CERTIFIED QUOTIENT NATURAL GRADIENTS IN WASSERSTEIN SPACE: RETRACTIONS, RAVINES, AND HIGHER-ORDER COLLISIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Degenerate Wasserstein pullback metrics arise from redundant parameters, unresolved transport potentials, and rank-changing generators. These mechanisms require different remedies. We develop a residual-certified quotient natural-gradient framework that distinguishes them. A matrix-valued Poisson residual bounds the true metric from above and below, certifies equality of nullspaces, and controls the induced distributional step without an absolute eigenvalue threshold in raw coordinates. Compatible horizontal lifts transfer retraction-based descent and local convergence to redundant generators. For genuine rank changes, we construct moment-balanced Ornstein–Uhlenbeck generators with collisions of arbitrarily high order. Their cumulant coordinates have nonzero Wasserstein tangents even when the raw metric and loss vanish to orders $2m-2$ and $2m$. Gauss–Hermite quadrature gives explicit K-component examples with $m = 2K$. We further reduce the true metric of rotated product generators to one- and two-dimensional Poisson problems, identifying precisely the rotations lost at Gaussian strata. Complete proofs cover inexact metric certificates, mixed-order ravines, chart contraction, angular cone limits, and statistical lower bounds. This is a pure theory study: it does not claim a globally regular chart for arbitrary neural generators or empirical superiority.

1 INTRODUCTION

A generator T_θ determines a distribution $\mu_\theta = (T_\theta)_{\#}\rho$, but a parameter velocity contains more information than the induced motion of μ_θ. It can move latent labels, circulate mass without changing density, or become invisible to the first derivative at a component collision. Consequently, the common-noise matrix $\mathbb{E}[DT_\theta^\top DT_\theta]$ need not be the true Wasserstein pullback, and replacing a singular matrix by a pseudoinverse does not resolve every degeneracy. Natural-gradient invariance on regular models is established theory (Chen & Li, 2020; van Oostrum et al., 2023); it does not imply uniform finite-step invariance through a change of rank.

This paper separates three questions. First, can approximate transport potentials *certify* the metric they approximate, including its kernel? Second, can the resulting step be implemented through a quotient-compatible retraction rather than a coordinate-dependent additive update? Third, when the generator differential really loses a direction, which new distribution coordinates restore it, and what information remains intrinsically absent?

Our first contribution is an a posteriori matrix inequality for the true metric. The scalar residual argument is classical (Repin, 2008); the object needed for generative optimization is a simultaneous quadratic-form certificate, compatible with redundant-coordinate congruences. It yields a nullspace test and an intrinsic step-error bound. Our second contribution combines this certificate with compatible lifts and a precise oracle model, giving descent and local convergence under explicit quotient smoothness assumptions. Our third contribution is an all-order collision construction and an exact pairwise Poisson reduction for rotated independent generators. The construction replaces a single fourth-cumulant example by a hierarchy of finite Gaussian mixtures, coupled orientation variables, and mixed-order ravines.

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.

The distinction is consequential. A small raw eigenvalue can reflect an arbitrary rescaling, whereas an angular direction at a Gaussian is not statistically identifiable. Our guarantees are invariant to redundant coordinates under fixed intrinsic assumptions; they are not independent of the conditioning of the identifiable distribution family or of the cost of solving its Poisson equations.

2 THE TRUE PULLBACK AND THE QUOTIENT

Let $\Theta \subset \mathbb{R}^P$ be open, $T : \Theta \to L^2(\rho; \mathbb{R}^d)$ be continuously differentiable, and p_θ a positive density. Boundary conditions are periodic, zero flux, or sufficient decay; the function spaces are specified in Appendix B. Define

$$W_\theta a = DT_\theta[a], \qquad M_\theta a(x) = \mathbb{E}[W_\theta a \mid T_\theta = x], \qquad A_\theta a = P_{\mu_\theta} M_\theta a, \quad G_\theta = A_\theta^* A_\theta, \tag{2.1}$$

where P_μ projects onto the $L^2(\mu)$ closure of gradients. For a differentiable energy $\mathcal{F}$, write $f(\theta) = \mathcal{F}(\mu_\theta)$ and $g = Df$. If $\xi_\mu = \nabla \delta\mathcal{F}/\delta\mu$ exists, then $g = A^*\xi_\mu$ and the exact induced natural velocity is $-P_{\operatorname{ran} A}\xi_\mu$.

The two losses of information are orthogonal:

$$\|Wa\|^2_{L^2(\rho)} = \|Aa\|^2_{L^2(\mu)} + \|(I - P_\mu)Ma\|^2_{L^2(\mu)} + \mathbb{E}\|Wa - \mathbb{E}[Wa \mid T]\|^2. \tag{2.2}$$

Thus the common-noise metric dominates G, but the excess can be entirely unrelated to useful distribution motion. Appendix C proves the differentiation and projection statements rather than identifying the two metrics by assumption.

A *regular quotient representation* is a submersion $\pi : \Theta \to \mathcal{X}$ together with an injectively parameterized distribution model $x \mapsto \mu_x$, such that $\mu_\theta = \mu_{\pi(\theta)}$ and its true metric G_x is positive definite. Then

$$G_\theta = D\pi_\theta^\top G_x D\pi_\theta, \quad \ker G_\theta = \ker D\pi_\theta, \quad D\pi_\theta G_\theta^\dagger D\pi_\theta^\top = G_x^{-1}. \tag{2.3}$$

These local quotient assumptions are not consequences of arbitrary constant-rank matrix observations; they include an actual model chart. Appendix D gives the proof and explains the necessary distinction from singular images.

3 A RESIDUAL CERTIFICATE FOR THE METRIC

At a fixed distribution, the potential for coordinate i solves

$$a_\mu(\phi_i, \psi) := \int \nabla\phi_i \cdot \nabla\psi \,\mathrm{d}\mu = \int \psi b_i \,\mathrm{d}x, \qquad b_i = \partial_i p. \tag{3.1}$$

Assume a Poincaré inequality $\operatorname{Var}_\mu \psi \le C_{\mathrm{P}} a_\mu(\psi, \psi)$ with a known valid constant. Let $\widetilde{\phi}_i$ be arbitrary admissible trial potentials and define

$$\begin{aligned} L_{ij} &= \int b_i \widetilde{\phi}_j \,\mathrm{d}x + \int b_j \widetilde{\phi}_i \,\mathrm{d}x - a_\mu(\widetilde{\phi}_i, \widetilde{\phi}_j), \\ r_i &= b_i/p + p^{-1} \operatorname{div}(p\nabla\widetilde{\phi}_i), \qquad R_{ij} = \int r_i r_j \,\mathrm{d}\mu. \end{aligned} \tag{3.2}$$

The residual is required to be in $L^2(\mu)$, to have zero mean, and to include any boundary mismatch; we use trial functions satisfying the stated boundary condition.

Theorem 3.1 (Simultaneous residual certificate)**.** *Under these assumptions,*

$$L \preceq G \preceq L + C_{\mathrm{P}} R. \tag{3.3}$$

If $L \succeq 0$ and $C_{\mathrm{P}} R \preceq \eta L$ on the whole *parameter tangent space, then*

$$L \preceq G \preceq (1 + \eta)L, \qquad \ker L = \ker G. \tag{3.4}$$

For any $g \in \operatorname{ran} G$, let $v = -AG^\dagger g$ and $v_L = -AL^\dagger g$. Then

$$\|v_L - v\|_{L^2(\mu)} \le \eta \|v\|_{L^2(\mu)}, \qquad g^T G^\dagger g \le g^T L^\dagger g \le (1 + \eta) g^T G^\dagger g. \tag{3.5}$$

All inequalities are preserved by coordinate congruences, including redundant submersions with the same represented tangent space.

Proof mechanism. Let $e_i = \phi_i - \widetilde{\phi}_i$. Expanding the energy gives the exact identity $G - L = (a_\mu(e_i, e_j))_{ij}$. For $e_a = \sum_i a_i e_i$ and $r_a = \sum_i a_i r_i$, $a_\mu(e_a, e_a) = \int r_a e_a \, \mathrm{d}\mu \le \sqrt{C_{\mathrm{P}}}\|r_a\|\sqrt{a_\mu(e_a, e_a)}$. This proves the matrix bound simultaneously for every a, not just its diagonal. The remaining assertions follow on the quotient by whitening with G; the eigenvalues of $G^{1/2}L^{-1}G^{1/2}$ lie in $[1, 1+\eta]$. Appendix E supplies the singular-space and congruence details. □

For exact Galerkin potentials, $L = B^T K^\dagger B$ where K is the trial Dirichlet matrix and B its source matrix. For inexact solves, the symmetrized L above remains a lower bound; the trial gradient Gram alone does not. A certificate restricted to $\operatorname{ran} L$ without checking $\ker L \subseteq \ker R$ can miss an entirely unresolved distribution direction. This is the central reason not to replace the test by an absolute eigenvalue cutoff.

Oracle model and algorithm. RCQ-NG assumes a known regular quotient chart or known gauge, trial-potential evaluations, a valid C_{P}, and certified energy/gradient and residual integrals. These are stronger oracles than samples from an arbitrary implicit network. At each iteration: construct L, R; refine the trial space until (3.4) is certified; solve the quotient system; apply a fixed quotient retraction and backtracking; lift the new point horizontally. A failed certificate means *unresolved geometry*, not convergence. At an analytically identified collision, the algorithm may change to a verified jet chart; no generic collision detector is claimed. Appendix G states the complete algorithm and a sufficient termination condition. Appendix F handles numerical integration and independent finite-sample validation.

3.1 An implementable acceptance test, not a rank heuristic

For a certified quotient representation, the following steps specify the method. They also describe what it means for the method to be unable to certify a proposed step. In particular, a finite potential basis does not determine the dimension of the true distribution tangent space.

Residual-certified quotient natural gradient (RCQ-NG). Fix a regular quotient chart, its retraction, a relative tolerance $\eta > 0$, and an admissible trial-space sequence. At an iterate x:

1. Obtain $g = Df_x$ and a valid Poincaré bound. Compute trial potentials, the lower matrix L, and a residual upper matrix $M \succeq C_{\mathrm{P}} R$.
2. Accept the geometry only after proving $L \succeq 0$ and $M \preceq \eta L$ on the entire represented tangent space. Otherwise refine the solves and integration. A finite-budget failure returns `unresolved`, not `stationary`.
3. Check that g annihilates the certified gauge. Solve $Lw = -g$ on the quotient, or use $w = -L^\dagger g$ in redundant coordinates.
4. Backtrack in the *fixed quotient retraction*. Accept only a step satisfying the stated exact or interval-validated energy test. Lift the accepted point with the compatible horizontal map.

At a known rank-changing point, use a separately verified resolving chart (JET-CQNG). This changes the differential being represented; it is not another pseudoinverse threshold.

The matrix test contains a kernel requirement that is easy to miss. Consider the exact rational example

$$L = \operatorname{diag}(1, 0), \qquad M = \operatorname{diag}(0, 1), \qquad G = L + M = I_2.$$

Every residual test restricted to $\operatorname{ran} L$ passes with zero error, yet the second direction is completely unresolved. The whole-space test fails for every finite η. By contrast, if $L = \operatorname{diag}(1, 0)$ and $M = \operatorname{diag}(\eta/2, 0)$, the null direction is genuinely shared by every G in the certified bracket. Both conclusions persist under $L \mapsto J^T L J$, $M \mapsto J^T M J$ for any full-row-rank redundant representation J.

For a fixed positive intrinsic gap, graph-norm convergence of the trial potentials makes the residual matrix converge to zero and eventually passes the relative test. Uniformity near a disappearing *intrinsic* eigenvalue requires a correspondingly more accurate solve; no absolute-tolerance argument removes that cost. With numerical integration, a lower enclosure L_- and upper enclosure U_+ replace L and $L + C_{\mathrm{P}} R$. The sufficient test is $0 \preceq L_-$ and $U_+ \preceq (1+\eta) L_-$, with vertical directions treated analytically. Appendix F proves a fresh-sample validation rule under bounded normalized integrands. It does not infer exact kernels from noisy matrix entries alone.

4 FINITE STEPS: RETRACTIONS, ERRORS, AND CONVERGENCE

Fix a quotient retraction Retr_x and suppose

$$f(\mathrm{Retr}_x v) \le f(x) + Df_x[v] + \frac{L_f}{2}\|v\|_{G_x}^2 \tag{4.1}$$

for the admissible steps in a neighborhood. This follows locally from smoothness on a compact regular set, but not automatically at a merely C^1 collision chart. A horizontal lift $\widetilde{\mathrm{Retr}}_\theta$ is *compatible* when $\pi(\widetilde{\mathrm{Retr}}_\theta v) = \mathrm{Retr}_{\pi(\theta)}(D\pi_\theta v)$ and its derivative is the identity on the chosen horizontal space.

Theorem 4.1 (Certified retraction descent)**.** *Suppose* (4.1) *holds,* $L \preceq G \preceq (1+\eta)L$ *on the quotient, and* $g = Df$ *is exact. For* $w = -L^{-1}g$ *and* $0 < \alpha \le [L_f(1+\eta)]^{-1}$ *with an admissible step,*

$$f(\mathrm{Retr}_x(\alpha w)) \le f(x) - \frac{\alpha}{2}\|\mathrm{grad}_G f(x)\|_G^2. \tag{4.2}$$

If $\|\mathrm{grad}_G f\|^2 \ge 2\mu(f - f_*)$ *and* $0 < \alpha\mu < 1$ *with all iterates in this neighborhood, then* $f(x_t) - f_* \le (1-\alpha\mu)^t(f(x_0) - f_*)$*. Compatible horizontal lifts produce exactly the same distribution iterates in any regular redundant representation. For a relatively perturbed metric and gradient, the analogous constants are given in Appendix I, together with sufficient conditions ensuring neighborhood invariance.*

Proof. Set $d = g^T L^{-1} g$. Then $\|w\|_G^2 \le (1+\eta)d$ and $d \ge \|\mathrm{grad}_G f\|^2$. Substitute into (4.1). Compatibility and (2.3) prove the distributional invariance. □

Compatible lifts exist locally by submersion coordinates. They are not generally additive parameter updates. For $z = a^m$ and a horizontal raw velocity w, the exact lift of an additive z-step is

$$\widetilde{\mathrm{Retr}}_a(w) = (a^m + ma^{m-1}w)^{1/m} = a + w - \frac{m-1}{2a}w^2 + O(w^3). \tag{4.3}$$

The coefficient blows up as $a \to 0$. No uniform second-order raw retraction bound follows from regular-stratum smoothness. Appendix H proves this and gives a Gaussian covariance-factor example.

4.1 PERTURBATIONS AND A VERIFIABLE LOCAL BASIN

The descent guarantee extends to simultaneous metric and covector errors, but their sizes must be measured intrinsically. Suppose

$$(1-\varepsilon)G \preceq \widehat{G} \preceq (1+\varepsilon)G, \qquad \|\widehat{g} - g\|_{G^{-1}} \le \sigma\|g\|_{G^{-1}}.$$

Let $a_0 = (1+\varepsilon)^{-1} - \sigma(1-\varepsilon)^{-1} > 0$ and $b_0 = (1+\sigma)/(1-\varepsilon)$. Then the update $w = -\widehat{G}^{-1}\widehat{g}$ satisfies

$$f(\mathrm{Retr}_x(\alpha w)) \le f(x) - \frac{\alpha a_0}{2}\|\mathrm{grad}_G f\|^2, \qquad 0 < \alpha \le \frac{a_0}{L_f b_0^2}. \tag{4.4}$$

This follows by whitening the metric before bounding the cross term $g^T\widehat{G}^{-1}(\widehat{g} - g)$; a raw-coordinate error bound would not be representation invariant.

A normal Hessian gap along a compact minimizer manifold S gives local PL, quadratic growth transverse to S, and an upper bound $\|\mathrm{grad}\, f\| \le B\sqrt{f - f_*}$. If the corresponding rate is $q = 1 - \alpha a_0 \mu \in (0,1)$ and retraction segments have length at most $C_R\|v\|_G$, their total length is bounded by

$$\sum_{t\ge 0} \mathrm{length}(x_t \rightsquigarrow x_{t+1}) \le \frac{C_R \alpha b_0 B\sqrt{f(x_0) - f_*}}{1 - \sqrt{q}}. \tag{4.5}$$

Requiring this bound to be smaller than an interior radius proves that the iterates remain in the region where the estimates hold. Appendix I gives the induction and the uniform Taylor bounds. Thus the local convergence argument does not assume its own neighborhood invariance. A fixed nonzero absolute covector error instead yields a validated small-gradient stopping region, not exact convergence.

4.2 A REGULAR GENERATIVE FITTING PROBLEM WITH VERIFIABLE ORACLES

The regular theorem is nonempty beyond Gaussian covariance models. On the unit flat torus let

$$p_z(x) = \exp\left(\sum_{i=1}^{r} z_i\varphi_i(x) - A(z)\right), \qquad A(z) = \log \int \exp\left(\sum_i z_i\varphi_i(x)\right) \mathrm{d}x, \tag{4.6}$$

where the smooth real features are independent modulo constants. Given a target feature vector t, such as empirical feature averages, fit the generative density by $f_t(z) = A(z) - t^T z$. Assume $t = \nabla A(z_*)$ for an interior point of the considered parameter neighborhood. The Hessian is $\mathrm{Cov}_{p_z}(\varphi) \succ 0$, so the fitting problem has local PL in the smooth true Wasserstein metric. This is a standard exponential-family objective, not a claimed new statistical loss.

The metric sources are the explicit centered functions $b_i = p_z(\varphi_i - \mathbb{E}_{p_z}\varphi_i)$. If the oscillation of $\sum_i z_i\varphi_i$ is bounded by D, then the known Poincaré upper bound is $e^D/(4\pi^2)$. Fourier Galerkin potentials converge in the strong-residual norm because the smooth elliptic solutions admit H^2 approximation; positivity of the intrinsic metric gives eventual certificate acceptance. Normalizer and residual integrals still need validated integration. A redundant submersion $z = \pi(\theta)$ changes neither the source problem nor the certified distribution step. Appendix R verifies these assertions and explains density sampling. Thus the method addresses a concrete density-learning task, while its assumptions are not incorrectly attributed to a general sample-only implicit network.

5 HIGHER-ORDER COLLISION GENERATORS

Let X be bounded, centered, of variance one, and let its moments through order $m-1$ agree with a standard Gaussian, with $m \geq 3$ and $c = \mathbb{E}H_m(X) \neq 0$. Here H_j are probabilists' Hermite polynomials. Define an Ornstein–Uhlenbeck smoothed generator

$$Y_a = aX + \sqrt{1-a^2}Z, \quad Z \sim \mathcal{N}(0,1), \quad 0 \leq a < a_0 < 1, \qquad \kappa = ca^m. \tag{5.1}$$

Its first $m-1$ moments remain Gaussian and its mth cumulant is κ. If X has K atoms, Y_a is a K-component Gaussian mixture. Write $q_\kappa = F_\kappa^{-1} \circ \Phi$.

Theorem 5.1 (All-order radial resolution)**.** *For the model* (5.1)*, in every finite Gaussian L^r space,*

$$q_\kappa = \mathrm{Id} + \frac{\kappa}{m!}H_{m-1} + O_{L^r}(|\kappa|^{1+1/m}), \qquad q'_\kappa = \frac{H_{m-1}}{m!} + O_{L^r}(|\kappa|^{1/m}). \tag{5.2}$$

For symmetric X the exponents $1/m$ improve to $2/m$. The one-sided chart is C^1 at zero and

$$h_m(0) = \|q'_0\|^2 = \frac{1}{m\,m!}, \quad G_{aa} \sim \frac{mc^2}{m!}a^{2m-2}, \quad \tfrac{1}{2}W_2^2(\mathrm{Law}(Y_a),\gamma) \sim \frac{c^2}{2m\,m!}a^{2m}. \tag{5.3}$$

On a sufficiently small cumulant interval, the projected chart natural gradient for a realizable target contracts uniformly up to and including $\kappa = 0$. The constants concern a fixed family and may depend on m and X.

The proof differentiates the quantile equation before estimating its remainder; an asymptotic density expansion alone is insufficient. Appendix K provides polynomial envelopes for every required derivative, while Appendix J proves the abstract C^1 contraction. Bounded smooth perturbations of a Gaussian with prescribed first nonzero Hermite coefficient give opposite-sign branches, so the one-sided curves can be joined with the same first tangent (Appendix M). The join is not asserted to be C^2.

Corollary 5.2 (An explicit finite-mixture hierarchy)**.** *Let X_K have the K-node Gauss–Hermite quadrature law for γ, with $K \geq 2$. Then $m = 2K$, $c = -K!$, and* (5.3) *becomes*

$$G_{aa} \sim \frac{2K(K!)^2}{(2K)!}a^{4K-2}, \qquad \tfrac{1}{2}W_2^2(\mathrm{Law}(Y_a),\gamma) \sim \frac{(K!)^2}{4K(2K)!}a^{4K}. \tag{5.4}$$

In particular, the two-component sixth-order metric/eighth-order loss is only the first member of the hierarchy.

Table 1: Exact asymptotic coefficients for the finite-mixture hierarchy. Entries are algebraic consequences of Corollary 5.2, not fitted numerical slopes. $G_{aa} \sim C_G a^{4K-2}$ and $\frac{1}{2}W_2^2(p_a, \gamma) \sim C_F a^{4K}$.

Components K	Cumulant order m	C_G	C_F	$h_m(0)$
2	4	$2/3$	$1/48$	$1/96$
3	6	$3/10$	$1/240$	$1/4320$
4	8	$4/35$	$1/1120$	$1/322560$
5	10	$5/126$	$1/5040$	$1/36288000$

5.1 A CONCRETE HIGHER-ORDER MODEL AND ITS UPDATE

For $K = 3$, take $X_3 \in \{-\sqrt{3}, 0, \sqrt{3}\}$ with probabilities $1/6, 2/3, 1/6$. This gives the explicit family

$$p_a = \tfrac{1}{6}\mathcal{N}(-\sqrt{3}a, 1-a^2) + \tfrac{2}{3}\mathcal{N}(0, 1-a^2) + \tfrac{1}{6}\mathcal{N}(\sqrt{3}a, 1-a^2), \quad \kappa = -6a^6. \tag{5.5}$$

It matches Gaussian moments through order five. Its raw metric behaves as $3a^{10}/10$ and its loss to the Gaussian as $a^{12}/240$, whereas the cumulant metric tends to $1/4320$. No components disappear in this limit: all three retain positive weights and variances tending to one. This differs from the small-variance regime of Gaussian-mixture transport.

Proposition 5.3 (A first-derivative chart guarantee)**.** *Let $Q : I \to \mathcal{H}$ be a C^1 curve in a real Hilbert space, with I a closed interval containing z_*, and suppose $\sup_{z\in I}\|Q'(z) - V\| \le \delta\|V\|$, where $V \neq 0$ and $\delta < 1/3$. For $f(z) = \frac{1}{2}\|Q(z) - Q(z_*)\|^2$, the projected step*

$$z^+ = \Pi_I\left(z - \frac{\langle Q(z) - Q(z_*), Q'(z)\rangle}{\|Q'(z)\|^2}\right) \tag{5.6}$$

satisfies $|z^+ - z_*| \le 2\delta(1-\delta)^{-1}|z - z_*|$. This includes an endpoint initialization and needs no second derivative at that endpoint.

Indeed $Q(z) - Q(z_*) = (z - z_*)\int_0^1 Q'(z_* + t(z - z_*))\,\mathrm{d}t$; compare the averaged tangent with the current tangent, and use nonexpansiveness of interval projection. Theorem 5.1 supplies this condition for the cumulant chart after shrinking its interval. In one dimension, $Q = q$ is the *exact* quantile representation of Wasserstein distance. A direct-sum version covers products with fixed coordinates; it is not an isometric representation of all multidimensional Wasserstein space.

For $K = 3$, equation (5.6) can leave $\kappa = 0$ toward any sufficiently close realizable negative target. Every raw a-gradient is zero at $a = 0$ because $q'_a = 0$ there. A constant chart metric $1/4320$ also gives a local contraction: resolving the coordinate is the essential change, while the varying metric follows the nonconstant geometry. The algorithm is therefore not presented as a new general Gauss–Newton principle.

Mixed-order ravines. For a regular identifiable energy $F(u, z)$ with positive-definite Hessian at a minimizer, profiling the fast variable gives $u = h(z)$. Under raw coordinates $z_j = a_j^{m_j}$, the graph $u = h(z(a))$ is a ravine with quadratic normal growth and restricted growth comparable to $\sum_j |a_j|^{2m_j}$. Different m_j need not admit a single two-sided growth order. Appendix N proves the graph and Hessian-kernel statements. Chart convergence uses the extra known generator structure; it is not a generic acceleration theorem for arbitrary flat objectives or a replacement for the ravine theory of Davis et al. (2025).

6 ROTATED PRODUCTS: A TRUE LOW-DIMENSIONAL METRIC REDUCTION

The model $\mu_{\kappa,U} = \mathrm{Law}(U(Y_1, \ldots, Y_d)^T)$ couples independent standardized generators by an unknown $U \in O(d)$. Let p_i be the marginal densities, $s_i = (\log p_i)'$, and $\mathcal{L}_i = -\partial_{ii} - s_i\partial_i$. For an infinitesimal coordinate-plane rotation define

$$b_{ij}(x_i, x_j) = x_j s_i(x_i) - x_i s_j(x_j), \qquad J_{ij} = \langle b_{ij}, (\mathcal{L}_i + \mathcal{L}_j)^{-1} b_{ij}\rangle_{p_i \otimes p_j}. \tag{6.1}$$

The inverse is on the mean-zero tensor-product subspace. Positive Poincaré gaps suffice for its existence.

Theorem 6.1 (Pairwise Poisson metric)**.** *For standardized smooth marginals with the stated integrability and Poincaré assumptions, the true metric of the rotated product model is*

$$\sum_i h_i(\kappa_i)\,\mathrm{d}\kappa_i^2 + \sum_{i<j} J_{ij}(\kappa_i,\kappa_j)\Omega_{ij}^2, \qquad \dot{U} = U\Omega, \quad \Omega^T = -\Omega. \tag{6.2}$$

There are no radial–angular or distinct-pair cross terms. Moreover, $J_{ij} = 0$ exactly when both marginals are standard Gaussian. If marginal i has first nonzero order $m_i \geq 3$, then near a joint Gaussian collision

$$J_{ij} = \frac{\kappa_i^2}{m_i!} + \frac{\kappa_j^2}{m_j!} + o(\kappa_i^2 + \kappa_j^2). \tag{6.3}$$

Thus a d-dimensional transport metric reduces to d one-dimensional radial calculations and at most $d(d-1)/2$ two-dimensional Poisson problems. This is a dimensional reduction of the PDEs, not a dimension-free complexity bound.

Appendix P proves the decomposition using centered tensor-product spaces. It also gives a residual certificate with coercivity gap $\lambda_i + \lambda_j$ for each pair. On a fixed-rank stratum, RCQ-NG can use the resulting block metric and the orthogonal retraction $U\exp(\Omega)$. Rotations inside the Gaussian block are gauge directions; small nonzero J_{ij} are not.

Relative accuracy despite a vanishing angular metric. The polynomial pair potential $\psi_{ij}^0 = \kappa_i H_{m_i-1}(x_i)x_j/m_i! - \kappa_j H_{m_j-1}(x_j)x_i/m_j!$ provides more than a formal leading term. With $s^2 = \kappa_i^2 + \kappa_j^2$, its energy is bounded above and below by positive multiples of s^2, and its strong residual is $O(s\tau)$, where $\tau = |\kappa_i|^{\beta_i} + |\kappa_j|^{\beta_j}$ and $\beta_i = 1/m_i$ (or $2/m_i$ in the symmetric case). A fixed positive pair Poincaré gap therefore implies a relative certificate

$$D_{ij} \leq J_{ij} \leq (1 + O(\tau^2))D_{ij}, \qquad D_{ij} > 0 \quad (s > 0 \text{ sufficiently small}). \tag{6.4}$$

Appendix U proves a quantitative version and describes the hybrid rule: try this explicit potential, validate its normalized residual, and enrich it only when the certificate fails. This connects the high-order collision construction to the certified quotient algorithm. At the collision the zero source is known analytically; near it the metric can be accurately approximated while its inverse remains intrinsically large.

For a single non-Gaussian axis u, the pair equation becomes a one-dimensional equation $(-\partial_{xx} - s_\kappa\partial_x + 1)w = mv_\kappa$ and $j(\kappa) = \kappa^2\int[(w')^2 + w^2]p_\kappa$. Hence $j(\kappa) \sim \kappa^2/m!$, while $h_m(0) > 0$. Appendix O derives the corresponding cone limit. A cumulant chart removes radial coordinate flattening, not the absence of orientation information at the Gaussian.

7 Scope, comparison, and conclusion

Natural-gradient least-squares formulations (Nurbekyan et al., 2023), parameterized Wasserstein flows (Jin et al., 2025), affine Wasserstein shape geometry (ichi Amari & Matsuda, 2024), and mixture information-matrix limits (Li & Zhao, 2023) are closely related. Our certificate concerns the *true* weighted-Poisson metric and all represented tangent directions simultaneously; it is not a claim that prior methods have no error analysis. Moment parameterizations and weak mixture identifiability are established (Agostini et al., 2021; Wu & Yang, 2020; Dwivedi et al., 2020). Recent Wasserstein ICA objectives (Laplante et al., 2026; Jha et al., 2026) do not by themselves identify the true pullback metric of a rotated generative family. Cumulant transport expansions also appear in Zenker (2026). We claim neither moment-based ICA nor the Gaussian linearization as new in isolation.

The framework supplies exact identities, conditional algorithmic guarantees, and constructive collision models. It does not provide a universally discoverable quotient or a globally regular neural-generator chart. Residual validation has explicit oracle costs; the statistical separation rate remains $n^{-1/(2m)}$ in the raw coordinate (Appendix Q). Purely additive raw updates, unjustified kernel truncation, and a C^2 assumption at a C^1 join are ruled out by counterexamples. The resulting principle is operational: certify the identifiable metric, implement its step through the quotient, and change charts only when a distributional jet has actually been established.

AI USE STATEMENT

The authors declare that no generative AI tools or LLMs were used to generate text, code, or ideas for this paper. Standard spell-checking and basic grammar-checking tools were used solely for proofreading.

REFERENCES

P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. URL https://press.princeton.edu/absil.

Daniele Agostini, Carlos Améndola, and Kristian Ranestad. Moment identifiability of homoscedastic Gaussian mixtures. *Foundations of Computational Mathematics*, 21:695–724, 2021. doi: 10.1007/s10208-020-09469-6. URL https://arxiv.org/abs/1905.05141.

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser, second edition, 2008. doi: 10.1007/978-3-7643-8722-8.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2014. doi: 10.1007/978-3-319-00227-9.

Yifan Chen and Wuchen Li. Optimal transport natural gradient for statistical manifolds with continuous sample space. *Information Geometry*, 3:1–32, 2020. doi: 10.1007/s41884-020-00028-0. URL https://arxiv.org/abs/1805.08380.

Damek Davis, Dmitriy Drusvyatskiy, and Liwei Jiang. Gradient descent with adaptive stepsize converges (nearly) linearly under fourth-order growth. *Mathematical Programming*, 2025. doi: 10.1007/s10107-025-02290-5. URL https://arxiv.org/abs/2409.19791.

Raaz Dwivedi, Nhat Ho, Koulik Khamaru, Martin J. Wainwright, Michael I. Jordan, and Bin Yu. Sharp analysis of expectation-maximization for weakly identifiable models. In *International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 1866–1876, 2020. URL https://proceedings.mlr.press/v108/dwivedi20a.html.

Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, second edition, 2010.

Walter Gautschi. *Orthogonal Polynomials: Computation and Approximation*. Oxford University Press, 2004. doi: 10.1093/oso/9780198506720.001.0001.

Aapo Hyvärinen and Erkki Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5):411–430, 2000. doi: 10.1016/S0893-6080(00)00026-5.

Shun ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998. doi: 10.1162/089976698300017746.

Shun ichi Amari and Takeru Matsuda. Information geometry of Wasserstein statistics on shapes and affine deformations. *Information Geometry*, 7:285–309, 2024. doi: 10.1007/s41884-024-00139-y.

Ashutosh Jha, Michel Besserve, and Simon Buchholz. Linear independent component analysis via optimal transport. arXiv:2607.14081, 2026. URL https://arxiv.org/abs/2607.14081.

Yijie Jin, Shu Liu, Hao Wu, Xiaojing Ye, and Haomin Zhou. Parameterized Wasserstein gradient flow. *Journal of Computational Physics*, 524:113660, 2025. doi: 10.1016/j.jcp.2024.113660. URL https://arxiv.org/abs/2404.19133.

Kaniav Kamary, Jeong Eun Lee, and Christian P. Robert. Weakly informative reparameterizations for location-scale mixtures. *Journal of Computational and Graphical Statistics*, 27(4):836–848, 2018. doi: 10.1080/10618600.2018.1438900. URL https://arxiv.org/abs/1601.01178.

Félix Laplante, Christophe Ambroise, and Pierre Humbert. Contrast-free ICA and causal inference via Wasserstein distances to the Gaussian. arXiv:2607.12832, 2026. URL https://arxiv.org/abs/2607.12832.

Wuchen Li and Jiaxi Zhao. Scaling limits of the Wasserstein information matrix on Gaussian mixture models. arXiv:2309.12997, 2023. URL https://arxiv.org/abs/2309.12997.

Levon Nurbekyan, Wanzhou Lei, and Yunan Yang. Efficient natural gradient descent methods for large-scale PDE-based optimization problems. *SIAM Journal on Scientific Computing*, 45(4): A1621–A1655, 2023. doi: 10.1137/22M1477805. URL https://arxiv.org/abs/2202.06236.

Sergey I. Repin. *A Posteriori Estimates for Partial Differential Equations*, volume 4 of *Radon Series on Computational and Applied Mathematics*. De Gruyter, 2008. doi: 10.1515/9783110203042.

Asuka Takatsu. Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics*, 48: 1005–1026, 2011. URL https://arxiv.org/abs/0801.2250.

Jesse van Oostrum, Johannes Müller, and Nihat Ay. Invariance properties of the natural gradient in overparametrised systems. *Information Geometry*, 6(1):51–67, 2023. doi: 10.1007/s41884-022-00067-9.

Yihong Wu and Pengkun Yang. Optimal estimation of Gaussian mixtures via denoised method of moments. *The Annals of Statistics*, 48(4):1981–2007, 2020. doi: 10.1214/19-AOS1873. URL https://arxiv.org/abs/1807.07237.

Connor P. Zenker. Cumulant perturbations of Gaussian optimal transport and Wasserstein-2 metric. Senior thesis, Princeton University, 2026. URL https://theses-dissertations.princeton.edu/handle/88435/dsp01v405sd89n.

A LOGICAL SCOPE, ASSUMPTIONS, AND NOTATION

This supplement gives proofs and constructions for the claims in the main paper. It is organized around three mathematically different objects: the differential of a parameterized probability law; the finite-dimensional regular quotient of that law; and a specified extension through a singular parameter value. None of these is silently identified with either of the others.

A.1 READING GUIDE

Section	Subject	Starts on page
B	Weighted Sobolev spaces and the minimum-energy velocity	11
C	Generator differentiation and the two projection losses	13
D	Regular quotients and singular Gram matrices	14
E	Matrix residual certificate and intrinsic step control	16
F	Inexact solves, integration errors, and validation	18
G	Complete certified algorithm and termination conditions	20
H	Compatible retractions and raw-coordinate obstructions	22
I	Descent, local convergence, and invariant neighborhoods	24
J	First-derivative chart contraction and higher-order jets	26
K	Quantile regularity for moment-balanced OU generators	29
L	Explicit finite-mixture collision hierarchy	31
M	Opposite-sign branches and the regularity of a join	33
N	Profiled ravines with different orders of growth	35
O	Single-axis geometry, residual bounds, and cone limits	37
P	Coupled rotations and pairwise Poisson equations	39
Q	What the chart does not change statistically	42
R	Explicit regular density and Gaussian factor models	44
S	Conditional finite-time flow consistency	47
T	Failure examples and sharp distinctions	48
U	Jet-normalized relative certificates near collisions	50
V	Proof dependencies and executable identity checks	52

A.2 ASSUMPTION LEDGER

The following ledger prevents conditions from migrating between results. A result that uses one row does not automatically inherit the conclusions of another row.

Tangent-space assumptions (T). At a fixed parameter, μ is a probability measure with positive density on $\mathbb{R}^d$, a flat torus, or a bounded domain with a specified no-flux boundary condition. The gradient space is the closure of admissible test gradients in $L^2(\mu)$. A differentiable generator has derivative in $L^2(\rho)$, and bounded smooth tests justify differentiation under the expectation. These assumptions suffice for the orthogonal-projection definition of the true tangent velocity. A Poincaré inequality is not needed for the projection itself.

Potential and certificate assumptions (P). The measure satisfies a Poincaré inequality with a finite, valid upper bound C_{P} for its constant. Sources have zero total mass and define bounded functionals on the weighted Dirichlet space. For the computable strong-residual bound, trial potentials satisfy the boundary condition and each strong residual belongs to $L^2(\mu)$. The numerical certificate additionally requires validated integrals or the probabilistic validation assumptions stated later. An empirical average without a confidence bound is not a deterministic certificate.

Quotient assumptions (Q). A local submersion $\pi : \Theta \to \mathcal{X}$ and a distribution model on $\mathcal{X}$ are supplied. The latter has a positive-definite true metric and is locally identifiable. Smooth dependence is assumed to the order needed for the chosen retraction. This does not follow merely because a numerical Gram matrix has an apparent rank. The distribution image of a rank-changing map can be a stratified set rather than a manifold.

Optimization assumptions (O). A quotient energy and its differential are available. The retraction pullback has the quadratic upper bound (4.1). Linear convergence additionally uses a local PL inequality and either an invariant sublevel neighborhood or the trapping lemma below. For perturbed gradients, error is measured in the dual intrinsic metric. Absolute Euclidean error in arbitrary raw coordinates is not an invariant substitute.

Jet assumptions (J). A curve or map into a Hilbert space has a differentiated higher-order expansion with a nonzero leading jet. The derivative remainder is part of the assumption; it is proved for the OU examples. The resulting chart need only be C^1. Its convergence theorem uses closeness of first derivatives, not a Lipschitz Hessian or a second-order retraction at the join.

Product assumptions (I). The sources are independent in a known coordinate frame, centered, and standardized. Their densities are positive and smooth, their scores and coordinate functions have the integrability used by the pairwise equations, and their weighted Poincaré constants are finite. Local asymptotics require uniform versions of these bounds. These conditions hold near the Gaussian for the finite bounded-mixture and bounded density-perturbation examples established below. They are not assumed for arbitrary heavy-tailed sources.

Statistical assumptions (S). Samples used in the lower-bound argument are independent observations from a specified one-parameter family. A local χ^2 expansion and uniform $2m$th moments are verified. The result is about estimating the raw separation, not the number of components, a general mixing distribution, or a minimax claim over all neural generators.

A.3 Notation and conventions

Vectors in the raw parameter space are denoted a or v; the scalar $a \geq 0$ in the OU construction is always identified by context. The letter m denotes the first nonzero non-Gaussian order, never the number of raw parameters. That number is P, the quotient dimension is r, and the number of trial potentials is k. G is the true pullback Gram matrix, L a certified lower matrix, and R a strong-residual Gram matrix. A dagger denotes the Moore–Penrose inverse with respect to the displayed Euclidean coordinates. Its *pushforward*, not the raw vector it selects, has the invariance asserted here.

For a positive-definite matrix S, $\|v\|_S^2 = v^T S v$ and $\|\ell\|_{S^{-1}}^2 = \ell^T S^{-1} \ell$. For a positive-semidefinite matrix these expressions are used on the quotient by its kernel, or equivalently on its Euclidean range. An inequality between singular matrices is always an inequality on the full displayed vector space unless an explicit quotient has already been taken. In particular, $R \preceq \eta L$ includes information about $\ker L$.

All big-O terms in Hilbert or L^r spaces refer to norm bounds with constants uniform in the small parameter on a specified compact one-sided interval. No assertion is uniform in collision order m, source support, or dimension unless so stated. Factorial constants in the leading terms are exact; constants bounding the remainders generally are not dimension- or order-free.

B Weighted spaces and minimum-energy velocities

B.1 The gradient closure

Fix a probability measure $\mu = p\,\mathrm{d}x$. Let $\mathscr{D}$ denote smooth admissible test functions: compactly supported smooth functions on $\mathbb{R}^d$, smooth periodic functions on a torus, or smooth functions on a bounded domain with the natural no-flux weak formulation. Define

$$\mathcal{H}_\mu = \overline{\{\nabla\psi : \psi \in \mathscr{D}\}}^{L^2(\mu;\mathbb{R}^d)}.$$

The closure is a Hilbert space. Therefore every $u \in L^2(\mu;\mathbb{R}^d)$ admits a unique orthogonal decomposition $u = P_\mu u + (I - P_\mu)u$. Two velocities u, v represent the same distributional source when

$$\int \nabla\psi \cdot (u - v)\,\mathrm{d}\mu = 0 \quad \text{for every } \psi \in \mathscr{D}.$$

This is the weak meaning of $\mathrm{div}(\mu(u - v)) = 0$. On a bounded domain the condition includes the boundary flux, not just the divergence in the interior.

Lemma B.1 (Least kinetic energy)**.** *For any $u \in L^2(\mu;\mathbb{R}^d)$, $P_\mu u$ is the unique minimum-norm velocity representing the same source as u. Every other representative is $P_\mu u + w$ with $w \in \mathcal{H}_\mu^\perp$, and*

$$\|P_\mu u + w\|^2_{L^2(\mu)} = \|P_\mu u\|^2_{L^2(\mu)} + \|w\|^2_{L^2(\mu)}.$$

Proof. Equality of sources is equivalent to orthogonality of the difference to every test gradient. By closure this is equivalent to orthogonality to $\mathcal{H}_\mu$. Orthogonal projection gives precisely the affine set of representatives just displayed, and Pythagoras proves both minimality and uniqueness. □

This elementary Hilbert-space statement is the local object needed in the paper. Its relation to the velocity of an absolutely continuous Wasserstein curve is standard (Ambrosio et al., 2008). We use the metric interpretation only for curves for which the weak continuity equation and finite kinetic energy are available. We do not posit a globally smooth infinite-dimensional manifold of all probability measures.

B.2 Potentials modulo constants

When a Poincaré inequality is available, it is useful to represent gradients by potentials. Let

$$\mathcal{V}_\mu = \left\{ \psi \in H^1(\mu) : \int \psi \,\mathrm{d}\mu = 0 \right\}, \qquad \|\psi\|^2_{\mathcal{V}_\mu} = a_\mu(\psi,\psi) = \int \|\nabla\psi\|^2 \,\mathrm{d}\mu.$$

We use the completion of smooth mean-zero functions under this norm when a classical weighted Sobolev description is unnecessary. Poincaré ensures a continuous embedding into $L^2(\mu)$ and makes the completion unambiguous modulo constants. Explicitly,

$$\|\psi\|_{L^2(\mu)} \le \sqrt{C_{\mathrm{P}}}\|\psi\|_{\mathcal{V}_\mu}. \tag{B.1}$$

The map $\psi \mapsto \nabla\psi$ is an isometry onto $\mathcal{H}_\mu$. If ℓ is a bounded linear functional on $\mathcal{V}_\mu$, the Riesz representation theorem gives a unique $\phi \in \mathcal{V}_\mu$ with

$$a_\mu(\phi,\psi) = \ell(\psi) \quad \text{for every } \psi \in \mathcal{V}_\mu.$$

For $r \in L^2(\mu)$ of zero mean, the functional $\ell_r(\psi) = \int r\psi \,\mathrm{d}\mu$ is bounded by (B.1). Its Riesz norm satisfies

$$\|\ell_r\|_{\mathcal{V}_\mu^*} \le \sqrt{C_{\mathrm{P}}}\|r\|_{L^2(\mu)}. \tag{B.2}$$

This is the only analytic inequality behind the strong-residual majorant. It is important that its constant is valid for the actual current measure.

Lemma B.2 (Energy identity for a source)**.** *Suppose $\ell(\psi) = \int u \cdot \nabla\psi \,\mathrm{d}\mu$ with $u \in L^2(\mu)$. Then its Riesz potential satisfies $\nabla\phi = P_\mu u$ and*

$$\|\phi\|_{\mathcal{V}_\mu} = \|\ell\|_{\mathcal{V}_\mu^*} = \|P_\mu u\|_{L^2(\mu)}.$$

Proof. The defining identity says that $u - \nabla\phi$ is orthogonal to every gradient. Thus the first assertion follows from Lemma B.1. The equality of the first two norms is the Riesz isometry, and the third is the definition of the Dirichlet norm. □

B.3 Boundary terms and strong residuals

For smooth $p, \widetilde{\phi}$, integration by parts gives

$$\int p\nabla\widetilde{\phi}\cdot\nabla\psi \,\mathrm{d}x = -\int \mathrm{div}(p\nabla\widetilde{\phi})\psi \,\mathrm{d}x + \int_{\partial\Omega} p\partial_n\widetilde{\phi}\,\psi \,\mathrm{d}S.$$

Thus the strong residual in the main paper represents the full weak residual only when the last term vanishes, agrees with the prescribed source flux, or is added to the residual functional and bounded separately. An interior L^2 residual by itself cannot certify a wrong boundary condition. In the periodic examples there is no boundary term; in the whole-space Gaussian examples the polynomial growth and Gaussian tails justify the integration by parts by cutoff approximation.

If no strong residual is available, one can use any reliable majorant for the *dual* residual Gram. For example, an equilibrated flux z_i with $\ell_i(\psi) = \int z_i \cdot \nabla\psi \,\mathrm{d}\mu$ yields

$$\|\ell_i - a_\mu(\widetilde{\phi}_i,\cdot)\|_{\mathcal{V}_\mu^*} \le \|z_i - \nabla\widetilde{\phi}_i\|_{L^2(\mu)}.$$

Applying this to every linear combination yields a matrix majorant. The flux must represent the source exactly; an arbitrary latent velocity with an uncontrolled flux error does not meet this requirement.

B.4 A USEFUL COMPARISON OF POINCARÉ CONSTANTS

Lemma B.3 (Bounded density change)**.** *Let ν satisfy Poincaré with constant C_ν, and suppose $0 < \underline{c} \le d\mu/d\nu \le \overline{c} < \infty$. Then μ satisfies Poincaré with constant at most $(\overline{c}/\underline{c})C_\nu$.*

Proof. Using the variational characterization of variance,

$$\operatorname{Var}_\mu \psi = \inf_c \int (\psi - c)^2 \,\mathrm{d}\mu \le \overline{c} \int (\psi - \mathbb{E}_\nu \psi)^2 \,\mathrm{d}\nu \le \overline{c} C_\nu \int \|\nabla \psi\|^2 \,\mathrm{d}\nu \le \frac{\overline{c}}{\underline{c}} C_\nu \int \|\nabla \psi\|^2 \,\mathrm{d}\mu.$$

The argument first applies to smooth functions and passes to the completion. □

The lemma supplies an explicit certificate constant for bounded density-perturbation and compact-domain examples. It is not a universal bound for separated mixtures. Even when the true tangent dimension stays fixed, a Poincaré constant may deteriorate along a family; any residual certificate based on it must reflect that deterioration.

C DIFFERENTIATING A GENERATOR AND SEPARATING PROJECTION ERRORS

Let $(\mathcal{Z}, \rho)$ be a probability space, $\Theta \subset \mathbb{R}^P$ open, and $T_\theta \in L^2(\rho; \mathbb{R}^d)$. At a fixed θ, suppose T is Fréchet differentiable as an L^2-valued map. Its derivative W_θ is a bounded linear map from $\mathbb{R}^P$ to $L^2(\rho; \mathbb{R}^d)$. Finite parameter dimension is used here; boundedness is automatic once the derivative exists.

C.1 THE WEAK DERIVATIVE OF THE OUTPUT LAW

Lemma C.1 (Pushforward differentiation)**.** *If $\psi \in C_b^2(\mathbb{R}^d)$, then for every $a \in \mathbb{R}^P$,*

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \int \psi \,\mathrm{d}\mu_{\theta+ta} = \mathbb{E}\big[\nabla\psi(T_\theta) \cdot W_\theta a\big].$$

Proof. Write $\Delta_t = T_{\theta+ta} - T_\theta = tW_\theta a + o_{L^2}(t)$. Taylor's theorem with bounded Hessian gives

$$|\psi(T_\theta + \Delta_t) - \psi(T_\theta) - \nabla\psi(T_\theta) \cdot \Delta_t| \le \tfrac{1}{2}\|D^2\psi\|_\infty \|\Delta_t\|^2.$$

After expectation and division by t, the right-hand side tends to zero. The linear term converges by Cauchy–Schwarz and the L^2 derivative. This proves the identity without pointwise differentiation of a density. □

Conditional expectation with respect to T_θ is an orthogonal projection from $L^2(\rho)$ onto its $\sigma(T_\theta)$-measurable subspace. There is a representative $M_\theta a \in L^2(\mu_\theta)$ such that $M_\theta a(T_\theta) = \mathbb{E}[W_\theta a \mid T_\theta]$. The representative is unique μ_θ-almost everywhere and is linear in a after identifying equivalent functions. The source functional is therefore

$$\ell_a(\psi) = \int \nabla\psi \cdot M_\theta a \,\mathrm{d}\mu_\theta.$$

By Lemma B.1, its minimum-energy velocity is $A_\theta a = P_{\mu_\theta} M_\theta a$. This proves that the definition in (2.1) uses the output law, not a particular coupling retained by the latent representation.

C.2 THE EXACT METRIC DISCREPANCY

Proposition C.2 (Two orthogonal losses)**.** *Let $\overline{G} = W^*W$ and $G = A^*A$. Define*

$$Da = Wa - (Ma) \circ T, \qquad Ca = (I - P_\mu)Ma.$$

Then $\overline{G} = G + C^*C + D^*D$. In particular, $\overline{G} \succeq G$. Equality holds exactly when $Da = 0$ and $Ca = 0$ for all parameter directions a.

Proof. Conditional expectation gives $\|Wa\|^2 = \|Ma\|^2_{L^2(\mu)} + \|Da\|^2$. Projection onto $\mathcal{H}_\mu$ gives $\|Ma\|^2 = \|Aa\|^2 + \|Ca\|^2$. These are identities of quadratic forms for every a. Polarization yields the stated matrix identity. A sum of two nonnegative quadratic forms vanishes identically exactly when both operators vanish. □

The first discarded quantity is conditional variance: the latent label can move differently at the same output point. The second is the energy of a density-preserving circulation. Neither is numerical error. Resolving a Poisson equation more accurately cannot remove them from a common-noise matrix unless the model or the metric definition itself is changed.

C.3 Energy derivatives and projected gradient flows

Assume that $\mathcal{F}$ has a first variation whose gradient $\xi_\mu \in \mathcal{H}_\mu$ satisfies the chain rule

$$D(\mathcal{F} \circ \mu)_\theta[a] = \langle \xi_\mu, A_\theta a \rangle_{L^2(\mu)}. \tag{C.1}$$

This chain rule is an assumption on the energy and the model curve. It holds for smooth potential energies by Lemma C.1; entropy and interaction energies require their own integrability. We do not infer it merely from a small empirical training loss.

Proposition C.3 (Induced natural velocity)**.** *Let $g = A^*\xi_\mu$. Then $g \in \operatorname{ran} G$, and*

$$-AG^\dagger g = -P_{\operatorname{ran} A}\xi_\mu.$$

Consequently an exact natural-gradient curve, whenever it exists and obeys the chain rule, dissipates energy at rate

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mu_{\theta_t}) = -\|P_{\operatorname{ran} A}\xi_\mu\|^2_{L^2(\mu)}.$$

Proof. $\ker G = \ker A$, so $\operatorname{ran} G = (\ker A)^\perp = \operatorname{ran} A^*$ in finite dimension. Write a thin singular value decomposition $A = U\Sigma V^*$ on its nonzero singular subspace. Then $AG^\dagger A^* = UU^*$, the orthogonal projection onto $\operatorname{ran} A$. Pairing the negative projected velocity with ξ_μ proves the dissipation identity. □

The projected flow need not be the full Wasserstein gradient flow. The omitted normal component $(I - P_{\operatorname{ran} A})\xi_\mu$ is a *model approximation* error, whereas the residual certificate controls the error in realizing the projected direction. Appendix S keeps these contributions separate in a finite-time comparison.

C.4 A concrete strictly excessive metric

Let $T_B(Z) = BZ$ with $Z \sim \mathcal{N}(0, I_k)$ and B full row rank. A skew-symmetric $k \times k$ matrix Ω gives a parameter direction $\dot{B} = B\Omega$. The covariance derivative is

$$\dot{\Sigma} = B\Omega B^T + B\Omega^T B^T = 0.$$

Hence the output distribution derivative is zero and the true metric assigns zero to $\dot{B}$. But the common-noise energy is $\mathbb{E}\|B\Omega Z\|^2 = \|B\Omega\|^2_F$, which is generally positive. This is already a counterexample to treating $\mathbb{E}[DT^T DT]$ as the true pullback in a smooth, non-colliding, fully Gaussian family. The general covariance formula is proved in Appendix R.

D Regular quotient geometry and singular Gram matrices

The word quotient is used locally and geometrically. We do not assume that every orbit space of a generator is a smooth manifold, or that identifying equal distributions automatically produces a Hausdorff smooth quotient. A supplied local submersion is the correct hypothesis for the finite-step results.

D.1 THE SUBMERSION FORMULATION

Definition D.1 (Regular distribution quotient)**.** A regular quotient representation consists of a smooth submersion $\pi : \Theta \to \mathcal{X}$, a smooth distribution model $\iota : x \mapsto \mu_x$, and a positive-definite pullback metric G_x on $T_x\mathcal{X}$, such that $\mu_\theta = \iota(\pi(\theta))$. The dimension $r = \dim \mathcal{X}$ is fixed on the neighborhood under consideration. The local distribution parameterization is injective, and its tangent map into the gradient-velocity space is injective.

Let $J = D\pi_\theta$. The tangent map factors as $A_\theta = A_x J$, whence

$$G_\theta = J^T G_x J.$$

Since G_x is positive definite and J is onto, $\ker G_\theta = \ker J$. This is the vertical tangent space. Choosing a Euclidean complement $H_\theta = (\ker J)^\perp$ gives a horizontal space, but this complement is not canonical under raw reparameterization. Any smooth complement works for the compatible-lift construction; the resulting distribution step is the object that is invariant.

Lemma D.2 (A rectangular pseudoinverse identity)**.** *If $J : \mathbb{R}^P \to \mathbb{R}^r$ is onto and S is positive definite on $\mathbb{R}^r$, then*

$$J(J^T S J)^\dagger J^T = S^{-1}. \tag{D.1}$$

Proof. Set $B = S^{1/2}J$, which has full row rank. A thin singular value decomposition $B = U\Sigma V^T$ gives $B(B^TB)^\dagger B^T = U\Sigma\Sigma^{-2}\Sigma U^T = I_r$. Multiply by $S^{-1/2}$ on both sides. The same calculation proves that the left side is unaffected by adding parameter columns that do not change the represented quotient tangent space. □

If $f = F \circ \pi$, its differential is $g_\theta = J^T g_x$. Applying Lemma D.2 proves

$$J[-G_\theta^\dagger g_\theta] = -G_x^{-1} g_x.$$

The selected raw vector can change when Euclidean coordinates on Θ change. Only its image under J is fixed. This is precisely why a claim of invariance must state which space contains the invariant object (van Oostrum et al., 2023).

D.2 THE QUOTIENT NORM AND ITS DUAL

The seminorm $\|v\|_{G_\theta}$ becomes a norm on $\mathbb{R}^P / \ker J$. The map $[v] \mapsto Jv$ is an isometric isomorphism to $(T_x\mathcal{X}, G_x)$. A covector on the quotient corresponds to a raw covector g_θ annihilating $\ker J$. Its dual norm is

$$\|g_\theta\|^2_{G_\theta^\dagger} = g_\theta^T G_\theta^\dagger g_\theta = g_x^T G_x^{-1} g_x.$$

Thus an intrinsic error condition is meaningful even when the raw Gram matrix is arbitrarily ill-conditioned or exactly singular. It does not assert that arbitrary floating-point raw linear algebra will be stable; the implementation must use a known quotient or a numerically controlled horizontal representation.

Lemma D.3 (Covectors of distribution objectives)**.** *If f is constant on local fibers of a submersion π, then $Df_\theta[v] = 0$ for $v \in \ker J$, and $Df_\theta \in \operatorname{ran} G_\theta$ after identifying covectors with Euclidean vectors.*

Proof. A vertical tangent vector is the derivative of a curve inside the fiber in submersion coordinates. The derivative of a fiber-constant function along that curve is zero. The orthogonal complement of $\ker J$ equals $\operatorname{ran} J^T = \operatorname{ran} G_\theta$. □

An arbitrary noisy raw gradient need not have this property. Projecting it onto an estimated kernel complement is not automatically equivalent to estimating the gradient of a distribution objective. Our perturbed-gradient theorem instead assumes a covector on the known quotient, and then pulls it back exactly.

D.3 What constant rank does and does not imply

In finite-dimensional Euclidean target spaces, the constant-rank theorem gives local normal forms for a smooth map of constant rank. A density map typically takes values in a function space; obtaining a finite-dimensional embedded image requires an appropriate split constant-rank theorem or a supplied parameterization. We avoid this hidden functional-analytic assumption by formulating (Q) directly.

Even in one dimension, changing rank is qualitatively different. Let $\mu_z = \mathcal{N}(z, 1)$ and parameterize it by $z = a^3$. The distribution image is a perfectly regular line and the z metric is one, but at $a = 0$ the differential of the raw map is zero. Formula (D.1) cannot be invoked there because $J = 3a^2$ is not onto. For the objective $F(z) = \frac{1}{2}(z - z_*)^2$ with $z_* \neq 0$, $Df(0) = 0$ while $DF(0) \neq 0$. Any update of the form $-B(a)Df(a)$ with a finite linear map $B(0)$ stays at zero at this point. The Moore–Penrose inverse and a fixed ridge are included in that statement; injected perturbations and higher-order oracles are not.

D.4 A local existence statement for the model differential

On a regular quotient neighborhood, suppose the potentials of coordinate sources depend continuously on x in the relevant Dirichlet spaces, after transporting them to a fixed reference representation. Then G_x is continuous. Positive definiteness at x_0 implies lower and upper spectral bounds on a sufficiently small compact chart neighborhood. If the dependence is C^k, so is the metric, and usual finite-dimensional Riemannian constructions apply. These are finite-dimensional consequences after the analytic dependence is proved or assumed; they are not uniform statements over all positive densities.

For the exponential family on a torus constructed in Appendix R, uniform ellipticity, smooth bounded coefficients, and independence of the sufficient statistics provide this regularity. For one-dimensional quantile models it follows directly from continuity of the L^2 quantile derivatives. For general implicit neural generators the conditions remain assumptions, not a theorem delivered by the present work.

E Complete proof of the metric certificate

We prove Theorem 3.1 in a form allowing arbitrary trial potentials and a general residual majorant. Fix μ, its Dirichlet space $\mathcal{V}_\mu$, and source functionals $\ell_i \in \mathcal{V}_\mu^*$, $1 \leq i \leq P$. Let ϕ_i be their Riesz potentials. Write $\Phi a = \sum_i a_i \phi_i$, so $G_{ij} = a_\mu(\phi_i, \phi_j)$.

E.1 The exact lower-matrix identity

Take arbitrary $\widetilde{\phi}_i \in \mathcal{V}_\mu$ and let $e_i = \phi_i - \widetilde{\phi}_i$. Define

$$L_{ij} = \ell_i(\widetilde{\phi}_j) + \ell_j(\widetilde{\phi}_i) - a_\mu(\widetilde{\phi}_i, \widetilde{\phi}_j).$$

By symmetry and the Riesz equations,

$$\begin{aligned} G_{ij} - L_{ij} &= a_\mu(\phi_i, \phi_j) - a_\mu(\phi_i, \widetilde{\phi}_j) - a_\mu(\widetilde{\phi}_i, \phi_j) + a_\mu(\widetilde{\phi}_i, \widetilde{\phi}_j) \\ &= a_\mu(e_i, e_j). \end{aligned}$$

Therefore

$$G = L + E, \qquad E_{ij} = a_\mu(e_i, e_j), \qquad E \succeq 0. \tag{E.1}$$

The identity holds without Galerkin orthogonality. The lower matrix L can be indefinite for poor trial potentials. Its positivity is a separate check, not a consequence of the formula.

For each i, define the weak residual $\mathfrak{r}_i = \ell_i - a_\mu(\widetilde{\phi}_i, \cdot)$. The Riesz representative of $\mathfrak{r}_i$ is e_i, hence

$$E_{ij} = \langle \mathfrak{r}_i, \mathfrak{r}_j \rangle_{\mathcal{V}_\mu^*}.$$

This is the exact residual Gram matrix. Any rigorously established matrix $M \succeq E$ gives $L \preceq G \preceq L + M$. The strong-residual version takes $M = C_{\mathrm{P}} R$.

E.2 THE STRONG-RESIDUAL MAJORANT

Suppose $\mathfrak{r}_i(\psi) = \int r_i \psi \, \mathrm{d}\mu$ with $r_i \in L^2(\mu)$ and $\mathbb{E}_\mu r_i = 0$. For every coefficient vector a,

$$a^T E a = \|\mathfrak{r}_a\|^2_{\mathcal{V}^*_\mu} \le C_{\mathrm{P}} \|r_a\|^2_{L^2(\mu)} = C_{\mathrm{P}} a^T R a.$$

This proves $E \preceq C_{\mathrm{P}} R$. Bounding each diagonal separately would yield only a cruder trace bound and would not justify the sharp generalized-eigenvalue test used in the algorithm.

Lemma E.1 (Kernel certification)**.** *If $L \succeq 0$ and $0 \preceq G - L \preceq M \preceq \eta L$, then* $\ker L = \ker G$. *Conversely, a finite inequality $M \preceq \eta L$ with $M \succeq 0$ requires* $\ker L \subseteq \ker M$.

Proof. For $v \in \ker L$, $0 \le v^T G v \le (1+\eta) v^T L v = 0$, so $Gv = 0$. If $Gv = 0$, $0 \le v^T L v \le v^T G v = 0$, hence $Lv = 0$. For the final statement, $v^T M v = 0$ on $\ker L$, and a positive-semidefinite matrix with zero quadratic form on a vector annihilates that vector. □

Accordingly, the computable exact-arithmetic test consists of two parts: verify $\ker L \subseteq \ker M$ and bound the largest generalized eigenvalue on the quotient. Replacing the first part by a numerical tolerance changes the logical nature of the guarantee. With finite-sample validation we work on a known quotient and do not try to infer an exact kernel from noisy data.

E.3 DISTRIBUTIONAL STEP CONTROL ON THE SINGULAR SPACE

Let $H = (\ker G)^\perp$. Both G and L are positive definite on H and vanish on its orthogonal complement. Restrict all inverses below to H. Define the isometry

$$U = AG^{-1/2} : H \to \operatorname{ran} A, \qquad U^* U = I_H.$$

Let $B = G^{1/2} L^{-1} G^{1/2}$. From $L \preceq G \preceq (1+\eta) L$,

$$I_H \preceq B \preceq (1+\eta) I_H.$$

For $z = G^{-1/2} g$, the exact and approximate velocities are $v = -Uz$ and $v_L = -UBz$. Hence

$$\|v_L - v\| = \|(B - I)z\| \le \eta \|z\| = \eta \|v\|.$$

Also $g^T L^{-1} g = z^T B z$, proving the two-sided dissipation bound. This argument does not estimate $\|L^\dagger - G^\dagger\|_{\mathrm{op}}$ in raw Euclidean coordinates, which may be enormous. It estimates the observable velocity after multiplying by the tangent map A.

Corollary E.2 (Speed and descent bounds)**.** *Under the same assumptions,*

$$\|v\|^2 \le \|v_L\|^2 \le (1+\eta)^2 \|v\|^2, \qquad g^T L^\dagger g \ge \frac{\|v_L\|^2}{1+\eta}.$$

Proof. The first inequalities follow from the spectrum of B. For the second, write the raw step $w = -L^\dagger g$, so $\|v_L\|^2 = w^T G w \le (1+\eta) w^T L w = (1+\eta) g^T L^\dagger g$. □

E.4 CONGRUENCES AND REDUNDANT REPARAMETERIZATIONS

Let $C : \mathbb{R}^{\widetilde{P}} \to \mathbb{R}^P$ describe a change of represented tangent directions, and let $\widetilde{A} = AC$. Sources and trial potentials transform by the same linear combinations. Then

$$\widetilde{G} = C^T G C, \qquad \widetilde{L} = C^T L C, \qquad \widetilde{M} = C^T M C.$$

Quadratic-form order is preserved under every such C, whether or not square. If $\operatorname{ran}(AC) = \operatorname{ran} A$, the quotient tangent space is unchanged and the metric solutions induce the same velocity. To see this, choose any coordinates on the common range, apply Lemma D.2 to the two full-row-rank representations, and push both solutions forward. There is no claim that the raw Moore–Penrose solution transforms as an ordinary vector under a nonorthogonal redundant parameter change.

For adaptive trial spaces, this invariance requires that the trial gradient space and validation rules depend on the represented distributional problem, not arbitrarily on parameter labels. Running two unrelated neural potential optimizers with different initializations can produce different certified approximations. Each has the error bound, but their finite steps need not coincide exactly. Exact invariance refers to corresponding congruent certificate data and the same quotient retraction.

E.5 Galerkin matrices and an exact completion of squares

Let $\psi_1, \dots, \psi_k$ span a finite-dimensional trial space. Define $K_{ab} = a_\mu(\psi_a, \psi_b)$ and $B_{ai} = \ell_i(\psi_a)$. If the gradients are linearly independent, K is positive definite; otherwise use its quotient by the nullspace, on which B vanishes. Exact Galerkin coefficients are $U_* = K^\dagger B$. Their metric is

$$G_V = B^T K^\dagger B.$$

For arbitrary coefficients U, substitution into the lower matrix gives

$$L(U) = B^T U + U^T B - U^T K U = G_V - (U - U_*)^T K (U - U_*). \tag{E.2}$$

Thus solve error decreases the variational lower matrix by an exactly known nonnegative form when K, B are exact. This explains why a highly inaccurate solve cannot be justified by merely observing that $U^T K U$ is positive definite. The latter is not a lower bound for the desired metric.

E.6 Finite-dimensional enrichment

Suppose trial gradient spaces V_n are nested and dense in $\mathcal{H}_\mu$. Then $A_n = P_{V_n} A$ converges to A in operator norm because its domain is finite dimensional. Indeed, with a Euclidean basis e_i,

$$\|(I - P_{V_n})A\|_{\mathrm{op}}^2 \le \sum_{i=1}^{P} \|(I - P_{V_n})Ae_i\|^2 \longrightarrow 0.$$

On the intrinsic quotient the same argument applies after whitening by G. Eventually A_n is injective on that quotient and $G_n \to G$ relatively. This proves eventual recoverability by exact Galerkin projection. It does *not* by itself prove that a chosen strong-residual majorant tends to zero: that requires approximation in an operator graph norm. The adaptive termination statement below explicitly adds that condition.

F Inexact solves, validated integrals, and finite-sample certificates

The preceding theorem is an exact statement about integrals. This section specifies how approximations can enter without turning a numerical diagnostic into an unjustified certificate. We work in an already identified r-dimensional quotient. Exact gauge directions can subsequently be reintroduced by a congruence.

F.1 Certified matrix intervals

Let $S \succ 0$ be a reference quadratic form on the quotient. Suppose numerical procedures return symmetric matrices $\widehat{L}, \widehat{R}$ and valid bounds

$$\|S^{-1/2}(\widehat{L} - L)S^{-1/2}\|_{\mathrm{op}} \le \varepsilon_L, \qquad \|S^{-1/2}(\widehat{R} - R)S^{-1/2}\|_{\mathrm{op}} \le \varepsilon_R. \tag{F.1}$$

The coefficients of the trial potentials can be inexact: L here is always the symmetrized variational matrix for the actual trial functions. Define

$$L_- := \widehat{L} - \varepsilon_L S, \qquad U_+ := \widehat{L} + \varepsilon_L S + C_{\mathrm{P}}(\widehat{R} + \varepsilon_R S). \tag{F.2}$$

Then $L_- \preceq G \preceq U_+$. If $L_- \succ 0$ and $U_+ \preceq (1+\eta)L_-$, Theorem 3.1 applies with L_-. This is a fully computable test once the integral error bounds and C_{P} are valid. In particular, it does not use an unknown smallest eigenvalue of G.

If the reference form is pulled back to redundant parameters, it becomes $S_\theta = J^T S J$. This matrix has the exact gauge kernel. Replacing it by an identity matrix in raw coordinates would enlarge the interval in vertical directions and generally destroy kernel certification. The need to preserve the gauge is not a cosmetic matter: an exact kernel cannot be reconstructed from an unconstrained absolute noise ball.

F.2 Inexact linear algebra

Suppose the trial matrix K and sources B are evaluated exactly but $KU = B$ is only approximately solved. Formula (E.2) gives a lower matrix for any U. Its strong residual is obtained from the actual functions $\sum_a U_{ai}\psi_a$, not from the residual of the finite linear system alone. The finite-system residual controls the error inside the trial space; it does not control the missing infinite-dimensional component. Both can be bounded by the full variational residual, or separately and then added as quadratic forms.

For example, if $K \succ 0$ and $E_K = B - KU$, then

$$G_V - L(U) = E_K^T K^{-1} E_K.$$

An upper bound $K^{-1} \preceq \tau S_K^{-1}$ supplies an explicit algebraic-solve majorant. It can be added to a discretization certificate, but it must not be counted twice if the full strong residual is already being used. This distinction is useful when the trial-space solve is performed iteratively.

F.3 Independent validation of fixed trial potentials

Here is a conservative finite-sample version. It is included as a theorem, not as a claim of statistical efficiency. Transform quotient coordinates so that $S = I$. Conditional on all prior information, suppose the trial potentials, sources, and current measure are fixed. Draw an independent validation batch $X_1, \dots, X_N \sim \mu$. Let $Y_{ij}^L(X)$ and $Y_{ij}^R(X)$ be unbiased integrands for L_{ij} and R_{ij}, and assume

$$|Y_{ij}^L(X)| \le B_L, \qquad |Y_{ij}^R(X)| \le B_R \quad \text{almost surely.}$$

The bounds may depend on the fixed trial functions and current parameter but must be known before viewing this batch.

Proposition F.1 (A finite-batch matrix envelope)**.** *For $\delta \in (0,1)$, put*

$$t_L = B_L\sqrt{\frac{2\log(4r^2/\delta)}{N}}, \qquad t_R = B_R\sqrt{\frac{2\log(4r^2/\delta)}{N}}.$$

With conditional probability at least $1-\delta$,

$$\|\widehat{L} - L\|_{\mathrm{op}} \le r t_L, \qquad \|\widehat{R} - R\|_{\mathrm{op}} \le r t_R.$$

Thus (F.2) *is valid with $\varepsilon_L = r t_L$ and $\varepsilon_R = r t_R$ on that event.*

Proof. For an entry of a bounded variable in $[-B, B]$, Hoeffding's inequality gives $\Pr(|\widehat{Y} - \mathbb{E}Y| > t) \le 2\exp[-Nt^2/(2B^2)]$. Apply the union bound to the $2r^2$ entries. On the resulting event every entrywise error is bounded by its corresponding t. A matrix with all entries bounded in absolute value by t has Frobenius norm at most rt, hence operator norm at most rt. □

This elementary bound avoids any assumption of independence among entries. Sharper matrix concentration can improve dimension factors, but is unnecessary for the logical guarantee. The dependence is on the quotient dimension r, not on arbitrary duplication of raw parameter coordinates. Bounds on the integrands and the number of potential evaluations can still depend on dimension and on the actual family.

Training a potential on the same batch invalidates the conditional-fixed-integrand argument. One must either use a fresh validation batch, establish uniform concentration over the fitted class, or account for dependence in another valid way. Likewise, density or score estimates inserted into the residual require additional validated error terms. The proposition does not convert black-box generator samples into an exact-density oracle.

F.4 Adaptive reuse and failure probabilities

Suppose validation attempt j is based on all previous attempts, followed by a fresh conditional validation batch, and is assigned failure probability δ_j. If $\sum_j \delta_j \le \delta$, then every accepted matrix interval is simultaneously valid with probability at least $1 - \delta$. Indeed, conditional validity gives

unconditional failure probability at most δ_j by the tower property; the union bound applies even when the attempted potentials and stopping time are adaptive. A permissible schedule is $\delta_j = 6\delta/[\pi^2(j+1)^2]$ for $j \geq 0$.

If instead one repeatedly inspects a single batch until a certificate appears favorable, the preceding argument does not apply. A time-uniform concentration bound would be required. Our algorithm specifies the fresh-batch route because it has a direct proof.

F.5 Gradient and energy validation

Let $\widehat{g}$ be a quotient covector with $\|\widehat{g} - g\|_{S^{-1}} \leq e_0$. If $\lambda S \preceq G \preceq \beta S$, then

$$\|\widehat{g} - g\|_{G^{-1}} \leq e_0/\sqrt{\lambda}, \qquad \|g\|_{G^{-1}} \geq (\|\widehat{g}\|_{S^{-1}} - e_0)/\sqrt{\beta}.$$

Thus the relative error test in Appendix I can itself be certified whenever the right-hand lower bound is positive. Close to stationarity, fixed absolute gradient uncertainty may prevent any relative certificate. The correct consequence is an error floor or a demand for a larger validation batch, not an unsupported claim of exact convergence.

For line search, suppose $f_-(x) \leq f(x) \leq f_+(x)$ and $d_0 > 0$ is a certified lower bound for $-Df_x[w]$. The acceptance condition

$$f_+(\mathrm{Retr}_x(\alpha w)) \leq f_-(x) - c\alpha d_0$$

guarantees genuine descent. Finite intervals can make a good step impossible to certify. Termination of such a line search therefore additionally requires refinement of energy and derivative bounds relative to the predicted decrease. This is separate from the exact-energy line-search theorem.

G A complete residual-certified quotient algorithm

The algorithm is best stated in terms of its oracles. Otherwise the word "certified" can conceal the hard part of evaluating a true Wasserstein metric. The regular mode uses a supplied quotient chart. The collision mode is enabled only for an explicitly constructed jet chart, such as those in Appendices K–M.

G.1 Required inputs

At a current quotient point x, the regular mode receives the following objects: sources for the coordinate density derivatives; a valid Poincaré constant; a sequence of admissible trial-potential spaces; reliable matrix intervals for the associated lower and residual matrices; a differential of the energy, exact or with an intrinsic error bound; and a quotient retraction. A horizontal local section is needed only when an updated raw generator parameter must be returned. The stopping tolerance concerns the certified quotient differential, not a raw Euclidean gradient that can vanish under a singular parameterization.

Algorithm 1: Residual-certified quotient natural gradient (RCQ-NG).
Inputs: current quotient point x, tolerance $\eta_0 > 0$, energy and source oracles, retraction, and validation/refinement rules.
1. For refinement index k, solve for admissible trial potentials. Form a lower matrix L_- and upper matrix U_+ using the actual trial functions and validated integrals.
2. Accept the metric only if $L_- \succ 0$ on the known quotient and $U_+ \preceq (1+\eta_0)L_-$. In a redundant representation verify the exact gauge kernel first and make this test on its quotient. Otherwise refine; at a fixed budget return `unresolved`, not `stationary`.
3. Compute $w = -L_-^{-1} g$ in quotient coordinates. With an inexact gradient use the certified relative-error test and the constants in Theorem I.1.
4. Choose an admissible step by the stated backtracking condition, and set $x^+ = \mathrm{Retr}_x(\alpha w)$. Return a compatible horizontal lift when raw parameters are requested.
5. If a verified rank-changing jet chart is supplied at the current distribution, replace the raw singular chart by that chart and apply its first-derivative convergence theorem. Do not infer a valid new chart solely from a failed metric test.

G.2 WHY EACH TEST IS NECESSARY

Positivity of L_- is not a raw eigenvalue threshold. It is checked after an identifiable tangent space has been specified, and the relative upper bound rules out unresolved directions. A tiny positive eigenvalue caused by raw coordinate scaling does not affect the congruence-invariant inequality. Conversely, if a genuine quotient eigenvalue is small, the residual majorant must resolve it to relative accuracy. The theorem does not promise that this will be cheap.

The distinction between an inaccurate trial space and a rank-changing generator cannot be decided from L_- alone. In the former case the true differential is nonzero but the test potentials do not see it; in the latter case the true first differential is zero and only higher-order information reveals an additional nearby distribution direction. Both may produce a zero trial Gram matrix. Only a model-level argument justifies switching to a new chart.

G.3 A SUFFICIENT REFINEMENT TERMINATION THEOREM

Let $\mathcal{L}_\mu \phi = -p^{-1} \operatorname{div}(p \nabla \phi)$ be the mean-zero weighted elliptic operator. Assume its domain is dense in the Dirichlet space and that the actual exact potentials ϕ_i belong to this operator domain.

Proposition G.1 (Graph-norm termination)**.** *At a fixed regular quotient point, suppose $G \succ 0$, $C_{\mathrm{P}} < \infty$, and trial potentials satisfy*

$$\|\phi_i - \widetilde{\phi}_{i,k}\|_{\mathcal{V}_\mu} \longrightarrow 0, \qquad \|\mathcal{L}_\mu(\phi_i - \widetilde{\phi}_{i,k})\|_{L^2(\mu)} \longrightarrow 0$$

for every i. Suppose also that the matrix-integration envelopes tend to zero. Then for every $\eta_0 > 0$, the metric test in Algorithm 1 passes for all sufficiently large k.

Proof. The exact identity gives $L_k \to G$ in operator norm, because there are finitely many potentials and their Dirichlet errors tend to zero. The strong residual Gram R_k tends to zero by the second assumption and Cauchy–Schwarz. The numerical lower and upper matrices therefore converge to the same positive-definite matrix G. Eventually $L_{-,k} \succeq \frac{1}{2}\lambda_{\min}(G) I$, while $\|U_{+,k} - L_{-,k}\|_{\mathrm{op}} \leq \eta_0 \lambda_{\min}(G)/2$. This implies the desired relative inequality. The unknown eigenvalue appears in this existence proof, not in the algorithmic test. □

The proposition is pointwise. Uniform termination over a compact quotient set requires uniform graph approximation, a uniformly bounded Poincaré constant, and a uniform intrinsic spectral gap. No raw-parameter spectral gap is required. Mere density of trial gradients in the Dirichlet norm is insufficient for the displayed strong-residual assumption; this is why the proposition states it separately.

G.4 AN ANALYTIC FAMILY FOR WHICH THE ASSUMPTIONS ARE NONEMPTY

On a flat torus, take smooth bounded sufficient statistics and a compact set of exponential-family parameters, as in Appendix R. The densities are uniformly bounded above and below, and all coefficient derivatives are bounded. Elliptic regularity gives smooth coordinate potentials. Trigonometric polynomial approximations converge in every fixed Sobolev norm for smooth potentials, so choosing approximate potentials converging in H^2 yields strong-residual convergence. Galerkin approximations with the corresponding regularity and approximation estimates provide one implementation; the termination proof does not depend on claiming every arbitrary neural potential optimizer reaches these approximants. Standard elliptic regularity is used only for this example and is identified as background (Evans, 2010).

G.5 COST AND WHAT IS NOT HIDDEN IN THE THEOREM

For r quotient directions and k trial potentials, forming a dense Dirichlet matrix from N integration points requires $O(Nk^2)$ pair products, source assembly $O(Nkr)$, and a dense factorization $O(k^3)$ plus $O(k^2 r)$ for multiple right-hand sides. The lower matrix and residual Gram add $O(kr^2)$ and $O(Nr^2)$ products, while the relative eigenvalue test costs $O(r^3)$. These counts exclude evaluating densities, scores, sources, trial derivatives, and the energy gradient itself. They are therefore algebraic operation counts, not end-to-end network training complexity.

The paired-product model replaces a single high-dimensional elliptic solve by lower-dimensional ones, but their number grows quadratically with the number of active coordinates. The Gaussian quadrature collision construction has increasingly high order and factorial normalization. Neither result says that arbitrary high-dimensional distributions become easy to fit.

G.6 A FORMAL TERMINATION VERSUS BUDGET DISTINCTION

If a finite computational budget is imposed, Algorithm 1 may return `unresolved`. This is a valid output rather than a failure of the mathematical definition. The graph-norm theorem states when unlimited refinement eventually succeeds; it does not bound the required k or N without approximation rates. Conversely, once a certificate passes, the subsequent step bound is independent of how the potentials were obtained. This separates *finding* a certificate from *using* it, an important distinction when the potentials are learned by a nonconvex optimization method.

H COMPATIBLE RETRACTIONS AND FINITE-STEP INVARIANCE

Infinitesimal invariance does not automatically imply finite-step invariance. A retraction specifies how a tangent vector is converted into a new point, and different raw additive rules can implement different distribution changes from the same first-order velocity.

H.1 CONSTRUCTION OF A COMPATIBLE LIFT

Let $\pi : \Theta \to \mathcal{X}$ be a smooth submersion, $x = \pi(\theta)$, and choose a horizontal complement H_θ of $\ker D\pi_\theta$. A local section s_θ through θ can be chosen with

$$\pi \circ s_\theta = \mathrm{Id}, \qquad s_\theta(x) = \theta, \qquad Ds_\theta(x) = (D\pi_\theta|_{H_\theta})^{-1}.$$

To justify the last condition, use submersion coordinates in which π is projection onto the first coordinates. A section is the graph of a function of those coordinates, and its derivative can be chosen to have the prescribed complementary graph. Smooth dependence on θ is available on a sufficiently small trivializing neighborhood when the horizontal distribution is smooth.

Given a quotient retraction Retr_x, define for $v \in H_\theta$ near zero

$$\widetilde{\mathrm{Retr}}_\theta(v) = s_\theta(\mathrm{Retr}_x(D\pi_\theta v)). \tag{H.1}$$

It satisfies $\widetilde{\mathrm{Retr}}_\theta(0) = \theta$, $D\widetilde{\mathrm{Retr}}_\theta(0)v = v$ on the horizontal space, and the exact compatibility identity. This is a horizontal retraction. It can be extended to a full parameter retraction in local product coordinates by adding a fiber retraction, but the vertical part is immaterial for the distribution update.

Proposition H.1 (Exact distributional invariance)**.** *Suppose two regular parameter representations describe the same quotient point, use corresponding intrinsic differential and certificate data, use the same quotient retraction, and use the same intrinsic step-selection rule. Their compatible horizontal lifts produce identical output distributions at every iteration for which both local lifts are defined.*

Proof. The induced tangent step agrees by Lemma D.2 and the congruence property of the certificate. The exact compatibility identity then gives the same quotient next point. Energy and intrinsic acceptance quantities agree, so the chosen step size agrees. Induction proves the claim. A change of local section on the next iteration does not change the quotient point. □

The proposition does not claim that the two raw parameter sequences coincide, or that a single local section covers an arbitrarily long trajectory. Sections may be changed as the quotient path moves. Such changes are gauge choices, not additional distributional motion.

H.2 FIRST-ORDER VERSUS SECOND-ORDER RETRACTIONS

In local coordinates, any C^2 retraction satisfies

$$R_x(v) = x + v + O(\|v\|^2).$$

A second-order Riemannian retraction has zero covariant acceleration at the origin; with enough regularity, its difference from the exponential map is $O(\|v\|^3)$. These are different statements. First-order retraction descent only needs a quadratic pullback upper bound. It does not require the stronger second-order condition. Conversely, a C^1 chart at a collision does not automatically supply either a uniform quadratic remainder or a smooth Hessian.

Lemma H.2 (Uniform comparison on a compact regular set)**.** *Let R be C^2 near the zero section over a compact subset of a finite-dimensional Riemannian manifold. There are $r, C > 0$ such that*

$$\operatorname{dist}(R_x(v), \exp_x(v)) \le C\|v\|^2 \quad \textit{for } \|v\| \le r.$$

If R is second order and C^3, the right side can be replaced by $C\|v\|^3$ after shrinking r.

Proof. Choose finitely many smooth coordinate neighborhoods covering the compact set and a common radius on which $\exp_x$ and R_x are defined. The coordinate map $v \mapsto \log_x R_x(v)$ has value zero and derivative identity at zero. Taylor's integral formula gives the quadratic bound with a uniform second-derivative constant. A local uniform Lipschitz bound for $\exp_x$ converts the coordinate remainder to distance. In the second-order case the second derivative at zero vanishes in normal coordinates; the third-order Taylor remainder gives the improvement. Compactness is used to make the radii and derivative bounds uniform. □

Nothing in this proof persists automatically as the chosen regular chart approaches a singular parameter value. Its derivatives, its inverse, or its metric comparison constants may diverge.

H.3 Power-coordinate lifts and their blow-up

Let $z = a^m$ for $a > 0$ and integer $m \ge 2$. The derivative is $Dz[a]w = ma^{m-1}w$. Lifting an additive z-retraction gives

$$R_a(w) = \left(a^m + ma^{m-1}w\right)^{1/m} = a\left(1 + \frac{mw}{a}\right)^{1/m}.$$

Expanding the last factor proves

$$R_a(w) = a + w - \frac{m-1}{2a}w^2 + \frac{(m-1)(2m-1)}{6a^2}w^3 + O_a(w^4). \tag{H.2}$$

In particular, no bound $|R_a(w) - (a+w)| \le Cw^2$ can hold with one constant C for all sufficiently small $a > 0$ and all sufficiently small admissible w. Taking the second derivative at $w = 0$ would require $C \ge (m-1)/(2a)$.

For a chart objective $F(z)$ and chart metric $h(z)$, the raw metric is $m^2a^{2m-2}h(z)$. Its natural velocity is $w = -F'(z)/(ma^{m-1}h(z))$. Substitution into the exact lift produces the chart step $z^+ = z - \alpha F'(z)/h(z)$ whenever it stays in the same branch. The corresponding additive raw update does not. At $a = 0$ this velocity is not defined for a nonzero target differential, and the actual raw pseudoinverse step is zero. Branch continuation is therefore a *change of chart*, not the continuous extension of a bounded raw retraction.

H.4 An exact covariance-compatible lift

For a centered Gaussian with covariance $\Sigma \succ 0$, let a symmetric matrix S solve $S\Sigma + \Sigma S = \dot{\Sigma}$. The Wasserstein geodesic associated with this tangent, while $I + tS \succ 0$, is

$$\Sigma_t = (I + tS)\Sigma(I + tS).$$

If $\Sigma = BB^T$, the parameter update $B_t = (I + tS)B$ realizes this covariance path exactly, regardless of the number of columns of B. This is a special case where a linear raw update is compatible with a chosen intrinsic retraction. Its validity follows from the symmetric transport velocity, not from a generic assumption that every additive parameter step respects quotient geometry. Appendix R gives the underlying Sylvester calculation.

I RETRACTION DESCENT, PERTURBATIONS, AND NEIGHBORHOOD INVARIANCE

This appendix concerns a finite-dimensional regular quotient with a smooth positive-definite metric G. It does not treat a rank-changing raw parameterization as a regular Riemannian manifold. For a covector g, write $\|g\|_*^2 = g^T G^{-1} g$. All matrix comparisons and covector errors below are on the quotient. Their pullbacks to redundant parameters preserve the vertical kernel exactly.

I.1 ONE-SIDED CERTIFIED METRIC ERROR

Let $L \preceq G \preceq (1+\eta)L$, $g = Df$, and $w = -L^{-1}g$. Define $D = g^T L^{-1} g$. Two inequalities suffice:

$$D \geq \|g\|_*^2, \qquad \|w\|_G^2 = w^T G w \leq (1+\eta) w^T L w = (1+\eta) D.$$

Applying the retraction smoothness inequality gives

$$f(\mathrm{Retr}_x(\alpha w)) - f(x) \leq -\alpha D + \frac{L_f \alpha^2}{2}(1+\eta) D.$$

For $\alpha L_f (1+\eta) \leq 1$, the right-hand side is at most $-\alpha D/2$, proving Theorem 4.1. This is stronger than estimating a raw inverse difference: no raw minimum positive eigenvalue occurs.

With Armijo parameter $c_A \in (0,1)$, every admissible step with

$$\alpha \leq \frac{2(1-c_A)}{L_f(1+\eta)}$$

passes $f(\mathrm{Retr}_x(\alpha w)) \leq f(x) + c_A \alpha g[w]$. Backtracking from $\bar{\alpha}$ with reduction factor $\beta \in (0,1)$ therefore terminates whenever a uniform admissible radius is available. Away from a zero gradient, its accepted step is bounded below by β times the minimum of the displayed bound, $\bar{\alpha}$, and the admissibility bound. Uniform convergence rates require that this last bound not collapse along the iterates.

I.2 SIMULTANEOUS METRIC AND COVECTOR ERROR

An integration certificate can produce a two-sided rather than a one-sided comparison. The following result states the precise allowable gradient error.

Theorem I.1 (Relative-error descent)**.** *Assume*

$$(1-\varepsilon) G \preceq \widehat{G} \preceq (1+\varepsilon) G, \qquad \|e\|_* \leq \sigma \|g\|_*, \qquad 0 \leq \varepsilon < 1.$$

Set $\widehat{g} = g + e$, $w = -\widehat{G}^{-1}\widehat{g}$, *and*

$$a_0 = \frac{1}{1+\varepsilon} - \frac{\sigma}{1-\varepsilon}, \qquad b_0 = \frac{1+\sigma}{1-\varepsilon}.$$

If $a_0 > 0$ *and* $0 < \alpha \leq a_0/(L_f b_0^2)$ *is admissible, then*

$$f(\mathrm{Retr}_x(\alpha w)) \leq f(x) - \frac{\alpha a_0}{2}\|g\|_*^2. \tag{I.1}$$

If the local PL inequality $\|g\|_*^2 \geq 2\mu(f - f_*)$ *holds at every iterate,* $0 < \alpha a_0 \mu < 1$, *and the steps stay in its neighborhood, then*

$$f(x_t) - f_* \leq (1 - \alpha a_0 \mu)^t (f(x_0) - f_*).$$

A uniform variable-step lower bound $\alpha_t \geq \underline{\alpha} > 0$ *gives the same conclusion with* α *replaced by* $\underline{\alpha}$.

Proof. Let $z = G^{-1/2} g$, $\epsilon = G^{-1/2} e$, and $B = G^{1/2} \widehat{G}^{-1} G^{1/2}$. Its spectrum lies in $[(1+\varepsilon)^{-1}, (1-\varepsilon)^{-1}]$. Thus

$$-g[w] = z^T B (z + \epsilon) \geq \left((1+\varepsilon)^{-1} - \sigma(1-\varepsilon)^{-1}\right) \|z\|^2,$$

and

$$\|w\|_G = \|B(z+\epsilon)\| \leq b_0 \|z\|.$$

Substitution into retraction smoothness proves $f(\mathrm{Retr}_x(\alpha w)) - f(x) \leq (-\alpha a_0 + L_f \alpha^2 b_0^2/2)\|z\|^2$. The step bound gives (I.1); applying PL and iterating proves the remaining statements. At $g = 0$, the relative error assumption forces $e = 0$, so the update is stationary as it should be. □

The condition $a_0 > 0$, equivalent to $\sigma < (1-\varepsilon)/(1+\varepsilon)$, is not merely technical. If the error is allowed to cancel and reverse the covector, a positive-definite metric cannot restore descent.

For a lower certificate $L \preceq G \preceq (1+\eta)L$, one may either use the sharper one-sided proof or rescale L to a centered approximation. Specifically, $\widehat{G} = (1+\eta/2)L$ satisfies the two-sided comparison with $\varepsilon = \eta/(2+\eta)$. The scaling changes the nominal step length, not its direction.

I.3 What an absolute error bound guarantees

Suppose only $\|e\|_* \le \zeta$. Choose any σ with $a_0 > 0$. Whenever $\|g\|_* \ge \zeta/\sigma$, Theorem I.1 applies. Therefore the method has certified descent outside a small-gradient region, provided step acceptance is checked against the true energy or an adequate interval certificate. This does not imply monotonicity for an unchecked noisy step inside that region.

A useful stopping rule is to stop only when a validated upper bound on $\|g\|_*$ is at most τ. Under PL this implies

$$f(x) - f_* \le \tau^2/(2\mu).$$

If only $\widehat{g}$ is available and $\|e\|_* \le \zeta$, use $\|\widehat{g}\|_* + \zeta$ as the upper bound. Approximate metric comparisons convert its computed dual norm to a true dual-norm bound. This is a controlled accuracy floor, not an exact convergence theorem in the presence of fixed nonzero oracle error.

I.4 Local PL from a nonisolated minimizer manifold

The next lemma supplies verifiable sufficient local assumptions. It is a finite-dimensional Morse–Bott argument, not a claim that all generator losses have this property.

Lemma I.2 (A normal Hessian gap gives local PL)**.** *Let S be a compact embedded submanifold of a smooth Riemannian quotient, and let $f \in C^3$ satisfy $f = f_*$ and $\operatorname{grad} f = 0$ on S. Suppose*

$$\ker \operatorname{Hess} f(y) = T_y S, \quad \lambda\|\nu\|^2 \le \langle \nu, \operatorname{Hess} f(y)\nu\rangle \le \Lambda\|\nu\|^2 \quad (\nu \in N_y S),$$

where $0 < \lambda \le \Lambda < \infty$. After shrinking a tubular neighborhood, with $d = \operatorname{dist}(x, S)$ one has

$$\frac{\lambda}{4}d^2 \le f(x) - f_* \le \Lambda d^2, \qquad \frac{\lambda}{2}d \le \|\operatorname{grad} f(x)\| \le 2\Lambda d.$$

Consequently PL holds with $\mu = \lambda^2/(8\Lambda)$, and $\|\operatorname{grad} f(x)\| \le B\sqrt{f(x) - f_*}$ with $B = 4\Lambda/\sqrt{\lambda}$.

Proof. In a sufficiently small tube, write $x = \exp_y \nu$ with $y \in S$ the unique nearest point and $\|\nu\| = d$. Taylor's theorem along this normal geodesic gives

$$f(x) - f_* = \tfrac{1}{2}\langle \nu, \operatorname{Hess} f(y)\nu\rangle + O(d^3).$$

Parallel transport to y gives

$$\tau_{x\to y} \operatorname{grad} f(x) = \operatorname{Hess} f(y)\nu + O(d^2).$$

Both remainders are uniform by compactness and C^3 regularity. Since the Hessian is self-adjoint and its kernel is $T_y S$, its normal restriction has eigenvalues in $[\lambda, \Lambda]$. The asserted estimates follow by choosing d small enough. The upper energy bound and lower gradient bound imply $\|\operatorname{grad} f\|^2 \ge \lambda^2(f - f_*)/(4\Lambda)$, which is PL with the stated μ. The remaining inequality uses the lower energy bound. □

There is no assumption that an off-manifold gradient is tangent or normal to S. The transported Taylor expansion treats its full vector and is the reason the argument is valid.

I.5 A noncircular trapping argument

A local rate whose proof assumes all iterates remain local is incomplete unless neighborhood invariance is separately supplied. Here is a sufficient condition.

Proposition I.3 (Finite-length trapping)**.** *Suppose the hypotheses of Theorem I.1, PL, and* $\|g\|_* \le B\sqrt{f - f_*}$ *hold in an open set* $\mathcal{U}$*. Suppose the chosen retraction segments satisfy* $\mathrm{length}(s \mapsto \mathrm{Retr}_x(sv),\ 0 \le s \le 1) \le C_R\|v\|_G$ *for the steps under consideration. Fix* α *with* $q = 1 - \alpha a_0 \mu \in (0, 1)$*. Let a closed metric ball of radius* R *about* x_0 *lie in* $\mathcal{U}$*, with uniform step admissibility there. If the initial energy* $E_0 = f(x_0) - f_*$ *satisfies both the admissible-radius bound and*

$$\frac{C_R \alpha b_0 B \sqrt{E_0}}{1 - \sqrt{q}} < R, \tag{I.2}$$

then all iterates stay in that ball, $E_t \le q^t E_0$*, and the sequence has finite total path length. If the closed ball is complete, it converges to a point in* $S = \{f = f_*\}$*.*

Proof. Assume the first t iterates satisfy the local conditions. Their energy bounds imply that step k has length at most $C_R \alpha b_0 B \sqrt{E_0} q^{k/2}$. Summing this bound, including the proposed next segment, gives a length strictly below R by (I.2). Thus that segment cannot be the first one to exit the ball. Uniform admissibility ensures its endpoint is defined. Induction proves all claims about remaining local and the geometric energy estimate. The summable path lengths make the sequence Cauchy. Its limit belongs to the closed ball, and continuity of f gives energy f_*. No uniform uniqueness of the minimizer in S is asserted. □

This proof explains the order of assumptions: a normal gap provides local PL and a gradient upper bound; retraction regularity provides a length bound; a small initial energy and interior starting point provide trapping; only then is a local asymptotic rate applied indefinitely.

I.6 Intrinsic conditioning and coordinate width

The constants $L_f, \lambda, \Lambda, C_R$ describe the quotient geometry and the objective there. Replacing a quotient coordinate by many redundant parameters does not change them when the lift is compatible. Changing the distribution family so that its intrinsic metric or normal Hessian gap degenerates can change them without bound. An algorithmic invariance theorem must preserve this distinction. In particular, the angular eigenvalues in Appendix P approach zero for nearly Gaussian factors; that is not merely a choice of parameter width.

J Distributional jets and contraction without a smooth Hessian

At a rank change, one must construct a new coordinate before using a regular quotient theorem. This section gives the analytic conditions for such a construction and a first-derivative convergence argument. The argument is a local Gauss–Newton argument in a Hilbert space; its role here is to avoid imposing a false C^2 hypothesis at a collision seam.

J.1 A derivative-controlled jet

Definition J.1 (Resolvable one-sided jet)**.** Let $\mathcal{H}$ be a real Hilbert space and $Q : [0, a_0) \to \mathcal{H}$. A resolvable jet of order $m \ge 2$ consists of $c \ne 0$, $V \ne 0$, and $\beta > 0$ such that

$$Q(a) = Q(0) + ca^m V + R(a), \quad R(0) = 0, \quad \|R'(a)\| \le Ca^{m+\beta-1} \quad (a > 0).$$

Here Q and R are continuously differentiable for $a > 0$, and the integral of R' represents R at zero. The latter follows, for example, if R is absolutely continuous on compact intervals including zero.

Lemma J.2 (Resolving the jet)**.** *With* $\kappa = ca^m$*, the curve* $\overline{Q}(\kappa) = Q((\kappa/c)^{1/m})$ *has a one-sided* C^1 *extension to zero and*

$$\overline{Q}(\kappa) = Q(0) + \kappa V + O(|\kappa|^{1+\beta/m}), \qquad \overline{Q}'(\kappa) = V + O(|\kappa|^{\beta/m}).$$

Its pullback metric tends to $\|V\|^2 > 0$*. In the raw coordinate,* $\|Q'(a)\|^2 \sim m^2 c^2 \|V\|^2 a^{2m-2}$ *and* $\frac{1}{2}\|Q(a) - Q(0)\|^2 \sim c^2 \|V\|^2 a^{2m}/2$*.*

Proof. Integrating the derivative bound gives $\|R(a)\| \le Ca^{m+\beta}/(m+\beta)$. For $a > 0$, the chain rule gives

$$\overline{Q}'(\kappa) = V + \frac{R'(a)}{mca^{m-1}},$$

whose remainder is $O(a^\beta) = O(|\kappa|^{\beta/m})$. Its limit equals the difference quotient at zero by the integrated estimate. The remaining assertions follow by taking Hilbert norms and using $V \neq 0$. □

The bound on R' is indispensable. Small function values do not imply a small derivative: Appendix T constructs an oscillatory remainder violating that inference. For the mixture models, Appendix K proves the derivative bounds directly from the implicit quantile equation.

J.2 One coordinate, possibly at a boundary

Let I be a closed interval, possibly one-sided with zero as an endpoint. Suppose $Q \in C^1(I; \mathcal{H})$, where derivatives at endpoints are one-sided, and let

$$\sup_{z \in I} \|Q'(z) - V\| \le \delta \|V\|, \qquad 0 \le \delta < 1/3.$$

For $f(z) = \frac{1}{2}\|Q(z) - Q(z_*)\|^2$, define $h(z) = \|Q'(z)\|^2$, $g(z) = \langle Q(z) - Q(z_*), Q'(z)\rangle$, and $z^+ = \Pi_I(z - g/h)$.

Proposition J.3 (Uniform scalar contraction)**.** *The update satisfies*

$$|z^+ - z_*| \le \rho |z - z_*|, \qquad \rho = \frac{2\delta}{1-\delta} < 1.$$

Moreover,

$$(1-\delta)\|V\|\,|z-y| \le \|Q(z) - Q(y)\| \le (1+\delta)\|V\|\,|z-y|.$$

Proof. For $d = z - z_* \neq 0$, the fundamental theorem of calculus gives $Q(z) - Q(z_*) = d\int_0^1 Q'(z_* + td)\,\mathrm{d}t$. Thus

$$\Big|\frac{g}{hd} - 1\Big| \le \frac{\sup_{s \in I}\|Q'(s) - Q'(z)\|}{\|Q'(z)\|} \le \frac{2\delta}{1-\delta}.$$

Projection onto an interval containing z_* is nonexpansive relative to z_*. Comparing the integral with dV gives the distance inequalities. The case $d = 0$ follows by $g = 0$. □

A nonzero realizable target close enough to zero therefore produces a nonzero first chart step from zero. This statement is local and model-specific. It does not assert that every arbitrary target distribution gives a nonzero gradient at a Gaussian.

J.3 A multidimensional version

The same argument extends to a nonflat finite-dimensional distribution submanifold in a Hilbert representation. It also makes clear which conditioning is intrinsic.

Theorem J.4 (Whitened chart contraction)**.** *Let $K \subset \mathbb{R}^r$ be closed and convex, and let Q be C^1 on a neighborhood of K into $\mathcal{H}$. Let $B : \mathbb{R}^r \to \mathcal{H}$ be injective, put $S = B^*B$, and assume*

$$\sup_{z \in K} \|(DQ_z - B)S^{-1/2}\|_{\mathrm{op}} \le \delta < 1/3. \tag{J.1}$$

For $z_* \in K$, let $J_z = DQ_z$, $G_z = J_z^* J_z$, and

$$N(z) = \Pi_K^S\left(z - G_z^{-1} J_z^*(Q(z) - Q(z_*))\right),$$

where Π_K^S is projection in the fixed S norm. Then

$$\|N(z) - z_*\|_S \le \frac{2\delta}{1-\delta}\|z - z_*\|_S.$$

Also $(1-\delta)\|z-y\|_S \le \|Q(z) - Q(y)\| \le (1+\delta)\|z-y\|_S$ on K.

Proof. Use $w = S^{1/2}z$ coordinates; after this change, B is an isometry and $S = I$. Every J_z has singular values in $[1-\delta, 1+\delta]$, so $\|J_z^\dagger\| \le (1-\delta)^{-1}$. Writing $d = z - z_*$,

$$Q(z) - Q(z_*) = J_z d + E, \quad E = \int_0^1 (J_{z_*+td} - J_z) d \, \mathrm{d}t, \quad \|E\| \le 2\delta\|d\|.$$

The unprojected error equals $-J_z^\dagger E$. Projection is nonexpansive and fixes z_*. This proves contraction. Comparing the same integral with Bd proves the norm equivalence. Undoing the change of coordinates proves the stated formula. □

The projection metric is fixed. Replacing it by an arbitrarily varying metric without proof would invalidate the nonexpansiveness step. No projection is needed when an invariant open chart neighborhood is already available.

The Hilbert representation is exact for one-dimensional Wasserstein space via quantiles, and for products with fixed independent coordinates by taking a direct sum of quantile spaces. It is not an isometric Hilbert representation of general multidimensional Wasserstein space. The rotated-product problem uses its Poisson metric instead; it does not automatically inherit Theorem J.4.

J.4 Approximate metric, approximate target, and gradient errors

Suppose $(1-\varepsilon)G_z \preceq \widehat{G}_z \preceq (1+\varepsilon)G_z$, and the covector is perturbed by e_z. Set $\rho = 2\delta/(1-\delta)$ and

$$c_\varepsilon = \frac{\varepsilon(1+\delta)(1+\rho)}{(1-\varepsilon)(1-\delta)}.$$

The corresponding projected unit step obeys

$$\|z^+ - z_*\|_S \le (\rho + c_\varepsilon)\|z - z_*\|_S + \frac{\|e_z\|_{G_z^{-1}}}{(1-\varepsilon)(1-\delta)}. \tag{J.2}$$

To prove this, subtract the exact and perturbed steps. Whiten by G_z to get

$$\|(\widehat{G}_z^{-1} - G_z^{-1})g_z\|_{G_z} \le \frac{\varepsilon}{1-\varepsilon}\|G_z^{-1}g_z\|_{G_z}, \quad \|\widehat{G}_z^{-1}e_z\|_{G_z} \le \frac{\|e_z\|_{G_z^{-1}}}{1-\varepsilon}.$$

Since $(1-\delta)\|v\|_S \le \|v\|_{G_z} \le (1+\delta)\|v\|_S$, and the exact unprojected step has S norm at most $(1+\rho)\|z - z_*\|_S$, the bound follows before projection. Projection preserves it. This proves a contraction with an explicit error floor when $\rho + c_\varepsilon < 1$ and the last term is uniformly bounded.

For a perturbed Hilbert target $Q(z_*) + r$, the induced covector error is $e_z = -J_z^* r$. The map $J_z G_z^{-1/2}$ is an isometry, hence $\|e_z\|_{G_z^{-1}} \le \|r\|$. In one-dimensional quantile fitting, $\|r\|$ is exactly the Wasserstein distance between the ideal target and the perturbed target. Thus sampling error and metric approximation enter different terms. A small training residual or a large number of optimization steps cannot eliminate the target-sampling term.

J.5 A constant metric is a necessary comparator

In the scalar setting, the constant metric $h_0 = \|V\|^2$ gives

$$\frac{g(z)}{h_0(z - z_*)} = 1 + \epsilon_z, \qquad |\epsilon_z| \le 2\delta + \delta^2.$$

Indeed both the averaged tangent and current tangent differ from V by at most $\delta\|V\|$. Consequently the projected constant-metric unit step contracts whenever $2\delta + \delta^2 < 1$. The varying metric is not necessary merely to move away from the seam. Its role is to track the actual geometry and to fit into the residual-certificate framework away from a fixed tangent approximation.

J.6 Sufficient Armijo acceptance near a scalar seam

A unit chart step also passes a conventional Armijo test sufficiently near a realizable target. Assume $\delta \le 1/10$ and $c_A = 10^{-4}$. Let $d = z - z_*$. The distance bounds and contraction give

$$f(z^+) \le \tfrac{1}{2}(1+\delta)^2\|V\|^2\rho^2 d^2, \quad f(z) \ge \tfrac{1}{2}(1-\delta)^2\|V\|^2 d^2.$$

Also $|g(z)(z^+ - z)| \le (1+\delta)^2(1+\rho)\|V\|^2 d^2$ because projection cannot move farther than the unprojected step. For $\delta \le 0.1$, the difference between the two energy bounds exceeds c_A times the last bound. Therefore $f(z^+) \le f(z) + c_A g(z)(z^+ - z)$. This proof uses only first derivatives and avoids a Lipschitz-Hessian assumption at a merely C^1 join.

K ALL-ORDER ORNSTEIN–UHLENBECK COLLISION ANALYSIS

We now prove the derivative-controlled statements for an actual family of generators. Let X be supported in $[-B, B]$, with mean zero and variance one, independent of $Z \sim \gamma = \mathcal{N}(0,1)$. Write $s(a) = \sqrt{1-a^2}$ and

$$Y_a = aX + s(a)Z, \qquad p_a(x) = \mathbb{E}\left[s(a)^{-1}\varphi\left(\frac{x - aX}{s(a)}\right)\right].$$

The variable a is first allowed to vary in a symmetric interval $[-a_0, a_0]$ with $a_0 < 1$, which simplifies Taylor expansion. The actual one-sided collision model uses $a \ge 0$.

K.1 MOMENT BALANCE AND THE HERMITE COEFFICIENT

Probabilists' Hermite polynomials are defined by

$$e^{tx - t^2/2} = \sum_{j=0}^{\infty} H_j(x) t^j / j!.$$

Conditioning on X and using the Gaussian moment-generating function gives

$$\mathbb{E}e^{tY_a - t^2/2} = \mathbb{E}e^{atX - a^2t^2/2}, \qquad \mathbb{E}H_j(Y_a) = a^j \mathbb{E}H_j(X).$$

The coefficient identity follows by differentiating at $t = 0$, justified by boundedness of X and Gaussian tails. Agreement of the first $m-1$ ordinary moments with Gaussian moments is equivalent to vanishing $\mathbb{E}H_j(X)$ for $1 \le j < m$. Let $c = \mathbb{E}H_m(X) \ne 0$. Since cumulants of order at least three vanish for Z and scale homogeneously,

$$\mathrm{cum}_j(Y_a) = a^j \,\mathrm{cum}_j(X), \quad j \ge 3.$$

Under the lower moment matching, $\mathrm{cum}_m(X) = \mathbb{E}H_m(X) = c$, so $\kappa = ca^m$ is exactly the first nonzero cumulant, not merely an asymptotically equivalent parameter.

K.2 QUANTILE ENVELOPES

Let $q_a(z) = F_a^{-1}(\Phi(z))$. Since $aX \in [-|a|B, |a|B]$ almost surely,

$$-|a|B + s(a)z \le q_a(z) \le |a|B + s(a)z.$$

The inequalities follow by stochastic ordering under a common Gaussian noise variable. In particular, $|q_a(z)| \le B + |z|$ on a fixed interval. The density is everywhere positive, so the implicit-function theorem gives pointwise smoothness of $q_a(z)$ in a.

Pointwise smoothness alone would not justify differentiating in $L^r(\gamma)$. We next prove uniform polynomial envelopes for the derivatives. This step prevents an unjustified conversion of a formal density expansion into a Wasserstein expansion.

K.3 AN EXACT QUANTILE DIFFERENTIAL EQUATION

Let $m_a(x) = \mathbb{E}[X \mid Y_a = x]$. Differentiating a Gaussian component CDF at a fixed x yields

$$\partial_a F_a(x) = \frac{p_a(x)}{1-a^2}\big(ax - m_a(x)\big).$$

For clarity, the derivative of $(x - aX)/s(a)$ equals $(-X + ax)/s(a)^3$; averaging its product with the standard Gaussian density and dividing by p_a gives the expression. Implicit differentiation therefore gives the exact identity

$$\partial_a q_a(z) = \frac{m_a(q_a(z)) - a q_a(z)}{1-a^2}. \tag{K.1}$$

The posterior of X is its original law tilted by

$$\exp\left(\frac{axX - a^2X^2/2}{1-a^2}\right).$$

All posterior moments of X are bounded by powers of B. Differentiating a posterior expectation is a covariance with the derivative of this exponent, together with any derivative of the integrand. Consequently, for each pair of nonnegative integers r, s, the mixed derivative $\partial_a^r \partial_x^s m_a(x)$ is bounded by $C_{r,s}(1+|x|)^{M_{r,s}}$, uniformly for $|a| \le a_0$. To verify this inductively, note that every derivative of the exponent is a polynomial in x, X of degree at most one in x, with coefficients rational in a and denominators bounded away from zero. A derivative of a posterior moment is a difference of two bounded posterior moments multiplied by such coefficients. Repetition preserves polynomial dependence on x and bounded dependence on X.

Lemma K.1 (Derivative envelopes)**.** *For each integer $j \ge 0$, there exist constants C_j, M_j such that*

$$\sup_{|a|\le a_0} |\partial_a^j q_a(z)| \le C_j(1+|z|)^{M_j}.$$

Thus $a \mapsto q_a$ is C^j into every finite $L^r(\gamma)$, for every fixed finite j and r.

Proof. The case $j = 0$ is the component envelope. Equation (K.1) proves the case $j = 1$. Differentiate that equation repeatedly. The chain rule expresses each new derivative in terms of derivatives of m_a evaluated at q_a, lower derivatives of q_a, and rational coefficients with denominator bounded on $[-a_0, a_0]$. The preceding posterior bounds and the induction hypothesis give a polynomial envelope. The pointwise derivatives are continuous; dominated convergence, using that every polynomial has every finite Gaussian moment, proves continuity in L^r. Integral difference quotients then prove differentiability in that norm at each order. □

The constants are allowed to depend on the derivative order, B, a_0, and r. No bound uniform in all collision orders is implied.

K.4 Identifying the first nonzero derivative

For each fixed x, expansion of the Gaussian kernel in a at zero gives

$$\partial_a^j p_a(x)\big|_{a=0} = \mathbb{E}H_j(X)\, H_j(x)\varphi(x). \tag{K.2}$$

One way to verify the coefficient without relying on an infinite series is to differentiate the characteristic function a finite number of times. Its expression is $e^{-t^2/2}\mathbb{E}e^{iatX + a^2t^2/2}$; the order-$j$ coefficient is $\mathbb{E}H_j(X)(it)^j/j!$. Fourier inversion of the Gaussian derivative identifies (K.2). All finite differentiations and inversion are justified by a polynomial times a Gaussian majorant for $|a|$ small, since X is bounded.

Integrating and using $(\varphi H_{j-1})' = -\varphi H_j$ gives, for $j \ge 1$,

$$\partial_a^j F_a(x)\big|_{a=0} = -\mathbb{E}H_j(X)\,\varphi(x)H_{j-1}(x).$$

In $F_a(q_a(z)) = \Phi(z)$, all parameter derivatives of orders $1, \dots, m-1$ vanish at zero, so recursive implicit differentiation gives

$$q_0(z) = z, \qquad \partial_a^j q_a(z)|_{a=0} = 0\ (1 \le j < m), \qquad \partial_a^m q_a(z)|_{a=0} = cH_{m-1}(z).$$

The derivative envelopes justify Taylor's theorem in L^r for both the function and its first derivative:

$$q_a = \mathrm{Id} + \frac{ca^m}{m!}H_{m-1} + O_{L^r}(a^{m+1}), \tag{K.3}$$

$$\partial_a q_a = \frac{ca^{m-1}}{(m-1)!}H_{m-1} + O_{L^r}(a^m). \tag{K.4}$$

If X is symmetric, $p_{-a} = p_a$ and $q_{-a} = q_a$, while the first nonzero m is even. The next odd Taylor coefficient vanishes. The remainders improve to $O(a^{m+2})$ and $O(a^{m+1})$, respectively. Dividing the second expansion by mca^{m-1} and applying Lemma J.2 proves (5.2), including its derivative statement.

K.5 THE METRIC AND ENERGY CONSTANTS

In one dimension the quantile map is an isometry from $(\mathcal{P}_2(\mathbb{R}), W_2)$ into $L^2(0,1)$. Equivalently, $W_2^2(p_a, p_b) = \|q_a - q_b\|^2_{L^2(\gamma)}$. Hermite orthogonality gives $\mathbb{E}H_j(Z)^2 = j!$. Hence the limiting chart tangent $V_m = H_{m-1}/m!$ has

$$\|V_m\|^2 = \frac{(m-1)!}{(m!)^2} = \frac{1}{m\, m!}.$$

Since $\partial_a q_a = mca^{m-1}\partial_\kappa q_\kappa$,

$$G_{aa} = m^2c^2a^{2m-2}h_m(\kappa), \qquad h_m(\kappa) = \|\partial_\kappa q_\kappa\|^2 \longrightarrow \frac{1}{m\, m!}.$$

Squaring the function expansion gives the energy asymptotic. A one-sided interval in the κ domain containing the target satisfies the flat-tangent condition of Proposition J.3 after shrinking it, proving the final assertion of Theorem 5.1.

K.6 UNIFORM POINCARÉ CONTROL NEAR THIS COLLISION

The posterior representation also gives

$$(\log p_a)''(x) = -\frac{1}{1-a^2} + \frac{a^2}{(1-a^2)^2}\,\mathrm{Var}(X \mid Y_a = x).$$

As $\mathrm{Var}(X \mid Y_a = x) \le B^2$, for all sufficiently small $|a|$ the right-hand side is at most $-c_0 < 0$, uniformly in x. Strong log-concavity yields a Poincaré constant at most $1/c_0$; the standard diffusion proof integrates the gradient contraction of the reversible semigroup (Bakry et al., 2014). Thus all members sufficiently close to the collision have a uniform positive Poincaré gap. This fact will justify the angular and pairwise residual limits. It is not a statement uniform as $a \uparrow 1$, where the component variance vanishes.

K.7 A USEFUL EXACT SCORE–VELOCITY IDENTITY

Let $s_\kappa = (\log p_\kappa)'$ and $v_\kappa = -\partial_\kappa F_\kappa / p_\kappa$. The Gaussian mixture score is

$$s_a(x) = \frac{a m_a(x) - x}{1-a^2}.$$

Comparing with the Eulerian version of (K.1) shows

$$s_\kappa(x) + x = a\, v_a(x) = m\kappa\, v_\kappa(x). \tag{K.5}$$

This is exact for every nonzero a in the model, not just its leading Hermite term. It also follows from the Ornstein–Uhlenbeck evolution: changing a to $e^{-t}a$ changes κ to $e^{-mt}\kappa$, while the forward Fokker–Planck operator is $\partial_x(p' + xp)$. The direct identity avoids any ambiguity about integration constants or tail flux.

L AN EXPLICIT HIERARCHY OF FINITE GAUSSIAN MIXTURES

The preceding theorem becomes constructive through Gaussian quadrature. The quadrature rule itself is classical (Gautschi, 2004); the consequence here is an explicit all-order degeneracy hierarchy for the true transport metric and a corresponding resolving coordinate.

L.1 CONSTRUCTION AND POSITIVITY OF THE MIXING LAW

Let H_K be the monic probabilists' Hermite polynomial of degree K. It is orthogonal in $L^2(\gamma)$ to all polynomials of degree less than K, and $\int H_K^2\,\mathrm{d}\gamma = K!$. Its K roots $x_1 < \cdots < x_K$ are real and simple. For completeness, if the number of real sign-changing roots were smaller than K, the product of the corresponding linear factors would have degree less than K and would have a product with H_K of constant nonzero sign away from finitely many points. Its Gaussian integral could not vanish, contradicting orthogonality. Thus all roots are real, simple, and sign-changing.

Let ℓ_i be the Lagrange interpolation polynomial of degree $K-1$ satisfying $\ell_i(x_j) = \delta_{ij}$, and define

$$w_i = \int \ell_i \, \mathrm{d}\gamma.$$

Polynomial division gives, for every P of degree at most $2K-1$, $P = H_K Q + R$ with $\deg Q \le K-1$ and $\deg R \le K-1$. Orthogonality and interpolation imply

$$\int P \, \mathrm{d}\gamma = \sum_{i=1}^{K} w_i P(x_i). \tag{L.1}$$

Taking $P = \ell_i^2$ shows $w_i = \int \ell_i^2 \, \mathrm{d}\gamma > 0$. Taking $P = 1$ gives $\sum_i w_i = 1$. Hence these weights define a genuine probability law, not signed coefficients.

The parity $H_K(-x) = (-1)^K H_K(x)$ implies a symmetric set of roots, and uniqueness of interpolation gives equal weights for reflected roots. Therefore the law $X_K \sim \sum_i w_i \delta_{x_i}$ is symmetric. For $K \ge 2$, it is centered and has unit variance by (L.1).

L.2 The first unmatched Hermite moment

The first $2K-1$ moments of X_K match those of γ. The first mismatch is especially simple.

Proposition L.1 (Exact leading coefficient)**.** *For the quadrature law,* $\mathbb{E}H_{2K}(X_K) = -K!$.

Proof. Both H_K^2 and H_{2K} are monic of degree $2K$, so their difference has degree at most $2K-1$. Quadrature is exact on that difference. Since $H_K(X_K) = 0$ and $\mathbb{E}_\gamma H_{2K} = 0$,

$$-\mathbb{E}H_{2K}(X_K) = \mathbb{E}[H_K(X_K)^2 - H_{2K}(X_K)] = \mathbb{E}_\gamma[H_K^2 - H_{2K}] = K!.$$

Thus the mismatch is nonzero and has the stated sign. □

No estimate of an exponentially small numerical moment is needed to identify this coefficient: it is an exact polynomial identity.

L.3 The generator and its collision

Define, for $0 \le a < 1$,

$$p_{K,a} = \sum_{i=1}^{K} w_i \, \mathcal{N}(a x_i, 1 - a^2), \qquad \kappa = -K! a^{2K}.$$

All weights stay strictly positive. The component variances remain close to one as $a \to 0$. The singularity is therefore a coalescence of means under exact moment balance, not a vanishing-component-variance limit and not a zero-weight limit. Theorem 5.1 applies with $m = 2K$ and $c = -K!$, proving Corollary 5.2.

The first four examples are summarized by exact constants:

K	m	c	$G_{aa}/a^{4K-2} \to$	$\frac{1}{2}W_2^2/a^{4K} \to$
2	4	-2	$2/3$	$1/48$
3	6	-6	$3/10$	$1/240$
4	8	-24	$4/35$	$1/1120$
5	10	-120	$5/126$	$1/5040$

For $K = 2$, the mixing law is the symmetric Rademacher law. For $K = 3$, its atoms are $-\sqrt{3}, 0, \sqrt{3}$ with weights $1/6, 2/3, 1/6$. Thus the twelve-order energy example is already a concrete three-component generator rather than an abstract reparameterized scalar function.

L.4 What is regularized and what is not

In the cumulant coordinate, the limiting radial metric is $[2K(2K)!]^{-1}$, strictly positive for each fixed K. This does not give constants uniform in K. Both the fixed intrinsic scaling and the necessary approximation order can become unfavorable for large K. The chart interval in the contraction theorem can shrink with K, the node magnitudes grow, and numerical moment recovery can become difficult. These are not contradicted by removal of the raw factor a^{4K-2}.

The construction also does not parameterize arbitrary K-component mixtures. The weights, standardized relative positions, and common variance are prescribed by the quadrature law. Allowing all component means, variances, and weights to vary adds shape directions, discrete label symmetries, and additional singular strata. The result supplies a nontrivial hierarchy of verified collision directions within that larger space, not a universal desingularization of it.

L.5 A general leading-Hermite interpretation

For any fixed m, the chart tangent is $H_{m-1}/m!$ regardless of the details of the bounded mixing law once the first nonzero cumulant is used as coordinate. Different mixing laws with the same first nonzero order have the same first transport tangent but generally differ at the next order. Thus a single tangent or a local metric constant does not identify an entire family. The higher coefficients determine chart curvature, seam regularity, and the neighborhood in which the local convergence bound applies.

M Constructing signed collision charts and controlling the seam

The finite quadrature family supplies a negative cumulant branch. To cross the Gaussian in a signed coordinate, a positive branch with the same normalized transport tangent must be constructed. Merely assigning a negative raw separation to an even-power map does not create that branch. This section gives a constructive density model and proves its derivative regularity.

M.1 A bounded moment-canceling perturbation

Choose a nonzero nonnegative C^∞ compactly supported function w, positive on an interval. For an even desired order m, choose w even. Let P_m be the monic degree-m polynomial orthogonal to all lower-degree polynomials in the measure $w(x)\varphi(x)\,\mathrm{d}x$. Its existence follows from Gram–Schmidt because this measure has support containing an interval, so its finite moment matrices are positive definite. Define $h = wP_m$. Then h is bounded, smooth, and compactly supported, and

$$\int x^j h(x)\varphi(x)\,\mathrm{d}x = 0 \quad (0 \le j < m).$$

Since $H_m - P_m$ has degree less than m,

$$C_m := \int H_m h\,\mathrm{d}\gamma = \int w P_m^2\,\mathrm{d}\gamma > 0. \tag{M.1}$$

For even m and even w, uniqueness of the monic orthogonal polynomial gives that P_m and h are even.

Choose $0 < \epsilon \le [2\|h\|_\infty]^{-1}$ and define two probability densities

$$r_\pm(x) = \varphi(x)(1 \pm \epsilon h(x)).$$

They are positive, integrate to one, and have Gaussian moments through order $m-1$. Their first nonzero Hermite coefficients are $c_\pm = \pm\epsilon C_m$. The perturbation parameter ϵ is fixed while the smoothing parameter approaches the collision.

This is an explicit existence construction: computing P_m requires a positive-definite finite moment system. Sampling $r_\pm$ is also possible exactly in the rejection-sampling sense using proposal γ and envelope $1 + \epsilon\|h\|_\infty$. The expected proposal count is at most that envelope. No experiment or practical efficiency claim follows from this existence observation.

M.2 Ornstein–Uhlenbeck smoothing of the perturbed laws

Let $X_\pm \sim r_\pm$, independent of Z, and set $Y_{\pm,a} = aX_\pm + \sqrt{1-a^2}Z$. Relative to γ, its density is

$$p_{\pm,a}(x)/\varphi(x) = 1 \pm \epsilon P_a h(x), \quad P_a h(x) = \mathbb{E}h(ax + \sqrt{1-a^2}Z). \tag{M.2}$$

This follows either from the symmetric Mehler kernel or from Gaussian conditional expectation. In particular, the ratio stays between $1/2$ and $3/2$, uniformly in x and $|a| \le a_0 < 1$. All moments are uniformly controlled, and the Poincaré comparison lemma gives a uniform constant at most three times the Gaussian one.

The bounded-X proof does not apply directly because $X_\pm$ is unbounded. We supply the necessary replacement rather than silently extending its hypothesis.

Lemma M.1 (Quantile differentiability for the perturbation model)**.** *For each fixed finite j, r, the quantile curve of* (M.2) *is C^j in a into $L^r(\gamma)$ on $[-a_0, a_0]$. Its first nonzero quantile coefficient is $c_\pm H_{m-1}/m!$ at order a^m, with the derivative-controlled remainder of Appendix K.*

Proof. Every finite derivative of $P_a h$ in a, x is bounded by a polynomial in $|x|$, uniformly on the stated interval: differentiate under the Gaussian expectation, use bounded derivatives of the compactly supported h, and integrate the resulting polynomial in Z. For each $j \ge 1$, its Gaussian integral is zero because $\int P_a h \,\mathrm{d}\gamma = 0$ for every a. Therefore the derivative of the CDF can be integrated from the nearer tail:

$$\partial_a^j F_a(x) = \int_{-\infty}^{x} \varphi(t)\partial_a^j(p_a/\varphi)(t)\,\mathrm{d}t = -\int_{x}^{\infty} \varphi(t)\partial_a^j(p_a/\varphi)(t)\,\mathrm{d}t.$$

Gaussian tail bounds for polynomial integrands imply $|\partial_a^j F_a(x)| \le C_j\varphi(x)(1+|x|)^{M_j}$. Derivatives involving at least one x derivative satisfy the same type of bound because they differentiate $p_a = \varphi$ times a polynomially bounded smooth factor.

The density ratio bounds give comparable Gaussian tails and hence a quantile envelope $|q_a(z)| \le C(1+|z|)$. Implicit differentiation of $F_a(q_a(z)) = \Phi(z)$ divides by $p_a(q_a) \ge \varphi(q_a)/2$. The factor $\varphi(q_a)$ cancels the corresponding factor in every differentiated CDF or density expression. Induction therefore gives polynomial envelopes for all finite a derivatives of q_a, just as in Lemma K.1. Dominated convergence proves the Banach-space differentiability.

Finally, finite differentiation of the Mehler kernel at $a = 0$ gives $\partial_a^j(p_a/\varphi)|_0 = \mathbb{E}H_j(X_\pm)H_j$. Lower coefficients vanish by construction. The same implicit Taylor argument as in Appendix K proves the coefficient and derivative remainder. Parity improves the remainder for even h. □

M.3 Joining the branches

Use $a = (\kappa/c_+)^{1/m}$ when $\kappa > 0$, and $a = (\kappa/c_-)^{1/m}$ when $\kappa < 0$. At zero set $p_0 = \gamma$. Both quantile curves have derivative

$$q_0' = V_m = H_{m-1}/m!.$$

The one-sided derivative estimates imply that the signed curve is C^1 into every finite $L^r(\gamma)$. Its metric is continuous and positive near zero. The scalar contraction and target-perturbation results therefore hold on a sufficiently small two-sided interval.

A finite quadrature negative branch may be joined instead to the positive perturbation branch. Their higher-order coefficients generally differ, but equality of the first normalized tangent suffices for C^1 gluing. The negative branch is then a finite Gaussian mixture and the positive branch is a smoothed continuous mixing law. We do not label the resulting two-sided family a finite-mixture model on both sides.

M.4 Fractional regularity at the seam

Suppose the next nonzero Hermite coefficient after order m occurs at order $m + s < 2m$. The implicit quantile expansion has the corresponding term

$$\frac{\mathbb{E}H_{m+s}(X)}{(m+s)!}a^{m+s}H_{m+s-1}.$$

Nonlinear compositions of the first order-m correction appear only at order $2m$, so they do not affect this coefficient. In the cumulant coordinate this term is proportional to $|\kappa|^{1+s/m}$. Its nonzero projection onto H_{m+s-1} prevents C^2 regularity at zero when $s < m$, provided the differentiated remainder is of higher order. The derivative estimates above supply that remainder control.

For example, $X \in \{-1, 1\}$ has $\mathbb{E}H_4(X) = -2$ and $\mathbb{E}H_6(X) = 16$. Hence

$$q_a(z) = z - \frac{a^4}{12}H_3(z) + \frac{a^6}{45}H_5(z) + O_{L^r}(a^8),$$

with the first derivative remainder $O(a^7)$. In $\kappa = -2a^4$, the nonzero $|\kappa|^{3/2}H_5$ term shows that this branch, and any join retaining it, is not C^2 at zero. A smooth chart assumption cannot be restored by drawing a smooth-looking parameter diagram.

Higher regularity is possible for specially designed families with further cancellations, but it must be proved from their actual coefficients. The results in this paper only need C^1 at a seam and smoothness separately on regular strata.

M.5 A distinction between a coordinate and an identifiable direction

Although the scalar cumulant is an identifiable quantity, a direction u attached to its vanishing non-Gaussian component is not identifiable at $\kappa = 0$. Gluing scalar branches does not choose that direction canonically. Two different orientations lead to different first perturbations of the same Gaussian when the scalar cumulant is turned on. Thus the signed scalar chart is a chart for a selected curve, not a smooth chart covering a neighborhood of every multidimensional mixture collision.

N Profiled normal forms and mixed-order ravines

The connection to ravines is structural: a well-conditioned identifiable objective can become extremely flat in raw shape coordinates, while fast nuisance directions retain quadratic curvature. We establish this directly. We do not infer an arbitrary-order theorem by substituting an exponent into the fourth-order results of Davis et al. (2025).

N.1 Profiling a regular identifiable objective

Let $F : \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}$ be C^3 near $(0, 0)$, with $DF(0, 0) = 0$ and positive-definite Hessian there. Subtract the minimum value so that $F(0, 0) = 0$. Write the variables as u and z. By the implicit-function theorem, after shrinking neighborhoods there is a unique C^2 map h with

$$h(0) = 0, \qquad D_u F(h(z), z) = 0.$$

Define $\Psi(z) = F(h(z), z)$. The Hessian at zero is the Schur complement

$$D^2\Psi(0) = F_{zz} - F_{zu}F_{uu}^{-1}F_{uz},$$

which is positive definite because the full Hessian is positive definite. Taylor's theorem and continuity therefore give constants $c_i, C_i > 0$ such that

$$c_1\|z\|^2 \le \Psi(z) \le C_1\|z\|^2, \tag{N.1}$$

$$c_2\|u - h(z)\|^2 \le F(u, z) - \Psi(z) \le C_2\|u - h(z)\|^2. \tag{N.2}$$

The second estimate follows by Taylor expansion in u with z fixed, using the exact first-order stationarity at $h(z)$ and uniform positive definiteness of F_{uu} in a small convex box. It is not a comparison of function values at an arbitrary projection.

N.2 Pulling the normal form back through powers

Let $m_j \ge 2$ be integers and define $z(a) = (a_1^{m_1}, \dots, a_s^{m_s})$. An even power can either be considered on all of $\mathbb{R}$ with its sign redundancy or on a chosen nonnegative branch. Set

$$f(u, a) = F(u, z(a)), \qquad H(a) = h(z(a)), \qquad \mathcal{T} = \{(H(a), a) : a \text{ near } 0\}.$$

Define the retraction onto the graph by $R(u, a) = (H(a), a)$. It fixes $\mathcal{T}$ and is C^2. If F is C^k, it is C^{k-1}; the examples built from analytic densities have as much raw-coordinate regularity as needed for a specified finite ravine order.

Proposition N.1 (Mixed-order ravine normal form)**.** *Near zero,*

$$f(u,a) - f(R(u,a)) \asymp \operatorname{dist}((u,a),\mathcal{T})^2, \qquad f(H(a),a) \asymp \sum_{j=1}^{s} |a_j|^{2m_j}. \tag{N.3}$$

Furthermore $T_0\mathcal{T} = \{0\} \times \mathbb{R}^s = \ker D^2 f(0)$*, and the Hessian is positive definite on the normal complement* $\mathbb{R}^r \times \{0\}$.

Proof. Equation (N.2) gives the first difference comparable to $\|u - H(a)\|^2$. If H has Lipschitz constant L_H, then for every nearby a',

$$\|u - H(a)\| \le \|u - H(a')\| + L_H\|a - a'\| \le \sqrt{1 + L_H^2}\|(u,a) - (H(a'),a')\|.$$

Taking the infimum and using the point $(H(a),a)$ gives $\operatorname{dist}((u,a),\mathcal{T}) \le \|u - H(a)\| \le \sqrt{1+L_H^2}\operatorname{dist}((u,a),\mathcal{T})$ in a suitably restricted local graph neighborhood. The shape estimate follows from $\|z(a)\|^2 = \sum_j |a_j|^{2m_j}$.

Since every $m_j \ge 2$, $Dz(0) = 0$, so $DH(0) = 0$ and the tangent graph is the a space. The chain rule for the Hessian of f has only the F_{uu} block at zero: terms involving $Dz(0)$ vanish, and the terms involving $D^2 z(0)$ are multiplied by $F_z(0) = 0$. The asserted kernel and normal positivity follow. □

Here the graph retraction is not claimed to be the nearest-point projection. Its derivative equals the tangent projection at the base point because $DH(0) = 0$, not because every graph projection is orthogonal everywhere.

N.3 Why a single growth exponent can fail

If all $m_j = m$, finite-dimensional norm equivalence gives $f|_{\mathcal{T}} \asymp \|a\|^{2m}$. If two orders differ, no single p gives a uniform two-sided estimate $f|_{\mathcal{T}} \asymp \|a\|^p$ around zero. Indeed, on the jth coordinate axis, the ratio is comparable to $|a_j|^{2m_j - p}$, which is bounded above and below by positive constants as $a_j \to 0$ only if $p = 2m_j$. Two different m_j make this impossible.

One-sided bounds still exist. For $\|a\| \le 1$, letting $m_{\min} \le m_j \le m_{\max}$,

$$c\|a\|^{2m_{\max}} \le \sum_j |a_j|^{2m_j} \le C\|a\|^{2m_{\min}}.$$

They are not a replacement for a two-sided single-order assumption. In particular, assigning the worst exponent to all directions can hide faster identifiable directions and does not justify a theorem requiring equality of growth orders.

N.4 Algorithmic consequences and their information requirement

The z objective is locally strongly convex and smooth in ordinary coordinates; a smooth positive-definite quotient metric preserves local PL and retraction smoothness. Thus the certified algorithm has geometric convergence in its identifiable energy, subject to the trapping conditions. Recovering raw coordinates incurs only a Hölder relation:

$$|a_j - a_{j,*}| \le |z_j - z_{j,*}|^{1/m_j}$$

on a fixed nonnegative branch, with an additional factor if z_j was scaled by a nonzero constant. Therefore geometric energy or z convergence still gives geometric raw-coordinate convergence with a slower exponent. This statement uses a known resolving map. It is not an oracle-free acceleration theorem for arbitrary nonconvex flat functions, and it does not remove a statistical inability to infer z accurately from data.

A transparent exact example is the Gaussian location family $\mu_{u,a} = \mathcal{N}((u, z(a)), I_{r+s})$. Its squared Wasserstein fitting objective to the standard Gaussian is

$$f(u,a) = \tfrac{1}{2}\|u\|^2 + \tfrac{1}{2}\sum_j a_j^{2m_j}, \qquad G_{u,a} = \operatorname{diag}\left(I_r, m_1^2 a_1^{2m_1-2}, \ldots, m_s^2 a_s^{2m_s-2}\right).$$

Its quotient in (u, z) is Euclidean and one exact unit natural step reaches the target. In raw coordinates at zero the differential can lose every a direction. This example checks the algebra but has a regular Gaussian image; the finite-mixture hierarchy supplies non-Gaussian images with the same order mechanism.

N.5 An exact shift–scale ravine for the mixture family

For the standardized quantile q_a from Appendix K, let

$$C(a) = \mathbb{E}[Z q_a(Z)], \qquad \mu_{b,s,a} = \mathrm{Law}(b + sY_a), \quad s > 0.$$

Using $\mathbb{E}q_a = 0$ and $\mathbb{E}q_a^2 = 1$,

$$\begin{aligned} \tfrac{1}{2}W_2^2(\mu_{b,s,a}, \gamma) &= \tfrac{1}{2}b^2 + \tfrac{1}{2}(s^2 + 1 - 2sC(a)) \\ &= \tfrac{1}{2}b^2 + \tfrac{1}{2}(s - C(a))^2 + \tfrac{1}{2}(1 - C(a)^2). \end{aligned} \tag{N.4}$$

Near zero, $C(a) > 0$, so the profiled scale is admissible. Write $E(a) = \frac{1}{2}\|q_a - \mathrm{Id}\|^2 = 1 - C(a)$. Then the restricted energy is exactly $E(a) - E(a)^2/2$, with leading term

$$\frac{c^2}{2m\, m!} a^{2m}.$$

The graph $\{(0, C(a), a)\}$ has exactly quadratic fast-variable energy. The derivative envelopes show smoothness in raw a; $C'(0) = 0$, so its base tangent agrees with the Hessian kernel. The distance-to-graph comparison from Proposition N.1 verifies the normal growth estimate. For the Gauss–Hermite K-component family, the restricted order is $4K$, with no appeal to a fourth-order-only theorem.

N.6 Normal curvature is not automatically uniform along a singular family

In (N.4), the fast-variable quadratic coefficients are exactly one in the chosen raw coordinates. In a general model, profiling and normal Hessian estimates may deteriorate when approaching a boundary where the identifiable metric changes rank or the local minimizer manifold ceases to be smooth. The regular quotient convergence theorem must be applied only where its hypotheses hold. A verified cumulant curve can bridge a selected radial collision, but it cannot justify a uniform normal gap for every surrounding mixture parameter.

O Angular transport geometry and the intrinsic cone limit

A radial coordinate can resolve an invisible first derivative while leaving other directions intrinsically unidentifiable. We compute this separation for a single non-Gaussian axis. The result extends the fourth-order calculation to every collision order covered by the preceding generator constructions.

O.1 The axisymmetric distribution and its source

Let $u \in S^{d-1}$ and let Y_κ be standardized, independent of $Z_d \sim \mathcal{N}(0, I_d)$. Define

$$X_{\kappa,u} = Z_d + (Y_\kappa - u^T Z_d)u.$$

The component orthogonal to u is standard Gaussian, independent of the component Y_κ. Its density is $p_\kappa(u^T x)\varphi_{d-1}(P_{u^\perp}x)$. If Y_κ is symmetric then u and $-u$ represent the same distribution, so the identifiable orientation is projective. Without symmetry the parameter is spherical and this sign identification is not made.

By rotational invariance it suffices to work at $u = e_1$ and differentiate $u(t) = \cos t\, e_1 + \sin t\, e_2$. Writing $s_\kappa = (\log p_\kappa)'$ gives the logarithmic density derivative

$$b(x) = x_2\big(s_\kappa(x_1) + x_1\big).$$

For an order-m OU family, (K.5) makes it exactly $m\kappa x_2 v_\kappa(x_1)$. This identity is why the angular Poisson problem is explicitly tractable.

O.2 Solving the true continuity equation

Consider the Hilbert space with norm

$$\|w\|_{1,p_\kappa}^2 = \int (w'^2 + w^2) p_\kappa \, \mathrm{d}x$$

and the coercive form $\mathfrak{a}_\kappa(w, \psi) = \int (w'\psi' + w\psi) p_\kappa$. There is a unique weak solution of

$$\mathfrak{a}_\kappa(w_\kappa, \psi) = \int m v_\kappa \psi p_\kappa \, \mathrm{d}x \quad \text{for all } \psi, \qquad \mathcal{A}_\kappa w_\kappa = m v_\kappa, \tag{O.1}$$

where $\mathcal{A}_\kappa = -\partial_{xx} - s_\kappa \partial_x + 1$. Existence follows from coercivity and $v_\kappa \in L^2(p_\kappa)$, with no Poincaré estimate required for this shifted operator.

Let $\Psi(x) = \kappa x_2 w_\kappa(x_1)$. For the product density at $u = e_1$, the weighted negative Laplacian acts as

$$-p^{-1} \operatorname{div}(p \nabla \Psi) = \kappa x_2 \mathcal{A}_\kappa w_\kappa = m \kappa x_2 v_\kappa(x_1) = b(x).$$

Thus $\nabla\Psi$ solves the infinitesimal continuity equation and is in the gradient closure. By the minimum-energy lemma, it is the true transport tangent rather than merely the velocity of a chosen rotating sample coupling. Its norm is

$$j(\kappa) = \kappa^2 \int \left(w_\kappa'^2 + w_\kappa^2 \right) p_\kappa \, \mathrm{d}x. \tag{O.2}$$

Every tangent direction to the orientation sphere has this same coefficient by rotational symmetry in the orthogonal Gaussian coordinates.

The radial potential depends only on x_1. Its inner product with $\nabla\Psi$ vanishes after integration in x_2, since $\mathbb{E}x_2 = 0$. Distinct angular coordinate directions are orthogonal because the corresponding Gaussian coordinates are orthogonal. The exact metric is therefore

$$h_m(\kappa) \, \mathrm{d}\kappa^2 + j(\kappa) \| \, \mathrm{d}u\|^2.$$

At zero every orientation gives the same standard Gaussian. The radial tangent at zero is nevertheless well-defined after a particular orientation has been selected.

O.3 A scalar residual enclosure

Put $f_\kappa = m v_\kappa$ and $J_\kappa = \mathfrak{a}_\kappa(w_\kappa, w_\kappa)$. For an admissible trial ψ with strong residual $r = f_\kappa - \mathcal{A}_\kappa \psi \in L^2(p_\kappa)$, define

$$D_\kappa(\psi) = 2\langle f_\kappa, \psi \rangle_{p_\kappa} - \mathfrak{a}_\kappa(\psi, \psi).$$

Completing the square and using $\mathcal{A}_\kappa \succeq I$ gives

$$D_\kappa(\psi) \le J_\kappa \le D_\kappa(\psi) + \|r\|_{L^2(p_\kappa)}^2. \tag{O.3}$$

Indeed $J - D = \mathfrak{a}(w - \psi, w - \psi) = \langle r, \mathcal{A}^{-1} r \rangle \le \|r\|^2$. This estimate provides both a practical variational bound and a convenient proof of the limiting metric. Rigorous numerical certification would additionally require rigorous integration of its two sides; ordinary floating-point evaluation is not an interval proof.

O.4 The normalized angular limit

Let $V_m = H_{m-1}/m!$. The Gaussian shifted operator satisfies

$$\mathcal{A}_0 V_m = (-\partial_{xx} + x\partial_x + 1) V_m = m V_m.$$

The Eulerian tangents v_κ converge to V_m in polynomially weighted finite $L^r(p_\kappa)$ spaces. Here is the relevant justification. The Taylor proof in Appendix K yields pointwise polynomial remainder bounds, not only integrated bounds, for q_κ' and $q_\kappa - \mathrm{Id}$. The component envelope also bounds the inverse quantile coordinate by a polynomial in $|x|$. Substituting $z = q_\kappa^{-1}(x)$ into $v_\kappa(x) = q_\kappa'(z)$ therefore gives a polynomially dominated difference $v_\kappa - V_m$ tending to zero. The perturbed-Gaussian construction has the analogous bounds by Lemma M.1. Uniform Gaussian-type moment bounds justify every finite polynomial weight used below.

With the fixed trial $\psi = V_m$, (K.5) gives

$$f_\kappa - \mathcal{A}_\kappa V_m = m(v_\kappa - V_m) + (s_\kappa + x)V_m' = m(v_\kappa - V_m) + m\kappa v_\kappa V_m'.$$

Its squared $L^2(p_\kappa)$ norm tends to zero. Also

$$D_\kappa(V_m) \longrightarrow 2m\|V_m\|_\gamma^2 - \int (V_m'^2 + V_m^2)\,\mathrm{d}\gamma = m\|V_m\|_\gamma^2 = \frac{1}{m!}.$$

Applying (O.3) proves

$$\frac{j(\kappa)}{\kappa^2} \longrightarrow \frac{1}{m!}. \tag{O.4}$$

This proof does not assume continuity of inverse operators between changing weighted Hilbert spaces. A fixed polynomial trial and a vanishing residual supply both bounds.

O.5 A CONE LIMIT OF THE INTRINSIC LENGTH METRIC

On one signed branch set $r = |\kappa|/\sqrt{m\,m!}$. Combining the radial and angular limits gives

$$g = (1 + o(1))\,\mathrm{d}r^2 + (m + o(1))r^2 g_{\mathcal{U}}, \tag{O.5}$$

where $\mathcal{U} = S^{d-1}$, or $\mathbb{R}P^{d-1}$ in the symmetric case, with its unit-sphere quotient metric. The errors depend only on r and tend uniformly to zero as $r \downarrow 0$.

To make the statement about distance precise, equip the punctured family with the length metric of its pullback and add the common Gaussian point by metric completion. On $0 < r \le \epsilon$, the quadratic form is bounded above and below by $(1 \pm \omega(\epsilon))$ times $g_C = \mathrm{d}r^2 + mr^2 g_{\mathcal{U}}$, with $\omega(\epsilon) \to 0$. Integrating along any absolutely continuous path proves the corresponding square-root bounds for its length, and taking infima proves them for the induced distances restricted to this region. Radial paths have finite length to the tip. After $r = \epsilon R$ and division of distances by ϵ, this gives the metric-cone limit on bounded radial regions. This is an intrinsic model-length statement, not a claim that all endpoint Wasserstein distances equal those intrinsic distances.

The cone may be singular at its tip, but one must not assert this for every dimension and order. For a symmetric order-four model in $d = 2$, the link $\mathbb{R}P^1$ has length π, and the factor $m = 4$ scales its length to 2π. The limiting intrinsic cone is then the Euclidean plane. Orientation remains unidentifiable at the tip even in this special metric limit. In dimensions with two-dimensional tangential planes, the cone metric has tangential sectional curvature $(1/m - 1)/r^2$ away from the tip, by the usual warped-product formula; radial sectional curvature is zero. These formulas refer to the limiting cone, not to an unproved curvature convergence for a merely C^1 family.

O.6 THE GAUSSIAN TANGENT EMBEDDING

At the Gaussian, the selected radial tangent field is

$$V_{m,u}(x) = \frac{H_{m-1}(u^T x)}{m!}u = \nabla\left(\frac{H_m(u^T x)}{m\,m!}\right).$$

For unit vectors u, v, the Gaussian Hermite covariance identity gives

$$\langle V_{m,u}, V_{m,v}\rangle_{\gamma_d} = \frac{(u^T v)^m}{m\,m!}. \tag{O.6}$$

It follows by expanding the joint Gaussian generating function $\mathbb{E}e^{tu^T Z - t^2/2}e^{sv^T Z - s^2/2} = e^{tsu^T v}$ and comparing coefficients. The derivative of this kernel along the orientation sphere recovers the factor m in (O.5). Hermite Gaussian linearization is an established technique; the complete nonlinear OU metric identity and its residual-controlled limit are what connect it to the algorithm here.

P ROTATED PRODUCT GENERATORS AND PAIRWISE POISSON REDUCTION

This section proves Theorem 6.1. It introduces a coupled generator family with unknown orthogonal mixing, rather than treating a known-coordinate product as evidence of solving orientation learning. The reduction is exact for the true transport metric, not for a common-noise surrogate.

P.1 Assumptions and tensor subspaces

Let p_i be positive C^2 densities on $\mathbb{R}$, centered with unit variance. Assume finite second moments of their scores $s_i = (\log p_i)'$, zero tail flux so that $\int s_i p_i = 0$, and Poincaré gaps $\lambda_i > 0$. Radial coordinate sources are assumed to have finite Dirichlet-dual norm and well-defined differentiable distribution curves. The near-collision asymptotic below is for the OU and smoothed-perturbation families already constructed, not for arbitrary densities with matching moments alone.

Let $\mu = \bigotimes_i p_i$. The weighted Laplacian on the product is $\mathcal{L} = \sum_i \mathcal{L}_i$, with $\mathcal{L}_i = -\partial_{ii} - s_i\partial_i$. Its Dirichlet form is the sum of coordinate gradient energies. In the two-coordinate space define

$$\mathcal{H}^0_{ij} = \{\psi \in L^2(p_i \otimes p_j) : \mathbb{E}[\psi(Y_i, Y_j) \mid Y_j] = 0, \\ \mathbb{E}[\psi(Y_i, Y_j) \mid Y_i] = 0\}.$$

This is the closed tensor product of the two marginal mean-zero spaces. Conditional expectation projections commute with the product form and the associated resolvent. Thus if a source belongs to $\mathcal{H}^0_{ij}$, its mean-zero weak solution under $\mathcal{L}_i + \mathcal{L}_j$ also belongs to that space.

For a direct proof of the last assertion, decompose a two-variable function as its constant, its centered x_i-only part, its centered x_j-only part, and its separately centered remainder. These four components are orthogonal for both the L^2 and Dirichlet forms. Testing the weak equation against the first three spaces makes the corresponding solution components vanish, except for the irrelevant constant fixed by mean zero.

On $\mathcal{H}^0_{ij}$, applying the marginal Poincaré inequality conditionally in each coordinate gives

$$\int (|\partial_i\psi|^2 + |\partial_j\psi|^2)\,\mathrm{d}\mu_{ij} \geq (\lambda_i + \lambda_j)\int \psi^2\,\mathrm{d}\mu_{ij}. \tag{P.1}$$

The sum, rather than the minimum, is valid because the function is centered separately in both coordinates.

P.2 The source of an infinitesimal rotation

Let $U(t) = U\exp(t\Omega)$, with $\Omega^T = -\Omega$, and use $y = U^T x$. Differentiating the density $\prod_i p_i((e^{-t\Omega}y)_i)$ gives

$$\partial_t \log p_t(x)|_{t=0} = -\sum_{i<j} \Omega_{ij} b_{ij}(y_i, y_j), \quad b_{ij} = y_j s_i(y_i) - y_i s_j(y_j).$$

The sign is relevant to the potential, although not to its squared metric coefficient. Since $\mathbb{E}y_i = 0$ and $\mathbb{E}s_i = 0$, b_{ij} is centered separately in i, j. Its L^2 norm is finite by the assumed score and second-moment integrability. The pair equation has a unique weak solution in $\mathcal{H}^0_{ij}$:

$$(\mathcal{L}_i + \mathcal{L}_j)\psi_{ij} = b_{ij}.$$

The potential associated with coefficient Ω_{ij} is $-\Omega_{ij}\psi_{ij}$. Rotation by U is an ambient isometry, so its gradient-energy norm is computed in the product coordinates without distortion.

P.3 Orthogonality of all blocks

A radial potential for coordinate i depends only on y_i and can be chosen mean zero. Different radial potentials are Dirichlet-orthogonal because their gradients have disjoint coordinate support. Their squared norms are $h_i(\kappa_i)$.

For a radial potential ϕ_i and a rotational pair potential ψ_{jk},

$$a_\mu(\phi_i, \psi_{jk}) = \int \phi_i b_{jk}\,\mathrm{d}\mu = 0.$$

If $i \notin \{j, k\}$, independence and the zero mean of b_{jk} give zero. If $i = j$ or $i = k$, integrate in the other coordinate and use separate centering. For two distinct pairs, say (i, j) and (k, l),

$$a_\mu(\psi_{ij}, \psi_{kl}) = \int b_{ij}\psi_{kl}\,\mathrm{d}\mu = 0.$$

There is at least one coordinate in the first pair not in the second; integration in that coordinate annihilates b_{ij}. The assertion also follows from the orthogonal tensor decomposition.

Since a linear combination of these potentials solves the full source equation, it is the minimum-energy solution. Its energy is exactly

$$\sum_i h_i(\kappa_i)\dot{\kappa}_i^2 + \sum_{i<j} J_{ij}(\kappa_i,\kappa_j)\Omega_{ij}^2, \quad J_{ij} = \int b_{ij}\psi_{ij}\,\mathrm{d}\mu_{ij}.$$

This proves the metric decomposition, including the absence of all asserted cross terms.

P.4 The exact zero set of a rotation block

Proposition P.1 (Only Gaussian pairs are rotationally invisible)**.** *Under the above assumptions, $J_{ij} = 0$ if and only if both p_i, p_j are standard Gaussian.*

Proof. The inverse operator is positive definite on $\mathcal{H}_{ij}^0$, so $J_{ij} = 0$ is equivalent to $b_{ij} = 0$ in L^2. Positivity and continuity of the densities and scores then give $ys_i(x) = xs_j(y)$ for all x, y. Fix a nonzero y; it follows that $s_i(x) = cx$. Varying y gives $s_j(y) = cy$ with the same constant. Integrating the score yields Gaussian-shaped densities proportional to $e^{cx^2/2}$; normalizability forces $c < 0$. Unit variance forces $c = -1$. Conversely two standard Gaussian scores equal $-x$ and $-y$, making $b_{ij} = 0$ identically. □

Identical non-Gaussian marginals do not produce a zero rotation coefficient. Equal covariance eigenvalues are not the criterion: every marginal here has the same variance, but non-Gaussian product densities still carry orientation information.

If precisely g marginals are Gaussian, the infinitesimal rotational kernel has dimension $g(g-1)/2$ and is the Lie algebra of rotations within that Gaussian block. Sign changes of symmetric components and permutations of exchangeable labels are discrete equivalences; they do not add tangent-kernel dimensions. A supplied regular quotient can remove the continuous Gaussian-block gauge on a fixed stratum. At a transition where another component becomes Gaussian, the kernel changes dimension and the regular quotient hypothesis must be reconsidered.

P.5 Asymptotic coefficients at a joint collision

For marginal i with order $m_i \geq 3$, write $V_i = H_{m_i-1}/m_i!$ and let v_i be its Eulerian cumulant tangent. The exact score identity gives

$$b_{ij} = m_i\kappa_i v_i(y_i)y_j - m_j\kappa_j v_j(y_j)y_i.$$

Consider the trial potential

$$\psi_{ij}^0 = \kappa_i V_i(y_i)y_j - \kappa_j V_j(y_j)y_i.$$

Lower moment matching gives $\mathbb{E}_{p_i}V_i = 0$ and $\mathbb{E}_{p_i}V_i' = 0$, so this potential is separately centered. Gaussian Hermite identities imply

$$(\mathcal{L}_i + \mathcal{L}_j)(V_i y_j) = m_i V_i y_j - (s_i + y_i)V_i' y_j - V_i(s_j + y_j).$$

Use the analogous expression with i, j interchanged. The polynomially dominated convergence $v_i \to V_i$, the exact identity $s_i + y_i = m_i\kappa_i v_i$, and uniform moments then imply

$$\|b_{ij} - (\mathcal{L}_i + \mathcal{L}_j)\psi_{ij}^0\|_{L^2(\mu_{ij})} = o\left(\sqrt{\kappa_i^2 + \kappa_j^2}\right). \tag{P.2}$$

This limit is uniform over relative approaches of the two coordinates to zero: each residual component is its corresponding $|\kappa_i|$ or $|\kappa_j|$ times a quantity tending to zero as both parameters approach zero. The Poincaré gaps are bounded below uniformly near these collisions by the constructions in Appendices K and M.

The trial energy has no cross term. For example one cross derivative integral contains $\mathbb{E}V_i'$ and the other contains $\mathbb{E}V_j'$, both zero. Hence

$$a(\psi_{ij}^0, \psi_{ij}^0) = \kappa_i^2\mathbb{E}_{p_i}(V_i'^2 + V_i^2) + \kappa_j^2\mathbb{E}_{p_j}(V_j'^2 + V_j^2) = \frac{\kappa_i^2}{m_i!} + \frac{\kappa_j^2}{m_j!} + o(\kappa_i^2 + \kappa_j^2).$$

Let $e = \psi_{ij} - \psi^0_{ij}$. The pair gap and (P.2) give $a(e,e) = o(\kappa_i^2 + \kappa_j^2)$. Since the trial energy is $O(\kappa_i^2 + \kappa_j^2)$, Cauchy–Schwarz makes its cross energy with e of smaller order as well. Expanding $a(\psi^0 + e, \psi^0 + e)$ proves (6.3).

If one marginal is exactly Gaussian, its score source vanishes and the equation reduces to the single-axis calculation of Appendix O. Thus the two-dimensional result is consistent with, and extends, that one-dimensional reduction.

P.6 A COMPUTABLE PAIR CERTIFICATE

For a trial $\widetilde{\psi} \in \mathcal{H}^0_{ij}$ with strong residual $r = b_{ij} - (\mathcal{L}_i + \mathcal{L}_j)\widetilde{\psi}$, set

$$D_{ij} = 2\langle b_{ij}, \widetilde{\psi}\rangle - a(\widetilde{\psi}, \widetilde{\psi}).$$

The same completion-of-squares argument gives

$$D_{ij} \le J_{ij} \le D_{ij} + \frac{\|r\|^2}{\lambda_i + \lambda_j}. \tag{P.3}$$

Separate centering is necessary for the sum-gap bound. If a numerical potential is not centered, it can first be projected by subtracting its two conditional means and adding its total mean. Alternatively one can use a weaker full-product gap, but not silently retain $\lambda_i + \lambda_j$.

Diagonal enclosures for the radial and pair blocks give a simultaneous matrix enclosure in the natural product-frame coordinates. Pulling that enclosure back through a redundant parameterization preserves the Loewner order. This provides a concrete family where the abstract residual certificate does not require solving a general d-dimensional Poisson equation for every parameter.

P.7 A QUOTIENT-COMPATIBLE BLOCK ALGORITHM

On a fixed-rank stratum, solve the radial and active pair systems to the prescribed relative certificate. For the quotient covector components (g_i, g_{ij}), set

$$\dot\kappa_i = -g_i/\widehat{h}_i, \qquad \Omega_{ij} = -g_{ij}/\widehat{J}_{ij} \quad \text{for active pairs,}$$

where g_{ij} is defined relative to the actual generator direction $U\Omega$, including the sign convention above. Set the certified Gaussian-gauge components to zero. Use a smooth radial chart update on its regular domain and the orthogonal retraction $U^+ = U\exp(\alpha\Omega)$. The product of these retractions is valid on a local regular quotient after fixing the known gauge. Backtracking and the certificates then meet the descent theorem's assumptions whenever the objective has the stated smoothness there.

A near-zero but nonzero J_{ij} is an intrinsically difficult direction, not a gauge direction. Its certified inverse can produce a large orientation step, requiring a smaller admissible step size. The present theorem neither discards that direction with an arbitrary threshold nor promises a uniform convergence rate through the joint Gaussian stratum. At a true collision, even the choice of which orientation supports a new radial cumulant is additional information; a scalar chart does not discover it.

Q STATISTICAL LIMITS OF A RESOLVING COORDINATE

Optimization regularity and statistical identifiability are different properties. This appendix gives a self-contained two-point lower bound and a matching elementary upper bound for the specified raw separation parameter. It is an all-order specialization of the weak-identifiability phenomenon, not a claim to have discovered the $n^{-1/8}$ phenomenon or moment-based mixture estimation (Dwivedi et al., 2020; Wu & Yang, 2020).

Q.1 A DENSITY-RATIO EXPANSION

For the bounded-mixing OU model, write $r_a = p_a/\varphi$. Its explicit expression is

$$r_a(x) = \frac{1}{\sqrt{1-a^2}} \exp\left(-\frac{a^2x^2}{2(1-a^2)}\right) \mathbb{E}\exp\left(\frac{axX - a^2X^2/2}{1-a^2}\right).$$

For each fixed derivative order and sufficiently small a_0, derivatives in a are bounded by a polynomial in $|x|$ times $Ce^{C|x|}$, uniformly for $|a| \le a_0$. Dropping the nonpositive quadratic exponential gives such an upper bound, and boundedness of X controls the remaining factors. This bound belongs to every finite $L^r(\gamma)$. Hence the finite Taylor coefficient calculation of Appendix K also holds in $L^2(\gamma)$ for the density ratio:

$$r_\kappa = 1 + \frac{\kappa}{m!} H_m + o_{L^2(\gamma)}(|\kappa|). \tag{Q.1}$$

The compactly supported perturbation construction has the same expansion directly from its Mehler representation and uniformly bounded density ratio.

Hermite orthogonality now gives

$$\chi^2(p_\kappa \| \gamma) = \frac{\kappa^2}{m!} + o(\kappa^2), \qquad \mathrm{KL}(p_\kappa \| \gamma) = \frac{\kappa^2}{2m!} + o(\kappa^2). \tag{Q.2}$$

To justify the second formula without a false uniform pointwise Taylor assumption, put $t = r_\kappa - 1$ and use $0 \le (1+t)\log(1+t) - t \le t^2$ for $t \ge -1$. The ratio of this expression to t^2 tends to $1/2$ as $t \to 0$. Since t/κ converges in L^2, its squares are uniformly integrable. Truncation followed by dominated convergence gives the stated limit. Normalization removes the linear term exactly.

The score of the signed cumulant chart at zero is $H_m/m!$, with Fisher information $1/m!$. Its Wasserstein metric is $1/(m\,m!)$. The difference is the Ornstein–Uhlenbeck eigenvalue m; Fisher information and the transport metric must not be interchanged.

Q.2 A LOWER BOUND FOR RAW SEPARATION

Fix the one-sided family $\kappa = ca^m$, $c \neq 0$, with $a \in [0, a_0]$. Suppose n observations are independent with this density, and let $\widehat{a}$ be any estimator.

Proposition Q.1 (Local raw-parameter lower bound)**.** *There are constants $h, c_1 > 0$, depending on the fixed family, such that for all sufficiently large n,*

$$\inf_{\widehat{a}} \sup_{a \in \{0, hn^{-1/(2m)}\}} \mathbb{E}_a |\widehat{a} - a| \ge c_1 n^{-1/(2m)}.$$

Consequently the same lower bound holds on any parameter interval containing these two points.

Proof. Choose $a_1 = hn^{-1/(2m)}$, so $\kappa_1 = ch^m n^{-1/2}$. By (Q.2), $n\chi^2(p_{\kappa_1}\|\gamma)$ is bounded by $2c^2h^{2m}/m!$ for large n. For product laws,

$$1 + \chi^2(p_{\kappa_1}^{\otimes n} \| \gamma^{\otimes n}) = \big(1 + \chi^2(p_{\kappa_1}\|\gamma)\big)^n.$$

Choose h small enough that the total variation distance of the two product laws is at most $1/2$, using $\mathrm{TV} \le \frac{1}{2}\sqrt{\chi^2}$. An estimator of a induces a test by choosing the nearer of $0, a_1$. A wrong test decision incurs absolute estimation error at least $a_1/2$. The sum of the two test error probabilities is at least $1 - \mathrm{TV} \ge 1/2$. Therefore the maximum of the two risks is at least $a_1/8$, proving the result. □

For $m = 4$ this is the familiar $n^{-1/8}$ local separation scale. For the K-node quadrature construction it is $n^{-1/(4K)}$. The statement concerns this specific one-dimensional path through a collision. It is not a minimax theorem for all parameters of an unrestricted K-component mixture.

Q.3 A MATCHING UPPER BOUND WITH KNOWN FAMILY STRUCTURE

The exact identity $\mathbb{E}H_m(Y) = \kappa$ suggests

$$\widehat{\kappa} = \Pi_I \left(\frac{1}{n} \sum_{l=1}^{n} H_m(Y_l) \right), \qquad \widehat{a} = (\widehat{\kappa}/c)^{1/m},$$

where I is the known one-sided cumulant interval. Uniform $2m$th moments give $\sup_{a \in [0,a_0]} \mathbb{E}|\widehat{\kappa} - \kappa|^2 \le C/n$; projection cannot increase the error relative to a true point in I. For nonnegative numbers x, y, concavity of the root implies $|x^{1/m} - y^{1/m}| \le |x - y|^{1/m}$. Jensen's inequality therefore yields

$$\sup_{a \in [0,a_0]} \mathbb{E}|\widehat{a} - a| \le |c|^{-1/m} (C/n)^{1/(2m)}. \tag{Q.3}$$

This upper bound assumes the order m, coefficient c, and branch are known. Estimating those additional structural quantities is a different problem. It also does not imply that a fourth-moment estimator is uniformly preferable to minimum-Wasserstein estimation; it only establishes the rate scale for the selected model.

In the resolving coordinate, the same estimator has root-n mean-square accuracy. A local bilipschitz Wasserstein chart converts this to a corresponding distribution error. The raw parameter remains a Hölder function of that regular quantity. This is the precise sense in which a chart can improve optimization without defeating statistical indistinguishability.

Q.4 Orientation lower bounds

For the axisymmetric model, at $\kappa = 0$ all u produce exactly the same standard Gaussian law. Under any orientation loss that separates two fixed nonequivalent axes, no estimator can have uniformly vanishing risk over those axes at zero: the data laws are identical.

There is also a quantitative near-zero statement. For two fixed axes u, v, each corresponding law has the same χ^2 divergence to γ_d as its one-dimensional non-Gaussian marginal has to γ. For product observations, the previous calculation implies that both laws are close in total variation to $\gamma_d^{\otimes n}$ whenever $n\kappa^2$ is sufficiently small. The triangle inequality then makes the two orientation experiments difficult to distinguish. A two-point testing argument gives a positive constant lower bound for their angular risk in this regime. This is a fixed-dimensional, fixed-pair lower bound; it is not a claimed sharp joint dependence on dimension and sample size.

The angular Fisher information also vanishes. Its score is $m\kappa x_2 v_\kappa(x_1)$, so

$$I_{uu}(\kappa) = m^2\kappa^2 \int v_\kappa^2 p_\kappa \sim \frac{m\kappa^2}{m!},$$

whereas the true angular transport coefficient is $j(\kappa) \sim \kappa^2/m!$. Both detect the lost orientation, although they are different geometric objects. No parameter retraction can manufacture observations distinguishing orientations at an exactly Gaussian distribution.

R Regular models with verifiable analytic assumptions

The abstract certificate is meaningful only when its assumptions can be met by an actual distribution model. We give two complementary examples: a smooth positive density family with a known Poincaré bound and convergent spectral approximations, and the Gaussian covariance-factor quotient with exact transport potentials. Neither model class is claimed as newly introduced here; they instantiate and test the logical reach of the framework.

R.1 A compact exponential density model

Let $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ carry normalized Haar measure. Choose real C^∞ features $\varphi_1, \dots, \varphi_r$ linearly independent modulo constants. For z in a compact parameter neighborhood define

$$V_z(x) = \sum_i z_i \varphi_i(x), \qquad A(z) = \log \int_{\mathbb{T}^d} e^{V_z(x)}\,\mathrm{d}x, \qquad p_z(x) = e^{V_z(x) - A(z)}.$$

Let D bound $\sup_x V_z - \inf_x V_z$ uniformly. The ratio of maximum to minimum density is at most e^D. The torus Poincaré constant for normalized Haar measure is $(4\pi^2)^{-1}$, as follows by Fourier expansion. The comparison lemma therefore yields the valid uniform bound

$$C_{\mathrm{P}}(p_z) \le e^D/(4\pi^2).$$

The normalizing constant need not be known to sample exactly by rejection from Haar measure: accept a proposal x with probability $e^{V_z(x) - M_z}$, where $M_z \ge \sup_x V_z(x)$ is a valid bound. This generates p_z and has finite expected cost. It does not imply a favorable dimension scaling of that cost.

The parameter sources are

$$b_i = p_z\big(\varphi_i - \mathbb{E}_{p_z}\varphi_i\big).$$

They are smooth, centered, bounded, and define Dirichlet-dual functionals. If a source combination vanishes, then the corresponding feature combination is constant. Independence modulo constants therefore makes the true pullback positive definite. The same argument proves injectivity of the distribution parameterization: equality of two densities makes the difference of their log potentials constant.

R.2 Smoothness of the metric

Fix the Haar-mean-zero H^1 space as a gauge for constants. The forms $a_z(u, v) = \int p_z \nabla u \cdot \nabla v$ are uniformly coercive there, with smoothly parameter-dependent bounded coefficients. The sources depend smoothly on z in its dual. Inverse-operator differentiation,

$$D(A_z^{-1})[h] = -A_z^{-1}(DA_z[h])A_z^{-1},$$

followed inductively gives smooth dependence of the potential in H^1. Here A_z denotes the elliptic form operator, distinct from the log-partition scalar above. The metric Gram entries are therefore smooth. Their positive definiteness and compactness of a sufficiently small closed parameter set give uniform intrinsic upper and lower spectral bounds.

Smooth elliptic regularity on the torus gives smooth potentials for these smooth sources and coefficients (Evans, 2010). This assumption is used only to justify graph-norm approximation, not to define the true metric in the first place.

R.3 Why a Fourier Galerkin sequence meets the residual requirement

Let V_N contain mean-zero trigonometric polynomials with frequencies bounded by N in each coordinate. Let ϕ be an exact smooth potential, ϕ_N its Galerkin approximation, and $P_N\phi$ its Fourier truncation. Uniform ellipticity gives Céa's estimate

$$\|\phi - \phi_N\|_{H^1} \le C\|\phi - P_N\phi\|_{H^1}.$$

On V_N, the inverse inequality $\|v\|_{H^2} \le CN\|v\|_{H^1}$ follows termwise from Fourier coefficients. Thus

$$\|\phi_N - P_N\phi\|_{H^2} \le CN\|\phi_N - P_N\phi\|_{H^1} \le C'N\|\phi - P_N\phi\|_{H^1}.$$

For $\phi \in H^s$ with $s > 2$, the last expression is $O(N^{-(s-2)})$. The H^2 truncation error also tends to zero, so $\phi_N \to \phi$ in H^2. Bounded smooth coefficients imply that the strong weighted-Poisson residual tends to zero in L^2. For finitely many coordinate sources, this convergence is simultaneous. Consequently the graph-norm termination hypothesis of Proposition G.1 holds for this concrete spectral hierarchy.

This argument includes the missing bridge between energy approximation and strong-residual approximation. Density of V_N in H^1 alone would not supply it. It also does not assign a universal rate to arbitrary neural trial potentials.

R.4 A locally convergent fitting problem

Choose a target parameter z_*. For $f(z) = \mathrm{KL}(p_z \| p_{z_*})$,

$$f(z) = (z - z_*)^T \nabla A(z) - A(z) + A(z_*), \qquad \nabla f(z) = \nabla^2 A(z)(z - z_*).$$

At the target, $\nabla^2 f(z_*) = \mathrm{Cov}_{p_{z_*}}(\varphi)$, positive definite by feature independence. Therefore f is locally strongly convex in Euclidean coordinates and locally PL in the smooth true transport metric. Ordinary coordinate retractions on a small convex chart have the required upper Taylor bound. The certified algorithm has local linear convergence once the residual and integration certificates meet their tolerances and the initial point satisfies the trapping condition.

The main-text objective can instead use a supplied target moment vector t:

$$f_t(z) = A(z) - t^T z, \qquad \nabla f_t(z) = \mathbb{E}_{p_z}\varphi - t, \qquad \nabla^2 f_t(z) = \mathrm{Cov}_{p_z}(\varphi).$$

If $t = \nabla A(z_*)$ at an interior parameter, z_* is a strict local minimizer, and the same PL and certificate analysis applies. The inverse-function theorem applied to ∇A supplies an open neighborhood of moment vectors with interior solutions. Thus an empirical feature vector sufficiently close to the

population vector also defines a locally regular fitting problem. Conditional on the target data, that vector is fixed; uncertainty in its relationship to a population law is distinct from validation error when integrating against the current model. We do not claim that every possible empirical vector admits an interior exponential-family fit. Boundary moment vectors are outside this local theorem.

A redundant map $z = \pi(\theta)$ with supplied full-row-rank derivative adds a regular gauge handled by the quotient theorem. A power parameterization $z_i = a_i^{m_i}$ at zero adds a genuine rank loss of that parameter differential. The latter is handled only after the resolving map is explicitly used; a numerical pseudoinverse does not infer it.

The model is a density-based generative model rather than an arbitrary implicit neural network. Positive conditional densities also permit a triangular inverse-CDF sampler on a cube representation of the torus. Conditional inverses exist because the conditional densities are positive. Differentiating these inverses on a compact regular parameter neighborhood supplies a differentiable generator representation, subject to a consistent treatment of the periodic boundary. The metric statement itself can be formulated directly from the density sources, independent of which sampler is used.

R.5 Gaussian covariance factors and exact potentials

Let $B \in \mathbb{R}^{d\times k}$ have full row rank, $k \geq d$, and set $\Sigma = BB^T$. The generator BZ_k, $Z_k \sim \mathcal{N}(0, I_k)$, has law $\mathcal{N}(0, \Sigma)$. For a factor variation $\dot{B}$,

$$\dot{\Sigma} = \dot{B}B^T + B\dot{B}^T.$$

The map to positive-definite covariance matrices is a submersion: a prescribed symmetric $\dot{\Sigma}$ is produced by $\dot{B} = \frac{1}{2}\dot{\Sigma}\Sigma^{-1}B$.

There is a unique symmetric solution S of the Sylvester equation

$$S\Sigma + \Sigma S = \dot{\Sigma}. \tag{R.1}$$

In a covariance eigenbasis, its entries are $S_{ij} = \dot{\Sigma}_{ij}/(\lambda_i + \lambda_j)$. The gradient field $v(x) = Sx$ has potential $x^T Sx/2$ and induces exactly the covariance derivative in (R.1). Direct differentiation of the Gaussian density verifies the continuity equation. Since it is a gradient, it is the minimum-energy velocity. The true metric is

$$\|v\|^2_{L^2(\mu)} = \operatorname{tr}(S\Sigma S) = \tfrac{1}{2}\operatorname{tr}(S\dot{\Sigma}) = \frac{1}{2}\sum_{i,j}\frac{\dot{\Sigma}^2_{ij}}{\lambda_i + \lambda_j}. \tag{R.2}$$

These are the classical Bures–Wasserstein formulas (Takatsu, 2011); their inclusion gives an exact calibration case for our definitions.

The Frobenius-horizontal factor directions are SB with S symmetric. They are orthogonal to every D with $DB^T + BD^T = 0$, because $\langle SB, D\rangle_F = \frac{1}{2}\operatorname{tr}(S(DB^T + BD^T))$. They have squared norm $\|SB\|^2_F = \operatorname{tr}(S\Sigma S)$, exactly the true metric. General factor directions have additional gauge motion and can have a larger common-noise energy.

For a small step with $I + \alpha S$ positive definite,

$$B^+ = (I + \alpha S)B, \qquad \Sigma^+ = (I + \alpha S)\Sigma(I + \alpha S).$$

The affine map $x \mapsto (I + \alpha S)x$ is a gradient of a convex quadratic, so its displacement interpolation is the Gaussian Wasserstein geodesic. The factor step is therefore a compatible lift in this special case. Replacing B by BO, $O \in O(k)$, changes B^+ by the same right rotation and leaves the distribution update fixed. The formula does not depend on latent width except through the factor representation, but its intrinsic conditioning still depends on the covariance eigenvalues.

Quadratic trial potentials span every tangent potential in this model. With exact integration, the residual matrix is zero and the certificate is exact. This shows both the potential strength and the proper limitation of the algorithm: some models admit a finite exact true-metric representation; others require a nontrivial certified PDE approximation.

S CONDITIONAL CONSISTENCY WITH AN AMBIENT WASSERSTEIN FLOW

A natural-gradient method on a model family follows the projection of an ambient gradient, not necessarily the full ambient gradient. This appendix separates model approximation from metric approximation and time discretization. The comparison is deliberately conditional on regular flow existence and Lipschitz assumptions; it is not asserted across arbitrary collision seams.

S.1 THE PROJECTED VELOCITY ERROR

Let the ambient descent velocity at ν be $b_\nu = -\xi_\nu$, and let P_ν denote the orthogonal projection in $L^2(\nu)$ onto the represented model tangent range. The exact model natural velocity is $v_\nu = P_\nu b_\nu$. A certified metric gives a velocity $\widehat{v}_\nu$ in the same range with

$$\|\widehat{v}_\nu - v_\nu\|_{L^2(\nu)} \le \eta \|v_\nu\|_{L^2(\nu)}.$$

The model residual $(I - P_\nu)b_\nu$ is orthogonal to the certificate error, hence

$$\|\widehat{v}_\nu - b_\nu\|^2_{L^2(\nu)} = \|(I - P_\nu)b_\nu\|^2_{L^2(\nu)} + \|\widehat{v}_\nu - v_\nu\|^2_{L^2(\nu)}. \tag{S.1}$$

A more accurate Poisson solve reduces only the second term. The first can remain nonzero even with an exact metric and arbitrarily small time steps. A separate gradient-oracle error can be added by the triangle inequality or bounded by its own projection decomposition when available.

S.2 A COUPLING ESTIMATE FOR PERTURBED CONTINUITY EQUATIONS

Assume $b(\mu, x)$ satisfies

$$|b(\mu, x) - b(\nu, y)| \le L_x|x - y| + L_\mu W_2(\mu, \nu)$$

on the distributions and points under consideration. Suppose exact and approximate curves admit the characteristic representations

$$\dot{X}_t = b(\mu_t, X_t), \quad \mathrm{Law}(X_t) = \mu_t, \qquad \dot{Y}_t = b(\nu_t, Y_t) + r_t(Y_t), \quad \mathrm{Law}(Y_t) = \nu_t,$$

with square-integrable velocities and $\|r_t\|_{L^2(\nu_t)} \le R_t$. Existence of these flows, or an equivalent admissible coupling construction, is part of the assumption.

Proposition S.1 (Finite-time distribution comparison)**.** *With $L = L_x + L_\mu$,*

$$W_2(\mu_t, \nu_t) \le e^{Lt} W_2(\mu_0, \nu_0) + \int_0^t e^{L(t-s)} R_s \, \mathrm{d}s.$$

Proof. Initialize the two characteristic systems under an optimal coupling of their initial laws. Let $D_t = (\mathbb{E}|X_t - Y_t|^2)^{1/2}$. Differentiating its square and using Cauchy–Schwarz gives its upper differential inequality $D_t' \le L_x D_t + L_\mu W_2(\mu_t, \nu_t) + R_t \le L D_t + R_t$. At $D_t = 0$ use a regularized square root and pass to the limit, or the upper Dini derivative. Grönwall's inequality proves the bound for D_t. Since the joint law is a coupling, $W_2(\mu_t, \nu_t) \le D_t$. □

For the certified model curve, (S.1) supplies

$$R_t \le \left(\|(I - P_{\nu_t})b_{\nu_t}\|^2 + \eta_t^2 \|P_{\nu_t} b_{\nu_t}\|^2\right)^{1/2}.$$

This is a consistency statement over a fixed finite horizon, not a long-time convergence claim when the Lipschitz factor can grow exponentially.

S.3 A CONDITIONAL DISCRETE VERSION

Suppose the exact ambient flow map Φ_h is e^{Lh}-Lipschitz in W_2, and a regular model retraction step satisfies a one-step comparison

$$W_2(\nu_{k+1}, \Phi_h(\nu_k)) \le h e^{Lh} R_k + C h^2.$$

The R_k term bounds model and metric velocity error over the step; the Ch^2 term is a separately established local integration defect. It can be obtained from a C^2 compatible retraction and locally Lipschitz regular velocity field with bounded derivatives. It is not automatic from consistency of first derivatives alone.

Writing $e_k = W_2(\nu_k, \mu_{kh})$, the triangle inequality gives

$$e_{k+1} \le e^{Lh} e_k + h e^{Lh} R_k + Ch^2.$$

If $R_k \le R$ and $Nh = T$, summing the geometric series yields, for $L > 0$,

$$e_N \le e^{LT} e_0 + e^{Lh} R \frac{e^{LT} - 1}{L} + Ch \frac{e^{LT} - 1}{L}.$$

For $L = 0$, the corresponding bound is $e_0 + TR + CTh$. Thus vanishing metric error and time step do not recover the ambient flow unless the model residual also vanishes or is controlled to the desired accuracy.

S.4 Why the seam is excluded from this flow theorem

A C^1 quantile curve with a Hölder-continuous derivative can support the discrete contraction proof of Appendix J. It need not generate a locally Lipschitz natural-gradient vector field for an arbitrary target. Nor does it automatically yield an $O(h^2)$ local defect for an additive chart update. The continuous-time and discrete-time arguments therefore have different hypotheses. Conflating them would falsely turn a valid first-derivative contraction theorem into an unproved high-order flow approximation theorem.

The model-limited nature of this comparison is closely related to existing analyses of parameterized Wasserstein flows (Jin et al., 2025). Our contribution is the explicit insertion of a true-metric residual certificate and its orthogonal separation from the model tangent defect, not the invention of Grönwall coupling estimates or of projected gradient flows.

T Counterexamples to stronger interpretations

The following examples make the boundaries of the results testable. They are not empirical counterexamples and do not depend on numerical tolerances.

T.1 A common-noise matrix can charge a pure gauge

Let $T_B(z) = Bz$ with $Z \sim \mathcal{N}(0, I_k)$, and take $\dot{B} = B\Omega$ for skew-symmetric Ω. The exact path $B(t) = B\exp(t\Omega)$ has constant covariance BB^T and hence constant output distribution. Its true transport speed is zero. Its chosen latent velocity is $B\Omega Z$, whose squared norm is $\|B\Omega\|_F^2$, potentially nonzero. Therefore $\mathbb{E}[DT^T DT]$ is not the true pullback in general. The difference is precisely one of the projection losses in Appendix C.

T.2 An absolute cutoff destroys coordinate invariance

Consider the regular family $\mu_{a,b} = \mathcal{N}((\epsilon a, b), I_2)$ for $\epsilon > 0$. Its true raw metric is $\mathrm{diag}(\epsilon^2, 1)$, while the identifiable mean coordinates have identity metric. A fixed spectral cutoff exceeding ϵ^2 deletes the first direction, even though both mean directions are perfectly identifiable and equally conditioned intrinsically. Taking a target with nonzero first mean then makes the truncated update fail in that direction. This is a reparameterization effect, not a singular statistical model.

T.3 A residual test on the estimated range can miss a direction

Take

$$G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad C_{\mathrm{P}} = 1.$$

The exact bound $L \preceq G \preceq L + R$ holds. Restricted to $\mathrm{ran}\, L$, the residual vanishes, so a range-only test would accept with relative error zero. On the whole space it fails because $\ker L$ is not contained in $\ker R$. The missing second direction is an unresolved true tangent, not a gauge. This justifies the whole-space language in Theorem 3.1.

T.4 A positive trial gradient Gram need not lower-bound the metric

In a one-dimensional Dirichlet problem normalize the exact potential ϕ to energy one. Choose the trial potential $c\phi$. Its gradient Gram is c^2, which exceeds the true metric when $|c| > 1$. The valid lower matrix is instead

$$L = 2c - c^2 = 1 - (1 - c)^2 \leq 1.$$

For $c = 3$, this lower bound is -3, correctly signaling a poor approximation, whereas the trial gradient Gram is the misleading positive value nine. Positivity alone is not evidence of a correct metric approximation.

T.5 No exact kernel identification under unrestricted absolute noise

The matrices $G_0 = \mathrm{diag}(1, 0)$ and $G_\epsilon = \mathrm{diag}(1, \epsilon)$ have different kernels and operator-norm distance ϵ. The same observed matrix $\mathrm{diag}(1, \epsilon/2)$ is within error $\epsilon/2$ of both. Any deterministic rule receiving only that matrix and the absolute error bound must return the same answer in both cases and therefore cannot identify both kernels correctly. A randomized rule also cannot be certain in both cases. Exact nullspace recovery requires additional structure, a verified gap, or a known gauge; it does not follow from a small absolute matrix error.

T.6 The Moore–Penrose inverse cannot recreate a lost differential

Let $\mu_a = \mathcal{N}(a^m, 1)$ with $m \geq 2$ and target mean $z_* \neq 0$. Then

$$f'(a) = ma^{m-1}(a^m - z_*), \qquad G(a) = m^2 a^{2m-2}.$$

At zero both are zero, and the scalar Moore–Penrose inverse of zero is zero. Thus the raw natural-gradient update is stationary. The identifiable coordinate $z = a^m$ has gradient $z - z_*$ and metric one, and can move. The raw map is not a submersion at zero, so this does not contradict regular natural-gradient invariance. For even m, choose the target in the image and a one-sided branch; a signed extension requires additional model structure.

T.7 A finite raw additive step need not match a chart step

For $z = a^m$ and $a > 0$, the raw vector corresponding to a unit z direction scales as a^{1-m}. The exact compatible lift is $(a^m + ma^{m-1}w)^{1/m}$, while an additive raw step is $a + w$. Their second derivatives at $w = 0$ differ by $-(m - 1)/a$. No bound with a constant independent of distance to the collision can make their second-order difference uniformly controlled. A small step can still be valid at every fixed $a > 0$; the counterexample excludes a uniform conclusion, not local smoothness.

T.8 A value expansion does not control a derivative

For $m = 4$, set $Q(a) = a^4 + a^5 \sin(a^{-2})$ for $a > 0$ and $Q(0) = 0$. This is C^1 at zero and $Q(a) = a^4 + O(a^5)$, but

$$Q'(a) = 4a^3 + 5a^4 \sin(a^{-2}) - 2a^2 \cos(a^{-2}).$$

The oscillatory derivative dominates the nominal leading derivative along some sequences. After $\kappa = a^4$, the derivative of $Q(\kappa^{1/4})$ does not approach one and is unbounded along subsequences. Taking $\mathcal{N}(Q(a), 1)$ makes this a valid generator example. A density or quantile value expansion without a differentiated remainder is therefore insufficient for a resolving chart theorem.

T.9 Distance or value equivalence does not imply descent alignment

Let $F(x) = x^2$ and $\widetilde{F}(x) = x^2(1 + \frac{1}{2}\sin(10x))$. Their values satisfy $F/2 \leq \widetilde{F} \leq 3F/2$ everywhere. At $x = 3\pi/10$, however,

$$\widetilde{F}'(x) = 2x - 5x^2 < 0,$$

whereas $F'(x) > 0$. A sufficiently small negative-gradient step for $\widetilde{F}$ increases F. Thus replacing a transport objective by a quantitatively comparable surrogate requires an additional derivative-consistency argument if one wants descent of the original energy.

T.10 Moment balance alone is not a uniform algorithmic theorem

The collision order fixes a leading Hermite direction, but it does not fix all higher coefficients or uniform tail bounds. Families with the same first few moments can have very different higher moments and remainder constants. In particular, the contraction neighborhood, Poincaré constant, and cost of estimating high-order statistics need not be uniform over all such families. Our constructions impose bounded support or a bounded smooth density perturbation precisely to supply the needed control.

T.11 Angular degeneracy is not removed by radial resolution

At $\kappa = 0$, the axisymmetric model equals $\mathcal{N}(0, I_d)$ for every u. There is no data-dependent or geometric rule that can recover a unique true orientation from this law alone. The resolved radial tangent depends on a selected u, so identifying all such points with one smooth full-rank parameter chart would require additional structure not provided here. The special Euclidean limiting cone for symmetric $m = 4, d = 2$ does not change this observation about labels at its vertex.

T.12 Known formulas are not isolated novelty claims

The weighted Poisson definition of Wasserstein natural gradient, Gaussian Bures geometry, Gauss–Hermite quadrature, moment parameterizations, Gauss–Newton contraction mechanisms, and weak mixture identifiability all have prior literature. The paper's claims concern their particular certified quotient integration, all-order derivative-controlled collision constructions, and exact rotated-product metric reduction. The bibliography and comparison notes do not establish a universal priority guarantee; they delimit the specific comparisons made after the documented search.

U Jet-normalized certificates near a singular stratum

The product metric theorem identifies a vanishing angular coefficient, but the general certificate is a relative statement. Does relative certification necessarily become harder as that coefficient vanishes? For the analytically specified OU models, there is a useful distinction: the direction becomes intrinsically ill-conditioned, yet a structured potential can approximate its metric to *increasing relative accuracy*. This section proves that claim with quantified assumptions. It connects the collision construction to the residual-certified algorithm rather than treating them as unrelated observations.

U.1 A normalized residual lemma

Work in a Hilbert space with a symmetric Dirichlet form a and an operator $\mathcal{L} \succeq \lambda_0 I$, where $\lambda_0 > 0$ is known. The space may be the separately centered pair space of Appendix P. Let ϕ solve $\mathcal{L}\phi = b$, let ψ be an admissible trial function, and set

$$J = a(\phi, \phi), \qquad r = b - \mathcal{L}\psi, \qquad D = 2\langle b, \psi\rangle - a(\psi, \psi).$$

Suppose there is a known amplitude $s > 0$ and constants $c, C > 0$ for which

$$cs^2 \le a(\psi, \psi) \le Cs^2, \qquad \|r\| \le \epsilon s. \tag{U.1}$$

The amplitude may approach zero; c, C, λ_0 must not depend on it in the region under consideration.

Theorem U.1 (A vanishing metric with a uniform relative certificate)**.** *Under* (U.1)*, if*

$$\epsilon \le \frac{c\sqrt{\lambda_0}}{4\sqrt{C}},$$

then

$$D \ge \frac{c}{2}s^2 > 0, \qquad D \le J \le (1+\eta)D, \qquad \eta = \frac{2\epsilon^2}{\lambda_0 c}. \tag{U.2}$$

The bounds are independent of the size of s. They do not bound J^{-1} independently of s.

Proof. Since $b = \mathcal{L}\psi + r$, integration by parts gives

$$D = a(\psi, \psi) + 2\langle r, \psi\rangle.$$

Coercivity implies $\|\psi\| \le \sqrt{a(\psi,\psi)/\lambda_0}$, so

$$\begin{aligned} D &\ge a(\psi,\psi) - 2\|r\|\sqrt{a(\psi,\psi)/\lambda_0} \\ &\ge \big(c - 2\epsilon\sqrt{C/\lambda_0}\big)s^2 \ge \frac{c}{2}s^2. \end{aligned}$$

The residual identity gives $0 \le J - D = \langle r, \mathcal{L}^{-1}r\rangle \le \epsilon^2 s^2/\lambda_0$. Dividing by the lower bound for D proves (U.2). The final observation follows because the same assumptions allow J to be of order s^2; taking a reciprocal can therefore amplify a covector by order s^{-2}. □

The proof uses neither a Taylor approximation of an inverse nor an absolute eigenvalue cutoff. It is an a posteriori statement once the three quantities in (U.1) are rigorously bounded. If the bounds are estimated by ordinary floating-point computation without error control, the resulting number is a diagnostic rather than the theorem's certificate.

U.2 Application to the Hermite pair potential

For two OU marginals of orders $m_i, m_j \ge 3$, set

$$s = (\kappa_i^2 + \kappa_j^2)^{1/2}, \quad V_i = H_{m_i-1}/m_i!, \quad V_j = H_{m_j-1}/m_j!, \quad \psi^0 = \kappa_i V_i(x_i)x_j - \kappa_j V_j(x_j)x_i.$$

This is the same polynomial potential used to identify the exact leading coefficient in Appendix P. It satisfies the necessary separate-centering conditions exactly, not approximately. Its energy is

$$a(\psi^0,\psi^0) = \kappa_i^2\mathbb{E}_{p_i}(V_i'^2 + V_i^2) + \kappa_j^2\mathbb{E}_{p_j}(V_j'^2 + V_j^2). \tag{U.3}$$

The cross terms vanish because the matched lower moments imply $\mathbb{E}V_i' = \mathbb{E}V_j' = 0$. The moment terms converge to $1/m_i!$ and $1/m_j!$. In a sufficiently small fixed neighborhood, for example, one can take

$$c = \tfrac{1}{2}\min\{1/m_i!, 1/m_j!\}, \qquad C = 2\max\{1/m_i!, 1/m_j!\}$$

provided that neighborhood has been chosen so the asserted moment bounds hold. Moment convergence proves existence of such a neighborhood; an implementation must certify its actual moment bounds rather than assume that any user-chosen radius works.

Let $\beta_i = 1/m_i$ and $\beta_j = 1/m_j$, improved to $2/m_i$ and $2/m_j$ for symmetric constructions. The pointwise polynomial remainders in the quantile proofs, followed by inverse-quantile substitution, imply, with the polynomial weights needed here,

$$\|v_i - V_i\|_{L^2(p_i;\,\mathrm{poly})} = O(|\kappa_i|^{\beta_i}), \qquad \|v_j - V_j\|_{L^2(p_j;\,\mathrm{poly})} = O(|\kappa_j|^{\beta_j}).$$

The notation means any one of the finitely many fixed polynomial weights occurring in the following residual, with its own finite constant. It is not a bound over all polynomials at once.

Expanding the operator on ψ^0 and using $s_i + x_i = m_i\kappa_i v_i$ gives

$$\begin{aligned} r = {}& m_i\kappa_i(v_i - V_i)x_j - m_j\kappa_j(v_j - V_j)x_i \\ &+ \kappa_i\big[(s_i + x_i)V_i'x_j + V_i(s_j + x_j)\big] \\ &- \kappa_j\big[(s_j + x_j)V_j'x_i + V_j(s_i + x_i)\big]. \end{aligned} \tag{U.4}$$

Uniform moments and independence bound each product term. Consequently, for a fixed sufficiently small neighborhood, there is a finite constant C_0 such that

$$\|r\|_{L^2(p_i\otimes p_j)} \le C_0 s\big(|\kappa_i|^{\beta_i} + |\kappa_j|^{\beta_j}\big). \tag{U.5}$$

The terms linear in $|\kappa_i| + |\kappa_j|$ are absorbed because $0 < \beta_i, \beta_j \le 1$ and the cumulants have magnitude at most one. This estimate is uniform over relative approaches to the origin, including approaches along either axis. At an exactly zero coordinate, its source and potential contribution vanish analytically.

The marginal constructions have uniform positive Poincaré gaps near zero, giving a fixed lower bound λ_0 for $\lambda_i + \lambda_j$. Theorem U.1 now implies the following relative statement.

Corollary U.2 (A Hermite certificate that improves toward the collision)**.** *For the verified OU families above and $(\kappa_i, \kappa_j) \neq (0, 0)$ sufficiently close to zero, the polynomial trial ψ^0 yields a positive lower bound D_{ij} and*

$$D_{ij} \leq J_{ij} \leq (1 + \eta_{ij}) D_{ij}, \qquad \eta_{ij} = O\Big(\big(|\kappa_i|^{\beta_i} + |\kappa_j|^{\beta_j} \big)^2 \Big).$$

At $(0, 0)$ both the exact source and the trial potential vanish, so $J_{ij} = D_{ij} = 0$. This zero is supplied by the model, not inferred by thresholding a small sample covariance.

The constants can depend on the collision orders, mixing laws, moment bounds, and the certified spectral gap. There is no guarantee uniform as the number of matched moments tends to infinity. In particular, factorial normalization must not be suppressed when comparing different orders.

U.3 A HYBRID STRUCTURED-POTENTIAL ALGORITHM

On a rotated-product model with known marginal family labels, the following rule complements RCQ-NG. Near a prescribed Gaussian stratum, first try the explicit pair polynomial ψ^0. Compute or enclose its lower bound and its normalized residual. If the relative certificate passes, use D_{ij} as the pair metric. If it fails, enrich the pair potential by a Galerkin correction and recheck. Away from the stratum, use the same general residual procedure without relying on the asymptotic trial. Radial blocks are computed from the one-dimensional quantile formula or certified independently.

For each accepted pair the step error satisfies Theorem 3.1. Orthogonality of the product-frame blocks makes the maximum of their relative tolerances a simultaneous tolerance for the full active angular block. A congruence transfers that certificate to any redundant representation. A pair known to be exactly Gaussian is a certified zero block; no solve or inversion is needed there.

Analytic normalization is important in numerical validation. With $s > 0$, write $\bar{\psi} = \psi^0/s$, $\bar{b} = b/s$, and $\bar{r} = r/s$. The normalized integrals can be evaluated directly, retaining the exact common factors rather than subtracting two unscaled numbers of order s^2. Once bounds for them are established, multiply the resulting matrix interval by s^2. This algebra preserves the certificate and its structural zero. It does not magically validate noisy estimates of κ_i, κ_j, or the density itself; uncertainties in these model inputs require their own enclosure.

U.4 WHY THIS DOES NOT REMOVE ANGULAR ILL-CONDITIONING

As a pair approaches the Gaussian, the true metric is of order s^2 even though the relative error above tends to zero. An accurate inverse still has size of order s^{-2}. Retraction radii, admissible rotation angles, energy Hessian conditioning, and statistical information can therefore force small steps or large sample sizes. The new certificate avoids mistaking *approximation failure* for that intrinsic effect, but does not remove it.

For example, with a pair of order-four symmetric marginals, the explicit relative error scales at most as $O((|\kappa_i|^{1/2} + |\kappa_j|^{1/2})^2)$ in the bound above, while $J_{ij} \sim (\kappa_i^2 + \kappa_j^2)/24$. These are different powers with different meanings. The former describes how accurately a known transport problem can be approximated by a structured potential; the latter describes how weakly the distribution responds to a rotation. The statistical lower bounds remain unaffected.

V PROOF DEPENDENCIES, COMPARISON LEDGER, AND EXECUTABLE CHECKS

The paper is theoretical. The accompanying code checks selected finite algebraic identities and implements the certificate acceptance rule for exact rational matrix data. It does not constitute a formal proof of the functional-analytic, asymptotic, or convergence theorems, and no machine-learning benchmark or empirical performance result is reported.

V.1 DEPENDENCY MAP FOR THE PRINCIPAL CONCLUSIONS

Conclusion	Necessary inputs and proof route
True pullback	L^2 generator differentiation; conditional expectation; Hilbert projection onto gradients. Appendices B–C.
Certified metric bracket	Poincaré bound; admissible potentials; full residual Gram; exact variational identity. Appendix E.
Finite-sample acceptance	Known quotient; independent validation; bounded integrands or separate valid intervals. Appendix F.
Finite-step invariance	Supplied submersion; corresponding intrinsic certificate data; compatible retraction and step-selection rule. Appendices D, H.
Local linear rate	Retraction smoothness; intrinsic relative oracle errors; PL; separate trapping condition. Appendix I.
Higher-order chart	A derivative-controlled quantile jet, proved from the actual OU generator. Appendices J–K.
Finite-mixture hierarchy	Positive Gauss–Hermite weights; exact first missing moment; fixed-family derivative bounds. Appendix L.
Signed continuation	A constructed opposite-sign source law; matching first transport tangent; no unproved C^2 seam. Appendix M.
Mixed-order ravine	Profiled fast-variable stationary graph; positive normal Hessian; explicit power map. Appendix N.
Angular and pair metric	Positive smooth standardized sources; tensor centering; coercive Poisson problems; verified uniform bounds for asymptotics. Appendices O–P.
Normalized near-stratum certificate	Analytic source factorization; uniform pair gap; residual of size $O(s\tau)$; positive normalized trial energy. Appendix U.
Raw statistical rate	Independent data; local density-ratio expansion; known fixed branch for the upper bound. Appendix Q.
Ambient-flow comparison	A regular flow representation, Lipschitz comparison, and separately established local discretization defect. Appendix S.

These dependencies matter operationally. For example, the jet contraction theorem does not require the retraction-smoothness assumption, while the regular flow comparison does. The statistical lower bound does not require the optimization algorithm to fail. A finite-dimensional certificate acceptance check cannot establish a missing Poincaré bound or validate a density source supplied incorrectly.

V.2 Prior-work comparison

The foundations of Wasserstein gradient flows are treated in Ambrosio et al. (2008); natural gradients and manifold retractions are established tools (ichi Amari, 1998; Absil et al., 2008). Continuous-density Wasserstein natural gradients and explicit one-dimensional information metrics are already developed by Chen & Li (2020). The present projection definitions and Gaussian calibration are consistent with those foundations rather than alternatives to them.

van Oostrum et al. (2023) distinguish parameter gradients from their pushforwards in overparameterized systems. Our corresponding finite-step result requires a common intrinsic retraction and congruent approximation data, which is stronger information than merely choosing a pseudoinverse in each parameterization. Nurbekyan et al. (2023) formulate efficient natural-gradient computations as least-squares problems. Our lower-matrix identity and full residual majorant address a different question: when do approximate true transport potentials certify the geometry, including unresolved null directions? The scalar functional-error principle is classical (Repin, 2008).

Jin et al. (2025) analyze parameterized Wasserstein flows and a computationally convenient pullback. We do not claim that parameterized flow error analysis was previously absent. Equation (S.1) separates the orthogonal model defect from our true-metric certificate error under explicit regular-flow assumptions. ichi Amari & Matsuda (2024) study Wasserstein shape and affine information geometry; Li & Zhao (2023) study mixture metric scaling in small-variance regimes. The finite quadrature construction here instead keeps component variances near one while means coalesce under increasingly high moment cancellation.

Moment-centered parameterizations, moment identification, and moment estimators for mixtures have prior literature (Kamary et al., 2018; Agostini et al., 2021; Wu & Yang, 2020). Hermite-based non-Gaussianity is also familiar from independent component analysis (Hyvärinen & Oja, 2000). Recent optimal-transport ICA studies (Laplante et al., 2026; Jha et al., 2026) optimize Wasserstein non-Gaussianity criteria. Theorem 6.1 computes the true metric of a specified rotated product generator, including its exact rotational zero set and pairwise Poisson certificates. These are distinct mathematical objects even when both are used in source-separation problems.

Zenker (2026) studies cumulant perturbations of Gaussian optimal transport through an Ornstein–Uhlenbeck linearization. Consequently neither a cumulant perturbation nor a Gaussian Hermite transport calculation is a defensible isolated novelty claim here. The all-order bounded-generator quantile proof, signed continuation, residual framework, and coupled product calculation are the specific results established in this manuscript. A literature search cannot certify that no related result exists anywhere; the comparisons refer to the documented sources rather than a universal priority assertion.

V.3 What the executable checks verify

The file `code/theory_checks.py` uses exact symbolic arithmetic to verify Hermite norms and Gaussian moments, the first missing quadrature coefficient, the leading constants in the mixture hierarchy, the root-retraction series, rational matrix completion-of-squares identities, and rectangular quotient identities. It also checks exact counterexamples for range-only certificates and false trial-Gram lower bounds. The source contains each input rather than depending on an external dataset.

The file `code/reference_algorithms.py` supplies a small exact-rational implementation of the algebraic certificate test. Its inputs are a proposed lower matrix, a proposed valid residual majorant, and a rational relative tolerance. It verifies symmetry, positive semidefiniteness, the whole-space relative comparison, and covector compatibility before returning a pseudoinverse direction. It can return `unresolved`. It assumes that the supplied matrices really arise from a valid analytic or validated numerical bracket; code cannot infer that fact from their entries alone.

The finite tests are regression checks for implementation and constants, not empirical experiments supporting general convergence. The proofs above remain the mathematical evidence. In particular, no numerical test replaces the quantile derivative envelope, the conditional validation argument, or the tensor-space elliptic proof.

V.4 Rebuilding and author verification

The source archive contains the manuscript, bibliography, style files, build script, identity checks, and source-provenance notes. The root `main.tex` builds the anonymous version; the separate author archive also contains `author_version.tex`. The proof guide uses dynamically resolved page references, so it remains consistent after compilation. The release audit records page boundaries, unresolved-reference checks, style checksums, and the exact-check output.

This manuscript was developed with substantial AI assistance, as disclosed in the main document. Symbolic identities and compilation checks do not establish independent peer review. Before submission, the named authors must verify the statements and proofs, confirm authorship and affiliations, approve the disclosure and source archive, and satisfy the conference's administrative rules. The archive does not assert that those approvals have already occurred or that the paper has been accepted.