STRUCTURAL COMPATIBILITY ENABLES EXACT SOLVABILITY IN WASSERSTEIN PROXIMAL DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Finite-dimensional Wasserstein updates replace an optimization over probability measures by one over a structured family. We ask when this restriction is mathematically lossless rather than merely convenient. The key property is structural compatibility: on every statistic fiber, the same canonical representative must minimize the energy and saturate the Wasserstein transport lower bound. We show that this compatibility is equivalent to all-step fiberwise exactness. For mean–covariance fibers, this yields full-space finite-step Gaussian and matched q-Gaussian closure for Shannon and power entropies under quadratic confinement. Our main converse result is a rigidity theorem: within regular fixed-generator radial location–scatter families, all-step exactness forces the internal energy into the Shannon/positive-power class and fixes the radial generator. When exactness fails, the normal Wasserstein gradient gives the first-order distance to the family in one-dimensional quantile geometry; a strongly convex counterexample shows that vanishing first-order defect still does not imply finite-step closure. The associated d-dimensional radial defect is directly estimable. Across 375 frozen Monte Carlo configurations with $d \leq 100$, every mismatch has positive estimated defect, matched $p = 1$ cases are numerically zero, the mean relative error at $d = 100, n = 10^5$ is 0.17%, and all six preregistered bootstrap intervals contain the analytic target. Together, the results separate exact finite-dimensional solvability, infinitesimal compatibility, and structurally misspecified approximation.

1 INTRODUCTION

Finite-dimensional Wasserstein updates are ubiquitous because optimizing directly over probability measures is costly. Gaussian/Bures variational inference is the canonical example: a full distributional proximal problem is replaced by an update of a mean and covariance. Yet tractability alone does not tell us whether the restricted update solves the original variational problem. Two cases must be separated. The family may contain the unrestricted minimizer, so the restriction is lossless, or it may define a genuinely different constrained problem. We call the first situation structural compatibility between the Wasserstein geometry, the energy, and the chosen family.

For an energy $\mathcal{F}$, the unrestricted Wasserstein proximal step is

$$\mathsf{J}_{\tau,\mathcal{F}}(\nu) \in \operatorname*{arg\,min}_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \mathcal{F}(\mu) + \frac{1}{2\tau} W_2^2(\mu, \nu) \right\}. \tag{1}$$

The central question is therefore:

> *When does a finite-dimensional family solve exactly the same Wasserstein proximal problem as the full probability space, and what error remains when it does not?*

This distinction is technically meaningful. Solver error can be reduced by tighter tolerances or more computation. A family error cannot: if the unrestricted minimizer leaves the chosen family, optimizing that family more accurately still returns the wrong variational object. Conversely, an exact family is an oracle reduction of the infinite-dimensional step. Figure 1 shows the distinction and the local defect used below.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

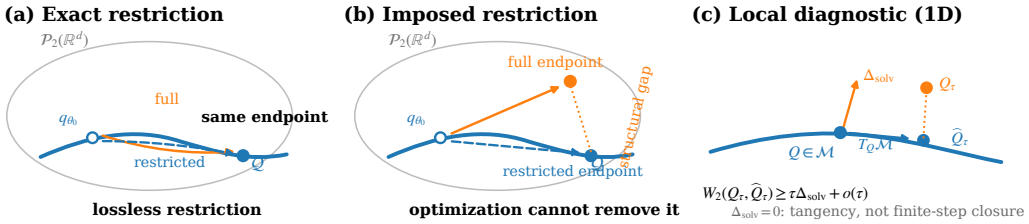


Figure 1: **Exact versus imposed restriction.** (a) Structural compatibility makes the unrestricted and family-restricted proximal steps coincide. (b) Under misspecification the endpoints differ, creating a representation error that cannot be removed by more accurate optimization within the same family. (c) In one-dimensional quantile geometry, the normal component Δ_{solv} gives the first-order family floor. Arrows denote alternative one-step updates, not continuous trajectories.

The paper develops this boundary in three steps. Section 2 gives a checkable all-step mechanism for exact reduction and a concrete d-dimensional Gaussian/q-Gaussian realization. Section 3 reverses the logic: within a regular radial fixed-generator class, all-step exactness is so restrictive that it classifies the admissible energy and generator. Section 4 quantifies failure through a normal-gradient defect and separates first-order tangency from genuine finite-step closure.

Main contributions.

- **Exactness decomposes into two independent fiberwise requirements.** All-step canonical replacement is equivalent to energy domination plus transport domination on each statistic fiber (Theorem 1). The two terms cannot hide behind a cancellation at one selected step size.
- **A rigidity classification is the central theoretical result.** In the stated regular radial location–scatter class, an all-step exact mean–covariance section forces the internal energy into the Shannon/positive-power class and forces the radial generator to be the corresponding generalized exponential (Theorem 3). For Shannon and $m > 1$ power entropies, Theorem 2 supplies the matching full-space exact branches.
- **Mismatch has an optimizer-independent first-order cost.** In one-dimensional quantile geometry, the distance from the unrestricted step to a smooth finite-dimensional family is $\tau\Delta_{\mathrm{solv}} + o(\tau)$ (Theorem 4); every family-constrained output is bounded below by the same quantity. Proposition 1 shows why $\Delta_{\mathrm{solv}} = 0$ is only a tangency statement, not an exactness certificate.

Relation to prior work. Gaussian/Bures methods analyze dynamics after restricting the state space to Gaussian measures (Lambert et al., 2022; Diao et al., 2023); their guarantees concern the resulting constrained geometry. Gelbrich's inequality gives the sharp mean–covariance transport bound (Gelbrich, 1990), while porous-medium and q-Gaussian invariances establish preservation for specified energy–family pairs (Carlen and Craig, 2013; Takatsu, 2012; 2013; Johnson and Vignat, 2007; Duong, 2015). Kim et al. (Kim et al., 2026) obtain a finite-dimensional reduction for a continuous Wasserstein gradient flow whose generalized-variance functional depends only on covariance. We ask a different finite-step question: when is a structured update exactly equal to the *unrestricted* resolvent, and what does that equality force about the energy and family? Thus Gaussian/q-Gaussian structure is not claimed as new; the contribution is the exactness criterion, its converse rigidity, and the structural error boundary when exactness fails.

2 Structural Compatibility and Exact Reduction

Let Θ denote the statistic space, let $S : \mathcal{P}_2(\mathbb{R}^d) \to \Theta$ be a statistic, and let $\theta \mapsto q_\theta$ be a chosen canonical section satisfying $S(q_\theta) = \theta$. For $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $q_{S\mu}$ denotes the selected representative of

the statistic fiber containing μ. For a canonical reference q_{θ_0} define

$$\mathcal{L}_{\tau,\theta_0}(\mu) := \mathcal{F}(\mu) + \frac{1}{2\tau} W_2^2(\mu, q_{\theta_0}).$$

The pointwise comparison between μ and $q_{S\mu}$ is stronger than merely asking where the global minimizer happens to lie.

Definition 1 (All-step fiberwise exactness)**.** The canonical section is all-step fiberwise exact if, for every μ, every canonical reference q_{θ_0}, and every $\tau > 0$,

$$\mathcal{L}_{\tau,\theta_0}(q_{S\mu}) \le \mathcal{L}_{\tau,\theta_0}(\mu). \tag{2}$$

Theorem 1 (Exactness decomposition)**.** *Definition 1 holds if and only if, for every μ and canonical reference q_{θ_0},*

$$\mathcal{F}(\mu) \ge \mathcal{F}(q_{S\mu}), \tag{3}$$

$$W_2^2(\mu, q_{\theta_0}) \ge W_2^2(q_{S\mu}, q_{\theta_0}). \tag{4}$$

Consequently, for every $\tau > 0$,

$$\inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{L}_{\tau,\theta_0}(\mu) = \inf_{\theta \in \Theta} \left\{ \mathcal{F}(q_\theta) + \frac{W_2^2(q_\theta, q_{\theta_0})}{2\tau} \right\}. \tag{5}$$

If (3) *is strict away from the canonical section, every full-space minimizer is canonical.*

Interpretation. Exactness for every relative weight between energy and transport forces the two preferences to agree separately. The large-τ limit isolates the energy comparison; the small-τ limit isolates the transport comparison. This is why the theorem is stronger than a statement at a single step size. It characterizes a fiber-checkable mechanism for exact reduction, not every conceivable reason a global minimizer could lie in a family. The full proof, including the quantifiers behind (5), is in Appendix A.1.

2.1 A FULL-SPACE d-DIMENSIONAL REALIZATION

For $m > 1$ and $m = 1$ let

$$U_m(\rho) = \frac{1}{m-1} \int \rho^m, \qquad U_1(\rho) = \int \rho \log \rho, \tag{6}$$

and let $V(x) = \frac{1}{2} x^\top H x + h^\top x + c$ with $H \succeq 0$ and $h \in \mathrm{Range}(H)$, so that V is bounded below. Denote by $\mathcal{Q}_m$ the location–scatter orbit of the covariance-normalized Rényi extremizer

$$B_m(x) = Z_{m,d}^{-1} [\alpha_m - \beta_m \|x\|^2]_+^{1/(m-1)}, \qquad \mathrm{Cov}(B_m) = I, \tag{7}$$

with $\mathcal{Q}_1$ the Gaussian family (Johnson and Vignat, 2007).

Gelbrich's inequality gives, for any law with mean a and covariance Σ,

$$W_2^2(\mu, q_{a_0,\Sigma_0}) \ge \|a - a_0\|^2 + d_B^2(\Sigma, \Sigma_0), \tag{8}$$

where equality is attained by the canonical member $q_{a,\Sigma} \in \mathcal{Q}_m$. On the same moment fiber, $q_{a,\Sigma}$ minimizes the matched internal energy, and the quadratic potential is constant.

Theorem 2 (Exact affine Gaussian/q-Gaussian closure)**.** *Let $m \ge 1$, let V be a bounded-below convex quadratic, and let $\nu = q_{a_0,\Sigma_0} \in \mathcal{Q}_m$ with $\Sigma_0 \succ 0$. For every $\tau > 0$, the unrestricted JKO problem admits a minimizer and every minimizer belongs to $\mathcal{Q}_m$. For $m > 1$, writing $C_{m,d} := U_m(B_m) > 0$, it reduces exactly to*

$$\Psi_\tau(a, \Sigma) = C_{m,d} (\det \Sigma)^{-(m-1)/2} + \frac{1}{2} a^\top H a + h^\top a + \frac{1}{2} \mathrm{tr}(H\Sigma) + \frac{\|a - a_0\|^2 + d_B^2(\Sigma, \Sigma_0)}{2\tau}. \tag{9}$$

For $m = 1$, the determinant term is replaced by $-\frac{1}{2} \log \det \Sigma$ up to an additive constant.

Theorem 2 is not presented as a discovery of Gaussian or q-Gaussian geometry. Its role is to show that the two fiberwise equalities in Theorem 1 can be simultaneously sharp in the *unrestricted* problem. Existence and all equality conditions are proved in Appendix A.2. For $H = \kappa I$, the covariance commutes with Σ_0 and the reduced covariance step further collapses to scalar root equations; these formulas are recorded in Appendix A.3.

3 MAIN RIGIDITY THEOREM: EXACTNESS LEAVES VERY FEW CHOICES

The previous section is a forward construction. The stronger question is converse: if a fixed radial location–scatter family is all-step exact, what can its energy and generator be?

For $U_\Phi(\rho) = \int \Phi(\rho(x))dx$, assume $\Phi \in C^3((0,\infty))$, $\Phi'' > 0$, and a finite lower-semicontinuous extension to 0 with $\Phi(0) = 0$. Let $\mathcal{Q}_g$ be the regular location–scatter orbit generated by $q_g(x) = c_g g(\|x\|^2)$, with the interior density range containing $(0, M)$.

Assumption 1 (First-order resolvent limit)**.** *At the centered isotropic reference q_g, the selected unrestricted proximal branch is locally unique. Write $\nabla_W \mathcal{F}(q_g)$ for the corresponding $L^2(q_g)$ Wasserstein gradient. Its optimal displacement velocity satisfies*

$$v_\tau \longrightarrow -\nabla_W \mathcal{F}(q_g) \quad \text{in } L^2(q_g) \qquad (\tau \downarrow 0). \tag{10}$$

This is the only local resolvent regularity used by the converse argument; compact-support free boundaries require it to be verified on the branch under consideration.

Theorem 3 (Rigidity classification of all-step exact radial closures)**.** *Let V be quadratic and suppose the mean–covariance canonical section of $\mathcal{Q}_g$ is all-step fiberwise exact for $\mathcal{F}_{\Phi,V} = U_\Phi + \int V\rho$. Assume the regularity above and Assumption 1. Then, up to normalization, radial rescaling, and addition of the irrelevant linear mass term bt, the internal energy is necessarily one of*

$$\Phi(t) = a\,t\log t \qquad \text{or} \qquad \Phi(t) = a\,t^m, \quad m > 0,\ m \neq 1, \tag{11}$$

with the coefficient sign fixed by strict convexity. Moreover the generator is forced by

$$\Phi'(c_g g(s)) = \alpha - \beta s, \qquad \beta > 0. \tag{12}$$

Thus, within this regular fixed-shape radial class, no other strictly convex finite-at-zero internal energy can support the all-step exact mean–covariance mechanism. In the Shannon and $m > 1$ power cases, the resulting Gaussian/q-Gaussian branches are also sufficient for full-space closure by Theorem 2.

Why the theorem is rigid. All-step exactness already implies that the canonical law minimizes the energy on every mean–covariance fiber. Local resolvent closure then forces the Wasserstein gradient of the internal energy to be an affine radial field, which yields (12). Requiring the same generator to remain the covariance-fiber minimizer over all isotropic scales produces the functional equation

$$h(ct) = A(c) + B(c)h(t), \qquad h = \Phi', \tag{13}$$

and composition forces $B(c_1c_2) = B(c_1)B(c_2)$. Continuity gives $B(c) = c^p$, hence $h'(t) = Kt^{p-1}$. Finiteness at the origin removes the barrier and negative-power branches, leaving (11). The complete proof separates the local generator lemma from the all-scale energy lemma and is given in Appendix A.5.

Theorem 3 is the paper's main classification result. It is d-dimensional and finite-step in its exactness premise. It does not classify arbitrary Wasserstein submanifolds, nonradial generators, or nonintegral energies; those are outside the stated class rather than hidden in the conclusion.

4 NEAR SOLVABILITY: AN IRREDUCIBLE STRUCTURAL ERROR

Exactness is binary, whereas misspecification can be small. At a regular $q \in \mathcal{Q}$, let $T_q\mathcal{Q}$ be the family tangent space and $N_q\mathcal{Q}$ its orthogonal complement in $L^2(q)$. In the one-dimensional quantile formulation below, for $Q \in \mathcal{M}$ we use $T_Q\mathcal{M}$ and $N_Q\mathcal{M}$ for the corresponding tangent and normal spaces.

Definition 2 (Solvability defect)**.**

$$\Delta_{\mathrm{solv}}(q; \mathcal{F}, \mathcal{Q}) := \|P_{N_q\mathcal{Q}} \nabla_W \mathcal{F}(q)\|_{L^2(q)}. \tag{14}$$

The coefficient is a geometric normal component; its role here is operational. It is the quantity that no in-family optimizer can remove at first order.

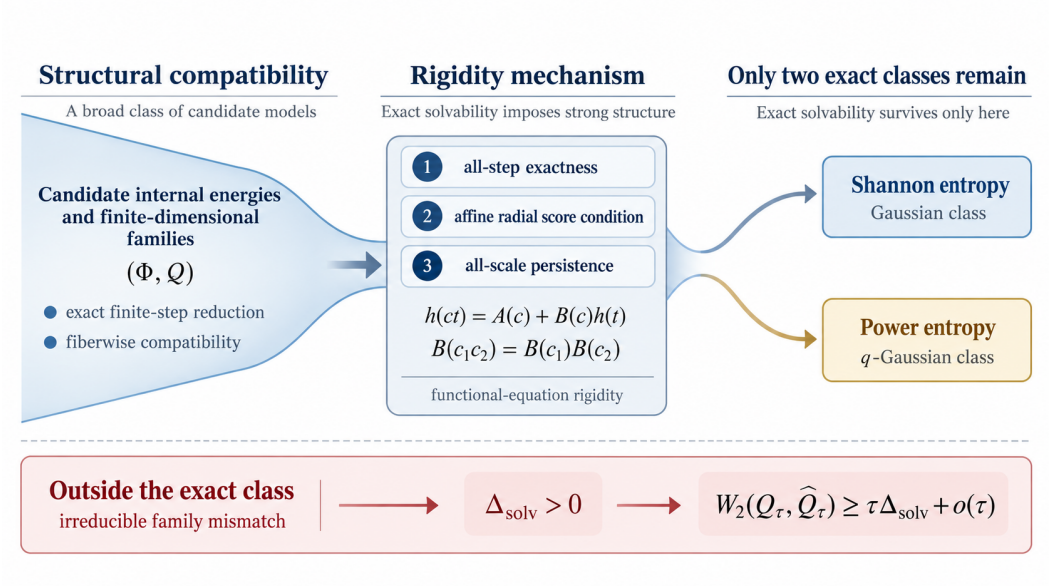


Figure 2: **Rigidity induced by structural compatibility.** Starting from a candidate internal energy Φ and a finite-dimensional family $\mathcal{Q}$, all-step fiberwise exactness first forces an affine radial score condition. Requiring this structure to persist across covariance scales yields the functional equations $h(ct) = A(c) + B(c)h(t)$ and $B(c_1c_2) = B(c_1)B(c_2)$. Under the regularity, scale-persistence, and finite-at-zero assumptions stated in the paper, only two exact branches remain: the Shannon/Gaussian and power/q-Gaussian classes. Outside these branches, the solvability defect Δ_{solv} is positive and induces the first-order lower bound $W_2(Q_\tau, \widehat{Q}_\tau) \geq \tau\Delta_{\mathrm{solv}} + o(\tau)$.

Theorem 4 (First-order structural error)**.** *Let Q be a one-dimensional quantile on a C^2 finite-dimensional quantile manifold $\mathcal{M}$. If the unrestricted proximal branch satisfies*

$$Q_\tau = Q - \tau G + o(\tau) \quad \text{in } L^2(0,1), \tag{15}$$

then

$$\mathrm{dist}_{W_2}(Q_\tau, \mathcal{M}) = \tau \|P_N G\|_{L^2} + o(\tau). \tag{16}$$

Under Assumption 1, the coefficient is Δ_{solv}.

Corollary 1 (Optimizer-independent family floor)**.** *For every family-constrained output $\widehat{Q}_\tau \in \mathcal{M}$,*

$$W_2(Q_\tau, \widehat{Q}_\tau) \geq \mathrm{dist}_{W_2}(Q_\tau, \mathcal{M}) = \tau\Delta_{\mathrm{solv}} + o(\tau). \tag{17}$$

The remainder belongs to the distance-to-manifold expansion itself; no uniformity over algorithms is required.

The one-dimensional restriction enters only through the exact quantile isometry used to convert the Wasserstein distance to a Hilbert-space distance. The abstract Hilbert normal-distance lemma and the full proof are in Appendix A.6. The exactness and rigidity theorems above do not use this one-dimensional device.

Proposition 1 (Tangency is strictly weaker than finite-step closure)**.** *Let $\mathcal{M} = \{(u, 0) : u \in \mathbb{R}\} \subset \mathbb{R}^2$ and*

$$F(u, v) = u + uv + \frac{\lambda}{2}(u^2 + v^2), \qquad \lambda > 1.$$

At $x_0 = (0, 0)$ the gradient is tangent to $\mathcal{M}$, so the normal first-order defect is zero. Nevertheless, for every $\tau > 0$ the unique proximal point has a positive second coordinate of order τ^2 and therefore lies outside $\mathcal{M}$.

This example prevents a common shortcut: $\Delta_{\mathrm{solv}} = 0$ certifies only the absence of an $O(\tau)$ obstruction. It does not certify a finite-step invariant family. The closed-form proximal point and strong-convexity check are in Appendix A.7.

For the centered isotropic radial models used below, let $X \sim q$, set $R = \|X\|$, and suppose the nonaffine radial part of the Wasserstein gradient is $v(x) = \psi(\|x\|^2)x$. Projection onto the dilation direction gives

$$\Delta^2_{\mathrm{solv}} = \mathbb{E}[R^2\psi(R^2)^2] - \frac{\mathbb{E}[R^2\psi(R^2)]^2}{\mathbb{E}[R^2]}. \tag{18}$$

Proposition 2 (Finite-sample estimation)**.** *Let $Y = (R^2, R^2\psi(R^2), R^2\psi(R^2)^2)$ have finite first moment. Replacing the expectations in* (18) *by empirical means yields $\widehat{\Delta_{\mathrm{solv}}}^2 \to \Delta^2_{\mathrm{solv}}$ almost surely. If Y has finite covariance, then the multivariate CLT and delta method give asymptotic normality for $\widehat{\Delta_{\mathrm{solv}}}^2$ and, when $\Delta_{\mathrm{solv}} > 0$, for $\widehat{\Delta_{\mathrm{solv}}}$.*

Corollary 2 (d-dimensional generalized-Gaussian calibration)**.** *Let $q_{p,d}(x) \propto \exp[-\alpha_{p,d}\|x\|^{2p}]$ be covariance-normalized, where*

$$\alpha_{p,d} = \left[\frac{\Gamma((d+2)/(2p))}{d\,\Gamma(d/(2p))}\right]^p. \tag{19}$$

For Shannon entropy, using $q_{p,d}$ as a diagnostic family, the radial solvability defect is

$$\Delta^2_{\mathrm{solv}}(d,p) = \frac{4p^2\Gamma((d+2)/(2p))\Gamma((d+4p-2)/(2p))}{d\,\Gamma(d/(2p))^2} - d. \tag{20}$$

It vanishes at the matched Gaussian point $p = 1$ and is positive for $p \neq 1$.

Here $q_{p,d}$ is a test family for the Shannon energy; only $p = 1$ is the matched Gaussian branch. Equation (20) provides an analytic d-dimensional calibration target for Proposition 2; it does not extend the sharp distance identity of Theorem 4 beyond one dimension.

5 What exactness, tangency, and approximation do—and do not—mean

The preceding results describe three different statements that are often conflated in finite-dimensional Wasserstein models. Exactness is a global property of the finite-step objective; tangency is an infinitesimal property of the gradient field; approximation quality is a quantitative property of a chosen family and step size. Their logical relation is strict.

Corollary 3 (Strict hierarchy of finite-dimensional structure)**.** *At a regular member where Assumption 1 applies:*

1. *all-step fiberwise exactness implies local tangency and hence $\Delta_{\mathrm{solv}} = 0$;*
2. *$\Delta_{\mathrm{solv}} = 0$ does not imply finite-step closure, even for a strongly convex proximal problem;*
3. *if $\Delta_{\mathrm{solv}} > 0$, every family-constrained output incurs $\tau\Delta_{\mathrm{solv}} + o(\tau)$ first-order error relative to the unrestricted step.*

The first implication follows because an exact proximal branch remains inside the family as $\tau \downarrow 0$. The strict failure of the converse is Proposition 1; the positive-defect statement is Corollary 1. Thus "the vector field is tangent" is not a finite-step certificate. This distinction matters for implicit JKO updates: a curved family can agree with the instantaneous direction and still fail to contain the next proximal point.

The local defect also has a direct implication for repeated constrained dynamics. A one-step structural mismatch need not disappear under iteration; its effect is propagated by the stability of the unrestricted proximal map.

Proposition 3 (Propagation over repeated constrained updates)**.** *Let P_τ be the unrestricted proximal map and R_τ any family-constrained update. If P_τ is L-Lipschitz in W_2 on the visited region,*

$$\mu_{k+1} = P_\tau(\mu_k), \qquad \widehat{\mu}_{k+1} = R_\tau(\widehat{\mu}_k),$$

and

$$\epsilon_k = W_2(P_\tau(\widehat{\mu}_k), R_\tau(\widehat{\mu}_k))\,,$$

Result	Scope	Direction	What it certifies
Thm. 1	d, all τ	equivalence for fiber-wise mechanism	exact reduction from separate energy/transport domination
Thm. 2	d, finite step	sufficient concrete branch	unrestricted Gaussian/q-Gaussian closure
Thm. 3	d, all scales	converse classification	only Shannon/positive-power radial branches survive
Thm. 4	1D, $\tau \downarrow 0$	quantitative local law	first-order distance to a misspecified family
Prop. 1	Hilbert model	strict separation	tangency alone does not imply finite-step invariance

Table 1: **Logical roles of the main results.** The paper does not merge global exactness, local tangency, and quantitative mismatch into a single if-and-only-if statement.

then

$$W_2(\mu_K, \widehat{\mu}_K) \le L^K W_2(\mu_0, \widehat{\mu}_0) + \sum_{j=0}^{K-1} L^{K-1-j} \epsilon_j.$$

Whenever the local hypotheses of Theorem 4 hold at $\widehat{\mu}_j$, Corollary 1 gives

$$\epsilon_j \ge \tau \Delta_{\mathrm{solv}}(\widehat{\mu}_j) + o(\tau).$$

This proposition does not imply monotone accumulation of local errors, nor do we estimate the Lipschitz constant L experimentally. Its role is structural: once a family-constrained update incurs a nonzero local mismatch, subsequent unrestricted dynamics can propagate that discrepancy. The 50-step experiment in Section 6 is therefore interpreted as an empirical consequence of misspecification rather than as a separate test of the proposition.

Why this differs from constrained Gaussian dynamics. A Gaussian Bures scheme may be mathematically well posed and convergent while still solving a constrained problem. Theorem 1 asks whether the constraint is lossless relative to the unrestricted objective; Theorem 3 then asks how restrictive that losslessness is. This is orthogonal to convergence-rate analyses inside the constrained state space. In particular, the classification theorem uses no learned model, optimizer architecture, or discretization-specific argument: it follows from the interaction of the full Wasserstein metric, moment fibers, and the internal energy.

6 Experiments: exactness, rigidity, and structural mismatch

The experiments follow the same question as the theory: when is the family restriction exact, and when does a structural gap remain? We use unrestricted one-dimensional quantile JKO as the numerical reference for finite-step claims and reserve the new d-dimensional study for the defect estimator, whose population target is available in closed form from Corollary 2. All decisive one-dimensional calculations use float64. The frozen protocols retain failed, weak-signal, boundary-sensitive, and initialization-sensitive cases rather than filtering them; the one-dimensional and high-dimensional audits use disjoint seed sets and independent code paths. Hardware/software details are summarized in Appendix C.

Exact branches and numerical floor. The independent replication contains 15 matched records across Shannon, $m = 1.5$, and $m = 2$, all feasible with no failed run. At $N = 256$, $\tau = 0.01$, $\kappa = 1$, the reproduced closure residuals are 2.80×10^{-4}, 7.32×10^{-5}, and 2.79×10^{-5}, respectively. Representative SLSQP/trust-constr cross-checks drive gradient RMS values from the 10^{-7} range to 10^{-9} without changing the closure conclusion. Endpoint and grid sweeps show that the remaining discrepancy is dominated by truncation/discretization rather than optimizer termination; Appendix B reports the retained rows.

Rigidity and first-order failure. The generator sweep contains ten matched/mismatched families over five step sizes. The matched Gaussian and matched power generator sit at zero population

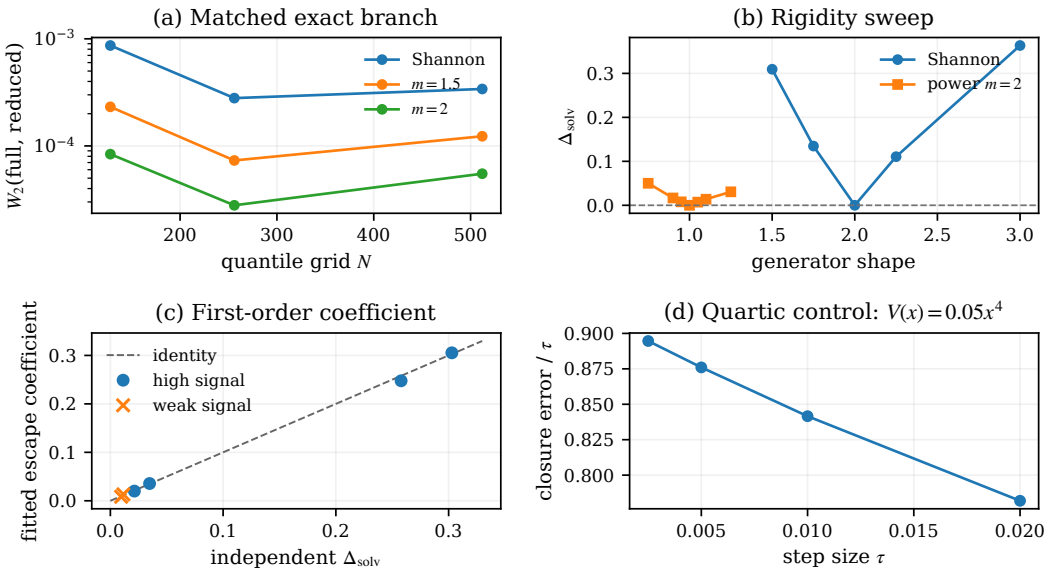


Figure 3: **Direct audit of the structural theory.** (a) Matched Shannon and power branches agree with the reduced solution at the measured discretization floor. (b) The solvability defect is minimized at the matched generator. (c) Independent defect calculations predict the small-step escape coefficient; four high-signal fits are used for the headline accuracy, while two additional weak-signal fits are retained rather than discarded. (d) Replacing quadratic confinement by $V(x) = 0.05x^4$ produces persistent first-order-sized escape.

defect, while moving the shape parameter away from the matched value produces a positive defect and a corresponding family escape. Six mismatch settings admit a numerical coefficient fit under the recorded signal-to-floor rule. Four are high-signal: Shannon $\beta = 1.5, 3$ and power $r = 0.75, 1.25$ have relative coefficient errors $3.96\%, 0.82\%, 2.46\%, 6.07\%$ (median 3.2%). Two additional weak-signal fits, $r = 0.9$ and $r = 1.1$, have 14.8% and 12.6% relative error; four closer-to-matched settings remain floor-limited. This separation is deliberate: the theory predicts a coefficient, but the numerical experiment can only resolve it above the measured solver floor.

A negative control breaks the quadratic-potential assumption with $V(x) = 0.05x^4$. The best Gaussian projection errors are $2.24, 4.38, 8.42, 15.64 \times 10^{-3}$ for $\tau = 0.0025, 0.005, 0.01, 0.02$, with error/$\tau$ between 0.78 and 0.89. The effect is far above the matched Gaussian floor and is reproduced in the final audit rerun, ruling out the explanation that the discretized solver mechanically prefers the Gaussian family.

Finite-sample and high-dimensional calibration. The original estimator audit uses two frozen mismatches, four sample sizes, and 30 independent repetitions per setting (240 runs, no failures). At $n = 20{,}000$, median relative errors are 0.50% for the Shannon mismatch and 1.94% for the compact-support power mismatch; the replication-mean 95% interval contains the quadrature truth for every sample-size/case pair.

The new high-dimensional audit directly targets Corollary 2. A separately rebuilt full-vector sampler evaluates $d \in \{5, 10, 20, 50, 100\}$, three sample sizes up to 10^5, five shapes around the matched point, and five independent seeds. The frozen grid contains 375 rows with no recorded failure. All 300 mismatched rows have $\widehat{\Delta_{\text{solv}}} > 0$, whereas every matched $p = 1$ estimate is numerically zero. For every fixed (p, d) series, RMSE at $n = 10^5$ does not exceed its value at $n = 5{,}000$. At $d = 100, n = 10^5$, mean relative error over mismatches is 0.17%. Six preregistered bootstrap checks use 1000 resamples each; all six analytic truths lie inside both the Δ_{solv} and Δ^2_{solv} 95% intervals. Sample covariance diagnostics approach the I_d target, with off-diagonal RMS near 3.1×10^{-3} at $d = 100, n = 10^5$.

One-step floor and repeated updates. For generalized-Gaussian mismatches $\beta = 1.5$ and $\beta = 3$ at $\tau = 0.01$, the predicted first-order floors are 2.58×10^{-3} and 3.03×10^{-3}; restricted-family JKO

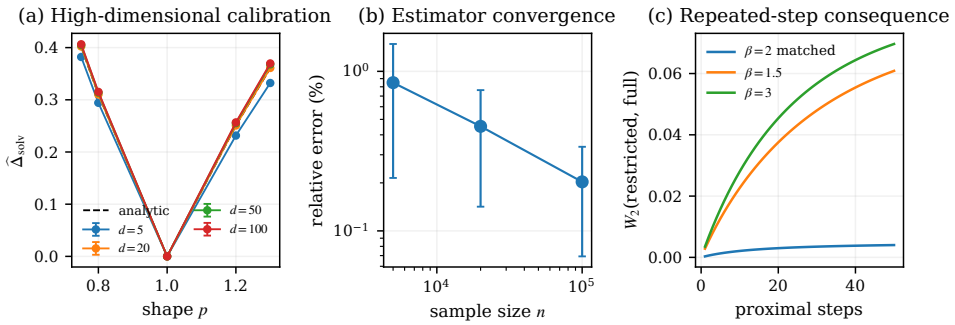


Figure 4: **Scalable defect diagnosis and long-horizon consequence.** (a) Monte Carlo estimates track the analytic generalized-Gaussian defect through $d = 100$ and preserve the exact zero at $p = 1$. (b) Relative estimation error decreases with sample size across the high-dimensional mismatch grid. (c) Over 50 proximal steps, the matched Gaussian family remains near the numerical reference floor, whereas mismatched families accumulate an order-of-magnitude larger discrepancy. The high-dimensional panels validate the defect and its estimator, not a multidimensional extension of Theorem 4.

gives 2.84×10^{-3} and 3.50×10^{-3}. After 50 steps the corresponding discrepancies reach 6.09×10^{-2} and 6.96×10^{-2}, while the matched $\beta = 2$ branch remains at 4.01×10^{-3}. The moment closure shows the same qualitative separation. These trajectories are consequences rather than theorem evidence: they show why a small local family mismatch matters when the same restricted representation is used repeatedly. Taken together, the experiments test the same hierarchy as the theory: exact reduction, rigidity under structural compatibility, first-order mismatch, high-dimensional estimability of the defect, and repeated-step consequences. Additional numerical audits are reported in Appendix B.

Computational consequence. The same matched run also illustrates why exact reduction matters computationally. Over 50 steps, the unrestricted quantile-JKO reference takes 3.92 seconds cumulatively, whereas the exact Gaussian family update takes 3.28×10^{-5} seconds in the recorded implementation. We report this only as an implementation-level consequence of replacing a grid optimization by a finite-dimensional update, not as a hardware-independent complexity claim.

Exploratory 2D solver check. A generic density-level 2D experiment evaluates a rotated anisotropic Gaussian on 64^2 and 96^2 grids with several entropic regularizations and initializations. All 72 recorded solves satisfy basic mass, nonnegativity, and finite-value checks. Several cells approach the analytic mean/covariance trend as regularization is reduced, but convergence is not monotone and some initializations remain poor. We therefore keep this block in the appendix as an implementation sanity check, not as evidence for the sharp one-dimensional distance law.

Additional numerical robustness. The headline plots are backed by checks that do not change the qualitative conclusions. On the $N = 256$ matched cases, independent trust-constr reruns reduce gradient RMS to 2.20×10^{-9} (Shannon), 4.03×10^{-9} ($m = 1.5$), and 5.29×10^{-9} ($m = 2$) while reproducing the same closure residuals. At $d = 100, n = 10^5$, the full-vector sampler has covariance-diagonal RMSE about 4.3×10^{-3} and off-diagonal RMS about 3.1×10^{-3}; all six 1000-resample bootstrap intervals contain the analytic defect target. The separate 2D block contains 72 feasible solves, with the best cells reaching mean error below 2×10^{-3} and covariance Frobenius error below 10^{-2}; because regularization and initialization trends are not uniformly monotone, we keep this block as a sanity check rather than theorem evidence.

7 Conclusion

Structural compatibility separates exact finite-dimensional solvability from convenient approximation. Theorem 1 requires energy and transport to select the same canonical representative, while Theorem 3 leaves only the matched Shannon/Gaussian and power/q-Gaussian branches in the stated radial class. Outside this regime, the normal defect gives the first-order obstruction, while tangency alone does not certify finite-step closure. Numerical audits recover this hierarchy through $d = 100$, showing that exact solvability is a property of the metric–energy–family triple rather than optimizer accuracy.

LIMITATIONS

The classification concerns regular fixed-generator radial location–scatter families, while the sharp proximal-distance equality uses one-dimensional quantile geometry. The high-dimensional experiment tests the d-dimensional defect and its estimator against an analytic target rather than extending that equality.

AI USE STATEMENT

Generative AI tools were used for language editing and code/figure prototyping. All mathematical statements, citations, and reported numerical values were checked by the authors, who take responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

The main text states the assumptions needed by each theorem and the appendix provides complete proofs. Experiments use frozen configurations and disjoint seed sets for the one-dimensional and high-dimensional audits. Runs were executed on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU (24 GB), using Python 3.10.20 and PyTorch 2.5.1 built with CUDA 12.1; the core one-dimensional and Monte Carlo calculations use float64 NumPy/SciPy routines. Raw arrays, CSV summaries, failed-run tables, optimizer and boundary-sensitivity records, frozen configurations, bootstrap samples, and figure-rebuild scripts are retained.

REFERENCES

F. Altekrüger, J. Hertrich, and G. Steidl. Neural Wasserstein gradient flows for discrepancies with Riesz kernels. *Proceedings of the 40th International Conference on Machine Learning*, PMLR 202, 2023.

L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser, 2008.

E. A. Carlen and K. Craig. Contraction of the proximal map and generalized convexity of the Moreau–Yosida regularization in the 2-Wasserstein metric. *Mathematics and Mechanics of Complex Systems*, 1(1):33–65, 2013.

P. Caucheteux, C. Bonet, and A. Korba. A unifying view of variational generative Wasserstein flows. arXiv preprint arXiv:2605.31369, 2026.

J. Choi, J. Choi, and M. Kang. Scalable Wasserstein gradient flow for generative modeling through unbalanced optimal transport. *ICML*, 2024.

M. Z. Diao, K. Balasubramanian, S. Chewi, and A. Salim. Forward-backward Gaussian variational inference via JKO in the Bures–Wasserstein space. *ICML*, 2023.

S. Di Marino, E. Naldi, and S. Villa. Inexact JKO and proximal-gradient algorithms in the Wasserstein space. arXiv preprint arXiv:2505.23517, 2025.

M. H. Duong. Asymptotic equivalence of the discrete variational functional and a rate-large-deviation-like functional in the Wasserstein gradient flow of the porous medium equation. *Asymptotic Analysis*, 92(1–2):85–106, 2015.

X. Feng, L. Wang, D. Needell, and R. Lai. Learn to evolve: Self-supervised neural JKO operator for Wasserstein gradient flow. arXiv preprint arXiv:2601.05583, 2026.

M. Gelbrich. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147:185–203, 1990.

Y. Jin, S. Liu, H. Wu, X. Ye, and H. Zhou. Parameterized Wasserstein gradient flow. *Journal of Computational Physics*, 524:113660, 2025.

O. Johnson and C. Vignat. Some results concerning maximum Rényi entropy distributions. *Annales de l'Institut Henri Poincaré B*, 43(3):339–351, 2007.

R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

D. Kim, H. Park, and W. Shim. Finite-dimensional reduction of a Wasserstein gradient flow and sharp decay rates. arXiv preprint arXiv:2606.21918, 2026.

M. Lambert, S. Chewi, F. Bach, S. Bonnabel, and P. Rigollet. Variational inference via Wasserstein gradient flows. *Advances in Neural Information Processing Systems*, 35:14434–14447, 2022.

A. Takatsu. Wasserstein geometry of porous medium equation. *Annales de l'Institut Henri Poincaré C*, 29(2):217–232, 2012.

A. Takatsu. Behaviors of φ-exponential distributions in Wasserstein geometry and an evolution equation. *SIAM Journal on Mathematical Analysis*, 45(4):2546–2556, 2013.

A Complete proofs and technical details

A.1 Proof of Theorem 1

Fix arbitrary $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and canonical reference q_{θ_0}. If (3)–(4) hold, then for every $\tau > 0$ their weighted sum is exactly (2). Conversely, assume (2) for every $\tau > 0$. Since both Wasserstein distances are finite, taking $\tau \to \infty$ gives (3). Multiplying (2) by 2τ gives

$$2\tau[\mathcal{F}(q_{S\mu}) - \mathcal{F}(\mu)] + W_2^2(q_{S\mu}, q_{\theta_0}) \le W_2^2(\mu, q_{\theta_0}).$$

Using the energy domination and sending $\tau \downarrow 0$ yields (4). For (5), every μ is dominated by a canonical point $q_{S\mu}$, so the full-space infimum is at least the canonical infimum; the reverse inequality holds because canonical points are feasible in the full problem. No compactness or interchange of nested infima is needed for this equality. If the energy inequality is strict away from the section, any noncanonical minimizer can be replaced by a strictly better canonical point on the same fiber.

A.2 Existence and proof of Theorem 2

For $m > 1$, U_m is nonnegative and convex in the density. For $m = 1$, let γ_s be the centered Gaussian with variance $s^2 I$. Nonnegativity of relative entropy gives

$$U_1(\rho) \ge -\frac{1}{2s^2} \int \|x\|^2 \rho(x) dx - C_{d,s}.$$

For any fixed reference $\nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$W_2(\mu, \nu) \ge \left| \sqrt{M_2(\mu)} - \sqrt{M_2(\nu)} \right|.$$

Choose s sufficiently large that the positive quadratic growth of $(2\tau)^{-1} W_2^2$ dominates the negative entropy coefficient. Since V is bounded below, minimizing sequences have uniformly bounded second moment. Tightness and lower semicontinuity yield a minimizer. Under the extended-value convention, finite internal energy gives absolute continuity; singular covariance would force concentration on an affine hyperplane and is therefore impossible.

Now fix a competitor μ with mean a and covariance $\Sigma \succ 0$. Gelbrich's inequality gives (8). The positive affine covariance map between two members of the same radial location–scatter orbit is optimal and attains the right-hand side. On the same moment fiber,

$$\int V(x) d\mu(x) = \tfrac{1}{2} a^\top H a + h^\top a + c + \tfrac{1}{2} \operatorname{tr}(H\Sigma),$$

so the potential depends only on (a, Σ). The Gaussian uniquely minimizes U_1 at fixed covariance; the covariance-normalized q-Gaussian minimizes U_m at fixed covariance for $m > 1$ (Johnson and Vignat, 2007). Thus the same canonical law satisfies both inequalities of Theorem 1, proving exact closure. Affine scaling gives

$$\int q_{a,\Sigma}^m dx = (\det \Sigma)^{-(m-1)/2} \int B_m^m dz,$$

and entropy scaling gives the Shannon $-\frac{1}{2} \log \det \Sigma$ term, yielding (9).

A.3 ISOTROPIC COVARIANCE ROOTS

If $H = \kappa I$, the Bures term is minimized when Σ shares eigenvectors with Σ_0. In their common eigenbasis, write $s_i = \sqrt{\lambda_i(\Sigma)}$ and $s_{0i} = \sqrt{\lambda_i(\Sigma_0)}$. Then $d_B^2 = \sum_i (s_i - s_{0i})^2$. The mean update is

$$a^\star = (I + \tau H)^{-1}(a_0 - \tau h).$$

In the Shannon case, differentiating the reduced objective gives

$$(\kappa + \tau^{-1})s_i - \tau^{-1}s_{0i} - s_i^{-1} = 0.$$

For $m > 1$, let $A = (m-1)C_{m,d}\prod_j s_j^{-(m-1)}$; then

$$(\kappa + \tau^{-1})s_i^2 - \tau^{-1}s_{0i}s_i - A = 0.$$

For fixed $A > 0$ each equation has a unique positive root, and substitution into the definition of A leaves a scalar consistency equation.

A.4 TWO RIGIDITY LEMMAS

We first record the local generator argument. For two regular radial location–scatter laws in one fixed-generator orbit, the optimal quadratic-cost map is positive affine. Hence the scaled displacement to any nearby family member belongs to the finite-dimensional affine tangent space $\{b + Mx : M = M^\top\}$, which is closed in $L^2(q_g)$. Under Assumption 1 and local proximal compatibility, $-\nabla_W \mathcal{F}_{\Phi,V}(q_g)$ lies in this affine space. Subtracting the affine quadratic-potential gradient shows that $\nabla_W U_\Phi(q_g)$ is affine. On the support interior,

$$\nabla_W U_\Phi(q_g)(x) = 2c_g\Phi''(c_g g(s))g'(s)x, \qquad s = \|x\|^2.$$

A radial affine field on a rotationally invariant law must equal λx: comparing x and $-x$ removes the translation, and commutation with every orthogonal matrix forces the linear part to be scalar. Thus

$$\frac{d}{ds}\Phi'(c_g g(s)) = \lambda/2.$$

Strict radial decrease and $\Phi'' > 0$ imply $\lambda < 0$, giving (12).

For the all-scale energy argument, strict convexity makes the covariance-fiber minimizer unique. All-step fiberwise exactness and Theorem 1 imply that the canonical member minimizes U_Φ on every mean–covariance fiber. On isotropic fibers, regularity gives the Euler–Lagrange relation

$$h(q_\sigma(x)) = \alpha_\sigma - \gamma_\sigma\|x\|^2, \qquad h = \Phi', \quad \gamma_\sigma > 0.$$

With $q_\sigma(\sigma y) = \sigma^{-d}q_1(y)$ and $c = \sigma^{-d}$, eliminate $\|y\|^2$ between the scale-1 and scale-σ equations to obtain (13). Because the interior range contains $(0, M)$, for arbitrary $c_1, c_2 > 0$ choose

$$0 < t < \frac{M}{\max\{1, c_2, c_1c_2\}}.$$

Then t, c_2t, c_1c_2t lie in the common range, so composing (13) gives $B(c_1c_2) = B(c_1)B(c_2)$. Continuity yields $B(c) = c^p$. Differentiation gives $h'(ct) = c^{p-1}h'(t)$ and therefore $h'(t) = Kt^{p-1}$ with $K > 0$. Integration produces the logarithmic, barrier, and power branches. The finite extension at zero eliminates the barrier and negative-power branches. The additive linear term bt is immaterial because $\int \rho = 1$; an additive constant in Φ is not retained because its integral over $\mathbb{R}^d$ need not be finite.

A.5 PROOF OF THEOREM 3

All-step exactness implies the energy domination of Theorem 1; since V is constant on a mean–covariance fiber, the canonical $q_{a,\Sigma}$ minimizes U_Φ on every such fiber. Strict convexity gives uniqueness whenever a density minimizer exists. At the centered isotropic member, exactness also gives local proximal compatibility. The first rigidity lemma in Appendix A.4, together with Assumption 1, therefore yields (12). The second rigidity lemma applies to the isotropic covariance fibers of the same fixed-generator section and yields the classification (11). Conversely, for Shannon and $m > 1$ power entropies, Theorem 2 verifies all-step exactness for the corresponding Gaussian/q-Gaussian location–scatter families under quadratic confinement. This establishes the claimed necessity in the full stated class and sufficiency for the matched branches treated concretely in the paper.

A.6 Proof of Theorem 4 and Corollary 1

Let $\mathcal{M}$ be a C^2 finite-dimensional submanifold of a Hilbert space $\mathcal{H}$, $x \in \mathcal{M}$, and $y_\tau = x - \tau G + o(\tau)$. In orthogonal coordinates $\mathcal{H} = T \oplus N$, a local graph representation has the form

$$\mathcal{M} = \{x + u + R(u) : u \in T,\ \|u\| \le r\}, \qquad R(u) \in N, \quad \|R(u)\| \le C\|u\|^2.$$

Write $G = G_T + G_N$. Choosing $u = -\tau G_T$ gives the upper bound $\operatorname{dist}(y_\tau, \mathcal{M}) \le \tau\|G_N\| + o(\tau)$. Conversely, any minimizing graph point has $u = O(\tau)$; otherwise its tangential discrepancy exceeds this upper bound. Hence $R(u) = O(\tau^2)$ and the normal residual is $-\tau G_N + o(\tau)$, giving the matching lower bound. For one-dimensional absolutely continuous laws, the quantile map is an exact W_2 isometry into $L^2(0,1)$, so (16) follows directly. Under Assumption 1, the normal velocity is the projected Wasserstein gradient and has norm Δ_{solv}. Finally, every $\widehat{Q}_\tau \in \mathcal{M}$ satisfies

$$W_2(Q_\tau, \widehat{Q}_\tau) \ge \operatorname{dist}_{W_2}(Q_\tau, \mathcal{M}),$$

which proves Corollary 1. The same distance remainder is used for every algorithm; no separate uniform algorithmic remainder is introduced.

A.7 Proof of Proposition 1

The Hessian of F is

$$\nabla^2 F = \begin{pmatrix} \lambda & 1 \\ 1 & \lambda \end{pmatrix},$$

whose eigenvalues are $\lambda \pm 1 > 0$; the proximal problem is therefore strongly convex and has a unique minimizer. At the origin $\nabla F(0,0) = (1,0)$ is tangent to $\mathcal{M}$. Solving the first-order conditions for $F(u,v) + (u^2+v^2)/(2\tau)$ gives

$$\operatorname{prox}_{\tau F}(0,0) = \frac{1}{(1+\lambda\tau)^2 - \tau^2} \begin{pmatrix} -\tau(1+\lambda\tau) \\ \tau^2 \end{pmatrix}.$$

The second coordinate is positive for every $\tau > 0$ and is $O(\tau^2)$, proving the claim.

A.8 Proof of Proposition 2

Let $m_1 = \mathbb{E}R^2$, $m_2 = \mathbb{E}[R^2\psi(R^2)]$, and $m_3 = \mathbb{E}[R^2\psi(R^2)^2]$. Then (18) is $g(m_1, m_2, m_3)$ with $g(u,v,w) = w - v^2/u$. The strong law gives convergence of the empirical moments and continuity of g on $u > 0$ gives strong consistency. Under finite covariance, the multivariate CLT and delta method yield the stated asymptotic normality. When $\Delta_{\mathrm{solv}} > 0$, apply the delta method once more to the square root.

A.9 Proof of Corollary 2

For $q_{p,d}(x) \propto \exp[-\alpha\|x\|^{2p}]$, write $R = \|X\|$. Radial integration gives

$$\mathbb{E}R^k = \alpha^{-k/(2p)} \frac{\Gamma((d+k)/(2p))}{\Gamma(d/(2p))}.$$

Imposing $\mathbb{E}R^2 = d$ yields (19). For Shannon entropy, $\nabla \log q_{p,d}(x) = \psi(R^2)x$ with $\psi(s) = -2p\alpha s^{p-1}$. The same moment identity gives $\mathbb{E}[R^2\psi(R^2)] = -d$, so the optimal dilation coefficient in (18) is -1. Substitution yields (20). The defect is a weighted variance after projection onto the dilation field, and therefore vanishes only when ψ is constant on the support, i.e. $p = 1$.

A.10 Proof of Proposition 3

The one-step triangle inequality gives

$$W_2(\mu_{k+1}, \widehat{\mu}_{k+1}) \le L W_2(\mu_k, \widehat{\mu}_k) + \epsilon_k.$$

Iterating this recursion yields this result. The local lower bound on ϵ_j is Corollary 1 applied at $\widehat{\mu}_j$.

B Extended numerical audit

B.1 Matched exactness and solver cross-checks

The independent matched replication contains 15 retained rows, all successful and feasible. Table 2 reports the main-grid rows at $\tau = 0.01$, $\kappa = 1$. Trust-constr cross-checks on the $N = 256$ representatives reduce gradient RMS to 2.20×10^{-9} (Shannon), 4.03×10^{-9} ($m = 1.5$), and 5.29×10^{-9} ($m = 2$), with the same closure conclusion. The final audit rerun reproduces the three stored $N = 256$ residuals to machine-level agreement.

Table 2: Representative independent matched checks.

Family	N	W_2(full,reduced)	gradient RMS
Shannon	128	8.64×10^{-4}	6.03×10^{-7}
Shannon	256	2.80×10^{-4}	5.21×10^{-7}
Shannon	512	3.30×10^{-4}	5.79×10^{-7}
$m = 1.5$	128	2.31×10^{-4}	7.70×10^{-7}
$m = 1.5$	256	7.32×10^{-5}	4.19×10^{-7}
$m = 1.5$	512	1.22×10^{-4}	4.01×10^{-7}
$m = 2$	128	8.35×10^{-5}	1.69×10^{-7}
$m = 2$	256	2.79×10^{-5}	1.08×10^{-7}
$m = 2$	512	5.49×10^{-5}	8.10×10^{-7}

B.2 Complete mismatch coefficient audit

Table 3: All ten mismatch settings. "Weak" denotes a fitted coefficient with relative error above 10%; "floor" denotes insufficient signal-to-floor points.

Case	Δ_{solv}	fitted a	rel. error	status
Shannon $\beta = 1.50$	.25786	.24764	3.96%	primary
Shannon $\beta = 1.75$	.11225	–	–	floor
Shannon $\beta = 2.25$	.09217	–	–	floor
Shannon $\beta = 3.00$	.30287	.30536	0.82%	primary
Power $r = .75$	.03478	.03564	2.46%	primary
Power $r = .90$	.01156	.01327	14.81%	weak
Power $r = .95$	.00549	–	–	floor
Power $r = 1.05$	.00499	–	–	floor
Power $r = 1.10$	.00957	.00836	12.61%	weak
Power $r = 1.25$	.02130	.02001	6.07%	primary

All fitted cases have $R^2 > 0.9999$; this is used only as a numerical fit diagnostic. The weak-signal rows are retained to expose the transition to solver-floor uncertainty.

B.3 Finite-sample estimator audit

All 240 runs complete. For the Shannon mismatch, median relative error decreases from 2.06% at $n = 500$ to 0.50% at $n = 20{,}000$; for the power mismatch it decreases from 14.20% to 1.94%. The replication-mean 95% interval contains the truth at every sample size for both families.

B.4 High-dimensional defect audit

The full-vector implementation contains 300 rows for $d \in \{5, 10, 20, 50\}$ and 75 further rows at $d = 100$. Shapes are $p \in \{0.75, 0.8, 1, 1.2, 1.3\}$ and sample sizes $n \in \{5{,}000, 20{,}000, 100{,}000\}$ with five independent seeds. There are no recorded failures. Across mismatches, mean relative error is 0.81%, 0.44%, and 0.21% as n increases. At $n = 100{,}000$, mean relative errors by dimension are

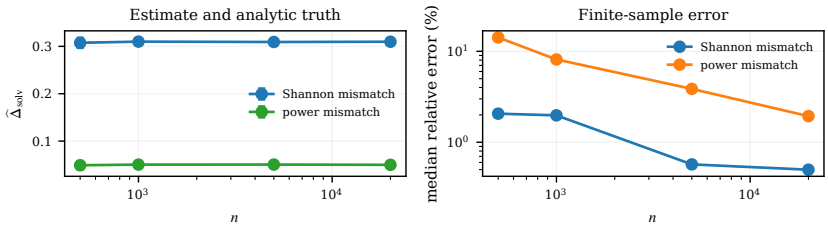


Figure 5: Finite-sample estimator audit. Thirty repetitions are used at each sample size for two frozen mismatch families.

Question	Evidence	What is resolved
Is the restriction exact?	matched full JKO vs. reduced update	residuals stay at the measured numerical floor across Shannon and two power branches
Is exactness structurally selective?	generator sweep + negative control	matched shapes minimize the defect; quartic confinement creates a persistent family gap
Does Δ_{solv} predict escape?	ten frozen mismatches	four high-signal coefficient fits agree within 0.8–6.1%; weak/floor-limited cases are retained
Is the defect estimable beyond 1D?	375 high-dimensional Monte Carlo settings	calibrated estimates through $d = 100$ with 0.17% mean relative error at $n = 10^5$
Does local mismatch matter repeatedly?	50-step restricted trajectories	mismatched errors reach 6–7 $\times 10^{-2}$ while the matched branch remains near 4×10^{-3}

Table 4: Evidence map linking each experiment to the corresponding theoretical question.

0.26%, 0.17%, 0.20%, and 0.22% for $d = 5, 10, 20, 50$, while $d = 100$ is 0.17%. Six 1000-resample bootstrap cases all contain the analytic truth in the 95% intervals for both Δ_{solv} and Δ^2_{solv}. At $d = 100, n = 100{,}000$, covariance-diagonal RMSE is about 4.3×10^{-3} and off-diagonal RMS about 3.1×10^{-3}.

B.5 Negative control and repeated-step utility

For $V(x) = 0.05x^4$, projection errors are $2.24, 4.38, 8.42, 15.64 \times 10^{-3}$ at $\tau = .0025, .005, .01, .02$. Duplicate initializations agree to the displayed precision and all gradient RMS values remain below 10^{-6}. At $\tau = .01$, one-step restricted-family errors are 3.23×10^{-4} (matched $\beta = 2$), 2.84×10^{-3} ($\beta = 1.5$), and 3.50×10^{-3} ($\beta = 3$). At 50 steps they become 4.01×10^{-3}, 6.09×10^{-2}, and 6.96×10^{-2}. All 150 stored full-solver trajectory rows are feasible and successful.

B.6 Two-dimensional implementation sanity check

The generic-density 2D block records 72 solves on 64^2 and 96^2 grids over $\tau \in \{.01, .05\}$, $\kappa \in \{.5, 1\}$, three regularizations, and three initializations. Every row passes the basic mass, nonnegativity, and finite-value checks. The best cells reach mean error below 2×10^{-3} and covariance Frobenius error below 10^{-2}, but regularization and initialization trends are not uniformly monotone; this block is therefore used only as an implementation sanity check.

C Reproducibility details

Experiments were run on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU (24 GB). The software environment uses Python 3.10.20 and PyTorch 2.5.1 built with CUDA 12.1; the core one-dimensional and Monte Carlo calculations use float64 NumPy/SciPy routines, while the GPU is used only where required by the 2D sanity check. No external benchmark dataset is used in this paper. The frozen one-dimensional protocol uses seeds $\{11, 23, 37, 53, 71\}$, and the independently rebuilt high-dimensional audit uses $\{89, 101, 131, 173, 211\}$. Six representative high-

dimensional cases use 1000 bootstrap resamples with settings fixed before the final aggregation. Raw arrays, CSV summaries, failed-run tables, optimizer and boundary-sensitivity records, command logs, environment files, SHA256 manifests, unit-test output, frozen configurations, bootstrap samples, and figure-rebuild scripts are retained. Figures are regenerated from persisted tables rather than by rerunning optimization.