BEYOND GRADIENT FLOW: IDENTIFIABILITY AND RECOVERY FROM DISTRIBUTION SNAPSHOTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Inferring dynamics from snapshots of evolving distributions is fundamentally underdetermined: the Fokker–Planck equation constrains the drift F only through its score-weighted divergence $\nabla \cdot F + F \cdot \nabla \log \rho$, leaving a ρ-solenoidal gauge invisible to any single-time constraint. Time-indexed transport formulations cannot resolve this ambiguity: every admissible marginal path admits a curl-free explanation, minimum-action reconstruction selects it, and marginal fit alone cannot distinguish dynamically inequivalent explanations. Requiring one autonomous field to explain several marginals instead makes part of the hidden circulation visible as $\nabla \log \rho$ changes across marginals. Separating instantaneous Fokker–Planck source constraints from the snapshot experiment, we show that the source constraints identify the field modulo the kernel of a stacked score-weighted divergence operator. For generic Gaussian shape variation, source constraints at $K \geq m$ time points in intrinsic dimension m eliminate every polynomial gauge direction, whereas finitely many density snapshots alone admit aliasing; we give the obstruction explicitly. At a Gaussian anchor, for Sobolev smoothness s and n samples per time point, we derive a conditional lower rate $(nK)^{-2s/(2s+m+1)}$ for the tangent snapshot experiment, with a matching upper rate in a degreewise benchmark. Strong-form fitting is non-orthogonal to score error and cannot be repaired by spectral filtering. Instead, we estimate using smooth test functions while retaining the known diffusion term, and derive a finite-sample bound that separates sampling error from fixed-grid quadrature bias. Planted-circulation experiments confirm the predicted gauge contraction and expose a design tension between cross-slice information and covariance-aware whitening.

1 INTRODUCTION

Inferring dynamical systems from *snapshots* of their time-evolving distributions is a central problem in fields ranging from single-cell biology to climate science. Suppose a latent state Z_t evolves according to

$$dZ_t = F_t(Z_t)\, dt + \sqrt{2\sigma^2}\, dW_t, \tag{1}$$

with W_t a standard Brownian motion and $\sigma \geq 0$ a *known, fixed* diffusion scale. Only independent samples from the marginals $\rho_{t_1}, \ldots, \rho_{t_K}$ are observed, and the goal is to recover the drift F_t. The problem is underdetermined: the Fokker–Planck equation constrains only the gradient component of the weighted Helmholtz decomposition $F_t = -\nabla\Phi_t + h_t$, with scalar potential Φ_t and $\nabla\cdot(\rho_t h_t) = 0$, leaving the weighted-solenoidal field h_t as a time-local gauge freedom.

At stationarity the gauge is precisely the probability current $J = \rho h$ separating equilibrium from nonequilibrium steady states: $h = 0$ is detailed balance, $h \neq 0$ permits persistent circulation without changing the stationary marginal. Such fluxes are widespread in living systems (Battle et al., 2016; Gnesotto et al., 2018), motivate landscape–flux descriptions of cell fate (Wang et al., 2008; Kwon et al., 2005), and drive irreversible Langevin samplers (Hwang et al., 1993; Lelièvre et al., 2013; Rey-Bellet & Spiliopoulos, 2015), game dynamics (Balduzzi et al., 2018) and stochastic optimization (Kunin et al., 2024). We ask whether it can be inferred when the dynamics are unobserved.

Existing methods resolve the underdetermination by imposing a selection principle — transport cost, a reference process, a coupling, or an action (Schiebinger et al., 2019; Tong et al., 2020; Huguet

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.

et al., 2022; Bunne et al., 2022; Terpin et al., 2024; De Bortoli et al., 2021; Shi et al., 2023; Albergo & Vanden-Eijnden, 2023; Lipman et al., 2023; Tong et al., 2024; Lee et al., 2025; Neklyudov et al., 2023; Benamou & Brenier, 2000). These select a representative compatible with the marginals rather than identifying the circulation; within an unrestricted *time-indexed* class every admissible marginal path admits a curl-free explanation (§2.1), so marginal fit cannot distinguish true circulation from a gauge choice.

The situation changes under *autonomy*, $F_t \equiv F$: one field must now explain all observed marginals. A direction invisible under one density ρ_t may become visible as its score $\nabla \log \rho_t$ changes, so the persistent gauge contracts to directions invisible across all source constraints. We formalize this through Stein operators, ask what becomes structurally identifiable, how the remaining near-gauge directions control statistical recovery, and how that structure dictates an estimator.

1. **Source constraints and snapshot identifiability** (§2). We separate known *sources* and finitely many *snapshots*. For a time-indexed drift, minimum-action reconstruction selects the gradient representative (Cor. 2). Under autonomy, the residual gauge under the source constraints consists of directions lying in every per-time Stein kernel and is contracted pointwise by changing scores (Prop. 6). Snapshots identify strictly less: this shared gauge is neither necessary nor sufficient there, and we give the aliasing obstruction (Ex. 5).
2. **Statistical limits and a conditional rate** (§3). For generic Gaussian shape variation, source constraints at $K \geq m$ time points eliminate every finite-degree polynomial gauge direction in intrinsic dimension m (Thm. 8). Linearizing the snapshot map around a reference system, we derive a conditional lower rate for the tangent snapshot experiment, with a matching upper rate in a degreewise benchmark (Thm. 12).
3. **A gauge-aware estimator, and a design tension** (§4, §5). Strong-form estimation is non-orthogonal to score error (Prop. 14) and spectral filtering cannot separate the resulting bias from signal (Prop. 35). A weak formulation removes the nuisance and, under independent slices, yields moment equations with exact block-tridiagonal covariance (Prop. 16). Empirically we find that the regime in which that covariance structure would pay is the regime in which the target ceases to be identifiable (Obs. 17).

Positioning. Trajectory inference is usually posed as transport between temporal marginals and closed by an optimal-transport, reference-process, flow-matching, or action-based principle. These construct dynamics consistent with prescribed distributions; we ask which aspects of the dynamics those distributions determine. Consistency results typically assume gradient drift (Lavenant et al., 2024) and action matching explicitly targets a gradient representative (Neklyudov et al., 2023); rotational ambiguity from population snapshots was noted by Weinreb et al. (2018). The closest identifiability result, Guan et al. (2024), characterizes linear additive-noise SDEs through generalized rotational symmetries of the initial law; we treat an infinite-dimensional field class, condition on the observed marginal shapes rather than the initial law, and study finite-sample recovery (App. I). Velocity, pairing and lineage measurements supply Lagrangian information we deliberately exclude (App. R). App. B gives the full discussion.

2 What can be identified from evolving marginals

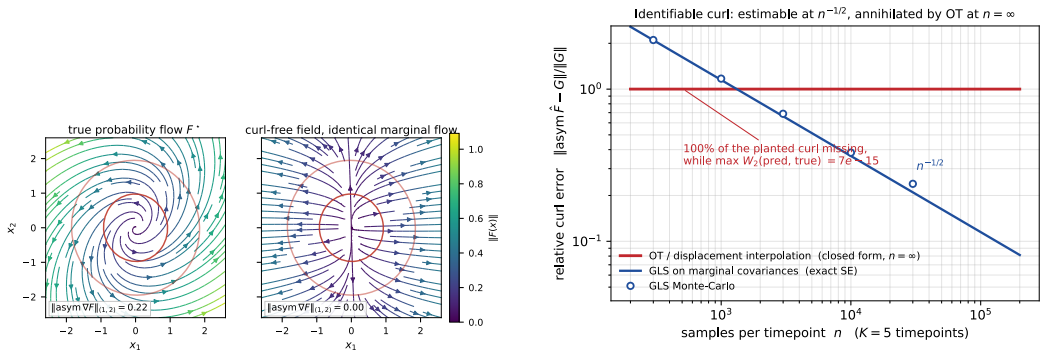
Marginal evolution does not reveal a vector field directly, only its effect on probability mass. A Stein operator captures this effect. We use it to separate source constraints from the snapshot experiment and to characterize time-indexed blindness. Proofs are in App. C.

Let ρ_t be the marginal of (1) and $s_t = \nabla \log \rho_t$ its score. Dividing the Fokker–Planck equation by ρ_t gives

$$\mathcal{A}_{\rho_t} F_t = b_t, \quad \mathcal{A}_\rho F := \nabla \cdot F + F \cdot \nabla \log \rho, \quad b_t = -\partial_t \log \rho_t + \sigma^2 \big(\nabla \cdot s_t + \|s_t\|^2\big). \tag{2}$$

The operator $\mathcal{A}_\rho$ is the vector Stein operator associated with ρ; we call $\mathcal{A}_\rho F$ the *footprint* of F on ρ. We fix σ throughout; joint drift–diffusion identifiability is discussed in App. X. A vector field h is invisible at ρ exactly when its footprint vanishes, $\ker \mathcal{A}_\rho = \{h : \nabla \cdot (\rho h) = 0\}$; we call such fields *$\rho$-solenoidal*[1].

[1]Throughout, *weighted-solenoidal* or *ρ-solenoidal* denotes this exact mathematical condition; *circulation* refers to its dynamical interpretation, while *antisymmetry* is used only in linear settings.



(a) (b)

Figure 1: **Identical marginal flow, different dynamics.** (a) A flow with a planted rotational component and a curl-free transport generate the same Gaussian marginals. (b) On the reference model of Tab. 1, linearized-GLS decays at the $n^{-1/2}$ rate whereas displacement interpolation has solenoidal relative error 1 for every n, including $n = \infty$. Both match the observed marginals to machine precision, so marginal fit cannot distinguish them.

Two information levels must be distinguished. At the *source-level*, one knows $(\rho_t, \partial_t \rho_t)$ at the times under consideration, so (2) is an observed population constraint; a smooth complete marginal path provides these constraints for every t. In the *snapshot experiment* one observes only independent samples from $\rho_{t_1}, \ldots, \rho_{t_K}$, so the temporal source is not directly available.

2.1 Time-indexed saturation and architectural blindness

Optimal transport and related models fit a separate field F_t at each time, letting every slice choose its own representative of the single-slice gauge.

Proposition 1 (Gradient saturation)**.** *Suppose $\{\rho_t\}$ is a smooth path of positive densities such that each ρ_t satisfies a Poincaré inequality and $\partial_t \log \rho_t \in L^2(\rho_t)$. Then for every t there is $\phi_t \in H^1(\rho_t)/\mathbb{R}$ with $\nabla \cdot (\rho_t \nabla \phi_t) = -\partial_t \rho_t$ weakly, so the path solves the continuity equation driven by the curl-free velocity $v_t = \nabla \phi_t$. The corresponding Itô drift $F_t = v_t + \sigma^2 s_t$ is also curl-free.*

Zero circulation is therefore unfalsifiable from marginal fit within a time-indexed gradient class, however strongly the underlying system rotates.

Corollary 2 (Minimum-action projection)**.** *At each t, among all velocities meeting the same marginal constraint, the minimum-$L^2(\rho_t)$ action solution lies in $(\ker \mathcal{A}_{\rho_t})^\perp = \overline{\{\nabla \phi\}}$ and has zero ρ_t-solenoidal component.*

Optimal transport is the closed-form instance: for Gaussian marginals the Benamou–Brenier displacement velocity is a gradient on every segment, so its solenoidal relative error equals 1 at every sample size while the induced marginal path matches the observed one exactly (Cor. 18, App. C). The same population floor holds for conditional flow matching with exact OT couplings; Schrödinger bridges add only a gradient correction to the reference drift and hence preserve its ρ_t-solenoidal projection (Cor. 19). Fig. 1 shows both halves of the blindness: identical marginal transport and an error floor that no amount of data removes.

This blindness is architectural: a time-conditioned curl head is compensable slice by slice, making circulation unidentifiable even given oracle nuisances (Prop. 21). Autonomy instead forces one compensator to satisfy the constraints across all selected times.

2.2 Autonomy, snapshot equivalence, and the residual gauge

This motivates an autonomy hypothesis at the representation level: the chosen state is sufficiently informative that a single field can explain the entire marginal evolution.

Assumption 3 (Representation-level autonomy)**.** On the chosen m-dimensional state space $\mathcal{M}$, the observed evolution is generated by a single time-independent Itô drift $F : \mathcal{M} \to T\mathcal{M}$.

Proposition 4 (Source-constraint equivalence). *Two autonomous fields satisfy the same source constraints (2) iff their difference lies in* $\ker \mathcal{A}_K := \bigcap_{k=1}^K \ker \mathcal{A}_{\rho_{t_k}}$, *where* $\mathcal{A}_K F := (\mathcal{A}_{\rho_{t_k}} F)_{k=1}^K$. *For a continuum of source times the intersection is taken over the whole interval; with the same initial law and forward uniqueness, this is equivalent to generating the same complete marginal path.*

Changing marginals can reveal directions hidden at a single slice: if $\mathcal{A}_{\rho_{t_1}} h = 0$ then $\mathcal{A}_{\rho_{t_k}} h = h \cdot (s_k - s_1)$, so cross-slice score variation creates a footprint for otherwise invisible directions.

Finite snapshots are weaker. Their population observation map is nonlinear, $\mathscr{S}_K(F, \rho_0) = (\rho^F_{t_1}, \dots, \rho^F_{t_K})$, and its *fibres*, not $\ker \mathcal{A}_K$, are the observational equivalence classes; an unknown initial law is a further nuisance. In a finite-dimensional parameterization, full column rank of the derivative of a continuously differentiable snapshot-statistic map at the truth gives local injectivity by the inverse function theorem applied to a full-rank square submap, but neither implies global injectivity nor follows from instantaneous score rank.

Example 5 (Temporal aliasing). Let $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $F_\omega(x) = (-aI + \omega J)x$ with $a > 0$, and $\rho_0 = \mathcal{N}(0, \Sigma_0)$ with Σ_0 anisotropic. Then $\Sigma_t = e^{-2at} R_{\omega t} \Sigma_0 R_{\omega t}^\top + \frac{\sigma^2}{a}(1 - e^{-2at}) I$ with $R_\theta = e^{\theta J}$. At $t_k = k\Delta$, the drifts F_ω and $F_{\omega + \pi/\Delta}$ give *identical* Gaussian snapshots, yet their difference $(\pi/\Delta) Jx$ is *not* in $\ker \mathcal{A}_{\rho_0}$.

Thus $\ker \mathcal{A}_K$ is neither sufficient for snapshot equivalence — vanishing footprints at the selected instants leave the path unconstrained in between — nor necessary, by Ex. 5. It characterizes source-level identifiability, whereas local snapshot identifiability is governed by $D\mathscr{S}_K$ (§3.2).

Unlike the time-indexed model, autonomy is falsifiable. If one autonomous field explains the slices but no autonomous gradient field does, then non-gradient structure is required within the specified model. The same cross-slice restriction therefore creates both falsifiability and additional identifiability.

Autonomy is a property of the chosen representation: failure of one field to explain several slices may instead signal missing dynamical state rather than genuinely time-varying dynamics. App. W gives both cases at the population level, including one where every stationary snapshot is compatible with a zero-rotation drift although the underlying system circulates. We return to this representation-level tradeoff in App. X.

Although not the finite-snapshot equivalence class, $\ker \mathcal{A}_K$ is the *residual gauge* of the stacked source constraints and the structural benchmark below. Its geometry has a pointwise reduction.

Proposition 6 (Residual fibre dimension). *Define* $\mathcal{C}_x := \operatorname{span}\{s_k(x) - s_1(x) : k = 2, \dots, K\}^\perp \subseteq T_x\mathcal{M}$ *and* $r(x) := \operatorname{rank}[s_2(x) - s_1(x), \dots, s_K(x) - s_1(x)]$. *Then* $\ker \mathcal{A}_K = \{h : \mathcal{A}_{\rho_{t_1}} h = 0,\ h(x) \in \mathcal{C}_x\ \rho_{t_1}\text{-a.e.}\}$, *with* $\dim \mathcal{C}_x = m - r(x)$.

The relevant quantity is the score-rank $r(x)$, not K alone, and at stationarity all constraints coincide. The count $m - r(x)$ is a *fibre* dimension only: the unknown is a square-integrable section, so even low-dimensional fibres support infinitely many such fields. A fixed finite-dimensional model class replaces this infinite-dimensional freedom by a finite-dimensional rank problem, though nonidentifiability can remain through null directions or temporal aliasing (App. I).

3 Source identifiability and tangent statistical limits

Section 2 separated the stacked *source constraints* from the finite snapshot map $\mathscr{S}_K$. We now characterize the Gaussian source gauge and the statistical recovery of locally visible directions in the tangent snapshot experiment. The lower bound is matched in a degreewise benchmark; transfer to sampled snapshots is not claimed. We work in intrinsic dimension $m \geq 2$. Proofs are in Apps. D–I, with ambient reduction, manifold concentration and the stationary specialization in Apps. J–L.

3.1 Gaussian shape change contracts the weighted gauge

Fix comparison slices $\bar\rho_{t_1}, \dots, \bar\rho_{t_K}$ and an anchor $t_\star$ such that, after centering and whitening, $\rho_\star := \bar\rho_{t_\star} = \mathcal{N}(0, I_m)$; in §3.2 these are the slices of the reference system at the observation times. Let

$\{\Psi_\beta\}$ be the normalized probabilists' Hermite basis, $\mathcal{P}_q$ the degree-q chaos, $D_q := \dim \mathcal{P}_q = \binom{q+m-1}{m-1}$, $W_q := \mathcal{P}_{q-1} \otimes \mathbb{R}^m$, and $S_q := \ker(\mathcal{A}_{\rho_\star}|_{W_q})$.

Proposition 7 (Anchor-weighted gauge). *$\mathcal{A}_{\rho_\star}(\Psi_\beta e_i) = -\sqrt{\beta_i + 1}\,\Psi_{\beta+e_i}$ for every multi-index β and coordinate i. Hence $\mathcal{A}_{\rho_\star} : W_q \to \mathcal{P}_q$ is onto and, for $q \geq 2$, $\dim S_q = mD_{q-1} - D_q$, with $\dim S_q / \dim W_q \to (m-1)/m$ as $q \to \infty$.*

Thus a single Gaussian anchor slice leaves most high-degree directions in its weighted gauge. With $\bar{s}_{t_k} := \nabla \log \bar{\rho}_{t_k}$ and the stacked reference operator $\bar{\mathcal{A}}_K h := (\mathcal{A}_{\bar{\rho}_{t_k}} h)_k$, the cross-slice footprint of $h \in S_q$ is $T_{K,q} h := (h \cdot (\bar{s}_{t_k} - s_\star))_{k \neq \star}$, and $\ker T_{K,q} = S_q \cap \ker \bar{\mathcal{A}}_K$.

For an explicit criterion, suppose further that the comparison slices are Gaussian with a common mean. After a common translation and whitening, write $\bar{\rho}_{t_k} = \mathcal{N}(0, \Sigma_k)$ with $\Sigma^\star = I_m$, and define the precision contrasts $\Xi_k := \Sigma_k^{-1} - I_m$. Then $T_{K,q} h = -(h(x)^\top \Xi_k x)_{k \neq \star}$.

Theorem 8 (Gauge contraction by shape change). *Let $\Xi_1, \dots, \Xi_r$ be the distinct precision contrasts among the comparison slices, and call the shapes* shape-rich *if $\operatorname{span}\{x, \Xi_1 x, \dots, \Xi_r x\} = \mathbb{R}^m$ for x in a nonempty open set. With $\mathcal{S}_{\leq Q} := \bigoplus_{2 \leq q \leq Q} S_q$, shape richness implies $\mathcal{S}_{\leq Q} \cap \ker \bar{\mathcal{A}}_K = \{0\}$ for every $Q \geq 2$. Generically, shape richness holds iff $r \geq m - 1$, and this threshold is sharp: for $r = m - 2$ the nonzero field $w(x) := \star(x \wedge \Xi_1 x \wedge \cdots \wedge \Xi_{m-2} x)$ lies in $\mathcal{S}_{\leq m} \cap \ker \bar{\mathcal{A}}_K$. Hence, with distinct generic comparison shapes, $K \geq m$ is necessary and sufficient to eliminate every finite-degree polynomial gauge.*

Thm. 8 is proved in App. E. Below the threshold the residual polynomial gauge can remain macroscopic: for $m = 3$ it contains the Euler rigid-body field $x \times \Xi_1 x$, and its degreewise size is quantified in App. E.1. More generally, with $K \leq m - 1$ generic shapes, the first surviving polynomial gauge appears at degree K (App. E.1). At $q = 2$, $S_2 = \{Gx : G^\top = -G\}$, and invisibility is equivalent to $[G, \Sigma_k^{-1}] = 0$ for every k. Thus one generic comparison shape eliminates the linear gauge; the threshold $K \geq m$ is imposed by higher polynomial degrees (cf. Guan et al. (2024); App. I).

3.2 A CONDITIONAL TANGENT BENCHMARK

Visibility to the source constraints does not determine how strongly a direction appears in snapshots. Linearize around a known reference $(\bar{F}, \bar{\rho}_t)$ whose observed slices are those of §3.1, with anchor $\bar{\rho}_{t_\star} = \rho_\star$ and fixed initial law. For $F^\delta = \bar{F} + \delta g$, write $\rho_t^\delta = \bar{\rho}_t(1 + \delta u_t) + o(\delta)$, defining the relative-density perturbation u_t. Let $\mathcal{L}_t$ be its linearized Fokker–Planck operator and $\mathcal{U}(t, s)$ the propagator of $\partial_t u = \mathcal{L}_t u$ (App. F). Then $\partial_t u_t = \mathcal{L}_t u_t - \mathcal{A}_{\bar{\rho}_t} g$ with $u_0 = 0$, and

$$u_{t_k} = -(\mathcal{B}_K g)_k, \qquad (\mathcal{B}_K g)_k := \int_0^{t_k} \mathcal{U}(t_k, s)\, \mathcal{A}_{\bar{\rho}_s} g \, ds. \tag{3}$$

Thus $\mathcal{B}_K = -D\mathscr{S}_K(\bar{F})$. Unlike $T_{K,q}$, $\mathcal{B}_K$ integrates source footprints along the reference path, so $\ker \mathcal{B}_K$ and $\ker \bar{\mathcal{A}}_K$ need not be nested: between-slice visibility and Duhamel cancellation can both occur (App. F). An unknown initial law would add a propagated nuisance term.

Let M_K map the stacked tangent densities to the retained snapshot statistics, let Γ_K be the covariance of their $\sqrt{n}$-scaled fluctuations, and set $\widetilde{\mathcal{B}}_K := \Gamma_K^{-1/2} \mathsf{M}_K \mathcal{B}_K$. The whitened *tangent snapshot experiment* observes $Y = \widetilde{\mathcal{B}}_K g + n^{-1/2} \xi$ with $\mathbb{E}\xi = 0$ and $\operatorname{Cov}(\xi) = I$. At the anchor, write the degree-q anchor-weighted Helmholtz split $W_q = G_q \oplus S_q$, with $G_q := S_q^\perp \cap W_q = \nabla \mathcal{P}_q$, and decompose $g_q = g_{q,\mathrm{grad}} + h_q$ accordingly. Treating the gradient part as an unrestricted nuisance and profiling it out degree by degree gives

$$\mathcal{J}_{\mathrm{sol},q} := \big(\widetilde{\mathcal{B}}_K|_{S_q}\big)^* (I - \Pi_{q,\mathrm{grad}}) \big(\widetilde{\mathcal{B}}_K|_{S_q}\big), \tag{4}$$

where $\Pi_{q,\mathrm{grad}}$ projects onto the range of $\widetilde{\mathcal{B}}_K|_{G_q}$. Each operator in (4) is self-adjoint and nonnegative on S_q; let $\kappa_{q,1}^2 \geq \cdots \geq \kappa_{q,r_q}^2 \geq 0$ be its eigenvalues, zeros included, so that $r_q = \dim S_q \asymp q^{m-1}$ by Prop. 7.

To align tangent with source identifiability we assume, on the degrees considered, the *faithfulness* condition $\ker \mathcal{J}_{\mathrm{sol},q} \subseteq \ker T_{K,q}$: tangent propagation creates no invisible direction beyond the source gauge. Under the shape-rich setting of Thm. 8 it makes $\mathcal{J}_{\mathrm{sol},q}$ injective on S_q, so all r_q

eigenvalues are positive and the tangent experiment targets exactly the circulation made visible by Thm. 8. The rate theorem below does not use it.

Assumption 9 (Block comparability)**.** Let $\mathcal{J}_{\mathrm{sol},\le Q}$ be the information on $\mathcal{S}_{\le Q}$ obtained by profiling the gradient directions of degree at most Q jointly, with all components above degree Q set to zero, and let $\mathcal{J}_Q^{\mathrm{blk}} := \bigoplus_{q=2}^{Q} \mathcal{J}_{\mathrm{sol},q}$. There is $C < \infty$, independent of Q, with $\mathcal{J}_{\mathrm{sol},\le Q} \preceq C\, \mathcal{J}_Q^{\mathrm{blk}}$.

Joint profiling therefore cannot raise information beyond the degreewise benchmark by more than a constant, which suffices for the lower bound; the reverse comparison can fail (App. F.4) and is not used. Let $\{e_{q,j}\}_{j=1}^{r_q}$ be an orthonormal eigenbasis of $\mathcal{J}_{\mathrm{sol},q}$ and write $h_q = \sum_{j=1}^{r_q} \theta_{q,j} e_{q,j}$. In these eigencoordinates, the block-diagonal *degreewise benchmark* is

$$Z_{q,j} = \kappa_{q,j}\theta_{q,j} + n^{-1/2}\xi_{q,j}, \qquad \sum_{q\ge 2} (1+q)^{2s} \|\theta_q\|^2 \le R^2, \tag{5}$$

which defines the Hermite–Sobolev ball $\mathcal{F}_m^s(R)$ for $h = \sum_{q\ge 2} h_q$ (conventions in App. H). The dimension law $r_q \asymp q^{m-1}$ is proved; the information scale is not.

Assumption 10 (First-moment information scale)**.** Uniformly over the degrees and designs considered, $\bar{\mu}_q := r_q^{-1} \sum_j \kappa_{q,j}^2 \asymp K/q$.

Here the factor K describes temporal excitation, not merely the number of snapshots, and does not follow from the structural threshold $K \ge m$; weakly excited designs can violate it (§5.2), and App. H.2 sketches where the scale comes from. By Markov's inequality, Asm. 10 implies that a constant fraction of degree-q directions have information $O(K/q)$, which drives the lower bound. A matching upper bound requires additional control near zero.

Assumption 11 (Uniform thin lower tail)**.** There exist $C' < \infty$ and $\gamma > 1$, independent of q, n, K, such that $\mathcal{N}_q(t) := r_q^{-1} \#\{j : \kappa_{q,j}^2 \le t\bar{\mu}_q\} \le C' t^\gamma$ for $0 < t \le 1$.

This requires no degree-uniform spectral gap, excludes zero eigenvalues, and, together with Asm. 10, yields $\sum_j \kappa_{q,j}^{-2} \asymp q^m/K$ (App. G).

Theorem 12 (Conditional tangent benchmark)**.** *Suppose the whitened noise in the tangent snapshot experiment is standard Gaussian. Then Asm. 10 and Asm. 9 imply that in that experiment, as $nK \to \infty$,*

$$\inf_{\widehat{h}} \sup_{h \in \mathcal{F}_m^s(R)} \mathbb{E}\big\|\widehat{h} - h\big\|_{L^2(\rho_\star)}^2 \gtrsim (nK)^{-2s/(2s+m+1)}.$$

If Asm. 11 also holds, then in the degreewise benchmark (5), for any noise $\xi_{q,j}$ there with $\mathbb{E}\xi = 0$ and $\mathrm{Cov}(\xi) = I$, a linear estimator with degree cutoff $Q_\star \asymp (nK)^{1/(2s+m+1)}$ attains the matching rate. Constants may depend on s, m, R and the hypothesis constants; at $\gamma = 1$ the upper bound incurs a logarithmic factor.

Indeed, $r_q \asymp q^{m-1}$ directions with average inverse information of order q/K give cumulative variance $\asymp Q^{m+1}/(nK)$, which balances the Sobolev bias $R^2 Q^{-2s}$ at the stated rate (App. H).

Remark 13 (What remains for sampled snapshots)**.** Thm. 12 gives the lower bound for the tangent snapshot experiment and the matching upper bound for the degreewise benchmark. By our argument, extending the upper bound to the full tangent experiment would need the reverse block comparison and control of omitted-degree leakage; transferring either result to nonlinear sampled snapshots would further need uniform control of the linearization remainder and of the sampling likelihood (App. H.7). We do not prove these transfers; the estimator of §4 instead has its own finite-basis guarantee (Thm. 37).

4 Estimation from snapshot moments

Sections 2–3 characterize the identifiable target and its local statistical difficulty. A direct implementation of (2) requires estimating the score and the source b_t, including the temporal log-density derivative, before fitting the field, so plug-in nuisance error enters at first order.

Proposition 14 (Strong-form nuisance amplification)**.** *Writing the strong-form moment against test functions χ_p as $m_p = \mathbb{E}[r_\theta \chi_p]$, its derivative with respect to the score nuisance is nonzero. Along a solenoidal eigendirection of the profiled strong-form Hessian (Prop. 15) with curvature $\lambda > 0$, the corresponding component of the first-order coefficient bias is amplified by λ^{-1}.*

Nuisance error is amplified along weakly visible solenoidal directions. Spectral filtering cannot fix this: it attenuates signal and first-order nuisance bias together (Prop. 35). A coefficient-side Riesz correction likewise leaves a nonzero nuisance derivative, so it is only a one-step estimating-equation correction, not a Neyman orthogonalization (App. M). We therefore remove the score nuisance altogether, retaining only a covariance-based strength audit from the strong-form geometry.

Proposition 15 (Covariance-only strength audit)**.** *For centered Gaussian slices and a linear field $F(x) = Hx$, the unweighted population loss has quadratic part $\frac{1}{K}\sum_{k=1}^{K} \mathrm{vec}(H)^\top \big(\Sigma_{t_k} \otimes \Sigma_{t_k}^{-1} + \mathsf{C}_m\big)\,\mathrm{vec}(H)$, with* vec *column-stacked and C_m the commutation matrix. Hence the profiled solenoidal curvature — the corresponding Schur complement after profiling the gradient part — depends only on the slice covariances; so does any version weighted deterministically from them.*

Thus, in the Gaussian linear model, zero eigenvalues mark source-level gauge directions and small positive ones weakly visible directions. We use this covariance-only geometry only as a second-moment *design* audit: beyond Gaussian slices, strong-form curvature depends on higher moments, while feasible-GLS information generally involves fourth moments (App. H). Estimation instead uses score-free weak moments.

4.1 Nuisance-free weak moments

For any smooth test function ψ in the generator domain, Itô's formula gives $\frac{d}{dt}\mathbb{E}_{\rho_t}[\psi] = \mathbb{E}_{\rho_t}[F \cdot \nabla\psi + \sigma^2\Delta\psi]$, and the trapezoidal rule between adjacent snapshots yields

$$\mathbb{E}_{k+1}[\psi_j] - \mathbb{E}_k[\psi_j] = \frac{\Delta_k}{2}(\mathbb{E}_k + \mathbb{E}_{k+1})\big[F \cdot \nabla\psi_j + \sigma^2\Delta\psi_j\big] + O(\Delta_k^3), \tag{6}$$

with $\Delta_k = t_{k+1} - t_k$ and $\mathbb{E}_k := \mathbb{E}_{\rho_{t_k}}$. The diffusion term requires only an empirical moment, so (6) estimates the Itô drift F: no score, density estimate, or pointwise source. With $F_a(x) = \sum_{p=1}^{P} a_p\phi_p(x)$ the system is linear in a.

Proposition 16 (Block-tridiagonal moment noise under independent slices)**.** *Let the slices be sampled independently with sizes n_k, let $\Psi = (\psi_1, \dots, \psi_J)^\top$, write $f_a := (F_a \cdot \nabla\psi_j + \sigma^2\Delta\psi_j)_j$ and $u_k^{\mp}(x;a) := \mp\Psi(x) - \frac{\Delta_k}{2} f_a(x)$. Then $V_{kk}(a) = \mathrm{Cov}_k(u_k^-)/n_k + \mathrm{Cov}_{k+1}(u_k^+)/n_{k+1}$, $V_{k,k+1}(a) = \mathrm{Cov}_{k+1}(u_k^+, u_{k+1}^-)/n_{k+1}$, and $V_{k\ell} = 0$ for $|k - \ell| > 1$. The stacked covariance is block tridiagonal, adjacent blocks sharing a slice with opposite sign.*

The covariance depends on the field through f_a, so feasible GLS needs a pilot (App. N). For the implemented finite representation this yields a snapshot-only guarantee of a different kind from Thm. 12: Thm. 37 bounds any linear readout by a sampling term of order $n_{\min}^{-1/2}$ plus a non-stochastic quadrature residual, both amplified by the smallest whitened singular value, so increasing n leaves a floor (App. O). Independence across slices is a hypothesis, not a formality: with lineage-tracked particles the true covariance is dominated by blocks this surrogate sets to zero (App. R).

Stacking (6) over intervals k and probes j, with $\widehat{\mathbb{E}}_k$ the empirical mean over slice k, gives the response and design

$$\widehat{b}_{kj} := \widehat{\mathbb{E}}_{k+1}[\psi_j] - \widehat{\mathbb{E}}_k[\psi_j] - \tfrac{\Delta_k}{2}(\widehat{\mathbb{E}}_k + \widehat{\mathbb{E}}_{k+1})[\sigma^2\Delta\psi_j], \qquad \widehat{X}_{kj,p} := \tfrac{\Delta_k}{2}(\widehat{\mathbb{E}}_k + \widehat{\mathbb{E}}_{k+1})[\phi_p \cdot \nabla\psi_j],$$

and $\widehat{V}$ is the covariance of Prop. 16 at a pilot fit. The estimator is $\widehat{a}_\zeta \in \arg\min_{a \in \mathcal{A}_{\mathrm{vis}}} \|\widehat{V}^{-1/2}(\widehat{b} - \widehat{X}a)\|_2^2 + \zeta\|a\|_2^2$, with $\mathcal{A}_{\mathrm{vis}}$ excluding the exact snapshot gauge, removed rather than weakly penalized so it carries neither an estimate nor spurious variance. We use the ridge penalty $\mathcal{R}(a) = \|a\|_2^2$ (App. Q) or a truncated-SVD cutoff of the whitened design (App. N). The local residual in (6) is $O(\Delta_k^3)$ whereas the corresponding design row is $O(\Delta_k)$, so under stable inversion fixed gaps leave the usual $O(\Delta_{\max}^2)$ coefficient-level quadrature floor, which increasing n does not remove (App. N).

4.2 Representation and stabilization

Three ingredients follow: GLS whitening, exact gauge exclusion before fitting, and a representation — a theory-aligned frame or fixed features — solved linearly rather than by descent. Regularization is then selected for the solenoidal target, not for held-out moment fit.

Table 1: **Planted-curl OU benchmark** (d=12, m=4, K=5, σ^2=ν=1). Entries are relErr. Only weak rows are nuisance-free: the parametric reference is given the symmetric–antisymmetric tie, while strong-form rows use the exact score and temporal source. Multi-seed entries are mean ± s.d.; the four snapshot-only arms share identical draws over 12 paired seeds, whereas strong-form rows use a separate 8-seed set. See App. Y for single-seed and seed-spread caveats and solver variants.

estimator	ingredients	n=40k	n=200k	n=40k, 12 seeds
gradient-only / displacement	—	1.000	1.000	—
weak, unweighted	moments	—	—	0.734 ± 0.157
weak, diagonal weights	moments	0.893	0.303	0.782 ± 0.199
weak, GLS	moments	**0.651**	**0.282**	0.751 ± 0.260
parametric reference	exact linear model	0.511	0.179	0.308 ± 0.088
strong form, oracle nuisances	exact score & source	—	—	0.535 ± 0.177
strong form, 10% score error	corrupted score	—	—	2.461 ± 0.689

Fit-based selection is aimed at the wrong target: the regularization strength minimizing moment-prediction risk and the one minimizing the risk of a linear functional of the field have different minimizers in general, with the gap governed by the near-gauge spectrum (Prop. 38, App. Q).

Because whitening makes the response law exactly $\mathcal{N}(0, I)$, the solenoidal risk is estimable without ground truth by perturbing the whitened response under its known law; its failure mode is documented in App. Q.

5 Experiments

Because marginal agreement cannot validate an ambiguous component, we use planted dynamics with known identifiable part and exact gauge, matched across data, seeds, and readouts. We report solenoidal relative error $\mathrm{relErr} := \|\widehat{h} - h^\star\|/\|h^\star\|$, so $\mathrm{relErr} = 1$ for a zero-circulation estimate. These finite-dimensional experiments test the structural predictions and estimator, not the rate exponent of Thm. 12. Protocol and receipts are in Apps. Y–Z.

Structural predictions. For a planted rotation of norm 0.5, exactly matched marginals yield antisymmetric displacement at the 10^{-16} level and gradient-only $\mathrm{relErr} = 1.0000$, illustrating Cor. 18 (Fig. 1b). An isotropic-block rotation leaves all marginals unchanged to machine precision (Prop. 22); operator checks are in App. S.

5.1 Recovery from weak moments

Our main benchmark is an autonomous planted-curl Ornstein–Uhlenbeck system with ambient $d = 12$, intrinsic $m = 4$, $K = 5$ snapshots, and diffusion σ^2 equal to the relaxation rate ν (Tab. 1). Slices are drawn independently, as Prop. 16 requires. The single-draw columns at $n = 40$k and 200k are consistent with the parametric $n^{-1/2}$ benchmark, but each is one draw from a heavy-tailed seed distribution, so we do not read them as a measured rate; the pre-estimation audit of Prop. 15 correctly refuses the $n = 4$k design, predicting $\mathrm{relErr} > 1$ against a realized 3.70.

Whitening. We find no point-estimation gain from weighting of any kind here. Paired differences are reported as mean ± standard error over seeds, and t is the paired t-statistic, their ratio. Over the twelve paired seeds diagonal-to-GLS is $+0.031 \pm 0.036$ ($t = +0.88$), unweighted-to-GLS is -0.017 ± 0.051 ($t = -0.32$), and unweighted is marginally better than diagonal (-0.048 ± 0.026, $t = -1.82$). Calibration is already good without whitening (1.03× diagonal, 0.87× GLS). We therefore report Prop. 16 as correct algebra for the independent-slice model, not a demonstrated gain; §5.2 says why.

The cost of not knowing the model class. The parametric reference, told the exact linear tie between symmetric and antisymmetric coefficients, beats every nuisance-free arm: 0.308 against 0.751 for GLS, paired $+0.443 \pm 0.075$ ($t = +5.87$). The price of model-class agnosticism is therefore substantial here.

5.2 A DESIGN TENSION

Prop. 16 gives whitening its largest potential efficiency gain when adjacent moment equations are strongly correlated through shared snapshots, which favors short gaps. Thm. 8 instead requires enough cross-slice shape variation to expose circulation, which on this system favors longer gaps.

Observation 17 (Efficiency and identifiability compete for the same knob)**.** *Compressing the five snapshot times sixfold, from a span of* 1.45 *to* 0.20 *while holding everything else fixed, raises solenoidal error from* 0.78 *to* 30.9 ± 18.3 *with diagonal weights and* 25.9 ± 13.5 *with* GLS*, roughly a* 40*-fold increase and far above the no-information level. The whitening contrast remains null there (*$t = +1.62$ *over six paired seeds).*

The collapse is consistent with the mechanism behind Thm. 8: compressing time suppresses the covariance contrasts that expose circulation, driving the planted component toward the near-gauge directions diagnosed by Prop. 15. In this design, the regime that strengthens the covariance exploited by whitening weakens the shape variation needed for recovery. We treat this as a limitation rather than a general law: we have not characterized when both requirements can hold simultaneously.

5.3 ABLATIONS

All ablations are reported in full in the appendices. With exact score and source, a closed-form strong-form solve recovers the planted circulation; fitting the score raises relative error to 0.862 and fitting both nuisances to 2.19. A probe-wise decomposition attributes 38% amplitude bias to the temporal log-density value channel versus 7% to the score channel; denoising score matching supplies $\nabla \log \rho_t$ but not the additive normalization of $\log \rho_t$, which (6) eliminates (App. M). On the nonlinear system of App. R, no snapshot-only recovery remains once slices are drawn independently: curl-channel error is 1.65 (diagonal) and 1.01 (GLS) against a no-information level of 1.0. The recovery under lineage tracking therefore came from particle identity rather than snapshot moments.

Polynomial frames scale as $P = 504$, $5{,}460$, $81{,}900$ at $d = 6, 12, 25$, with the $O(P^3)$ whitening Cholesky as the bottleneck. Random features reach parity with the correctly specified degree-three frame at $m = 6$ (t=0.49, six seeds), while a random rotation of all 12 features leaves recovery unchanged and the audit recovers the $m = 4$ tangent space to $2.58°$. Noise, parameterization, conditioning, latent-regime, and autonomy stress tests are in Apps. P, T, U, V and W.

6 DISCUSSION

Snapshot data supports a smaller estimand than the literature implicitly assumes, parts of its visibility can be diagnosed from the snapshot geometry before fitting, and the resulting estimator is a whitened linear solve rather than a trained field. Reporting zero circulation is a property of a parameterization, not a finding: within a time-indexed class it is unfalsifiable from marginal fit (Prop. 1) and least-action reconstruction actively selects it (Cor. 2). Marginal metrics therefore cannot validate non-gradient structure; planted-circulation benchmarks can. Autonomy converts an untestable model into a testable one, but it is an assumption about the chosen state representation rather than necessarily about the underlying biology, and App. W shows it can fail invisibly. The price is explicit: the rate is governed by intrinsic dimension m, and $K \geq m$ generic shapes eliminate all finite-degree polynomial gauge directions at the source level (Thm. 8); fewer shapes can already identify lower-degree components. Assumptions 9–11 and the faithfulness condition of §3.2 remain unproved; App. A tags every claim.

The current implementation is computationally limited by the intrinsic representation dimension: polynomial frames grow combinatorially in m, and dense whitening has $O(P^3)$ Cholesky cost. Thus the exact-solve estimator is most practical when the dynamics admit a low-dimensional representation or a compact feature basis, consistent with the intrinsic-dimensional formulation of the theory. All validation here is synthetic. We therefore do not claim a biological benchmark result or a neural-baseline leaderboard; methods supplied with velocity, pairing, or lineage information answer a strictly richer observation problem, while snapshot-only neural transport baselines remain useful empirical comparisons for future work.

REPRODUCIBILITY STATEMENT

All theoretical claims are stated with their hypotheses in the main text and proved in the appendices; App. A tabulates every numbered claim against its status, distinguishing results proved as stated, results proved conditional on stated assumptions, results with numerical support only, and open problems. Assumptions that remain unproved are identified as such at the point of use, and Rem. 13 states explicitly what Thm. 12 does and does not establish.

All experiments are synthetic with closed-form ground truth. The supplementary material contains the estimator library, the benchmark suite, every experiment script, and machine-readable JSON receipts from which each number and table in this paper is generated; App. Z maps each reported quantity to its receipt and generating command, and the weak-form path runs on NumPy and SciPy alone, while the strong-form and neural comparison arms also require PyTorch. Gates were fixed before each run. App. Y records the negative results and every retracted or superseded claim, including two corrections that changed reported numbers, with superseded intermediate values retained in the receipts rather than deleted. The complete suite runs on a single CPU core.

AI USE STATEMENT

Large language models (Claude, Anthropic, and ChatGPT, OpenAI) were used in preparing and conducting parts of this work. Their roles are disclosed below by subtask, following the ICLR 2027 AI Policy for Authors.

Manuscript preparation. The models revised the experiments and discussion sections, drafted appendices from existing proof notes and experimental receipts, repaired cross-references, and audited reported numbers against the machine-readable receipts in the supplement. They also assisted with a feedback-driven algebra and exposition audit, including corrections to the Gaussian fourth-moment calculation, the finite-basis proof, and scope statements.

Conceptual and theoretical development. The models assisted in developing and refining the conceptual and theoretical framework of the work, including stress-testing the formulation of the identifiability problem and proposing or refining hypotheses and intermediate conjectures considered during development. The authors evaluated these suggestions and determined the final framework, assumptions, and claims.

Mathematical claims and proofs. The models assisted in checking and refining mathematical claims, derivations, and proof arguments, and in drafting portions of proofs from the authors' existing arguments and notes. This included identifying gaps or errors and suggesting corrections during the proof audit. The authors checked the final statements and proofs and take responsibility for their correctness.

Methodology, implementation, and interpretation. The models provided feedback on experimental methodology and benchmark design and assisted with implementation, diagnostic code, and interpretation of the resulting experiments. In particular, a model identified two defects in the shipped experimental code—an omitted diffusion term in the weak-form response and a lineage-tracked sampling design inconsistent with the independence hypothesis of Prop. 16—wrote the corrections and diagnostic scripts that quantified them, and re-ran the affected benchmarks. The resulting retractions in App. Y and the design tension reported in Obs. 17 follow from those runs. All corrections, re-runs, and interpretations were verified by the authors against the receipts.

Synthetic data. The synthetic data used in the experimental benchmarks were generated by simulation code whose design, implementation, or debugging was assisted by the models. The authors verified the data-generating mechanisms and experimental configurations used for the reported results.

Literature retrieval and bibliography. The models assisted in searching for and identifying potentially relevant literature and in checking bibliographic records. The authors verified cited works against primary sources and checked titles, author lists, identifiers, and the relevance of the cited results.

Other uses. Generative AI was not used for translation, cleaning or reformatting an external dataset, or qualitative or thematic data analysis; these tasks were not applicable to this work.

The authors reviewed all AI-assisted work and take full responsibility for the final content of this paper, including all text, mathematical claims, proofs, code, experimental results, citations, and conclusions produced with model assistance.

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023.

David Balduzzi, Sébastien Racanière, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, 2018.

Christopher Battle, Chase P. Broedersz, Nikta Fakhri, Veikko F. Geyer, Jonathon Howard, Christoph F. Schmidt, and Fred C. MacKintosh. Broken detailed balance at mesoscopic scales in active biological systems. *Science*, 352(6285):604–607, 2016.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

Volker Bergen, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12): 1408–1414, 2020.

Nicholas M. Boffi and Eric Vanden-Eijnden. Probability flow solution of the Fokker–Planck equation. *Machine Learning: Science and Technology*, 4(3):035012, 2023.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

Charlotte Bunne, Laetitia Papaxanthos, Andreas Krause, and Marco Cuturi. Proximal optimal transport modeling of population dynamics. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 6511–6528. PMLR, 2022. URL https://proceedings.mlr.press/v151/bunne22a.html.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, 2021.

David L. Donoho, Richard C. Liu, and Brenda MacGibbon. Minimax risk over hyperrectangles, and implications. *The Annals of Statistics*, 18(3):1416–1437, 1990.

Anna Frishman and Pierre Ronceray. Learning force fields from stochastic trajectories. *Physical Review X*, 10(2):021009, 2020.

Federico S. Gnesotto, Federica Mura, Jannes Gladrow, and Chase P. Broedersz. Broken detailed balance and non-equilibrium dynamics in living systems: a review. *Reports on Progress in Physics*, 81(6):066601, 2018.

Vincent Guan, Joseph Janssen, Hossein Rahmani, Andrew Warren, Stephen Y. Zhang, Elina Robeva, and Geoffrey Schiebinger. Identifying drift, diffusion, and causal structure from temporal snapshots. *arXiv preprint arXiv:2410.22729*, 2024.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

Tatsunori Hashimoto, David Gifford, and Tommi Jaakkola. Learning population-level diffusions with generative RNNs. In *International Conference on Machine Learning*, 2016.

Guillaume Huguet, Daniel S. Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy. Manifold interpolating optimal-transport flows for trajectory inference. In *Advances in Neural Information Processing Systems*, 2022.

Chii-Ruey Hwang, Shu-Yin Hwang-Ma, and Shuenn-Jyi Sheu. Accelerating Gaussian diffusions. *The Annals of Applied Probability*, 3(3):897–913, 1993.

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

Ildar A. Ibragimov and Rafail Z. Has'minskii. *Statistical Estimation: Asymptotic Theory*. Springer, 1981.

Daniel Kunin, Javier Sagastuy-Brena, Lauren Gillespie, Eshed Margalit, Hidenori Tanaka, Surya Ganguli, and Daniel L. K. Yamins. The limiting dynamics of SGD: Modified loss, phase-space oscillations, and anomalous diffusion. *Neural Computation*, 36(1):151–174, 2024. doi: 10.1162/neco_a_01626.

Chulan Kwon, Ping Ao, and David J. Thouless. Structure of stochastic dynamics near fixed points. *Proceedings of the National Academy of Sciences*, 102(37):13029–13033, 2005.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. RNA velocity of single cells. *Nature*, 560(7719):494–498, 2018.

Hugo Lavenant, Stephen Zhang, Young-Heon Kim, and Geoffrey Schiebinger. Towards a mathematical theory of trajectory inference. *The Annals of Applied Probability*, 34(1A):428–500, 2024.

Justin Lee, Behnaz Moradijamei, and Heman Shakeri. Multi-marginal stochastic flow matching for high-dimensional snapshot data at irregular time points. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 33476–33498. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/lee25x.html.

Tony Lelièvre, Francis Nier, and Grigorios A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152(2):237–274, 2013.

Xiang Li, Zebang Shen, Ya-Ping Hsieh, and Niao He. When scores learn geometry: Rate separations under the manifold hypothesis. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/4406cafe40d3ca4a7511a8996b6bd1b6-Abstract-Conference.html. arXiv:2509.24912.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Dimitra Maoutsa, Sebastian Reich, and Manfred Opper. Interacting particle solutions of Fokker–Planck equations through gradient–log-density estimation. *Entropy*, 22(8):802, 2020.

Daniel A. Messenger and David M. Bortz. Weak SINDy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021.

Kirill Neklyudov, Rob Brekelmans, Daniel Severo, and Alireza Makhzani. Action matching: Learning stochastic dynamics from samples. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25858–25889. PMLR, 2023. URL https://proceedings.mlr.press/v202/neklyudov23a.html.

Katarina Petrović, Lazar Atanackovic, Viggo Moro, Kacper Kapuśniak, İsmail İlkan Ceylan, Michael M. Bronstein, Avishek Joey Bose, and Alexander Tong. Curly flow matching for learning non-gradient field dynamics. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-4053. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/b00692865928625ace212e89ded2decd-Abstract-Conference.html. arXiv:2510.26645.

Luc Rey-Bellet and Konstantinos Spiliopoulos. Irreversible Langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7):2081–2103, 2015.

Jack Richter-Powell, Yaron Lipman, and Ricky T. Q. Chen. Neural conservation laws: A divergence-free perspective. In *Advances in Neural Information Processing Systems*, 2022.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019. doi: 10.1016/j.cell.2019.01.006.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2717. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/c428adf74782c2092d254329b6b02482-Abstract-Conference.html.

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

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

Antonio Terpin, Nicolas Lanzetti, Martín Gadea, and Florian Dörfler. Learning diffusion at lightspeed. In *Advances in Neural Information Processing Systems*, volume 37, pp. 6797–6832, 2024. doi: 10.52202/079017-0218. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/0ce1eb87dbb03fdfa872a93d15cfe333-Abstract-Conference.html.

Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, and Smita Krishnaswamy. TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. In *International Conference on Machine Learning*, 2020.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

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

Jin Wang, Li Xu, and Erkang Wang. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences*, 105(34):12271–12276, 2008.

Caleb Weinreb, Samuel Wolock, Betsabeh K. Tusi, Merav Socolovsky, and Allon M. Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, 2018.

Grace Hui Ting Yeo, Sachit D. Saksena, and David K. Gifford. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nature Communications*, 12(1):3222, 2021.

APPENDIX

The appendices are organized as follows. App. A is the hypothesis ledger: every numbered claim against its status. App. B extends the positioning. Apps. C–E prove the structural results of §2–3.1. Apps. F–L develop the tangent analysis, its Gaussian specialization and the intrinsic-dimension reduction. Apps. M–Q cover the estimator, including its finite-basis guarantee (App. O). Apps. R–Z contain the remaining experiments, stress tests, scope, negative results and reproduction instructions.

A HYPOTHESIS LEDGER

Tab. 2 labels each claim **P** (proved as stated), **C** (proved conditional on stated assumptions), **N** (numerically supported, not proved), or **O** (open).

Table 2: **Hypothesis ledger.** Status of every numbered claim and what it rests on (App. A).

claim	status	scope and what it rests on
Prop. 1	P	Poincaré inequality, $\partial_t \log \rho_t \in L^2(\rho_t)$; App. C
Cor. 2	P	$(\ker \mathcal{A}_\rho)^\perp = \overline{\{\nabla\phi\}}$; single time
Cor. 18	P	Gaussian marginals; closed-form Benamou–Brenier
Cor. 19	P	population level; (i) Gaussian slices, exact OT coupling, noiseless interpolant; (ii) any (possibly non-gradient) reference drift, smooth positive Schrödinger potentials
Prop. 4	P	source constraints only, at finitely many or a continuum of times; forward uniqueness for the continuum converse
Ex. 5	P	explicit; shows $\ker \mathcal{A}_K$ is not necessary for snapshot equivalence
Prop. 6	P	smooth h; fibre dimension, not function-space dimension
Prop. 21	P	time-conditioned head; per-slice Poisson solvability
Prop. 22	P	isotropic block; exact marginal invariance
Prop. 7	P	Gaussian anchor slice; Hermite ladder (App. D)
Thm. 8	P	compared slices Gaussian with a common mean (the reference slices in the local analysis); exact, no linearization; sharpness by the explicit field of Lem. 29
Prop. 28	P	leading-degree bound; its attainment below threshold is numerical (App. E.1), open in general
degree cap (App. E.1)	P/N	exact gauge of degree K for $K \le m-1$ proved; identification up to degree $K-1$ proved for $K = m-1$, certified exactly for $m \le 6$, open in general
faithfulness (§3.2)	O	bridges source identifiability to the tangent statistical limits; not used by Thm. 12; implied by Asm. 11; holds with equality in the shape-rich cases computed (App. F)
Asm. 9	O	one-sided Loewner comparison at every cutoff; used for the lower bound through Lem. 32; not verified
Asm. 10	O	$r_q \asymp q^{m-1}$ proved in Prop. 7; the K/q scale is a heuristic factorization with four stated gaps (App. H.2)
Asm. 11	N	tail exponent measured only on the source operator; the implied trace law measured on the propagated design (App. G)
Thm. 12	C	lower bound in the tangent snapshot experiment under Asm. 10 and 9; upper bound in the degreewise benchmark under Asm. 10 and 11; *not* a sampled-snapshot theorem (Rem. 13)
Thm. 37	P	snapshot-only, finite basis, finite sample; needs $\alpha = s_{\min}(WX) > 0$ on the retained subspace
Prop. 14	P	nonzero solenoidal component; first-order expansion
Prop. 35	P	diagonal filters in the profiled basis only
Prop. 15	P	centered Gaussian slices, covariance-determined weighting; fails for general densities
Prop. 16	P	independent slices, fixed a; violated by lineage-tracked designs (App. R)
Prop. 38	P	whitened linear experiment; ridge penalty; risks averaged over coordinate signs of $a^\star$; linear functional
extensions (App. X)	P	known state-dependent diffusion, for §2 and the weak moments only; not §3.2
Prop. 34	C	smoothed-manifold regime of App. K
Obs. 17	N	one system, $m = 4$; no characterization of when both requirements can hold

Three cautions carry across the paper. First, every rate statement lives in the tangent snapshot experiment or its degreewise benchmark, never in the sampled-snapshot experiment. Second, the finite-dimensional experiments of §5 exercise the structural predictions and the estimator; they do not validate the nonparametric exponent. Third, ρ-solenoidal, Euclidean divergence-free, and "antisymmetric linear coefficient" are three different things, and results proved for one do not transfer to another.

B Extended related work

Population dynamics and generative transport. Trajectory inference is commonly posed as transport between temporal marginals: optimal transport (Schiebinger et al., 2019), neural-ODE and JKO approaches (Tong et al., 2020; Huguet et al., 2022; Bunne et al., 2022; Terpin et al., 2024), Schrödinger bridges (De Bortoli et al., 2021; Shi et al., 2023), and flow-matching or generative formulations (Hashimoto et al., 2016; Yeo et al., 2021; Lipman et al., 2023; Tong et al., 2024; Lee et al., 2025). These construct dynamics consistent with prescribed distributions; we ask which aspects of the dynamics those distributions determine, and Cor. 19 shows that OT-coupled flow matching returns a gradient field at the population level and that Schrödinger-bridge matching, for any reference drift, only ever adds a gradient correction to it, so neither method can manufacture circulation a curl-free reference lacks. Consistency results often assume gradient drift (Lavenant et al., 2024), and action matching explicitly targets a gradient representative (Neklyudov et al., 2023); rotational ambiguity from population snapshots was noted by Weinreb et al. (2018). Scores near a smoothed manifold encode support geometry strongly (Li et al., 2026), which we connect to the present setting in App. K.

Breaking the gauge with dynamical information. Autonomy couples Eulerian marginals through one shared field; velocity, pairing or lineage measurements instead supply Lagrangian information (Petrović et al., 2025; La Manno et al., 2018; Bergen et al., 2020). We isolate what marginals alone determine; a lineage-tracked variant of our nonlinear benchmark is in App. R.

Identifiability from marginals. The closest identifiability results either impose gradient dynamics (Hashimoto et al., 2016; Lavenant et al., 2024; Neklyudov et al., 2023) or specialize to linear-Gaussian models. Guan et al. (2024) characterize when the drift and diffusion of a linear additive-noise SDE are identifiable from population marginals, in terms of generalized rotational symmetries of the initial law. We treat an infinite-dimensional field class, condition on the observed shape family rather than on the initial law, and study finite-sample recovery; App. I makes the comparison precise.

Score-based and weak-form estimation. Strong-form approaches combine score estimation (Hyvärinen, 2005; Vincent, 2011; Song & Ermon, 2019) with continuity or Fokker–Planck equations (Maoutsa et al., 2020; Song et al., 2021; Boffi & Vanden-Eijnden, 2023); we show score error is a non-orthogonal nuisance there. Weak formulations are classical in system identification (Brunton et al., 2016; Messenger & Bortz, 2021) and correspond to generalized method of moments (Hansen, 1982); we adapt them to unpaired snapshots. Related nonequilibrium methods infer forces from full trajectories (Frishman & Ronceray, 2020; Gnesotto et al., 2018).

C Proofs for Section 2

Throughout, densities are smooth and positive with enough decay for the displayed integrations by parts, and $\mathcal{A}_\rho F = \rho^{-1}\nabla\cdot(\rho F)$.

C.1 The adjoint and the two orthogonal complements

Scalars live in $L^2(\rho)$ and vector fields in $L^2(\rho;\mathbb{R}^m)$, with $\langle F, G\rangle_{L^2(\rho)} := \int F\cdot G\,\rho$; orthogonal complements and closures of sets of fields are taken in $L^2(\rho;\mathbb{R}^m)$. For scalar ϕ and field F in the relevant domains, $\langle \mathcal{A}_\rho F, \phi\rangle_{L^2(\rho)} = \int \nabla\cdot(\rho F)\,\phi = -\int \rho\, F\cdot\nabla\phi = -\langle F, \nabla\phi\rangle_{L^2(\rho)}$, so $\mathcal{A}_\rho^*\phi = -\nabla\phi$ and

$$(\ker \mathcal{A}_\rho)^\perp = \overline{\operatorname{range}\mathcal{A}_\rho^*} = \overline{\{\nabla\phi\}}. \tag{7}$$

This is a statement about *one* measure and does not survive stacking: for the stacked operator, with fields in $L^2(\mu;\mathbb{R}^m)$ for a reference density μ and the k-th component in $L^2(\rho_{t_k})$, $\mathcal{A}_K^*(\phi_k)_k =$

$-\mu^{-1}\sum_k \rho_{t_k}\nabla\phi_k$, whose range is the span of *reweighted* gradients — strictly larger than $\overline{\{\nabla\phi\}}$ as soon as two slices differ. Much of the folklore about L^2 actions losing the curl is (7) applied at a single time; §2.2 is what happens when it is not.

Proof of Prop. 1. Fix t, write $\rho = \rho_t$ and $f = \partial_t\rho_t$, and let C_P be the Poincaré constant of ρ, so that $\mathrm{Var}_\rho(\psi) \le C_P \int |\nabla\psi|^2\rho$. On $\mathcal{H} := H^1(\rho)/\mathbb{R}$ let

$$a(\phi,\psi) := \int \nabla\phi\cdot\nabla\psi\,\rho, \qquad \ell(\psi) := \int \psi f.$$

Mass conservation gives $\int f = \frac{d}{dt}\int\rho_t = 0$, so ℓ annihilates constants and is well defined on $\mathcal{H}$. The form a is bounded and, by the Poincaré inequality, coercive on $\mathcal{H}$, and

$$|\ell(\psi)| = \left|\int (\psi - \mathbb{E}_\rho\psi)\,\partial_t\log\rho\,\rho\right| \le C_P^{1/2}\,\|\nabla\psi\|_{L^2(\rho)}\,\|\partial_t\log\rho\|_{L^2(\rho)},$$

so ℓ is bounded. Lax–Milgram gives a unique $\phi_t \in \mathcal{H}$ with $a(\phi_t,\cdot) = \ell$, which is the weak form of $\nabla\cdot(\rho_t\nabla\phi_t) = -\partial_t\rho_t$. Thus $v_t := \nabla\phi_t$ satisfies $\partial_t\rho_t + \nabla\cdot(\rho_t v_t) = 0$, and testing with $\psi = \phi_t$ gives $\|v_t\|_{L^2(\rho_t)} \le C_P^{1/2}\|\partial_t\log\rho_t\|_{L^2(\rho_t)}$. That transporting ρ_0 along v recovers $\{\rho_t\}$, rather than merely being consistent with it, would further require uniqueness for the continuity equation, which needs regularity of v beyond L^2; the unfalsifiability conclusion uses only the consistency proved here. □

Proof of Cor. 2. The velocities compatible with the observed path at time t form the affine set $v_t^0 + \ker\mathcal{A}_{\rho_t}$ for any particular solution v_t^0. The element of least $L^2(\rho_t)$ norm in an affine set is its projection onto the orthogonal complement of the direction space, which is $\overline{\{\nabla\phi\}}$ by (7). Its ρ_t-solenoidal component is therefore zero. Note this is a projection, not a bias: least-action reconstruction does not shrink the circulation, it deletes it. □

Corollary 18 (Displacement interpolation is curl-blind in closed form)**.** *For Gaussian marginals with a planted rotation, the Benamou–Brenier displacement velocity between consecutive slices is a gradient on every segment. Its solenoidal relative error equals* 1 *at every sample size, while the induced marginal path matches the observed one exactly.*

Proof. Between consecutive centered Gaussians $\mathcal{N}(0,\Sigma_k)$ and $\mathcal{N}(0,\Sigma_{k+1})$ the quadratic-cost optimal map is the linear map $T_k = \Sigma_k^{-1/2}(\Sigma_k^{1/2}\Sigma_{k+1}\Sigma_k^{1/2})^{1/2}\Sigma_k^{-1/2}$, which is symmetric positive definite. Displacement interpolation uses $X_s = ((1-s)I + sT_k)X_0$, whose velocity field at time s is $v_s(x) = (T_k - I)((1-s)I + sT_k)^{-1}x$. All three factors are symmetric and commute (they are polynomials in T_k), so v_s is a symmetric matrix field, hence the gradient of a quadratic form and exactly curl-free. Writing the readout as $\mathrm{relErr} = \|\widehat{h} - h^\star\|/\|h^\star\|$ with $\widehat{h} = 0$ gives 1 at every sample size, while by construction the interpolation matches both endpoint marginals exactly. □

Corollary 19 (Flow matching and Schrödinger bridges cannot manufacture circulation)**.** *(i) Between consecutive centered Gaussian slices, conditional flow matching with the exact quadratic-cost OT coupling and the noiseless linear interpolant has population regression target equal to the displacement velocity v_s of Cor. 18, whose ρ_s-solenoidal component vanishes, so its solenoidal relative error is* 1 *at every sample size. (ii) Let a reference process obey* (1) *with any smooth, possibly time-dependent, possibly non-gradient drift $F_{\mathrm{ref},t}$, and diffusion σ^2 as fixed throughout. A Schrödinger bridge between two marginals relative to this reference, with smooth positive Schrödinger potentials (as holds for Gaussian marginals), is Markov with Itô drift $F_{\mathrm{ref},t} + 2\sigma^2\nabla\log\eta_t$: the reference drift plus a gradient correction. Consequently the bridge's ρ_t-solenoidal component at every t equals that of $F_{\mathrm{ref},t}$ exactly, whether or not $F_{\mathrm{ref},t}$ is itself a gradient: a Schrödinger bridge can neither create circulation a curl-free reference lacks nor remove circulation a rotational reference has, also for multi-marginal constructions built segment by segment. In particular, a curl-free reference (Brownian motion, or any gradient drift) forces the bridge to be curl-free, with solenoidal relative error* 1 *at every sample size, as for (i); with a circulating reference the fitted circulation is whatever the modeller built into the reference, not information extracted from the marginals.*

Proof. (i) The OT plan between the two Gaussians is deterministic, $X_1 = T_kX_0$ (proof of Cor. 18). The interpolant $X_s = (1-s)X_0 + sX_1 = ((1-s)I + sT_k)X_0$ is an invertible linear image of

X_0, and conditional flow matching regresses on $X_1 - X_0 = (T_k - I)X_0$, so its L^2-minimizer is $\mathbb{E}[X_1 - X_0 \mid X_s = x] = (T_k - I)((1-s)I + sT_k)^{-1}x = v_s(x)$, the velocity of Cor. 18; the corresponding Itô drift $v_s + \sigma^2 s_s$ is a gradient as well (§2.1). (ii) The bridge law is the reference path law reweighted by $f_0(X_0)f_1(X_1)$ for the Schrödinger potentials f_0, f_1. Conditionally on $X_t = x$ the future is the reference reweighted by $f_1(X_1)$, a Doob transform with $\eta_t(x) := \mathbb{E}_{\mathrm{ref}}[f_1(X_1) \mid X_t = x]$, which is space–time harmonic, $(\partial_t + \mathcal{L}_{\mathrm{ref},t})\eta_t = 0$ with $\mathcal{L}_{\mathrm{ref},t}u := F_{\mathrm{ref},t} \cdot \nabla u + \sigma^2 \Delta u$ for the (possibly non-gradient) reference generator. For space–time harmonic η_t and any smooth u, the product rule gives, at every (x,t),

$$\begin{aligned}\eta_t^{-1}\big(\partial_t + \mathcal{L}_{\mathrm{ref},t}\big)(\eta_t u) &= \eta_t^{-1}\big[(\partial_t\eta_t)u + \eta_t\partial_t u + F_{\mathrm{ref},t}\cdot(u\nabla\eta_t + \eta_t\nabla u)\big] \\ &\quad + \sigma^2\eta_t^{-1}\big(u\Delta\eta_t + 2\nabla\eta_t\cdot\nabla u + \eta_t\Delta u\big) \\ &= u\,\eta_t^{-1}\big[(\partial_t + \mathcal{L}_{\mathrm{ref},t})\eta_t\big] + \partial_t u + \mathcal{L}_{\mathrm{ref},t}u + 2\sigma^2\nabla\log\eta_t\cdot\nabla u \\ &= \partial_t u + \mathcal{L}_{\mathrm{ref},t}u + 2\sigma^2\nabla\log\eta_t\cdot\nabla u,\end{aligned}$$

the bracketed harmonicity term vanishing identically; this uses only that $F_{\mathrm{ref},t}\cdot\nabla$ is a first-order operator, not that $F_{\mathrm{ref},t}$ is a gradient. Hence the bridge Itô drift is $F_{\mathrm{ref},t} + 2\sigma^2\nabla\log\eta_t$. By (7), $\overline{\{\nabla\phi\}} = (\ker\mathcal{A}_{\rho_t})^\perp$ for every ρ_t, so adding a gradient field to *any* vector field leaves its ρ_t-solenoidal projection unchanged: the bridge and the reference drift have the same ρ_t-solenoidal component at every t. When $F_{\mathrm{ref},t}$ is itself a gradient this component is zero and the readout of Cor. 18 gives 1. □

Proof of Prop. 4. Finitely many times: if $\mathcal{A}_{\rho_{t_k}}F = b_{t_k}$ and $\mathcal{A}_{\rho_{t_k}}F' = b_{t_k}$ for all k, subtracting gives $\mathcal{A}_{\rho_{t_k}}(F - F') = 0$ for every k, i.e. $F - F' \in \ker\mathcal{A}_K$; the converse is immediate by linearity. A continuum of times: the same argument at every t gives the intersection over the interval. For the converse, let $h := F' - F$ satisfy $\nabla\cdot(\rho_t h) = 0$ for all t. Then $\partial_t\rho_t + \nabla\cdot(\rho_t F') - \sigma^2\Delta\rho_t = \partial_t\rho_t + \nabla\cdot(\rho_t F) - \sigma^2\Delta\rho_t = 0$, so ρ_t solves the Fokker–Planck equation for F' as well; with the same initial law and uniqueness of the forward equation, F' generates exactly the path $\{\rho_t\}$. □

The converse fails for finitely many instants, which is the content of Rem. 20 below and the reason §2 distinguishes source constraints from the snapshot experiment.

Remark 20 (Path operator, not snapshot operator)**.** Always $\ker\mathcal{A}_{[0,T]} \subseteq \ker\mathcal{A}_K \subseteq \ker\mathcal{A}_{\rho_{t_1}}$, where $\mathcal{A}_{[0,T]}F := (\mathcal{A}_{\rho_t}F)_{t\in[0,T]}$. Membership in $\ker\mathcal{A}_K$ is not sufficient for path equivalence: annihilating $\nabla\cdot(\rho_{t_k}h)$ at K instants leaves ρ_t free to solve a different equation between them, so the paths — and hence later marginals — separate. Only the continuum condition makes ρ_t solve the perturbed equation. At first order around a *stationary* base all three operators collapse onto $\mathcal{A}_\rho$, which is why the distinction is easy to miss.

Verification of Ex. 5. For $F_\omega(x) = (-aI + \omega J)x$ with isotropic diffusion $\sigma^2 I$, the covariance solves $\dot\Sigma = (-aI + \omega J)\Sigma + \Sigma(-aI + \omega J)^\top + 2\sigma^2 I$, whose solution with $R_\theta := e^{\theta J}$ is the displayed Σ_t, using $J^\top = -J$ and $R_\theta R_\theta^\top = I$. At $t_k = k\Delta$, replacing ω by $\omega + \pi/\Delta$ replaces $R_{\omega t_k}$ by $R_{\omega t_k + \pi k} = (-1)^k R_{\omega t_k}$, and $(-1)^k R\Sigma_0 R^\top (-1)^k = R\Sigma_0 R^\top$; every centered Gaussian snapshot is therefore unchanged. The drift difference is $(\pi/\Delta)Jx$. A linear field Gx is ρ-solenoidal for $\rho = \mathcal{N}(0,\Sigma)$ iff $\operatorname{tr} G = 0$ and $\Sigma^{-1}G$ is antisymmetric (Lem. 26). For $G = J$ and $\Sigma_0 = \operatorname{diag}(\sigma_1^2, \sigma_2^2)$, $\Sigma_0^{-1}J = \left(\begin{smallmatrix} 0 & -\sigma_1^{-2} \\ \sigma_2^{-2} & 0 \end{smallmatrix}\right)$, antisymmetric iff $\sigma_1 = \sigma_2$. For anisotropic Σ_0 the difference is therefore not in $\ker\mathcal{A}_{\rho_0}$, so $\ker\mathcal{A}_K$ is not necessary for snapshot equivalence. With $a = 0.4$, $\sigma^2 = 0.3$, $\Delta = 0.5$, $\omega = 0.8$ and $\Sigma_0 = \operatorname{diag}(2, 0.7)$, the maximum covariance discrepancy over five sampled times is 2.0×10^{-14} while the discrepancy halfway to the next snapshot is 1.505. □

Proof of Prop. 6. If $h \in \ker\mathcal{A}_K$ then $\nabla\cdot h + h\cdot s_k = 0$ for every k. Subtracting the $k = 1$ equation from the k-th eliminates $\nabla\cdot h$ and leaves $h(x)\cdot(s_k(x) - s_1(x)) = 0$ pointwise. Conversely these $K - 1$ conditions together with $\mathcal{A}_{\rho_{t_1}}h = 0$ reconstitute all K equations. At each x the conditions say $h(x) \perp \operatorname{span}\{s_k(x) - s_1(x)\}_{k\geq 2} =: \mathcal{C}_x^\perp$, a subspace of dimension $r(x)$, so $h(x)$ is confined to $\mathcal{C}_x$ of dimension $m - r(x)$. This is a constraint on the *value* at each point; the unknown is a section $x \mapsto h(x) \in \mathcal{C}_x$ in $L^2(\rho_{t_1})$, and the space of such sections is infinite-dimensional whenever $m - r(x) \geq 1$ on a set of positive measure. The differential condition $\mathcal{A}_{\rho_{t_1}}h = 0$ then removes

further sections; Thm. 8 shows that on polynomials it removes all of them once the shapes are rich. □

Proposition 21 (A time-conditioned curl head is exactly compensable)**.** *Let a model represent $F_t = -\nabla\Phi_t + h_t$ with h_t an unconstrained time-conditioned ρ_t-solenoidal head, and suppose the marginal path satisfies the hypotheses of Prop. 1. Then for* every *choice of $\{h_t\}$ there is a $\{\Phi_t\}$ reproducing the observed path exactly. The circulation is therefore unidentifiable within such a model even given oracle nuisances.*

Proof. Fix $\{h_t\}$ with $\nabla\cdot(\rho_t h_t) = 0$. We need Φ_t with $\partial_t\rho_t + \nabla\cdot(\rho_t(-\nabla\Phi_t + h_t)) - \sigma^2\Delta\rho_t = 0$. Since $\nabla\cdot(\rho_t h_t) = 0$ this reduces to $\nabla\cdot(\rho_t\nabla\Phi_t) = \partial_t\rho_t - \sigma^2\Delta\rho_t$, whose right-hand side integrates to zero, so Prop. 1 applies verbatim at each t and supplies Φ_t. The compensation is per-slice and independent across t. □

Autonomy is exactly what breaks this: a single F must satisfy K constraints with one potential, so the compensator would have to solve K Poisson problems simultaneously with a shared solution, which Thm. 8 shows is impossible on polynomials under shape richness.

Proposition 22 (Rotation in an isotropic block is an exact gauge)**.** *Let the state split as $\mathbb{R}^m = E\oplus E^\perp$ with $\Sigma_0|_{E^\perp} = \varsigma^2 I$, let G be antisymmetric supported in $E^\perp$, and let the reference dynamics act blockwise. Then G leaves every marginal unchanged, so it lies in* $\ker\mathcal{A}_K$ *for any K and any snapshot times.*

Proof. With $Q_t = e^{Gt}$ orthogonal and supported in $E^\perp$, the covariance contribution from that block is $Q_t(\varsigma^2 I)Q_t^\top = \varsigma^2 Q_t Q_t^\top = \varsigma^2 I$, unchanged for every t; the isotropic part of the stationary covariance is likewise Q_t-invariant. Hence Σ_t and every centered Gaussian marginal are identical with and without G. Numerically, planting a rotation of Frobenius norm 0.4 in the isotropic normal block of the $d = 12$ benchmark leaves the maximal off-diagonal entry of that block below 10^{-16} at all five snapshot times, whereas a tangent-block rotation is visible. □

D Weighted-solenoidal Hermite algebra

Fix $\rho_\star = \mathcal{N}(0, I_m)$. Write $\delta_j := x_j - \partial_j$ for the Gaussian creation operator, so that $\delta_j\Psi_\beta = \sqrt{\beta_j + 1}\,\Psi_{\beta+e_j}$ and $\partial_j\Psi_\beta = \sqrt{\beta_j}\,\Psi_{\beta-e_j}$; the δ_j commute, and $x_j = \delta_j + \partial_j$ on polynomials.

Prop. 7 is Lem. 23(i) together with Lem. 24.

Lemma 23 (Degree raising and the Koszul model)**.** *(i) $\mathcal{A}_{\rho_\star}h = -\delta\cdot h := -\sum_i \delta_i h_i$ for every polynomial field h; in particular $\mathcal{A}_{\rho_\star}(\Psi_\beta e_i) = -\sqrt{\beta_i+1}\,\Psi_{\beta+e_i}$. (ii) The linear map $\iota : \Psi_\beta \mapsto x^\beta/\sqrt{\beta!}$, extended componentwise to fields, is an isomorphism of polynomial spaces intertwining δ_j with multiplication by x_j and ∂_j with ∂_j. Consequently ι maps S_q onto the homogeneous fields g of degree $q-1$ with $x\cdot g \equiv 0$.*

Proof. (i) $\mathcal{A}_{\rho_\star}h = \nabla\cdot h + h\cdot\nabla\log\rho_\star = \sum_i(\partial_i - x_i)h_i = -\sum_i\delta_i h_i$, and the action on $\Psi_\beta e_i$ is the creation identity. (ii) On unnormalized Hermite polynomials $\delta_j H_\beta = H_{\beta+e_j}$ and $\partial_j H_\beta = \beta_j H_{\beta-e_j}$, which are the actions of $x_j\cdot$ and ∂_j on x^β; normalization is a diagonal change of basis. By (i), $h\in S_q$ iff $\delta\cdot h = 0$, which ι carries to $x\cdot g = 0$. □

Lemma 24 (Surjectivity and the gauge dimension)**.** *For $q\ge 1$, $\mathcal{A}_{\rho_\star}|_{W_q} : W_q \to \mathcal{P}_q$ is onto, so for $q\ge 2$ $\dim S_q = mD_{q-1} - D_q \asymp q^{m-1}$ and $\dim S_q/\dim W_q \to (m-1)/m$.*

Proof. Given $|\alpha| = q$ choose i with $\alpha_i \ge 1$; then $\mathcal{A}_{\rho_\star}(\Psi_{\alpha-e_i}e_i) = -\sqrt{\alpha_i}\,\Psi_\alpha$ by Lem. 23(i). Rank–nullity gives the dimension, and $D_q/D_{q-1} = (q+m-1)/q \to 1$ gives the ratio. □

The limit $(m-1)/m$ is a reference-weighted gauge fraction and should not be confused with any fraction computed in the Euclidean divergence-free class, a different subspace with a different Stein operator. On S_q the reference Stein operator vanishes identically by definition, which is why a stationary spectrum computed in that other class cannot verify anything about S_q.

Proposition 25 (Structure of the single-slice gauge). *For $q \geq 2$, S_q is exactly the image of an antisymmetric matrix potential,*

$$S_q = \left\{ h_i = \textstyle\sum_j \delta_j a_{ij} \;:\; a_{ij} = -a_{ji} \in \mathcal{P}_{q-2} \right\} = \left\{ \rho_\star^{-1} \nabla \cdot (\rho_\star a) \;:\; a^\top = -a,\; a_{ij} \in \mathcal{P}_{q-2} \right\},$$

and the parameterization $a \mapsto h$ has kernel $\{(\sum_k \delta_k b_{ijk})_{ij} : b \text{ totally antisymmetric}\}$.

Proof. If $h_i = \sum_j \delta_j a_{ij}$ with a antisymmetric then $\delta \cdot h = \sum_{ij} \delta_i \delta_j a_{ij} = 0$, since the δ's commute and a is antisymmetric, so the image lies in S_q. Under ι the claim becomes exactness of the Koszul complex of $(x_1, \dots, x_m)$ on polynomials: a homogeneous field g with $x \cdot g = 0$ is $g_i = \sum_j x_j a_{ij}$ for some antisymmetric a of one lower degree, unique modulo $a_{ij} = \sum_k x_k b_{ijk}$ with b totally antisymmetric. The second description is $\rho_\star^{-1} \partial_j (\rho_\star a_{ij}) = (\partial_j - x_j) a_{ij} = -\delta_j a_{ij}$, a sign convention. □

Lemma 26 (Linear fields). *For $\rho = \mathcal{N}(0, \Sigma)$ and $h(x) = Gx$, h is ρ-solenoidal iff $\Sigma^{-1} G$ is antisymmetric; equivalently $G = \Sigma N$ with $N^\top = -N$, equivalently $G\Sigma + \Sigma G^\top = 0$. For $\Sigma = I$ this is $G^\top = -G$.*

Proof. $\mathcal{A}_\rho(Gx) = \operatorname{tr} G - x^\top G^\top \Sigma^{-1} x$, which vanishes identically iff $\operatorname{tr} G = 0$ and $G^\top \Sigma^{-1} + \Sigma^{-1} G = 0$. With $N := \Sigma^{-1} G$ the second condition is $N^\top = -N$, and then $\operatorname{tr} G = \operatorname{tr}(\Sigma N) = 0$ automatically, as the trace of a symmetric times an antisymmetric matrix. Multiplying $G^\top \Sigma^{-1} + \Sigma^{-1} G = 0$ by Σ on both sides gives $G\Sigma + \Sigma G^\top = 0$. □

Lem. 26 is the bridge to Guan et al. (2024): their Σ-generalized rotations solve $A\Sigma + \Sigma A^\top = 0$, so the ρ-solenoidal linear fields are exactly those rotations (App. I).

Multiplication by x_j is not degree-preserving in the Hermite basis, since $x_j = \delta_j + \partial_j$ raises and lowers. This is what makes the cross-slice operator banded rather than block diagonal.

Lemma 27 (Raising/lowering split). *For symmetric B and $h \in W_q$ put $R_B h := -\delta \cdot (Bh) \in \mathcal{P}_q$ and $L_B h := -\nabla \cdot (Bh) \in \mathcal{P}_{q-2}$. Then $-h \cdot Bx = R_B h + L_B h$, the two summands are orthogonal in $L^2(\rho_\star)$, and under ι the operator R_B becomes $g \mapsto -g \cdot Bx$.*

Proof. By symmetry of B, $-h \cdot Bx = -\sum_{ij} B_{ij} x_j h_i = -\sum_j x_j (Bh)_j = -\sum_j (\delta_j + \partial_j)(Bh)_j$. The two sums lie in different chaoses, hence are orthogonal, and the last claim is Lem. 23(ii). □

E GAUGE CONTRACTION UNDER CHANGING SHAPE

Throughout, the compared slices are Gaussian with a common mean and are whitened at $t_\star$ as in §3.1 — the reference slices in the local analysis, or the observed marginals when these are themselves Gaussian — and $\mathcal{A}_K$ denotes their stacked Stein operator ($\bar{\mathcal{A}}_K$ in §3.1), so $s_k - s_\star = -\Xi_k x$ exactly. Write $T_K := \bigoplus_q T_{K,q}$ on $\mathcal{S}_{\leq Q}$ and $R_{K,q} := (R_{\Xi_k})_{k \neq \star}$ for its raising part.

Proposition 28 (Structure of the persistent gauge). *(i)* $\mathcal{S}_{\leq Q} \cap \ker \mathcal{A}_K = \ker(T_K|_{\mathcal{S}_{\leq Q}})$, *and* $\ker \mathcal{A}_K \subseteq \ker \mathcal{A}_{\rho_\star}$.

(ii) $\dim \ker(T_K|_{\mathcal{S}_{\leq Q}}) \leq \sum_{q \leq Q} \dim \ker R_{K,q}$.

(iii) Under ι, $\ker R_{K,q} \cong \{g \in \mathbb{R}[x]^m_{q-1} : g(x) \perp \operatorname{span}\{x, \Xi_1 x, \dots, \Xi_r x\}\ \forall x\}$.

(iv) $\ker R_{K,q} = \{0\}$ *for every q iff the shapes are shape-rich in the sense of Thm. 8; generically this holds iff $r \geq m - 1$.*

(v) For $r = m - 2$ and generic shapes, $\ker R_{K,q} = \{p\, w : p \in \mathcal{P}_{q-m}\}$ *under ι, of dimension D_{q-m}, with w as in Lem. 29.*

Proof. (i) For $h \in S_q$, $\mathcal{A}_{\rho_k} h = \mathcal{A}_{\rho_\star} h + h \cdot (s_k - s_\star) = h \cdot (s_k - s_\star)$, so $\mathcal{A}_K h = (T_K h, 0)$; and $\rho_\star$ is itself observed.

(ii) By Lem. 27, T_K maps S_q into $\mathcal{P}_q \oplus \mathcal{P}_{q-2}$, with the $\mathcal{P}_q$ part equal to $R_{K,q}$. If $0 \neq h = \sum_{q \le Q'} h_q \in \ker T_K$ with top component $h_{Q'} \neq 0$, the only contribution to the $\mathcal{P}_{Q'}$ component of $T_K h$ is $R_{K,Q'} h_{Q'}$, so $h_{Q'} \in \ker R_{K,Q'}$. The map $h \mapsto h_{Q'}$ has kernel $\ker(T_K|_{\mathcal{S}_{\le Q'-1}})$; induct on Q'.

(iii) By Lem. 27, $R_{K,q} h = 0$ iff $g \cdot \Xi_k x \equiv 0$ for all k, and by Lem. 23, $h \in S_q$ iff $g \cdot x \equiv 0$; by Prop. 25 such g are exactly the Koszul images $g_i = \sum_j x_j a_{ij}$ of antisymmetric potentials.

(iv) If the span is $\mathbb{R}^m$ on an open set, a polynomial field orthogonal to it vanishes there, hence identically. If $r \le m-2$ the span has dimension at most $r+1 < m$ everywhere, and (v) or its analogue with fewer shapes gives nonzero orthogonal polynomial fields. For genericity when $r = m-1$, the tuples with $\det[\,x \,|\, \Xi_1 x \,|\, \cdots \,|\, \Xi_{m-1} x\,] \equiv 0$ form an algebraic subset, and it is proper: for $\Xi_j = \Xi^j$ with $\Xi = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ and distinct λ_i, the determinant is $\prod_i x_i$ times the Vandermonde determinant of the λ_i, not identically zero.

(v) For $r = m-2$ the pointwise orthogonal complement of the span is generically the line through $w(x)$, whose entries are the maximal minors of $[\,x \,|\, \Xi_1 x \,|\, \cdots \,|\, \Xi_{m-2} x\,]$, homogeneous of degree $m-1$ and without common polynomial factor for generic shapes. A polynomial field proportional to w at every x is then $p\,w$ with p a polynomial, of degree $(q-1)-(m-1) = q-m$. □

Prop. 28(ii) bounds the full gauge by its leading-degree part but does not show that leading elements extend to full kernel elements, because the lowering part of a degree-$(q+2)$ component can cancel the raising part of a degree-q one. Sharpness therefore needs an explicit element.

Lemma 29 (An explicit residual gauge field)**.** *Let $r = m-2$ and define $w_i(x) := \det[\,e_i \,|\, x \,|\, \Xi_1 x \,|\, \cdots \,|\, \Xi_{m-2} x\,]$. Then $\nabla \cdot w = 0$, $x \cdot w = 0$ and $\Xi_k x \cdot w = 0$ for every k, so $w \in \ker \mathcal{A}_K$; w is a polynomial field of degree $m-1$, hence lies in $\mathcal{S}_{\le m}$, and it is nonzero for generic shapes. More generally $p(\varphi_0, \ldots, \varphi_{m-2})\, w \in \ker \mathcal{A}_K$ for every polynomial p, where $\varphi_0 := |x|^2/2$ and $\varphi_k := x^\top \Xi_k x/2$.*

Proof. For any v, $v \cdot w = \det[\,v \,|\, x \,|\, \Xi_1 x \,|\, \cdots\,]$, which vanishes when v repeats a column; this gives the three orthogonality relations. For the divergence, differentiate the determinant column by column: $\partial_i w_i$ is a sum of terms $\sum_i \det[\,e_i \,|\, \cdots \,|\, M e_i \,|\, \cdots\,]$ with $M \in \{I, \Xi_1, \ldots, \Xi_{m-2}\}$ in one column and the others fixed. Each such sum is $\sum_i \Omega(e_i, M e_i) = \mathrm{tr}(\Omega M)$ for the antisymmetric bilinear form Ω obtained by fixing the other columns, and it vanishes because M is symmetric. Hence $\mathcal{A}_{\rho_\star} w = \nabla \cdot w - x \cdot w = 0$ and $\mathcal{A}_{\rho_k} w = \nabla \cdot w - w \cdot (I + \Xi_k) x = 0$. For the multiplied field, $\mathcal{A}_{\rho_\star}(pw) = \nabla p \cdot w + p\, \mathcal{A}_{\rho_\star} w = 0$ because $\nabla p \in \mathrm{span}\{x, \Xi_1 x, \ldots, \Xi_{m-2} x\}$, and the cross-slice conditions are unchanged. Nonvanishing: $w(x) = 0$ iff the $m-1$ columns are dependent, which for generic shapes fails on an open set. □

Proof of Thm. 8. Sufficiency: under shape richness Prop. 28(iv) gives $\ker R_{K,q} = \{0\}$ for every q, and (ii) then gives $\ker(T_K|_{\mathcal{S}_{\le Q}}) = \{0\}$, which by (i) is the claim. Genericity is (iv). Sharpness is Lem. 29: with $r = m-2$, i.e. $K = m-1$ distinct shapes, a nonzero polynomial gauge direction survives. □

For $m = 3$ and one shape change, $w(x) = x \times \Xi_1 x$ is the vector field of Euler's equations for a free rigid body. It is divergence-free, tangent to the spheres $|x| = \text{const}$, and tangent to the ellipsoids $x^\top \Xi_1 x = \text{const}$, which is exactly why neither Gaussian slice can see it.

E.1 One shape below the threshold

By Prop. 28(ii) and (v), for $r = m-2$ the residual gauge gains at most D_{q-m} new directions at each degree q, so $\dim(\mathcal{S}_{\le Q} \cap \ker \mathcal{A}_K) \le \sum_{q \le Q} D_{q-m}$. Since $D_{q-m} \sim q^{m-1}/(m-1)!$ while $\dim S_q \sim (m-1) q^{m-1}/(m-1)!$, the bound is a fraction tending to $1/(m-1)$ of $\dim \mathcal{S}_{\le Q}$ — one half at $m = 3$. The bound is attained in every case we computed. Solving the full linear system $\{\nabla \cdot h - x \cdot h = 0,\ h \cdot \Xi_k x = 0\}$ on polynomial fields of degree at most ℓ with random integer shapes gives

m, r	degrees ℓ	$\dim(\mathcal{S}_{\le \ell+1} \cap \ker \mathcal{A}_K)$
3, 0	1–5	$3, 11, 26, 50, 85 = \sum_q \dim S_q$
3, 1	1–8	$0, 1, 4, 10, 20, 35, 56, 84 = \sum_q D_{q-3}$
3, 2	1–8	0 throughout
4, 2	1–6	$0, 0, 1, 5, 15, 35 = \sum_q D_{q-4}$
4, 3	1–6	0 throughout

Whether equality holds in general is open. The fields of Lem. 29 account for only a small part of the residual gauge: their number grows like q^{m-2}, against q^{m-1} for the bound, so most residual directions have leading part $p\,w$ with p not a function of the quadratic invariants, completed at lower degrees.

A degree cap below the threshold. Fewer shapes do not remove identifiability; they cap its degree. For $r \le m-2$ and constant vectors $c_1, \dots, c_{m-r-2}$, the field $(w_c)_i(x) := \det[\, e_i \,|\, x \,|\, \Xi_1 x \,|\, \cdots \,|\, \Xi_r x \,|\, c_1 \,|\, \cdots \,|\, c_{m-r-2}\,]$ lies in $\ker \mathcal{A}_K$ by the proof of Lem. 29 verbatim, constant columns not entering the divergence; it is homogeneous of degree $r+1$ and nonzero for generic shapes and c. With $K = r+1$ distinct generic slices an exact polynomial gauge of degree K therefore survives. Conversely, if $\ker R_{K,q} = \{0\}$ for every $q \le K$, then $\mathcal{S}_{\le K} \cap \ker \mathcal{A}_K = \{0\}$ by Prop. 28(ii), and every polynomial drift of degree at most $K-1$ is identified by the source constraints, since the difference of two such drifts lies in that intersection. For $r = m-2$ this is (v), for every m. For $r < m-2$ we certify it exactly for $m \le 6$: for one draw of integer shapes the homogeneous system of (iii) has full column rank modulo the prime $2^{31}-1$ at every degree $q-1 \le r$, which implies full rank over $\mathbb{Q}$ and hence, full rank being a Zariski-open condition, for generic shapes; at degree $r+1$ its kernel is nonzero, as w_c requires. The general case is open. The number of distinct shapes thus sets the polynomial degree to which the drift is identified, and $K \ge m$ removes the cap.

This is the algebraic counterpart of Obs. 17. When the shapes stop moving the failure is not a gradual loss of precision: one shape short of the threshold a fixed fraction of every high-degree block reappears in the gauge, and with no shape change ($r = 0$) all of it does.

F Linearization, propagation, and the bridge assumptions

F.1 The tangent equation

Write $\rho_t^\delta = \bar\rho_t(1 + \delta u_t) + o(\delta)$ for the system with drift $F^\delta = \bar F + \delta g$. Substituting into $\partial_t \rho = -\nabla\cdot(\rho F) + \sigma^2 \Delta \rho$, subtracting the reference equation and dividing by $\bar\rho_t$ gives, to first order,

$$\partial_t u_t = \mathcal{L}_t u_t - \mathcal{A}_{\bar\rho_t} g, \qquad \mathcal{L}_t u := -\bar F \cdot \nabla u + \sigma^2 \big(\Delta u + 2\bar s_t \cdot \nabla u\big),$$

which is the first line of (3); with $u_0 = 0$ Duhamel's formula gives the second. If the initial law is unknown, $u_0 \ne 0$ propagates as $\mathcal{U}(t_k, 0)u_0$, an additional nuisance to be profiled. The map $g \mapsto (u_{t_1}, \dots, u_{t_K})$ is the derivative of the snapshot map $\mathscr{S}_K$ at $\bar F$, so local identifiability in the snapshot experiment (§2.2) is a statement about the column rank of $\mathcal{B}_K$.

Lemma 30 (Hermite truncation is closed under propagation)**.** *For the Gaussian reference, $\mathcal{L}_t$ maps $\mathcal{P}_q$ into $\mathcal{P}_q \oplus \mathcal{P}_{q-2}$ and never raises degree. Hence $\mathcal{U}(t,s)$ preserves $\mathcal{P}_{\le Q}$ and is invertible on it, and for $g \in W_{\le Q}$ the Duhamel integral lies in $\mathcal{P}_{\le Q}$.*

Proof. Each term of $\mathcal{L}_t$ is a differential operator with constant or linear coefficients: $-\bar F \cdot \nabla = -\sum_{ij} \bar A_{ij} x_j \partial_i$ for a linear reference drift, $\sigma^2\Delta$, and $2\sigma^2 \bar s_t \cdot \nabla = -2\sigma^2 \sum_{ij} (\bar\Sigma_t^{-1})_{ij} x_j \partial_i$. Each ∂ lowers degree by one and $x_j = \delta_j + \partial_j$ changes it by ± 1, so each term changes degree by 0 or -2. A degree-nonincreasing generator has a degree-nonincreasing, block lower triangular propagator, invertible because its diagonal blocks are, and $\mathcal{A}_{\bar\rho_s}$ maps $W_{\le Q}$ into $\mathcal{P}_{\le Q}$ by Lem. 27. □

Lem. 30 says a *truncated input* propagates exactly. It does not say that low-degree observations exclude high-degree inputs, and they do not (App. F.3).

F.2 How the snapshot kernel relates to the source kernel

Three kernels are in play. If $\mathcal{A}_{\bar{\rho}_s} h = 0$ for every s — the path constraints — then every Duhamel integrand vanishes and $\mathcal{B}_K h = 0$, so the path kernel lies in $\ker \mathcal{B}_K$ and in $\ker \bar{\mathcal{A}}_K$. Between $\ker \mathcal{B}_K$ and $\ker \bar{\mathcal{A}}_K$ there is no inclusion in general.

Source-invisible but snapshot-visible. A direction with $\mathcal{A}_{\bar{\rho}_{t_k}} h = 0$ at every observed time can have $\mathcal{A}_{\bar{\rho}_s} h \neq 0$ for s between them, and propagation accumulates that footprint. The oracle design is informed by the continuum of shapes $\{\bar{\Sigma}_s\}$ traversed along the path, whereas $T_{K,q}$ sees only the K observed shapes; this is the precise sense in which an oracle that knows the reference trajectory is stronger than an estimator that sees only snapshots.

Source-visible but snapshot-invisible. Because $\mathcal{U}(t, s)$ is invertible on each truncation (Lem. 30), $\mathcal{B}_K$ can fail to be injective on directions the source constraints see only through cancellation in the s-integral — an analytic condition on the path. That is a first-order form of the aliasing in Ex. 5, and it is what the faithfulness condition of §3.2 excludes, together with the analogous possibility that the profiling projection in (4) cancels a solenoidal direction exactly.

Numerically, for an autonomous Gaussian reference with $\dot{\bar{\Sigma}} = \bar{A}\bar{\Sigma} + \bar{\Sigma}\bar{A}^\top + 2\sigma^2 I$, building $\mathcal{B}_K$ by integrating the tangent equation from $u_0 = 0$ gives $\ker T_{K,q} = \ker \mathcal{J}_{\mathrm{sol},q} = \{0\}$ in every shape-rich case tested ($m = 3$, $K = 4$, $Q \leq 6$; $m = 4$, $K = 5$, $Q \leq 5$; two reference paths each), so the faithfulness condition holds there with equality. Without shape richness the inclusion is strict in the favourable direction: at $m = 3$, $K = 2$ the source gauge has dimension 10 while the profiled kernel has dimension 1 and 0 on two paths, and at $m = 4$, $K = 2$ the dimensions are 18 against 7 and 6. The directions propagation recovers are recovered weakly — the profiled spectrum has $\kappa_{\min}/\kappa_{\max} \approx 10^{-5}$ against $\approx 10^{-1}$ for the source operator — so shape-poor paths convert exact gauge into near-gauge rather than into usable information.

F.3 Propagated aliasing: the guard band does not transfer

The instantaneous source has only raising and lowering bands, and $P_{\leq Q-2} T_{>Q} = 0$ for the source operator. This does *not* justify the same statement for $\mathcal{B}_K$: exponentiating a degree-lowering generator produces arbitrarily many downward bands, so invariance of $\mathcal{P}_{\leq Q}$ proves exact propagation of a truncated *input* but not that projecting observations to low degree excludes high-degree inputs.

The leakage is nonzero. Take the Brownian reference $\bar{F} = 0$, $\sigma^2 = 1$, $\bar{\Sigma}_t = (1+2t) I_2$, and with H_j the unnormalized probabilists' Hermite polynomials set $h = (x_2 H_5(x_1), -H_6(x_1)) \in S_7$, which satisfies $\mathcal{A}_{\rho_\star} h = 0$ at $\rho_\star = \mathcal{N}(0, I_2)$. The transient source is $\mathcal{A}_{\bar{\rho}_t} h = 5(1 - \frac{1}{v}) x_2 H_4(x_1)$ with $v = 1 + 2t$. Writing $u_t = x_2[c_0(t) + c_2(t) H_2(x_1) + c_4(t) H_4(x_1)]$, the tangent generator gives

$$\dot{c}_4 = -\tfrac{10 c_4}{v} - 5\big(1 - \tfrac{1}{v}\big), \quad \dot{c}_2 = -\tfrac{6 c_2}{v} + 12\big(1 - \tfrac{2}{v}\big) c_4, \quad \dot{c}_0 = -\tfrac{2 c_0}{v} + 2\big(1 - \tfrac{2}{v}\big) c_2.$$

From zero initial conditions, $c_0(0.4) = -7.7213162 \times 10^{-3}$: a degree-7 field reaches scalar degree 1. The degreewise benchmark (5) has no such term by construction, since it treats degrees as independent; in the linearized experiment high-degree leakage is an additional bias, one of the transfer steps of Rem. 13, and its control is open (App. A).

F.4 Joint versus blockwise information

Asm. 9 bounds the jointly profiled information above by the blockwise one. The reverse bound, which would be needed to carry the upper bound of Thm. 12 from the benchmark to the tangent experiment (App. H.7), is not assumed, and it can fail: good information in every block does not control the joint inverse trace. With

$$A = \begin{pmatrix} 1 & 1 \\ 0 & \delta \end{pmatrix},$$

the two coordinates have individual information 1 and $1 + \delta^2$, whereas $\mathrm{tr}[(A^\top A)^{-1}] = 1 + 2/\delta^2$, which diverges as $\delta \to 0$. Because propagation is not Hermite-degree diagonal — $\mathcal{U}(t, s)$ mixes degrees, and profiling removes only G_q within each block rather than gradients at other degrees or guard-band nuisance fields — overlapping output degrees can in principle create exactly this configuration. Neither direction of the comparison has been verified numerically, and both are open.

G NEAR-GAUGE SPECTRAL GEOMETRY AND THE THIN TAIL

G.1 FROM A THIN TAIL TO THE INVERSE-TRACE LAW

Lemma 31 (Trace law). *Under Asm. 10 and 11, every $\kappa_{q,j}^2$ is positive and*

$$\frac{r_q}{\bar{\mu}_q} \le \operatorname{tr}\big(\mathcal{J}_{\mathrm{sol},q}^{-1}\big) \le \Big(1+\frac{C'}{\gamma-1}\Big)\frac{r_q}{\bar{\mu}_q} \asymp \frac{q^m}{K}.$$

Proof. Since $\mathcal{N}_q(t) \le C't^\gamma \to 0$ as $t \to 0$ and $\mathcal{N}_q$ takes values in $r_q^{-1}\mathbb{Z}$, no eigenvalue is zero. The lower bound is Cauchy–Schwarz, $r_q^2 = (\sum_j \kappa_{q,j}\kappa_{q,j}^{-1})^2 \le (\sum_j \kappa_{q,j}^2)(\sum_j \kappa_{q,j}^{-2})$. For the upper bound put $t_j := \kappa_{q,j}^2/\bar{\mu}_q$; the layer-cake formula gives $r_q^{-1}\sum_j t_j^{-1} = \int_0^\infty r_q^{-1}\#\{j : t_j < 1/u\}\,du \le 1 + \int_1^\infty \mathcal{N}_q(1/u)\,du \le 1 + C'\int_1^\infty u^{-\gamma}du$, finite exactly because $\gamma > 1$. At $\gamma = 1$ the integral is cut off at $u \le C'r_q$, since $\mathcal{N}_q(t) < r_q^{-1}$ forces $\mathcal{N}_q(t) = 0$, giving an extra $\log r_q \asymp \log q$. □

The thin tail bounds no individual eigenvalue: what it excludes is a large population of near-gauge directions, and it needs no degree-uniform spectral gap.

G.2 WHAT HAS BEEN MEASURED, AND ON WHICH OPERATOR

The diagnostics bearing on Asm. 10 and 11 were computed by direct construction in the tensor-Hermite basis, and they are not all computed on the operator the assumptions concern. We separate them.

On the propagated, profiled operator $\mathcal{J}_{\mathrm{sol},q}$. Fitting $\operatorname{tr}(\mathcal{J}_{\mathrm{sol},q}^{-1}) \sim q^p$ with propagation and gradient profiling included gives $p = 3.20$ at $m = 3$, 4.08 at $m = 4$ and 4.76 at $m = 5$, each with $r = m-1$ shapes, against the predicted $p = m$. This is the *conclusion* of Lem. 31, measured where it is used, and it is the evidence the upper bound actually rests on.

On the source operator $T_{K,q}^* T_{K,q}$. The tail exponent itself has been measured only here: $\gamma \approx 1.6$ at $m = 3$ and $\gamma \approx 2.0$–2.2 at $m = 4$ (degrees up to 40 and 14), comfortably above the threshold. For the source operator the inverse trace scales as q^{m-2}, and one shape short of richness it drifts only logarithmically while two short it gains a power. Asm. 11 is stated for $\mathcal{J}_{\mathrm{sol},q}$, so these source-level exponents support it only through the structural relation between the two operators; they are not a measurement of it.

Why the symbol does not settle it. The raising part of the source operator alone has a semiclassical symbol whose lower-tail exponent is $\gamma_{\mathrm{sym}} = 1/2$, below the threshold, and the exponent measured for the raising part drifts downward toward that value as q grows. What restores $\gamma > 1$ for the full source operator is the lowering part of Lem. 27, which the symbol discards. That is also where a proof would have to come from, and no fixed-degree computation substitutes for a bound uniform in q.

H THE TANGENT BENCHMARK: REDUCTION, RATE PROOFS, AND TRANSFER

H.1 REDUCTION TO A GAUSSIAN SHIFT, AND COMPARISON

In the tangent snapshot experiment the gradient part of g is an unrestricted nuisance. For the lower bound it suffices to consider, at a cutoff Q, the subproblem in which $h \in \mathcal{S}_{\le Q}$, the gradient directions of degree at most Q are unrestricted, and all components above degree Q vanish: its parameter set is contained in the full one, so its minimax risk is no larger. With standard Gaussian noise the subproblem is a linear Gaussian model with unrestricted linear nuisance. Projecting Y onto the orthocomplement of the range of the retained gradient columns leaves a Gaussian shift for h with information operator $\mathcal{J}_{\mathrm{sol},\le Q}$ of Asm. 9, and this reduction loses nothing for minimax risk, by the usual limit of least favourable priors on the nuisance with growing variance.

Lemma 32 (Comparison of Gaussian shifts). *For $J \succeq 0$ on a finite-dimensional parameter space, let $\mathcal{E}_J$ be the Gaussian shift with sufficient statistic $S \sim \mathcal{N}(J\theta, J)$. If $J_2 \preceq J_1$, then $\mathcal{E}_{J_2}$ is a*

randomization of $\mathcal{E}_{J_1}$, so every minimax risk in $\mathcal{E}_{J_1}$ is at most the corresponding one in $\mathcal{E}_{J_2}$. For estimators linear in S the same conclusion holds when the Gaussian law is replaced by any noise with the same mean and covariance.

Proof. If $J_1 x = 0$ then $0 \le x^\top J_2 x \le x^\top J_1 x = 0$, so $J_2 x = 0$: hence $\ker J_1 \subseteq \ker J_2$ and $J_2 J_1^+ J_1 = J_2$, with J_1^+ the pseudo-inverse. Put $M := J_2 J_1^+$. Then MS_1 has mean $J_2\theta$ and covariance $J_2 J_1^+ J_2$, and $J_2 - J_2 J_1^+ J_2 = J_2^{1/2}(I - J_2^{1/2} J_1^+ J_2^{1/2}) J_2^{1/2} \succeq 0$, because $J_2 \preceq J_1$ gives $(J_1^+)^{1/2} J_2 (J_1^+)^{1/2} \preceq I$ and the nonzero spectra of XX^* and X^*X coincide for $X = (J_1^+)^{1/2} J_2^{1/2}$. Adding independent Gaussian noise with that covariance yields exactly the law of the sufficient statistic of $\mathcal{E}_{J_2}$, so any estimator in $\mathcal{E}_{J_2}$ can be run in $\mathcal{E}_{J_1}$ with the same risk. The risk of a linear estimator depends only on the first two moments, which gives the last claim. Scaling J by a constant is the same as scaling n. □

Under Asm. 9 the subproblem at cutoff Q is therefore a randomization of the benchmark (5), restricted to degrees at most Q, with n replaced by Cn. The comparison is stated at every cutoff and one-sidedly because that is exactly what the lower bound uses. A statement about inverse traces alone would not suffice: comparable traces do not prevent the joint experiment from being more informative along the weak directions on which the lower bound is built.

Two conventions. The weights $(1+q)^{2s}$ define the smoothness class and are not identified with a conventional s-derivative norm. And θ are coordinates of h itself: if one writes $F^\delta = \bar{F} + \delta g$ with g fixed, the signal is $\delta\kappa\theta$, n becomes $n\delta^2$, and the risk must be rescaled accordingly. Coordinates with $\kappa_{q,j} = 0$ are unidentified, which only helps the lower bound.

H.2 Where the K/q scale comes from

Asm. 10 is assumed, not proved. The heuristic behind it is a factorization of each slice's contribution to $\operatorname{tr} \mathcal{J}_{\mathrm{sol},q}$ into a source strength and a propagator gain.

Source strength $\asymp q$. For $h \in S_q$ the footprint at time s is $\mathcal{A}_{\bar{\rho}_s} h = -h^\top \Xi_s x$ with $\Xi_s := \bar{\Sigma}_s^{-1} - I$. By Lem. 27 its raising part multiplies by x, and on degree-q Hermite modes multiplication by x_j has norm $\asymp \sqrt{q}$, so $\|\mathcal{A}_{\bar{\rho}_s} h\|^2 \lesssim \|\Xi_s\|^2 q \, \|h\|^2$. A matching lower bound uniform over S_q is exactly what shape richness does not quantify: it makes $R_{K,q}$ injective (Prop. 28) without bounding its smallest singular value.

Propagator gain $\asymp q^{-2}$ per slice. For an Ornstein–Uhlenbeck-type reference with relaxation rate $\nu > 0$, the diagonal part of $\mathcal{L}_t$ on $\mathcal{P}_q$ is $\approx -q\nu$, so on degree-q components $\int_0^{t_k} \mathcal{U}(t_k, s)\, ds \approx (q\nu)^{-1}$ once $q\nu t_k \gg 1$, and the squared gain is $\asymp q^{-2}$. Summing over the K observed slices gives the factor K, provided each slice sees a shape perturbation Ξ_s of order one on the preceding window of length $\asymp (q\nu)^{-1}$.

Multiplying, $\bar{\mu}_q \asymp q \cdot K/q^2 = K/q$. The gaps are explicit: the uniform lower bound on the source strength, uniformity of the damping along a transient path, the requirement that every slice be preceded by genuine shape change — which fails in the short-gap design of Obs. 17 — and a non-degeneracy of profiling against G_q, which can only decrease the trace and could in principle decrease it by more than a constant. Without diffusion there is no damping and the factorization changes. Numerically, the trace law that Asm. 10 and 11 jointly imply is observed on the propagated design (App. G).

Corollary 33 (Bulk consequence)**.** *Under Asm. 10, at least half the degree-q coordinates satisfy $\kappa_{q,j}^2 \le 2\bar{\mu}_q \lesssim K/q$.*

Proof. Markov's inequality for the empirical measure of $\{\kappa_{q,j}^2\}_j$, whose mean is $\bar{\mu}_q$. No tail hypothesis is used. □

H.3 Upper bound

Estimate $\theta_{q,j}$ by $Z_{q,j}/\kappa_{q,j}$ for $q \le Q$ and by zero beyond; this is well defined because Asm. 11 excludes zero eigenvalues (Lem. 31). The risk is

$$\sum_{q\le Q} \frac{1}{n}\operatorname{tr}\big(\mathcal{J}_{\mathrm{sol},q}^{-1}\big) + \sum_{q>Q} \|\theta_q\|^2 \;\lesssim\; \sum_{q\le Q} \frac{q^m}{nK} + R^2(1+Q)^{-2s} \;\asymp\; \frac{Q^{m+1}}{nK} + R^2 Q^{-2s},$$

using Lem. 31 for the first term and the Sobolev constraint in (5) for the second. Taking $Q_\star \asymp (nK)^{1/(2s+m+1)}$ gives risk $\asymp (nK)^{-2s/(2s+m+1)}$. At $\gamma = 1$, Lem. 31 carries an extra $\log q$, which costs a logarithmic factor in the rate. This is the only place Asm. 11 is used, and since the estimator is linear the calculation uses only second moments of the noise.

H.4 Lower bound

The lower bound uses only Cor. 33. Fix a small constant $c > 0$ and consider the band of degrees $q \in [Q_\star/2, Q_\star]$ with $Q_\star := (nK)^{1/(2s+m+1)}$. In each such degree keep the at-least-half of the coordinates with $\kappa_{q,j}^2 \le 2\bar{\mu}_q$, whose noise variance $1/(n\kappa_{q,j}^2)$ is $\gtrsim q/(nK)$, and let $\theta_{q,j}$ range over $[-\tau_q, \tau_q]$ with $\tau_q^2 := c\,q/(nK)$, setting all other coordinates to zero. For c small this hyperrectangle lies in the Sobolev ball, since

$$\sum_{q\in[Q_\star/2,Q_\star]} (1+q)^{2s}\, r_q\, \tau_q^2 \;\lesssim\; c\,Q_\star^{2s}\cdot Q_\star \cdot Q_\star^{m-1}\cdot \frac{Q_\star}{nK} \;=\; c\,\frac{Q_\star^{2s+m+1}}{nK} \;=\; c \;\le\; R^2.$$

The minimax risk over a hyperrectangle in a Gaussian sequence model is at least a constant multiple of $\sum_j \min(\tau_j^2, \sigma_j^2)$ (Donoho et al., 1990; Ibragimov & Has'minskii, 1981), and here $\tau_q^2 \lesssim \sigma_{q,j}^2$ on every kept coordinate, so the risk is

$$\gtrsim \sum_{q\in[Q_\star/2,Q_\star]} \frac{r_q}{2}\cdot\frac{c\,q}{nK} \;\asymp\; \frac{Q_\star^{m+1}}{nK} \;\asymp\; (nK)^{-2s/(2s+m+1)}.$$

Restricting to a single degree $q \asymp Q_\star$ would give only $Q_\star^m/(nK)$, short of the rate by a factor $Q_\star$; the band of $\asymp Q_\star$ degrees is what recovers it.

H.5 Proof of Thm. 12

Lower bound. The hyperrectangle of the lower-bound subsection lives in degrees $q \le Q_\star$. By the reduction above, the minimax risk of the tangent experiment is at least that of the subproblem at cutoff $Q_\star$, a Gaussian shift with information $\mathcal{J}_{\mathrm{sol},\le Q_\star}$. By Asm. 9 and Lem. 32 this is at least the minimax risk of (5) restricted to degrees at most $Q_\star$ at sample size Cn, which the lower-bound subsection bounds below by a constant multiple of $(CnK)^{-2s/(2s+m+1)}$. Constants in n do not change the rate.

Upper bound. This is the upper-bound subsection, stated in (5); the estimator is linear, so only the first two moments of the noise enter.

In both parts the $L^2(\rho_\star)$ loss equals the coordinate loss, because the eigenbases are orthonormal in $L^2(\rho_\star)$ and the Sobolev weights are constant on each degree.

H.6 Where the $+1$ comes from

Direct regression on the $\asymp Q^m$ Hermite coefficients up to degree Q, each with unit inverse information, gives exponent $2s + m$. Here each degree-q direction has inverse information $\asymp q/K$, one power of q worse, which multiplies the cumulative variance by Q and gives $2s + m + 1$. In the factorization of App. H.2, the propagator's degree-proportional damping costs two powers and the degree-raising footprint returns one. This is bookkeeping under Asm. 10; the proof is the two bounds above.

H.7 From the benchmark to the tangent experiment, and to sampled snapshots

The upper bound of Thm. 12 is stated in the benchmark. Carrying it to the tangent experiment along the route of this appendix would need two further ingredients. The first is the reverse comparison $\mathcal{J}_{\mathrm{sol},\le Q} \succeq c\,\mathcal{J}_Q^{\mathrm{blk}}$, uniformly in Q; with it, the second-moment part of Lem. 32 transfers the variance of the cutoff estimator with n replaced by cn. It is not assumed, and it can fail (App. F.4). The second is control of omitted-degree leakage. Components above the cutoff reach the retained statistics through propagation, which lowers degree (App. F.3), so the bias of a cutoff estimator is not just the Sobolev tail. For the solenoidal part this is a boundedness condition on the leakage map; for the gradient part it cannot hold while the nuisance is unrestricted in every degree, so the gradient nuisance above the cutoff must be either restricted or profiled jointly with the retained degrees. Neither ingredient is established.

The nonlinear sampled-snapshot experiment observes $X_i^{(k)} \overset{\mathrm{iid}}{\sim} \rho_{t_k}^F$. Transfer from the tangent experiment would need uniform control of the linearization remainder and of the sampling likelihood through the growing cutoff. A direct lower bound, for example, would need $\mathrm{KL}(\rho_{t_k}, \rho'_{t_k}) = \frac{1}{2}\|u_k - u'_k\|^2_{L^2(\bar\rho_{t_k})}(1+o(1))$ uniformly over the hypercube used in the proof. Finally, the tangent oracle is credited with the reference path entering $\mathcal{B}_K$ up to each observation time, whereas $T_{K,q}$ uses only the source footprints at the observed slices: the oracle knows shapes that the data never show.

H.8 Proof of the covariance-only audit

Proof of Prop. 15. Let the slices be centered Gaussians $\mathcal{N}(0,\Sigma_k)$ and the field class linear, $F = Hx$. By Lem. 26 the footprint at slice k is $\mathcal{A}_{\rho_k}(Hx) = \operatorname{tr} H - x^\top H^\top \Sigma_k^{-1} x$, affine in H, and b_k does not depend on H, so the Hessian of $\frac{1}{2K}\sum_k \mathbb{E}_k[(\mathcal{A}_{\rho_k}(Hx) - b_k)^2]$ is the quadratic form $G \mapsto \frac{1}{K}\sum_k \mathbb{E}_k[(\mathcal{A}_{\rho_k}(Gx))^2]$. Put $M := G^\top \Sigma_k^{-1}$, which is not symmetric in general, and $M_s := (M + M^\top)/2$. Since $x^\top M x = x^\top M_s x$ and $\mathbb{E}_k[x^\top M x] = \operatorname{tr}(M\Sigma_k) = \operatorname{tr} G$, the footprint $\mathcal{A}_{\rho_k}(Gx) = -(x^\top M_s x - \mathbb{E}_k x^\top M_s x)$ is centered. The consolidated fourth-moment identity $\mathbb{E}[(x^\top Ax)(x^\top Bx)] = \operatorname{tr}(A\Sigma)\operatorname{tr}(B\Sigma) + 2\operatorname{tr}(A\Sigma B\Sigma)$ from Isserlis' theorem holds for *symmetric* A, B — for general A, B the last term is $\operatorname{tr}(A\Sigma B\Sigma) + \operatorname{tr}(A\Sigma B^\top \Sigma)$ — so it must be applied to M_s, not M:

$$\mathbb{E}_k[(\mathcal{A}_{\rho_k}(Gx))^2] = 2\operatorname{tr}(M_s\Sigma_k M_s \Sigma_k) = \operatorname{tr}(G^2) + \operatorname{tr}(G^\top \Sigma_k^{-1} G \Sigma_k).$$

Every term is a function of Σ_k alone. With $\vec G := \operatorname{vec}(G)$ column-stacked, $\operatorname{vec}(\Sigma_k^{-1} G \Sigma_k) = (\Sigma_k \otimes \Sigma_k^{-1})\vec G$ and $\operatorname{tr}(G^2) = \vec G^\top \mathsf{C}_m \vec G$, so the Hessian is $\frac{1}{K}\sum_k[\Sigma_k \otimes \Sigma_k^{-1} + \mathsf{C}_m]$; in row-stacked coordinates the Kronecker factors appear in the opposite order. Profiling against the symmetric block is the Schur complement. As a check, the exact gauge $G = \Sigma_k N$ with $N^\top = -N$ (Lem. 26) gives $\operatorname{tr}(\Sigma_k N \Sigma_k N) + \operatorname{tr}(N^\top \Sigma_k N \Sigma_k) = 0$, a null direction of the slice-k term, as it must be. □

The Gaussian hypothesis is essential and its failure is not subtle. In one dimension with $g(x) = x$, a variance-one Gaussian gives $\mathbb{E}[(\mathcal{A}_\rho g)^2] = 2$, whereas a variance-one rescaling of a density proportional to e^{-ax^4} gives 4. The two densities have the same covariance, so no covariance-determined expression can distinguish them: for a general density, linearity of Hx does not make $\mathcal{A}_\rho(Hx)$ covariance-determined. Nonzero means and an arbitrary weighting operator add further dependence. Prop. 15 is therefore a second-moment *design* audit under a Gaussian reference, not a general information audit; a GLS information audit for the weak form needs fourth moments as well.

I The Gaussian specialization

I.1 The commutant criterion

Take the compared slices Gaussian with a common mean, whitened as in §3.1 so that $\rho_\star = \mathcal{N}(0, I_m)$ and the others are $\mathcal{N}(0,\Sigma_k)$; the statements below are exact. Restrict to $q = 2$, so $h(x) = Gx$ with $G^\top = -G$ by Lem. 26. With $\Xi_k = \Sigma_k^{-1} - I$, the cross-slice footprint is $T_{K,2}h = -(x^\top G^\top \Xi_k x)_k = (x^\top G \Xi_k x)_k$, which vanishes for all x iff the symmetric part of $G\Xi_k$ vanishes, i.e.

$$G\Xi_k + \Xi_k^\top G^\top = G\Xi_k - \Xi_k G = [G, \Xi_k] = 0 \iff [G, \Sigma_k^{-1}] = 0.$$

So the degree-two gauge is the commutant of the observed precision family inside the antisymmetric matrices. It is $\{0\}$ exactly when the Σ_k^{-1} have no common nontrivial antisymmetric commutant — and at $q = 2$ a single generic shape difference already suffices: a symmetric matrix with distinct eigenvalues has only symmetric commutant, whose intersection with the antisymmetric matrices is trivial, so $K = 2$ is enough. The $K \geq m$ threshold of Thm. 8 is therefore *not* driven by $q = 2$; it comes from shape richness (Thm. 8), which needs $r \geq m - 1$ and governs all degrees jointly. The theorem gives a sufficient condition uniform in q, and the low-degree blocks are cleared sooner. Both statements are verified in App. S.

The common-mean hypothesis is where the degree separation lives. For a general mean, an affine field $Hx + c$ is invisible under $\mathcal{N}(\mu, \Sigma)$ iff $H\mu + c = 0$ and $H\Sigma + \Sigma H^\top = 0$: with $y := x - \mu$, $\mathcal{A}_\rho(Hx + c) = \operatorname{tr} H - y^\top H^\top \Sigma^{-1} y - (H\mu + c)^\top \Sigma^{-1} y$, and the three terms have different degrees in y. Differing means across slices mix these degrees, which is why Thm. 8 assumes a common one.

I.2 Comparison with Guan et al. (2024)

By Lem. 26 the ρ-solenoidal linear fields are exactly the Σ-generalized rotations of Guan et al. (2024), so the algebra agrees. The statistical experiments do not, in three ways, and the comparison is worth making precisely because it is easy to overstate.

What is conditioned on. Their criterion is a property of the *initial law*: they characterize when the pair (drift, diffusion) is identifiable for the *whole* linear class given p_0, and a Gaussian p_0 is always auto-rotationally invariant, so their theorem declines to certify it. Ours is a property of a fixed *family of Gaussian shapes* — the reference slices in the local analysis, the observed marginals when these are Gaussian — for a given system: we ask whether the particular Σ_k move. There is no contradiction — their non-identifiability witness has zero base drift, so every traversed shape commutes with the perturbation and shape richness fails — but the two statements are not the same statement, and ours is the per-system refinement their own limitations section leaves open.

What is estimated. They identify drift *and* diffusion jointly. We take σ known throughout, which removes their second clause, the mean-zero Gaussian cross-section, since that clause encodes a drift–diffusion rescaling ambiguity that a known σ eliminates. Our comparison is therefore only with the rotational clause.

What is observed. Their result uses marginals on a continuum of times; ours uses K discrete shapes, and Ex. 5 shows the discrete case admits aliasing that the continuum case does not. We therefore do not claim to recover their theorem, and we do not claim our criterion implies theirs.

I.3 Fibrewise conditions become algebraic

Under a fixed finite-dimensional model class the infinite-dimensional difficulty of Prop. 6 disappears: with $F = Hx$ the gauge directions are $h = Gx$, $G^\top = -G$, and the conditions $h(x) \in \mathcal{C}_x$ become $[G, \Sigma_k^{-1}] = 0$, finitely many linear equations in the $\binom{m}{2}$ free parameters of G. This is why the linear benchmark of §5 has an exactly computable gauge and why Prop. 15 can refuse a design before any field is fitted.

J Intrinsic dimension and the normal scale

The rate of Thm. 12 is governed by m, not the ambient dimension d. Suppose the slices concentrate near an m-dimensional affine subspace E with normal width $\tau_\perp$, so that $\Sigma_k = \Sigma_k^{\parallel} \oplus \tau_\perp^2 I_{d-m}$ with $\Sigma_k^{\parallel}$ of order one. Two consequences follow, and we keep the diffusion amplitude σ and the normal width $\tau_\perp$ notationally distinct because they have different physical dimensions.

First, by Prop. 22 every rotation supported in the isotropic normal block is an exact gauge, so the estimable antisymmetric parameters number $\binom{m}{2}$ rather than $\binom{d}{2}$, and the remaining $\binom{d-m}{2}$ directions are declared, not estimated. Second, the typical score magnitude in the normal directions is $O(\tau_\perp^{-1})$ and the corresponding block of the information is $\Theta(\tau_\perp^{-2})$; these are a score magnitude and a precision respectively, and conflating them is how the normal block acquires a spurious reputation

for carrying information about the field. It carries information about the *support*, not the circulation, and recovery is correspondingly flat in d (Fig. 2).

The subspace is recovered rather than assumed: the isotropic bulk of each Σ_k is its maximal-multiplicity eigenvalue cluster, and the curl-visible subspace is the average of the complements' projectors across k. This is basis-free, and §5.3 reports that a random orthogonal rotation of all 12 ambient coordinates leaves the recovered tangent space correct to a principal angle of $2.58°$ and the solenoidal error essentially unchanged (0.6373 against 0.6361, both under the velocity convention of App. Y).

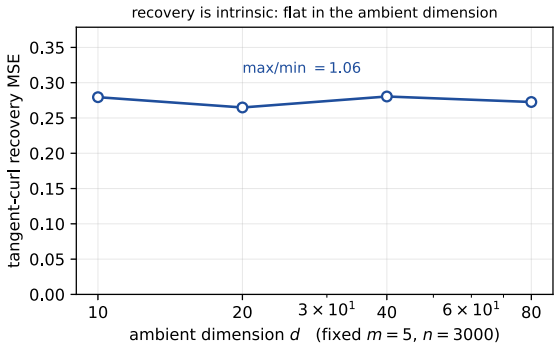


Figure 2: **Recovery is intrinsic.** Tangent-curl recovery mean squared error against ambient dimension d at fixed intrinsic $m = 5$ and $n = 3000$, over 300 trials per point with the parametric estimator. The error is flat in d: the ratio of largest to smallest across $d \in \{10, 20, 40, 80\}$ is close to one. Ambient dimension enters the cost of assembly, not the statistical difficulty of the solenoidal channel.

We do not claim a solution to intrinsic-dimension selection. The audit above assumes a spectral gap between tangent and normal blocks; when that gap is absent the cluster is ill-defined and the method has no principled fallback. Nonlinear intrinsic geometry is discussed in App. K.

K Manifold concentration and rate separation

When the slices concentrate near a curved m-dimensional submanifold rather than an affine subspace, the score acquires a large normal component that encodes support geometry rather than dynamics. Li et al. (2026) show that near a smoothed manifold the score is dominated by the normal projection at scale $\tau_\perp$, with the tangential component an order smaller. In our setting this produces a clean separation.

Proposition 34 (Rate separation under manifold concentration)**.** *In the smoothed-manifold regime with normal width $\tau_\perp \to 0$ and curvature bounded on the region carrying the mass, the footprint decomposes as $\mathcal{A}_\rho g = \mathcal{A}^{\parallel}_\rho g^{\parallel} - \tau_\perp^{-2}\, r\, \langle g^\perp, e_\perp\rangle + O(1)$, where r is the signed normal coordinate, $e_\perp$ the outward unit normal, $\mathcal{A}^{\parallel}$ the intrinsic Stein operator, and the $O(1)$ term is controlled by the curvature bound and the C^1 norm of g. Since $r = O(\tau_\perp)$ on the mass, the normal term has size $\tau_\perp^{-1}$ and its block of the information is $\Theta(\tau_\perp^{-2})$, against $\Theta(1)$ tangentially. Hence the normal component of g is identified at rate $\tau_\perp^2$ times the tangential rate, and the solenoidal estimand, which is tangential by Prop. 22, is governed by the intrinsic operator alone.*

Sketch. Write ρ in tubular coordinates as a product of an intrinsic density and a Gaussian factor of width $\tau_\perp$ in the normal fibre. Then $\nabla \log \rho = \nabla^{\parallel} \log \rho^{\parallel} - \tau_\perp^{-2} r e_\perp + O(1)$ with r the signed normal coordinate, and substituting into $\mathcal{A}_\rho g = \nabla\cdot g + g\cdot\nabla \log\rho$ gives the stated split; the curvature bound controls the cross terms at $O(1)$ — the mean-curvature term from the volume element and the $O(r)$ curvature distortion of tangential derivatives off the manifold. (For higher codimension read $re_\perp$ as the normal displacement.) Because $\mathbb{E}[r^2] = \tau_\perp^2$ under the normal factor, $\mathbb{E}[(\tau_\perp^{-2} r\langle g^\perp, e_\perp\rangle)^2] = \tau_\perp^{-2}\,\mathbb{E}\langle g^\perp, e_\perp\rangle^2$: the normal block of the information is $\Theta(\tau_\perp^{-2})$, which is the claimed separation. Without the factor r it would be $\Theta(\tau_\perp^{-4})$. □

Two cautions. The Euclidean diffusion in (1) is not silently projected onto an estimated manifold; a genuinely intrinsic treatment requires intrinsic differential operators and boundary conditions, which we do not develop. And the separation is a statement about scales, not a selection procedure: choosing m and $\tau_\perp$ from data remains open.

Numerically, on a curved benchmark with normal width $\tau_\perp$ swept over $1, 0.1, 0.03, 0.01$, the within-manifold and moving-frame readouts separate as predicted. At $\tau_\perp = 1$ the two are comparable (0.210 ± 0.030 and 0.201 ± 0.087 over five seeds); at $\tau_\perp = 0.01$ with 10% score error they differ by more than an order of magnitude (0.954 ± 0.085 against 0.028 ± 0.030), and the score-free weak form gives 0.116 and 0.054 on the same design.

L Stationary separable-OU specialization

For a stationary reference with separable space–time structure the propagated operator factorizes, and the temporal channel can be analysed in closed form. Let the reference be an OU process at stationarity, so $\bar{\rho}_t \equiv \bar{\rho}$, and let the temporal profile of the perturbation be $\varpi(t)$ with relaxation rate ν. Then $\mathcal{B}_K$ separates into a spatial factor, which is the stationary Stein operator, and a temporal factor determined by ϖ.

This regime is a warning, not a target. With $\bar{\rho}_t$ constant, all snapshot shapes coincide, $r(x) \equiv 0$ in Prop. 6, and Thm. 8 gives no contraction at all: the entire degree-q gauge S_q survives. The separable calculation therefore describes how much *temporal* information a shared field supplies when the spatial shapes supply none, and it should not be read as a rate for the transient problem.

L.1 Cost of estimating the temporal spectrum

Take one spatial mode of amplitude θ that is an eigenfunction of the stationary tangent generator with eigenvalue $-\nu$. From $u_0 = 0$, the Duhamel formula (3) gives it the temporal profile $\varpi(t;\nu) = (1 - e^{-\nu t})/\nu$, so slice k carries $\theta\,\varpi(t_k;\nu)$. Suppose ν is itself unknown. With $\varpi := (\varpi(t_k;\nu))_{k\le K}$ and $\varpi_\nu := (\partial_\nu \varpi(t_k;\nu))_{k\le K}$ in $\mathbb{R}^K$, the joint Fisher information for (θ,ν) is proportional to $\left(\begin{smallmatrix} \|\varpi\|^2 & \theta\langle \varpi,\varpi_\nu\rangle \\ \theta\langle \varpi,\varpi_\nu\rangle & \theta^2\|\varpi_\nu\|^2 \end{smallmatrix}\right)$, and profiling out ν multiplies the variance of $\widehat{\theta}$ by

$$\frac{1}{1-\rho_{\mathrm{temp}}^2}, \qquad \rho_{\mathrm{temp}} := \frac{\langle \varpi, \varpi_\nu\rangle}{\|\varpi\|\,\|\varpi_\nu\|}, \tag{8}$$

where ρ_{temp} is the cosine between the profile and its derivative in the *parameter* ν, not in time, and ρ_{temp}^2 is the squared cosine. With $K \ge 2$ distinct times ϖ_ν/ϖ is not constant in k, so $\rho_{\mathrm{temp}}^2 < 1$ and the inflation is finite for every νt_1. It is nevertheless large once the mode has relaxed before the first observation: the slices then see it only near its plateau θ/ν, in which amplitude and rate are confounded, and separating them requires the transient, of size $e^{-\nu t_1}$. Numerically the inflation tracks $e^{2\nu t_1}/(\nu t_1)^2$ up to a constant (Fig. 3).

This is a nuisance-conditioning statement about one scalar, and we state it as such. Earlier drafts of this work claimed a phase transition in which co-estimating the dynamics collapses the polynomial rate to a logarithmic one. That claim was tied to a spectral model this paper no longer uses, and we withdraw it: nothing here establishes a dichotomy, and (8) is a finite factor for each mode, not a change of rate.

M Strong-form diagnostics and failed repairs

M.1 First-order nuisance sensitivity

Proof of Prop. 14. Write the strong-form residual at slice k as $r_\theta = \mathcal{A}^\eta_{\rho_k} F_\theta - b^\eta_k$, where η collects the score and temporal-source nuisances, and the moment as $m_p(\theta,\eta) = \mathbb{E}[r_\theta \chi_p]$. Then $\partial_\eta m_p = \mathbb{E}[(\partial_\eta \mathcal{A}^\eta F_\theta - \partial_\eta b^\eta)\chi_p]$, which contains $\mathbb{E}[F_\theta \cdot \delta s\, \chi_p]$ and does not vanish at the truth unless F_θ is orthogonal to the score perturbation in every probe direction. Solving the moment equations to first order, the induced parameter bias is $-\mathcal{H}^{-1}\partial_\eta m\, \delta\eta$ with $\mathcal{H}$ the profiled Hessian of Prop. 15. Reading out a solenoidal coordinate with curvature λ multiplies the bias by λ^{-1}. Since $\lambda \to 0$ precisely

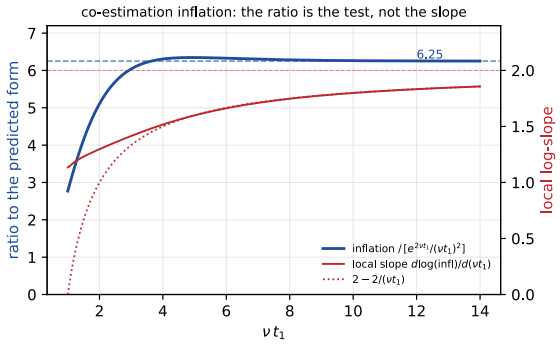


Figure 3: **Cost of profiling the temporal nuisance.** Ratio of the exact inflation (8) to the surrogate $e^{2\nu t_1}/(\nu t_1)^2$ against νt_1, for $K = 5$ observation times with $t_k/t_1 \in \{1, 1.6, 2.4, 3.4, 5\}$: the ratio settles at $\approx$6.25 for $\nu t_1 \gtrsim 4$ while the inflation grows by more than seven orders of magnitude, so the surrogate captures the growth and misses only a constant; the local log-slope approaches 2 only as $2 - 2/(\nu t_1)$. The inflation is finite at every νt_1 and large once the mode has relaxed before the first observation.

along near-gauge directions, nuisance error is amplified exactly where the signal is weakest. Exact orthogonalization of this moment is unavailable whenever F has a nonzero solenoidal component, because $\partial_\eta m$ then has a component along the solenoidal block that no reweighting of χ_p removes. □

Proposition 35 (Filtering cannot separate signal from nuisance bias)**.** *Let $\mathcal{H} = \sum_i \lambda_i v_i v_i^\top$ be the profiled Hessian in the solenoidal basis and let a spectral filter act diagonally, $\widehat{\theta}_\phi = \sum_i \phi(\lambda_i)\lambda_i^{-1} v_i^\top z$, where z is the right-hand side of the profiled normal equations. Then both the signal and the nuisance bias along v_i are multiplied by the same factor $\phi(\lambda_i)$, so the ratio of bias to signal in every coordinate is invariant under ϕ. No diagonal filter improves it.*

Proof. Both the signal contribution and the first-order nuisance bias enter $v_i^\top z$ linearly and are acted on by the same scalar $\phi(\lambda_i)\lambda_i^{-1}$; their ratio is therefore unchanged. Filtering trades variance against bias uniformly, which is useful, but it cannot separate two contributions that occupy the same eigendirection. □

Fig. 4 separates the two failures: the optimization one is a solver choice, the nuisance one is structural.

M.2 The coefficient-side correction is not an orthogonalization

A natural repair solves $\mathcal{H}^\top r_c = \ell$ for a Riesz representer and reports $\ell^\top a - r_c^\top m(a, \eta)$. In the local model $m(a, \eta) = \mathcal{H}(a - a^\star) + b(\eta)$ with $b(\eta^\star) = 0$, the corrected functional is exactly $\ell^\top a^\star - r_c^\top b(\eta)$, whose nuisance derivative is $-r_c^\top D_\eta b$ — generally nonzero. The correction removes first-order sensitivity to the *initial coefficient estimate*, not to the nuisance, and at a same-sample exact moment root it is identically zero. Cross-fitting does not change this derivative. We therefore describe the implemented method as a one-step estimating-equation correction and report its behaviour empirically; we do not claim Neyman orthogonality, and Prop. 14 is not a proof that every orthogonal construction must fail.

Proposition 36 (The targeting window)**.** *The correction's own sampling error is amplified by the same λ^{-1} that amplifies the bias it removes. The corrected estimator improves on the plug-in only when the nuisance error lies below a threshold set by that amplification.*

Measured on the planted-curl benchmark over eight paired seeds (Fig. 5), the threshold sits below the practical denoising-score-matching error floor: the targeting study's plug-in at $\varepsilon = 0.10$ score error gives 2.547 ± 1.201 (the Gauss–Newton arm of Tab. 1, on a different eight-seed set, gives 2.461 ± 0.689) and the targeted variant does not recover the score-free weak form's performance. The

two failure modes — amplitude bias from the temporal log-density value channel, and gradient error from the score — separate cleanly: a probe-wise decomposition attributes 38% to the former against 7% to the latter. Denoising score matching supervises $\nabla \log \rho_t$ but not the additive normalization of $\log \rho_t$, which is exactly the quantity (6) eliminates.

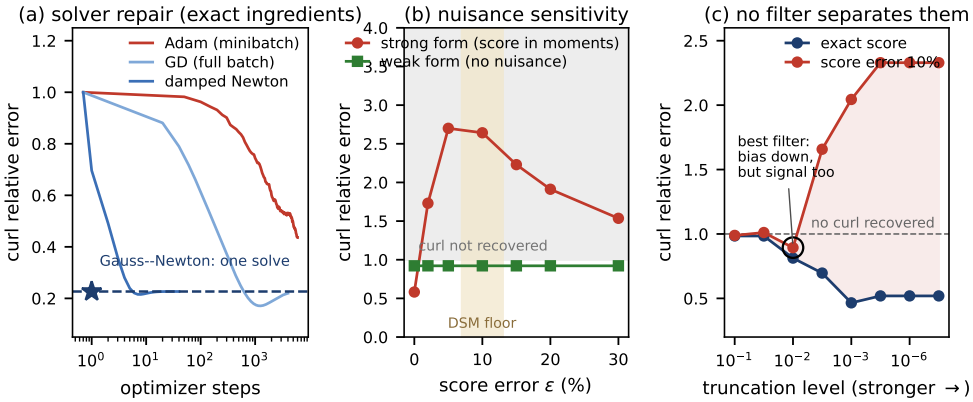


Figure 4: **Repairing the strong form, and the limit of repair.** (a) The optimization failure is a solver choice: Gauss–Newton returns the exact minimizer in one linear solve, damped Newton in $\sim$5 steps, first-order methods in 10^3–10^4. (b) The nuisance failure is structural: as score error grows the strong form degrades through the "no curl recovered" line while the weak form stays flat, its moments containing no score; the band is the 7–13% error floor of denoising score matching. (c) Filtering cannot rescue it (Prop. 35) — the truncation that best suppresses the nuisance bias suppresses the curl with it.

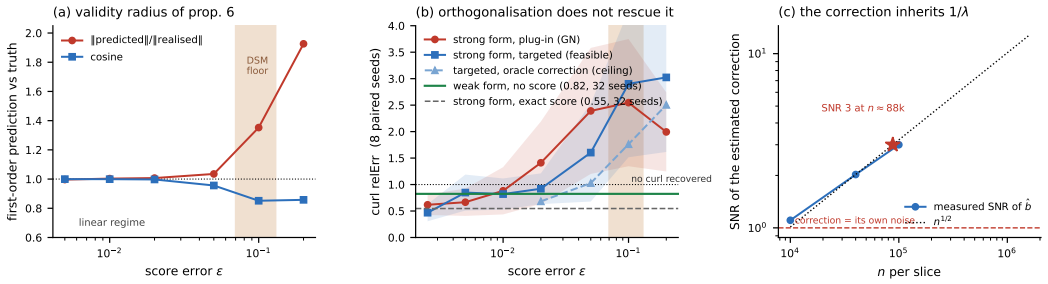


Figure 5: **The targeting window.** Correcting the strong-form moment helps only for score error $\varepsilon \lesssim 0.05$, and the window shuts below the practical denoising-score-matching floor, so the one-step route cannot rescue the strong form at feasible nuisance quality (Prop. 36). The correction's own sampling error is amplified by the same λ^{-1} that amplifies the bias it removes.

M.3 Reference arms

With exact score and temporal source, a closed-form linear least squares reaches solenoidal relative error 0.351; a Gauss–Newton solve on the same objective gives 0.535 ± 0.177 over eight seeds and 0.547 ± 0.182 over thirty-two; Adam on the same objective gives 0.510 ± 0.123. Fitting the score alone raises the closed-form arm to 0.862, and fitting both score and source to 2.19. These arms are handed oracle nuisances and are reported as references, not as competitors to the weak form.

N WEAK-FORM IMPLEMENTATION

N.1 ASSEMBLY

Probes are tensor frames ψ_e, either monomial $\prod_a x_a^{e_a}$ or normalized probabilists' Hermite $\prod_a \Psi_{e_a}(x_a/\hat{\varsigma}_a)$, with Ψ_k the univariate case of the basis of §3.1 and $\hat{\varsigma}$ the pooled per-coordinate standard deviation. Both require gradients and Laplacians; for the Hermite frame $\Psi'_k = \sqrt{k}\,\Psi_{k-1}$ gives $\partial_a\psi_e = \hat{\varsigma}_a^{-1}\sqrt{e_a}\,\psi_{e-e_a}$ and $\Delta\psi_e = \sum_a \hat{\varsigma}_a^{-2}\sqrt{e_a(e_a-1)}\,\psi_{e-2e_a}$, and for the monomial frame $\Delta x^e = \sum_a e_a(e_a-1)x^{e-2e_a}$. The Laplacian is what carries the $\sigma^2\Delta\psi$ term of (6); omitting it silently changes the estimand, as §5 quantifies. For degree-two monomial probes $\Delta\psi$ is constant, so the correction is a deterministic offset to the response and leaves the noise model untouched.

The design entry for field basis element $\phi_p = x^{e_p}e_{i_p}$ and probe ψ_j is $\frac{\Delta_k}{2}(\mathbb{E}_k + \mathbb{E}_{k+1})[\partial_{i_p}\psi_j \cdot x^{e_p}]$, an empirical moment of a single monomial; the response is $\mathbb{E}_{k+1}[\psi_j] - \mathbb{E}_k[\psi_j] - \frac{\Delta_k}{2}(\mathbb{E}_k + \mathbb{E}_{k+1})[\sigma^2\Delta\psi_j]$. Both sides are computable from raw samples with no score, density, or pointwise source.

N.2 FEASIBLE GLS

Prop. 16 gives the covariance at a fixed coefficient vector, and the covariance depends on a through f_a. We use a same-sample unweighted least-squares pilot $\widehat{a}_0$, build $\widehat{V}(\widehat{a}_0)$, and take one frozen GLS step; the one-step feasible variant that rebuilds $\widehat{V}(\widehat{a}_1)$ is reported separately where it differs. The covariance is regularized by an eigenvalue floor at 10^{-10} times the largest eigenvalue before the Cholesky factorization, and singular blocks are handled by the same truncation as the design. The distinction between the exact population covariance algebra and the estimated whitening matters: the finite-sample coverage of a same-sample pilot is not the coverage the fixed-weight proposition describes, and we do not claim it.

N.3 TRUNCATION, GAUGE EXCLUSION, AND THE QUADRATURE FLOOR

Fig. 6 collects the recovery, calibration and conditioning diagnostics. The whitened design is column-equilibrated before a truncated SVD solve, so the truncation metric is comparable across monomial columns spanning orders of magnitude. The cutoff is selected by split-sample scoring on independently assembled halves, with a tie-break preferring the least truncation among candidates within 15% of the best score: the half-sample fit has a $\sqrt{2}$-higher noise floor, so its optimum is biased toward over-truncation.

Gauge columns are excluded structurally rather than penalized. For the linear benchmark these are the zero-curvature directions of Prop. 15; for the Hermite construction they are the blockwise kernels of §3.1. A downstream monitor verifies that the fitted field has exactly zero component along the declared gauge.

Because the snapshot times are fixed, increasing n does not remove the trapezoidal bias. The residual of (6) is $O(\Delta_k^3)$ per interval, but the design entries carry the factor $\Delta_k/2$, so the induced bias in the coefficients, and hence in the drift, is $O(\Delta^2)$. On a two-dimensional OU control with exact moments, the population drift error falls as Δ^2 — 3.0×10^{-2}, 7.5×10^{-3}, 1.9×10^{-3}, 4.6×10^{-4} at gaps 0.2, 0.1, 0.05, 0.025 — confirming the order and showing that the floor is a property of the design, not of the sample size. On the main benchmark the corresponding floor is 0.033 in solenoidal relative error.

O A SNAPSHOT-ONLY FINITE-BASIS ERROR BOUND

Thm. 12 is a statement about the tangent experiment and its degreewise benchmark, under unproved assumptions. This appendix gives a complementary guarantee of a different kind: finite sample, snapshot only, no Gaussian-reference or spectral hypothesis, for the estimator actually implemented. It is much weaker than the oracle rate — it says nothing about a nonparametric exponent — but it is unconditional in the assumptions the rate needs, and it makes the quadrature obstruction explicit.

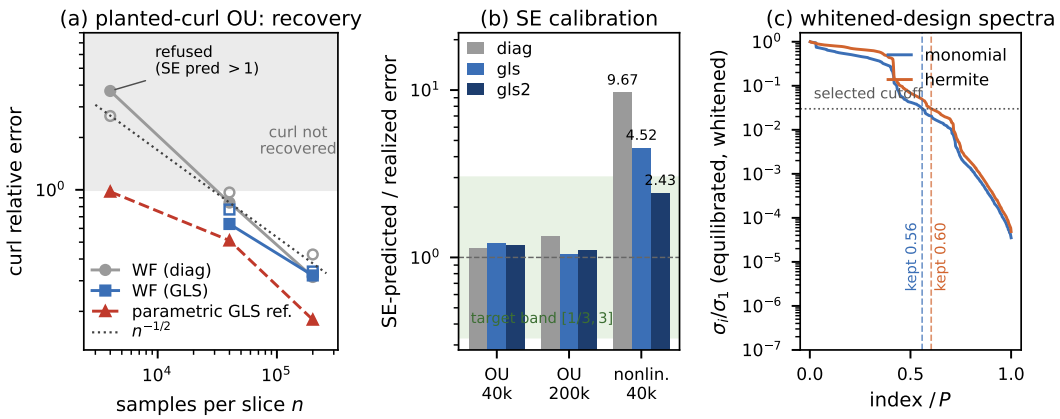


Figure 6: **Recovery, calibration, conditioning.** (a) Curl error against n, filled for realized and open for standard-error predicted; the power check of Prop. 15 refuses the n=4k design. (c) Equilibrated spectra of the whitened nonlinear design: Hermite frames retain more of the spectrum and decay more slowly, which is where their advantage lies. **Panel (b) is retained as a record of a withdrawn claim** and should not be read as a current result: its $9.7\times$ calibration gap was measured on the lineage-tracked design of App. R, where the slices are not independent, and with independent slices the diagonal predictor is calibrated at $1.00\times$ with nothing to repair (App. Y). Panels (a) and (c) were computed under the velocity convention and are qualitatively unchanged under the correction, but their absolute values are those of App. Y's first column.

Fix the probe set, the field basis, and a retained coefficient subspace with orthonormal coordinates, and restrict $\widehat{X}$ to that subspace. Write $X = \mathbb{E}\widehat{X}$, $b = \mathbb{E}\widehat{b}$, and $r = b - Xa^\star$ for the population residual of a reference coefficient vector $a^\star$.

Theorem 37 (Snapshot-only finite-basis error bound)**.** *Let $a^\star$ and the whitener W be deterministic, and let $\alpha := s_{\min}(WX) > 0$ on the retained subspace. Suppose the required second moments are finite and $\Pr(\|W(\widehat{X} - X)\|_{\mathrm{op}} > \alpha/2) \le \eta$. Then for the unregularized solution $\widehat{a}_0$ of $\min_a \|W(\widehat{b} - \widehat{X}a)\|^2$ and any linear readout C, with probability at least $1 - \eta - \delta$ and $0 < \delta < 1$,*

$$\|C(\widehat{a}_0 - a^\star)\| \le \frac{2\|C\|_{\mathrm{op}}}{\alpha}\left(\|Wr\| + \sqrt{\frac{\mathrm{tr}\big(WV(a^\star)W^\top\big)}{\delta}}\right). \tag{9}$$

One may take $\eta \le 4\,\mathbb{E}\|W(\widehat{X} - X)\|_F^2/\alpha^2$. The statement holds conditionally on an independent pilot fixing W and the retained subspace, provided the conditional hypotheses hold.

Proof. Write $E := W(\widehat{X} - X)$ and $\varepsilon := W(\widehat{b} - b)$, and let $M := WX$. On the event $\mathcal{E} := \{\|E\|_{\mathrm{op}} \le \alpha/2\}$, which has probability at least $1 - \eta$, Weyl's inequality gives $s_{\min}(M + E) \ge \alpha - \|E\|_{\mathrm{op}} \ge \alpha/2 > 0$, so $M + E$ has full column rank on the retained subspace, the least-squares solution is unique, $\widehat{a}_0 = (M+E)^+ W\widehat{b}$, and $(M+E)^+(M+E) = I$ there with $\|(M+E)^+\|_{\mathrm{op}} \le 2/\alpha$. The product $(M + E)\widehat{a}_0$ is only the projection of $W\widehat{b}$ onto the range of $M + E$, so we do not equate it with $W\widehat{b}$; instead, since $M + E = W\widehat{X}$,

$$\widehat{a}_0 - a^\star = (M+E)^+\big(W\widehat{b} - (M+E)a^\star\big), \qquad W\widehat{b} - (M+E)a^\star = W(\widehat{b} - \widehat{X}a^\star) = Wr + Z,$$

with $Z := W(\widehat{b} - \widehat{X}a^\star) - Wr = \varepsilon - Ea^\star$, using $Wb = Wr + Ma^\star$. Hence, on $\mathcal{E}$, $\|C(\widehat{a}_0 - a^\star)\| \le \frac{2\|C\|_{\mathrm{op}}}{\alpha}(\|Wr\| + \|Z\|)$. Since $\mathbb{E}[\widehat{b} - \widehat{X}a^\star] = b - Xa^\star = r$, Z is the centred whitened empirical moment at $a^\star$, with mean zero and covariance $WV(a^\star)W^\top$, V the block-tridiagonal covariance of Prop. 16. Chebyshev's inequality in the form $\Pr(\|Z\|^2 > \mathrm{tr}\,\mathrm{Cov}(Z)/\delta) \le \delta$ for mean-zero Z gives (9) on the intersection, of probability at least $1 - \eta - \delta$. For the bound on η, Markov's inequality applied to $\|E\|_{\mathrm{op}}^2 \le \|E\|_F^2$ gives $\Pr(\|E\|_{\mathrm{op}} > \alpha/2) \le 4\mathbb{E}\|E\|_F^2/\alpha^2$. □

Three consequences deserve statement, because they are what the theorem is for.

Sampling error and quadrature bias are different terms. The second bracket in (9) is $O(n_{\min}^{-1/2})$ through $V(a^\star)$; the first, $\|Wr\|$, is not stochastic at all. At fixed snapshot times r is the whitened trapezoidal residual, of order $\Delta_{\max}^3$ in moment scale by (6), and no amount of data removes it. The design entries carry the factor $\Delta_k/2$ while the whitener does not shrink with the gaps (the moment noise of Prop. 16 contains the gap-independent term $\mp\Psi$), so α scales with the gaps and, for comparable gaps, the amplified bias $\|Wr\|/\alpha$ is $O(\Delta_{\max}^2)$ in coefficient scale — the formal version of the Δ^2 floor measured in App. N.

Both terms are amplified by $1/\alpha$. The same near-gauge ill-conditioning that governs the oracle spectrum governs the finite-basis bound, which is why gauge exclusion — removing directions with $\alpha = 0$ rather than penalizing them — is not a numerical convenience.

The population target need not be the truth. With W fixed the population least-squares target is $(WX)^+Wb$, which equals $a^\star$ only when $Wr \perp \mathrm{range}(WX)$. Consistency for the drift along refining grids therefore requires the amplified bias $\|Wr\|/\alpha \to 0$ as well as design stability; increasing n alone does not deliver it. A readout in a moment-null direction is not identified at all unless C annihilates that null space, and excluding such a direction is a modelling convention rather than recovery of its coefficient.

Two scope limits. The theorem needs $\alpha > 0$ on the retained subspace, and empirically small singular values indicate weak identification, not a proved structural gauge: the exclusion in $\mathcal{A}_{\mathrm{vis}}$ is certified by Prop. 15 only in the centered-Gaussian linear case, so for the frame constructions the retained subspace is a modelling choice. And the field metric must be specified — orthonormality of coordinates is imposed through a field Gram matrix under a chosen evaluation measure, not assumed for arbitrary features.

P Measurement noise

Suppose $Y = X + \varepsilon$ with ε independent of X, mean zero and covariance Σ_ε. Independence, or at least zero cross-covariance, is required: mean-zero noise alone does not give $\mathbb{E}[YY^\top] = \mathbb{E}[XX^\top] + \Sigma_\varepsilon$.

It is tempting to argue that the antisymmetric readout is immune, because shifting a covariance by cI leaves $G\Sigma - \Sigma G$ unchanged. That argument is incomplete. For $H = S + G$ the same shift moves the drift moment by $2cS$, so an uncorrected symmetric block can bias a *jointly* profiled estimate of G. The effect is real at the population level: with

$$H = \begin{pmatrix} -1 & -0.6 \\ 0.6 & -2 \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} 2 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}, \quad D = 0.4I,$$

regressing exact diffusion-corrected covariance derivatives on $H\Sigma + \Sigma H^\top$ at $t = 0.1, 0.3, 0.7, 1$ recovers skew coefficient 0.600000 with clean covariances and 0.441094 after adding $0.2I$ to the design covariances without deconvolution — a 26% bias with no sampling noise at all. We therefore do not claim universal immunity (Fig. 7). Moment deconvolution of the design, using a known or separately estimated Σ_ε, restores the clean value; without it, the symmetric nuisance reaches the solenoidal readout.

A transparent second-moment variant of the estimator isolates the design perturbation (clean-data curl 0.849, against 0.844 and 0.636 for the full pipeline at diagonal and GLS weighting). Under isotropic contamination the curl readout degrades from 0.8494 clean to 1.145 naive, and deconvolution does *not* repair it (1.1501); under a curl-aligned contamination the same numbers are 0.8494, 0.9345 and 1.0274. What deconvolution does repair is the symmetric channel, where the distance to the truth falls from 0.2823 to 0.1134. This is the honest version of the earlier claim: the fix is exact at second moments on the block where the bias lands, and the solenoidal channel is not protected.

Q Target-aware stabilization

Proposition 38 (Selection misalignment)**.** *In the whitened linear experiment with ridge penalty, and with risks averaged over the signs of the coordinates of the true coefficient vector in the eigenbasis of*

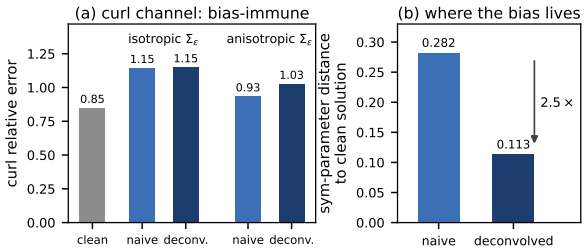


Figure 7: **Measurement noise.** The symmetric channel absorbs the constant design shift and its residual bias is removed by design deconvolution; the solenoidal channel is *not* exactly immune, contrary to an earlier version of this claim (App. Y). Isotropic contamination is the adverse case: deconvolution of the design covariances repairs the symmetric block without restoring the curl readout.

the profiled Hessian, the regularization strengths minimizing moment-prediction risk and the risk of a linear functional of the field differ in general; they coincide when the functional's squared weights are proportional to the curvatures or the coordinate signal is constant across the spectrum. For the solenoidal readout the gap is governed by the near-gauge spectrum.

Proof of Prop. 38. Write the whitened experiment as $y = Xa^\star + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, I)$, let $\mathcal{H} := X^\top X = \sum_i \lambda_i v_i v_i^\top$, and write the ridge path as $\widehat{a}_\zeta = (\mathcal{H}+\zeta I)^{-1} X^\top y = (\mathcal{H}+\zeta I)^{-1}\mathcal{H}a^\star + (\mathcal{H}+\zeta I)^{-1}\xi$ with $\xi := X^\top \varepsilon \sim \mathcal{N}(0, \mathcal{H})$. In the coordinates $a_i^\star := v_i^\top a^\star$ and $\ell_i := v_i^\top \ell$, coordinate i has bias $-\zeta a_i^\star/(\lambda_i+\zeta)$ and variance $\lambda_i/(\lambda_i+\zeta)^2$, independently across i. The moment-prediction risk is

$$\mathbb{E}\|X(\widehat{a}_\zeta - a^\star)\|^2 = \sum_i \frac{\zeta^2 \lambda_i (a_i^\star)^2 + \lambda_i^2}{(\lambda_i+\zeta)^2},$$

which has no cross terms. The risk of $\ell^\top a$ is $\left(\sum_i \ell_i \zeta a_i^\star/(\lambda_i+\zeta)\right)^2 + \sum_i \ell_i^2 \lambda_i/(\lambda_i+\zeta)^2$; averaging over independent signs of the $a_i^\star$, which leaves the moment risk unchanged, removes the cross terms and gives $\sum_i \ell_i^2 (\zeta^2 (a_i^\star)^2 + \lambda_i)/(\lambda_i+\zeta)^2$. Since $\frac{d}{d\zeta}\frac{\zeta^2 c^2 + \lambda}{(\lambda+\zeta)^2} = \frac{2\lambda(\zeta c^2 - 1)}{(\lambda+\zeta)^3}$, the two stationarity conditions are

$$\sum_i w_i \frac{\lambda_i\big(\zeta (a_i^\star)^2 - 1\big)}{(\lambda_i+\zeta)^3} = 0, \qquad w_i = \lambda_i \text{ (moment)}, \quad w_i = \ell_i^2 \text{ (functional)}.$$

If $(a_i^\star)^2 \equiv c^2$ both are solved by $\zeta = 1/c^2$; otherwise the two conditions weight the same terms differently, and their roots differ unless $\ell_i^2 \propto \lambda_i$. For the solenoidal readout ℓ is supported on the near-gauge coordinates where λ_i is smallest, which the moment criterion downweights by λ_i, so the two criteria disagree most exactly where the estimand lives. □

Because whitening makes the response law exactly $\mathcal{N}(0, I)$, target risk can be estimated without ground truth: perturb the whitened response under its known law, re-run the estimator along its regularization path, and average the resulting functional deviations. The rule needs only that the estimand be a linear functional of the field. Its failure mode is a plain estimate that is already near-unbiased, in which case the perturbation estimate is dominated by variance and the selected ζ collapses to zero; on the main benchmark split-sample selection picks $\zeta = 0$ on every seed, and a fixed ridge at $\zeta = 0.03$ outperforms the selected value. We report the rule as correctly motivated and unreliable in this regime.

R Single snapshots and Lagrangian information

R.1 One snapshot

With $K = 1$ the gauge is the full $\ker \mathcal{A}_{\rho_{t_1}}$ and Prop. 6 gives $r(x) \equiv 0$: nothing about the circulation is identified. A single snapshot constrains only the gradient part, and any reported circulation is a property of the prior, not of the data.

R.2 LINEAGE-TRACKED OBSERVATION

Autonomy is one way to couple slices; particle identity is another, and it is a different observation model with different statistics. If the same particles are observed at every time, the slices are deterministic images of one another and the independence hypothesis of Prop. 16 fails badly.

Our nonlinear benchmark is of exactly this type, and we report it here rather than in §5 for that reason. The system is a deterministic ODE ($\sigma = 0$) in $d = 6$ with a planted degree-three solenoidal field, $F(x) = -0.7x + (3cx_1x_2^2, -cx_2^3, 0, \dots)$ with $c = 0.2$, integrated by RK4 from a Gaussian initial ensemble to five closely spaced times. Because $\sigma = 0$ the continuity identity is exact and the diffusion correction of §4.1 does not apply.

Measuring the true moment-residual covariance over 250 independent replications at $n = 1500$ against the surrogates the estimator uses:

	$\operatorname{tr} V$	ratio to truth
true Monte-Carlo covariance	2.11×10^{-3}	—
block-tridiagonal surrogate	1.08×10^{-1}	$51.3\times$
diagonal surrogate	1.08×10^{-1}	$51.3\times$

Moreover 54.5% of the true covariance mass lies in non-adjacent blocks, which both surrogates set to zero by construction. That the two surrogates agree to four digits is itself diagnostic: with coupled particles the off-diagonal $-C_{k+1}$ correction is negligible against a covariance dominated by cross-slice terms neither model contains.

The consequences for recovery are large. Under the coupled design the estimator attains curl-channel error 0.4027 with diagonal weights and 0.2095 with GLS, with a null-circulation control at 0.038 and standard-error ratios of $9.67\times$ and $4.52\times$. Re-drawing each slice independently — the observation model of §2 — gives 1.6455 and 1.0142 respectively, against a no-information level of 1.0, with a null control at 0.1835 and standard-error ratios of $1.00\times$ and $1.09\times$.

Three readings follow. The recovery was a consequence of particle identity, not of the snapshot moments. The apparent $9.67\times$ conservatism of the diagonal predictor was not a pathology to be fixed: with independent slices that predictor is calibrated, and its own power check, which prescribed 6.7M samples per slice, was correct. And the lineage-tracked design is a legitimate and arguably more realistic observation model — it is what pairing or lineage barcoding supplies — but it belongs to this appendix, because Lagrangian information is precisely what §2 sets out to do without. An estimator for it would need a covariance model containing the cross-slice blocks measured above; we do not develop one.

S NUMERICAL VERIFICATION OF THE ANALYTIC IDENTITIES

Each identity below is checked by an executable entry point listed in App. Z; these are machine-precision checks of the stated algebra, not validation of the rate theorem.

Reference-slice algebra. $\dim S_q = mD_{q-1} - D_q$ from Lem. 24 is checked against explicit kernels of $\mathcal{A}_{\rho_\star}|_{W_q}$ for $m = 2, \dots, 6$ and $q \le 8$; surjectivity is checked by rank. The linear characterization of Lem. 26 is checked by sampling random Σ and verifying $\|\mathcal{A}_\rho(\Sigma N x)\| = 0$ to machine precision for antisymmetric N, and $\neq 0$ otherwise.

Gauge contraction. The $q = 2$ commutant is computed by explicit nullspace sweep over K for $m = 3, \dots, 6$: it has full dimension $\binom{m}{2}$ at $K = 1$ and collapses to $\{0\}$ already at $K = 2$, confirming that the degree-two block is cleared well before the uniform threshold. The shape-richness condition of Thm. 8 is checked separately by sampling x and computing $\operatorname{rank}[x \,|\, \Xi_1 x \,|\, \cdots \,|\, \Xi_r x]$: it is deficient at $r = m - 2$ and full at $r = m - 1$, i.e. exactly at $K = m$. The two gates are distinct and the paper should not be read as claiming otherwise.

Propagated spectrum. The first-moment law $\bar{\mu}_q \asymp K/q$ is checked by assembling $\mathcal{B}_K$ on the reference path and computing $\operatorname{tr} \mathcal{J}_{\mathrm{sol},q}/r_q$ across q; the fitted inverse-trace powers are 3.20, 4.08 and 4.76 at $m = 3, 4, 5$ against the predicted m, and the fitted lower-tail exponent lies in $[1.6, 2.2]$.

Estimator assembly. The Laplacian tables of App. N are checked against central finite differences for both frames, with maximum deviation 8×10^{-8}. The block-tridiagonal covariance of Prop. 16 is checked against a 1500-replication Monte-Carlo covariance on an independent-slice design, agreeing to 4.6% in relative Frobenius norm with exactly zero non-adjacent blocks; the same check on the lineage-tracked design fails by the factor reported in App. R, which is how that defect was found.

T How the curl is parameterized

The solenoidal component can be carried by several parameterizations of the same field class; the antisymmetric potential of Prop. 25 is the ρ-weighted analogue of the divergence-free parameterization of Richter-Powell et al. (2022). They differ enormously in conditioning, and — because first-order optimization is sensitive to conditioning while an exact solve is not — in how close a descent method gets to the exact minimizer. Five variants on the linear benchmark, five seeds each (Tab. 3), against Gauss–Newton on the same objective:

Table 3: **Curl parameterizations.** Solenoidal relative error after 8000 descent steps; "reached" counts seeds whose loss came within 2.7% of the Gauss–Newton value. Conditioning is of the whitened design in that parameterization.

parameterization	params	rank	cond.	error	reached
antisymmetric potential	5226	144	3.1×10^4	0.542 ± 0.105	$4/5$
generator frame	144	144	2.7×10^4	0.521 ± 0.092	$5/5$
+ gauge reduction	116	116	1.0×10^3	0.521 ± 0.092	$5/5$
+ audit preconditioning	116	116	1.0	0.963 ± 0.385	$2/5$
+ tangent-only support	84	84	9.2×10^2	$\mathbf{0.520 \pm 0.081}$	$5/5$
Gauss–Newton (exact solve)	—	—	—	0.479 ± 0.194	—

Three readings. Over-parameterization is not free: the antisymmetric potential carries 5226 parameters for a design of rank 144, and is the only variant that fails to reach the exact minimizer on every seed. Gauge reduction buys a factor of 26 in conditioning at no cost in error, which is the quantitative case for excluding gauge columns rather than penalizing them. And perfect conditioning is not the objective: audit preconditioning drives the condition number to 1.000 and makes the result *worse*, because it whitens the parameterization rather than the residual, destroying the correspondence between descent geometry and the estimand. The exact solve is insensitive to all of this, which is the argument for using one.

Under a 10% score error the ordering changes and the margins compress: the antisymmetric potential gives 2.470, the generator frame 2.468, Gauss–Newton 2.397, and tangent-only support 1.019. Restricting the support to the audit-identified tangent block is the only variant that retains a meaningful advantage once the nuisance is corrupted, which is consistent with Prop. 22: the normal-block coordinates it removes were never identified, so they contribute variance and nuisance sensitivity without signal.

U Conditioning of the linear design

The Schur complement of Prop. 15 is computable before any field is fitted, and its spectrum is the design's own account of which solenoidal directions are visible. On the $d = 12$, $m = 4$ benchmark the six tangent antisymmetric directions carry eigenvalues $\{5.1, 7.8, 14.0, 23.7, 40.8, 71.4\} \times 10^{-3}$, a spread of $14\times$ between the best- and worst-determined direction within a single design. Recovery error is not uniform across the curl; it is concentrated in the small end of this spectrum, which is what makes a scalar readout of "curl error" a summary rather than a diagnosis.

Adding snapshots helps, but with sharply diminishing returns, and the gain is in the weakest direction rather than the strongest:

	$K = 3$	$K = 5$	$K = 9$
smallest Schur eigenvalue	1.79×10^{-3}	3.77×10^{-3}	3.85×10^{-3}
largest Schur eigenvalue	2.38×10^{-2}	5.47×10^{-2}	5.65×10^{-2}
pinned-direction eigenvalue	7.56×10^{-2}	7.05×10^{-2}	6.82×10^{-2}

Going from three to five snapshots roughly doubles the weakest eigenvalue; going from five to nine adds 2%. This is Thm. 8 seen numerically: once $K \geq m$ the polynomial gauge is gone and further snapshots buy only the shape diversity they happen to add, which for a fixed reference trajectory saturates quickly. The pinned direction, fixed by a constraint rather than by the data, is flat as expected and serves as a control.

A composite check, run under the velocity convention of App. Y, closes the loop between the audit and the fit. Predicting the solenoidal error from the Schur spectrum and the quadrature residual alone gives 0.6326 against a realized 0.6323 at $n = 40$k, a consistency of 0.001; the cosine between the estimated and planted curl rises from 0.825 at $n = 40$k to 0.950 at $n = 200$k; and the gauge monitor returns exactly 0, confirming the excluded directions carry no estimate. A gradient-head control on the same data, handed the true generator, returns 0.2196 — the residual is genuinely in the solenoidal channel and not an artifact of the gradient fit.

V Additional robustness and negative results

Frames. Monomial and Hermite frames recover comparably at $n = 40$k (full-field 0.062 against 0.067, curl channel 0.210 against 0.226); the Hermite frame mainly improves conditioning, with reciprocal condition number 0.03 against 0.01 at $n = 100$k.

Degree ladder. At $m = 6$ with a planted degree-three field, fitting degrees one through five gives 0.896, 0.935, 0.535, 2.513 and an infeasible solve. Under-specification is graceful; over-specification is not, and the failure is conditioning rather than variance.

Representation scaling. P grows as 504, $5{,}460$, $81{,}900$, $1{,}171{,}300$ at $d = 6, 12, 25, 50$ for the ambient degree-three construction. The binding cost is the $O(P^3)$ whitening Cholesky, not enumeration. A neural field at $d = 25$ reaches 0.989, essentially the no-information level.

Hard latent regime. With $d_{\text{latent}} = 8$, $n = 3000$ and four snapshots, latent-space cosine similarity is 0.83 and 0.86 on two seeds, against 0.51 and 0.48 for the earlier strong-form pipeline on the same data.

Optimization. Holding whitened moments and representation fixed, first-order descent on the same objective continues to reduce moment loss after the solenoidal error has turned over; over a long trace the final-minus-best gap is $+0.040$ with standard error 0.015 ($t = +2.68$, six seeds). We report this as near-null overfitting only. An earlier draft attributed it to the fitted field leaking into the exact design gauge; that attribution was wrong. For Euclidean descent on $\|Aa - z\|^2$ every update lies in $\operatorname{range}(A^\top)$, so $P_{\ker A} a_{r+1} = P_{\ker A} a_r$ and an exact design-null component cannot move. The monitor that appeared to grow was measuring a population gauge against a fitted design whose gauge columns had been excluded in the other arm, so the two procedures did not share a feasible model. The corrected comparison — projected descent against the direct solve on identical columns, initialization, whitening and penalty — shows the exact solve avoiding the failure by never building those columns, and shows no motion into the design kernel.

W Hidden-state stress tests

Asm. 3 is what converts an untestable model into a testable one, so it deserves adversarial testing. Both diagnostics below are exact-moment population calculations at the benchmark parameters $\nu = \sigma^2 = 1$: there is no sampling and no quadrature error, so any residual is misspecification and nothing else. They are population contrasts, not calibrated hypothesis tests.

W.1 Misspecification that is detectable

Take the autonomous two-dimensional system $H = -I + 1.5J$ with $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, isotropic diffusion $\sigma^2 I$, and $\Sigma_0 = \mathrm{diag}(2, 0.7)$, and retain only the first coordinate. Fitting the exact variance derivative on 101 times in $[0, 2]$ to the scalar autonomous linear model $\dot{v} = 2av + 2\sigma^2$ gives $\widehat{a} = -1.1083$ with relative diffusion-corrected residual 0.1204, while fitting the full two-dimensional covariance derivatives gives residual 3.8×10^{-16}.

The projection is therefore *visibly* inadequate: the shape of the observed variance path is incompatible with any autonomous scalar linear drift, and the diagnostic says so without reference to the truth. This is the favourable case, and it is the one a practitioner hopes for. Note its scope: it rejects the scalar *linear* model, not autonomy against every nonlinear scalar drift.

W.2 Misspecification that is not

Now start the same system at its stationary covariance. Every snapshot is then exactly compatible with the zero-rotation drift $-x$: the endpoint-moment residual against that drift is 0 to machine precision, at every time. An analyst fitting autonomous models to these marginals would accept a detailed-balance explanation of a system that is in fact circulating at $\omega = 1.5$.

The marginals are not merely uninformative, they are actively misleading, and nothing in the snapshot experiment reveals it. The projected process is not Markov: its lag autocovariance oscillates, changing sign three times over $[0, 6]$, whereas a scalar Ornstein–Uhlenbeck autocovariance is a decaying exponential with no sign change at all. Paired full-state measurements would separate the two models immediately — their lag-0.3 cross-covariances differ by 0.4675 in Frobenius norm — but such measurements are exactly the Lagrangian information §2 sets out to do without.

This is the sharp form of the caution in §2.2, and it connects to two results already in the paper. It is the extreme case of Obs. 17: stationary shapes give $r(x) \equiv 0$ in Prop. 6, so Thm. 8 yields no contraction whatsoever and the entire gauge survives, which is also why App. L is a warning rather than a target. Enforcing autonomy on an inadequate representation does not here manufacture spurious circulation; it does something quieter and worse, returning the confident answer zero.

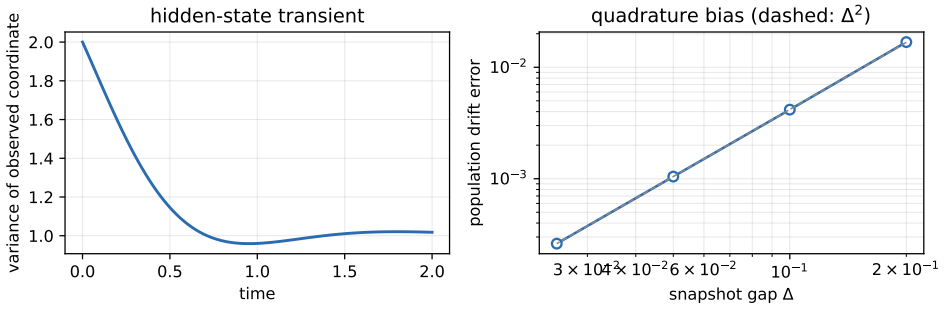


Figure 8: **Controlled diagnostics.** Left: transient variance of the projected coordinate in the hidden-state example, whose shape no autonomous scalar linear drift reproduces. Right: population error of the diffusion-corrected weak fit as the time grid is refined, against a Δ^2 reference; sampling alone cannot remove the quadrature bias at fixed gap (Thm. 37).

Fig. 8 shows the two mechanisms. A sampling-calibrated misspecification test and large-scale nonlinear stress studies remain future work; neither is represented by these population checks.

We therefore do not claim a calibrated test for representation adequacy. What the paper offers is narrower: shape movement is necessary for any recovery at all, is computable in advance from second moments by Prop. 15, and its absence is detectable even when the inadequacy of the representation is not.

X SCOPE: AUTONOMY, DIFFUSION, AND DESIGN SIZE

X.1 AUTONOMY IS A PROPERTY OF THE REPRESENTATION

Asm. 3 is not the biological claim that a regulatory program is fixed. Suppose the observed coordinates obey $dX_t = G(X_t, Z_t)\,dt + \sqrt{2\sigma^2}\,dW_t$, with Z_t latent regulatory state — chromatin, protein levels, signalling context — evolving in any way. If Z_t is part of the modelled state, the enlarged system is autonomous however much regulation changes along Z. If it is omitted, Itô's formula gives $\frac{d}{dt}\mathbb{E}[\psi(X_t)] = \mathbb{E}[F_t(X_t)\cdot\nabla\psi(X_t) + \sigma^2\Delta\psi(X_t)]$ with $F_t(x) := \mathbb{E}[G(x, Z_t) \mid X_t = x]$, so the marginals of X solve the Fokker–Planck equation with drift F_t. This drift is autonomous when the conditional law of Z_t given $X_t = x$ does not change with t — more precisely, when the conditional mean of $G(x, Z_t)$ does not — and genuinely time-dependent otherwise, the projected process then being non-Markov. Asm. 3 is therefore a state-sufficiency assumption on the chosen representation. Its failure is sometimes detectable (App. W, first case) and sometimes not (second case), so it can be falsified but not certified from snapshots. Restoring it by enlarging the state raises the intrinsic dimension m, the quantity that sets the threshold of Thm. 8 and the rate of Thm. 12: representation sufficiency and statistical efficiency pull in opposite directions. Any model that predicts responses from the current state, perturbation prediction included, presupposes the same sufficiency.

X.2 DIFFUSION: KNOWN, STATE-DEPENDENT, UNKNOWN

With $dZ_t = F(Z_t)\,dt + \sqrt{2}\,\Lambda(Z_t)\,dW_t$ and known $D := \Lambda\Lambda^\top$, the Fokker–Planck equation is $\partial_t\rho = -\nabla\cdot(\rho F) + \sum_{ij}\partial_i\partial_j(D_{ij}\rho)$, and dividing by ρ_t gives $\mathcal{A}_{\rho_t}F = b_t^D$ with $b_t^D := -\partial_t\log\rho_t + \rho_t^{-1}\sum_{ij}\partial_i\partial_j(D_{ij}\rho_t)$. The operator acting on the drift is unchanged; only the known source changes. The results of §2 that use the diffusion only through the source therefore hold verbatim: Prop. 4, Prop. 6, Cor. 2 applied to drifts, and Prop. 21, whose Poisson problem becomes $\nabla\cdot(\rho_t\nabla\Phi_t) = \partial_t\rho_t - \sum_{ij}\partial_i\partial_j(D_{ij}\rho_t)$, again with zero integral. The relation $F_t = v_t + \sigma^2 s_t$ of §2.1 becomes $F_t = v_t + \rho_t^{-1}\nabla\cdot(\rho_t D)$, which is not a gradient for anisotropic or state-dependent D, so drift-level saturation should then be read from Prop. 21 with $h_t = 0$. Thm. 8 concerns the source constraints of Gaussian slices and involves no diffusion. The weak moments extend as well: Itô's formula gives $\frac{d}{dt}\mathbb{E}_{\rho_t}[\psi] = \mathbb{E}_{\rho_t}[F\cdot\nabla\psi + \mathrm{tr}(D\nabla^2\psi)]$, so (6) and Prop. 16 hold with $\sigma^2\Delta\psi_j$ replaced by the empirical moment of $\mathrm{tr}(D\nabla^2\psi_j)$. What does not extend is §3.2: the Hermite structure of the tangent generator (Lem. 30) uses a Gaussian reference with constant isotropic diffusion, and no rate is claimed beyond it.

Unknown diffusion is a different problem. The drift is then identified only jointly with D, which already for linear SDEs requires conditions on the initial law (Guan et al., 2024; App. I), and every identifiability statement here is conditional on D. Misspecifying D biases the weak form: with σ^2 replaced by 0 it fits an autonomous approximation to the time-dependent probability-flow velocity, which on the main benchmark raises the population floor from 0.0332 to 0.3545 (App. Y).

X.3 DESIGN SIZE WHEN $K < m$

The threshold $K \geq m$ of Thm. 8 is uniform over all polynomial degrees. With fewer distinct shapes, as in many single-cell time courses, the source constraints still identify every polynomial drift of degree at most $K-1$ and leave an exact gauge at degree K (App. E.1; proved for $K = m-1$, certified for $m \leq 6$). Three consequences follow. First, m is the intrinsic dimension of the chosen representation, not the number of measured features; ambient dimension does not enter (App. J). Second, for fixed K one may restrict the drift class to degree at most $K-1$; the linear class needs only $K = 2$ (App. I). Third, the factor K in Asm. 10 measures temporal excitation rather than the snapshot count, and designs whose shapes barely move lose it even when $K \geq m$ (Obs. 17).

Y EXPERIMENTAL PROTOCOL AND NEGATIVE-RESULT LOG

Y.1 PROTOCOL

Gates were fixed before each run and are recorded with the receipts. All comparisons within a table share seeds and draws; seed sets are stated wherever arms come from different runs. Readouts are

relative errors in the solenoidal channel unless marked otherwise, normalized by the Frobenius norm of the planted circulation; for null-circulation controls that denominator vanishes, so those values are reported against the norm of the symmetric part instead and are not comparable to the main readout. Standard errors are plug-in and noise-only in the whitened metric. The full suite runs on a single CPU core.

Multi-seed entries are mean $\pm$ standard deviation over seeds. Paired differences are mean $\pm$ standard error of the per-seed differences, and t is their ratio, the paired t-statistic with one fewer degree of freedom than seeds. Seed counts are small and the spread is heavy-tailed. A 32-seed replication of the weak-form GLS arm in the targeting study gives 0.824 ± 0.313, a standard deviation about $1.2\times$ the 12-seed value in Tab. 1 and $1.8\times$ an earlier 8-seed estimate, so we treat any paired difference below roughly two standard errors as no effect, and we do not report equivalence without an equivalence test.

Y.2 Reading Tab. 1

The four snapshot-only arms share one 12-seed set with identical draws, so their differences are paired; the two strong-form rows come from a separate 8-seed set and are not paired against them. The single-seed columns are one draw and are noisier than they look: seed 0 is a high draw for the parametric arm (0.511 against a 12-seed mean of 0.308), so the paired column is the reliable comparison. Seed spread is heavy-tailed relative to these counts (App. Y, protocol), so small paired differences should not be read as effects; for reference, the strong-form oracle gives 0.547 ± 0.182 over thirty-two seeds against 0.535 ± 0.177 over eight. The strong-form oracle row uses a Gauss–Newton solve; a closed-form linear least squares on exact ingredients reaches 0.351 and Adam on the same objective 0.510 ± 0.123 (App. M).

Y.3 Retracted and superseded claims

The estimand was the wrong field. The shipped weak form used the deterministic continuity identity, omitting the $\sigma^2\Delta\psi$ term of (6), and therefore fitted an autonomous approximation to the probability-flow velocity $v_t = F - \sigma^2 s_t$, which is time-dependent even for autonomous F. On the main benchmark this has a population floor of 0.3545 against 0.0332 with the term retained. The velocity-convention numbers are retained here in full. The within-run comparisons reported in Apps. J, P, U and V (frames, latent regime) come from receipts computed under the same velocity convention and have not been rerun under the correction.

n	weights	velocity convention	drift convention
40k	diagonal	0.8442	0.8928
40k	GLS	0.6361	0.6505
40k	GLS^2	0.6514	0.6534
200k	diagonal	0.3153	0.3030
200k	GLS	0.3217	0.2818
200k	GLS^2	0.3067	0.2556
population floor	—	0.3545	0.0332

At $n = 40$k sampling error dominates and the numbers barely move. At $n = 200$k the uncorrected estimator has converged to its own floor, so that column measured misspecification rather than sample size, and the fitted $n^{-1/2}$ slope of -0.61 was measuring saturation. Under the correction the floor is ten times lower, so the two sample sizes are no longer separated by a floor; with one draw at each they are consistent with $n^{-1/2}$ scaling but do not measure it (§5.1). A side effect: under the old convention GLS was *worse* than diagonal weighting at $n = 200$k, an artifact of a bias-dominated error in which variance-optimal whitening buys nothing.

Whitening was claimed as an efficiency gain, twice, in both directions. An early single-seed reading ($0.844 \to 0.636$) was reported as a gain; an eight-seed replication reduced it to $t = 1.67$ and the claim was retracted. A subsequent draft reinstated a stronger claim from statistics that compared unweighted against GLS rather than diagonal against GLS. The corrected twelve-seed comparison shows no effect of any weighting (§5.1), and §5.2 gives the structural reason.

A calibration pathology that did not exist. The $9.67\times$ conservatism of the diagonal predictor, and its repair by GLS, were measured on a lineage-tracked design in which the slices are not independent. With independent slices the diagonal predictor is calibrated at $1.00\times$ and there is nothing to repair; its power check, which prescribed 6.7M samples per slice and was overridden, was correct. See App. R.

Descent into the exact gauge. Withdrawn; see App. V.

A co-estimation phase transition. Withdrawn; see App. L.

Superseded spectral model. An earlier version analysed a stationary ladder inside the Euclidean divergence-free class $\ker(\mathrm{div}) \cap W_q$. The present theory uses $S_q = \ker \mathcal{A}_{\rho_\star} \cap W_q$ and a transient cross-slice operator, on which the reference Stein operator vanishes identically. Machine-precision checks of the former are not validation of the latter, and the associated lower bound, figures and verification claims have been removed rather than re-labelled.

Y.4 Negative results retained

Split-sample selection picks $\zeta = 0$ on every seed while a fixed ridge at $\zeta = 0.03$ does better (0.468 ± 0.112, $t = -5.09$). Random features and the correctly specified degree-three frame show no detectable difference at six seeds. A neural field at $d = 25$ returns the no-information level. Targeting the strong form does not close the gap to the score-free weak form at any score error we could reach.

Z Reproducibility

The supplement contains the estimator library, the benchmark suite, every experiment script, and machine-readable JSON receipts. The weak-form path depends only on NumPy and SciPy; the strong-form comparison arms require PyTorch and are imported lazily, so every number in Tab. 1 can be regenerated without it.

reported quantity	script (in `scripts/`)	receipt (in `results/`)
Tab. 1 scaling columns; population floors	`diffusion_rerun.py`	`diffusion_rerun.json`
Tab. 1 paired column	`paired_rerun.py`, `parametric_matched.py`	`table1_paired.json`
Obs. 17	`ou_shortgap.py`	`ou_shortgap.json`
Fig. 1(b); structural predictions of §5	`ot_blindness.py`	printed to stdout
App. R	`nonlinear_independent.py`	`nonlinear_independent.json`
strong-form arms (App. M)	`repair_strongform.py`, `paired_bench.py`	`../repair/*.json`
targeting study (Fig. 5)	`targeting/`	`../scripts/targeting/`
App. W; Fig. 8	`hidden_state_stress.py`	`hidden_state_stress.json`
App. S	`verify_identities.py`	`verification.json`
degree cap (App. E.1)	`check_gauge_degree.py`	`gauge_degree.json`

Superseded intermediate values are retained in the receipts rather than deleted, so the velocity-convention numbers of App. Y and the lineage-tracked numbers of App. R can be reproduced alongside the corrected ones. The repository README maps each table and each figure regenerated from the current receipts to its generating command.