and a lower bound realized by probability laws.

Fix $d \ge 3$, $m \ge 2$, $q \in \mathbb{N}$, $0 \le \beta_1 < 1$ and a rank-r matrix Z with exactly r active rows, $1 \le r \le d$. Let the isotropic reference D° on $\|v\|^2 = d$ satisfy $D^\circ \ge c^\circ \sigma_d$ for $c^\circ > 0$ and uniform surface measure σ_d. Define the contracted class by the finite constraints

$$\mathcal{C}_m^Z(D^\circ) = \left\{ D \text{ on } \|v\|^2 = d : (\mathbb{E}_D - \mathbb{E}_{D^\circ}) \left[v_j^{|\alpha|} \prod_{i=1}^{r} (z_i^\top v)^{\alpha_i} \right] = 0,\ j \le d,\ 1 \le |\alpha| \le 2m-1 \right\}, \tag{8}$$

where $\alpha \in \mathbb{N}_0^r$. By the multinomial identity these match $\mathbb{E}_D X_j(p, v)^k$ throughout the mass simplex, for $1 \le k \le 2m-1$, including state derivatives. They preserve aggregate estimator MSE on the data span, but do not impose full isotropy or full estimator covariance on every member (Appendix A.2). The same reading characterization and residual order hold in the smaller class matching all query moments through degree $4m - 1$, which fixes the zero-radius estimator mean and full covariance at a common state for fixed q.

Set $f_\lambda(s) = (\lambda + \sqrt{s})^{-1}$, with derivatives in s, and

$$\nu_{q,j}(x) = (1 - q^{-1}) a_j(x)^2 + q^{-1} \mathbb{E}_{D^\circ} X_j(x, v)^2, \qquad H_{0,q,j}(x) = a_j(x) f_{\lambda(x)}(\nu_{q,j}(x)).$$

Assume its physical slow field has an interior Hurwitz root x° with $\lambda^\circ > 0$. Write $a^\circ = a(x^\circ)$, $E^\circ = D_x H_{0,q}(x^\circ)$ and $P^\circ = I - E^\circ(ZE^\circ)^{-1}Z$; stability ensures the inverse exists.

Theorem 1 (Root-adjusted minimal information)**.** *Under these hypotheses, for all sufficiently small $\delta > 0$ every $D \in \mathcal{C}_m^Z(D^\circ)$ has a unique attracting root in a common local domain. Uniformly over the class,*

$$U_{D,q} = U_q^\circ + \delta^{m-1} \mathsf{B}_{m,q} \theta(D) + O(\delta^m), \qquad \theta_j(D) = (\mathbb{E}_D - \mathbb{E}_{D^\circ}) X_j(x^\circ, v)^{2m}, \tag{9}$$

where $U_q^\circ = U_{D^\circ,q}$ is the reference law's finite-memory own-root response (its δ dependence is suppressed), and

$$\mathsf{B}_{m,q} = q^{1-2m} P^\circ \operatorname{diag}_j \left[\frac{a_j^\circ f_{\lambda^\circ}^{(m)}(\nu_{q,j}(x^\circ))}{m\, m!} \right]. \tag{10}$$

The realizable span of $\theta(D)$ is $\mathbb{R}^d$. Supplied readings of the form $\mathcal{A}\theta$ remove leading ambiguity on every fiber exactly when $\ker \mathcal{A} \subseteq \ker \mathsf{B}_{m,q}$. Among all bounded measurable single-direction expectation readings, the sharp count for uniform $O(\delta^m)$ prediction is

$$k_{\min} = \operatorname{rank} \mathsf{B}_{m,q}, \tag{11}$$

attained by bounded degree-$4m$ polynomial readings. If $a_j^\circ \neq 0$ for every j and $r < d$, the count equals $d - r$. If $r = d$, $ZU_{D,q} = \kappa \mathbf{1}$ determines the response exactly. Constants and thresholds may depend on fixed data, reference, q, m, d, β_1 and the chosen root.

Which information must be recovered? Corollary 5 supplies an explicit stable family for every fixed batch size and an open data neighborhood. The row constraint gives only the upper rank $d - r$. Under coordinate nondegeneracy, the diagonal factor in (10) is invertible, and P° has rank $d - r$. Legal realizability ensures that these directions can all be excited. The sharp count concerns the response-relevant image rather than the size of the hidden moment vector. The kernel criterion specifies what readings must retain: every hidden-moment change invisible to the readings must also be invisible to the leading response. Repeated readings or readings aimed at the wrong directions can fail even when their number equals $d - r$.

For example, when $d = 3$ and $r = 2$, response differences are confined to the line $\ker Z$. Under the theorem's nondegeneracy conditions, one scalar amplitude specifies the leading difference along that line. An aligned polynomial reading recovers this amplitude even though the hidden moment vector has three coordinates. A reading insensitive to moment changes that excite this line leaves distinct responses indistinguishable. The count is one because the response varies in one direction; sufficiency also requires measuring that direction.

Why fewer readings fail. With $B = \mathsf{B}_{m,q}$, positivity-preserving harmonic density perturbations give a legal finite-dimensional chart t with hidden moments $\theta = \Gamma t$, where Γ is invertible. Arbitrary bounded readings restrict to linear measurements of t, while the leading response is $B\Gamma t$. Fewer than $\operatorname{rank} B$ readings leave opposite legal laws with identical information and a nonzero leading gap. This chart argument does not assume that arbitrary readings globally factor through θ. The full proof and attaining construction are in Appendix A.2–A.3.

An attaining predictor. Put $M(v) = (X_j(x^\circ, v)^{2m})_{j=1}^d$, and let N have orthonormal columns spanning $\operatorname{range} B$. Then

$$\phi(v) = N^\top B(M(v) - \mathbb{E}_{D^\circ} M), \qquad \widehat{U}_D = U_q^\circ + \delta^{m-1} N \mathbb{E}_D \phi, \tag{12}$$

has uniform error $O(\delta^m)$. Indeed, $\mathbb{E}_D \phi = N^\top B\theta(D)$ and $NN^\top B = B$, so the correction reconstructs precisely the surviving image. The functions are public: they depend on the reference root and operator, not on the unknown law's own root. Since X_j is quadratic in v, each coordinate has degree $4m$ and is bounded on the sphere.

What remains after recovery. Under the same coordinate nondegeneracy with $r < d$, the aligned central fiber retains radius $\Theta(\delta^m)$. Higher-order harmonic perturbations preserve these readings but vary the next coefficient.

4.3 Recovery limits and sampled information

The remaining deterministic ambiguity has order δ^m. On the separate three-dimensional circle–sign central fiber, Theorem 20 identifies its sharp coefficient and a next reading that improves prediction by one order. In the sphere-supported moment box, inaccurate base moments and aligned readings instead give the three contributions $e + \delta^m + \delta^{m-1}\xi$. Appendices B.2 and B.3 state these finer classes and results; Appendix F.2 connects their deterministic tolerances to confidence widths.

Exact readings identify the leading response; sampling them adds a separate statistical cost. Draw n_s independent directions from D, replace $\mathbb{E}_D \phi$ in (12) by their average, and denote the result by $\widehat{U}_{n_s}$. With $r_D = U_{D,q} - [U_q^\circ + \delta^{m-1} B\theta(D)]$, exact public quantities give

$$\mathbb{E}_D \|\widehat{U}_{n_s} - U_{D,q}\|_2^2 = \|r_D\|_2^2 + \frac{\delta^{2m-2}}{n_s} \operatorname{tr}[B \operatorname{Cov}_{v \sim D}(M(v)) B^\top]. \tag{13}$$

Rank determines how many response-relevant coordinates must be retained; projected moment variance determines their estimation cost. All readings use the same directions. Averaging full moments and then projecting gives exactly the same estimator. Appendix F.2 proves the identity and a bounded-reading confidence bound, including the $k = 0$ case and separate public-reference and numerical costs. A sufficient number of independent direction samples for $O(\delta^m)$ stochastic error scales as δ^{-2} times the squared public envelope and a logarithmic factor.

Section 5.2 tests this variance identity and aligned recovery on a matched angular mixture. Appendix F.4 examines the sampling cost in a low-signal spherical instance.

The characterization also extends locally beyond the zero-radius oracle. For the original octagon pair at $m = 2, q = 1$,

$$U_{A,\delta,\mu} - U_{B,\delta,\mu} = \delta C_{\text{seed}} + O(\delta^2 + \delta\mu^2 + \mu^4), \qquad C_{\text{seed}} = P_0 \Delta h \neq 0.$$

The bound is uniform on a common local branch (Appendix F.3). Thus $\mu = O(\sqrt{\delta})$ preserves the $O(\delta^2)$ remainder. At positive radius the exponential-loss estimator summaries are quantitatively close rather than exactly matched; the radius is held fixed within each stationary history.

4.4 Connection to initialized dynamics

We now turn from own-root responses to expected displacement accumulated from initialization. Under local contraction and coupled high-memory, long-time, and small-step limits, the same operator governs this response on a fixed law chart and metric.

Fix a finite legal chart $dD_\vartheta = (1 + \vartheta^\top h)\, d\sigma_d$, $\|\vartheta\|_2 \leq \rho$, with bounded linearly independent h_j orthogonal to all degree-at-most-seven polynomials and $1/2 \leq 1 + \vartheta^\top h \leq 3/2$. Keep this ball and its Euclidean metric fixed. Define $F_n(\vartheta) = \mathbb{E}_\vartheta[(w_n - w_0)/S_n]$, $S_n = \sum_{t<n} \eta_t$: the expectation is unconditional. Use common w_0, zero raw filters, $\epsilon_0 = \lambda_0 L(w_0) > 0$, $\eta_t = \varepsilon(t+1)^{-3/4}$ and $\mu_t = \bar{\mu}(t+1)^{-1/4}$, with stabilizer annealing as in (3). The step multiplier ε differs from ϵ_t.

Theorem 2 (Local weak expected-response bridge)**.** *Assume the common initialized contracting-domain conditions for the fixed-q population field $H_{0,q}$ in Appendix C.3, with the common initialization bounds of Appendix C.1.1, positive masses and relative stabilizer, nonzero aggregate coordinates, and the declared initializer and schedules. Fix q, β_1 with $\beta_1^2 < 1 - \delta$. For $D^\circ = \sigma_d$ at its root x_q, set $K_{j\ell}(x_q) = \mathbb{E}_{\sigma_d}[X_j(x_q, v)^4 h_\ell(v)]$. Then $\theta(D_\vartheta) = K(x_q)\vartheta$ and the normalized chart operator is $B_q = q^3 \mathsf{B}_{2,q} K(x_q) \in \mathbb{R}^{d\times s}$. Along the coupled limit $\delta \downarrow 0$, $S_n \to \infty$ and $C_{q,\delta} q^3 \varepsilon/\delta \to 0$,*

$$\sup_{\|\vartheta\|\leq\rho} \left\| \frac{q^3}{\delta}[F_n(\vartheta) - F_n(0)] - B_q\vartheta \right\| \longrightarrow 0. \tag{14}$$

Here $C_{q,\delta}$ is the tracking constant in (80)*, finite for fixed memory. Under the same limit, the optimal worst-fiber radius R_k^{act} of F_n on the fixed chart ball after k preassigned bounded readings satisfies*

$$(q^3/\delta) R_k^{\text{act}} \longrightarrow \rho\sigma_{k+1}(B_q). \tag{15}$$

Appendix C.3 proves the fixed-q bridge and radius transfer; Appendix C.4 gives finite-error spectral resolution conditions. For $q = 1$, $m = 2$, and $k_* = \operatorname{rank} B_1$ aligned readings, the sufficient scales $S_n \gtrsim \delta^{-1}$ and $C_{1,\delta}\varepsilon \lesssim \delta^2$ yield an $O(\delta^2)$ prediction-error upper bound.

5 Numerical Validation

We test three stationary predictions: root adjustment selects the surviving law effect, aligned information improves finite-sample prediction on a specified matched family, and the information operator determines response dimension. Independent own-root computations supply the response truths; an initialized-tail replication examines a late-window directional response.[1]

[1] Anonymized reproducibility materials: `https://anonymous.4open.science/r/reproducibility-materials-112C/`.

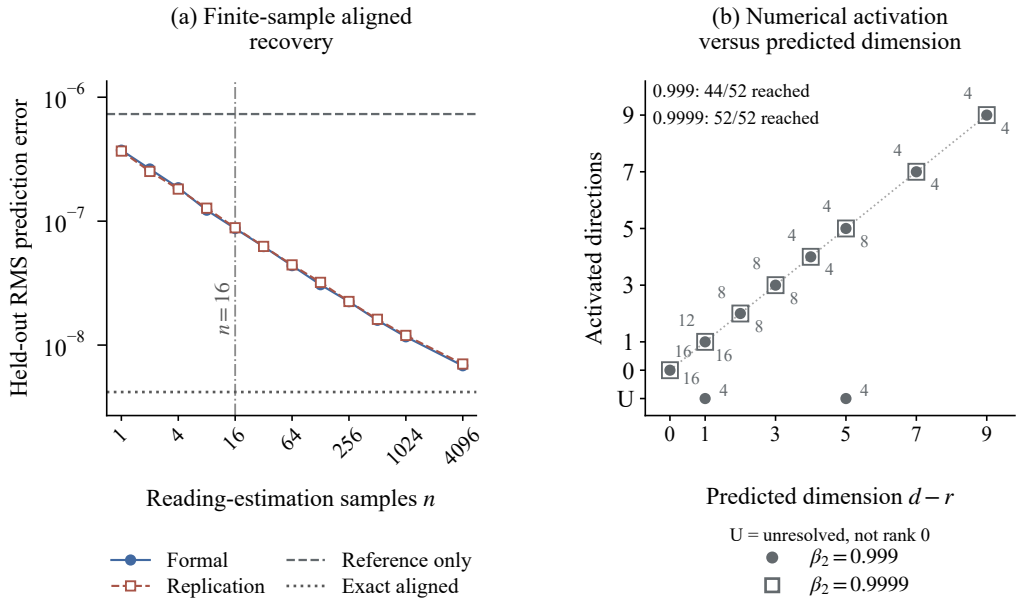


Figure 2: **Finite-sample aligned recovery and response dimension.** (a) Pooled held-out RMS prediction error on matched $D_{\pm 0.9}$ at $\beta_2 = 0.9995$. Target and $n = 16$ were frozen at $\beta_2 = 0.999$. Horizontal lines mark reference-only and exact-reading errors. Here n counts i.i.d. direction samples used to estimate the public reading, not optimizer objective calls. At $n = 16$, MSE ratios are 0.0142 and 0.0146. (b) Activation reaches $d - r$ in $44/52$ positive-rank configurations at $\beta_2 = 0.999$ and $52/52$ at 0.9999; eight shorter-memory cases remain unresolved at the enclosure floor. Marker counts, zero-rank controls and assurance are detailed in Appendix E.3–E.4.

5.1 Root adjustment

The matched octagon laws have distinct own-root responses, and root adjustment predicts their surviving difference (Figure 1). For this pair (Section 4.1), take $q = 1$, $\beta_1 = 0.9$ and eight tested memories. Each law's own root is solved independently of the prediction. With $\Delta U = U_A - U_B$ and $C_{\text{seed}} = P_0 \Delta h$, the projected-amplitude ratio is $R = \langle C_{\text{seed}}, \Delta U \rangle / (\delta \|C_{\text{seed}}\|^2)$ and the vector error is $E = \|\Delta U/\delta - C_{\text{seed}}\| / \|C_{\text{seed}}\|$. Figure 1 compares raw Δh, orthogonal $\Pi \Delta h$ and adjusted $P_0 \Delta h$, where $\Pi = I - Z^\top (Z Z^\top)^{-1} Z$. The adjusted predictor incorporates state sensitivity through the population-root derivative.

The adjusted predictor has smaller midpoint relative vector error than both controls at every displayed memory. Its full-vector error decreases from about 5.37% to 0.55% as memory increases, while the absolute response gap also decreases. For this three-dimensional, two-row seed, the exact gap and C_{seed} lie in the same one-dimensional $\ker Z$, so $E = |R - 1|$: panels (c) and (d) show amplitude agreement and competing predictors for the same gap. Root enclosures establish nonzero response and unit-direction gaps at each memory; the adjusted uncertainty comes from these deterministic enclosures, whereas the two controls are evaluated at interval midpoints. Appendix A.8.2 and D.3.3 extend this comparison across rotations, including analytically extremal and first-order cancellation pairs.

5.2 Finite-sample reading recovery

The angular mixture makes the reading concrete. Let D_+ and D_- be the extremal matched octagon laws, and set $D_t = (1 + t) D_+/2 + (1 - t) D_-/2$ for $-1 \le t \le 1$. Varying t preserves their moments through degree seven while changing the response-relevant eighth-order moment. The public harmonic reading takes values $+1$ and -1 on the two supports, so its expectation is t and its variance is $1 - t^2$. Averaging it over sampled directions estimates the scalar amplitude; multiplying

by the known first-order response coefficient gives the correction to the equal-mixture reference's own-root prediction. Appendix F.5 specifies the reading and predictor.

Figure 2(a) tests actual prediction on the legal matched angular mixtures $D_{\pm 9/10}$ of Appendix F.5. The target magnitude and primary $n = 16$ were fixed using only $\beta_2 = 0.999$ development roots; n counts i.i.d. direction samples used to estimate the public reading. Evaluation uses independently enclosed finite-memory own-root truths at held-out $\beta_2 = 0.9995$.

The disjoint formal and replication phases, each with 1024 paired seed clusters, gave sampled/reference MSE ratios of 0.0142 and 0.0146. Paired-gain 95% intervals were $[5.269, 5.283] \times 10^{-13}$ and $[5.268, 5.279] \times 10^{-13}$; both signs improved. Same-sample full-moment projection matches the reading; the orthogonal control misses the response.

Empirical MSE follows the finite-n variance identity, linking the observed recovery to the projected moment variance in (13).

5.3 Response Dimension Across Stable Local Instances

Numerical activation reaches the predicted dimension in $44/52$ positive-rank configurations at $\beta_2 = 0.999$ and in all 52 at $\beta_2 = 0.9999$ (Figure 2(b)). The eight remaining shorter-memory configurations lie below the numerical resolution threshold and are marked unresolved. The 32 primary $r = d$ configurations serve as analytic zero-rank controls.

The $m = 2, q = 1, \beta_1 = 0.9$ study covers 68 data instances and 136 primary instance–memory configurations in small neighborhoods of 17 stable seeds (Appendix E.1). Each of the 17 settings contains an analytic seed and three preassigned data perturbations of Frobenius norm 10^{-5}, screened for the common local domain before response evaluation. The density chart preserves spherical moments through degree seven and has an invertible moment map K, with $\theta(D_t) - \theta(D_0) = Kt$. We compare $\delta^{-1} S_{\delta,h}$ to $\mathsf{B}_{2,1} K$, where $S_{\delta,h} e_j = [U(D_{he_j}, \delta) - U(D_{-he_j}, \delta)]/(2h)$ is the central secant used for the numerical comparison. The resolved dimension combines numerically separated singular directions with the analytic $d - r$ upper bound (Appendix E.3–E.4).

Initialized tail response. Two independently seeded replicas of 2,048 initialized trajectories each reproduce a normalized directional law response: their descriptive 95% intervals and the pooled interval lie within the prespecified 25% tolerance of the fixed-q high-memory target, with the pooled mean approximately 9.54% above it. The transport estimator measures a preassigned direction over updates 1,179–13,911, inheriting optimizer states and the absolute schedule clock continuously from initialization without resetting at the window boundary. This comparison measures the late-window directional response; Appendix D.3.2 reports the window protocol and the complete-prefix comparison.

6 Conclusion

Equal estimator summaries need not identify Adam's response, even at the first fixed-stabilizer update. In our local annealed model, root adjustment identifies the surviving information: its operator rank is the sharp exact-reading count, conditionally $d - r$, with legal laws establishing necessity and explicit readings attaining recovery. Finite-radius stability and the initialized bridge specify further regimes. The sampled-reading identity separates information dimension from estimation cost, and a legal matched angular mixture demonstrates replicated held-out recovery with response-aligned readings.

AI USE STATEMENT

Generative AI tools were used to provide ideas and feedback during the development of some mathematical arguments, to aid and polish the writing of this manuscript, and for retrieval and discovery, including identifying relevant related work. All mathematical claims and proofs in the paper were reviewed and verified by the authors. The authors take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

The supplement states the local hypotheses, proof dependencies, legal query-law constructions, numerical protocols, and statistical protocols. The accompanying source package links reported results to archived inputs and distinguishes analytic proofs, interval-certified root computations, high-precision numerical diagnostics, and Monte Carlo experiments according to their evidence scope.

REFERENCES

Beomhan Baek, Minhak Song, and Chulhee Yun. Implicit bias of per-sample Adam on separable data: Departure from the full-batch regime. In *International Conference on Learning Representations*, pp. 120180–120229, 2026.

Dimitris Bertsimas and Ioana Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005.

Matias D. Cattaneo and Boris Shigida. The effect of mini-batch noise on the implicit bias of Adam. *arXiv preprint arXiv:2602.01642*, 2026.

Matias D. Cattaneo, Jason Matthew Klusowski, and Boris Shigida. On the implicit bias of Adam. In *International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 5862–5906, 2024.

Xiangyi Chen, Sijia Liu, Kaidi Xu, Xingguo Li, Xue Lin, Mingyi Hong, and David Cox. ZO-AdaMM: Zeroth-order adaptive momentum method for black-box optimization. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Hassan Dbouk, Nidham Gazagnadou, Matthias Reisser, and Christos Louizos. On adaptivity in zeroth-order optimization. *arXiv preprint arXiv:2605.03869*, 2026.

Steffen Dereich and Arnulf Jentzen. Convergence rates for the Adam optimizer. *arXiv preprint arXiv:2407.21078*, 2024. doi: 10.48550/arXiv.2407.21078. URL https://arxiv.org/abs/2407.21078.

Steffen Dereich and Arnulf Jentzen. Central limit theorem for the averaged Adam optimizer. *arXiv preprint arXiv:2606.21433*, 2026. doi: 10.48550/arXiv.2606.21433. URL https://arxiv.org/abs/2606.21433.

Steffen Dereich, Thang Do, Arnulf Jentzen, and Philippe von Wurstemberger. Adam symmetry theorem: Characterization of the convergence of the stochastic Adam optimizer. *arXiv preprint arXiv:2511.06675*, 2025a.

Steffen Dereich, Arnulf Jentzen, and Sebastian Kassing. ODE approximation for the Adam algorithm: General and overparametrized setting. *arXiv preprint arXiv:2511.04622*, 2025b.

David L. Donoho. Statistical estimation and optimal recovery. *The Annals of Statistics*, 22(1): 238–270, 1994.

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(61):2121–2159, 2011.

Simon Foucart and Chunyang Liao. Radius of information for two intersected centered hyperellipsoids and implications in optimal recovery from inaccurate data. *Journal of Complexity*, 83: 101841, 2024.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. Characterizing implicit bias in terms of optimization geometry. In *International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1832–1841, 2018.

Ziwei Ji and Matus Telgarsky. Characterizing the implicit bias via a primal-dual analysis. In *International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pp. 772–804, 2021.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Sijia Liu, Pin-Yu Chen, Xiangyi Chen, and Mingyi Hong. signSGD via zeroth-order oracle. In *International Conference on Learning Representations*, 2019.

Shaocong Ma and Heng Huang. Revisiting zeroth-order optimization: Minimum-variance two-point estimators and directionally aligned perturbations. In *International Conference on Learning Representations*, pp. 13267–13293, 2025.

Sadhika Malladi, Kaifeng Lyu, Abhishek Panigrahi, and Sanjeev Arora. On the SDEs and scaling rules for adaptive gradient algorithms. In *Advances in Neural Information Processing Systems*, volume 35, pp. 7697–7711, 2022.

Sadhika Malladi, Tianyu Gao, Eshaan Nichani, Alex Damian, Jason D. Lee, Danqi Chen, and Sanjeev Arora. Fine-tuning language models with just forward passes. In *Advances in Neural Information Processing Systems*, volume 36, pp. 53038–53075, 2023.

Mor Shpigel Nacson, Jason Lee, Suriya Gunasekar, Pedro Henrique Pamplona Savarese, Nathan Srebro, and Daniel Soudry. Convergence of gradient descent on separable data. In *International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 3420–3428, 2019.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of Adam and beyond. In *International Conference on Learning Representations*, 2018.

Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Suriya Gunasekar, and Nathan Srebro. The implicit bias of gradient descent on separable data. *Journal of Machine Learning Research*, 19(70): 1–57, 2018.

Shuo Xie, Mohamad Amin Mohamadi, and Zhiyuan Li. Adam exploits ℓ_∞-geometry of loss landscape via coordinate-wise adaptivity. In *International Conference on Learning Representations*, pp. 7937–7965, 2025.

Chenyang Zhang, Difan Zou, and Yuan Cao. The implicit bias of Adam on separable data. In *Advances in Neural Information Processing Systems*, volume 37, pp. 23988–24021, 2024.

SUPPLEMENTARY CONTENTS

Appendices A–C provide the proofs for minimal information, recovery resolution, and initialized dynamics. Appendix D contains numerical enclosures, information–precision slices (Appendix D.4), and initialized-tail replication (Appendix D.3.2); Appendix E examines response geometry across stable instances. The angular extension and its tests are in Appendices A.8.2 and D.3.3; the concrete initialized pair is in Appendix C.1.2, and paired reading tolerances in Appendix D.4.3. Appendix F gives the fixed-stabilizer example, sampled-information and finite-radius connections, and contrasting low-SNR and SNR-aware independent sampling checks (Appendices F.4 and F.5). PDF bookmarks expose the individual proof subsections within each entry.

A Core information: shared history, root adjustment, and reading rank . 14

A.1 Proof guide and common notation 14

A.2 Information rank beyond a prescribed seed 15

A.3 Legal moment directions 18

A.4 The uniform shared-history expansion 19

A.5 The first two shared-filter coefficients 20

A.6 The physical two-row Jacobian . 22

A.7 Matched query laws and the first moment they leave free 22

A.8 The exact two-row seed: stability and non-cancellation 24

B Proofs of residual sharpness and precision 26

B.1 Exact-class support, moments and witnesses 27

B.2 The first residual order on the entire exact central fiber 29

B.3 Inexact moments and an additional expectation 30

C Proof of the initialized-response bridge 34

C.1 A law-uniform initialized domain 34

C.2 Scope of the transfer results . . . 37

C.3 The fixed-batch expected-response bridge 37

C.4 Spectral Interpretation and Prediction Resolution 39

D Numerical response bounds and experimental evidence 39

D.1 Moment computation and response enclosures 40

D.2 Computing stationary responses at finite memory 40

D.3 Stationary comparisons and initialized-tail replication 43

D.4 Numerical Resolution under Exact and Inexact Information . . . 47

E Additional response-geometry evidence across stable instances 54

E.1 Configurations, data instances and query-law coordinates 54

E.2 Theoretical operator, finite secants and rank information 54

E.3 Resolution limits across memory scales 54

E.4 Numerical assurance for response and operator calculations 55

E.5 Reading geometry and actual named-law evidence 55

F Connections to fixed stabilizers and sampled information 56

F.1 An exact first update 56

F.2 Sampling the attaining readings . 56

F.3 Positive-radius own-root responses 57

F.4 Independent sampling check . . . 58

F.5 SNR-aware finite-sample recovery on a matched angular mixture 58

Supplementary references 60

A CORE INFORMATION: SHARED HISTORY, ROOT ADJUSTMENT, AND READING RANK

Theorem 1 combines a uniform shared-history expansion, equilibrium adjustment, and legal moment perturbations. Appendix A.2 assembles these ingredients; the following subsections prove the required tools and construct explicit matched laws and stable seeds.

A.1 PROOF GUIDE AND COMMON NOTATION

All norms are Euclidean unless a different norm is specified. Absolute values, products, divisions, and square roots of vectors are coordinatewise. The rows of Z are $z_i^\top$, p lies in the mass simplex, and $a(p) = Z^\top p$. The symbol H_0 denotes the population high-memory field, whereas H_{ref} includes the complete finite-memory shared history of the reference law. A root is always a root of the full physical slow system, even when an equivalent fixed-row equation is used to solve for it.

The pre-query sigma-field contains the initializer and all previous states and queries; the new query is independent of it. State coordinates and schedules are predictable. Both the fresh-query and Poisson martingale terms are retained in the initialization proof, and are not presumed independent. Every first-exit argument includes the update that could leave the domain.

A.1.1 TECHNICAL DEFINITIONS USED IN THE MAIN TEXT

Recursion and inverse coordinates. For the recursion (2)–(3), the bias corrections are $\widehat{m}_t = m_t/(1-\beta_1^{t+1})$ and $\widehat{v}_t = v_t/(1-\beta_2^{t+1})$. The inverse slow chart is $p_i = e^{x_i}/(1+\sum_{j<r} e^{x_j})$ for $i < r$, $p_r = (1+\sum_{j<r} e^{x_j})^{-1}$ and $\lambda = e^{x_r}$. Since $\log(p_i/p_r) = \log(r_i/r_r) + (z_r - z_i)^\top w$, differentiation with $\dot{w} = H$ gives the mass drift in Section 3. Likewise $\dot{\log}\,\epsilon = -\kappa$ and $\nabla \log L = -a$ give $\dot{\log}\,\lambda = a^\top H - \kappa$. Initializing $\epsilon_0 = \lambda_0 L(w_0)$ may require one startup query if $L(w_0)$ is unavailable.

Positive-radius input and stationary history. The zero-radius extension of the two-value query is

$$\widehat{g}_0(w) = vv^\top \nabla L(w). \tag{16}$$

For positive radius, the exponential-loss identity is $-\widehat{g}_\mu/L = v\sum_i p_i \sinh(\mu z_i^\top v)/\mu$. The schedules $\eta_t = \varepsilon(t+1)^{-3/4}$ and $\mu_t = 0.01(t+1)^{-1/4}$ satisfy $\sum_t \eta_t = \infty$, $\sum_t \eta_t^2 < \infty$, and $\sum_t \eta_t \mu_t^2 < \infty$. Here 0.01 is an example of the fixed $\bar{\mu} > 0$ used in Section 4.4. At fixed p, λ, put $X_j(p, v) = v_j v^\top a(p)$ and

$$A_j = (1-\beta_1)\sum_{k\geq 0} \beta_1^k X_{j,-k}, \qquad V_j = \delta \sum_{k\geq 0} (1-\delta)^k X_{j,-k}^2, \qquad H_D(p,\lambda) = \mathbb{E}_D \frac{A}{\sqrt{V}+\lambda}.$$

These displayed filters use the $q = 1$ input. For fixed q, replace the history input $X_{j,-k}$ by the same batch average $G_{q,j,-k}$ in both filters, leaving the single-direction polynomial $X_j(p, v)$ unchanged. The row count in the following display is $n = r$ of the main text. Both filters use the same two-sided IID history. In coordinates $x = (s_1, \ldots, s_{n-1}, \ell)$, with $s_i = \log(p_i/p_n)$ and $\ell = \log\lambda$, the full physical slow field is

$$G_D(x) = \begin{pmatrix} (z_n - z_1)^\top H_D \\ \vdots \\ (z_n - z_{n-1})^\top H_D \\ a(p)^\top H_D - \kappa \end{pmatrix}. \tag{17}$$

Its interior roots satisfy $ZU_D = \kappa \mathbf{1}$; stability refers to the full physical field G_D.

Exact root adjustment, including a numerical center. Changing the law changes both the field and the state at which it is evaluated. We separate these effects before imposing the row constraint. Let x_D be a root and let $\widehat{x}, \widehat{U}$ be a reference state and response in the common convex differentiability domain. Set

$$e_D = H_D(\widehat{x}) - \widehat{U}, \qquad \varrho = Z\widehat{U} - \kappa\mathbf{1}, \qquad \overline{E}_D = \int_0^1 D_x H_D(\widehat{x} + t(x_D - \widehat{x}))\, dt.$$

The fundamental theorem of calculus gives $U_D - \widehat{U} = e_D + \overline{E}_D(x_D - \widehat{x})$. Since $ZU_D = \kappa\mathbf{1}$, multiplication by Z determines the displacement whenever $Z\overline{E}_D$ is invertible:

$$x_D - \widehat{x} = -(Z\overline{E}_D)^{-1}(Ze_D + \varrho).$$

Substitution yields

$$U_D - \widehat{U} = [I - \overline{E}_D(Z\overline{E}_D)^{-1}Z]e_D - \overline{E}_D(Z\overline{E}_D)^{-1}\varrho. \tag{18}$$

For an exact reference root, take $\widehat{x} = x_{\mathrm{ref}}$, $\widehat{U} = U_{\mathrm{ref}}$ and $\varrho = 0$ to obtain

$$U_D - U_{\mathrm{ref}} = [I - \overline{E}_D(Z\overline{E}_D)^{-1}Z][H_D(x_{\mathrm{ref}}) - H_{\mathrm{ref}}(x_{\mathrm{ref}})]. \tag{19}$$

Thus state adjustment removes the component of the direct law effect absorbed by the root equation. A numerical row residual contributes the separate final term in (18). The identity requires the joining segment to lie in the common domain and its averaged row derivative to be invertible; the local root results below establish those conditions for each application.

Appendix A.6 verifies its relation to the physical Jacobian, including the product-rule term.

A.2 Information rank beyond a prescribed seed

We now prove Theorem 1. Data contractions first identify the shared-history moments and their derivatives; the resulting field comparison gives common roots and a projected response. The realizable harmonic family then turns the algebraic information criterion into a lower bound against arbitrary bounded readings. The stable family at the end supplies explicit instances of the hypotheses.

In this proof route, the subscript-circle symbols $D_\circ, x_\circ, a_\circ, E_\circ, P_\circ, U_{\circ,q}$ denote the main text's $D^\circ, x^\circ, a^\circ, E^\circ, P^\circ, U_q^\circ$; $\mathbb{E}_\circ = \mathbb{E}_{D^\circ}$. In particular $U_{\circ,q}$ is the finite-memory reference own-root response. These aliases also apply in Appendix A.3 and do not identify the seed, fixed-q or infinite-batch roots.

We now allow nonsymmetric data and a finite IID batch size. Fix $d \geq 3$, $m \geq 2$, $q \geq 1$, $0 \leq \beta_1 < 1$, and a rank-r matrix $Z \in \mathbb{R}^{r\times d}$ with exactly r active rows, $1 \leq r \leq d$. Write σ_d for uniform measure on $\|v\|^2 = d$. The public reference $D_\circ$ is a probability law on this sphere, is isotropic, and has a surface component:

$$\mathbb{E}_\circ vv^\top = I_d, \qquad D_\circ \geq c_\circ\sigma_d \quad \text{as measures, for some } c_\circ > 0. \tag{20}$$

The surface component permits moment perturbations that remain probability measures; mixtures of σ_d with other isotropic sphere laws qualify.

A.2.1 Moments adapted to the data

For a multi-index $\alpha \in \mathbb{N}_0^r$, define $F_{j,\alpha}(v) = v_j^{|\alpha|}\prod_{i=1}^r (z_i^\top v)^{\alpha_i}$. Let $\mathcal{C}_m^Z(D_\circ)$ consist of all probability laws on the sphere whose expectations of these polynomials agree with $D_\circ$, for $j = 1,\dots,d$ and $1 \leq |\alpha| \leq 2m-1$. This is a convex class. An explicit, possibly redundant, descriptor table has at most

$$N_{Z,m} = d\left[\binom{r+2m-1}{2m-1} - 1\right] \tag{21}$$

entries. This is a count of supplied expectations, not of samples or bits. The full query-moment class matching $D_\circ$ through degree $4m-1$ is a subclass. We do not impose full isotropy on every member of the larger class.

Lemma 3 (Data contractions determine the needed input moments)**.** *For $X_j(p,v) = v_j v^\top Z^\top p$ and $1 \leq k \leq 2m-1$,*

$$\mathbb{E}_D X_j(p,v)^k = \sum_{|\alpha|=k}\binom{k}{\alpha}p^\alpha \mathbb{E}_\circ F_{j,\alpha}(v), \qquad D \in \mathcal{C}_m^Z(D_\circ). \tag{22}$$

Thus these moments and all their state derivatives agree on the entire mass simplex. For $G_{q,j} = q^{-1}\sum_{\ell=1}^q X_j(p,v_\ell)$ with IID directions, the batch moments through degree $2m-1$ also agree, and

$$(\mathbb{E}_D - \mathbb{E}_\circ)G_{q,j}^{2m} = q^{1-2m}(\mathbb{E}_D - \mathbb{E}_\circ)X_j^{2m}. \tag{23}$$

Every member has $\mathbb{E}_D(vv^\top)a = a$ for $a \in \operatorname{range} Z^\top$. Consequently, on that subspace, its zero-radius batch input has the same aggregate mean-square error $(d-1)\|a\|^2/q$ as the reference.

Proof. The multinomial formula gives (22) as a polynomial identity in p, hence also for state derivatives. Grouping a batch moment by query index, all terms below order $2m$ use matched moments. At order $2m$, only the single-index term can differ, with coefficient q/q^{2m}, proving (23) and its derivatives. For $\alpha = e_i$, $\mathbb{E}_D[v_j(z_i^\top v)] = z_{i,j}$, so $\mathbb{E}_D(vv^\top)a = a$ throughout the row span. Sphere support gives $\mathbb{E}_D\|vv^\top a - a\|^2 = (d-2)a^\top \mathbb{E}_D(vv^\top)a + \|a\|^2 = (d-1)\|a\|^2$. Averaging q independent centered inputs divides this error by q. This establishes aggregate error matching on the row span, without full covariance matching or isotropy outside it. Summing $\binom{r+k-1}{k}$ over the degrees gives (21). □

Use the same batch input G_q in both infinite filters, and denote the resulting response by $H_{D,q}(x, \delta)$. Here x has $r-1$ mass logits and $\ell = \log \lambda$. Set $a = Z^\top p$,

$$T_j^\circ = \mathbb{E}_\circ[v_j^2 vv^\top], \qquad \nu_{q,j} = (1-q^{-1})a_j^2 + q^{-1}a^\top T_j^\circ a, \qquad H_{0,q,j} = a_j f_\lambda(\nu_{q,j}).$$

The formula for ν_q follows by separating equal and unequal batch indices. They are public on the contracted class. Define $G_{0,q}$ by (17), with $n = r$ and H_D replaced by $H_{0,q}$.

A.2.2 Uniform local roots and their response

Lemma 4 (Stable-root expansion for the contracted class)**.** *Suppose $G_{0,q}$ has an interior root $x_\circ$ with $\lambda_\circ > 0$ and Hurwitz Jacobian. Then $a_\circ \neq 0$, and there is a common convex local domain in which every $D \in \mathcal{C}_m^Z(D_\circ)$ has a unique attracting root for sufficiently small $\delta > 0$. Put*

$$E_\circ = D_x H_{0,q}(x_\circ), \quad P_\circ = I - E_\circ(ZE_\circ)^{-1}Z, \quad D_{m,q} = q^{1-2m}\,\mathrm{diag}_j\,\frac{a_{\circ,j} f_{\lambda_\circ}^{(m)}(\nu_{q,j}(x_\circ))}{m\,m!}.$$

The inverse exists. Uniformly over this whole contracted class,

$$U_{D,q} = U_{\circ,q} + \delta^{m-1} P_\circ D_{m,q}\theta(D) + O(\delta^m), \qquad \theta_j(D) = (\mathbb{E}_D - \mathbb{E}_\circ)X_j(x_\circ, v)^{2m}. \tag{24}$$

All constants and thresholds may depend on the fixed data, reference, q, m, d, β_1, and the chosen local root.

Proof. Write $T(p)$ for the square matrix whose first $r-1$ rows are $(e_r - e_i)^\top$ and last row is $p^\top$. It is invertible: a vector with all coordinate differences zero is constant, and its p-weighted mean then determines that constant. Exactly,

$$G_{D,q}(x) = T(p)(ZH_{D,q}(x) - \kappa\mathbf{1}).$$

At a root the product-rule term containing $D_x T$ vanishes, so $D_x G_{0,q}(x_\circ) = T(p_\circ)ZE_\circ$. Hurwitz stability implies invertibility of $ZE_\circ$. The root equation and $\kappa > 0$ exclude $a_\circ = 0$. The reference surface component gives positive $\nu_{q,j}$ for every coordinate on a sufficiently small closed neighborhood; indeed $\mathbb{E}_\circ X_j^2 \geq c_\circ \mathbb{E}_{\sigma_d} X_j^2 > 0$ when $a \neq 0$.

Apply the coordinatewise expansion of Appendix A.4 to the scalar batch input $G_{q,j}$. Its proof uses bounded inputs and derivatives, positive second moments and stabilizer, and matching moments through degree $2m-1$; all hold here by sphere support, fixed q and Lemma 3. It does not require uncontracted query moments once these input conditions hold. The sole unmatched leading block has coefficient $\delta^{m-1} a_j f_\lambda^{(m)}(\nu_{q,j})/(m\,m!)$. Equation (23) supplies its factor q^{1-2m}. The remainder is uniformly $O_{C^1}(\delta^m)$, and the same expansion gives $H_{D,q} - H_{0,q} = O_{C^1}(\delta)$.

For root existence and attraction, choose $W_\circ \succ 0$ satisfying $J_\circ^\top W_\circ + W_\circ J_\circ = -I$, where $J_\circ = D_x G_{0,q}(x_\circ)$. On a small convex $W_\circ$-ball, continuity and uniform C^1 comparison give every physical field a strictly negative derivative bound. The root map $x \mapsto x - (ZE_\circ)^{-1}(ZH_{D,q}(x) - \kappa\mathbf{1})$ has derivative norm uniformly below one and center displacement $O(\delta)$, so it maps the ball into itself for small δ. Contraction gives a unique root with displacement $O(\delta)$; the physical derivative bound gives attraction.

Apply (19). The averaged response derivative is $E_\circ + O(\delta)$ with uniformly invertible row Jacobian. Bounded moment derivatives permit replacing the reference root by $x_\circ$ at cost $O(\delta)$ in the coefficient. This proves (24), including mass and stabilizer motion. □

A.2.3 REALIZABILITY IN ANY NONZERO POPULATION DIRECTION

Apply Lemma 6 with $a = a_\circ \neq 0$ and the reference in (20). Every contracted constraint has degree at most $4m-2$, so each perturbation in its proof is legal for $\mathcal{C}_m^Z(D_\circ)$. The same family is legal for the smaller full query-moment class. Its positive definite Gram matrix realizes the required open ellipsoid in $\mathbb{R}^d$.

Proof of Theorem 1. The difficulty in the lower bound is that moment directions must come from legal laws, while a competing observer may choose arbitrary bounded readings. Restricting those readings to a nonnegative harmonic family makes them linear in its parameters; the response rank then determines both an indistinguishable pair and a matching constructive predictor.

Put $B = P_\circ D_{m,q}$ and $k_* = \operatorname{rank} B$. The open ellipsoid from Lemma 6 makes the realizable moment span $\mathbb{R}^d$. For readings $\mathcal{A}\theta$, if $\ker \mathcal{A} \subseteq \ker B$, the map $C(\mathcal{A}h) = Bh$ is well defined and extends linearly to the reading space. Thus $U_{\circ,q} + \delta^{m-1} C\mathcal{A}\theta(D)$ has uniform error $O(\delta^m)$. Conversely, if $\mathcal{A}h = 0$ but $Bh \neq 0$, the ellipsoid contains the legal moment vectors $\pm th$ for sufficiently small fixed $t > 0$. Their readings agree while their response gap is $2t\delta^{m-1}Bh + O(\delta^m)$, proving necessity. For arbitrary bounded readouts restricted to D_t, their shifts are $\mathcal{A}_{\mathrm{chart}} t$, where $(\mathcal{A}_{\mathrm{chart}})_{\ell j} = \int \varphi_\ell H_j \, d\sigma_d$, whereas the leading response is $B\Gamma t$. Since Γ is invertible, fewer than $\operatorname{rank} B$ readings leave a legal opposite pair with a nonzero order-δ^{m-1} gap.

Conversely, choose an orthonormal basis N_* for $\operatorname{range} B$; the k_* polynomial readings

$$\varphi(v) = N_*^\top B\big((X_j(x_\circ, v)^{2m})_j - \mathbb{E}_\circ(X_j(x_\circ, v)^{2m})_j\big)$$

give a public center $U_{\circ,q} + \delta^{m-1} N_* \mathbb{E}_D \varphi$ with the uniform remainder from (24).

The Stieltjes formula $f_\lambda(s) = \pi^{-1} \int_0^\infty \sqrt{t}\,[(\lambda^2 + t)(s + t)]^{-1}\, dt$ gives $(-1)^m f_\lambda^{(m)}(s) > 0$ for $s, \lambda > 0$. Hence coordinatewise nonzero $a_\circ$ makes $D_{m,q}$ invertible. $P_\circ$ is a projection of rank $d - r$, so $k_* = d - r$. In this case take $N_* = N$ to be an orthonormal basis of $\ker Z$. Repeat the harmonic construction at order $m + 1$ about the same $D_\circ$. These laws match all moments through degree $4m+3$, including every specified degree-$4m$ reading. Applying the expansion at order $m+1$ yields a nonzero order-δ^m legal pair in the central fiber, because $P_\circ D_{m+1,q} \Gamma_{m+1}$ still has rank $d - r$. Together with the aligned-center upper bound, this proves $\Theta(\delta^m)$ on the central fiber. The pair also belongs to the full query-moment class, which inherits these lower and upper bounds.

Finally $ZU = \kappa \mathbf{1}$ gives $\|U\| \geq \kappa\sqrt{r}/\|Z\|$. On the local affine response plane, normalization has the smooth inverse $u \mapsto \kappa u/(z_1^\top u)$, with denominator bounded away from zero. Both maps are locally Lipschitz, preserving the information orders and nonzero lower gaps. For $r = d$, $U = \kappa Z^{-1}\mathbf{1}$ exactly. □

A.2.4 AN EXPLICIT STABLE FAMILY

Corollary 5 (An explicit zero-initialized family for every fixed batch)**.** *Let D_{R} be uniform on $\{-1, 1\}^d$ and choose $D_\circ = \tau\sigma_d + (1 - \tau)D_{\mathrm{R}}$, $0 < \tau \leq 1$. For $0 < h < 1$, set*

$$z_i = \mathbf{1}_d + h(e_i - \bar{e}_r), \quad \bar{e}_r = r^{-1} \sum_{i=1}^r e_i, \quad v_q = \frac{d + q - 1}{q}, \quad \kappa_q = \frac{d}{1 + \sqrt{v_q}}.$$

At uniform masses and $\lambda = 1$, the population root is strictly stable for every fixed $q \geq 1$. For unit loss weights, $w_0 = 0$, $\epsilon_0 = L(0) = r$, and zero raw filters, all hypotheses of the initialized information theorem hold for sufficiently small fixed δ and then sufficiently small steps, uniformly over $\mathcal{C}_m^Z(D_\circ)$. The conclusion persists on an open data neighborhood for each fixed (d, r, m, q, τ, h).

Proof. Put $c_d = d/(d + 2)$, $s_q = (1 - \tau + \tau c_d)/q$ and $t_q = 1 - q^{-1} + 2\tau c_d/q$. Fourth-moment averaging gives $\nu_{q,j} = s_q\|a\|^2 + t_q a_j^2$ and $ds_q + t_q = v_q$. The rows are positive for $0 < h < 1$. Their differences span the $(r-1)$-dimensional zero-sum subspace V on the first r coordinates; their mean $\mathbf{1}_d$ is perpendicular to V, so their rank is r. At $a = \mathbf{1}_d$, $H_0 = \mathbf{1}_d/(1 + \sqrt{v_q})$, so every row has margin κ_q. On $\mathbf{1}_d^\perp$, the response derivative in a is scalar multiplication by

$$\alpha_{q,\tau} = \frac{1}{1 + \sqrt{v_q}} - \frac{t_q}{\sqrt{v_q}(1 + \sqrt{v_q})^2} > 0,$$

because $v_q - t_q = ds_q > 0$. In orthonormal mass tangent coordinates and $\ell = \log \lambda$, the full physical Jacobian is

$$\operatorname{diag}\left(-\frac{\alpha_{q,\tau} h^2}{r} I_{r-1}, -\frac{d}{(1+\sqrt{v_q})^2}\right).$$

Mass tangents are perpendicular to $\mathbf{1}_d$, while H_0 and its ℓ derivative are parallel to it. Thus both off-diagonal blocks and the product-rule term $(D_y a)^\top H_0$ vanish at this root. The positive independent rows and surface component $\tau\sigma_d$ complete the hypotheses of Lemma 4. Its common ball has initializer drift $O(\delta)$, so Lemma 25 applies with each batch as one bounded history input. Continuity preserves the root, derivative margins and nonzero aggregate coordinates on a small open data neighborhood, with the same initializer and fixed κ_q. □

The admissible-root and initialization conditions remain essential. Here q is fixed before the memory and step limits, and the information count concerns single-direction expectations. Growing batches, arbitrary batch-functional readings and finite-time training advantages require separate results.

A.3 Legal moment directions

The harmonic tools used here follow the spherical-polynomial framework of Dai & Xu (2013). The main proof uses an open moment ellipsoid realized by probability laws. The following harmonic construction supplies it for both the full query-moment and data-contracted classes.

A.3.1 A common harmonic construction for every nonzero direction

The following tool applies to both the full sphere class and the data-contracted class. Fix $d \geq 3$, $m \geq 2$ and $a \neq 0$.

Lemma 6 (Full moment freedom at a nonsymmetric root)**.** *For any $a \in \mathbb{R}^d \setminus \{0\}$, the highest degree-$4m$ harmonic parts of $p_j(v) = v_j^{2m}(a^\top v)^{2m}$, $j = 1, \ldots, d$, are linearly independent. Under* (20)*, legal perturbations of $D_\circ$ preserving all query moments through degree $4m-1$ realize an open ellipsoid of the moment vector $((\mathbb{E}_D - \mathbb{E}_\circ)p_j)_j$ about zero.*

Proof. Write $n = 4m$, $Q_{\mathrm{sph}}(v) = \sum_j v_j^2$, and $p_j(v) = v_j^{2m}(a^T v)^{2m}$. Define their top homogeneous harmonic projections explicitly by

$$H_j(v) = \sum_{k=0}^{\lfloor n/2 \rfloor} c_k Q_{\mathrm{sph}}(v)^k \Delta^k p_j(v), \quad c_0 = 1, \quad c_k = -\frac{c_{k-1}}{2k(2n-2k+d-2)}.$$

The identity $\Delta(Q_{\mathrm{sph}}^k f) = Q_{\mathrm{sph}}^k \Delta f + 2k(2n-2k+d-2)Q_{\mathrm{sph}}^{k-1} f$ for homogeneous f of degree $n-2k$ proves $\Delta H_j = 0$ and $p_j - H_j$ is divisible by Q_{sph}. Iteration gives the homogeneous harmonic decomposition into degrees $n, n-2, \ldots$ on a sphere. Restrictions of harmonic homogeneous polynomials of different degrees are orthogonal: in polar coordinates the spherical Laplacian eigenvalue is $-k(k+d-2)$ and integration by parts makes that operator self-adjoint. Decomposing each homogeneous part of an arbitrary lower degree polynomial therefore proves $H_j \perp \{\text{polynomials of degree} < n\}$.

The H_j are linearly independent. Otherwise Q_{sph} divides $(a^T v)^{2m} \sum_j b_j v_j^{2m}$. For $d \geq 3$, Q_{sph} is an irreducible real quadratic (a product of real linear forms cannot be positive definite), hence prime in the polynomial ring and coprime to $a^T v$. Thus Q_{sph} divides $\sum_j b_j v_j^{2m}$. Evaluation on complex zeros with coordinates $(1, i)$ at positions j, k gives $b_j + (-1)^m b_k = 0$. If m is even, three distinct positions force all coefficients to vanish. If m is odd they are equal; evaluation at $(1, 1, i\sqrt{2})$ gives $b(2 - 2^m) = 0$. Here odd $m \geq 3$, so again all vanish.

The Gram matrix $\Gamma_{ij} = \int H_i H_j \, d\sigma_d$ is therefore positive definite. Orthogonality gives $\int p_i H_j \, d\sigma_d = \Gamma_{ij}$, because $p_i - H_i$ restricts to lower harmonic degrees on the sphere.

Choose $\rho > 0$ such that $|\sum_j t_j H_j| \leq c_\circ/2$ whenever $\|t\| \leq \rho$, and set $D_t = D_\circ + (\sum_j t_j H_j)\sigma_d$. The surface component makes these measures nonnegative; harmonic orthogonality preserves total mass and all moments through degree $4m-1$. Their first unmatched moment shifts are exactly Γt. This is the required open ellipsoid. Any additional constraints given by polynomials of degree at most $4m-1$ are preserved as well. □

A.4 THE UNIFORM SHARED-HISTORY EXPANSION

The first unmatched input moment determines the leading field difference. The coordinatewise derivation retains the filters' shared history and applies in any fixed dimension, as used in Lemma 4. Put $D_m(x) = \mathrm{diag}_j[a_j(x) f^{(m)}_{\lambda(x)}(\nu_{\mathrm{ref},j}(x))/(m\,m!)]$, where $\nu_{\mathrm{ref},j} = \mathbb{E}_{\mathrm{ref}} X_j^2$. Inexact information allows noncentral low-moment errors, which we retain throughout.

A.4.1 COMMON HISTORY IDENTITIES AND REMAINDER CONTROL

Lemma 7 (Exact weighted-cumulant identities)**.** *Let X_k be IID bounded scalar inputs, put $Y_k = X_k^2 - \mathbb{E}X_k^2$, $\delta = 1 - \beta_2$, and define the stationary weighted filters*

$$W_\delta = \delta \sum_{k\ge 0} \beta_2^k Y_k, \qquad A = (1-\beta_1)\sum_{k\ge 0} \beta_1^k X_k.$$

For every integer $\ell \ge 2$ and every $0 \le \beta_1, \beta_2 < 1$,

$$\mathrm{cum}_\ell(W_\delta) = \frac{\delta^\ell}{1-\beta_2^\ell}\,\mathrm{cum}_\ell(Y), \tag{25}$$

$$\mathrm{cum}(A, \underbrace{W_\delta, \dots, W_\delta}_{\ell}) = \frac{(1-\beta_1)\delta^\ell}{1-\beta_1\beta_2^\ell}\,\mathrm{cum}(X, \underbrace{Y, \dots, Y}_{\ell}). \tag{26}$$

If two input laws match the moments needed to identify the cumulants through order $m-1$, then their first unmatched mth cumulant contribution to W_δ is

$$\frac{\delta^{m-1}}{m}\,\Delta\,\mathrm{cum}_m(Y) + O(\delta^m).$$

The identity is exact for the stationary filters; it is not by itself a finite-memory or initialized-recursion transfer statement.

Proof. Multilinearity and independence leave only terms with a common history index. Summing their weights gives $\delta^\ell \sum_{k\ge 0} \beta_2^{\ell k} = \delta^\ell/(1-\beta_2^\ell)$ and $(1-\beta_1)\delta^\ell \sum_{k\ge 0}(\beta_1\beta_2^\ell)^k = (1-\beta_1)\delta^\ell/(1-\beta_1\beta_2^\ell)$. Moment matching cancels the lower-order partitions, and $1-\beta_2^m = m\delta + O(\delta^2)$ gives the leading term. □

Marked partitions. Write $d_r = \delta^r/(1-\beta_2^r)$ and $t_r = (1-\beta_1)\delta^r/(1-\beta_1\beta_2^r)$. The mixed identity also holds for $r = 1$ by the same common-index sum. The moment–cumulant formula expresses $\mathbb{E}[AW^k]$ as a sum over partitions of one marked A slot and k unmarked W slots. An unmarked singleton vanishes; the marked singleton is $\mu = \mathbb{E}X$. An unmarked r-block contributes $d_r\,\mathrm{cum}_r(Y)$; a block containing the mark and r unmarked slots contributes $t_r\,\mathrm{cum}(X, Y, \dots, Y)$. The weights are $O_r(\delta^{r-1})$ and $O_r(\delta^r)$, uniformly for $0 \le \beta_1 \le \bar\beta_1 < 1$. These identities apply to any bounded IID input law, including continuous laws, independently of the kernel.

Centered moments and the Adam cutoff. Suppose inputs and their first state derivatives are uniformly bounded on a compact domain, the law is state-independent, and $\nu = \mathbb{E}X^2 \ge \nu_- > 0$, $\lambda \ge \lambda_- > 0$. Set $Y = X^2 - \nu$, $W = \delta\sum_k \beta_2^k Y_k$. For each fixed even integer r, the unmarked partitions have blocks of size at least two, so their total weight is $O(\delta^{r/2})$. The same argument applies to the centered ∂Y; interpolation gives $\mathbb{E}|W|^r + \mathbb{E}|\partial W|^r = O(\delta^{r/2})$ for every integer $r \ge 1$. Extend f_λ smoothly below a fixed positive fraction of ν_-. For Taylor degree N, choose the extension with bounded derivatives through order $N+2$ on the bounded range. Since $A, \partial A$ are bounded, the value and differentiated Taylor remainders are bounded by

$$C\{\mathbb{E}|W|^{N+1} + \mathbb{E}(|W|^N|\partial W|)\} = O(\delta^{(N+1)/2}).$$

This includes derivatives of the coefficients. Bounded-sum concentration gives $\Pr(V < \nu_-/2) \le C\exp(-c/\delta)$. The original and extended integrands agree above the cutoff, and $|\partial V| \le 2\sup|\partial X|\sqrt{V}$ controls the original derivative below it because $\lambda_- > 0$. An infinite IID history has $V > 0$ almost surely when $\nu > 0$. Dominated differentiation of the smooth extension therefore changes the response by only an exponentially small C^1 error. Constants are uniform on a compact β_1 interval; no joint limit $\beta_1 \uparrow 1$ is used.

A.4.2 THE LEADING COMPARISON WITH INEXACT MOMENTS

Lemma 8 (A full-sphere comparison with an extra integer order)**.** *On a sufficiently small fixed compact state neighborhood, uniformly over $D \in \mathcal{C}_{m,e}$ for sufficiently small $e, \delta > 0$,*

$$H_D(x) - H_{\mathrm{ref}}(x) = \delta^{m-1} D_m(x)\big(\mathbb{E}_D(X_j(x))^{2m} - \mathbb{E}_{\mathrm{ref}}(X_j(x))^{2m}\big)_{j=1}^{3} + O_{C^1_x}(e + \delta^m). \quad (27)$$

The moment vector here is taken on the full sphere; no cylinder projection is made when $e > 0$.

Proof. Expand the response at its true second moment, then group shared-history terms by their lag partitions. Bounded support and the compact state set control all required state derivatives. For a polynomial $p(v) = \sum_\alpha c_\alpha v^\alpha$ of degree at most $4m-1$,

$$|(\mathbb{E}_D - \mathbb{E}_{\mathrm{ref}})p| \le e \sum_{\alpha \neq 0} |c_\alpha| 3^{|\alpha|/2}.$$

Applied to the input polynomials and their derivatives, this gives $O_{C^1}(e)$ error for every input moment through degree $2m-1$. In particular, the true means μ_D and second moments ν_D have this error; small enough e gives common positive ν_- and λ_-. Work coordinatewise, with

$$A = (1-\beta_1)\sum_{k\ge 0} \beta_1^k X_k, \qquad W = \delta \sum_{k \ge 0} \beta_2^k Y_k, \qquad Y = X^2 - \nu_D.$$

The input and derivative bounds, state-independent laws and positive floors meet the common cutoff conditions above. Taking Taylor degree $N = 2m$ gives remainder $O_{C^1}(\delta^{m+1/2})$ for the original Adam kernel.

Apply the marked-partition formula in Appendix A.4.1 for $0 \le k \le 2m$, with the true mean μ_D and centered Y. Unmarked blocks with $r < m$ and marked blocks with $r < m$ use only input moments of degrees at most $2m-2$ and $2m-1$; all their values and first state derivatives differ by $O(e)$. Their products and the coefficients $f_\lambda^{(k)}(\nu_D)/k!$ have bounded derivatives, so they cost $O(e)$ without an inverse power of δ.

The only unknown block of order δ^{m-1} is an unmarked block of size m with a separate marked singleton, at Taylor index $k = m$. Its unknown raw moment has coefficient

$$\mu_D \frac{f_\lambda^{(m)}(\nu_D)}{m!} d_m = \delta^{m-1} \frac{a_j f_\lambda^{(m)}(\nu_{\mathrm{ref},j})}{m\, m!} + O_{C^1}(e\delta^{m-1} + \delta^m),$$

since $d_m = \delta^{m-1}/m + O_m(\delta^m)$. The lower raw moments in this cumulant are already matched to $O(e)$. Every other partition containing unknown moments contributes $O(\delta^m)$: an unmarked block of size $m+1$ does so alone; a size-m block plus any other nontrivial block gains an extra δ; and an unknown marked block has at least m unmarked slots. Uniform bounds on the finitely many cumulants, products and derivatives now give the stated C^1 remainder $O(e + \delta^m)$. □

Reusing the expansion. The calculation is coordinatewise. It uses IID history inputs, uniformly bounded inputs and state derivatives, positive second moments and stabilizers, and matching input moments through order $2m-1$ (including their first state derivatives). Thus it also applies to the bounded scalar inputs in the sphere and contracted classes used here. The leading coefficient uses the reference input mean, which is a_j in those applications; dimension changes the constants, not the lag weights or the remainder order.

A.5 THE FIRST TWO SHARED-FILTER COEFFICIENTS

Lemma 8 compares two moment-matched laws. Here we compute the individual first-order coefficient, including the shared numerator–denominator correction used by the seed. The same expansion gives the second-order residual coefficient for bounded inputs, with finite-support families as a special case.

Let $X(x)$ be a scalar input on a fixed probability space, with X and its first state derivatives uniformly bounded on a compact domain. Assume $\nu = \mathbb{E}X^2 \ge \nu_- > 0$ and $\lambda \in [\lambda_-, \lambda_+]$

with $\lambda_- > 0$. Use the shared IID-history filters A, V of Lemma 7, with $\beta_2 = 1 - \delta$ and $0 \le b = \beta_1 \le \bar{\beta}_1 < 1$. Write

$$\mu = \mathbb{E}X, \quad Y = X^2 - \nu, \quad r_k = \mathbb{E}Y^k, \quad s_k = \mathbb{E}[XY^k], \qquad f^{(k)} = f_\lambda^{(k)}(\nu).$$

Finite-support C^3 input families with state-independent probabilities satisfy these conditions as a special case.

Theorem 9 (The first two shared-filter coefficients)**.** *Uniformly on the stated domain and compact β_1 interval,*

$$H = \mathbb{E}[Af_\lambda(V)] = \mu f_\lambda(\nu) + \delta h + \delta^2 k + O_{C^1}(\delta^3), \tag{28}$$

where

$$h = f'_\lambda(\nu)\,\mathrm{Cov}(X, X^2) + \frac{\mu}{4} f''_\lambda(\nu)\,\mathrm{Var}(X^2), \tag{29}$$

$$k = -\frac{b}{1-b} f' s_1 + f'' \left(\frac{s_2}{2} - \frac{3\mu r_2}{8} \right) + f''' \left(\frac{\mu r_3}{18} + \frac{s_1 r_2}{4} \right) + \frac{\mu f^{(4)}}{32} r_2^2. \tag{30}$$

Here O_{C^1} controls the value and first derivatives in the state and $\log\lambda$. In particular the first-order expansion has remainder $O_{C^1}(\delta^2)$. No joint limit $\beta_1 \uparrow 1$ is asserted.

Proof. Put $\mu = \mathbb{E}X$ and $\nu = \mathbb{E}X^2$. The common marked-partition identities give exactly

$$\mathbb{E}[A(V - \nu)] = t_1\,\mathrm{Cov}(X, X^2), \tag{31}$$

$$\mathbb{E}[A(V - \nu)^2] = \mu d_2\,\mathrm{Var}(X^2) + t_2 \mathbb{E}[(X - \mu)(X^2 - \nu)^2]. \tag{32}$$

Uniformly for $\beta_1 \le \bar{\beta}_1 < 1$, $t_1 = \delta + O(\delta^2)$, $d_2 = \delta/2 + O(\delta^2)$ and $t_2 = O(\delta^2)$. Taylor expansion through degree two therefore gives (29).

Use the shared-history block weights d_s, t_s from Appendix A.4, with $b = \beta_1$ and $W = V - \nu$, and retain terms through order δ^2. The exact identities (31)–(32) become $\mathbb{E}[AW] = t_1 s_1$ and $\mathbb{E}[AW^2] = \mu d_2 r_2 + t_2(s_2 - \mu r_2)$ in the present notation. Grouping the remaining independent time indices gives

$$\mathbb{E}[AW^3] = \mu d_3 r_3 + 3 t_1 d_2 s_1 r_2 + O_{C^1}(\delta^3),$$
$$\mathbb{E}[AW^4] = 3\mu d_2^2 r_2^2 + O_{C^1}(\delta^3).$$

The omitted terms have order at least δ^3: in the first line this is the marked three-block, and in the second all partitions except two pure pairs with a separate mark. Also

$$d_2 = \tfrac{1}{2}\delta + \tfrac{1}{4}\delta^2 + O(\delta^3), \quad d_3 = \tfrac{1}{3}\delta^2 + O(\delta^3), \quad t_1 = \delta - \tfrac{b}{1-b}\delta^2 + O(\delta^3), \quad t_2 = \delta^2 + O(\delta^3).$$

Substitution gives the displayed coefficients. The cutoff and C^1 remainder argument is the bounded-input calculation of Appendix A.4, expanded here through degree six: $\mathbb{E}|W|^7 + \mathbb{E}(|W|^6|\partial W|) = O(\delta^{7/2})$. The fifth and sixth Taylor terms are $O_{C^1}(\delta^3)$ by the same block weights (pure singletons vanish). The positive second-moment and stabilizer floors control the cutoff uniformly, including histories with arbitrarily small V. The weights defining A and its state derivative are convex, so their bounds are independent of b. Moreover $1 - b\beta_2^j \ge 1 - \bar{\beta}_1$ and each fixed-order geometric derivative sum is uniformly bounded for $b \le \bar{\beta}_1$. Thus all marked-partition coefficients and the cutoff remainder are uniform on the stated compact memory interval.

□

If two coordinate-input laws match their first three raw moments, then their means, second moments, and covariance term in (29) cancel. Consequently

$$\Delta h_j = \frac{a_j}{4} f''_\lambda(\nu_{q,j}) \{\mathbb{E}_A G_{q,j}^4 - \mathbb{E}_B G_{q,j}^4\}. \tag{33}$$

A.6 THE PHYSICAL TWO-ROW JACOBIAN

Physical two-row specialization. For independent rows z_1, z_2, use $x = (s, \ell)$, $p = (1 + e^{-s})^{-1}$, $\lambda = e^{\ell}$, $\Delta z = z_1 - z_2$, and $a(p) = pz_1 + (1 - p)z_2$. The physical drift is $\mathcal{G}_D = (-\Delta z^\top H_D, a(p)^\top H_D - \kappa)^\top$. At its common interior limiting root x_0, put $L_0 = (-\Delta z, a(p_0))^\top$ and $H_x = D_x H_0(x_0)$. The extra product-rule term in the logit column is $p_0(1 - p_0)\Delta z^\top H_0(x_0) = 0$, so the physical Jacobian is $J_0 = L_0 H_x$. Equal margins also make their p-weighted average independent of p; hence the constant-row equation has precisely the physical drift's nearby zeros. If J_0 is invertible and $H_D = H_0 + \delta h_D + o_{C^1}(\delta)$, the root argument of Lemma 4 and the exact identity (19) give

$$x_A - x_B = -\delta J_0^{-1} L_0 \Delta h(x_0) + o(\delta), \qquad U_A - U_B = \delta P_0 \Delta h(x_0) + o(\delta).$$

Its projector has range $\ker Z$ and kernel $\operatorname{range} H_x$; thus precisely the forcing in $\operatorname{range} H_x$ cancels. The product-rule derivative is still retained away from the root.

In dimension three, if n spans $\ker Z$ and ϖ spans $\ker(H_x^\top)$, then

$$P_0 \Delta h = n \frac{\varpi^\top \Delta h}{\varpi^\top n}. \tag{34}$$

Invertibility ensures $\varpi^\top n \neq 0$. If $\kappa > 0$, then $Z H_0(x_0) = \kappa(1, 1)^\top$, so a nonzero vector in $\ker Z$ cannot be a scalar multiple of the common direction. The surviving component is therefore directional, not merely a speed change.

A.7 MATCHED QUERY LAWS AND THE FIRST MOMENT THEY LEAVE FREE

For the motivating example in Section 4.1, finite Fourier sums verify the matching exactly. We first state the arbitrary order construction, then identify the octagon pair and its leading forcing. The next subsection verifies that this forcing survives at a stable physical root.

A.7.1 FINITE-ORDER INFORMATION MATCHING BY ROTATED REGULAR POLYGONS

The octagon pair is the first member of a finite-order hierarchy. For an integer $N \geq 3$ and a phase ϕ, define

$$V_{N,\phi}^{(2)} = \sqrt{2}\big(\cos(\phi + 2\pi K/N), \sin(\phi + 2\pi K/N)\big), \qquad K \sim \mathrm{Unif}\{0, \ldots, N - 1\}. \tag{35}$$

Let R be an independent Rademacher variable and write $V_{N,\phi} = (V_{N,\phi}^{(2)}, R)$.

Lemma 10 (Finite-order Fourier matching)**.** *For every real polynomial P in two variables of total degree strictly less than N,*

$$\mathbb{E}P\big(V_{N,\phi}^{(2)}\big) \quad \textit{is independent of } \phi.$$

Consequently, for every three-variable monomial of total degree less than N, its expectation under $V_{N,\phi}$ is independent of ϕ. In particular, the two lifted octagon laws $V_{8,0}$ and $V_{8,\pi/8}$ have identical query moment tensors of orders $1, \ldots, 7$. For fixed a and an IID batch $G_q = q^{-1} \sum_{\ell=1}^{q} V_\ell V_\ell^\top a$, every joint moment of total degree k with $2k < N$ is phase-independent. At $N = 8$ this includes all moments through degree three and the full estimator covariance matrix.

Proof. Restrict a monomial $x^a y^b$ to the circle in (35) and write the resulting trigonometric polynomial in the basis $\{e^{i\ell\theta}\}$. Its frequencies satisfy $|\ell| \leq a + b < N$. For every integer ℓ,

$$\frac{1}{N} \sum_{k=0}^{N-1} e^{i\ell(\phi + 2\pi k/N)} = e^{i\ell\phi}\, \mathbf{1}_{\{N|\ell\}}.$$

Thus all nonconstant frequencies vanish and only the rotation-invariant constant coefficient remains. Multiplication by R^c contributes the common factor $\mathbb{E}R^c$, so the same conclusion holds for lifted monomials. Linearity proves the polynomial statement. A joint degree-k monomial of G_q expands into independent batch-index blocks, each a query polynomial of degree at most $2k < N$. Their expectations are therefore phase-independent. □

This matching concerns the zero-smoothing polynomial oracle. Positive-radius inputs contain $\sinh(\mu z^\top V)/\mu$ and need not match.

A.7.2 THE FIRST UNMATCHED MOMENT IN THE FINITE-ORDER HIERARCHY

Fix $m \geq 2$, put $N = 4m$ and write $(a_1, a_2) = r(\cos\psi, \sin\psi)$, and use the two fixed laws $D_A = \mathrm{Law}(V_{N,\psi/2})$ and $D_B = \mathrm{Law}(V_{N,\psi/2+\pi/N})$. These phases are chosen once from the reference a and remain fixed as the state evolves. If $r = 0$, choose any ψ; the displayed differences are then zero.

Lemma 11 (Exact first unmatched scalar input moment)**.** *Let $X_j = v_j v^\top a$ under the two laws above. Their moments through degree $2m-1$ agree, while*

$$\mathbb{E}_A X_1^{2m} - \mathbb{E}_B X_1^{2m} = 2^{2-2m} r^{2m}, \qquad \mathbb{E}_A X_2^{2m} - \mathbb{E}_B X_2^{2m} = (-1)^m 2^{2-2m} r^{2m},$$
$$\mathbb{E}_A X_3^{2m} - \mathbb{E}_B X_3^{2m} = 0. \tag{36}$$

For an IID batch average $G_{q,j} = q^{-1}\sum_{\ell=1}^{q} X_{\ell,j}$, all joint moments of total degree below $2m$ agree and

$$\mathbb{E}_A G_{q,j}^{2m} - \mathbb{E}_B G_{q,j}^{2m} = q^{-(2m-1)}\big(\mathbb{E}_A X_j^{2m} - \mathbb{E}_B X_j^{2m}\big). \tag{37}$$

Proof. The matching follows from Lemma 10. For arbitrary fixed phases ϕ_A, ϕ_B, put $c = a_1 - ia_2$ and $z = e^{i\theta}$. The highest planar terms of X_1, X_2 are $cz^2/2$ and $cz^2/(2i)$, respectively. Terms with an odd power of R vanish, while a positive even power reduces the planar degree below $4m$. Only the frequencies $\pm 4m$ in X_j^{2m} can differ, giving

$$\Delta\mathbb{E}X_1^{2m} = 2^{1-2m}\Re\{c^{2m}(e^{4mi\phi_A} - e^{4mi\phi_B})\}, \qquad \Delta\mathbb{E}X_2^{2m} = (-1)^m \Delta\mathbb{E}X_1^{2m}. \tag{38}$$

For $X_3 = a_3 + R(a_1 v_1 + a_2 v_2)$, sign averaging leaves planar degree at most $2m < 4m$, so its moment is common. At the specified phases, $c^{2m}(e^{4mi\phi_A} - e^{4mi\phi_B}) = 2r^{2m}$, proving (36). In the batch expansion, every partition except the all-equal one has blocks below order $2m$, whose moments are matched. The remaining weight is $q/q^{2m} = q^{-(2m-1)}$, proving (37). □

A.7.3 THE EQUAL-INFORMATION QUERY LAWS

The phases 0 and $\pi/8$ at $N = 8$ have the algebraic planar supports

$$\mathcal{D}_A^{(2)} = \frac{1}{2}\,\mathrm{Unif}\{\pm\sqrt{2}e_1, \pm\sqrt{2}e_2\} + \frac{1}{2}\,\mathrm{Unif}\{(\pm 1, \pm 1)\}, \tag{39}$$

where all sign pairs occur, and let $c_8 = \sqrt{1 + 2^{-1/2}}$ and $s_8 = \sqrt{1 - 2^{-1/2}}$,

$$\mathcal{D}_B^{(2)} = \mathrm{Unif}\{(\pm c_8, \pm s_8), (\pm s_8, \pm c_8)\}. \tag{40}$$

For $V_A^{(2)} \sim \mathcal{D}_A^{(2)}$ and $V_B^{(2)} \sim \mathcal{D}_B^{(2)}$, append an independent Rademacher coordinate R and let $\mathcal{D}_A = \mathrm{Law}(V_A^{(2)}, R)$ and $\mathcal{D}_B = \mathrm{Law}(V_B^{(2)}, R)$. Each is a sixteen-atom law.

Proposition 12 (Matched lifted information)**.** *Both laws are centered and isotropic, satisfy $\|v\|_2^2 = 3$ almost surely, and have identical aggregate raw MSE. Their coordinate contractions are*

$$T_1 = \mathrm{diag}(3/2, 1/2, 1), \qquad T_2 = \mathrm{diag}(1/2, 3/2, 1), \qquad T_3 = I_3.$$

For $a = (a_1, a_2, a_3)$ and $X_j = v_j v^\top a$, the joint input moments through degree three agree by the Fourier lemma, while

$$\mathbb{E}_A X_1^4 - \mathbb{E}_B X_1^4 = \mathbb{E}_A X_2^4 - \mathbb{E}_B X_2^4 = \frac{P(a_1, a_2)}{4}, \qquad \mathbb{E}_A X_3^4 - \mathbb{E}_B X_3^4 = 0, \tag{41}$$

where $P(x, y) = x^4 - 6x^2y^2 + y^4$. For an IID average of fixed size q, the same lower moments agree and every difference in (41) *is divided by q^3.*

Proof. The common circle moments $\mathbb{E}v_1^4 = \mathbb{E}v_2^4 = 3/2$, $\mathbb{E}v_1^2 v_2^2 = 1/2$ and $\mathbb{E}v_1^2 = \mathbb{E}v_2^2 = 1$, with vanishing odd moments and independent R, give T_j, centering and isotropy. Since $(vv^\top)^2 = 3vv^\top$, the IID batch raw MSE is $\mathbb{E}\|G_q - a\|^2 = 2\|a\|^2/q$ for either law. Equation (38) at $m = 2$, $\phi_A = 0$, $\phi_B = \pi/8$ gives $\Re\{(a_1 - ia_2)^4\}/4 = P(a_1, a_2)/4$; the batch factor follows from (37). □

Combining Proposition 12 with (33) yields

$$\Delta h(a, \lambda) = \frac{P(a_1, a_2)}{16q^3}\begin{pmatrix} a_1 f_\lambda''(\nu_{q,1}) \\ a_2 f_\lambda''(\nu_{q,2}) \\ 0 \end{pmatrix}. \tag{42}$$

A.8 THE EXACT TWO-ROW SEED: STABILITY AND NON-CANCELLATION

The sharp and finite-precision results use a common seed: a uniform angle on the radius-$\sqrt{2}$ circle, lifted by an independent sign. This circle–sign reference differs from uniform sphere measure.

Population response and active rows. The reference has the same fourth-moment contractions T_j as the lifted laws in Proposition 12. Consequently, for $\nu_j(a) = a^\top T_j a$, $H_{0,j}(a,\lambda) = a_j/(\sqrt{\nu_j(a)}+\lambda)$. Set

$$a_0 = (1,2,1), \quad \nu_0 = (9/2, 15/2, 6), \quad U_0 = \frac{a_0}{1+\sqrt{\nu_0}}, \quad h_\lambda = -\frac{a_0}{(1+\sqrt{\nu_0})^2},$$

$$b = \frac{U_0 \times h_\lambda}{4\|U_0 \times h_\lambda\|}, \qquad z_1 = a_0 + b, \quad z_2 = a_0 - b, \quad p_0 = 1/2, \quad \lambda_0 = 1, \quad \kappa = a_0^\top U_0. \tag{43}$$

The coordinate ratios $h_{\lambda,j}/U_{0,j}$ are unequal, so the cross product is nonzero. Since $\|b\| = 1/4$, both rows are strictly positive. They are independent because $b \neq 0$ and $b^\top U_0 = 0$ whereas $a_0^\top U_0 > 0$. Both have margin κ at U_0.

Put $M = D_a H_0(a_0, 1)$. Its full derivative, including the dependence of ν_j on every coordinate of a, is

$$M_{jk} = \frac{\mathbf{1}_{j=k}}{1+\sqrt{\nu_{0,j}}} - \frac{a_{0,j}(T_j a_0)_k}{\sqrt{\nu_{0,j}}(1+\sqrt{\nu_{0,j}})^2}. \tag{44}$$

In $x = (s, \ell) = (\operatorname{logit} p, \log\lambda)$ coordinates, $H_x = (Mb/2,\ h_\lambda)$. The complete slow Jacobian has

$$J_{0,11} = -b^\top M b, \qquad J_{0,12} = 0, \qquad J_{0,22} = a_0^\top h_\lambda < 0.$$

The product-rule term $(\partial_s a)^\top H_0$ vanishes at this root because the two active margins agree; it is retained off root. The rational bounds below show $b^\top M b > 0$. Thus J_0 is Hurwitz and ZH_x is invertible. The projector used throughout is $P_0 = I - H_x(ZH_x)^{-1}Z$, with range $\ker Z$ and kernel $\operatorname{range} H_x$.

Proposition 13 (All-order non-cancellation for the exact seed)**.** *For every $m \geq 2$, let*

$$b_m = \frac{(f^{(m)}(9/2),\ 2(-1)^m f^{(m)}(15/2),\ 0)^\top}{m\, m!}, \qquad f(x) = (1+\sqrt{x})^{-1}.$$

At the seed (43)*, $P_0 b_m \neq 0$ for every m.*

Proof. We reduce cancellation to a scalar ratio, then show that its sign or magnitude excludes the required value at every order. Write $g_m(x) = (-1)^m f^{(m)}(x) > 0$, as justified by the integral representation below, and $c_m = m\, m! b_m$. On the plane $c_3 = 0$, cancellation is equivalent to $c_{m,2}/c_{m,1} = q_*$, where $q_* = -\varpi_1/\varpi_2$ and $\varpi = (Mb) \times h_\lambda$. Indeed, $P_0 c = 0$ precisely when $c \in \operatorname{range} H_x$, or $\varpi^\top c = 0$; the signs below ensure the relevant denominators are nonzero.

For the rational check, write $r_j = \sqrt{\nu_{0,j}}$, $d_j = 1 + r_j$, $u = U_0$, $h = h_\lambda$ and $v = u \times h$; then $b = v/(4\|v\|)$ and M is the derivative displayed above. Squaring endpoints verifies

$$10^8 r_1 \in [212132034, 212132035], \qquad 10^8 r_2 \in [273861278, 273861279],$$
$$10^8 r_3 \in [244948974, 244948975].$$

Substituting these intervals into the displayed rational operations, with positive denominators and outward endpoints, gives

quantity	containing interval
$10^8 v_1$	$[-347682\,, -347681]$
$10^8 v_2$	$[-283082\,, -283081]$
$10^8 v_3$	$[906615\,, 906616]$
$10^8 v^\top M v$	$[2986\,, 2987]$
$10^8((Mv)\times h)_1$	$[43495\,, 43496]$
$10^8((Mv)\times h)_2$	$[-36779\,, -36778]$
$10^4\{-((Mv)\times h)_1/((Mv)\times h)_2\}$	$[11826\,, 11827]$
$10^4 f''(9/2)/f''(15/2)$	$[29543\,, 29545]$

Here $f''(r^2) = (1+3r)/(4r^3(1+r)^3)$. In particular $b^\top Mb = (v^\top Mv)/(16v^\top v) > 0$. Positive rescaling from $(Mv) \times h$ to ϖ gives

$$\varpi_1 > 0, \quad \varpi_2 < 0, \quad \frac{59}{50} < q_* < \frac{6}{5}, \qquad \frac{2}{3} < \frac{2g_2(15/2)}{g_2(9/2)} < \frac{17}{25}.$$

The Stieltjes representation

$$f(x) = \int_0^\infty \frac{\sqrt{s}}{\pi(1+s)} \frac{ds}{x+s}$$

gives $g_m(x) = m! \int_0^\infty \sqrt{s}/[\pi(1+s)(x+s)^{m+1}]\, ds$. For $y > x$, put $R(s) = (x+s)/(y+s)$ and $q(s) = (x+s)^{-1}$. Relative to the probability density proportional to $\sqrt{s}/[(1+s)(x+s)^{m+1}]$, passing to order $m+1$ reweights by the decreasing function q, while R^{m+2} is increasing. The covariance inequality, followed by $0 < R < 1$, yields

$$0 < \frac{g_{m+1}(y)}{g_{m+1}(x)} < \frac{g_m(y)}{g_m(x)}.$$

For odd m, $c_{m,2}/c_{m,1} < 0$ and cannot equal $q_* > 0$. For even m, that ratio is at most $2g_2(15/2)/g_2(9/2) < 1 < q_*$. Cancellation is therefore impossible for every $m \geq 2$. □

These strict inequalities give population stability and projected nonvanishing. The uniform comparison and initialization lemmas extend them to a common finite-memory root domain and initialized transfer.

A.8.1 The matched octagon pair at this seed

The general stability and non-cancellation proof above applies with $q = 1$ and $\beta_1 = 0.9$. For the matched octagons, the forcing is an explicit multiple of the non-cancelling order-two direction. Equation (42) becomes

$$\Delta h = \left(-\frac{7}{16} f_1''(9/2), -\frac{7}{8} f_1''(15/2), 0 \right)^\top.$$

Thus stability and survival reduce to

$$G_1 = b^\top Mb > 0, \qquad G_2 = ((Mb) \times h_\lambda)^\top \Delta h \neq 0. \tag{45}$$

Since $\Delta h = -(7/4)b_2$, Proposition 13 and the seed stability calculation above prove both conditions. With $n = a_0 \times b$ and $\varpi = (Mb) \times h_\lambda$, the coefficient

$$C_{\text{seed}} = n \frac{\varpi^\top \Delta h}{\varpi^\top n} \tag{46}$$

is approximately $(-7.65940, 4.58289, -1.50638) \times 10^{-4}$, has norm at least $9.05198897077934 \times 10^{-4}$, lies in $\ker Z$, and has nonzero derivative after unit normalization. These bounds quantify the stability and non-cancellation of this matched pair.

Application to every polygon order. Lemma 11 makes the rotated $4m$-gon forcing c_m a nonzero scalar multiple of b_m at this same seed, with coordinate ratio $(1, (-1)^m, 0)$. Thus Proposition 13 gives $P_0 c_m \neq 0$ for every $m \geq 2$. For this finite-support polynomial input family, the bounded inputs and state derivatives, matched moments and positive floors satisfy Lemma 8. Combining its C^1 expansion with the root identity (19) transfers this signal to the own-root response.

A.8.2 Angular response within the matched octagon family

The two laws in Figure 1(a) belong to a continuous rotation family with the same supplied estimator summaries. Its first unmatched angular moment determines the leading response and allows both extremal separation and first-order cancellation.

Let $\mathcal{C}_{\text{ang}}$ consist of $v = (\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, R)$ with an antipodal angular law, an independent Rademacher R, and $\mathbb{E}e^{2ik\theta} = 0$ for $k = 1, 2, 3$. This angular subclass of $\mathcal{C}_2$ has fixed query-direction covariance I_3 and matches all query moments through degree seven. It therefore fixes the

zero-radius estimator mean, MSE and full covariance at a common state. Write $m_8(D) = \mathbb{E}_D e^{8i\theta}$. The rotated octagon D_ϕ has angles $\phi + k\pi/4$, $0 \le k < 8$, so $m_8(D_\phi) = e^{8i\phi}$; the original pair is $D_A = D_0$, $D_B = D_{\pi/8}$.

Proposition 14 (The complete angular response interval)**.** *In the three-dimensional, two-row model, fix finite q, $\beta_1 < 1$, and a common interior Hurwitz population root for the circle–sign class, with positive second moments and stabilizer. At this same fixed-q root put*

$$\gamma = (a_1 - ia_2)^4, \quad r_a = (a_1^2 + a_2^2)^2, \quad c_q = \frac{1}{32q^3} P_0 \begin{pmatrix} a_1 f''_{\lambda_0}(\nu_{q,1}) \\ a_2 f''_{\lambda_0}(\nu_{q,2}) \\ 0 \end{pmatrix}.$$

Let the reference in $\mathcal{C}_{\mathrm{ang}}$ have $m_8 = 0$. On the common nearby root branches, uniformly over this class,

$$U_D(\delta) - U_{\mathrm{ref}}(\delta) = \delta c_q \Re\{\gamma m_8(D)\} + O(\delta^{3/2}). \tag{47}$$

The attainable leading coefficients and their diameter are

$$\{c_q \Re(\gamma m_8(D)) : D \in \mathcal{C}_{\mathrm{ang}}\} = \{t c_q : -r_a \le t \le r_a\},$$
$$\sup_{D,E \in \mathcal{C}_{\mathrm{ang}}} \|U_D - U_E\| = 2\delta r_a \|c_q\| + O(\delta^{3/2}).$$

For $r_a > 0$, set $\alpha = \arg(a_1 + ia_2)$. Then

$$U_\phi - U_{\mathrm{ref}} = \delta r_a c_q \cos(8\phi - 4\alpha) + O(\delta^{3/2}), \qquad \phi_+ = \alpha/2, \quad \phi_- = \alpha/2 + \pi/8$$

attain the two leading endpoints. If $r_a c_q = 0$, the first-order interval degenerates.

Proof. The Fourier matching argument in Appendix A.7 applies to every stipulated angular law. Rotations cover the unit circle in m_8; their mixtures fill its closed disk, and $|m_8| \le 1$ proves equality with the attainable set. The first unmatched scalar moment formula, or expansion of the fourth power, gives $\Delta\mathbb{E}X_j^4 = \Re(\gamma m_8)/8$ for $j = 1, 2$ and zero for $j = 3$. In an IID batch fourth moment, only the all-equal index partition can differ; its weight is $q/q^4 = q^{-3}$. Equation (33) therefore gives fixed-state forcing $(a_1 f''(\nu_{q,1}), a_2 f''(\nu_{q,2}), 0)^\top \Re(\gamma m_8)/(32q^3)$.

For fixed q, the common sphere bounds inputs and their state derivatives uniformly. Positive second-moment and stabilizer floors persist on one compact root neighborhood. The shared-history expansion in Theorem 9 is uniform on this state-independent class. The fixed inverse margin and Hurwitz root give common local root branches at distance $O(\delta)$ from the population root. Applying (19) to the forcing yields (47), retaining the $O(\delta^{3/2})$ remainder. The scalar linear map sends the disk exactly to $[-r_a, r_a]$; uniformity gives the diameter upper bound, and the two displayed rotations give its lower bound. Their phase choices also give the cosine rule. □

Extremal separation and cancellation at the original seed. For $a_0 = (1, 2, 1)$, $q = 1$, $\lambda_0 = 1$ and the data in (43),

$$\gamma = -7 + 24i, \quad r_a = 25, \quad C_{\mathrm{seed}} = -14c_1, \qquad C_{\mathrm{opt}} = 50c_1 = -\frac{25}{7} C_{\mathrm{seed}}.$$

The analytically extremal pair has a leading separation coefficient whose norm is $25/7$ times that of the original A/B pair. The angles $\phi_+ + \pi/16$ and $\phi_+ + 3\pi/16$ both have zero leading coefficient; Appendix D.3.3 measures the remaining finite-memory gap. Angles are understood modulo $\pi/4$. All coefficients use the same fixed-q reference root; changing q changes that root as well as the factor q^{-3}. Thus fixed standard summaries permit a continuous interval of adapted responses: root adjustment retains the scalar image of the first unmatched angular moment.

B PROOFS OF RESIDUAL SHARPNESS AND PRECISION

Theorem 20 uses a circle–sign class. The precision bounds in (64) allow laws on the full sphere. We first establish the exact-class witnesses needed by both proofs.

Table 1: Information supplied, remaining own-root response uncertainty, and the effect of additional readings. The rows describe distinct information classes under their stated local root conditions. Here $B = \mathsf{B}_{m,q}$, k is the number of additional readings, K is the highest supplied moment degree, and e, ξ are deterministic information tolerances.

Information class	Remaining uncertainty	Additional information and evidence
Contracted class $\mathcal{C}_m^Z(D^\circ)$	Leading ambiguity $\Theta(\delta^{m-1})$ when $B \neq 0$	$k_{\min} = \operatorname{rank} B$; under non-degeneracy and full realizability, $k_{\min} = d - r$. Aligned readings reduce the error to $O(\delta^m)$. Appendix D.2.1 gives fixed-instance spherical recovery evidence.
Separate 3D central fiber $\mathcal{F}_0$	$\frac{1}{2}(u_2 - \ell_2)\delta^2 + O(\delta^3)$	One further exact reading $\mathbb{E}_D\omega_2$ reduces the error to $O(\delta^3)$; the limiting extrema are attained.
Separate 3D moment box, $K = 4m - 1$	$\Theta_m(\delta^{m-1} + e)$	Increasing moment degree cannot remove the order-e precision floor. Proofs are in Appendix B.3; numerical bounds in Appendix D.4.
Central augmented fiber $\mathcal{F}_{m,e}(0, \xi)$	$\Theta_m\left(e + \delta^m + \delta^{m-1}\xi\right)$	One additional response-aligned reading with tolerance ξ; the base precision floor remains even at $\xi = 0$.

B.1 Exact-class support, moments and witnesses

Fix $m \geq 2$, $q = 1$, $\beta_1 = 0.9$ and the two-row seed of Appendix A.8. The reference has uniform angle and an independent sign in $v = (\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, R)$; define

$$\mathcal{C}_m = \{D : \|v\|^2 = 3 \ \ D\text{-a.s.}, \quad \mathbb{E}_D v^{\otimes k} = \mathbb{E}_{\mathrm{ref}} v^{\otimes k}, \ 1 \leq k \leq 4m - 1\}. \tag{48}$$

This circle–sign reference does not satisfy the positive uniform surface-component assumption of Theorem 1.

Lemma 15 (Uniform roots and their projected comparison)**.** *For all sufficiently small $\delta > 0$, the circle–sign class $\mathcal{C}_m$ has unique attracting roots in one common convex domain containing the specified initializer, with $x_D - x_0 = O(\delta)$ uniformly. For $\Delta M_D(x) = (\mathbb{E}_D - \mathbb{E}_{\mathrm{ref}})(X_j(x)^{2m})_{j=1}^3$,*

$$U_D = U_{\mathrm{ref}} + \delta^{m-1} P_0 D_m(x_0) \Delta M_D(x_0) + O_m(\delta^m). \tag{49}$$

For every finite preselected list and $\rho > 0$, one sufficiently small common step multiplier gives the initialized limits of the recursion (2)–(3) *jointly with probability at least $1 - \rho$.*

Proof. The seed in Appendix A.8 has a Hurwitz physical Jacobian, an invertible row Jacobian and positive masses, second moments and stabilizer. Sphere support uniformly bounds the inputs and derivatives; matched degree-four moments give the common population field. Appendix A.4 gives $H_D - H_0 = O_{C^1}(\delta)$ uniformly. The Lyapunov-ball and root-map argument of Lemma 4 therefore applies: its surface component supplied positive floors, which the explicit seed gives here. It yields a common attracting domain and $O(\delta)$ root displacements.

In (19) the averaged derivative is $E_0 + O(\delta)$, hence its projector is $P_0 + O(\delta)$. Lemma 8 at $e = 0$ and bounded coefficient derivatives give the asserted comparison at x_0. The ball also meets Lemma 25: its initializer drift is $O(\delta)$ and the derivative margins are uniform. Reduce δ so $\beta_1^2 < \beta_2$, then fix memory, finite list and failure budget before choosing the step threshold. □

Lemma 16 (Finite circle representation, including singular matrices)**.** *If $c_0 = 1$, $c_{-k} = \overline{c_k}$ and $T = (c_{k-j})_{0 \leq j,k \leq N} \succeq 0$, there is a probability measure on the unit circle with these moments and at most $\operatorname{rank} T \leq N + 1$ atoms.*

Proof. Use the classical finite Toeplitz construction (Yang & Xie, 2018, Theorem 1). Realize T as the Gram matrix of $e_0, \dots, e_N$ in a Hilbert space of dimension $\operatorname{rank} T$, with inner product conjugate-linear in its first entry. Toeplitz symmetry makes $e_j \mapsto e_{j+1}$ a well-defined isometry on the spans for $j < N$: shifted inner products preserve both norms and linear relations. Extend it to a unitary U using the equal-dimensional orthogonal complements. Then $e_k = U^k e_0$ for $k \le N$, and the spectral theorem gives $\langle e_0, U^k e_0\rangle = \sum_\ell w_\ell z_\ell^k = c_k$, where $|z_\ell| = 1$, $w_\ell = \|\Pi_\ell e_0\|^2 \ge 0$ and $\sum w_\ell = 1$. Discarding zero weights gives the atom count, also when T is singular. □

Theorem 17 (Support and the two free moment modes)**.** *Put $N = 4m$. A law belongs to $\mathcal{C}_m$ exactly when it is supported on the circle–sign set and satisfies*

$$\mathbb{E}e^{ik\theta} = 0 \ (1 \le |k| \le N-1), \qquad \mathbb{E}[Re^{ik\theta}] = 0 \ (|k| \le N-2). \tag{50}$$

Its remaining pair $\tau = \mathbb{E}e^{iN\theta}$, $\chi = \mathbb{E}[Re^{i(N-1)\theta}]$ ranges over $|\tau| + |\chi|^2 \le 1$.

Proof. Moment matching gives $\mathbb{E}(v_3^2 - 1)^2 = 0$, hence $v_3^2 = 1$ and $v_1^2 + v_2^2 = 2$. Writing $z = (v_1 + iv_2)/\sqrt{2}$ and reducing powers of R proves the equivalence with the Fourier constraints. In particular, both signs have probability $1/2$.

Set $\zeta = \mathbb{E}[Rz^N]$. Conditional on $R = \sigma$, the Gram matrix of $1, z, z^{N-1}, z^N$ is

$$\begin{pmatrix} I_2 & B_\sigma \\ B_\sigma^* & I_2 \end{pmatrix}, \qquad B_\sigma = \begin{pmatrix} \sigma\chi & \tau + \sigma\zeta \\ 0 & \sigma\chi \end{pmatrix}.$$

Its Schur complement gives $|\tau + \sigma\zeta| \le 1 - |\chi|^2$. Averaging the two bounds proves necessity. Conversely, prescribe conditional moments $c_0^\sigma = 1$, $c_1^\sigma = \cdots = c_{N-2}^\sigma = 0$, $c_{N-1}^\sigma = \sigma\chi$ and $c_N^\sigma = \tau$. Their Toeplitz matrices consist of the displayed block with $\zeta = 0$ and an identity block. The same Schur complement makes them positive semidefinite when $|\tau| + |\chi|^2 \le 1$. Lemma 16 realizes each by at most $N + 1$ circle atoms; mixing with equal sign weights proves sufficiency, including singular boundary cases. □

Theorem 18 (The first unmatched forcing and legal endpoints)**.** *For $c = a_1 - ia_2$, put*

$$Q_m(D; a) = 2^{1-2m}\Re\{c^{2m}\tau_D + 2m\sqrt{2}a_3 c^{2m-1}\chi_D\}. \tag{51}$$

The $2m$th input-moment difference from the reference is $Q_m(1, (-1)^m, 0)^\top$. At $a_0 = (1, 2, 1)$ its scalar range is $[-Q_m^*, Q_m^*]$, where

$$Q_m^* = 2^{1-2m}2m\sqrt{2}\,5^{m-1/2}. \tag{52}$$

The signed endpoints $D_m^\pm$ are uniform on

$$\begin{gathered} \ell = 4m - 1, \quad \phi_m = \frac{(2m-1)\arg(1+2i)}{\ell}, \\ v_k^\pm = (\sqrt{2}\cos(\phi_m + k\pi/\ell), \sqrt{2}\sin(\phi_m + k\pi/\ell), \ \pm(-1)^k), \quad 0 \le k < 2\ell. \end{gathered} \tag{53}$$

Proof. The highest planar and sign terms of X_1 are $cz^2/2$ and $a_3 Rz/\sqrt{2}$. In X_1^{2m} only frequencies $4m$ and $4m-1$ remain unmatched; summing conjugates gives (51). For X_2 their coefficients acquire $(-i)^{2m} = (-1)^m$. Every reduced term of X_3^{2m} has degree at most $2m + 1 \le 4m - 1$.

Aligning the two free phases reduces maximization of $|Q_m|$ to $2^{1-2m}\max_{0\le u\le 1}[A(1-u^2) + Bu]$, with $A = |c|^{2m}$ and $B = 2m\sqrt{2}|a_3||c|^{2m-1}$. At the seed $B/(2A) = m\sqrt{2/5} > 1$, so the maximum is $2^{1-2m}B$. For the stated atoms, ordinary Fourier sums vanish unless the frequency is a multiple of 2ℓ; sign-weighted sums vanish unless it is an odd multiple of ℓ. They satisfy (50), with $\tau = 0$ and $\chi = \pm e^{i\ell\phi_m}$. The chosen phase gives $c^{2m-1}e^{i\ell\phi_m} = 5^{m-1/2}$, proving both endpoint claims. □

Theorem 19 (Exact-class response width)**.** *Let b_m be from Proposition 13 and set*

$$w_m = P_0 b_m, \qquad n_m = w_m/\|w_m\|, \qquad C_m^* = Q_m^*\|w_m\| > 0. \tag{54}$$

Uniformly on $\mathcal{C}_m$,

$$U_D - U_{\mathrm{ref}} = \delta^{m-1}Q_m(D; a_0)w_m + O_m(\delta^m), \qquad R(\mathcal{C}_m) = C_m^*\delta^{m-1} + O_m(\delta^m). \tag{55}$$

Proof. The moment identity and Lemma 15 give the expansion; Proposition 13 gives $w_m \ne 0$. The attained opposite endpoints supply its matching half-gap lower bound and the reference supplies the upper bound. More precisely, all responses lie on the affine line $ZU = \kappa\mathbf{1}$, whose enclosing radius equals half its diameter. These endpoint laws will supply the higher-order lower bounds below. □

B.2 THE FIRST RESIDUAL ORDER ON THE ENTIRE EXACT CENTRAL FIBER

Return to the original two-row seed, $q = 1$, $m = 2$ and circle–sign reference. Its exact fiber $\mathcal{F}_0 = \mathcal{F}_{2,0}(0,0) = \{D \in \mathcal{C}_2 : \mathbb{E}_D\psi_2 = 0\}$ uses the normalized polynomial (65) at $m = 2$ on this circle–sign support. This class differs from the contracted class of Appendix A.2. The order-3 pair already gives a positive δ^2 lower bound; we determine the sharp residual coefficient by finite moment optimization.

B.2.1 A SECOND-ORDER EXPANSION RETAINING THE SHARED HISTORY

Apply the second-order expansion of Theorem 9 coordinatewise, using (29) and (30).

Second-order root elimination. For a uniform expansion $H_D = H_0 + \delta h_D + \delta^2 k_D + O_{C^1}(\delta^3)$ on the common root domain, assume H_0 is C^3, h_D uniformly C^2 and k_D uniformly C^1. Put $E_0 = D_x H_0(x_0)$, $J = ZE_0$, with J invertible and $ZH_0(x_0) = \kappa\mathbf{1}$, and set

$$y_D = -J^{-1}Zh_D(x_0), \qquad P_0 = I - E_0J^{-1}Z.$$

All following coefficients are evaluated at x_0. The trial root $x_0 + \delta y_D + \delta^2 z_D$, with $z_D = -J^{-1}Z[k_D + D_x h_D y_D + \frac{1}{2}D_x^2 H_0[y_D, y_D]]$, has row residual $O(\delta^3)$. Uniform invertibility of the averaged row derivative gives the same error in the already established root. Expanding its response therefore gives

$$U_D = U_0 + \delta P_0 h_D(x_0) + \delta^2 P_0\left[k_D(x_0) + D_x h_D(x_0) y_D + \tfrac{1}{2}D_x^2 H_0(x_0)[y_D, y_D]\right] + O(\delta^3). \quad (56)$$

The coefficient derivative bounds make this remainder uniform.

B.2.2 A LINEAR MOMENT FUNCTIONAL ON THE EXACT FIBER

At the public seed ($\lambda_0 = 1$), evaluate h_{ref} using (29) and define

$$y = -(ZD_xH_0(x_0))^{-1}Zh_{\mathrm{ref}}(x_0), \qquad a_y = D_x a(x_0)y. \quad (57)$$

The coefficients use only public quantities. Reference moments satisfy

$$\mathbb{E}_{\mathrm{ref}} v_1^{2a} v_2^{2b} v_3^{2c} = \frac{(2a)!(2b)!}{2^{a+b}a!b!(a+b)!}, \qquad a, b, c \in \mathbb{N}_0,$$

with any monomial having an odd coordinate exponent equal to zero. Extracting the constant circle Fourier coefficient and averaging the independent sign gives this formula. Put $X_j^0(v) = v_j v^\top a_0$, $Y_j^0(v) = v_j v^\top a_y$, and define the degree-at-most-12 polynomial

$$g_j(v) = \frac{f''(\nu_{0,j})}{2}(X_j^0)^5 + \frac{a_{0,j}f'''(\nu_{0,j})}{18}(X_j^0)^6 + a_{0,j}f''(\nu_{0,j})(X_j^0)^3 Y_j^0, \qquad f(s) = (1+\sqrt{s})^{-1}. \quad (58)$$

Choose a public unit vector n spanning $\ker Z$, for example $n = P_0 b_2/\|P_0 b_2\|$, and set

$$\omega_2(v) = n^\top P_0\big(g(v) - \mathbb{E}_{\mathrm{ref}} g\big). \quad (59)$$

B.2.3 EXACT REDUCTION TO A FINITE TOEPLITZ PROBLEM

For $\sigma \in \{-1, 1\}$ introduce complex numbers $c_0^\sigma, \dots, c_{12}^\sigma$ and define $c_{-k}^\sigma = \overline{c_k^\sigma}$ and $T_{12}(c^\sigma) = (c_{k-j}^\sigma)_{0\le j,k\le 12}$. For any polynomial g used below, substitute $v = (\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \sigma)$ and expand $g_\sigma(\theta) = \sum_{k=-12}^{12} g_{\sigma,k} e^{ik\theta}$. Define the real linear expectation functional

$$\mathcal{L}_c(g) = \frac{1}{2}\sum_{\sigma=\pm 1}\sum_{k=-12}^{12} g_{\sigma,k} c_k^\sigma.$$

The finite feasible set $\mathcal{T}_0$ is

$$\begin{gathered} T_{12}(c^+) \succeq 0, \quad T_{12}(c^-) \succeq 0, \quad c_0^+ = c_0^- = 1, \\ c_k^+ = c_k^- = 0 \quad (1 \le k \le 6), \qquad c_7^+ + c_7^- = 0, \qquad \mathcal{L}_c(\psi_2) = 0. \end{gathered} \quad (60)$$

The difference $c_8^+ - c_8^-$ and higher moments remain free variables. Set

$$\ell_2 = \min_{c \in \mathcal{T}_0} \mathcal{L}_c(\omega_2), \qquad u_2 = \max_{c \in \mathcal{T}_0} \mathcal{L}_c(\omega_2), \qquad C_{2,\mathrm{res}} = (u_2 - \ell_2)/2. \tag{61}$$

Both programs have two Hermitian 13×13 Toeplitz blocks (equivalently, real representations) and coefficients determined by the public seed.

Theorem 20 (Sharp residual on the original central fiber)**.** *Uniformly for every $D \in \mathcal{F}_0$,*

$$U_D - U_{\mathrm{ref}}(\delta) = \delta^2 n \mathbb{E}_D \omega_2 + O(\delta^3). \tag{62}$$

No order-3 moment-class membership is required. In particular, on this original fiber one further exact expectation $\mathbb{E}_D\omega_2$ gives a public center with error $O(\delta^3)$. A reported half-width ζ instead gives error $O(\delta^3) + \delta^2\zeta$. On the doubly central fiber $\{D \in \mathcal{F}_0 : \mathbb{E}_D\omega_2 = 0\}$, the information radius is $\Theta(\delta^3)$.

The set $\mathcal{T}_0$ is exactly the set of conditional moments through degree 12 of laws in $\mathcal{F}_0$. Its two extrema are attained by laws in $\mathcal{F}_0$ with at most 26 atoms each. Moreover,

$$R(\mathcal{F}_0) = C_{2,\mathrm{res}}\delta^2 + O(\delta^3), \qquad C_{2,\mathrm{res}} > 0. \tag{63}$$

The public center $U_{\mathrm{ref}}(\delta) + \frac{1}{2}\delta^2(\ell_2 + u_2)n$ has worst error $C_{2,\mathrm{res}}\delta^2 + O(\delta^3)$ and is asymptotically minimax. No symmetry $u_2 = -\ell_2$ is presumed.

Proof. Every $D \in \mathcal{C}_2$ has the same input moments through degree three at every state. The exact cylinder identity gives

$$(\mathbb{E}_D - \mathbb{E}_{\mathrm{ref}})(X_j(x,v)^4)_j = Q_2(D; a(x))(1,1,0)^\top.$$

The fiber constraint $\mathbb{E}_D\psi_2 = 0$ is $Q_2(D; a_0) = 0$, so fourth input moments also agree at x_0. In (29), μ, ν, s_1 agree as state functions and r_2 agrees at x_0. Hence $h_D(x_0) = h_{\mathrm{ref}}(x_0)$, giving every law the same first root displacement y.

Subtract (30) at x_0. Since $s_2 = \mathbb{E}X^5 - 2\nu\mathbb{E}X^3 + \nu^2\mu$ and $r_3 = \mathbb{E}X^6 - 3\nu\mathbb{E}X^4 + 2\nu^3$, only the fifth and sixth moments remain, giving the first two terms of g_j in (58). Differentiating $h_D - h_{\mathrm{ref}}$ along the common y gives its third term: the prefactor derivative vanishes because fourth moments agree at x_0, while differentiating X^4 supplies $4X^3Y$. The quadratic root term in (56) is common and cancels. Hence the response difference is $\delta^2 P_0(\mathbb{E}_D - \mathbb{E}_{\mathrm{ref}})g + O(\delta^3)$. Since $\operatorname{range} P_0 = \operatorname{span}\{n\}$, this is (62), uniformly by compact support and the common root domain.

Centering at $U_{\mathrm{ref}}(\delta) + \delta^2 n\widehat{\omega}_2$ gives the exact- and finite-width reading bounds. For sharpness on the doubly central fiber, the order-4 class matches query moments through degree fifteen, including ψ_2 and ω_2. Its correlated endpoints have a nonzero order-δ^3 gap by Theorem 19 at sufficiently small memory parameter. Combined with the upper bound, this proves the radius order.

Cylinder support and $\mathbb{E}R = 0$ give sign probabilities $1/2$. Theorem 17 identifies its Fourier constraints with the low-mode equations in (60); the last equation is precisely the extra reading. Thus every fiber law is feasible. Conversely, Lemma 16 realizes each positive semidefinite block on at most 13 circle atoms. Mixing them with weights $1/2$ and applying the same Fourier equivalence gives a law in $\mathcal{F}_0$ that realizes every degree-twelve objective. This proves exactness and the 26-atom bound, also for singular blocks.

The feasible set is nonempty (the reference is feasible), closed and bounded since principal minors give $|c_k^\sigma| \le 1$. Its extrema are attained. The response expansion just proved gives response endpoints $n^\top U_{\mathrm{ref}} + \delta^2\ell_2 + O(\delta^3)$ and $n^\top U_{\mathrm{ref}} + \delta^2 u_2 + O(\delta^3)$. On the affine response line, the midpoint and half-range give the center and radius exactly. Finally the order-3 class lies in $\mathcal{F}_0$ and its correlated endpoints have a nonzero δ^2 gap by Theorem 19, so $u_2 > \ell_2$. □

B.3 Inexact moments and an additional expectation

B.3.1 Resolution with inexact moment information

Fix $m \ge 2$, $K = 4m - 1$, the exact seed, $q = 1$, and $\beta_1 = 0.9$. For $\phi_\alpha(v) = (v/\sqrt{3})^\alpha$ define

$$\mathcal{C}_{m,e} = \{D : \|v\|^2 = 3 \text{ a.s.}, \quad |\mathbb{E}_D\phi_\alpha - \mathbb{E}_{\mathrm{ref}}\phi_\alpha| \le e,\ 1 \le |\alpha| \le K\}.$$

The reference belongs to this class for every $e \geq 0$. The query-direction means and covariances of its other members are not assumed exact. Constants in this subsection may depend on the fixed m, seed and chosen compact state neighborhood, but never on D, e, δ in the stated domain.

For the central augmented fiber $\mathcal{F}_{m,e}(0,\xi) = \{D \in \mathcal{C}_{m,e} : |\mathbb{E}_D \psi_m| \leq \xi\}$, with ψ_m defined in (65), independently sufficiently small capped tolerances and δ give

$$R(\mathcal{C}_{m,e}) = \Theta_m(\delta^{m-1} + e), \qquad R(\mathcal{F}_{m,e}(0,\xi)) = \Theta_m(e + \delta^m + \delta^{m-1}\xi). \tag{64}$$

The polynomial estimate in Lemma 8 also gives $H_D - H_0 = O_{C^1}(e+\delta)$: the true means and second moments differ from their reference values by $O(e)$. The Lyapunov-ball argument of Lemma 15 therefore gives common attracting roots for independently small e, δ, with $x_D - x_0 = O(e + \delta)$. This comparison takes place on the full sphere.

Lemma 21 (A non-neutral precision direction)**.** *Let $D_t = (R_t)_{\#} D_{\mathrm{ref}}$ where R_t rotates coordinates $1, 3$ with infinitesimal velocity $(v_3, 0, -v_1)$. Then $D_{\pm e/K} \in \mathcal{C}_{m,e}$. Their query-direction means are zero and covariances are I, and sphere radii are unchanged. At the seed,*

$$\nu_0' = a_1 a_3 (3, 1, -4)^T, \qquad h_{\mathrm{rot,j}} = a_j f_1'(\nu_j)\nu_{0,j}', \qquad g_{\mathrm{rot}} = P_0 h_{\mathrm{rot}} \neq 0.$$

Uniformly for small t, δ the own-root responses satisfy

$$U_{D_t} - U_{D_{-t}} = 2t\{g_{\mathrm{rot}} + O(\delta + |t|)\}.$$

Proof. A normalized degree-k monomial has rotation derivative at most k: each differentiated factor and coordinate velocity is bounded by one. Integration from zero to t proves membership in $\mathcal{C}_{m,e}$. Orthogonality preserves the query-direction means and covariance. To compute ν_0', differentiate $v_j^2 (v^T a)^2$ and use $\mathbb{E}v_1^4 = \mathbb{E}v_2^4 = 3/2$, $\mathbb{E}v_3^4 = 1$, $\mathbb{E}v_1^2 v_2^2 = 1/2$, $\mathbb{E}v_1^2 v_3^2 = \mathbb{E}v_2^2 v_3^2 = 1$, and zero odd moments. This gives $3a_1 a_3, a_1 a_3, -4a_1 a_3$; the mean $\mathbb{E}X = a$ stays fixed.

Use the seed notation and rational radical enclosures of Appendix A.8: $r_j = \sqrt{\nu_j}$, $d_j = 1 + r_j$, $h_j = -a_j/d_j^2$, $v = U_0 \times h$, and $M = D_a H_0$. The kernel of P_0 is spanned by Mv, h. Substituting the same outward radical intervals and $h_{\mathrm{rot},j} = -a_j(3,1,-4)_j/(2r_j d_j^2)$ gives the distinct check

$$10^{11}((Mv) \times h)^T h_{\mathrm{rot}} \in [-1544280, -1544278].$$

Its positive denominators and nonzero interval prove projection survival.

Realize the rotations on the same reference probability space. Their inputs and mixed state/rotation derivatives are uniformly bounded, so the order-two cutoff argument gives $H_t = \mu_t f_\lambda(\nu_t) + O_{C^1_{x,t}}(\delta)$. This also applies to the continuous circle law: it uses boundedness and positive second moments, independently of support size. The uniform root inverse yields $\partial_t U_{D_t} = g_{\mathrm{rot}} + O(\delta + |t|)$. Integrating from $-t$ to t gives the stated bound, with its factor t even when t is much smaller than a power of δ. □

Proof of the first bound in (64). The uniformly bounded degree-$4m$ moments and the shared-history comparison give $H_D - H_{\mathrm{ref}} = O_{C^1}(e + \delta^{m-1})$. The exact segment identity (19), with its common inverse margin, gives the same upper order for $U_D - U_{\mathrm{ref}}$. The exact endpoint laws $D_m^{\pm}$ belong to every enlarged class, so Theorem 19 supplies a half-gap of order δ^{m-1}. Lemma 21 supplies a separate half-gap of order e for small δ, e. Taking the larger lower bound gives at least a fixed multiple of their sum. Hence $R(\mathcal{C}_{m,e}) = \Theta_m(\delta^{m-1} + e)$ uniformly in their ratio. □

Appendix B.3.2 gives the upper bound for every nonempty reported fiber. Appendix B.3.3 proves both exact and noisy central-fiber sharpness, including a precision witness with zero extra expectation.

B.3.2 One selector-aware expectation and a higher resolution order

Keep the seed of Appendix A.8, with $q = 1$, $\beta_1 = 0.9$ and fixed $m \geq 2$; constants may depend on m and the seed. Use the population root x_0, aggregate a_0, projector P_0 and constants of (54).

Put $D_m^0 = D_m(x_0)$, where D_m is the coefficient matrix in Appendix A.4. For any point of the full sphere, define $X_j^0(v) = v_j v^T a_0$ and $M_{2m,j}^{\mathrm{ref}} = \mathbb{E}_{\mathrm{ref}}(X_j^0)^{2m}$. The new public polynomial is

$$\psi_m(v) = \frac{n_m^T P_0 D_m^0 \big(((X_j^0(v))^{2m})_{j=1}^3 - M_{2m}^{\mathrm{ref}}\big)}{C_m^*}. \tag{65}$$

This public degree-$4m$ polynomial adds one expectation to the original degree-$4m-1$ information. Its coefficients use the population seed, not a target root; this reading uses the exact-expectation information model of Section 3.

Expectation identity and a pointwise bound. For $D \in \mathcal{C}_m$, Theorem 18 gives $\Delta M_{2m} = (1,(-1)^m,0)^T Q_m(D;a_0)$ and $D_m^0(1,(-1)^m,0)^T = b_m$. Consequently

$$\mathbb{E}_D \psi_m = Q_m(D;a_0)/Q_m^* \in [-1,1]. \tag{66}$$

This is an expectation bound on the exact class, not a pointwise bound. On $\|v\|^2 = 3$, $|X_j^0(v)| \le B_X := 3\|a_0\|_2$, so the explicit pointwise bound

$$|\psi_m(v)| \le B_{\psi,m} := \frac{2B_X^{2m}\|n_m^T P_0 D_m^0\|_1}{C_m^*} \tag{67}$$

holds on the entire sphere. This conservative constant can be large; descriptor dimension alone implies no statistical sample saving.

The augmented information contract. For $e, \xi \ge 0$ and a reported real center $\widehat{q}$, set

$$\mathcal{F}_{m,e}(\widehat{q},\xi) = \{D \in \mathcal{C}_{m,e} : |\mathbb{E}_D \psi_m - \widehat{q}| \le \xi\}.$$

We state upper bounds only for nonempty fibers. At $\widehat{q} = 0$ the public reference is a witness for every $e, \xi \ge 0$. Arbitrary centers need a feasible witness or an explicit infeasible/unknown status. Let $U_{\mathrm{ref}}(\delta)$ be the exact reference own-root response and put

$$\widetilde{U} = U_{\mathrm{ref}}(\delta) + \delta^{m-1} C_m^* \widehat{q}\, n_m. \tag{68}$$

Minimax centers are allowed in $\mathbb{R}^3$, so $\widetilde{U}$ need not be the response of a law in the fiber. It satisfies $Z\widetilde{U} = \kappa \mathbf{1}$.

Theorem 22 (One expectation improves the resolution order)**.** *For fixed m and sufficiently small $e, \delta > 0$, all nonempty fibers satisfy*

$$\sup_{D \in \mathcal{F}_{m,e}(\widehat{q},\xi)} \|U_D - \widetilde{U}\| \le C_m\{e + \delta^m + \delta^{m-1}\xi\}. \tag{69}$$

Normalization gives the same upper order on the common nonzero response domain. A finite preselected list has the initialized consequence of Lemma 25. The exact and noisy central radii are established together in Theorem 23.

Proof. The population comparison gives common attracting roots with $x_D - x_0 = O(e+\delta)$ and $x_{\mathrm{ref}} - x_0 = O(\delta)$. Apply the exact integrated root identity (18) at the exact reference root. Its row residual is zero, and its averaged projector is $P_0 + O(e+\delta)$ by the full C^1 comparison and the fixed inverse margin of ZE_0. Lemma 8 therefore gives

$$U_D - U_{\mathrm{ref}} = \delta^{m-1} P_0 D_m^0 \Delta M_{2m}(x_0) + O(e+\delta^m).$$

The evaluation shift costs $O(\delta^m)$; changing the projector costs $O((e+\delta)(e+\delta^{m-1})) = O(e+\delta^m)$. Since every vector in $\operatorname{range} P_0$ is a multiple of the unit vector n_m,

$$U_D - U_{\mathrm{ref}} = \delta^{m-1} C_m^* (\mathbb{E}_D \psi_m) n_m + O(e+\delta^m). \tag{70}$$

This holds also for laws off the cylinder and with nonexact query-direction covariance. Subtracting (68) proves the upper bound.

All responses and the proposed center satisfy $ZU = \kappa\mathbf{1}$, so their norms are at least $u_- := \kappa\sqrt{2}/\|Z\|_{\mathrm{op}} > 0$ and normalization is $2/u_-$-Lipschitz. Its derivative is injective on $\ker Z$, since a nonzero kernel vector cannot be parallel to the population response. Continuity also gives a local inverse Lipschitz bound on the affine response line, which transfers the witness gaps used below. The common domain and bounded inputs give finite-list transfer by Lemma 25, with its stated order of choices. □

Numerical precision of the center. An implemented descriptor with pointwise error ϵ_ψ enlarges the information width to $\xi + \epsilon_\psi$; coefficient intervals give $\epsilon_\psi \le \sum_\alpha \epsilon_\alpha 3^{|\alpha|/2}$. If the reference response, $k_m = C_m^* n_m$ and reported center have errors $\epsilon_{\mathrm{ref}}, \epsilon_k, \epsilon_q$, respectively, a numerical center adds at most

$$\epsilon_{\mathrm{ref}} + \delta^{m-1}\{|\widehat{q}|\epsilon_k + (\|k_m\| + \epsilon_k)\epsilon_q + C_m^*\epsilon_\psi\} + \epsilon_{\mathrm{row}}.$$

A row correction costs at most $\epsilon_{\mathrm{row}} = \|Z^\dagger\|\|Z\widehat{U} - \kappa\mathbf{1}\|$; the uncorrected center instead needs its full error bound. Thus $e = O(\delta^m)$ and $\xi = O(\delta)$ preserve the improved order when this numerical allowance is also $O(\delta^m)$.

B.3.3 The precision floor with an extra expectation

Retain the seed, reference, normalized moment constraints and degree-$4m$ polynomial ψ_m of Appendix B.3.2. We now set the reported center to zero in the same fiber $\mathcal{F}_{m,e}(\widehat{q}, \xi)$. The reference makes this fiber nonempty. The upper bound, normalization and finite-list initialization implications are those proved there; the new step is a precision witness whose extra reading stays exactly zero.

Theorem 23 (Complete central noisy-fiber resolution)**.** *For fixed $m \ge 2$ and any fixed $0 < \xi_0 \le 1$, uniformly for independently sufficiently small $e \ge 0$, $\delta > 0$ and $0 \le \xi \le \xi_0$,*

$$R(\mathcal{F}_{m,e}(0, \xi)) = \Theta_m(e + \delta^m + \delta^{m-1}\xi).$$

In particular, at $e = 0$,

$$R(\mathcal{F}_{m,0}(0, \xi)) = \Theta_m(\delta^m + \delta^{m-1}\xi). \tag{71}$$

The unit-direction radius has the same order. The precision lower bound admits witnesses with identical query-direction covariance I_3, identical zero-radius estimator MSE at a common state, and exactly zero additional expectation. Any finite preselected list has the initialized consequence at sufficiently small steps, with the original annealed stabilizer and positive vanishing smoothing.

Proof. The upper bound is Theorem 22; the reference witnesses feasibility. For the precision lower bound, we need a perturbation that changes the response while leaving the extra reading exactly zero. Lemma 21 supplies the nonzero population forcing g_{rot} from rotation in coordinates $1, 3$. Choose a fixed small positive angle ϑ_0 and put $E = (R_{\vartheta_0})_\# D_{\mathrm{ref}}$. With $\Delta\nu_E = (\mathbb{E}_E X_j(x_0)^2 - \nu_{0,j})_j$, set $g_E = P_0 \operatorname{diag}(a_{0,j} f_1'(\nu_{0,j}))\Delta\nu_E$. This angle can be chosen from public population quantities alone: for $B = 3\|a_0\|_2$ and $K_\nu = 16B^2\|P_0 \operatorname{diag}(a_{0,j} f_1'(\nu_{0,j}))\|_{2\to 2}$, Taylor's theorem gives $\|g_E - \vartheta_0 g_{\mathrm{rot}}\| \le K_\nu \vartheta_0^2$. Indeed $|X_j| \le B$, $|X_j'| \le 2B$, $|X_j''| \le 4B$, so $|(X_j^2)''| \le 16B^2$. Taking $\vartheta_0 \le \min\{1/4, \|g_{\mathrm{rot}}\|/(2K_\nu)\}$ when $K_\nu > 0$ ensures $\|g_E\| \ge \vartheta_0\|g_{\mathrm{rot}}\|/2 > 0$.

Write $b_E = \mathbb{E}_E\psi_m$ and choose the exact endpoint $D_s = D_m^{-\operatorname{sign}(b_E)}$ when $b_E \ne 0$. The mixture

$$D_t = [1 - (1 + |b_E|)t]D_{\mathrm{ref}} + tE + |b_E|tD_s, \qquad 0 \le t \le (1 + |b_E|)^{-1}, \tag{72}$$

omits its last summand if $b_E = 0$. Its reading is exactly zero. Since D_s matches every capped moment, the capped error is $t(\mathbb{E}_E - \mathbb{E}_{\mathrm{ref}})\phi_\alpha$, bounded by $2t$. Thus $t = e/2$ gives a legal law in $\mathcal{F}_{m,e}(0, 0)$ for $e \le 2/(1 + |b_E|)$. All components are centered, isotropic and have squared norm three, so the mixture preserves query-direction covariance and the zero-radius raw MSE $2\|a\|^2$ at any common fixed state a.

The required comparison retains the perturbation size:

$$U_{D_t} - U_{\mathrm{ref}} = tg_E + O_m(t\delta + t^2 + \delta^m), \tag{73}$$

uniformly for small t, δ, without restricting their ratio. To see this, $\mu_{D_t} = a$ and $\nu_{D_t}(x) = \nu_{\mathrm{ref}}(x) + t\Delta\nu_E(x)$ exactly, because D_s matches the second and fourth query moments. The population difference is therefore $th_E + O_{C^1}(t^2)$, where $h_{E,j} = a_j f_\lambda'(\nu_{\mathrm{ref},j})\Delta\nu_{E,j}$. Use the degree-$2m$ cutoff expansion from Appendix A.4. Each finite raw moment and state derivative is affine in t; its centered cumulants and Taylor coefficients are smooth with bounded mixed derivatives. Every nonpopulation block carries at least one power of δ, so its difference is $O_{C^1}(t\delta)$. Subtracting the two uniform Taylor remainders costs $O_{C^1}(\delta^{m+1/2})$ and hence at most $O_{C^1}(\delta^m)$. The common root displacement is $O(t + \delta)$. Evaluating at $x_{\mathrm{ref}} = x_0 + O(\delta)$ and applying (19), whose averaged projector is $P_0 + O(t + \delta)$, proves (73). This uses the full physical state derivative and its uniform inverse margin.

For the two exact-information lower bounds, take $\xi_0 \le 1$. First $\mathcal{C}_{m+1} \subset \mathcal{C}_m$, and its members match the polynomial ψ_m of degree $4m$. Thus the correlated $(m+1)$-order endpoints lie in $\mathcal{F}_{m,0}(0,0)$; Theorem 19 at order $m+1$ gives half-gap at least $c_m\delta^m$ for small δ. Next let $D_m^{\pm}$ be the exact m-order correlated endpoints, and use $D_\xi^{\pm} = (1-\xi)D_{\mathrm{ref}} + \xi D_m^{\pm}$. They retain all degree-$4m-1$ moments, have $\mathbb{E}\psi_m = \pm\xi$, and hence lie in the same central fiber. Equation (70) gives their half-gap at least $C_m^*\delta^{m-1}\xi - C_m'\delta^m$. Choose a fixed $A \ge \max(1, 2C_m'/C_m^*)$. For $\xi \ge A\delta$ this is at least $(C_m^*/2)\delta^{m-1}\xi$, which also bounds a fixed multiple of the sum of the two target terms. For $\xi < A\delta$, the $(m+1)$ pair gives a fixed multiple of their sum, since $\delta^m + \delta^{m-1}\xi \le (1+A)\delta^m$. This proves (71), including $\xi = 0$, using the upper bound of Theorem 22.

Both pairs remain legal for every larger e. Finally compare (72), at $t = e/2$, with the reference. Shrink the separate e, δ thresholds so that the $O(t\delta + t^2)$ error is at most $t\|g_E\|/4$. For a sufficiently large fixed A_2, $e \ge A_2\delta^m$ makes the remaining error at most another $t\|g_E\|/4$, giving a gap at least $e\|g_E\|/4$. If $e < A_2\delta^m$, the first pair already supplies a multiple of e. The maximum of these three lower bounds bounds a fixed multiple of their sum, proving the radius order for all small-parameter ratios.

The bounded witness mixtures remain in the same stable domain and on its affine response line, with the same positive norm floor. Thus the normalization and finite-list argument of Theorem 22 applies, with its step threshold chosen after the fixed witness laws, memory and failure budget. □

C Proof of the initialized-response bridge

Theorem 2 follows from three steps: a uniform initialized-domain estimate, the fixed-q variational equation, and continuity of the optimal fiber radius.

C.1 A law-uniform initialized domain

Lemma 24 (A global bound for the initialized update)**.** *For zero raw filters and arbitrary scalar inputs $g_0, \dots, g_t$, if $\beta_1^2 < \beta_2 < 1$ and $\widehat{v}_t > 0$, then*

$$\frac{|\widehat{m}_t|}{\sqrt{\widehat{v}_t}} \le \frac{1-\beta_1}{1-\beta_1^{t+1}} \sqrt{\frac{(1-\beta_2^{t+1})[1-(\beta_1^2/\beta_2)^{t+1}]}{(1-\beta_2)[1-\beta_1^2/\beta_2]}}.$$

A positive outside-root stabilizer preserves this bound, including $\widehat{v}_t = 0$, and the right side is uniformly bounded in t.

Proof. The two exact filter sums and weighted Cauchy–Schwarz give

$$|m_t| \le (1-\beta_1)\sqrt{\frac{1-(\beta_1^2/\beta_2)^{t+1}}{1-\beta_1^2/\beta_2}}\sqrt{\frac{v_t}{1-\beta_2}}.$$

Applying the two absolute-counter bias corrections proves the formula. If $v_t = 0$, every contributing input and m_t vanish, so the stabilized update is zero. The bound is independent of the law and state, including paths that leave the local domain. It therefore controls the exceptional event in the unconditional expectation below. □

C.1.1 Uniform initialization on a bounded family of query laws

Lemma 25 (A law-uniform initialized-transfer interface)**.** *Fix data, a positive annealing rate, $\beta_1^2 < \beta_2 < 1$ and bounded IID query laws on a common sphere. Translate the common initializer to zero in the local coordinates. Suppose their full physical slow fields F_D satisfy, on a convex P-norm ball B_R about the common initializer, with $P \succ 0$,*

$$\langle v, DF_D(x)v\rangle_P \le -c\|v\|_P^2, \quad \|DF_D(x)\|_P \le L, \quad \|F_D(0)\|_P \le cR/8,$$

with $c, L, R > 0$ uniform in D. Assume the ball is inside the mass simplex and $\lambda \ge \lambda_- > 0$. If the normalized zero-radius input and its state derivative have common bounds there, then for every finite J preselected laws and $\rho > 0$ there is one $\varepsilon_* > 0$, uniform over those laws, such that the

zero-filter, two-bias-corrected algorithm with $\eta_t = \varepsilon\bar{\eta}_t$ stays in the ball and converges to its own roots and responses jointly with probability at least $1-\rho$, for $0 < \varepsilon \le \varepsilon_*$, under the following fixed schedules: $\bar{\eta}_t > 0$ is nonincreasing, $\sum_t \bar{\eta}_t = \infty$, $B := \sum_t \bar{\eta}_t^2 < \infty$, while $\mu_t > 0$, $\mu_t \to 0$ and $\sum_t \bar{\eta}_t \mu_t^2 < \infty$. The threshold may depend on these schedules and the fixed memory. In particular this includes $\bar{\eta}_t = (t+1)^{-3/4}$ and $\mu_t = 0.01(t+1)^{-1/4}$ used in the information-class applications. Neither isotropy, a zero query mean, finite support, nor lower atom masses are assumptions of this interface.

Proof. The argument has two parts: control the cumulative difference from the averaged recursion, then use contraction to rule out exit. Constants below are uniform over the query laws; norm-equivalence factors are absorbed.

First, the history observable and its state derivative forget old queries geometrically. To see this without a denominator singularity, put $r_0 = \sqrt{\beta_2}$, $\omega_k = (1-\beta_2)\beta_2^k$, $c_k = (1-\beta_1)\beta_1^k/\sqrt{\omega_k}$ and $z_k = \sqrt{\omega_k}X_k$. Then

$$\Phi(z,\lambda) = \frac{\langle c, z\rangle}{\|z\| + \lambda}, \qquad \|c\|^2 = C_\beta^2 = \frac{(1-\beta_1)^2}{(1-\beta_2)(1-\beta_1^2/\beta_2)} < \infty.$$

The derivative at zero is c/λ. The quotient rule and $|\langle c, z\rangle| \le C_\beta\|z\|$ bound $D_z\Phi$ and its Lipschitz constant by C_β/λ_- and $5C_\beta/\lambda_-^2$; the apparent singularity in the differentiated unit vector cancels. The same calculation bounds the history Lipschitz constant of $\partial_\lambda\Phi$ by $3C_\beta/\lambda_-^2$. Histories agreeing on their latest n queries differ by $O(r_0^n)$ in both input and input derivative. The observable and its full state derivative therefore have oscillations at most $C_0 r_0^n, C_1 r_0^n$. The bounded smooth matrices defining the physical field preserve these estimates.

Let ξ_{t-1} denote the pre-query past, ξ_t the updated past, $f_x(\xi_t)$ the frozen complete-history field and $h_x = \mathsf{P}f_x$ its pre-query conditional average, where P is the IID-history transition operator. Geometric forgetting gives a bounded, state-differentiable Poisson corrector $q_x = \sum_{k\ge 0}\mathsf{P}^k(h_x - F_D(x))$, with uniform bounds $\|q_x\| \le Q = C_0/(1-r_0)$ and $\|D_x q_x\| \le Q_x = C_1/(1-r_0)$ after enlarging the constants. At a predictable x_t,

$$\begin{aligned} f_{x_t}(\xi_t) - F_D(x_t) &= M_t^{\mathrm{fresh}} + M_t^{\mathrm{P}} + q_{x_t}(\xi_{t-1}) - q_{x_t}(\xi_t), \\ M_t^{\mathrm{fresh}} &= f_{x_t}(\xi_t) - h_{x_t}(\xi_{t-1}), \qquad M_t^{\mathrm{P}} = q_{x_t}(\xi_t) - \mathsf{P}q_{x_t}(\xi_{t-1}). \end{aligned}$$

Both martingale terms are centered given the pre-update sigma-field. Their sum, which need not have independent components, is bounded by $D_M = 2B_F + 2Q$, where B_F bounds f.

We next compare frozen histories with the actual filters. The shared-input ratio bound, including both bias corrections, is $C_\beta/(1-\beta_1)$, so log-loss and state increments are bounded by $K_L\eta_t, K_x\eta_t$, including the first possible exit update. Choose $\varepsilon_{\mathrm{tr}} \le 1$ so that the Euler maps contract and $r_0 e^{K_L\varepsilon_{\mathrm{tr}}\bar{\eta}_0} \le r = (1+r_0)/2 < 1$. The positive-radius estimate is uniform in the law and in the state. Indeed, for $\|z_i\| \le Z_b$, $\|v\| \le V_b$, put $\phi_\mu(p,v) = \sum_i p_i \sinh(\mu z_i^\top v)/\mu$ and $\phi_0(p,v) = v^\top \sum_i p_i z_i$. The inequality $|\sinh u - u| \le |u|^3\cosh(|u|)/6$ gives

$$|\phi_\mu - \phi_0| \le \mu^2 Z_b^3 V_b^3 \cosh(\mu Z_b V_b)/6 \le \mathrm{e}\mu^2 Z_b^3 V_b^3/6 \quad (0 \le \mu Z_b V_b \le 1). \tag{74}$$

A simplex-tangent derivative replaces p_i by $\dot{p}_i$, so the same bound gains a factor $\|\dot{p}\|_1$. Multiplication by $|v_j| \le V_b$ gives coordinate-input bound $\mathrm{e}\mu^2 Z_b^3 V_b^4/6$ in the stated radius range, and the analogous $\cosh(\mu Z_b V_b)$ bound at any bounded radius. The smooth mass chart on the compact domain preserves the $O(\mu^2)$ state-derivative bound. Since $\mu_{\max} = \sup_t \mu_t < \infty$, this also controls the initialized prefix.

Write $Q_t = (v_{t,1}, \dots, v_{t,q})$ for a query batch and $\mathcal{X}_\mu(x,Q) = -\overline{g}_\mu/L$ for its normalized input; at zero radius this is G_q. For lag $k \le t$, set $g_k = (L_{t-k}/L_t)\mathcal{X}_{\mu_{t-k}}(x_{t-k}, Q_{t-k})$ and $g_k^\circ = \mathcal{X}_0(x_t, Q_{t-k})$. If B_G, D_G bound the normalized input and its state derivative, Taylor's inequality gives

$$|g_k - g_k^\circ| \le e^{kK_L\varepsilon\bar{\eta}_0}\left[(B_G K_L + D_G K_x)\sum_{s=t-k}^{t-1}\eta_s + B_\mu\mu_{t-k}^2\right], \tag{75}$$

where $B_\mu = R_v M^3 \cosh(\mu_{\max} M)/6$, R_v is the sphere radius and $M = R_v \max_i \|z_i\|$. Minkowski and Tonelli sum these lag errors: monotonicity gives $\sum_{t\ge k} \eta_t \sum_{s=t-k}^{t-1} \eta_s \le k \sum_t \eta_t^2$ and $\sum_{t\ge k} \eta_t \mu_{t-k}^2 \le \sum_t \eta_t \mu_t^2$; the remaining geometric sums are finite. Missing history is $O(r_0^{t+1})$, and the two bias corrections cost $O(\beta_1^{t+1}/(1-\beta_1) + \beta_2^{t+1})$. The log-loss Euler error is $O(\eta_t^2)$. All are summable after the required step weighting.

Summation by parts of the corrector, including its state motion, gives

$$A_\delta = \bar\eta_0 + \sum_t |\bar\eta_{t+1} - \bar\eta_t| + \sum_t \bar\eta_t \mu_t^2 + \sum_t \bar\eta_t (\beta_1^{t+1} + \beta_2^{(t+1)/2}), \qquad \sup_n \|E_n^{\det}\|_P \le C_b(\varepsilon A_\delta + \varepsilon^2 B).$$

All sums are finite: monotonicity bounds the variation and $\bar\eta_t$, while missing-history weights decay geometrically. Stop the martingales just before exit. Their possibly correlated sum has terminal second moment at most $C_M \varepsilon^2 B$. Choose $r_* < R/4$ with $L r_* \le cR/8$. A sufficient common budget is

$$\varepsilon \le \min\left\{ \varepsilon_{\mathrm{tr}}, \frac{r_*}{2C_b(A_\delta + B + 1)}, r_* \sqrt{\frac{\rho}{4JC_M B}} \right\}. \tag{76}$$

The deterministic primitive is then at most $r_*/2$. For each stopped martingale primitive M_n, the maximal inequality for the nonnegative submartingale $\|M_n\|_P^2$ gives $\Pr(\sup_n \|M_n\|_P > r_*/2) \le 4C_M \varepsilon^2 B / r_*^2$. A union bound over the J laws is therefore at most ρ. No independence of the runs is needed. Thus the complete primitive E_n is at most r_* for every run. For the power schedules, $\sum_t (t+1)^{-3/2} \le 3$ and $\sum_t (t+1)^{-5/4} \le 5$, giving the $O_\delta(\varepsilon + \varepsilon^2)$ deterministic and $O_\delta(\varepsilon^2)$ squared martingale budgets used below.

On this event put $y_n = x_n - E_n$. Before exit, $y_{n+1} = y_n + \eta_n F_D(y_n + E_n)$. For small steps this map contracts by $1 - c\eta_n/2$ and sends zero to a point of norm at most $\eta_n(cR/8 + Lr_*) \le c\eta_n R/4$. Thus the y-ball of radius $R - 2r_*$ is invariant, including the potential exit update, and x_n remains strictly inside B_R.

Finally, the deterministic series converge absolutely and the weighted martingale converges almost surely. The primitive therefore has vanishing tail oscillations. Summation by parts and discrete Gronwall on fixed slow-time blocks compare the perturbed recursion with the contracting Euler flow. Iterating blocks gives $x_n \to x_D$. The same decomposition for response coordinates, divided by $S_n = \sum_{t<n} \eta_t \to \infty$, has vanishing error; weighted Cesàro convergence gives $w_n/S_n \to H_D(x_D)$ on the stated joint event. □

C.1.2 Initialized separation for the matched octagons

The original A/B octagons give different limiting directions from a common initialization under the same annealed algorithm.

Corollary 26 (Initialized matched-law separation)**.** *Use the two laws in Appendix A.7 and the two-row seed* (43)*, with $q = 1$ and $\beta_1 = 0.9$. For every fixed sufficiently small $\delta > 0$, set $\beta_2 = 1 - \delta$. The nearby Hurwitz roots have responses $U_A(\delta) - U_B(\delta) = \delta C_{\mathrm{seed}} + o(\delta)$ and noncollinear directions. Give both runs equal positive loss weights, $w_0 = 0$, zero raw filters, $\lambda_0 = 1$ and $\epsilon_0 = \lambda_0 L(w_0)$. Use the same annealed stabilizer and fixed schedules allowed by Lemma 25. For every $\zeta \in (0,1)$ there is $\varepsilon_*(\delta,\zeta) > 0$ such that, with the common multiplier $0 < \varepsilon \le \varepsilon_*$, on a joint event of probability at least $1 - \zeta$,*

$$x_t^D \longrightarrow x_D(\delta), \qquad \frac{w_t^D}{S_t} \longrightarrow U_D(\delta), \quad D \in \{A, B\}, \qquad S_t = \sum_{s<t} \varepsilon \bar\eta_s.$$

The runs have identical update counts, slow-time clocks and per-update query costs.

Proof. The seed stability calculation and (45) give a Hurwitz population root and a nonzero projected coefficient whose unit-normalized derivative is nonzero. The uniform C^1 comparison in Theorem 9 preserves these properties for small fixed δ and yields one Lyapunov ball containing both roots and the common initializer. The initializer drift is $O(\delta)$. On a smaller ball the stabilizer has a positive floor, while sphere support bounds the inputs and their state derivatives. Here $\beta_1^2 < \beta_2$ for small δ. Lemma 25, with $J = 2$ and $\rho = \zeta$ in (76), then controls the zero-filter prefix, both bias corrections and the first potential exit update, and gives the two limits on one event. The step multiplier is chosen after fixing δ. □

C.2 Scope of the transfer results

Stationary upper bounds use common root domains and invertible averaged Jacobians. Initialized transfer additionally fixes a finite law list and a common sufficiently small step multiplier. Its threshold may depend on dimension, information order, schedules and fixed memory.

Finite-budget boundary. The transfer theorem is qualitative because its constants depend on the fixed memory and the local contracting domain. A separate six-dimensional budget calculation did not resolve a positive signal at $256{,}000$ value queries.

C.3 The fixed-batch expected-response bridge

This subsection proves Theorem 2 directly at fixed q. The chart parameter is ϑ from Section 4.4, not the hidden-moment vector $\theta(D)$ of Theorem 1. Keep its closed ball and Euclidean metric fixed. With $\mathcal{T}(p)$ as in the proof of Lemma 4, write $G_H = \mathcal{T}(p)(ZH - \kappa\mathbf{1})$.

Common initialized domain. Use the population field $H_{0,q}$ in Section 4.2. Assume a compact convex ball of radius R about the common initializer, in a fixed P-norm, lies inside the positive mass and stabilizer domain, with $|a_j| \geq a_- > 0$, $\lambda \geq \lambda_- > 0$, and

$$\langle z, DG_{0,q}(x)z\rangle_P \leq -2\gamma\|z\|_P^2.$$

Require a strict inward margin and the common initializer in a smaller trapping ball, with the initialization bounds stated next. The input and state derivatives are bounded on a neighborhood of this ball. Take the zero-filter initializer, absolute bias-correction counters and annealed schedules of Theorem 2; the initialization bounds in Appendix C.1.1 hold there. Uniform C^1 closeness preserves contraction and the inward margin for small δ. All constants in this subsection may depend on fixed q; stochastic tracking constants may also depend on δ.

Forcing and the stationary operator. At the root x_q of $G_{0,q}$, put

$$E_q = D_x H_{0,q}(x_q), \quad P_q = I - E_q(ZE_q)^{-1}Z.$$

$$D_q(x) = \operatorname{diag}_j \frac{a_j f_\lambda''(\nu_{q,j})}{4}, \quad K_{j\ell}(x) = \mathbb{E}_{\sigma_d}[X_j(x,v)^4 h_\ell(v)].$$

The invertible row Jacobian follows from contraction. Set $L_q = D_q K$ and $B_q = P_q L_q(x_q) = q^3 \mathsf{B}_{2,q} K(x_q)$. Here $H_{\delta,q,\vartheta}$ is the shared stationary field (4) for D_ϑ. Each chart law matches the first three scalar input moments. The batch identity (23) and the shared-history expansion give

$$H_{\delta,q,\vartheta} - H_{\delta,q,0} = \delta q^{-3} L_q \vartheta + O_{C^1,q}(\delta^2), \qquad H_{\delta,q,0} = H_{0,q} + O_{C^1,q}(\delta). \tag{77}$$

Both estimates are uniform over the fixed law ball.

C.3.1 From finite slow time to root adjustment

Let $x(s)$ solve $\dot{x} = G_{0,q}(x)$ from x_{in}, and define

$$\begin{aligned} \dot{Y} &= DG_{0,q}(x(s))Y + \mathcal{T}(p(s))ZL_q(x(s)), \quad Y(0) = 0, \\ M_{q,S}^{\mathrm{hm}} &= \frac{1}{S}\int_0^S [L_q(x(s)) + D_x H_{0,q}(x(s))Y(s)]\, ds. \end{aligned} \tag{78}$$

Contraction gives $x(s) - x_q = O_q(e^{-\gamma s})$. At equilibrium,

$$Y_q = -(ZE_q)^{-1}ZL_q(x_q), \qquad L_q(x_q) + E_q Y_q = B_q.$$

The equation for $Y - Y_q$ has forcing $O_q(e^{-\gamma s})$. Convolution with the contracting evolution operator yields $Y(s) - Y_q = O_q((1+s)e^{-\gamma s})$, hence $\|M_{q,S}^{\mathrm{hm}} - B_q\| \leq C_q/S$ for $S \geq 1$. The full derivative $DG_{0,q}$ includes $D\mathcal{T}$ away from the root.

For the actual stationary fields let $x_{\delta,q,\vartheta}(s)$ be their averaged flows from the same initializer and put $\overline{F}_{\delta,q,S}(\vartheta) = S^{-1}\int_0^S H_{\delta,q,\vartheta}(x_{\delta,q,\vartheta}(s))\, ds$. Their reference-flow displacement is $O_q(\delta)$, and their

mutual law displacement is $O_q(\delta)$. Subtracting $\delta q^{-3} Y(s)\vartheta$ from the law-flow difference leaves a contracting residual equation: (77), the $O_q(\delta)$ reference displacement and the quadratic law displacement bound its forcing by $O_q(\delta^2)$. The reference field is smooth; only a C^1 bound on its random-field discrepancy is used. With zero initial residual, convolution gives an $O_q(\delta^2)$ state and output error uniformly in slow time. Thus

$$\sup_{\|\vartheta\|\le\rho} \left\| \frac{q^3}{\delta}[\overline{F}_{\delta,q,S}(\vartheta) - \overline{F}_{\delta,q,S}(0)] - B_q\vartheta \right\| \le C_q(\delta + S^{-1}). \tag{79}$$

C.3.2 Unconditional tracking of initialized Adam

Write $F_n(\vartheta) = \mathbb{E}_\vartheta[(w_n - w_0)/S_n]$ for the unconditional expectation in the main theorem. Both filters use the same current batch.

Lemma 27 (Weak tracking at fixed batch and memory)**.** *Under the common initialized-domain conditions, for each fixed sufficiently small δ there are $C_{q,\delta} < \infty$ and $\varepsilon_0(q,\delta) > 0$ such that, for $0 < \varepsilon \le \varepsilon_0$ and $S_n \ge 1$,*

$$\sup_{\|\vartheta\|\le\rho} \|F_n(\vartheta) - \overline{F}_{\delta,q,S_n}(\vartheta)\| \le C_{q,\delta}\varepsilon(1 + S_n^{-1}). \tag{80}$$

Proof. Treat each batch as one history symbol in Lemma 25. Averaging preserves the input and state-derivative bounds; the geometric history oscillation $C_{q,\delta}\beta_2^{k/2}$ therefore bounds the pre-query corrector and its derivative uniformly over the law. The contraction and inward margin supply that lemma's common trapping domain.

Stop immediately before its first exit. The same decomposition, for both state and response, bounds the cumulative error E_n by a deterministic part $O_\delta(\varepsilon + \varepsilon^2)$ and a martingale with squared maximal expectation $O_\delta(\varepsilon^2)$. It includes $\sum \eta_j^2 = O(\varepsilon^2)$, smoothing $\sum \eta_j \mu_j^2 = O(\varepsilon)$, loss-ratio and parameter tracking, initial-filter decay, both bias corrections, and corrector boundaries. As in that lemma, apply the martingale inequality to the possibly correlated sum of fresh-query and Poisson increments. Thus, for a fixed small trapping margin r_* and sufficiently small ε,

$$\mathbb{E}\sup_n \|E_n\| \le C_{q,\delta}\varepsilon, \qquad \Pr(\sup_n \|E_n\| > r_*) \le C_{q,\delta}\varepsilon^2.$$

Lemma 24 bounds $(w_n - w_0)/S_n$ globally. Its bound applies to the zero-filter prefix and both absolute-counter bias corrections. Missing history and loss-ratio tracking are already included in the cumulative estimates above.

On $\sup \|E_n\| \le r_*$, Lemma 25's shifted-variable argument prevents exit, including the first potential exit update, and contraction bounds the distance to the averaged flow by $C \sup \|E_n\|$ plus Euler error $O_\delta(\varepsilon)$. Averaging the Lipschitz response adds $C_{q,\delta}\varepsilon/S_n$. On the exceptional event, the global ratio bound and bounded averaged output cost $O_\delta(\varepsilon^2)$ in expectation. This yields the unconditional, law-uniform bound. □

Proof of Theorem 2: response limit. Apply Lemma 27 to the target and reference laws, then use (79). After enlarging the tracking constant to cover both laws,

$$\sup_{\|\vartheta\|\le\rho} \left\| \frac{q^3}{\delta}[F_n(\vartheta) - F_n(0)] - B_q\vartheta \right\| \le C_q(\delta + S_n^{-1}) + C_{q,\delta}\frac{q^3\varepsilon}{\delta}(1 + S_n^{-1}) =: e_n. \tag{81}$$

The declared coupled limit makes all terms vanish. In particular, taking long slow time alone does not remove the fixed-memory tracking error. □

C.3.3 Transfer of the information radius

For a map F on the fixed chart ball let $\mathcal{R}_k(F)$ be its optimal worst-fiber enclosing radius after k preassigned bounded expectation readings. If $\|F - G\|_\infty \le e$, each fiber's two response sets are within Hausdorff distance e, so their enclosing radii differ by at most e. Taking the worst fiber and then optimizing the readings preserves the inequality.

On this density chart every bounded reading is affine in ϑ. Conversely, the positive definite Gram matrix of the bounded independent h_j realizes every linear reading by a bounded linear combination of these functions. For a linear map B, the first k right singular coordinates give radius at most $\rho\sigma_{k+1}(B)$. Any k readings have a kernel vector in the first $k+1$ right singular subspace; the two opposite radius-ρ points in that direction give the matching lower bound. Thus $\mathcal{R}_k(B\vartheta) = \rho\sigma_{k+1}(B)$, including zero beyond the rank.

Translation invariance, positive scaling and (81) therefore give

$$\left| \frac{q^3}{\delta} R_k^{\mathrm{act}} - \rho\sigma_{k+1}(B_q) \right| \le e_n \longrightarrow 0,$$

which completes Theorem 2.

C.4 Spectral Interpretation and Prediction Resolution

The uniform response bound controls singular values and therefore the number of directions resolved above a separated threshold. Let $\mathcal{F}_n(\vartheta) = q^3\{F_n(\vartheta) - F_n(0)\}/\delta$ and retain e_n from (81). For a fixed $0 < h_{\mathrm{sec}} \le \rho$, form the central-secant matrix from exact expectations:

$$M_n^{(h_{\mathrm{sec}})} e_\ell = \frac{\mathcal{F}_n(h_{\mathrm{sec}} e_\ell) - \mathcal{F}_n(-h_{\mathrm{sec}} e_\ell)}{2h_{\mathrm{sec}}}, \qquad 1 \le \ell \le s.$$

Here s is the chart dimension.

Corollary 28 (Spectral resolution at fixed batch size)**.** *Under the hypotheses of Appendix C.3,*

$$\|M_n^{(h_{\mathrm{sec}})} - B_q\|_{\mathrm{op}} \le e_M := \frac{\sqrt{s}\, e_n}{h_{\mathrm{sec}}}, \qquad |\sigma_j(M_n^{(h_{\mathrm{sec}})}) - \sigma_j(B_q)| \le e_M.$$

If $k_* = \mathrm{rank}(B_q) > 0$, put $\gamma = \sigma_{k_*}(B_q)$. When $e_M < \gamma/2$, every threshold $e_M < \tau < \gamma - e_M$ selects exactly k_* singular directions. For $k_* = 0$, the operator-norm bound gives $\|M_n^{(h_{\mathrm{sec}})}\|_{\mathrm{op}} \le e_M$.

Proof. Each column error has norm at most e_n/h_{sec} by the uniform response bound at the two endpoints. The Frobenius norm bounds the operator norm by $\sqrt{s}\, e_n/h_{\mathrm{sec}}$. The singular-value perturbation inequality then gives the displayed estimate and the threshold claim. □

Threshold dimension is stable under this separated-error condition; exact algebraic rank need not be. Numerically estimated secants require an additional sampling and numerical error bound.

A sufficient scale for stationary prediction order. For $q = 1$, $m = 2$ and $k_* = \mathrm{rank}(B_1)$, the radius transfer in Appendix C.3.3 gives $R_{k_*}^{\mathrm{act}} \le \delta e_n$, since $\sigma_{k_*+1}(B_1) = 0$. Hence, for $S_n \ge 1$ and after absorbing $1 + S_n^{-1} \le 2$ into the tracking constant,

$$R_{k_*}^{\mathrm{act}} \le C\left(\delta^2 + \frac{\delta}{S_n}\right) + C_{1,\delta}\varepsilon.$$

Under the same initialization, local-domain and step-size conditions, $S_n \gtrsim \delta^{-1}$ and $C_{1,\delta}\varepsilon \lesssim \delta^2$ are sufficient for the $O(\delta^2)$ prediction upper order. The index equals $d - r$ only when the chart's realizability and nondegeneracy give $\mathrm{rank}(B_1) = d - r$. These scale relations separate the transient and tracking contributions; they do not determine the constants for the finite tail-replication budget.

D Numerical response bounds and experimental evidence

Finite-memory response bounds quantify how query-law differences survive adjustment to each law's stationary root. We derive the response enclosures used in the main figures, then report the stationary comparisons and initialized-tail replication.

D.1 Moment computation and response enclosures

These results turn permitted moments into finite-memory enclosures. The approximation error, movement of the reference center, and root residual have different origins and are bounded separately.

Computing shared polynomial moments. For $|X| \le B$ and fixed $\beta_1, \beta_2 < 1$, expanding AV^k by lag indices gives an absolute sum at most B^{2k+1}. Fubini therefore reduces its expectation to raw input moments through degree $2k+1$. The triangular recursions (83) compute these moments exactly, without truncating history. For a fixed IID batch, the input $X_j(v) = v_j v^\top a$ requires query moments through degree $4k+2$ by multinomial expansion.

D.1.1 Shared-history moments at a public anchor

At a public state x, choose a positive differentiable anchor $\nu(p)$ and put $r_k = \mathbb{E}_D(X^2-\nu)^k$ and $s_k = \mathbb{E}_D[X(X^2-\nu)^k]$. Write $M_k = \mathbb{E}_D(V-\nu)^k$ and $J_k = \mathbb{E}_D[A(V-\nu)^k]$. Starting from $M_0 = 1$, $J_0 = s_0$, compute to every required order $k \ge 1$

$$M_k = \frac{\sum_{j<k} \binom{k}{j} \beta_2^j \delta^{k-j} M_j r_{k-j}}{1-\beta_2^k}, \tag{82}$$

$$J_k = \frac{\beta_1 \sum_{j<k} \binom{k}{j} \beta_2^j \delta^{k-j} J_j r_{k-j} + (1-\beta_1) \sum_{j\le k} \binom{k}{j} \beta_2^j \delta^{k-j} M_j s_{k-j}}{1-\beta_1\beta_2^k}. \tag{83}$$

They follow by splitting the same fresh input from both stationary filters. In particular r_1 and s_0 are not set to zero and a, respectively, on a noisy fiber. The response approximation is $\sum_{k<N} f_\lambda^{(k)}(\nu) J_k/k!$.

The exact remainder. For $W = V - \nu$ and positive ν, λ, the Stieltjes representation of $f_\lambda(v) = (\lambda + \sqrt{v})^{-1}$ gives

$$R_N(V) = \int_0^\infty \frac{\sqrt{s}}{\pi(\lambda^2+s)} \frac{(-W)^N}{(\nu+s)^N (V+s)}\, ds. \tag{84}$$

This follows by a finite geometric expansion of $(V+s)^{-1}$ about $\nu+s$ and is exact without a small-W assumption. The next subsection bounds its expectation and derivatives using even moments from (83).

D.2 Computing stationary responses at finite memory

We now evaluate the zero-smoothing stationary responses and slow roots at two fixed memories, independently of the asymptotic coefficient C_{seed}.

For one coordinate, use the shared filters above and set $Y = X^2 - \nu$, $W = V - \nu$, $M_k = \mathbb{E}W^k$ and $J_k = \mathbb{E}[AW^k]$. Use the general fresh-input recursion (83) with $r_k = \mathbb{E}Y^k$ and $s_k = \mathbb{E}[XY^k]$. Its initial values are $M_0 = 1$ and $J_0 = \mathbb{E}X$. The fixed anchor need not equal the current second moment, so r_1 is retained. The response uses $N = 24$ Taylor terms at the fixed coordinate anchor $\nu \in \{9/2, 15/2, 6\}$: the polynomial runs from $k = 0$ through $k = N-1$, and the remainder starts at degree N. Use the exact remainder (84) at this constant anchor. Let $|X| \le M$, $|\partial_p X| \le L$ uniformly, $\lambda \ge \lambda_- > 0$, and let E_N, E_{N-2} bound the even moments $\mathbb{E}W^N, \mathbb{E}W^{N-2}$. With $E_{N-1} = \sqrt{E_N E_{N-2}}$, the omitted response and derivatives obey

$$|\mathbb{E}[AR_N]| \le \frac{ME_N}{\lambda_-\nu^N},$$

$$|\partial_p \mathbb{E}[AR_N]| \le \frac{1}{\nu^N}\left\{\left(\frac{L}{\lambda_-} + \frac{ML}{\lambda_-^2}\right)E_N + \frac{2M^2LN}{\lambda_-} E_{N-1}\right\},$$

$$|\partial_\lambda \mathbb{E}[AR_N]| \le \frac{2ME_N}{\lambda_-^2\nu^N}.$$

For the value bound, $(\nu+s)^{-N} \le \nu^{-N}$ and $f_\lambda(V) \le \lambda_-^{-1}$. The Stieltjes density's λ derivative is bounded by $2/\lambda_-$ times that density. For the state derivative, the constant anchor gives $W' = V'$, $|A'| \le L$ and $|W'| \le 2ML$. Cauchy–Schwarz gives $\mathbb{E}|W|^{N-1} \le \sqrt{E_N E_{N-2}}$, yielding the displayed power-derivative term; $|V'| \le 2L\sqrt{V}$ controls the denominator as in the remainder formula. These integrable bounds justify differentiation. This coarser estimate requires no moments of order $2N-2$. The implementation uses $M = 9, L = 2$ and moments through N. The log-scale derivative uses the upper endpoint of λ times the last bound.

Two independent computations enclose the response and its Jacobian. The first uses (83), mpmath intervals and the global Stieltjes remainder. The second uses Fraction/integer outward arithmetic, cumulants, Bell recurrences and a separately bounded logarithm. Both establish strict Krawczyk inclusion in physical (p, λ) coordinates. The root boxes are transformed to $(\operatorname{logit} p, \log \lambda)$ for reporting. Local stability follows from negative trace and positive determinant of the full slow Jacobian, including the state-coordinate transformation. Appendix D.3.1 reports the resulting root separation and mechanism comparisons.

Diagnostics for the main root-adjustment comparison. Section 5.1 defines the projected-amplitude ratio R and full-vector error E. For the three controls $c_{\mathrm{raw}} = \Delta h$, $c_{\mathrm{orth}} = \Pi\Delta h$ and $c_{\mathrm{adj}} = C_{\mathrm{seed}}$ in Figure 1(d), the displayed quantity is

$$E_j^{\mathrm{mid}} = \frac{\|\Delta U^{\mathrm{mid}} - \delta c_j^{\mathrm{mid}}\|_2}{\delta \|C_{\mathrm{seed}}^{\mathrm{mid}}\|_2}. \tag{85}$$

For this $d = 3, r = 2$ seed, $Z\Delta U = 0$, $C_{\mathrm{seed}} \in \ker Z \setminus \{0\}$ and $\dim \ker Z = 1$. Hence $\Delta U = \gamma\delta C_{\mathrm{seed}}$, $R = \gamma$ and $E = |R-1|$ for the exact responses. Independently rounded enclosures and midpoint quantities need not reproduce this identity numerically; their saved records are retained. Panel (d) compares predictors, not independent transverse response components. All three use the same independently enclosed $A-B$ gap and denominator. The raw and orthogonal controls carry no independent certified error bars. Appendix D.3.1 reports the root boxes, eight-memory endpoints, control comparisons and primary/held-out/follow-up provenance.

D.2.1 Fixed-instance reading constructions and diagnostics

The configurations in Figure 3 use $m = 2$, $(d, r) = (3,1), (4,2), (6,3)$ and $\beta_2 = 0.999, 0.9999$. Insufficient readings have $k = d - r - 1$; redundant readings have $k = d - r$ with one repeated direction; nonaligned readings have $k = d - r$ but leave a response direction unobserved. All query laws share the prescribed lower-order moments. The accompanying `data/reading_law_protocol.json` specifies the density coefficients, Gram matrices and reading rows; `data/READING_PROTOCOL.md` defines their harmonic basis.

In the frozen spherical chart, density parameters t induce hidden moments $\theta = Gt$ through the exact harmonic Gram matrix G. The selected minus/plus laws satisfy

$$\mathcal{A}G(t_+ - t_-) = 0, \qquad \mathsf{B}_{m,q}G(t_+ - t_-) \neq 0. \tag{86}$$

Their independently enclosed response gaps supply pair-based half-gap lower bounds. Including the reference law gives three unique equal-reading comparisons per setting (18 in total), missed by each of the three defective maps. These shared witnesses instantiate the indistinguishability mechanism of Theorem 1. The stored aligned rows $\mathcal{A}_\star$ span the response-relevant subspace, so a fixed change of Gram coordinates gives the correction in (12). Its baseline is the reference law's own-root response at the same finite memory.

The six settings contain 46 diagnostic rows: seven laws per memory for $(3,1)$ and $(4,2)$, and nine for $(6,3)$. Maximum reference-only upper bounds across settings lie on the 10^{-13}–10^{-11} scale and aligned bounds on the 10^{-16}–10^{-13} scale. These fixed-instance comparisons share truth roots and norm; the whole-class statement is Theorem 1. No saved identical-reading pair is supplied for aligned configurations, so the absence of such an aligned witness is not a zero-radius assertion. Appendix E.5 distinguishes this named-law evidence from retrospective linear combinations of axis secants.

Figure 3 displays $b_i^{\mathrm{ref}}/b_i^{\mathrm{aligned}}$ for the 40 non-reference records, pairing the same target response, memory and Euclidean norm. The ratio is formed from the saved decimal reference bound and exact

rational aligned bound before conversion to plotting coordinates. It compares two upper bounds, rather than actual errors or a guaranteed error-reduction factor. Table 2 retains the absolute scales for every target; exact source values and ratios accompany the figure as a deterministic record table.

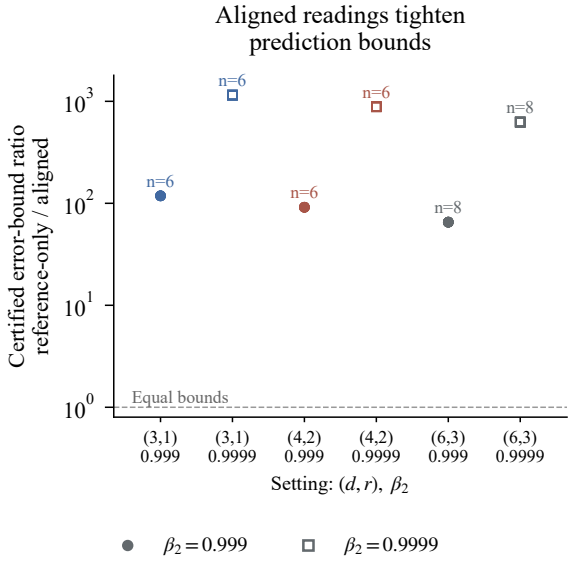


Figure 3: Certified bound tightening for all 40 target records. Ratios compare reference-only and aligned prediction-error upper bounds at the same law, memory and norm; they are not actual prediction-error ratios. Absolute bounds and legal equal-reading pairs are retained below.

Table 2: Absolute prediction upper bounds for all 40 targets in Figure 3, grouped into 20 minus/plus pairs. IDs retain the original non-reference order. Direction j names laws direction$j$_minus and direction$j$_plus, in that order. Each bound is rounded upward independently to four significant digits. A single displayed value applies to both listed signs; distinct displays are separated by a semicolon. Shared displays do not assert equality at full precision.

IDs $(-,+)$	Direction	b^{ref} $(-,+)$	b^{aligned} $(-,+)$
Setting: $(d,r) = (3,1)$, $\beta_2 = 0.999$			
1, 2	0	2.222×10^{-11}	1.877×10^{-13}
3, 4	1	2.222×10^{-11}	1.877×10^{-13}
5, 6	2	1.924×10^{-11}	1.626×10^{-13}
Setting: $(d,r) = (3,1)$, $\beta_2 = 0.9999$			
7, 8	0	2.205×10^{-12}	1.916×10^{-15}
9, 10	1	2.205×10^{-12}	1.916×10^{-15}
11, 12	2	1.910×10^{-12}	1.660×10^{-15}
Setting: $(d,r) = (4,2)$, $\beta_2 = 0.999$			
13, 14	0	1.245×10^{-11}	1.360×10^{-13}
15, 16	1	1.245×10^{-11}	1.360×10^{-13}
17, 18	2	1.174×10^{-11}	1.283×10^{-13}
Setting: $(d,r) = (4,2)$, $\beta_2 = 0.9999$			
19, 20	0	1.233×10^{-12}	1.393×10^{-15}
21, 22	1	1.233×10^{-12}	1.393×10^{-15}
23, 24	2	1.162×10^{-12}	1.313×10^{-15}

IDs $(-,+)$	Direction	b^{ref} $(-,+)$	b^{aligned} $(-,+)$
Setting: $(d,r)=(6,3)$, $\beta_2=0.999$			
25, 26	0	3.889×10^{-12}	5.951×10^{-14}
27, 28	1	3.889×10^{-12}	5.951×10^{-14}
29, 30	2	3.889×10^{-12}	5.951×10^{-14}
31, 32	3	3.689×10^{-12}	5.646×10^{-14}
Setting: $(d,r)=(6,3)$, $\beta_2=0.9999$			
33, 34	0	3.835×10^{-13}	6.128×10^{-16}
35, 36	1	3.835×10^{-13}	6.128×10^{-16}
37, 38	2	3.835×10^{-13}	6.128×10^{-16}
39, 40	3	3.639×10^{-13}	5.814×10^{-16}

D.3 Stationary comparisons and initialized-tail replication

The own-root certificates (Appendix D.3.1) support the stationary root-adjustment comparison. The initialized-tail replication (Appendix D.3.2) then examines a late segment of trajectories evolved continuously from initialization.

D.3.1 Own-root interval certificates

Interval enclosures establish distinct responses at the matched laws' own attracting roots. We use the three-dimensional seed (43), with $a_0=(1,2,1)$, $\nu=(9/2,15/2,6)$ and $\lambda_0=1$. The matched laws of Appendix A.7.3 lift uniform octagons of rotations 0 and $\pi/8$ by an independent Rademacher sign; each has sixteen atoms of probability $1/16$ and the same per-update query count.

The pair has matching moments and nonzero projected forcing, so it tests whether the law difference survives adjustment to each law's own root. Normalization, moment matching and the projector identity are checked exactly. We use $q=1$, $\beta_1=9/10$, $\beta_2=999/1000$ and $1999/2000$, and the physical slow field on $|s|,|\ell|\le 1/100$. The two evaluators and their remainder bounds are described in Appendix D.2. Table 3 encloses both evaluations by rounding lower endpoints down and upper endpoints up; exact rational endpoints accompany the source data. Response-norm gaps and unit-direction gaps are reported separately. These certificates describe the fixed pair's local stationary roots and separation.

Two-memory root and primary-memory contraction results.

Proposition 29 (Finite-memory root and contraction certificates)**.** (i) Roots and responses at two memories. *For each law, at both $\beta_2=0.999$ and the preregistered held-out $\beta_2=0.9995$, the slow drift has a unique zero inside its certified root box contained in $|\operatorname{logit} p|,|\log\lambda|\le 0.01$. Every root is Hurwitz. The root-adjusted response enclosures are reported in Table 3.*

The signed alignment with C_{seed} is positive at both points, and

$$R_{0.999}\in[1.0536104944,1.0538677613],\qquad R_{0.9995}\in[1.0269697332,1.0274836150],$$

where $R_\beta=\langle C_{\mathrm{seed}},U_A-U_B\rangle/((1-\beta_2)\|C_{\mathrm{seed}}\|_2^2)$. The full validated uncertainty diameter is less than one quarter of every reported signal; the held-out ratio lies inside $[1/2,3/2]$. (ii) Contraction at the primary memory only. *At $\beta_2=0.999$ and $\beta_1=0.9$, both validated stationary slow fields have strictly inward faces on*

$$\mathcal{D}=\{p\in[0.499,0.5015],\ \lambda\in[0.998,1.001]\}.$$

After transforming derivatives to $(s,\ell)=(\operatorname{logit}p,\log\lambda)$, an order-64 outward interval enclosure gives symmetric-part Gershgorin upper bounds approximately -0.01772 for both laws. The validated roots at $\beta_2=0.999$ and 0.9995 lie in this domain. Consequently, at $\beta_2=0.999$ only, each deterministic stationary mean ODE is contractive on the domain and has a unique attracting root there. Containment of the 0.9995 roots in $\mathcal{D}$ asserts no domain-wide contraction certificate at that held-out memory. The proposition certifies stationary roots and contraction on $\mathcal{D}$ at $\beta_2=0.999$; stochastic tracking and initialized transfer are treated separately in Appendix C.

At $\beta_2 = 0.999$, the smallest inward face margin is $7.3928883257420196 \times 10^{-5}$. Thus a deterministic additive slow-drift perturbation with componentwise sup norm below $3.6964441628710098 \times 10^{-5}$ preserves all face signs on $\mathcal{D}$. This supplies a deterministic input to a stopped-noise argument, whose martingale and Euler budgets must still be established.

Finite-memory profile. At all eight tested memories, the two evaluators enclose positive alignment, nonzero response and unit-direction gaps, $R \in [0.5, 1.5]$ and $E \leq 0.10$ (Table 3). The initial comparisons at $\beta_2 = 0.999$ and 0.9995 served as the primary and held-out points; the six additional endpoints were specified before their evaluation. Each law has its own root, with $\Delta U = U_A - U_B$, $\Delta h = h_A - h_B$ and $C_{\text{seed}} = P_0 \Delta h$.

β_2	R	E	$10^6 G$	$10^6 G_{\text{unit}}$
0.999	$[1.0536, 1.0539]$	$[0.053610, 0.053868]$	$[0.95372, 0.95396]$	$[1.3833, 1.3841]$
0.99925	$[1.0403, 1.0408]$	$[0.040399, 0.040743]$	$[0.70632, 0.70656]$	$[1.0244, 1.0252]$
0.9995	$[1.0269, 1.0275]$	$[0.026970, 0.027484]$	$[0.46480, 0.46504]$	$[0.67409, 0.67476]$
0.999625	$[1.0201, 1.0209]$	$[0.020146, 0.020831]$	$[0.34628, 0.34653]$	$[0.50216, 0.50284]$
0.99975	$[1.0136, 1.0138]$	$[0.013698, 0.013710]$	$[0.22939, 0.22941]$	$[0.33276, 0.33278]$
0.9998125	$[1.0102, 1.0104]$	$[0.010288, 0.010302]$	$[0.17147, 0.17148]$	$[0.24873, 0.24875]$
0.999875	$[1.0068, 1.0069]$	$[0.0068644, 0.0068850]$	$[0.11392, 0.11393]$	$[0.16526, 0.16527]$
0.9999	$[1.0054, 1.0056]$	$[0.0054906, 0.0055163]$	$[0.091016, 0.091020]$	$[0.13202, 0.13204]$

Table 3: Stationary separation across eight memory values. R measures signed alignment with the first-order prediction and E its relative vector error. $G = \|\Delta U\|_2$ and $G_{\text{unit}} = \|U_A/\|U_A\| - U_B/\|U_B\|\|_2$. Each interval encloses both independent computations, with endpoints rounded outward to five significant digits. Alignment and both gaps are positive at every tested memory.

Table 3 reports outward-rounded hulls of the two independent enclosures. Across these eight memories, the absolute response gap decreases as normalized agreement improves. The adjusted midpoint lies within the enclosure for E and below both controls at every memory. The endpoint enclosures for E are approximately $[5.361, 5.387]\%$ at $\delta = 10^{-3}$ and $[0.5491, 0.5517]\%$ at $\delta = 10^{-4}$; nonzero unit-direction gaps exclude a pure speed change.

Projected forcing and independent finite-memory comparison. The mechanism comparison separates the direct field difference from the change induced by moving the root. With sign convention $A - B$, the population-root coefficients are $\Delta h \simeq (-0.0027745704, -0.0018782766, 0)$ and $C_{\text{seed}} \simeq (-0.00076594035, 0.00045828894, -0.00015063754)$; the first-order adjustment is $-(I - P_0)\Delta h$. Each law's finite-memory root is solved independently using the evaluators in Appendix D.2; response comparisons use R and E as defined in Appendix D.2.

The unit-direction tangent has norm about 0.00131309, excluding pure rescaling. Frozen-λ, Euclidean-projection and sign-flipped alternatives give different responses; independent numerator and denominator histories erase a nonzero shared-history term. These comparisons distinguish the projected, shared-history mechanism from those alternatives.

Separate deterministic bridge validation and stochastic boundary. The density-chart calculation tests the off-root variational formula: the physical response $L_q + D_x H_{0,q} Y$ agrees with the reduced formulation to relative discrepancy 2.30×10^{-14}. An off-root rank-two control detects omission of $D\mathcal{T}(p)$, and fixed-q roots are solved separately. The causal crossover, conditional-axis and multistate comparisons did not meet their predeclared criteria. Earlier direct estimators did not resolve the prescribed initialized complete-prefix target within their tested budgets. The subsequent transport study measures a same-trajectory tail response with independent replicas (Appendix D.3.2); the earlier inconclusive result remains unchanged.

D.3.2 Independent same-trajectory tail replication

Two independently seeded replicas resolve a late-window law response within a prespecified tolerance of the fixed-q high-memory prediction. We retain the initialized trajectory throughout and compare a fixed segment of its displacement with the corresponding deterministic target.

Observable and configuration. Let $\vartheta_\pm = \pm\rho v_1$ be the endpoints of the registered law-chart direction, and retain $F_k(\vartheta) = \mathbb{E}_\vartheta[(w_k - w_0)/S_k]$ from Section 4.4. Define the normalized law contrasts and window target by

$$\mathcal{C}_k = \frac{q^3}{2\rho\delta}\big[F_k(\vartheta_+) - F_k(\vartheta_-)\big], \tag{87}$$

$$\mathcal{C}_{b,n} = \frac{S_n\mathcal{C}_n - S_b\mathcal{C}_b}{S_n - S_b}, \qquad T_{b,n} = \frac{S_n M^{\mathrm{hm}}_{q,S_n} v_1 - S_b M^{\mathrm{hm}}_{q,S_b} v_1}{S_n - S_b}. \tag{88}$$

Here $M^{\mathrm{hm}}_{q,S}$ is the fixed-q high-memory variational response operator, retaining the fixed-q second moment, and $\mathcal{C}$ denotes a law contrast. The reported scalar is $v_1^\top \mathcal{C}_{b,n}$. We use the three-dimensional spherical chart with active row $a = (1, 1, 3/2)$, $q = 1$, $v_1 = (1, -1, 0)/\sqrt{2}$, $\beta_1 = 0.9$, $\beta_2 = 0.98$ and $\delta = 0.02$. The amplitude is $\rho \simeq 6.19434 \times 10^{-4}$. The schedules are $\eta_t = 0.01(t+1)^{-3/4}$ and $\mu_t = 0.01(t+1)^{-1/4}$, with outside-root stabilizer $\epsilon_t = \epsilon_0 e^{-\kappa S_t}$, $\kappa = 1.9$ and initial relative stabilizer $\lambda_0 = \epsilon_0/L(w_0) \simeq 0.126580257$. Parameters and both raw filters start at zero. The fixed window has $b = 1,179$, $n = 13,911$, $S_b \simeq 0.200002$ and $S_n \simeq 0.400001$. Optimizer states, tangent states and absolute schedule counters are inherited continuously from initialization, with no reset at the window boundary.

The laws and their coordinates. Let σ_3 be uniform probability measure on $\{v \in \mathbb{R}^3 : \|v\| = \sqrt{3}\}$. For $j = 1, 2, 3$ in this coordinate order, define

$$p_j(v) = v_j^4 (a^\top v)^4, \qquad H_j = \mathcal{H}_8 p_j, \qquad \Gamma = \mathbb{E}_{\sigma_3}[HH^\top], \qquad h = \Gamma^{-1} H.$$

Here $\mathcal{H}_8$ is the homogeneous harmonic projection: $\mathcal{H}_8 p = \sum_{k=0}^4 c_k \|v\|^{2k} \Delta^k p$, where $c_0 = 1$ and $c_k = -c_{k-1}/[2k(17-2k)]$. The positive definite Gram matrix fixes the signs and scales without further normalization. Degree-eight harmonicity gives orthogonality to all polynomials of degree at most seven, and $K_{j\ell} = \mathbb{E}_{\sigma_3}[p_j h_\ell] = \delta_{j\ell}$. The compared laws are therefore uniquely specified by

$$dD_{\vartheta_\pm} = (1 + \vartheta_\pm^\top h)\, d\sigma_3, \qquad \vartheta_\pm = \pm\rho(1, -1, 0)^\top/\sqrt{2}.$$

For $h_j(v) = \sum_\alpha b_{j,\alpha} v^\alpha$, the implementation uses $B_j = \sum_\alpha |b_{j,\alpha}| 3^{|\alpha|/2}$ and $\rho = 0.9/(2\sqrt{\sum_j B_j^2})$, so both endpoint densities are positive. The exact rational coefficients, Gram matrix and coordinate order are in `data/tail_chart_coefficients.json` in the source package. The chart and output both have three coordinates; the experiment deliberately uses the same vector $v_1 = (1, -1, 0)/\sqrt{2}$ for the law direction and output projection. The target uses (78) at $q = 1$, with $K = I_3$ and $\nu_{1,j} = \frac{3}{5}(\|a\|^2 + 2a_j^2)$ throughout the variational flow. Its equilibrium limit is $B_1 = P_1 D_1 K$; the reported finite-window target retains the stated initializer and stabilizer.

Estimation and independent replication. An exact spherical-transport identity represents the finite-amplitude contrast as an expectation of complete-path directional derivatives. For each complete trajectory, a law amplitude is drawn uniformly along the registered chart segment; query directions are sampled conditionally on that amplitude. We use the construction's exchange symmetry and a zero-mean derivative moment control with coefficient $\alpha \simeq 1.0268665086$, frozen from a separate training sample before either replica is sampled. The optimizer uses two function values per update; the transport measurement additionally differentiates the known loss simulator and optimizer recurrence. Independent differentiation, finite-difference and state-resume checks support the complete-path derivatives and window identity.

R1 and R2 use disjoint random streams, each with 2,048 complete initialized trajectories; their independence is conditional on the shared frozen control, which is not refitted. The window, direction, target and tolerance $\tau = (25/100)\|T_{b,n}\|$ were fixed before sampling. Each replica and their pooled estimate must have its entire descriptive Student-t 95% interval for $v_1^\top(\mathcal{C}_{b,n} - T_{b,n})$ inside $[-\tau, \tau]$; the independent replica-difference Welch interval must also lie inside $[-\tau/2, \tau/2]$. Both replicas satisfy the predeclared outcome-independent hash-block, fourth-moment and individual-influence checks. The hash-block threshold uses an approximate chi-square calibration. The earlier shard-based inconclusive study used a different reliability design and remains separate from this preregistered sample set.

Table 4: Independent replication of the registered same-trajectory tail response over updates 1,179–13,911. Means and interval endpoints are in units of 10^{-3}; the target is 5.81819 and the prespecified tolerance is ± 1.45455 in these units. Intervals are descriptive 95% Student-t intervals. The pooled row aggregates R1 and R2.

Replica	Trajectories	Mean	95% interval
R1	2,048	6.38594	[6.31900, 6.45288]
R2	2,048	6.36083	[6.29080, 6.43086]
Pooled	4,096	6.37339	[6.32497, 6.42181]

Results and complete-prefix context. Both replicas and the pooled estimate satisfy the tolerance criterion (Table 4). The pooled response is 0.00637339, compared with the target 0.00581819, a positive excess of approximately 9.54%. The descriptive intervals resolve this remaining offset above the target. At the same final update, complete-prefix responses are 0.065560 and 0.064286 in R1 and R2, respectively. Their larger values are consistent with retained early contributions. This experiment concerns one registered directional window; empirical validation of the complete-prefix uniform response and information-radius limits in Theorem 2 remains open.

D.3.3 Angular response and first-order cancellation

The fixed angular grid resolves the continuous response predicted in Appendix A.8.2; the analytically extremal and cancellation pairs test its amplitude and the higher-order effect left when the first-order term vanishes. The data and algorithm are the original two-row seed, $q = 1$ and $\beta_1 = 0.9$. The design uses primary $\beta_2 = 0.999$ and held-out 0.9995. All eight grid points $\phi = k\pi/32$, $0 \le k < 8$, are retained at 0.999; $k = 0, 4$ reuse the original A/B roots. The extremal pair (ϕ_+, ϕ_-) and cancellation pair $(\phi_+ + \pi/16, \phi_+ + 3\pi/16)$ are evaluated at both memories. These are fourteen angular law–memory root records plus four reused A/B records. The twelve plotted designs share these endpoints; the design choices precede evaluation.

Enclosures and observable. The two order-24 interval engines in Appendix D.2 use the specified algebraic atoms, preserving their order and signs. On the frozen box, $|X_j| \le 3(\sqrt{6} + 0.005/4) < 9$ and $|\partial_p X_j| \le 3\|2b\|_2 = 3/2 < 2$, so the same input and derivative remainder bounds apply. The stored records give strict root inclusion and physical stability; the response intervals used here are outward hulls of the two engines. For the primary grid the ordinate is

$$Y(\phi) = \frac{\langle c_1, U_\phi - (U_A + U_B)/2 \rangle}{\delta r_a \|c_1\|^2}, \qquad r_a = 25.$$

This diagnostic center agrees with a zero-eighth-mode reference to leading order. Numerical markers use only independently enclosed own-root responses. The solid curve evaluates the analytic cosine rule.

β_2	Gap (10^{-6})	R_{opt}	Gap ratio to A/B	Cancellation (10^{-8})
0.999	[3.26965, 3.26989]	[1.01138, 1.01146]	[3.42745, 3.42854]	[3.97868, 4.00209]
0.9995	[1.62559, 1.62584]	[1.00567, 1.00582]	[3.49561, 3.49788]	[1.00107, 1.02444]

Table 5: Extremal and cancellation pairs at the primary and held-out memories. Bounds are outward rounded. The gap columns report the Euclidean own-root response separations of the extremal and cancellation pairs, respectively; the gap ratio divides the extremal separation by the original A/B separation at the same memory. Complete response and unit-direction enclosures are in `data/angular_records.json`.

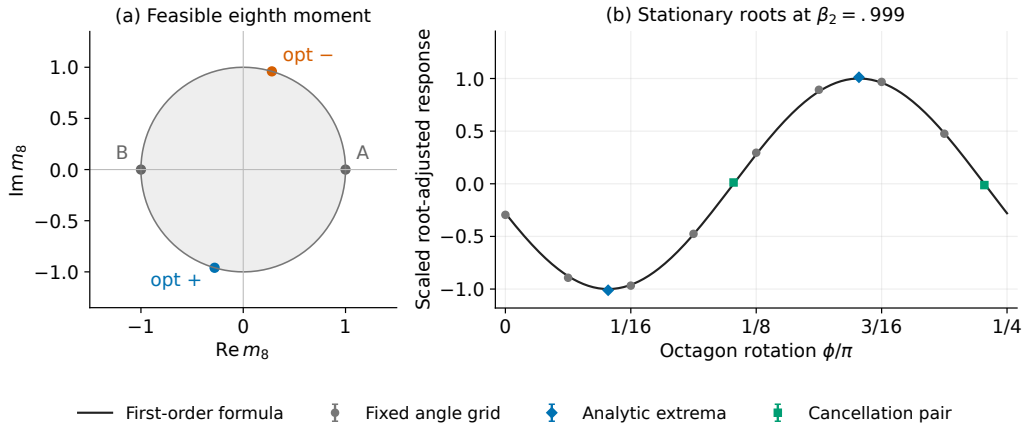


Figure 4: Angular response under matched estimator summaries. Left: the attainable eighth-moment disk, original A/B laws and analytically extremal pair. Right: all eight fixed grid points and four extremal/cancellation designs at $\beta_2 = 0.999$, compared with $\cos(8\phi - 4\alpha)$. The ordinate is $Y(\phi)$ above; interval bars propagate the response enclosures and are generally smaller than the markers.

For the extremal pair, both response and unit-direction gap intervals exclude zero. Its normalized amplitude is $R_{\mathrm{opt}} = \langle C_{\mathrm{opt}}, U_+ - U_- \rangle / (\delta \|C_{\mathrm{opt}}\|^2)$; the held-out interval is closer to the leading prediction one. At both memories, the measured separation ratio to the original A/B pair is below the asymptotic coefficient-norm ratio $25/7$. The cancellation pair has zero leading coefficient and a smaller, nonzero finite-memory response gap, resolving the higher-order effect left after first-order cancellation.

D.4 Numerical Resolution under Exact and Inexact Information

Increasing the supplied moment degree narrows the remaining stationary response uncertainty. With inexact moments, longer memory eventually exposes a precision floor. The following comparisons quantify these patterns on the fixed three-dimensional information classes of Section 4.3, using full-class upper enclosures and legal-pair lower bounds.

D.4.1 Information Classes and Computable Enclosures

The query batch is $q = 1$, $\beta_1 = 0.9$, and $K = 4m - 1$. We use $\mathcal{C}_{m,e}$ from Appendix B.3: probability laws on $\|v\| = \sqrt{3}$ whose normalized monomials through degree K differ from the circle–sign reference by at most e. At $e = 0$ the support identities reduce this to $\mathcal{C}_m$. All outputs below are zero-radius stationary own-root responses in the Euclidean norm at the fixed two-row seed. The public-anchor recursion (83) and its accompanying exact Stieltjes identity supply the finite calculation; the following bounds control its remainder and the movement to each law's root.

Lemma 30 (State-dependent forcing amplitude)**.** *For the exact class $\mathcal{C}_m$, let $Q_m(D; a)$ be* (51)*. Writing $c = a_1 - ia_2$, $A = |c|^{2m}$ and $B = 2m\sqrt{2}|a_3||c|^{2m-1}$, its attainable amplitude is*

$$\mathfrak{q}_m(a) = \sup_{D \in \mathcal{C}_m} |Q_m(D; a)| = 2^{1-2m} \begin{cases} A + B^2/(4A), & A > 0,\ B \le 2A, \\ B, & A > 0,\ B \ge 2A, \\ 0, & A = 0. \end{cases}$$

This bound is evaluated over the state box; the seed value Q_m^* alone does not suffice.

Proof. The moment body in Theorem 17 permits independent phases with $|\tau| + |\chi|^2 \le 1$. Aligning both phases reduces the maximum to $2^{1-2m} \max_{0 \le t \le 1}[A(1 - t^2) + Bt]$. Its maximizer is $t = \min\{B/(2A), 1\}$ for $A > 0$; both coefficients vanish otherwise. Negating the two moments attains the opposite endpoint. □

Cylinder comparison and perturbed Gram bound. Put $N = 4m$. For a polynomial $p = \sum c_\alpha v^\alpha$, write $\|p\|_3 = \sum_{\alpha\neq 0} |c_\alpha| 3^{|\alpha|/2}$, so every capped polynomial has expectation error at most $e\|p\|_3$. To bound the first unknown moment sharply, project to the cylinder. The second and fourth moment constraints give $0 \le \mathbb{E}_D(v_3^2-1)^2 \le 9e + 6e = 15e$. Set $R = \mathrm{sign}(v_3)$, with $R = 1$ at zero, and project the planar vector radially to radius $\sqrt{2}$; at a planar zero choose angle zero. This is measurable and

$$\|v - \widetilde{v}\| \le \sqrt{3/2}|v_3^2 - 1|, \qquad \mathbb{E}\|v - \widetilde{v}\| \le \sqrt{45e/2}.$$

The planar inequality follows by rationalizing the difference of the radii, and the vertical one from $||v_3| - 1| \le |v_3^2 - 1|$. For each of the finitely many polynomials $z^k = ((v_1 + iv_2)/\sqrt{2})^k$ ($1 \le k \le N-1$) and $v_3 z^k$ ($0 \le k \le N-2$), their projected low-mode residual is at most

$$e\|p\|_3 + L_p\sqrt{45e/2}, \qquad L_p = \sup_{\|v\|\le\sqrt{3}} \|\nabla p(v)\|.$$

One explicit upper bound for L_p is $\deg(p)\|p\|_3/\sqrt{3}$. Here $\widetilde{D}$ is the law of the projected point, and $Q_m(\widetilde{D}; a)$ denotes the same Fourier expression as (51). Let η be the largest of these bounds; $\eta = O_m(\sqrt{e})$. In particular $|\mathbb{E}R| \le \eta$, so both conditional probabilities are bounded away from zero for $\eta < 1/2$.

For the projected law write $\tau = \mathbb{E}z^N$ and $\chi = \mathbb{E}[Rz^{N-1}]$. For the projected Gram bound, put $r = |\chi|$, $\zeta = \mathbb{E}[Rz^N]$, and $t_\sigma = |\tau + \sigma\zeta|$ for the projected measure. The conditional four-function Gram matrix on $1, z, z^{N-1}, z^N$ differs in every entry by at most $4\eta/(1-\eta)$ from $\left(\begin{smallmatrix} I & B_\sigma \\ B_\sigma^* & I \end{smallmatrix}\right)$, where $B_\sigma = \left(\begin{smallmatrix} \sigma\chi & \tau+\sigma\zeta \\ 0 & \sigma\chi \end{smallmatrix}\right)$. Indeed the actual conditional modes are $(\mathbb{E}z^k + \sigma\mathbb{E}Rz^k)/(1+\sigma\mathbb{E}R)$; the low modes, the $N-1$ mode, and the N mode give these bounds directly, using $|\chi| \le 1$, $|\tau + \sigma\zeta| \le 2$. The operator error is at most $s = 16\eta/(1-\eta)$. Positivity therefore implies $\|B_\sigma\| \le 1 + s$. The exact singular-value identity $\|B_\sigma\| = (\sqrt{t_\sigma^2 + 4r^2} + t_\sigma)/2$ gives $t_\sigma + r^2/(1+s) \le 1 + s$. Average and use $|\tau| \le (t_+ + t_-)/2$ to obtain $|\tau| + r^2 \le 1 + 2s$. Maximizing the same quadratic as in Lemma 30 gives $|Q_m(\widetilde{D}; a)| \le \mathfrak{q}_m(a) + 2^{1-2m} 2As$.

Lemma 31 (Value and derivative tails at a public anchor)**.** *Suppose $X = X(p)$, $|X| \le B$, $|\partial_p X| \le D$, and the law is independent of the state. Assume $0 < \beta_2 < 1$ and $\mu_2 = \mathbb{E}_D X^2 > 0$, as on the validated domains below, and take public bounds $b \ge |\mu_2 - \nu|$, $b_p \ge |\partial_p(\mu_2 - \nu)|$, $R_c \ge \max(\mu_2, B^2 - \mu_2)$ and $V_c \ge \mathbb{E}_D(X^2 - \mu_2)^2$. Define $A_0 = 1$, $A_1 = 0$ and*

$$A_k = \frac{\sum_{j=0}^{k-2} \binom{k}{j} \beta_2^j \delta^{k-j} A_j R_c^{k-j-2} V_c}{1 - \beta_2^k} \quad (k \ge 2).$$

Set $L_{2j} = A_{2j}$, $L_{2j+1} = \sqrt{A_{2j}A_{2j+2}}$ and $T_k = \sum_{j=0}^k \binom{k}{j} b^{k-j} L_j$. For even $N \ge 2$, positive floors ν_-, λ_-, and a ceiling $\mu_{2,+} \ge \mu_2$, put

$$V_p = \left\{ T_{2N-2}\left[4D^2\mu_{2,+}\frac{\delta}{2-\delta} + b_p^2 \right] \right\}^{1/2}.$$

The response polynomial $H_N = \sum_{k<N} f_\lambda^{(k)}(\nu) J_k/k!$ has value and derivative errors bounded by

$$t_0 = \frac{BT_N}{\lambda_-\nu_-^N}, \tag{89}$$

$$t_p = \frac{(D/\lambda_- + BD/\lambda_-^2 + NB|\partial_p\nu|/(\lambda_-\nu_-))T_N + NBV_p/\lambda_-}{\nu_-^N}, \tag{90}$$

$$t_\lambda = \frac{2BT_N}{\lambda_-^2\nu_-^N}. \tag{91}$$

When $\nu = \mu_2$, take $b = b_p = 0$; in particular $A_2 = \delta V_c/(2-\delta)$. All quantities may be outward upper bounds computed from capped moments and support, rather than unknown higher moments of a target law.

Proof. For $W_c = \delta \sum_{i\geq 0} \beta_2^i (X_i^2 - \mu_2)$, expand the stationary recursion after separating its fresh input. Centering removes the size-one term, and $|\mathbb{E}(X^2 - \mu_2)^h| \leq R_c^{h-2} V_c$ for $h \geq 2$ gives $|\mathbb{E} W_c^k| \leq A_k$ by induction. Even moments are absolute moments; Cauchy–Schwarz gives the odd bounds. Since $V - \nu = W_c + (\mu_2 - \nu)$, the binomial inequality gives $\mathbb{E}|V - \nu|^k \leq T_k$.

The Stieltjes formula $f_\lambda(v) = \int_0^\infty \sqrt{s}\{\pi(\lambda^2 + s)(v + s)\}^{-1} ds$ and the finite geometric identity give the exact Taylor remainder with integrand $(-(V - \nu))^N / ((\nu + s)^N (V + s))$. Multiplying by $|A| \leq B$ gives t_0, including $V = 0$ by continuity. For a state derivative, the centered part of $\partial_p (V - \nu)$ is $\delta \sum_i \beta_2^i (2X_i X_i' - \partial_p \mu_2)$. Its second moment is at most $4D^2 \mu_2 \delta / (2 - \delta)$; its mean offset is bounded by b_p. Thus the derivative of the numerator power is bounded by NV_p. The remaining derivatives cost DT_N / λ_- from A, $NB|\nu'|T_N / (\lambda_- \nu_-)$ from the anchor, and BDT_N / λ_-^2 from the denominator: here $|V'| \leq 2D\sqrt{V}$ cancels the apparent singularity at zero. Dividing by ν_-^N proves t_p. The Stieltjes density's λ derivative is at most $2/\lambda_-$ times that density, giving t_λ. Since $\mu_2 > 0$ and $\beta_2 > 0$, an IID infinite history has $V > 0$ almost surely. The displayed integrable bounds justify differentiation under the expectation, including histories with arbitrarily small V. □

A computable projection-aware full-class radius. The fixed-row constraint removes part of a fixed-state response error when the root moves. The identity below retains the shared scalar uncertainty through this projection and bounds the stationary response over the public cylinder class.

Root adjustment for a computed center. Let the fixed-row equation be $LH_D(x_D) = b$ on a certified convex box, and let $\hat{U}$ be a public response center at $\hat{x}$ in that box. Apply (18) with Z replaced by L and $\kappa \mathbf{1}$ by b. The segment-averaged derivative belongs to the convex interval enclosure of DH_D on the box. An interval inverse therefore encloses the own-root response error, including $r = L\hat{U} - b$, whenever every averaged row derivative is invertible.

Finite polynomial information with a shared scalar uncertainty. At a fixed state, write $\mu_j = \mathbb{E}X_j = a_j, \nu_j = \mathbb{E}X_j^2, Y_j = X_j^2 - \nu_j, W_j = V_j - \nu_j$ and, suppressing j, $r_k = \mathbb{E}Y_j^k$, $s_k = \mathbb{E}[X_j Y_j^k]$. The capped information fixes μ_j, ν_j. Use the general fresh-input recursion (83), with $M_k = \mathbb{E}W_j^k$ and $J_k = \mathbb{E}[A_j W_j^k]$, $M_0 = 1$ and $J_0 = s_0$. Here the true-second-moment anchor gives $r_1 = 0$ and $s_0 = \mu_j$; these simplifications must not be imposed on noisy fibers. The recursion conditions on the current IID query in both updates, so replacing J_k by $\mu_j M_k$ would lose the shared-history dependence.

For $k < m$, both r_k and s_k agree with the reference, because their input degrees are at most $2m - 1$. At $k = m$,

$$r_m - r_m^{\mathrm{ref}} = \sigma_j Q_m(D; a), \qquad (\sigma_1, \sigma_2, \sigma_3) = (1, (-1)^m, 0).$$

The same Q_m enters all coordinates. Its exact attainable radius is $\mathfrak{q}_m(a)$, with $\mathfrak{q}_m$ given in Lemma 30. That quadratic maximum increases with A, B, so their outward upper bounds give a public enclosure without target-law information.

Bounds for every remaining input moment and derivative. For $a_\perp = (a_1, a_2)$, legitimate coordinatewise support bounds are

$$B_j = |a_j| + \|a_\perp\| + \sqrt{2}|a_3| \quad (j = 1, 2), \qquad B_3 = |a_3| + \sqrt{2}\|a_\perp\|.$$

The same formulas applied to $\partial_p a$ give $D_j \geq |\partial_p X_j|$. For interval states use their outward suprema. Let $R \geq \max(\nu, B_j^2 - \nu)$ and let v bound $\mathbb{E}Y^2$. For $m \geq 3$ this variance is matched and public; for $m = 2$ it is at most the reference variance plus $\mathfrak{q}_2(a)$ in coordinates one and two, and is matched in coordinate three. Set $L_0 = 1$, $L_1 = \sqrt{v}$ and $L_k = R^{k-2} v$ for $k \geq 2$. Then $\mathbb{E}|Y|^k \leq L_k$, $|s_k| \leq B_j L_k$, and $|r_k| \leq L_k$; for even k, $r_k \geq 0$. With $D_Y = 2B_j D_j + |\partial_p \nu|$, differentiation of these moments gives

$$|\partial_p r_k| \leq k D_Y L_{k-1}, \qquad |\partial_p s_k| \leq D_j L_k + k B_j D_Y L_{k-1}.$$

Known moments and their derivatives are evaluated exactly as public polynomials instead of using these loose bounds; in particular $r_1 = 0$ and $\partial_p r_1 = 0$ exactly. The third coordinate also retains its matched r_m. These rules produce finite interval inputs to (83) and its differentiated recursion.

Infinite-history and derivative tails. Apply Lemma 31 coordinatewise with $(B, D, \mu_2, R_c, V_c) = (B_j, D_j, \nu, R, v)$ and $b = b_p = 0$. Choose even $N \geq 2$ and compute its majorants A_k through $2N-2$; write $\mathcal{M}_k = A_k$. With outward state-box floors ν_-, λ_- and ceiling $\mu_{2,+} = \nu_+$, define $(e_0, e_p, e_\lambda) = (t_0, t_p, t_\lambda)$ from (89)–(91). For $c_k = f_\lambda^{(k)}(\nu)/k!$, these respectively bound the value, physical p derivative and λ derivative errors of $\sum_{k<N} c_k J_k$. This evaluates the same infinite-history remainder without a history truncation.

Conservative affine propagation. Represent a scalar uncertainty as $a+bQ+E$, with $|Q| \leq Q_{\max}$ and E an interval. Addition and multiplication by a known interval are immediate. For two such quantities, retain constant a_1a_2 and linear coefficient $a_1b_2 + a_2b_1$; enclose the remaining terms by

$$a_1E_2 + a_2E_1 + E_1E_2 + (b_1E_2 + b_2E_1)[-Q_{\max}, Q_{\max}] + b_1b_2[0, Q_{\max}^2].$$

This encloses the quadratic terms of the exact product. Use reference moments for the constants, the common signed Q for r_m, and the above support intervals for all further unknowns. Every denominator in (83) is known and positive. Thus a finite sequence of these operations produces, at $\hat{x}$,

$$H_D(\hat{x}) - \hat{U} \;\in\; cQ + \mathbb{E}_h, \qquad |Q| \leq \mathfrak{q}_m(a(\hat{x})), \tag{92}$$

including the specialized bound (89) and outward rounding in $\mathbb{E}_h$. The center $\hat{U}$ is a rounded public-reference polynomial response. Its full-response approximation is paid by the same tail bound; it is not replaced by an unknown true root. Increasing N or arithmetic precision changes computational accuracy, not the public information cap.

Theorem 32 (Public root and response certification)**.** *Assume every compatible H_D is continuously differentiable on a neighborhood of the box below, and let $\mathbb{E}$ be a convex interval matrix enclosure of all $DH_D(x)$ for compatible D and x in that box. Suppose an invertible numerical preconditioner C and a closed convex box $B = \hat{x} + [-r_x, r_x]$ validate, uniformly over the entire public information class, the inclusion*

$$\hat{x} - C(LH_D(\hat{x}) - b) + (I - CL\mathbb{E})(B - \hat{x}) \subset \operatorname{int} B$$

and a weighted infinity-norm bound $\|I - CL\mathbb{E}\| < 1$. Then each compatible law has a unique root in B. If the full physical slow Jacobian enclosure there has negative trace and positive determinant, these two-dimensional roots are attracting. Let $\mathbb{P}$ enclose $I - E(LE)^{-1}L$ and $\mathbb{A}$ enclose $E(LE)^{-1}$ for every $E \in \mathbb{E}$, with inverse existence validated. Let $\mathcal{E}$ enclose $H_D(\hat{x}) - \hat{U}$ for every compatible law and $r = L\hat{U} - b$. A certified response radius is

$$R = \sup_{P \in \mathbb{P}, e \in \mathcal{E}} \|Pe\| + \sup_{A \in \mathbb{A}} \|Ar\|.$$

If $\mathcal{E} = c[-Q_{\max}, Q_{\max}] + \mathbb{E}_h$, as in (92)*, this gives the computable bound*

$$R_{\text{cert}} = Q_{\max} \sup_{P \in \mathbb{P}} \|Pc\| + \sup_{P \in \mathbb{P}, e \in \mathbb{E}_h} \|Pe\| + \sup_{A \in \mathbb{A}} \|Ar\|. \tag{93}$$

Every supremum on the right can be upper bounded by elementary interval matrix multiplication followed by a Euclidean norm bound. If two legal endpoint own-root responses have distance at least $g > 0$, then $\mathrm{CalRatio} = R_{\text{cert}}/(g/2)$ is a certified competitive factor relative to the full-class Euclidean minimax radius.

Proof. For each law, $x \mapsto x - C(LH_D(x) - b)$ is a contraction of the complete box into itself. Its unique fixed point is a root because C is invertible. Negative trace and positive determinant give local stability of the full physical Jacobian, including the derivative of the mass vector in $a(p)^T H_D$. Apply the segment identity (18) with the fixed rows L and right side b, then take suprema over the derivative and response-error enclosures. The triangle inequality gives the scalar specialization; for the exact cylinder take $Q_{\max} = \mathfrak{q}_m(a(\hat{x}))$. The bound includes root residual, derivative uncertainty, series tail and rounding. Any enclosing ball has radius at least $g/2$ for a legal pair of gap g, which gives the competitive factor; the upper bound covers the full class. □

Direction error. The public cap $K = 4m - 1$ fixes the information order independently of the target law. Provided $\hat{U} \neq 0$, a response-norm floor $u_- > 0$ gives unit-direction error at most $\min\{2, 2R_{\text{cert}}/\max(u_-, \|\hat{U}\|)\}$, using outward floors. A zero center does not certify a predicted unit direction. An unsupported root or projection domain gives no certificate, rather than a zero radius; a valid radius need not attain a requested tolerance.

The full-sphere class with inexact moments. We certify the reference-centered spherical box $\mathcal{C}_{m,e}$, whose feasibility witness is the reference law. A different center needs its own feasibility check. All bounds use outward interval arithmetic; floating-point root locators supply centers and preconditioners.

Deterministic anchor and shared symbols. At each public state $x = (p, \lambda)$ set $n_j = \mathbb{E}_{\mathrm{ref}} X_j^2$, $Y_j = X_j^2 - n_j$ and $d_j = \mathbb{E}_D Y_j$. For every capped polynomial $f = \sum c_\alpha v^\alpha$, use

$$\mathbb{E}_D f = \mathbb{E}_{\mathrm{ref}} f + e \sum_{\alpha \neq 0} c_\alpha 3^{|\alpha|/2} \xi_\alpha, \qquad |\xi_\alpha| \leq 1.$$

One symbol is shared wherever the same normalized moment occurs. The interval box is an outer relaxation of realizable moment vectors. In particular d_j is retained, and $\mathbb{E}_D X_j$ is computed from the second-moment symbols. Only the *reference* identities $\mathbb{E}_{\mathrm{ref}} Y_j = 0$ and $\mathbb{E}_{\mathrm{ref}} X_j = a_j$ are simplified exactly. Their state derivatives are simplified before interval evaluation as well.

For the first unmatched input moment, there is an explicit polynomial L_j of degree at most $4m - 2$ such that on the cylinder

$$X_j^{2m} = s_j H_m(v; a) + L_j(v; a), \qquad (s_1, s_2, s_3) = (1, (-1)^m, 0),$$

where $c = a_1 - i a_2$, $z = v_1 + i v_2$ and

$$H_m = 2^{1-4m} \Re(c^{2m} z^{4m}) + 2m 2^{2-4m} a_3 v_3 \Re(c^{2m-1} z^{4m-1}).$$

An explicit construction expands $X_j^{2m} - s_j H_m$ and replaces v_3^{2r} by 1 and v_2^{2r} by $(2 - v_1^2)^r$, retaining the odd factors. The two leading harmonics cancel, so the degree drops to $4m - 2$. This is the polynomial form of the Fourier identity. Writing $h = \sqrt{45e/2}$, projection to the cylinder yields

$$\mathbb{E}_D X_j^{2m} = s_j Q + \mathbb{E}_D L_j + \epsilon_j, \quad Q = Q_m(\widetilde{D}; a), \quad |\epsilon_j| \leq \{\mathrm{Lip}(X_j^{2m}) + \mathrm{Lip}(L_j)\} h.$$

Indeed evaluate the cylinder identity under $\widetilde{D}$ and compare only X_j^{2m} and L_j under the two measures. The same scalar Q is kept in all three coordinates. Its bound is the smaller of the unconditional $2^{1-2m}(A + B)$ and the perturbed-Gram bound in the cylinder comparison above when $\eta < 1/2$. No exact support identity is imposed on D itself.

Higher moments and derivative bounds. On the sphere, valid pointwise bounds are

$$B_j = \tfrac{3}{2}(\|a\| + |a_j|), \quad D_j = \tfrac{3}{2}(\|a'\| + |a_j'|), \quad L_{X_j} = \sqrt{3}(\|a\| + |a_j|).$$

The first two follow from the extremal eigenvalues of $(e_j a^T + a e_j^T)/2$. Let R bound $|X^2 - n|$ on the sphere and let v bound $\mathbb{E}_D Y^2$. Then $\mathbb{E}_D |Y|^k \leq R^{k-2} v$ for $k \geq 2$, $\mathbb{E}_D |Y| \leq \sqrt{v}$, and $|\partial_p Y| \leq 2BD + |n'|$. These bound all uncapped $\mathbb{E}_D Y^k$, $\mathbb{E}_D X Y^k$ and their derivatives, including noncentered Y.

For sharper bounds as $e \to 0$, intersect them with cylinder comparisons. If R_c, B_c, D_c are the exact cylinder bounds, $L_Y = 2BL_X$, then $v_c = v + 2RL_Y h$ bounds $\mathbb{E}_{\widetilde{D}} Y^2$, and

$$\mathbb{E}_D |Y|^k \leq R_c^{k-2} v_c + k R^{k-1} L_Y h.$$

For XY^k, replace the right side by $B_c R_c^{k-2} v_c + (L_X R^k + kBR^{k-1} L_Y) h$. Derivative versions follow by differentiating the integrands: $\partial_p Y^k = k Y^{k-1} Y'$ and $\partial_p (XY^k) = X' Y^k + k X Y^{k-1} Y'$. Their Lipschitz constants are bounded by the product rule with $L_{Y'} = 2(L_X D + B L_{X'})$. All these inequalities compare expectations only; the full sphere remains the pointwise support used in the infinite-history tail.

Shared-history evaluation and its remainder. Use the exact recurrences (83), with $r_k = \mathbb{E}_D Y^k$, $s_k = \mathbb{E}_D X Y^k$, $M_0 = 1$ and $J_0 = s_0$. They split the same fresh input from both filters and retain r_1. The response polynomial is $\sum_{k<N} f_\lambda^{(k)}(n) J_k / k!$. Represent each expression by a constant, a linear form in the shared moment symbols, and an interval remainder; multiplication encloses all products beyond the linear terms, without assuming independence of symbols.

Lemma 31, with anchor $\nu = n$, completes the infinite-history enclosure. Take $b \geq |\mathbb{E}_D X^2 - n|$, $b_p \geq |\partial_p (\mathbb{E}_D X^2 - n)|$, and true-centered full-sphere bounds R_c, V_c. For any even $N \geq 2$, the allowances t_0, t_p, t_λ retain these offsets even when they dominate the memory fluctuations. Cylinder comparisons sharpen expectations; the tail's pointwise support remains the full sphere.

Validity of the noisy certificate. The preceding moment and tail calculations enclose the response and its full derivative throughout the proposed box, uniformly over $\mathcal{C}_{m,e}$. Apply Theorem 32 with the resulting shared-symbol error set $\mathcal{E}$. The theorem's inclusion and invertibility conditions establish each root and its own-root response enclosure; its Jacobian conditions establish local physical stability. Project each shared-symbol coefficient before taking norms, then include the projected remainder and numerical row residual. The row equation $ZU_D = \kappa\mathbf{1}$ gives a positive response-norm floor; the bound in the direction-error paragraph gives direction errors. Monotonicity of log and logit supplies logarithmic state bounds from the same root box. These conclusions require the root inclusion, inverse and normalization conditions above. Competitiveness additionally requires legal lower witnesses, as in the common theorem; no initialized trajectory budget follows from these stationary enclosures.

Arithmetic precision. The reported grid uses 80-digit outward interval arithmetic, including rounding error in geometric denominators evaluated by subtraction. For very small rational gaps, cancellation can widen these enclosures and prevent root certification at this precision. The factored identity $1-(1-\delta)^k = \delta\sum_{j=0}^{k-1}(1-\delta)^j$ offers an alternative evaluation; higher precision may also be needed. The reported bounds apply to the evaluated grid.

The frozen full-class computation uses the public-anchor majorants through $2N-2$, including the noncentral value and derivative offsets. This differs from the named-law implementation remainder in the preceding finite-memory computation subsection, which uses its own coarser derivative bound.

D.4.2 Information and finite-precision effects

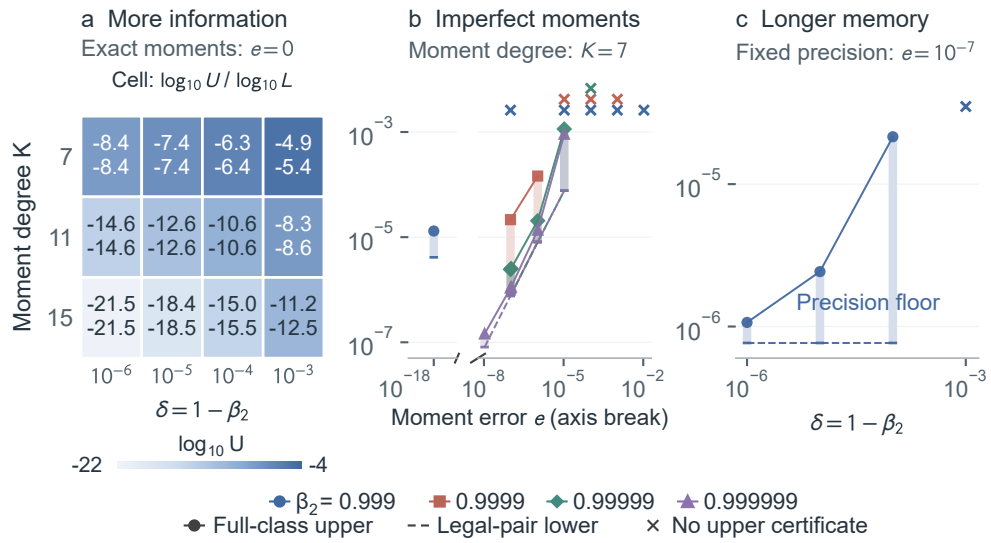


Figure 5: More information improves prediction; finite precision limits it. Compatible query laws leave a range of possible responses. (a) The discrete certificate map reports $\log_{10}$ upper/lower bounds (rounded to one decimal); higher K shrinks ambiguity. (b) Moment error at $K = 7$ limits resolution; the broken log axis separates the $e = 10^{-18}$ check without omitting grid cells. (c) At fixed $e = 10^{-7}$, the legal-pair lower bound reveals a precision floor. Solid and dashed curves are deterministic full-class upper and legal-pair lower bounds; crosses mark cells without a public upper certificate. Lines connect tested grid values only. The source package contains all plotted values.

Complete values and certificate statuses are supplied in `data/precision_diagnostic_tables.tex`, with exact sources in `data/precision_evidence_index.json`.

D.4.3 AUGMENTED-INFORMATION ENCLOSURES AT PAIRED READING TOLERANCES

The augmented-information calculation compares exact readings with tolerance $\xi = \delta$ at each fixed moment and memory configuration. The source records contain 48 configurations paired into 24 comparisons, with whole-fiber upper enclosures, legal-pair lower bounds and specified-law prediction errors. At the nonzero tolerance, $\delta^{m-1}\xi = \delta^m$, so the reading allowance is on the same scale as the next-order residual in the theoretical bound.

Public information and numerical center. The fiber is $\mathcal{F}_{m,e}(0, \xi)$ from Appendix B.3.2, with $m \in \{2, 3\}$, $q = 1$, $\beta_1 = 0.9$, and the original two-row seed. All monomials in $v/\sqrt{3}$ through degree $K = 4m - 1$ have coordinatewise error at most e. The extra reading is the global degree-$4m$ polynomial ψ_m, normalized by the seed projected coefficient and Q_m^* as in that appendix; its reported center is zero. The circle–sign reference witnesses feasibility for every displayed fiber. Coefficient interval error enlarges the reading width to $\xi + \epsilon_\psi$. The public reference own-root response supplies the prediction center, with its numerical error paid as in Appendix B.3.2.

Augmented-fiber enclosure. The evaluator retains a separate shared symbol for each first unknown raw input moment at the population seed, as well as the supplied low-degree moments. It bounds their motion to the reference state and the remaining moments on the full sphere. A validated public root domain encloses the averaged response derivative and the root-adjustment projector. After projection, the degree-$4m$ term is removed using the supplied expectation of ψ_m; the finite-memory coefficient and projector differences remain in the error budget. Thus the upper radius is the sum of five computable allowances: descriptor width, remaining first-unknown coefficients, supplied-moment uncertainty, higher-order/state/tail terms, and the center's root residual. The latter includes the complementary projection of any difference between the recomputed Taylor center and the localizer's reference center. This retains the numerical-center allowance rather than assuming exact row feasibility. The same public calculation can use either a validated domain for the unaugmented superset or one for the augmented fiber; it selects the smaller valid radius, paying each candidate's own center residual. The extra-reading reduction is therefore applied explicitly, in addition to the unaugmented method of Appendix D.4.

Meaning of the three bounds. For each valid configuration, U encloses the entire augmented fiber about the public numerical center. The lower bound L is half a certified response gap between two legal laws in that same fiber. For the specified finite law set, the two independent root engines give outward response hulls; T is the maximum of the resulting upper bounds on distance to that center. The complete records, including certificate statuses and missing endpoints, are supplied in `data/augmented_full_records.csv`; `data/augmented_pair_map.json` specifies the pairing.

E ADDITIONAL RESPONSE-GEOMETRY EVIDENCE ACROSS STABLE INSTANCES

E.1 CONFIGURATIONS, DATA INSTANCES AND QUERY-LAW COORDINATES

The primary grid contains $(d, r) = (3, 1)$ and, for each $d \in \{4, 6, 8, 10\}$, the four row ranks $r \in \{1, d/2, d-1, d\}$. There are 17 analytic seeds. Three preassigned, independently seeded Gaussian matrix perturbations per seed are normalized to Frobenius norm 10^{-5}, yielding 51 perturbation units and 68 total data instances. The same data instance is used at $\beta_2 = 0.999$ and 0.9999; these two uses are configurations, not independent random datasets. The original screening checks full row rank, positive masses and stabilizer, nonzero aggregate coordinates, population-root residuals, and a positive stability margin, without using response comparisons. The scope is this local neighborhood of stable seeds.

The model and local response are those of Sections 3 and 4.1. The regime is $m = 2$, $q = 1$, $\beta_1 = 0.9$. A harmonic-density family on $\|v\|^2 = d$ preserves all moments through degree seven and varies degree-eight moments. The density coefficient chart uses the full standard basis $C = I_d$; its first unmatched moment map K is nonidentity and invertible. The secant amplitude is one eighth of the declared density-positivity radius, fixed separately at each dimension. It equals $715/1401919488$ for $d = 3$, $35/979632128$ for $d = 4$, $7/8285825664$ for $d = 6$, $15/251591131136$ for $d = 8$, and $33/4305650000000$ for $d = 10$.

Of the 136 primary configurations, 104 have positive predicted rank and 32 are analytic zero-rank controls. The secondary operator analysis also includes $\beta_2 = 0.99$ at the canonical $(4, 1)$, $(6, 3)$, $(10, 9)$ and $(4, 4)$ instances.

E.2 THEORETICAL OPERATOR, FINITE SECANTS AND RANK INFORMATION

The operator in Theorem 1 acts on hidden moments. The experiment's density parameters therefore require the explicit map

$$\theta(D_t) - \theta(D_0) = Kt, \qquad B_{\mathrm{moment}} = P_0 D_2, \qquad M_{\mathrm{chart}} = B_{\mathrm{moment}} K C, \qquad C = I_d. \tag{94}$$

Here D_2 denotes the diagonal $m = 2, q = 1$ forcing coefficient in (10); it is not a query-law name. All norms use the fixed standard Euclidean metrics. The independently solved axis responses form

$$S_{\delta,h} e_j = [U(D_{he_j}, \delta) - U(D_{-he_j}, \delta)]/(2h). \tag{95}$$

The recorded operator error compares $\delta^{-1} S_{\delta,h}$ with M_{chart} and normalizes by $\|M_{\mathrm{chart}}\|_{\mathrm{op}}$. The source records in `data/rank_configurations.csv` report the numerical upper value, with response-box and theoretical-operator uncertainty included, on positive-rank configurations. A zero operator does not receive an artificial relative-error denominator.

The theoretical projector P_0 is the oblique equilibrium-adjustment projector. The matrix $\Pi = I - Z^\top (Z Z^\top)^{-1} Z$ is an orthogonal representation of the row constraint. These matrices have different roles; substituting Π for P_0 would change the prediction.

The experiment reports a constraint-aware response-dimension endpoint. The numerical lower activation count k_- records singular directions separated from the frozen floor after accounting for the response-box radius, while $k_A = d - r$ is the analytic upper rank after the common row and response contracts are checked: $ZU = \kappa \mathbf{1}$ implies $Z S_{\delta,h} = 0$. The endpoint is resolved when these two counts agree. The raw-box count k_{box} and midpoint effective rank k_{eff} are retained as auxiliary numerical diagnostics. The latter uses a roundoff floor plus four response-box uncertainty radii. Exact algebraic rank is not inferred from either auxiliary thresholded count.

For the 32 primary $r = d$ controls, $U = Z^{-1} \kappa \mathbf{1}$ forces the secant to vanish identically.

E.3 RESOLUTION LIMITS ACROSS MEMORY SCALES

At the two displayed memories, activation reaches the predicted dimension in $44/52$ positive-rank configurations at $\beta_2 = 0.999$ and in all 52 at $\beta_2 = 0.9999$ (Table 6). The eight shorter-memory cases are the four instances of $(10, 5)$ and $(10, 9)$; they lie at the enclosure floor and are unresolved rather than counted as rank zero.

Table 6: Resolution of the predicted response dimension. Positive-rank rows combine numerical nonzero activation with the analytic $d - r$ upper bound; invertible-row controls have analytically zero response secants.

Primary subset	Rank resolved	Unresolved
Positive rank, $\beta_2 = 0.999$	44/52	8
Positive rank, $\beta_2 = 0.9999$	52/52	0
Analytic zero-rank controls	32/32	0

Per-configuration activation counts and secondary operator-error diagnostics are supplied in `data/rank_configurations.csv`. The repository retains the complete operator-error analysis and its numerical checks.

E.4 Numerical assurance for response and operator calculations

Response boxes are assembled from exact rational root-certificate endpoints. Singular-value calculations were checked at 100 and 180 decimal digits, together with reconstruction and orthogonality residuals. The activation count includes only singular directions separated from the response-box uncertainty and numerical floor; the common row equation supplies the analytic upper rank. These are high-precision numerical spectral checks, with exact interval assembly.

E.5 Reading geometry and actual named-law evidence

The axis-secant study characterizes local response geometry. For a direction v, $S_{\delta,h}v$ is a linear combination of axis secants and is not, in general, the independently evaluated mixed-law response difference

$$\frac{U(D_{hv}, \delta) - U(D_{-hv}, \delta)}{2h}.$$

A readings-only predictor therefore additionally requires mixed-direction truth roots and a remainder linking those responses to the secant combination. The six canonical named-law comparisons detailed in Appendix D.2.1 provide actual legal equal-reading response tests. They have exact matching readings and independently enclosed response gaps at $(d, r) = (3, 1), (4, 2), (6, 3)$ and $\beta_2 = 0.999, 0.9999$, with law and root identities specified in Appendix D.2.1.

F CONNECTIONS TO FIXED STABILIZERS AND SAMPLED INFORMATION

F.1 AN EXACT FIRST UPDATE

Proposition 33 (Fixed-stabilizer insufficiency)**.** *For $F(w) = \|w\|^2/2$, $w_0 = e_1$ and the original sixteen-atom octagon laws D_A, D_B, every fixed $\epsilon, \mu > 0$ and $\beta_1, \beta_2 \in [0, 1)$ give different expected first normalized Adam updates, despite identical estimator mean, full covariance and MSE.*

Proof. The centered two-value estimator is exactly $g = vv^\top e_1$ for any $\mu > 0$. Octagon matching gives $\mathbb{E}g = e_1$, $\mathrm{Cov}(g) = \mathrm{diag}(1/2, 1/2, 1)$ and MSE 2. From zero raw filters, the usual bias corrections (Kingma & Ba, 2015) give $\widehat{m}_0 = g$, $\widehat{v}_0 = g^{\odot 2}$. For $R_\epsilon = g/(|g| + \epsilon)$, symmetry cancels coordinates two and three. Under A, $v_1^2 = 0, 1, 2$ with probabilities $1/4, 1/2, 1/4$; under B, $v_1^2 = 1 \pm 1/\sqrt{2}$ with equal probabilities. Hence

$$\mathbb{E}_A R_{\epsilon,1} = \frac{2\epsilon + 3}{2(\epsilon + 1)(\epsilon + 2)}, \quad \mathbb{E}_B R_{\epsilon,1} = \frac{2\epsilon + 1}{2\epsilon^2 + 4\epsilon + 1},$$

$$\mathbb{E}_A R_\epsilon - \mathbb{E}_B R_\epsilon = \frac{-e_1}{2(\epsilon + 1)(\epsilon + 2)(2\epsilon^2 + 4\epsilon + 1)}.$$

At $\epsilon = 1$ the means are $(5/12)e_1$ and $(3/7)e_1$; displacement is $-\eta_0 R_\epsilon$. □

This first-step result is independent of memory. It does not extend the interior-root theorem to fixed ϵ: setting $\kappa = 0$ leaves $a^\top H_{0,q} = \sum_j a_j^2/(\sqrt{\nu_{q,j}} + \lambda) > 0$ for $a \neq 0$. The accompanying symbolic and two-function-value checks verify the algebra and implementation.

F.2 SAMPLING THE ATTAINING READINGS

Use the hypotheses and public quantities of Theorem 1, with $B = \mathsf{B}_{m,q}$, $k = \operatorname{rank} B$, and N, ϕ from (12). For n_s IID directions from D, set $\widehat{U}_{n_s} = U_q^\circ + \delta^{m-1} N \bar{\phi}_{n_s}$, where $\bar{\phi}_{n_s} = n_s^{-1} \sum_i \phi(V_i)$ and U_q° is the finite-memory reference own-root response.

Corollary 34 (Sampled recovery)**.** *If $|\phi_\ell(v)| \leq R_\ell$ and $R_\Sigma^2 = \sum_\ell R_\ell^2$, then for $k \geq 1$ and each allowed law, with probability at least $1 - \alpha$,*

$$\|\widehat{U}_{n_s} - U_{D,q}\|_2 \leq C\delta^m + \delta^{m-1}\sqrt{2R_\Sigma^2 \log(2k/\alpha)/n_s}. \tag{96}$$

With exact public quantities, the mean squared error is exactly (13)*. For $k = 0$, use U_q° with deterministic $O(\delta^m)$ error and zero variance.*

Proof. Hoeffding's inequality (Hoeffding, 1963) bounds coordinate error by $R_\ell\sqrt{2\log(2k/\alpha)/n_s}$ with failure probability α/k. Union bounding and the isometry of N give (96). The centered sampling error has zero mean, so its cross term with the deterministic residual vanishes; $NN^\top B = B$ gives the covariance trace in (13). Coordinates may share samples. Constants are uniform across laws, with probability interpreted under each law separately. □

The sphere bound $|X_j| \leq d\|a^\circ\|$ gives $R_\ell = 2(d\|a^\circ\|)^{2m}\|(N^\top B)_{\ell,:}\|_1$. Thus $n_s \geq 2R_\Sigma^2 \log(2k/\alpha)/(c^2\delta^2)$ suffices for stochastic error at most $c\delta^m$. This retains the data and projector scale; it is not a sharp function-query bound.

Implementation and base information. Reference-response error e_0 and correction-vector error e_c add $e_0 + \delta^{m-1}e_c$ to (96); e_c includes operator, reference-moment, polynomial and summation errors. All readings share the same directions, and $N\bar{\phi}_{n_s} = B(\bar{M} - \mathbb{E}_{D^\circ}\bar{M})$: full-moment projection is the same estimator. Evaluating public polynomials needs no objective calls. Reference roots, public operators, base moments and truth calculations are separate costs from an optimizer's $2qT$ calls.

For the separate 3D moment box in Appendix B.3, its pointwise envelope $B_{\psi,m}$ gives $\xi = B_{\psi,m}\sqrt{2\log(2/\alpha)/n_s}$ in the deterministic bound $C_m(e + \delta^m + \delta^{m-1}\xi)$. The expectation range

$[-1,1]$ is not a pointwise bound. For P normalized base polynomials in $[-1,1]$, an exact reference and n_b IID samples give

$$e_{\mathrm{conf}} = \max_{j\le P} |\widehat{b}_j - b_j^\circ| + \sqrt{2\log(2P/\alpha_b)/n_b}$$

with confidence $1-\alpha_b$. An estimated reference adds its own width; split failure probability between base and aligned readings and respect the tolerance caps. Reducing the base confidence width to $O(\delta^m)$ has sufficient sample scale $\delta^{-2m}\log(P/\alpha_b)$, not a statistical lower bound.

Tracking cost. The bridge bound (80) retains fixed-memory constants depending on input/derivative bounds, stabilizer and second-moment floors, root inverse and contraction margins, history forgetting, and Poisson/martingale estimates. Factors such as $(1-\sqrt{\beta_2})^{-1}$ preclude treating them as memory-independent. Since $S_T \sim 4\varepsilon T^{1/4}$, the sufficient condition $S_T \gtrsim \delta^{-1}$ corresponds to update scale $(\varepsilon\delta)^{-4}$, with ε still constrained by tracking. This is a sufficient-condition interpretation, not a practical lower bound; the distinct tail measurement and its resolved bias are discussed in Appendix D.3.2.

F.3 Positive-radius own-root responses

Retain the original exponential loss, annealed stabilizer, fixed octagon laws D_A, D_B, $q=1$ and $0 \le \beta_1 \le \bar{\beta}_1 < 1$. Replace X_0 in both stationary filters by $X_{\mu,j} = v_j \sum_i p_i(x)\sinh(\mu z_i^\top v)/\mu$, holding the radius fixed throughout the shared IID history.

Proposition 35 (Octagon response at positive radius)**.** *At the stable seed of Appendix A.8, both laws have unique Hurwitz roots on a common local branch for sufficiently small $\delta>0$, $\mu \ge 0$. Uniformly there,*

$$U_{A,\delta,\mu} - U_{B,\delta,\mu} = \delta C_{\mathrm{seed}} + O(\delta^2 + \delta\mu^2 + \mu^4), \qquad C_{\mathrm{seed}} = P_0(h_{A,0} - h_{B,0})(x_0) \neq 0. \quad (97)$$

Thus $\mu = o(\delta^{1/4})$ preserves the leading signal and $\mu = O(\sqrt{\delta})$ preserves the $O(\delta^2)$ remainder.

Proof. The coefficient of $\mu^{2\ell}$ in X_μ is $v_j \sum_i p_i(x)(z_i^\top v)^{2\ell+1}/(2\ell+1)!$, of query degree $2\ell+2$. Finite support gives uniform convergence with state derivatives on a fixed compact neighborhood. Seven-degree matching therefore cancels the mean terms through μ^4, joint second-moment terms through μ^2, and cubic constant term. With $\Delta = \mathbb{E}_A - \mathbb{E}_B$,

$$\Delta X_\mu = O_{C^1}(\mu^6), \quad \Delta(X_\mu X_\mu^\top) = O_{C^1}(\mu^4), \quad \Delta X_{\mu,j}^3 = O_{C^1}(\mu^2). \quad (98)$$

Choose the neighborhood with positive λ and second-moment floors, preserved at small μ by continuity. Input and state-derivative bounds are uniform, and law probabilities are state-independent. Hence the cutoff proof and Theorem 9 have common constants and give $H_{D,\delta,\mu} = F_{D,\mu} + \delta h_{D,\mu} + O_{C^1}(\delta^2)$, where $F_{D,\mu,j} = (\mathbb{E}_D X_{\mu,j}) f_\lambda(\mathbb{E}_D X_{\mu,j}^2)$. Equation (98) gives $F_{A,\mu} - F_{B,\mu} = O_{C^1}(\mu^4)$; the explicit coefficient (29) and $X_\mu - X_0 = O_{C^1}(\mu^2)$ give $h_{D,\mu} - h_{D,0} = O_{C^1}(\mu^2)$. Thus

$$H_A - H_B = \delta\Delta h_0 + O_{C^1}(\delta^2 + \delta\mu^2 + \mu^4), \qquad H_D = H_0 + O_{C^1}(\delta + \mu^2). \quad (99)$$

Since $-\widehat{g}_\mu/L = X_\mu$, the physical drift retains its original form and root equation $ZH_D = \kappa\mathbf{1}$. If J_0 is the population physical Jacobian, $x \mapsto x - J_0^{-1}\mathcal{G}_D(x)$ is a contraction mapping a fixed small closed ball into itself for small δ, μ. It gives unique roots $x_D = x_0 + O(\delta+\mu^2)$, with Hurwitz Jacobians by continuity. On the connecting segment, $E_A = \int_0^1 D_x H_A(x_B + t(x_A - x_B))dt = E_0 + O(\delta+\mu^2)$. Invertibility of ZE_0 implies $P_A = P_0 + O(\delta+\mu^2)$. Substituting (99) into the exact root identity (19), and using $\Delta h_0(x_B) = \Delta h_0(x_0) + O(\delta+\mu^2)$, proves the claim. Constants depend on the fixed domain, data, laws, floors, derivative bounds, inverse and stability margins, and $\bar{\beta}_1$, not on δ, μ. □

Single-law perturbation. The same branch satisfies $\|U_{D,\delta,\mu} - U_{D,\delta,0}\| \le C\mu^2$ uniformly in δ. Indeed, coupling histories gives $|A_\mu - A_0| \le K\mu^2$ and, by the reverse triangle inequality in weighted ℓ_2, $|\sqrt{V_\mu} - \sqrt{V_0}| \le (\sum_k \delta(1-\delta)^k |X_{\mu,-k} - X_{0,-k}|^2)^{1/2} \le K\mu^2$. The stabilizer floor gives a uniform C^0 field difference. Segment-averaged physical Jacobians stay close to J_0 and have uniformly bounded inverses, transferring this bound to roots; the uniform state Lipschitz bound then transfers it to responses. At positive radius estimator moments are close, not exactly matched. The result concerns stationary histories, and fixed positive μ need not preserve the memory signal as $\delta \to 0$.

F.4 INDEPENDENT SAMPLING CHECK

We used the frozen sphere instance $(d, r, m, q) = (3, 1, 2, 1)$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and its `direction2_minus/plus` laws. Interval certificates supply independent finite-memory truth and the uniform-law reference; no roots or coefficients were fitted to samples. Public $a^\circ = \mathbf{1}$, $\lambda^\circ = 1$ give $M_j = (v_j \sum_i v_i)^4$ and $B = \frac{1}{4} f''(3)(I - \mathbf{1}\mathbf{1}^\top/3)$, $f(s) = (1 + \sqrt{s})^{-1}$. Two aligned readings span $\mathbf{1}^\perp$. The same-count comparator reads $\sum_j M_j$ and $M_2 - M_1$, then uses public minimum-norm reconstruction and B; it misses the pair's $(1, 1, -2)$ direction.

We froze five sample sizes $64, 256, 1024, 4096, 16384$ and 256 independent replicates per law and size. Rejection sampling uses the frozen harmonic density and public $3/2$ envelope, with independent random streams and no nested samples. Exact sphere polynomial integration supplies the projected variance 0.09213. Figure 6 compares actual errors and the variance identity.

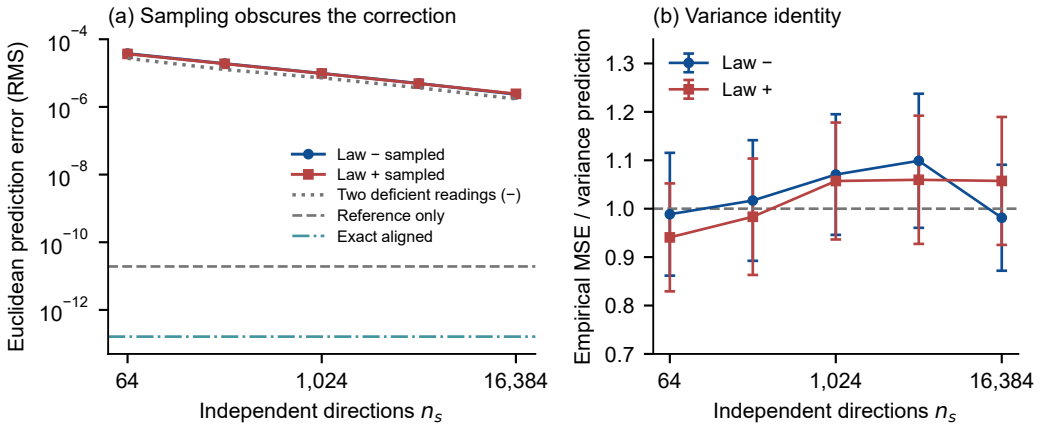


Figure 6: Independent sampling: RMS prediction errors against certified truth (left) and empirical MSE divided by the variance-identity prediction (right). Bars are 1.96 Monte Carlo standard errors, not certified coverage intervals. The deficient estimator has less variance but omits a necessary response direction.

Reference-only and exact-aligned errors are 1.924×10^{-11} and 1.625×10^{-13}; combined reference/truth enclosure radii are below 8.73×10^{-23}. Sampled RMS remains 2.35–2.44×10^{-6} at the largest size, so recovery is unresolved in this grid. Across ten cells, MSE/variance prediction is 0.941–1.099. The identity extrapolates baseline and residual-floor crossovers near 2.49×10^{14} and 3.49×10^{18} directions; neither is an executed budget or a statistical lower bound. Same-sample full-moment projection and aligned reconstruction agree to 6.71×10^{-20}.

The sampling seed, independent-stream construction and reading maps are specified in `data/sampled_reading_protocol.json`; the law coefficients and harmonic basis are given in `data/READING_PROTOCOL.md` and its linked protocol. The Monte Carlo stage uses floating-point arithmetic; imported root enclosures account for truth uncertainty but do not certify the sampling computation.

Appendix F.5 gives a complementary SNR-aware angular family resolved at a finite frozen sample size.

F.5 SNR-AWARE FINITE-SAMPLE RECOVERY ON A MATCHED ANGULAR MIXTURE

The low-SNR example above shows that a small reading count need not make estimation easy. We now consider a legal family with an observable response. Let D_+, D_- be the analytically extremal rotated octagons of Appendix A.8.2, and define

$$D_t = \tfrac{1+t}{2} D_+ + \tfrac{1-t}{2} D_-, \qquad -1 \le t \le 1.$$

Here $D_0 = (D_+ + D_-)/2$ is the mixture reference, distinct from the unrotated octagon D_A. Every mixture matches query moments through degree seven and hence the zero-radius estimator mean, full covariance and MSE. At the public seed, the normalized eighth-harmonic reading $\psi(\theta) = \cos(8\theta - 4\alpha)$, $\alpha = \arg(a_1 + ia_2)$, equals $+1$ on D_+ and -1 on D_-. Thus $\mathbb{E}_{D_t}\psi = t$ and $\mathrm{Var}_{D_t}\,\psi = 1 - t^2$; at $t = \pm 9/10$ this is $19/100$, strictly positive.

Predictor and protocol. C_{opt} is the endpoint-pair separation coefficient. The reference-centered correction is therefore half of it:

$$\widehat{U}_n = U_{D_0} + \delta \frac{C_{\mathrm{opt}}}{2} \bar{\psi}_n, \qquad \delta = 1 - \beta_2 = 1/2000 \quad \text{at held-out memory.} \tag{100}$$

U_{D_0} is the finite-memory own-root response. Only development roots at $\beta_2 = 0.999$ selected $|t| = 9/10$ and $n = 16$. The fixed-$q = 1$, $\beta_1 = 0.9$ evaluation used independently enclosed truths at $\beta_2 = 0.9995$, with 1024 independent paired seed clusters in each of two disjoint phases. Within a cluster, the two target-sign gains are averaged before constructing the two-sided Student-t 95% interval. The displayed sample-size grid uses nested prefixes within clusters; these are not additional independent replicates. There were no exclusions or significance-based stopping.

Table 7: Held-out primary endpoint, $n = 16$, $t = \pm 9/10$. Gain is reference MSE minus sampled MSE; gain columns are in units of 10^{-13}.

Phase	Mean paired gain	95% interval	MSE ratio
Formal	5.276003	$[5.269477, 5.282529]$	0.01419
Replication	5.273740	$[5.268185, 5.279296]$	0.01462

Both signs have positive point estimates, and both phase intervals remain positive after truth-interval sensitivity. Empirical/predicted MSE is 0.968–1.006 at the primary endpoint, consistent with exact binomial expectation. The same-count orthogonal eighth harmonic vanishes on these supports and misses the response; its MSE is about 68–71 times the aligned MSE. Same-sample reconstruction of all nine degree-eight monomials followed by projection agrees with direct ψ to 1.22×10^{-15}. At $t = 0$ sampled correction gives no false gain; $t = \pm 1$ controls have zero variance and agree with exact readings. The cancellation pair has zero leading eighth-mode coefficient and a higher-order finite-memory gap.

Protocol and scope. The target, sample size, endpoint, and inferential criteria were fixed from development evidence. Formal and replication used fresh disjoint streams under a corrected frozen protocol; complete protocol and amendment histories are archived with the code. These direction samples estimate a public reading for stationary prediction on this named family; they do not constitute optimizer objective-call savings, training improvement, or a uniform black-box sample-complexity guarantee.

SUPPLEMENTARY REFERENCES

Feng Dai and Yuan Xu. Spherical harmonics. In *Approximation Theory and Harmonic Analysis on Spheres and Balls*, pp. 1–27. Springer New York, New York, NY, 2013.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Zai Yang and Lihua Xie. Frequency-selective vandermonde decomposition of Toeplitz matrices with applications. *Signal Processing*, 142:157–167, 2018.