CERTIFYING OUTPUT-SPACE SEPARATION OF CONDITIONAL LOCATION-SCALE MIXTURES

Anonymous authors
Paper under double-blind review

ABSTRACT

Analyses of conditional generative models routinely assume that conditions far apart are mapped to output distributions far apart in Wasserstein distance. We ask when this separation can instead be certified for a learned conditional location-scale mixture. For each pair of conditions, our decision procedure returns either an explicit lower bound on output distance or an honest refusal. Three certificates cover complementary regimes. A spectral moment certificate combines mean displacement with the full Bures distance between output covariances, using exact covariance addition under convolution. It has constant one, is dimension-free, and detects anisotropic changes that a trace-only certificate misses. A heat-semigroup certificate converts an integrated absolute density gap into an output-transport lower bound with a three-halves power law for a fixed Laplace kernel in every dimension; scale interpolation extends it to different kernels with an additive Bures slack. A latent inverse certificate feeds these output bounds. Our direct proof extends to every kernel admitting finite-order PDE inversion and yields an exponent twice the inversion order, exactly matching the detour through the first Wasserstein distance. Thus direct inversion does not improve the exponent within this proof template. We also make two normally implicit side conditions checkable. The globally Lipschitz gradient needed by the quadratic refinement reduces to strong convexity of the dual potential; it holds for single-point targets and in the classical smooth-density regime but fails for multi-cell atomic transports. A mollifier uncertainty principle forces spatial moment and derivative size to grow jointly at least at the square-root-of-dimension scale. Consequently, the best latent constant within this proof architecture deteriorates like the reciprocal of dimension to the three-halves power, regardless of mollifier choice. Numerically, we verify the heat bridge in dimensions two, four, and eight without empirical optimal transport, exhibit an anisotropic pair that defeats the trace certificate, and certify a trained mixture density network. Throughout, we distinguish certification from refusal: the exponents are dimension-free, the latent constants are not, and the composed guarantee is polynomial rather than a linear coarse embedding.

1 INTRODUCTION

Conditional generative models learn a map from a condition space $\mathcal{C}$ to a family of output distributions, and a recent line of work makes out-of-distribution generalisation precise by treating $c \mapsto \mu_c \in (\mathcal{P}_2(\mathbb{R}^d), W_2)$ as a curve in Wasserstein space (Pham, 2026). The resulting bound is only as good as one geometric premise: that the map is a *coarse embedding*,

$$a\, d_{\mathcal{C}}(c, c') - b \;\le\; W_2(\mu_c, \mu_{c'}) \;\le\; A\, d_{\mathcal{C}}(c, c') + B. \tag{1}$$

The upper bound is a Lipschitz-type assumption on the learned map; the lower bound carries the geometry, and it is almost always assumed. We ask when it can instead be *certified* - computed from a model one already has, with a proof attached - and answer with a decision procedure (Algorithm 1) returning either an explicit lower bound on $W_2(\mu_c, \mu_{c'})$ or an explicit refusal. Since $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is snowflake-universal, no general bi-Lipschitz statement exists without structure on the model family (Andoni et al., 2018), and our compact-support hypothesis forces a condition space of finite diameter (Remark 3), so every conclusion is a quantitative separation guarantee on a bounded region.

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.

We work with location-scale mixtures $p_{G_c}(x) = \int f(x \mid \theta, \Sigma(c)) P_c(d\theta)$, one scale matrix shared across the components of each model. Three certificates cover complementary regimes; Table 1 collects their constants.

(i) A spectral moment certificate. Convolution propagates moments exactly: $\Sigma_{\mu_c} = \Sigma_{P_c} + \Sigma_{\nu_c}$, with kernel covariance $\Sigma_{\nu_c} = \Sigma(c)$ for Laplace. Feeding this into the Gelbrich lower bound gives $W_2^2(\mu_c, \mu_{c'}) \geq \|\Delta m\|^2 + \mathfrak{B}^2(\Sigma_{\mu_c}, \Sigma_{\mu_{c'}})$ with $\mathfrak{B}$ the Bures distance (Proposition 2): constant one, dimension-free, closed form in the model parameters, and linear in $d_{\mathcal{C}}$ whenever that displacement is (Assumption 6). It is *strictly* stronger than the trace-variance form used in prior analyses - on a rotated anisotropic pair the trace certificate is identically zero while Bures equals the true W_2 (Example 1).

(ii) A density-to-output bridge. When the moments are stationary that certificate is silent and one must convert a density gap $V = \|p_{G_c} - p_{G_{c'}}\|_1$ into transport distance - not automatic, since a mean-zero high-frequency ripple keeps V fixed while $W_2 \to 0$. For a fixed Laplace kernel a heat-semigroup argument gives $W_1 \geq c_{\mathrm{heat}} V^{3/2}$ in *every* dimension with c_{heat} independent of d (Proposition 4), and scale interpolation extends it to different kernels at the cost of an additive Bures slack (Proposition 5).

(iii) A latent inverse bound to feed the bridge. For the multivariate Laplace kernel, Bariletto et al. (2026, Theorem 8) prove

$$\|p_G - p_{G'}\|_1 \;\geq\; \bar{C}_d \big[W_1(P, P')^2 + \|\Sigma - \Sigma'\|_{\mathrm{op}} \big], \tag{2}$$

which with $W_2^2 \leq 2RW_1$ already implies a quartic latent W_2 inequality; we do not claim that consequence as new. We add a proof acting directly on W_2^2 (Theorem 1) and, from it, a meta-theorem (Theorem 2) covering every kernel with order-β PDE inversion, with exponent 2β. The W_1 detour gives the same 2β, so the exponent is route-independent: within this template no order-β argument does better, so extending beyond Laplace buys coverage, not rate.

Certificates are useful only if their side conditions are checkable, and two of ours are the hypotheses such theorems usually leave dangling. The quadratic latent refinement needs global $C^{1,1}$ regularity of the Brenier potential; we reduce it to strong convexity of the dual potential and give regimes where that holds or provably fails (Proposition 6). The latent constant depends on the mollifier through M_ψ, M'_ψ; an L^1 uncertainty principle $M_\psi M'_\psi \geq \sqrt{d}/2$ (Lemma 4) shows $C_d = \Theta(d^{-3/2})$ is the limit of this proof architecture, not of our mollifier choice.

Contributions.

- A Bures/Gelbrich output certificate that tracks $\Sigma_\mu = \Sigma_P + \Sigma_\nu$, dominates the trace-variance bridge, is attained on Gaussian pairs, and detects anisotropy trace certificates cannot (Proposition 2, Example 1).
- A main-text proof of the all-dimensional fixed-Laplace heat bridge pinpointing where $\lambda_{\min}$ enters, plus its extension to $\Sigma \neq \Sigma'$ with an explicit activation threshold (Proposition 5); and a meta-theorem giving exponent 2β for order-β inversion, with the comparison showing the W_1 detour reaches the same exponent (Theorem 2).
- Closure of the two standing hypotheses: sufficient and failing regimes for $C^{1,1}$ (Proposition 6) and a matching lower bound on the mollifier constants (Lemma 4).
- A certification recipe (Algorithm 1) and a numerical pipeline: the heat-bridge chain verified at $d = 2, 4, 8$ without empirical optimal transport, the anisotropic pair on which the trace certificate is blind, the mollifier trade-off against its proved floor, a trained mixture density network certified end to end, and three sharpness constructions reported as failures (Section 4).

2 Related Work and Background

Mixture identifiability and inverse bounds. Classical results establish identifiability of finite location-scale mixtures (Teicher, 1963); Ho & Nguyen (2016) sharpen this to explicit polynomial rates linking the L^1 distance between mixture densities to the Wasserstein distance between mixing

measures, with the achievable order governed by how many moments a near-degenerate configuration can match - the mechanism behind our sharpness probes in Section 4. Bariletto et al. (2026) derive inverse bounds for compactly supported, possibly infinite mixing measures and give posterior applications to Dirichlet-process mixtures. Their model has an unknown scale shared within each mixture, and they introduce the PDE-inversion condition used here: an operator T_Σ with $T_\Sigma p_G = P$. Their Laplace theorem already permits $\Sigma \neq \Sigma'$ and, together with bounded support, implies our quartic latent W_2 inequality. We therefore use it as a cross-check and present Theorem 1 only as an independent direct proof. Theorem 2 extends the direct proof to any kernel admitting order-β inversion - for instance the variance-gamma / generalised hyperbolic family with characteristic function $(1 + t^\top \Sigma t/2)^{-\alpha}$, $\alpha \in \mathbb{N}$ (Madan & Seneta, 1990; Barndorff-Nielsen, 1997), where $\beta = 2\alpha$ - and shows that the resulting exponent 2β coincides with what the W_1 route of Bariletto et al. (2026) would give for the same operator order, so the two routes differ in constants and in which metric they act on, not in exponent. Our Proposition 7 replaces a coordinatewise summation for the second-order Laplace Hessian term by an operator/trace-norm bound; it does not compare or determine the complete hidden constants.

Wasserstein geometry of conditional models. Pham (2026) formalise OOD generalisation for conditional flow matching by viewing $c \mapsto \mu_c$ as a curve in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and bounding error via Wasserstein barycentric interpolation, under the coarse embedding relaxation of Equation (1) - justified because $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is snowflake-universal (Andoni et al., 2018; Pritchard & Weighill, 2025). That framework *assumes* the lower bound in Equation (1). Here we give sufficient moment conditions for that lower bound and a more limited same-kernel polynomial bridge. The heat-semigroup version covers a fixed Laplace kernel in every dimension, but not comparisons whose kernels differ. We do not derive an unconditional coarse embedding for the full condition-dependent-scale class.

Moment-based Wasserstein bounds. That two measures with given mean and covariance are at least as far apart as the Gaussians matching those moments is classical: the Gaussian closed form is due to Dowson & Landau (1982), the general lower bound to Olkin & Pukelsheim (1982) and Gelbrich (1990), and the resulting Bures-Wasserstein geometry on positive definite matrices is surveyed by Bhatia et al. (2019). Proposition 2 *invokes* that bound rather than the trace-norm corollary obtained by Cauchy-Schwarz, which Example 1 shows is strictly and sometimes totally lossy. What is ours here is the mixture-specific part: the convolution identities $m(\mu) = m(P) + m(\nu)$ and $\Sigma_\mu = \Sigma_P + \Sigma_\nu$ that make the certificate computable from model parameters, the kernel scale drift entering Σ_μ additively so that heteroscedastic families are covered, and the consequences for when the certificate is silent (Assumption 6 and Corollary 3).

Mixture model and Laplace kernel. A model is specified by a mixing distribution P over locations θ and a scale matrix Σ shared by all mixture components. We encode this pair as $G = P \times \delta_\Sigma$, where δ_Σ is the Dirac mass at Σ; thus, the scale is fixed within each model rather than mixed. The corresponding mixture density is

$$p_G(x) = \int f(x \mid \theta, \Sigma)\, P(d\theta).$$

Since f is a location kernel, $p_G = P * f_\Sigma$ with $f_\Sigma = f(\cdot, \Sigma) := f(\cdot \mid 0, \Sigma)$; f_Σ denotes a generic kernel unless a statement specifies the Laplace density below. We assume that $\Sigma \in \mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$, meaning that Σ is positive definite and all its eigenvalues lie in $[\lambda_{\min}, \lambda_{\max}]$.

Following Bariletto et al. (2026), our main example is the multivariate Laplace location kernel. Conditional on θ, a draw from this kernel can be represented as

$$X = \theta + \sqrt{E}\, Z, \qquad E \sim \mathrm{Exp}(1), \qquad Z \sim N(0, \Sigma),$$

where E and Z are independent. The centred noise $\sqrt{E}\, Z$ has characteristic function

$$\mathbb{E}\left[e^{it^\top \sqrt{E} Z}\right] = \left(1 + \frac{1}{2} t^\top \Sigma t\right)^{-1}.$$

Consequently, the mixture satisfies the distributional inversion identity

$$T_\Sigma p_G = P, \qquad T_\Sigma := \mathrm{Id} - \frac{1}{2}\,\Sigma : \mathrm{Hess}.$$

Here $A : B := \mathrm{tr}(A^\top B)$ denotes the Frobenius contraction. Thus, the Laplace kernel admits PDE inversion of order $\beta = 2$. For two models G and G', we abbreviate the L^1 distance between their densities by $V := \|p_G - p_{G'}\|_1$.

Assumptions and transport potential. The condition-indexed results additionally require uniform compact support of the mixing distributions (Assumption 1), Lipschitz regularity of the map from conditions to mixing laws (Assumption 2), and Lipschitz variation of the scale matrix (Assumption 3). Their precise statements appear in Section A.1. These are modelling assumptions and are not consequences of the PDE inversion identity.

To compare two mixing distributions P and P' in quadratic Wasserstein distance, we use an optimal Kantorovich potential in Brenier form, $\phi^*(x) = \frac{1}{2}\|x\|^2 - u(x)$, where u is convex. For the canonical c-transform representative, $\partial u(x) \subseteq \mathrm{conv}(\mathrm{supp}(P'))$. Hence, if P' is supported in $B(0, R)$, every subgradient of u has norm at most R, and at every differentiability point $x \in B(0, R)$, $\|\nabla\phi^*(x)\| = \|x - \nabla u(x)\| \le \|x\| + \|\nabla u(x)\| \le 2R$. This bound is dimension-free, but it is only local and first-order. In particular, it does not imply a global Lipschitz bound or global $C^{1,1}$ regularity for ϕ^*. The cutoff construction in Section A.2 is needed precisely because of this distinction.

3 METHOD

Throughout, T_Σ denotes the *explicit* Laplace inversion operator $\mathrm{Id} - \frac{1}{2}\Sigma : \mathrm{Hess}$ of Section 2, and $V := \|p_G - p_{G'}\|_1$; Assumption 4 states the two adjoint-action inequalities used below, which for this operator follow from operator/trace-norm duality.

3.1 A ONE-SIDED KANTOROVICH REPRESENTATION

The W_1 dual uses a single witness, which is what PDE inversion needs; the W_2^2 dual uses two. The following collapses the latter to the former at the cost of an inequality, which suffices because we need only an *upper* bound on W_2^2 to invert into a *lower* bound on V.

Lemma 1 (One-sided Kantorovich representation)**.** *Let $P, P' \in \mathcal{P}_2(\mathbb{R}^d)$ and let (ϕ^*, ψ^*) be optimal Kantorovich potentials for the cost $c(x, y) = \frac{1}{2}\|x - y\|^2$, with ϕ^* integrable under both P and P'. Then $\frac{1}{2}W_2^2(P, P') \le \int \phi^*\, d(P - P')$, with equality iff $\psi^* = -\phi^*$ P'-a.e.*

The proof is two lines (Section A.1): the last dual term $\int(\phi^* + \psi^*)dP'$ is nonpositive by feasibility at (y, y).

3.2 W_2 INVERSE BOUNDS WITH A DIMENSION-INDEPENDENT EXPONENT

Theorem 1 (Quartic inverse bound)**.** *Let $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_\Sigma$ share the same Laplace scale matrix. Suppose that P and P' are supported in $B(0, R)$ and that the adjoint-action bounds in Assumption 4 hold. Then there exists $C_d > 0$ such that*

$$\|p_G - p_{G'}\|_1 \;\ge\; C_d\, W_2(P, P')^4.$$

The exponent 4 is independent of the ambient dimension. The constant C_d, however, may depend on d, R, $\lambda_{\max}$, and the fixed cutoff and mollifier. An explicit admissible value of C_d is given in Section A.2.

Proof idea. Lemma 1, localisation of the optimal potential, mollification at scale Λ^{-1}, and distributional PDE inversion give, for $\Lambda \ge 1$,

$$\tfrac{1}{2}W_2^2(P, P') \;\le\; 2C_{\mathrm{op}}\Lambda\, V \;+\; \frac{C_0}{\Lambda}. \tag{3}$$

Balancing the two terms yields $W_2^2(P, P') \lesssim V^{1/2}$ and hence the quartic lower bound. The cutoff construction, the optimisation over Λ, the small-distance case, and the discussion of why the substitution $\Lambda = 1/W_2(P, P')$ is invalid are deferred to Section A.2.

Theorem 2 (Order-β inverse bound)**.** *Let the kernel admit order-β PDE inversion in the quantitative sense of Assumption 5, and let Assumption 1 hold. Then for all P, P' supported in $B(0, R)$,*

$$V \;\ge\; C_d(\beta)\, W_2(P, P')^{2\beta}.$$

The exponent 2β is dimension-independent, whereas the constant $C_d(\beta)$ need not be. Its explicit form and the complete proof are given in Section A.4.

Remark 1 (The exponent is route-independent). The same three steps against the W_1 dual give $V \geq C'W_1^{\beta}$, and $W_1 \geq W_2^2/(2R)$ turns this into $V \geq C'(2R)^{-\beta}W_2^{2\beta}$: the *same* exponent (Section A.4). Theorem 2 therefore buys coverage and constants, not rate. Kernels with characteristic function $(1 + t^{\top}\Sigma t/2)^{-\alpha}$, $\alpha \in \mathbb{N}$, invert with $T_{\Sigma} = (\mathrm{Id} - \frac{1}{2}\Sigma : \mathrm{Hess})^{\alpha}$, giving $\beta = 2\alpha$ and exponent 4α; Gaussian kernels admit no finite-order inversion and fall outside the template entirely, leaving them only the output bridges of Section 3.3.

A Lipschitz gradient yields a stronger conditional conclusion.

Theorem 3 (Conditional quadratic bound)**.** *Under the same shared-scale hypotheses as Theorem 1, suppose in addition that $\nabla\phi^*$ is globally M-Lipschitz (equivalently, $D^2\phi^* \in L^{\infty}$ as a weak derivative with operator norm at most M). Then $\|p_G - p_{G'}\|_1 \geq C'_d W_2(P, P')^2$ with $C'_d = \left(15R^2 + \lambda_{\max} d\widetilde{M}\right)^{-1}$ and $\widetilde{M} = M + 8C_1 + \frac{15}{2}C_2$.*

Global $C^{1,1}$ regularity of a Brenier potential is delicate, and densities merely bounded above and below on a convex domain do *not* give it (Caffarelli, 1992; 1996; De Philippis & Figalli, 2014). Rather than leave the hypothesis dangling, we isolate checkable sufficient and failure conditions in Proposition 6, proved and discussed in Section A.6. In particular, the quadratic regime covers single-Dirac and Caffarelli-type regular settings, while multi-cell atomic targets remain in the quartic regime.

Condition-dependent shared scale. When $\Sigma(c)$ varies, the quartic exponent persists, with an additive scale-mismatch slack. In the pairwise form,

$$\|p_{G_c} - p_{G_{c'}}\|_1 \geq C_{4,d} W_2(P_c, P_{c'})^4 - L_d \|\Sigma(c) - \Sigma(c')\|_{\mathrm{op}},$$

and under Assumption 3 this yields $\|p_{G_c} - p_{G_{c'}}\|_1 \geq C_{4,d} W_2(P_c, P_{c'})^4 - L_d L_{\Sigma} d_{\mathcal{C}}(c, c')$. The operator-mismatch decomposition and dimension bookkeeping are deferred to Section B.

The refined Hessian bound and the mollifier uncertainty principle are stated in Proposition 7 and Lemma 4 and proved in Sections B.3 and B.4. Together they imply that $M'_{\psi} = O(1)$ can make the Hessian term $O(d)$, but necessarily with $M_{\psi} = \Omega(\sqrt{d})$, so

$$C_d \;=\; \Theta\Big(d^{-3/2}\Big) \tag{4}$$

for this first-order mollification architecture. Full constant bookkeeping is deferred to Sections A.2 and B.4.

3.3 Bridging to the output space

Theorems 1, 2 and 4 constrain $\|p_{G_c} - p_{G_{c'}}\|_1$; the geometry needs $W_2(\mu_c, \mu_{c'})$, and the naive implication is false - a mean-zero, high-frequency ripple keeps $\|p - q\|_1$ fixed while the transport cost vanishes with the wavelength (Figure 3). Any valid bridge must use structure of the mixture class; our first two sidestep L^1 entirely, using moment identities that convolution propagates exactly.

Proposition 1 (First-moment bridge)**.** *Let $P, P', \nu, \nu' \in \mathcal{P}_1(\mathbb{R}^d)$ and set $\mu = P * \nu$, $\mu' = P' * \nu'$. Then $W_1(\mu, \mu') \geq \|(m(P)+m(\nu))-(m(P')+m(\nu'))\|$; in particular, if the kernels are centred (or shared), this is $\|m(P) - m(P')\|$. If all four measures lie in $\mathcal{P}_2$, then also $W_2(\mu, \mu') \geq W_1(\mu, \mu')$.*

Proposition 2 (Bures bridge)**.** *Let $P, P', \nu, \nu' \in \mathcal{P}_2(\mathbb{R}^d)$, $\mu = P * \nu$, $\mu' = P' * \nu'$, and write $\Sigma_{\mu} = \Sigma_P + \Sigma_{\nu}$, $\Sigma_{\mu'} = \Sigma_{P'} + \Sigma_{\nu'}$. Then*

$$W_2(\mu, \mu')^2 \;\geq\; \|m(\mu) - m(\mu')\|^2 + \mathfrak{B}^2(\Sigma_{\mu}, \Sigma_{\mu'}),$$

where $\mathfrak{B}^2(A, B) = \mathrm{tr}A + \mathrm{tr}B - 2\mathrm{tr}[(A^{1/2}BA^{1/2})^{1/2}]$ is the squared Bures distance, with equality when μ, μ' are Gaussian. Since $\mathrm{tr}[(A^{1/2}BA^{1/2})^{1/2}] \leq \sqrt{\mathrm{tr}A\,\mathrm{tr}B}$, the right-hand side dominates the trace-variance form $\|m(\mu) - m(\mu')\|^2 + (\sqrt{\mathrm{tr}\Sigma_{\mu}} - \sqrt{\mathrm{tr}\Sigma_{\mu'}})^2$ (Corollary 1).

Example 1 (The trace certificate can be identically zero)**.** Take $d = 2$, $\Sigma_P = \mathrm{diag}(\sigma^2, 0)$, $\Sigma_{P'} = \mathrm{diag}(0, \sigma^2)$ and a shared isotropic kernel $\Sigma_{\nu} = \varepsilon^2\mathrm{Id}$, admissible for every $\varepsilon > 0$. The two traces

agree, so the trace certificate is 0 for every ε, while $\mathfrak{B}^2 = 2(\sigma^2 + 2\varepsilon^2) - 4\varepsilon\sqrt{\sigma^2 + \varepsilon^2} \to 2\sigma^2$; for centred Gaussians the latter is exactly W_2^2 (Dowson & Landau, 1982). Figure 1 (left) measures both.

Corollary 1 (Trace-variance bridge)**.** $W_2(\mu, \mu')^2 \geq \|m(\mu) - m(\mu')\|^2 + (\sqrt{\mathrm{tr}\Sigma_\mu} - \sqrt{\mathrm{tr}\Sigma_{\mu'}})^2$.

Both are proved in Section C; neither needs bounded support, PDE machinery, or a dimension-dependent constant, and both cover condition-varying kernels $\nu_c = f(\cdot, \Sigma(c))$. Together they give a constant-one, *linear* coarse lower bound whenever the mean-and-covariance displacement is quasi-isometric in $d_{\mathcal{C}}$ (Assumption 6 and Corollary 3). Crucially Assumption 2 does *not* imply Assumption 6 - the standard moment-matched construction has $W_1(P_c, P_{c'}) = \Theta(d_{\mathcal{C}})$ with mean and covariance exactly fixed - so this is a real gap, not a corner case.

For that residual case we mollify a witness built from the density difference itself. The key asymmetry is that $g := p_G - p_{G'} = (P - P') * f(\cdot, \Sigma)$ inherits the kernel's regularity *however rough* P, P' are, since only $\|P - P'\|_{\mathrm{TV}} \leq 2$ is used; so, unlike Theorem 1, the mollified object is regular and pays $O(1/\Lambda^2)$ while the discontinuous sign costs Λ^{+1} in the Lipschitz constant (Figure 4). The relevant kernel constant is *not* the Laplacian norm but a directional second-derivative total variation,

$$K_2(\Sigma, d) \;:=\; \sup_{\|e\|=1} \big\|\partial_{ee}^2 f(\cdot, \Sigma)\big\|_{\mathrm{TV}}. \tag{5}$$

Proposition 3 (Sign-mollification bridge)**.** *Let G, G' share a fixed kernel density f_Σ (any location kernel, not necessarily Laplace) with $P, P' \in \mathcal{P}_2(\mathbb{R}^d)$ and $f_\Sigma(x)dx \in \mathcal{P}_2(\mathbb{R}^d)$; set $\mu = P * f_\Sigma$, $\mu' = P' * f_\Sigma$, $V = \|p_G - p_{G'}\|_1$. Fix a nonnegative even probability mollifier $\psi \in W^{1,1}$ with finite $m_2(\psi), M_1(\psi)$ and assume $0 < K_2(\Sigma, d) < \infty$. Then $W_1(\mu, \mu') \geq C_{\mathrm{sign}} V^{3/2}$ with $C_{\mathrm{sign}} = \frac{2}{3} M_1(\psi)^{-1} (3 m_2(\psi) K_2)^{-1/2}$, and the same for W_2. No non-degeneracy condition on P, P' is required, and they may have infinitely many atoms.*

Remark 2 (When K_2 is finite). $K_2 < \infty$ for any kernel with integrable second derivative, in particular the Gaussian in *all* dimensions. For the multivariate Laplace kernel it holds *only in* $d = 1$, where $f_s'' = (2/s)(f_s\, dx - \delta_0)$ and $K_2 = 4/s$; for $d \geq 2$ a fixed directional second derivative has logarithmically divergent total variation at the origin (Section C.5). This restricts the argument, and is not an impossibility result.

Proposition 4 (Fixed-Laplace heat bridge)**.** *Let $P, P' \in \mathcal{P}_2(\mathbb{R}^d)$, fix $\Sigma \in \mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$, and let f_Σ now denote the common multivariate Laplace density of Section 2. With $\mu = P * f_\Sigma$, $\mu' = P' * f_\Sigma$ and $V = \|p_G - p_{G'}\|_1$, in every dimension*

$$W_2(\mu, \mu') \;\geq\; W_1(\mu, \mu') \;\geq\; c_{\mathrm{heat}}\, V^{3/2}, \qquad c_{\mathrm{heat}} = \frac{\sqrt{\pi \lambda_{\min}}}{3\sqrt{6}}.$$

Proof sketch (full proof in Section C.5). Write $Q = P - P'$, $g = p_G - p_{G'}$, $L_\Sigma = \Sigma : \mathrm{Hess}$. **(i)** Laplace inversion gives the generator identity $L_\Sigma g = 2(g\, dx - Q)$, hence $\|L_\Sigma g\|_{\mathrm{TV}} \leq 2(\|g\|_1 + \|Q\|_{\mathrm{TV}}) \leq 8$. **(ii)** With $(H_t)_{t \geq 0} = \exp(tL_\Sigma)$, convolution with $N(0, 2t\Sigma)$, the semigroup identity and L^1-contraction give $\|H_t g - g\|_1 \leq \int_0^t \|H_s L_\Sigma g\|_{\mathrm{TV}} ds \leq 8t$. **(iii)** For $h_t = H_t \,\mathrm{sign}(g)$ and any unit e, $\|\partial_e q_t\|_1 = \sqrt{e^\top \Sigma^{-1} e/(\pi t)} \leq (\pi \lambda_{\min} t)^{-1/2}$, so $\mathrm{Lip}(h_t) \leq (\pi \lambda_{\min} t)^{-1/2}$. *This is the only place $\lambda_{\min}$ enters*, and the quantity that really appears is $\sup_{\|e\|=1} e^\top \Sigma^{-1} e = \lambda_{\min}(\Sigma)^{-1}$, so the constant improves whenever the actual smallest eigenvalue exceeds the class bound of Assumption 1. **(iv)** Self-adjointness of H_t gives $\int h_t g = \int \mathrm{sign}(g) H_t g \geq V - 8t$; normalising h_t in the Kantorovich-Rubinstein dual and maximising $(V - 8t)\sqrt{\pi \lambda_{\min} t}$ at $t = V/24$ gives the claim. □

Step (i) is where a shared kernel is used: for $\Sigma \neq \Sigma'$ the operator L_Σ does not invert $p_{G'}$, so the identity fails. Interpolating in the scale repairs this, at the price of an additive Bures slack.

Proposition 5 (Scale-interpolation bridge, $\Sigma \neq \Sigma'$)**.** *Let $f_\Sigma, f_{\Sigma'}$ be multivariate Laplace densities, $\mu = P * f_\Sigma$, $\mu' = P' * f_{\Sigma'}$ with $\Sigma, \Sigma' \in \mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$, $\Delta = \Sigma' - \Sigma$, $\kappa_\Sigma = \lambda_{\max}/\lambda_{\min}$. If $V \geq C_\delta \|\Delta\|_F$ then $W_1(\mu, \mu') \geq c_{\mathrm{heat}} V^{3/2} - C_1 \|\Delta\|_F V^{1/2} - C_2 \|\Delta\|_F$ with $C_1 = \frac{3}{2} c_{\mathrm{heat}} C_\delta$, $C_2 = (2\sqrt{\lambda_{\min}})^{-1}$, $C_\delta = \sqrt{\gamma_{\kappa_\Sigma}}/\lambda_{\min}$ and $\gamma_\kappa = (\kappa - \kappa^2 + \kappa^2 \log \kappa)/(\kappa - 1)^2$ ($\gamma_1 = \frac{1}{2}$). The right-hand side is nondecreasing in V on that range, so any lower bound on V - for instance Theorem 4, the statement available when $\Sigma \neq \Sigma'$ - may be substituted.*

Table 1: Summary of the available certificates. Exact constants are given in the referenced statements. The moment and heat constants are dimension-independent; the latent constants are not.

Certificate	Lower bound	Main requirement
Moment (Propositions 1 and 2)	$W_2^2 \geq \|\Delta m\|^2 + \mathfrak{B}^2$	Finite second moments
Sign mollification (Proposition 3)	$W_1 \geq C_{\mathrm{sign}} V^{3/2}$	Shared kernel; $K_2(\Sigma, d) < \infty$
Fixed Laplace (Proposition 4)	$W_1 \geq c_{\mathrm{heat}} V^{3/2}$, $c_{\mathrm{heat}} = \sqrt{\pi \lambda_{\min}}/(3\sqrt{6})$	Shared Laplace scale; any d
Scale interpolation (Proposition 5)	$W_1 \geq c_{\mathrm{heat}} V^{3/2} - C_1 \|\Delta\|_F V^{1/2} - C_2 \|\Delta\|_F$	Laplace; $V \geq C_\delta \|\Delta\|_F$
Quartic inverse (Theorem 1)	$V \geq C_d W_2(P, P')^4$	Compact support; shared Σ
Scale-drift inverse (Theorem 4)	$V \geq C_{4,d} W_2(P, P')^4 - L_d \|\Delta\|_{\mathrm{op}}$	Compact support; $\Sigma \neq \Sigma'$
Order-β inverse (Theorem 2)	$V \geq C_d(\beta) W_2(P, P')^{2\beta}$	Order-β PDE inversion
Quadratic inverse (Theorem 3)	$V \geq C'_d W_2(P, P')^2$	Global $C^{1,1}$ potential
Composed certificate (Corollary 2)	$W_2(\mu, \mu') \geq c_{\mathrm{heat}} C_d^{3/2} W_2(P, P')^6$	Quartic inverse and fixed Laplace

Algorithm 1 Certifying output separation for a pair of conditions

Require: conditions c, c'; fitted $P_c, P_{c'}$; kernels $\nu_c = f(\cdot, \Sigma(c))$ with covariance Σ_{ν_c} ($= \Sigma(c)$ for Laplace)

1: $\rho \leftarrow \left(\|m(\mu_c) - m(\mu_{c'})\|^2 + \mathfrak{B}^2(\Sigma_{P_c} + \Sigma_{\nu_c}, \Sigma_{P_{c'}} + \Sigma_{\nu_{c'}})\right)^{1/2}$ ▷ Proposition 2: constant one, dimension-free
2: **if** $\rho > 0$ **then return** linear certificate $W_2(\mu_c, \mu_{c'}) \geq \rho$
3: $\Delta \leftarrow \Sigma(c') - \Sigma(c)$; $v \leftarrow$ lower bound on V from Theorem 1 (Laplace, $\Delta = 0$), Theorem 2 (order-β, $\Delta = 0$) or Theorem 4 (Laplace, $\Delta \neq 0$); else $v \leftarrow 0$
4: **if** $v \leq 0$ **then return no certificate**
5: **if** Laplace and $\Delta = 0$ **then return** $W_2 \geq c_{\mathrm{heat}} v^{3/2}$ ▷ Proposition 4, any d
6: **else if** $\Delta = 0$ and $K_2(\Sigma(c), d) < \infty$ **then return** $W_2 \geq C_{\mathrm{sign}} v^{3/2}$ ▷ Proposition 3
7: **else if** Laplace and $v \geq C_\delta \|\Delta\|_F$ **then return** $W_2 \geq c_{\mathrm{heat}} v^{3/2} - C_1 \|\Delta\|_F v^{1/2} - C_2 \|\Delta\|_F$ ▷ Proposition 5
8: **return no certificate**

The hypothesis $V \geq C_\delta \|\Delta\|_F$ is a "signal above scale drift" condition - drift alone can account for the whole density gap - and Section C.6 gives the resulting activation threshold in $d_{\mathcal{C}}$, mirroring Definition 1. At $\Delta = 0$ it degrades to Proposition 4, and composing with Theorem 1 recovers:

Corollary 2 (Composed fixed-Laplace lower bound)**.** *Under the hypotheses of Theorem 1 and Proposition 4 with a shared fixed Laplace kernel,* $W_2(\mu, \mu') \geq W_1(\mu, \mu') \geq c_{\mathrm{heat}} C_d^{3/2} W_2(P, P')^6$*, hence at least* $c_{\mathrm{heat}} C_d^{3/2} (\alpha d_{\mathcal{C}}(c, c') - \kappa)_+^6$ *under Assumption 2. This is polynomial separation, not a linear coarse embedding.*

3.4 Constants at a glance, and which certificate to use

Table 1 collects every bound above with its constant and its coverage, and Algorithm 1 turns the collection into a procedure: try the spectral certificate first, since it is linear and dimension-free; fall back to a density bridge when the moments are stationary; and report *no certificate* rather than a vacuous one when neither applies.

4 Experiments

The Bures certificate detects separation the trace certificate cannot. On the anisotropic pair of Example 1 at $\sigma = 1$, sweeping ε from 1 to 10^{-3} (Figure 1, left), the trace-variance certificate

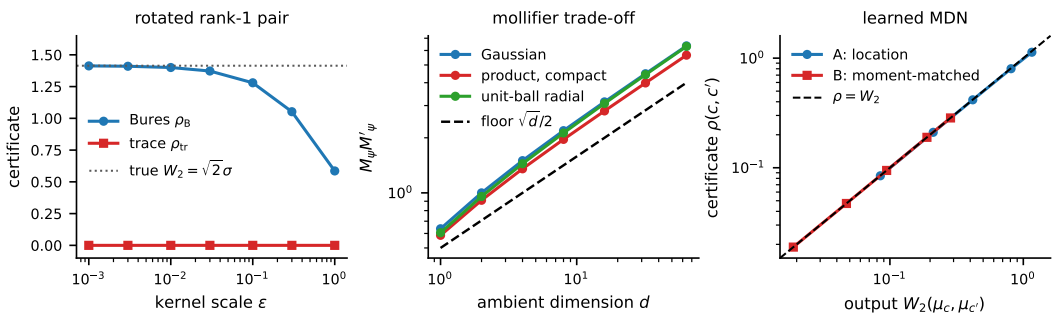

Figure 1: Left: on the rotated anisotropic pair the trace certificate is identically zero while Bures attains the true W_2 (Example 1). Middle: the mollifier trade-off $M_\psi M'_\psi$ against the floor $\sqrt{d}/2$ of Lemma 4; no family escapes it. Right: the certificate of Proposition 2 on a *trained* mixture density network against the exact output W_2; points on or below the diagonal are valid.

is exactly 0 at every ε - not small, zero - while Bures rises to 1.4128 against $\sqrt{2}\sigma = 1.4142$. Both measures are centred Gaussians, so that value *is* the true W_2 (Dowson & Landau, 1982): the certificate is not merely nonzero but attained. A wide kernel, $\Sigma_\nu = 0.5\,\mathrm{Id}_2$ ($\varepsilon^2 = 0.5$), leaves trace at 0 and Bures at 0.732 - kernel variance shrinks the eigenvalue gap without hiding it.

No mollifier escapes the $\sqrt{d}$ trade-off. We evaluate M_ψ, M'_ψ in closed form for a Gaussian, a compactly supported product family and a unit-ball radial family, $d = 1, \ldots, 64$ (Figure 1, middle; Table 12). The first two have M'_ψ exactly constant in d ($\sqrt{2/\pi}$ and $15/8$) with $M_\psi = \Theta(\sqrt{d})$; the radial family reverses the roles. Every family sits within a factor 1.17-1.59 of the floor of Lemma 4, and none has both constants $O(1)$: $M'_\psi(d) = O(1)$ *is* available, so Proposition 7's factor is genuinely $O(d)$, but the product cannot be, so $C_d = \Theta(d^{-3/2})$ is this proof's ceiling.

The heat bridge, verified step by step at $d = 2, 4, 8$. The proof of Proposition 4 is a chain of four inequalities, and we check each rather than only the conclusion, since a conclusion holding while an intermediate step fails would hide a proof error. At $d = 2$ all four are exact on a grid (Bessel closed form for the density, Gauss-Laguerre for the heat flow): $\|L_\Sigma g\|_{\mathrm{TV}} = 6.05 \le 8$; $\|H_t g - g\|_1 = 0.171, 0.270, 0.401, 0.574$ against $8t$ at $t = 0.05, 0.1, 0.2, 0.4$; $\mathrm{Lip}(h_t) = 1.442 \le 1.591$; and the witness value 0.292 against the predicted $c_{\mathrm{heat}} V^{3/2} = 0.204$. At $d = 4, 8$ the integrands stay exact and only the integrals are sampled (Section E.8); every inequality holds, with witness values $0.600 \ge 0.418$ and $0.459 \ge 0.280$. Because the witness is a genuine Kantorovich-Rubinstein test function, its normalised pairing is itself a lower bound on W_1: the high-dimensional corroboration empirical optimal transport could not deliver, without estimating a transport plan. Two caveats: bands are 3 standard errors, and $\widehat{V}$ enters both sides of the last inequality through $t^* = \widehat{V}/24$, so those errors are correlated.

Certifying a trained model. We train a mixture density network (Bishop, 1994) with five Laplace components and a shared scale - exactly the model class of Section 2 - and run Algorithm 1 on the *learned* parameters (Figure 1, right). On task A (a location encoding) the certificate is linear in $d_{\mathcal{C}}$ (fitted log-log slope 0.962 against 0.969 for the true output W_2), and $\rho \le W_2$ at every pair with a gap of at most 2.4%. On task B the network is trained on the moment-stationary family of Section E.4, where the certificate should be silent; instead it is active and matches W_2 to within 0.01%. The reason is instructive: the network did not reproduce the moment-matched structure, and the conditionals it learned form a location-scale family in c, for which $W_2^2 = \|\Delta m\|^2 + (\sigma - \sigma')^2$ is an equality. The certificate reports a property of the model actually fitted - what a practitioner needs - and correctly flags that this model is not the target.

The moment bridges are silent exactly where predicted, and three constructions still fail to settle sharpness. Exact 1D quantile computation over four perturbation families (Table 9) shows

translation and small-mass transfer matching the orders of the mean bound, split/merge matching the covariance bound, and moment-matched vanishing at strictly higher order - the regime where Propositions 1 and 2 are silent and only the density bridges apply. In that regime we tested three ways to reach the quartic floor. Higher-order stencils leave the Laplace L_1 order pinned near 2; a fixed-radius weight perturbation gives $L_1 \sim W_2^2$ exactly, for an identified algebraic reason; and a new six-atom family that is moment-stationary *exactly at every* ϵ gives slopes $W_1 \sim \epsilon^{1.001}$, $W_2 \sim \epsilon^{0.501}$, $V \sim \epsilon^{1.000}$, so $V/W_2^2 \to 0.49$ while $V/W_2^4 \to \infty$ (Section E.4). All three stay a factor $\Theta(W_2^{-2})$ above the floor - consistent with Theorem 1 and unable to reach it - which suggests these families are the wrong place to look, not that the exponent is sharp. Separately, the rate the sign bridge needs, $\|g_\Lambda - g\|_1 = O(\Lambda^{-2})$, is observed asymptotically for both kernels (Section E.6), and on $d = 1$ Laplace families the fitted W_1-versus-V slopes are 1.501-1.507 (Section E.7).

5 Limitations

The quartic latent rate is prior, and the exponent is route-independent. Bariletto et al. (2026, Theorem 8) with $W_2^2 \le 2RW_1$ already gives it; our contribution is a direct W_2 mechanism, and Theorem 2 reaches the same exponent 2β as the W_1 detour, buying coverage and constants rather than rate. Gaussian kernels stay outside it. Relatedly, the quadratic refinement is conditional: Proposition 6(c) needs smooth uniformly convex domains and two-sided density bounds with a non-explicit constant, while (d) shows the hypothesis genuinely fails for the multi-cell atomic transports Dirichlet-process posteriors produce, and narrow smoothing of a sparse atomic measure does not repair this (Section A.6).

The output bridges have explicit coverage conditions. Proposition 2 strictly enlarges the certified set and Example 1 shows the enlargement can be total, but a family matching the full covariance still defeats it: the certificate is sufficient, never necessary. Proposition 5 compares different kernels only above an activation threshold and pays $O(\sqrt{d}L_\Sigma d_\mathcal{C})$ slack, so it is polynomial separation with slack, not a linear coarse embedding. And Algorithm 1 evaluates the *learned* mixture, so a nonzero certificate says the model's conditionals separate, not the target's - task B of Section 4 is exactly that case.

Only the powers, not the latent constants, are dimension independent. Lemma 4 makes $C_d = \Theta(d^{-3/2})$ this architecture's ceiling rather than our mollifier's, but it lower-bounds *the proof*, not the truth; c_{heat}, by contrast, depends only on $\lambda_{\min}$. Separately, Assumption 1 caps $W_1(P_c, P_{c'})$ at $2R$, so Assumption 2 with $\alpha > 0$ forces $\mathrm{diam}(\mathcal{C}) \le (2R + \kappa)/\alpha$: all results are separation statements on a bounded condition region. And sharpness remains unresolved - three constructions fail to reach the quartic floor for three identified reasons (Section E.4), which is not a proof of sharpness.

6 Conclusion

We treated the coarse-embedding lower bound that analyses of conditional generative models usually assume as something to be *certified*, and gave a decision procedure returning an explicit bound or an explicit refusal. The spectral certificate is the practical core - two moment computations, constant one, no dimensional loss - and it detects anisotropic separation on which the trace-variance certificate is identically zero; when moments are stationary the heat-semigroup bridge takes over in every dimension for a fixed Laplace kernel, and scale interpolation extends it to different kernels above an explicit threshold. Feeding those bridges needs a latent inverse bound, where the accounting is that our direct W_2 proof recovers a rate already implied by Bariletto et al. (2026) plus bounded support, and its order-β generalisation yields the same exponent 2β as the W_1 detour: the value is coverage and constants, not rate. We closed the two standing hypotheses - $C^{1,1}$ reduces to strong convexity of the dual potential, and $M_\psi M'_\psi \ge \sqrt{d}/2$ fixes $C_d = \Theta(d^{-3/2})$ as this architecture's ceiling.

AI Use Statement

In this work, generative AI tools were used to (i) create or modify visual figures based on our experimental results, (ii) edit English grammar, phrasing, and vocabulary, and (iii) assist in producing Lean formalizations for checking selected mathematical equations. The tools were not used to generate

data, formulate the research questions or theoretical claims, design or run experiments, analyze or interpret experimental results, or make scientific decisions. They were also not used to replace the authors' judgment about the scientific content.

We reviewed all AI-assisted work. The experimental data and results were checked against the original runs and stored result files; AI-assisted figures were checked against those results; and the Lean formalizations were compiled and checked by the authors to ensure that they corresponded to the equations stated in the paper. AI-assisted language edits were manually reviewed for meaning and accuracy. The authors take responsibility for the final content of this work, including all text, claims, figures, and formalizations produced with AI assistance.

REFERENCES

Alexandr Andoni, Assaf Naor, and Ofer Neiman. Snowflake universality of Wasserstein spaces. *Annales scientifiques de l'École normale supérieure*, 51(3):657–700, 2018. doi: 10.24033/asens.2363.

Nicola Bariletto, Dung Le, Alessandro Rinaldo, and Nhat Ho. Convergence rates for latent mixing measures in infinite homoscedastic location-scale mixture models. *arXiv preprint arXiv:2605.06845*, 2026. doi: 10.48550/arXiv.2605.06845. URL https://arxiv.org/abs/2605.06845.

Ole E. Barndorff-Nielsen. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13, 1997. doi: 10.1111/1467-9469.00045.

Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019. doi: 10.1016/j.exmath.2018.01.002.

Christopher M. Bishop. Mixture density networks. Technical Report NCRG/94/004, Neural Computing Research Group, Aston University, 1994. URL https://publications.aston.ac.uk/id/eprint/373/.

Luis A. Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992. doi: 10.1090/S0894-0347-1992-1124980-8.

Luis A. Caffarelli. Boundary regularity of maps with convex potentials. II. *Annals of Mathematics*, 144(3):453–496, 1996. doi: 10.2307/2118564.

Guido De Philippis and Alessio Figalli. The Monge–Ampère equation and its link to optimal transportation. *Bulletin of the American Mathematical Society*, 51(4):527–580, 2014. doi: 10.1090/S0273-0979-2014-01459-4.

D. C. Dowson and B. V. Landau. The Fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis*, 12(3):450–455, 1982. doi: 10.1016/0047-259X(82)90077-X.

Matthias Gelbrich. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990. doi: 10.1002/mana.19901470121.

Nhat Ho and XuanLong Nguyen. On strong identifiability and convergence rates of parameter estimation in finite mixtures. *Electronic Journal of Statistics*, 10(1):271–307, 2016. doi: 10.1214/16-EJS1105.

Dilip B. Madan and Eugene Seneta. The variance gamma (V.G.) model for share market returns. *The Journal of Business*, 63(4):511–524, 1990. doi: 10.1086/296519.

Ingram Olkin and Friedrich Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263, 1982. doi: 10.1016/0024-3795(82)90112-4.

Long HC Pham. Decomposing out-of-distribution error in conditional flow matching via Wasserstein geometry. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026. URL https://openreview.net/forum?id=orzcDIBxa0.

Neil Pritchard and Thomas Weighill. Coarse embeddability of Wasserstein space and the space of persistence diagrams. *Discrete & Computational Geometry*, 74(2):358–373, 2025. doi: 10.1007/s00454-024-00674-6.

R. Tyrrell Rockafellar. *Convex Analysis*, volume 28 of *Princeton Mathematical Series*. Princeton University Press, 1970.

Henry Teicher. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34(4):1265–1269, 1963. doi: 10.1214/aoms/1177703862.

A Setup, Assumptions, and Proofs of the Inverse Bounds

A.1 Assumptions

Assumption 1 (Support)**.** P, P' are supported in a closed ball $B(0,R) \subset \mathbb{R}^d$ with $R > 0$, and $\Sigma \in \mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$ with $0 < \lambda_{\min} \le \lambda_{\max} < \infty$. For a condition family, $\mathrm{supp}(P_c) \subseteq B(0,R)$ for every $c \in \mathcal{C}$.

Assumption 2 (Mixing Lipschitz)**.** For a condition family $\{P_c\}_{c\in\mathcal{C}}$, the map $c \mapsto P_c$ is (α,κ)-quasi-isometric in W_1: $\alpha d_{\mathcal{C}}(c,c') - \kappa \le W_1(P_c, P_{c'}) \le L_P d_{\mathcal{C}}(c,c')$.

Assumption 3 (Scale Lipschitz)**.** $\Sigma(c) \in \mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$ for every c, and $c \mapsto \Sigma(c)$ is L_Σ-Lipschitz: $\|\Sigma(c) - \Sigma(c')\|_{\mathrm{op}} \le L_\Sigma d_{\mathcal{C}}(c,c')$.

PDE inversion, quantitatively. A kernel admits PDE inversion if a differential operator T_Σ satisfies $T_\Sigma p_G = P$ distributionally. For the Laplace kernel, $T_\Sigma = \mathrm{Id} - \frac{1}{2}\Sigma : \mathrm{Hess}$. Merely listing its coordinate coefficients obscures the relevant matrix norm, so we state directly the adjoint-action bounds used in the proofs.

Assumption 4 (Quantitative adjoint control)**.** Uniformly over $\Sigma, \Sigma' \in \mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$ and $h \in C_c^\infty(\mathbb{R}^d)$,

$$\|T_\Sigma^* h\|_\infty \le \|h\|_\infty + \frac{\lambda_{\max}}{2}\|\mathrm{Hess}(h)\|_{\mathrm{tr},\infty},$$

and

$$\|(T_\Sigma^* - T_{\Sigma'}^*)h\|_\infty \le \frac{1}{2}\|\Sigma - \Sigma'\|_{\mathrm{op}}\|\mathrm{Hess}(h)\|_{\mathrm{tr},\infty},$$

where $\|H\|_{\mathrm{tr},\infty} := \sup_x \|H(x)\|_{\mathrm{tr}}$.

For the explicit Laplace operator these inequalities are immediate from $|A : B| \le \|A\|_{\mathrm{op}}\|B\|_{\mathrm{tr}}$. They make the dimension dependence explicit: a coordinatewise coefficient sum is *not* independent of d.

Assumption 5 (Order-β adjoint control)**.** The kernel admits a differential operator $T_\Sigma = \sum_{|\nu|\le\beta} c_\nu(\Sigma)\partial^\nu$ with $T_\Sigma p_G = P$ distributionally, and there is a pair-independent constant C_{op} such that for the cutoff-localised potential $\tilde\phi = \chi\phi^*$ of Section A.2 (Step 0) of every pair P, P' supported in $B(0,R)$, and every $\Lambda \ge 1$,

$$\big\|T_\Sigma^*(\tilde\phi * \psi_\Lambda)\big\|_\infty \;\le\; C_{\mathrm{op}}\big(1 + \Lambda^{\beta-1}\big).$$

This is the single estimate the inversion argument needs, isolated so that the kernel class enters only through it. It holds whenever the coefficients $c_\nu(\Sigma)$ are bounded on $\mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$ and the mollifier has $\|\partial^\gamma\psi\|_1 < \infty$ for $|\gamma| \le \beta - 1$, using only $\|\tilde\phi\|_\infty \le B_R$ and $\mathrm{Lip}(\tilde\phi) \le L_R$ (Step 0): transferring one derivative onto $\tilde\phi$ and the remaining $k-1$ onto ψ_Λ bounds an order-k term by $L_R\Lambda^{k-1}\max_{|\gamma|=k-1}\|\partial^\gamma\psi\|_1$, and $\Lambda^{k-1} \le \Lambda^{\beta-1}$ for $\Lambda \ge 1$. For the Laplace kernel ($\beta = 2$), Step 3 of Section A.2 verifies it with the sharper $C_{\mathrm{op}} = B_R + \frac{\lambda_{\max}}{2}c_\chi d(1 + 2RM'_\psi)$, which also uses the convex structure of ϕ^* (Lemma 3).

Remark 3 (Bounded condition space). Assumption 1 forces $W_1(P_c, P_{c'}) \le 2R$ for all c, c', so Assumption 2 with $\alpha > 0$ implies $d_{\mathcal{C}}(c,c') \le (2R + \kappa)/\alpha$. The two assumptions are therefore compatible only on a condition space of finite diameter, and every coarse-embedding statement in this paper should be read as a quantitative separation guarantee on such a bounded region – not as an embedding of an unbounded condition space. Extending beyond it would require the support radius to grow with the condition, which we do not pursue.

The Brenier form and its locality. For the quadratic cost, $\phi^*(x) = \frac{1}{2}\|x\|^2 - u(x)$ with u convex. We choose the canonical c-transform representative obtained by restricting the transform to $\mathrm{supp}(P')$. Then u is a supremum of affine functions whose slopes lie in $\mathrm{supp}(P')$, so $\partial u(x) \subseteq \mathrm{conv}(\mathrm{supp}(P')) \subseteq B(0,R)$ wherever its subdifferential is nonempty. Hence $\|\nabla u\| \le R$ a.e. globally. This conclusion need not hold for an arbitrary off-support extension of an optimal potential. In contrast, $\|\nabla\phi^*(x)\| = \|x - \nabla u(x)\| \le \|x\| + R$ is bounded by $2R$ only *on* $B(0,R)$; it grows like $\|x\|$ off that ball. Section A.2 Step 0 handles this explicitly.

Proof of Lemma 1. For this cost the dual optimum equals $\inf_\pi \int c\, d\pi = \frac{1}{2}W_2^2$, so $\frac{1}{2}W_2^2 = \int \phi^* dP + \int \psi^* dP' = \int \phi^* d(P - P') + \int (\phi^* + \psi^*) dP'$. Dual feasibility gives $\phi^*(y) + \psi^*(y) \le c(y, y) = 0$ for every y, so the last integral is nonpositive; since its integrand is nonpositive it vanishes exactly when $\phi^* + \psi^* = 0$ P'-a.e. □

A.2 Proof of Theorem 1 (quartic bound)

Step 0 (cutoff localisation). ϕ^* is not globally Lipschitz, so we may not use the $2R$ bound as a global L^∞ estimate – which the pairing against $p_G - p_{G'}$ in Step 4 genuinely requires, since $p_G - p_{G'}$ has full (if exponentially decaying) support. Let $\chi \in C_c^\infty(\mathbb{R}^d)$ be radially symmetric with $0 \le \chi \le 1$, $\chi \equiv 1$ on $B(0, 2R)$, $\chi \equiv 0$ off $B(0, 3R)$, and $\|\nabla^k \chi\|_\infty \le C_k / R^k$; set $\tilde{\phi}^* := \chi \phi^*$. Since $\chi \equiv 1$ on $B(0, 2R) \supset B(0, R) = \mathrm{supp}(P - P')$,

$$\int \tilde{\phi}^*\, d(P - P') = \int \phi^*\, d(P - P') \quad \text{exactly,}$$

so the cutoff introduces no error in Lemma 1. On $B(0, 3R)$ we have $\|\nabla \phi^*\| \le 4R$. We fix the additive normalization $u(0) = 0$; this does not change $\int \phi^* d(P - P')$, since both measures have unit mass. Then $|\phi^*| \le \frac{9}{2}R^2 + 3R^2 =: B_R = 15R^2/2$. By the product rule $\nabla \tilde{\phi}^* = \chi \nabla \phi^* + \phi^* \nabla \chi$ is bounded everywhere by the pair-independent constant $L_R := R(4 + 15C_1/2)$, and $\tilde{\phi}^*$ is compactly supported: it is genuinely globally Lipschitz and bounded. We use $\tilde{\phi}^*$ throughout below; it agrees with ϕ^* on $\mathrm{supp}(P - P')$, which is exactly what the mixing-measure pairing requires.

Step 1 (first-order mollification). Fix a nonnegative, even probability mollifier $\psi \in C_c^\infty(\mathbb{R}^d)$ and set $\psi_\Lambda(x) = \Lambda^d \psi(\Lambda x)$. Then $\tilde{\phi}_\Lambda^* = \tilde{\phi}^* * \psi_\Lambda \in C_c^\infty$, as required by the distributional inversion identity. Since $\tilde{\phi}^*$ is only Lipschitz, $\|\tilde{\phi}^* - \tilde{\phi}_\Lambda^*\|_\infty \le L_R M_\psi / \Lambda$ with $M_\psi = \int \|y\| \psi(y) dy$. Substituting into Lemma 1 and bounding the total-variation factor by 2,

$$\tfrac{1}{2}W_2^2(P, P') \le \int \tilde{\phi}_\Lambda^*\, d(P - P') + \frac{C_0}{\Lambda}, \qquad C_0 := 2L_R M_\psi,$$

the factor $\frac{1}{2}$ coming from Lemma 1 under the cost convention $c = \frac{1}{2}\|x - y\|^2$. It propagates through the constants below and changes none of the exponents. This is the only mollification bound available without further regularity; uniform Hessian control requires Theorem 3's hypothesis.

Step 2 (PDE inversion). Because both mixtures use the same scale Σ, $T_\Sigma p_G = P$ and $T_\Sigma p_{G'} = P'$ distributionally. Since $\tilde{\phi}_\Lambda^* \in C_c^\infty$, $\int \tilde{\phi}_\Lambda^* d(P - P') = \langle T_\Sigma^* \tilde{\phi}_\Lambda^*,\, p_G - p_{G'} \rangle$.

Step 3 (adjoint bound). Transferring one derivative to $\tilde{\phi}^*$ and one to ψ_Λ for $|\nu| = 2$ – now a genuine global sup norm, by Step 0 – gives $\|T_\Sigma^* \tilde{\phi}_\Lambda^*\|_\infty \le C_{\mathrm{op}}(1 + \Lambda)$ with, for example, the pair-independent choice

$$C_{\mathrm{op}} := B_R + \frac{\lambda_{\max}}{2} c_\chi d(1 + 2RM'_\psi).$$

Here the Hessian part of C_{op} is $O(d[1 + RM'_\psi(d)])$ by Proposition 7, which is proved for the *cutoff-localised* potential $\tilde{\phi}^*$ used in Step 0 rather than for ϕ^* itself: multiplying by χ destroys the convexity structure the trace bound relies on, and the resulting $\nabla\chi \otimes \nabla\phi^*$ and $\phi^* D^2 \chi$ terms must be carried explicitly. Section B.3 does so.

Step 4 (assemble). By Hölder, $\frac{1}{2}W_2^2(P, P') \le C_{\mathrm{op}}(1 + \Lambda)V + C_0/\Lambda$.

Step 5 (correct optimisation). For $\Lambda \ge 1$, $(1 + \Lambda) \le 2\Lambda$, so with $A := 2C_{\mathrm{op}}V$, $B := C_0$, $\frac{1}{2}W_2^2 \le A\Lambda + B/\Lambda$. If $V = 0$, letting $\Lambda \to \infty$ gives $W_2 = 0$, so assume $V > 0$. The right-hand side is minimised at $\Lambda^* = \sqrt{B/A}$ with value $2\sqrt{AB}$. *The balancing point matters:* setting $\Lambda = 1/W_2$ and discarding B/Λ would be invalid because $B/\Lambda = BW_2$ dominates W_2^2 as $W_2 \to 0$; the balancing point is $\Lambda^* \sim V^{-1/2}$, not W_2^{-1}. For $V > 0$ small enough that $\Lambda^* \ge 1$, $\frac{1}{2}W_2^2 \le 2\sqrt{2C_{\mathrm{op}}C_0}\sqrt{V}$.

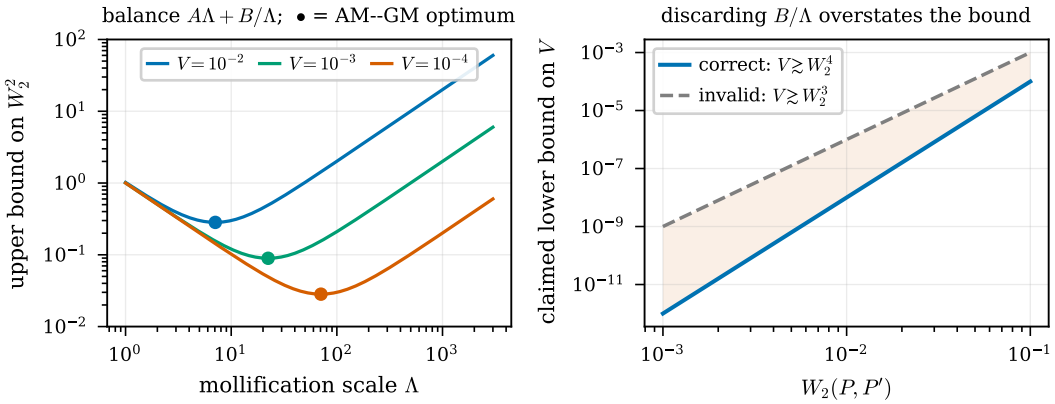

Figure 2: Left: the two error terms in Equation (3) trade off, and the AM–GM optimum (•) sits at $\Lambda^* \sim V^{-1/2}$, not at $1/W_2$. Right: substituting $\Lambda = 1/W_2$ and dropping the remainder claims the shaded region, which the argument does not support.

Step 6 (rearrange, both regimes). Squaring, $\frac{1}{4}W_2^4 \le 8C_{\mathrm{op}}C_0V$ (using Lemma 1's factor $\frac{1}{2}$), i.e.

$$V \;\ge\; \frac{W_2^4}{32\,C_{\mathrm{op}}C_0} \qquad \text{whenever } \Lambda^* \ge 1, \;\; \text{i.e. } V \le V_0 := \frac{C_0}{2C_{\mathrm{op}}}.$$

The complementary regime is immediate rather than absent: if $V > V_0$ then, since Assumption 1 gives $W_2(P, P') \le 2R$,

$$V \;>\; V_0 \;=\; V_0\frac{(2R)^4}{(2R)^4} \;\ge\; \frac{V_0}{(2R)^4}\,W_2^4.$$

Taking $C_d = \min\{(32C_{\mathrm{op}}C_0)^{-1},\, V_0(2R)^{-4}\}$ covers all V and proves Theorem 1. □

The same two-regime split applies to Theorem 4; its small-discrepancy optimisation is valid only where the interior optimum satisfies $\Lambda^* \ge 1$.

Detailed constant scaling for Theorem 1. The asymptotic order of the quartic constant can be read from the mollifier-only quantities in Steps 1 and 3. Every product mollifier has $M'_\psi = \int |\rho'|$, independent of d, so $M'_\psi = O(1)$ *is* attainable and Proposition 7's Hessian factor is genuinely $O(d)$. Lemma 4 gives the complementary obstruction: any such choice forces $M_\psi = \Omega(\sqrt{d})$, hence $C_{\mathrm{op}}C_0 = \Omega(d^{3/2})$. Therefore $C_d = O(d^{-3/2})$ for *every* mollifier family, and this order is attained, i.e. $C_d = \Theta(d^{-3/2})$, by the Gaussian and by compactly supported product families. So the only way to improve this scaling within the architecture is to bypass the first-order mollification error, not to tune the mollifier constants.

The cutoff contributes only through dimension-free profile constants: $c_\chi \le 1 + 8C_1 + \frac{15}{2}C_2$ and $L_R = R(4 + 15C_1/2)$. For the Hermite cubic ramp on $[2R, 3R]$, one has $C_1 = 3/2$, $C_2 = 6$, hence $c_\chi \le 58$ and $L_R = 15.25R$ (Section B.4). Because $\chi \equiv 1$ on $\mathrm{supp}(P - P')$, the cutoff adds no pairing error; it affects only constants.

A.3 An independent cross-check of the quartic exponent

Bariletto et al. (2026, Theorem 8) prove the stronger pairwise inequality $V \ge \bar{C}_d[W_1(P, P')^2 + \|\Sigma - \Sigma'\|_{\mathrm{op}}]$ for the Laplace kernel (Equation (2)). On bounded support, for the W_1-optimal coupling π we have $\|x - y\| \le 2R$ pointwise, hence $\|x - y\|^2 \le 2R\|x - y\|$, and integrating against π gives $W_2(P, P')^2 \le \int \|x - y\|^2 d\pi \le 2R \int \|x - y\| d\pi = 2R\, W_1(P, P')$. Therefore $W_1 \ge W_2^2/(2R)$ and

$$V \;\ge\; \bar{C}_d\Big(\frac{W_2^2}{2R}\Big)^2 \;=\; \frac{\bar{C}_d}{4R^2}\,W_2^4.$$

This reaches exponent 4 by a route that shares no step with Section A.2. It also shows that the quartic inequality itself is prior; the value of Section A.2 is its direct mechanism and constant bookkeeping.

A.4 Proof of Theorem 2 (order-β inversion)

Steps 0–2 of Section A.2 are unchanged: cutoff localisation is exact on $\mathrm{supp}(P - P')$, first-order mollification costs C_0/Λ with $C_0 = 2L_R M_\psi$, and PDE inversion converts the mixing-measure pairing into $\langle T_\Sigma^* \tilde{\phi}_\Lambda^*, p_G - p_{G'} \rangle$. Step 3 is replaced by Assumption 5, and Hölder gives, for $\Lambda \geq 1$,

$$\tfrac{1}{2} W_2^2(P, P') \;\leq\; C_{\mathrm{op}}(1 + \Lambda^{\beta-1})V + \frac{C_0}{\Lambda} \;\leq\; A\Lambda^{\beta-1} + \frac{B}{\Lambda}, \qquad A := 2C_{\mathrm{op}}V, \quad B := C_0,$$

using $1 + \Lambda^{\beta-1} \leq 2\Lambda^{\beta-1}$ for $\Lambda \geq 1$, $\beta \geq 1$.

$\beta = 1$. The first term does not depend on Λ, so letting $\Lambda \to \infty$ gives $\frac{1}{2} W_2^2 \leq 2C_{\mathrm{op}}V$, i.e. $V \geq W_2^2/(4C_{\mathrm{op}})$, with no regime restriction.

$\beta > 1$. At the interior optimum $A(\beta-1)\Lambda^{\beta-2} = B/\Lambda^2$, so $\Lambda^* = (B/(A(\beta-1)))^{1/\beta}$ and

$$A(\Lambda^*)^{\beta-1} + \frac{B}{\Lambda^*} = A^{1/\beta} B^{(\beta-1)/\beta} \Big[(\beta-1)^{-(\beta-1)/\beta} + (\beta-1)^{1/\beta} \Big] = c_\beta A^{1/\beta} B^{(\beta-1)/\beta}.$$

Hence $W_2^2 \leq 2c_\beta (2C_{\mathrm{op}}V)^{1/\beta} C_0^{(\beta-1)/\beta}$ and, rearranging,

$$V \;\geq\; \frac{W_2^{2\beta}}{2^\beta c_\beta^\beta \, 2C_{\mathrm{op}} \, C_0^{\beta-1}} \qquad \text{whenever } \Lambda^* \geq 1, \text{ i.e. } V \leq V_0(\beta) = \frac{C_0}{2C_{\mathrm{op}}(\beta-1)}.$$

If $V > V_0(\beta)$ then $W_2 \leq 2R$ gives $V > V_0(\beta)(2R)^{-2\beta} W_2^{2\beta}$, and taking the minimum of the two constants proves Theorem 2. At $\beta = 2$, $c_2 = 2$ and the constant is $(2^2 \cdot 2^2 \cdot 2C_{\mathrm{op}}C_0)^{-1} = (32C_{\mathrm{op}}C_0)^{-1}$, which is exactly Theorem 1; the threshold likewise reduces to $V_0 = C_0/(2C_{\mathrm{op}})$. □

Explicit admissible constants (including optimisation and case split). The optimisation step for $\beta > 1$ is the interior minimiser $\Lambda^* = (B/(A(\beta-1)))^{1/\beta}$ of $A\Lambda^{\beta-1} + B/\Lambda$, with $A := 2C_{\mathrm{op}}V$ and $B := C_0$. This gives the small-discrepancy branch

$$V \;\geq\; \frac{W_2(P, P')^{2\beta}}{2^\beta c_\beta^\beta \, 2C_{\mathrm{op}} C_0^{\beta-1}} \quad \text{when } \Lambda^* \geq 1 \;\Leftrightarrow\; V \leq V_0(\beta) := \frac{C_0}{2C_{\mathrm{op}}(\beta-1)},$$

where $c_\beta = (\beta-1)^{-(\beta-1)/\beta} + (\beta-1)^{1/\beta}$. The complementary branch uses $W_2 \leq 2R$:

$$V \;>\; V_0(\beta)(2R)^{-2\beta} W_2(P, P')^{2\beta}.$$

Hence one admissible global constant is

$$C_d(\beta) = \Big(2^\beta c_\beta^\beta \, 2C_{\mathrm{op}} C_0^{\beta-1} \Big)^{-1} \wedge \frac{V_0(\beta)}{(2R)^{2\beta}} \quad (\beta > 1).$$

For $\beta = 1$, letting $\Lambda \to \infty$ yields $V \geq W_2(P, P')^2/(4C_{\mathrm{op}})$ with no regime split.

The W_1 route gives the same exponent. Run the identical steps against the Kantorovich–Rubinstein dual, whose witness h may be taken 1-Lipschitz and, after the same Step-0 cutoff, bounded and compactly supported. Then $\|\tilde{h} - \tilde{h}_\Lambda\|_\infty \leq L'_R M_\psi/\Lambda$ and, since $\|P - P'\|_{\mathrm{TV}} \leq 2$, $W_1(P, P') \leq C'_{\mathrm{op}}(1 + \Lambda^{\beta-1})V + 2L'_R M_\psi/\Lambda$, so the same optimisation yields $W_1 \lesssim V^{1/\beta}$, i.e. $V \gtrsim W_1^\beta$ – the shape of Equation (2) at $\beta = 2$. Composing with $W_1 \geq W_2^2/(2R)$ from Section A.3 gives $V \gtrsim (2R)^{-\beta} W_2^{2\beta}$. Both routes therefore reach exponent 2β; they differ in constants – the direct route avoids the $(2R)^{-\beta}$ factor but pays Lemma 1's $\frac{1}{2}$ – and in which dual is used, not in rate. In particular, acting directly on W_2 does not improve the exponent to $2\beta/(\beta-1)$; that rate would follow only from assigning V to the mollification term rather than to the operator term, which is incorrect.

Instantiation beyond Laplace. Normal-variance mixtures with $\mathrm{Gamma}(\alpha, 1)$ mixing have characteristic function $(1 + t^\top \Sigma t/2)^{-\alpha}$ (Madan & Seneta, 1990; Barndorff-Nielsen, 1997) and therefore invert with $T_\Sigma = (\mathrm{Id} - \frac{1}{2}\Sigma : \mathrm{Hess})^\alpha$: $\beta = 2\alpha$ and the exponent is 4α, with $\alpha = 1$ the Laplace case. The Gaussian kernel has characteristic function $e^{-t^\top \Sigma t/2}$, whose reciprocal is not polynomial, so no finite-order T_Σ exists and Theorem 2 does not apply at any β; for Gaussian kernels only the output bridges of Section 3.3 are available, and there $K_2 < \infty$ in every dimension.

A.5 Proof of Theorem 3 (conditional quadratic bound)

Apply the cutoff of Section A.2 before mollifying. On $B(0,3R)$, $\|\nabla\phi^*\| \le 4R$ and $|\phi^*| \le 15R^2/2$. Hence the product rule gives

$$D^2(\chi\phi^*) = \chi D^2\phi^* + \nabla\chi \otimes \nabla\phi^* + \nabla\phi^* \otimes \nabla\chi + \phi^* D^2\chi$$

a.e., and therefore

$$\|D^2\tilde{\phi}^*\|_{\mathrm{op},\infty} \le \widetilde{M} := M + 8C_1 + \tfrac{15}{2}C_2.$$

Thus $\tilde{\phi}^* = \chi\phi^*$ is compactly supported and globally $C^{1,1}$. For an even mollifier,

$$\|\tilde{\phi}^*_\Lambda - \tilde{\phi}^*\|_\infty \le \frac{\widetilde{M} m_2(\psi)}{2\Lambda^2}, \qquad \|D^2\tilde{\phi}^*_\Lambda\|_{\mathrm{tr},\infty} \le d\widetilde{M}.$$

The second inequality is uniform in Λ, which is stronger than the first-derivative transfer used in Theorem 1. PDE inversion and Assumption 4 therefore yield

$$\tfrac{1}{2}W_2(P,P')^2 \le \left(\|\tilde{\phi}^*\|_\infty + \frac{\lambda_{\max}}{2}d\widetilde{M}\right)V + \frac{\widetilde{M} m_2(\psi)}{\Lambda^2}.$$

Letting $\Lambda \to \infty$ proves

$$V \ge \frac{W_2(P,P')^2}{15R^2 + \lambda_{\max} d\widetilde{M}}.$$

This proves Theorem 3. □

A.6 Proof of Proposition 6 (when the $C^{1,1}$ hypothesis holds)

We deliberately do not assert that densities bounded above and below on a convex domain suffice: those hypotheses give strict convexity of the potential and interior $C^{1,\alpha}$ regularity of the transport map, *not* a bounded Hessian, and upgrading interior $C^{2,\alpha}$ estimates to global, up-to-the-boundary Hessian control requires substantially stronger domain hypotheses (Caffarelli, 1992; 1996; De Philippis & Figalli, 2014). Proposition 6 instead routes the hypothesis through a single convexity statement about the dual potential.

The reduction must be stated for u, not for ϕ^*: $\phi^* = \frac{1}{2}\|\cdot\|^2 - u$ is a difference of convex functions and is in general neither convex nor concave, so $(\phi^*)^*$ is not the object convex duality applies to. The convex Brenier potential u is, and $D^2\phi^* = \mathrm{Id} - D^2u$ transfers any bound on D^2u to the hypothesis of Theorem 3.

Proposition 6 (When the $C^{1,1}$ hypothesis holds)**.** *Write $\phi^* = \frac{1}{2}\|\cdot\|^2 - u$ with u the convex Brenier potential (∇u pushes P to P'), and let u^* be its Legendre conjugate – the convex potential of the reverse map, $\nabla u^*{}_\# P' = P$.*

- *(a) If u^* is m-strongly convex on its domain, then ∇u is $(1/m)$-Lipschitz on all of $\mathbb{R}^d$, so $0 \preceq D^2u \preceq m^{-1}\mathrm{Id}$ and Theorem 3 applies with $M = \max\{1,\, m^{-1} - 1\}$.*
- *(b)* (Single-Dirac target.) *If $P' = \delta_\theta$ then $u(x) = \theta\cdot x + \mathrm{const}$, so $D^2u = 0$, $\phi^*(x) = \frac{1}{2}\|x-\theta\|^2 + \mathrm{const}$ and $M = 1$.*
- *(c)* (Caffarelli regime.) *If P, P' have $C^{0,\alpha}$ densities bounded above and below on bounded, smooth, uniformly convex $\Omega, \Omega' \subseteq B(0,R)$, then $u^* \in C^{2,\alpha}(\overline{\Omega'})$ and is uniformly convex, so (a) holds with a finite, non-explicit $m = m(\alpha, \text{density bounds}, \Omega, \Omega')$, compatible with Assumption 1.*
- *(d)* (Failure.) *If P' is atomic with at least two active Laguerre cells, u^* is piecewise affine on $\mathrm{conv}(\mathrm{supp}P')$ and hence not strongly convex; ∇u jumps across cell interfaces and $M = \infty$, so Theorem 1 is the available statement.*

Proof of Proposition 6. (a) u is convex, proper and lower semicontinuous, so $u = (u^*)^*$ and $p \in \partial u(x) \iff x \in \partial u^*(p)$ (Rockafellar, 1970). Let $p = \nabla u(x)$, $q = \nabla u(y)$. Then $x \in \partial u^*(p)$, $y \in \partial u^*(q)$, and m-strong convexity of u^* gives $\langle x-y,\, p-q\rangle \ge m\|p-q\|^2$; Cauchy–Schwarz yields

$\|x-y\|\|p-q\| \geq m\|p-q\|^2$, i.e. ∇u is $(1/m)$-Lipschitz on $\mathbb{R}^d$. (The conclusion is global even when $\operatorname{dom} u^*$ is bounded: u^* is m-strongly convex as an extended-real function, its conjugate is then finite and $C^{1,1}$ everywhere.) Convexity gives $D^2u \succeq 0$, so $0 \preceq D^2u \preceq m^{-1}\mathrm{Id}$ and the eigenvalues of $D^2\phi^* = \mathrm{Id} - D^2u$ lie in $[1-m^{-1}, 1]$, whence $\|D^2\phi^*\|_{\mathrm{op}} \leq \max\{1, m^{-1}-1\} =: M$.

(b) The canonical c-transform is taken over $\operatorname{supp}(P') = \{\theta\}$, so $u(x) = \theta \cdot x - v_\theta$ identically, $D^2u = 0$, $\phi^*(x) = \frac{1}{2}\|x-\theta\|^2 + \text{const}$ and $D^2\phi^* = \mathrm{Id}$; note that C'_d is then free of M.

(c) Apply Caffarelli's boundary regularity theory to the reverse map $\nabla u^*{}_{\#}P' = P$, whose hypotheses are symmetric in P, P': it gives $u^* \in C^{2,\alpha}(\overline{\Omega'})$. The Monge–Ampère equation $\det D^2u^* = p'/(p \circ \nabla u^*)$ has right-hand side bounded below by some $c_0 > 0$, while $\|D^2u^*\|_{\mathrm{op}} \leq \Lambda_0 < \infty$ by the $C^{2,\alpha}$ bound; hence every eigenvalue of D^2u^* is at least $c_0\Lambda_0^{-(d-1)} =: m > 0$, so u^* is m-strongly convex on $\overline{\Omega'}$ and (a) applies. The constant is finite but not explicit, since Caffarelli's estimates are not constructive.

(d) If $P' = \sum_j w_j\delta_{y_j}$ has at least two atoms receiving mass, the convex potential is $u(x) = \max_j[x \cdot y_j - v_j]$, so $\nabla u(x) = y_{j(x)}$ jumps by $y_j - y_{j'} \neq 0$ across the interface between two active Laguerre cells and $\nabla\phi^* = x - \nabla u$ jumps with it. A function with a gradient jump is not $C^{1,1}$; equivalently u^* is the lower convex envelope of the points (y_j, v_j), hence piecewise affine on $\operatorname{conv}\{y_j\}$ and not strongly convex. □

Smoothing does not automatically enter regime (c). It is tempting to say that replacing an atomic P by $P * K_h$ puts one in regime (c). The resulting density is $C^{0,\alpha}$ when K_h is, and stays supported in $B(0, R+h)$, but it is bounded *below* on a convex Ω only if the atoms are $O(h)$-dense in Ω: for well-separated atoms the smoothed density is exponentially small in between, and the lower bound in (c) degenerates. Regime (c) therefore covers deliberately regularised posteriors – smoothed at a bandwidth comparable to the spacing of the posterior mass – not narrow smoothing of a sparse atomic measure.

Independently of which sufficient condition one adopts, none follows merely from atomicity: regime (b) is atomic and satisfies the hypothesis, while regime (d) is the real obstruction – a singular distributional Hessian in multi-cell configurations. Theorem 3 is therefore a conditional statement not guaranteed in the Dirichlet-process regime this framework targets; Theorem 1 is what downstream work should use there. Note also that C'_d depends on M, which we track explicitly rather than absorb.

B Condition-Dependent Shared Scales and Operator Bookkeeping

B.1 A perturbative comparison

Let $G_c = P_c \times \delta_{\Sigma(c)}$, where every $\Sigma(c)$ belongs to $\mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$, and

$$T_{\Sigma(c)} = \mathrm{Id} - \tfrac{1}{2}\Sigma(c) : \mathrm{Hess}.$$

Although $T_{\Sigma(c)}p_{G_c} = P_c$ distributionally for each c, the two operators in a comparison need not agree. The following result records the cross term produced by this mismatch. It is a self-contained perturbative calculation, not an extension of the stronger pairwise Laplace theorem of Bariletto et al. (2026).

Theorem 4 (Perturbative condition-dependent shared-scale bound)**.** *Assume $P_c, P_{c'}$ are supported in $B(0,R)$, that $\Sigma(c), \Sigma(c') \in \mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$, and that Assumption 4 holds. Then there are finite constants $C_{4,d} > 0$ and $L_d \geq 0$, specified in the proof, such that*

$$\|p_{G_c} - p_{G_{c'}}\|_1 \;\geq\; C_{4,d}W_2(P_c, P_{c'})^4 - L_d\|\Sigma(c) - \Sigma(c')\|_{\mathrm{op}}.$$

Under Assumption 3, the final term is at least $-L_dL_\Sigma d_{\mathcal{C}}(c,c')$. Each G_c still has one scale shared across all of its components; this is not a component-wise heteroscedastic-mixture theorem.

For the Laplace class, Bariletto et al. (2026, Theorem 8) already allow $\Sigma(c) \neq \Sigma(c')$ and give the stronger positive term $+\bar{C}_d\|\Sigma(c) - \Sigma(c')\|_{\mathrm{op}}$. Thus the negative term above is an artefact of treating the operator mismatch perturbatively, not an intrinsic loss of the model.

Proof. Let ϕ^* be the one-sided potential used in Section A.2, let $\tilde{\phi}^* = \chi\phi^*$, and put $h_\Lambda = \tilde{\phi}^* * \psi_\Lambda$. The cutoff is essential, since the density pairing is over all of $\mathbb{R}^d$. With $V = \|p_{G_c} - p_{G_{c'}}\|_1$ and $\delta = \|\Sigma(c) - \Sigma(c')\|_{\mathrm{op}}$, PDE inversion gives

$$\int h_\Lambda \, d(P_c - P_{c'}) = \underbrace{\langle T^*_{\Sigma(c)} h_\Lambda, p_{G_c} - p_{G_{c'}} \rangle}_{I_1} + \underbrace{\langle (T^*_{\Sigma(c)} - T^*_{\Sigma(c')}) h_\Lambda, p_{G_{c'}} \rangle}_{I_2}.$$

By Lemma 3, for $\Lambda \geq 1$,

$$\|\mathrm{Hess}(h_\Lambda)\|_{\mathrm{tr},\infty} \leq c_\chi d(1 + 2RM'_\psi\Lambda) \leq c_\chi d(1 + 2RM'_\psi)\Lambda.$$

Using the pair-independent bound $B_R = 15R^2/2$ from Section A.2 and defining

$$a_d := B_R + \frac{\lambda_{\max}}{2} c_\chi d(1 + 2RM'_\psi), \qquad b_d := \frac{1}{2} c_\chi d(1 + 2RM'_\psi),$$

the two adjoint-action inequalities in Assumption 4 imply

$$|I_1| \leq a_d \Lambda V, \qquad |I_2| \leq b_d \Lambda \delta.$$

The one-sided Kantorovich inequality and the mollification estimate of Section A.2 therefore yield

$$\frac{1}{2} W_2(P_c, P_{c'})^2 \leq \Lambda\big(a_d V + b_d \delta\big) + \frac{C_0}{\Lambda}. \tag{6}$$

Set $U = a_d V + b_d \delta$. If $U \leq C_0$, the minimiser $\Lambda = \sqrt{C_0/U}$ is admissible when $U > 0$ and Equation (6) gives $U \geq W_2^4/(16C_0)$. If $U > C_0$, then $W_2 \leq 2R$ gives $U > C_0 \geq C_0 W_2^4/(2R)^4$. Hence the theorem holds with

$$C_{4,d} := \min\left\{ \frac{1}{16 a_d C_0}, \frac{C_0}{a_d (2R)^4} \right\}, \qquad L_d := \frac{b_d}{a_d}.$$

If $U = 0$, letting $\Lambda \to \infty$ in Equation (6) gives $W_2 = 0$, so the same conclusion is trivial. Since $a_d = C_{\mathrm{op}}$, at $\delta = 0$ this gives $C_{4,d} = 2C_d$: bounding each term by its $\Lambda \geq 1$ value, rather than using $1 + \Lambda \leq 2\Lambda$ as in Step 5 of Section A.2, gains a factor 2 and changes no exponent. No assumption connecting P_c to the condition metric is needed for this pairwise statement. □

Proposition 7 (Refined Hessian contribution)**.** *For the cutoff-localised potential $\tilde{\phi}^* = \chi\phi^*$ of Section A.2, $\|\mathrm{Hess}(\tilde{\phi}^*_\Lambda)(x)\|_{\mathrm{tr}} \leq c_\chi \, d \, (1 + 2RM'_\psi\Lambda)$ for all x, with c_χ depending only on the fixed cutoff profile and $M'_\psi := \max_i \|\partial_i \psi\|_1$. Consequently the Hessian contribution is $O(d[1 + RM'_\psi(d)])$ rather than the entrywise $O(d^2)$ count, and it is $O(d)$ whenever $M'_\psi(d) = O(1)$.*

B.2 When the perturbative lower bound is active

Combining Theorem 4 with Assumption 2 gives the scalar lower envelope

$$C_{4,d}(\alpha r - \kappa)_+^4 - L_d L_\Sigma r, \qquad r = d_{\mathcal{C}}(c, c').$$

Because Assumptions 1 and 2 force $D_{\mathcal{C}} := \mathrm{diam}(\mathcal{C}) \leq (2R + \kappa)/\alpha$, activation must be defined on the accessible interval, not by sending r to infinity.

Definition 1 (Scale-induced activation threshold)**.** Let

$$\mathcal{A} := \left\{ \rho \in [0, D_{\mathcal{C}}] : C_{4,d}(\alpha r - \kappa)_+^4 \geq L_d L_\Sigma r \text{ for every } r \in [\rho, D_{\mathcal{C}}] \right\}.$$

Set $r_\Sigma = \inf \mathcal{A}$ when $\mathcal{A} \neq \varnothing$, and $r_\Sigma = +\infty$ otherwise. Thus $r_\Sigma = +\infty$ records a vacuous perturbative bound on the available condition region.

When the positive root lies inside the accessible interval, its small L_Σ behaviour is

$$r_\Sigma = \left(\frac{L_d L_\Sigma}{C_{4,d} \alpha^4} \right)^{1/3} \qquad (\kappa = 0),$$

whereas for $\kappa > 0$ it is

$$r_\Sigma = \frac{\kappa}{\alpha} + \left(\frac{L_d L_\Sigma \kappa}{C_{4,d}\alpha^5}\right)^{1/4} + o(L_\Sigma^{1/4}).$$

In particular, the cube-root law is not valid when the original additive slack κ is positive. For a target separation $\rho_0 \in (\kappa/\alpha, D_{\mathcal{C}}]$, activation by ρ_0 requires

$$L_\Sigma \le \frac{C_{4,d}}{L_d}\,\frac{(\alpha\rho_0 - \kappa)^4}{\rho_0},$$

with the condition interpreted as automatic when $L_d = 0$.

B.3 Proof of Proposition 7

Lemma 2 (Trace norm before localisation)**.** *Let $\phi^* = \frac{1}{2}\|\cdot\|^2 - u$, where u is convex and $\|\nabla u\| \le R$ a.e., and let ψ_Λ be an even probability mollifier with $M'_\psi := \max_i \|\partial_i \psi\|_1$. Then*

$$\|\mathrm{Hess}(\phi^* * \psi_\Lambda)(x)\|_{\mathrm{tr}} \le d(1 + 2RM'_\psi\Lambda).$$

Proof. Convolution preserves the Hessian of the quadratic term. Moreover $u * \psi_\Lambda$ is convex, so its Hessian is positive semidefinite. Distributionally, and then pointwise after convolution,

$$[\mathrm{Hess}(u * \psi_\Lambda)]_{ij}(x) = \int (\nabla u)_j(x - z)\partial_i \psi_\Lambda(z)\, dz,$$

for any bounded a.e. representative of ∇u. Since $\|\nabla u\| \le R$ and $\|\partial_i \psi_\Lambda\|_1 = \Lambda \|\partial_i \psi\|_1$, the trace of this positive semidefinite Hessian is at most $RM'_\psi d\Lambda$, and hence also at most the looser $2RM'_\psi d\Lambda$ used in the displayed statement. If its eigenvalues are $\nu_i \ge 0$, the eigenvalues of $\mathrm{Hess}(\phi^* * \psi_\Lambda)$ are $1 - \nu_i$, and $\sum_i |1 - \nu_i| \le d + \sum_i \nu_i$, proving the claim. □

Lemma 3 (Trace norm after localisation)**.** *With the cutoff χ from Section A.2,*

$$\big\|\mathrm{Hess}((\chi\phi^*) * \psi_\Lambda)(x)\big\|_{\mathrm{tr}} \le c_\chi d(1 + 2RM'_\psi\Lambda),$$

where c_χ depends only on the fixed cutoff profile, not on d, Λ, P, or P'.

Proof. Write $\chi\phi^* = \chi\|x\|^2/2 - \chi u$ and use the distributional product rule. For the quadratic part,

$$D^2(\chi\|x\|^2/2) = \chi\mathrm{Id} + \nabla\chi \otimes x + x \otimes \nabla\chi + \tfrac{1}{2}\|x\|^2 D^2\chi.$$

Its trace norm is bounded by a constant multiple of d, using $\|x\| \le 3R$, $\|\nabla\chi\|_\infty \le C_1/R$, and $\|D^2\chi\|_\infty \le C_2/R^2$ on the cutoff support.

For the convex part,

$$D^2(\chi u) = \chi D^2 u + \nabla\chi \otimes \nabla u + \nabla u \otimes \nabla\chi + u D^2\chi.$$

The matrix measure $\chi D^2 u$ is positive semidefinite and dominated by $D^2 u$; after convolution its trace is bounded as in Lemma 2 by $2RM'_\psi d\Lambda$. The remaining terms have bounded trace norm because $\|\nabla u\| \le R$ and, after fixing $u(0) = 0$, $|u| \le 3R^2$ on $B(0, 3R)$. Convolution with the nonnegative probability mollifier preserves these bounds. Absorbing the cutoff constants into c_χ proves the claim. □

By operator/trace-norm duality,

$$|\Sigma : \mathrm{Hess}(h_\Lambda)| \le \|\Sigma\|_{\mathrm{op}} \|\mathrm{Hess}(h_\Lambda)\|_{\mathrm{tr}},$$

so Lemma 3 proves Proposition 7. A coordinatewise sum can contain $O(d^2)$ terms; the trace argument bounds this particular Laplace-operator contribution by $O(d[1 + RM'_\psi(d)])$. It is $O(d)$ only for a mollifier family with $M'_\psi(d) = O(1)$. This does not establish the dimension dependence of the complete inverse constant, which also contains $M_\psi(d)$ and the cutoff constants.

B.4 THE MOLLIFIER AND THE CUTOFF: WHAT THEY CAN AND CANNOT DO

Proposition 7 makes the Hessian contribution $O(d)$ as soon as $M'_\psi(d) = O(1)$. That is attainable: for any product mollifier $\psi(x) = \prod_i \rho(x_i)$ one has $\|\partial_i\psi\|_1 = \int|\rho'|$, independent of d, and a smooth bump profile on $[-1,1]$ keeps $\psi \in C_c^\infty$ as Step 1 of Section A.2 requires. The Gaussian gives $M'_\psi = E|X_1| = \sqrt{2/\pi}$ likewise, at the cost of leaving C_c^∞ for the Schwartz class, which is harmless here because p_G and P are finite measures with finite second moments so the inversion identity extends by density. The remaining question – whether both mollifier constants can be $O(1)$ at once – has a negative answer.

Lemma 4 (L^1 uncertainty principle for mollifiers)**.** *Let ψ be a probability density on $\mathbb{R}^d$ with $M_\psi = \int\|y\|\psi < \infty$ whose distributional partials are finite measures, and let $M'_\psi = \max_i\|\partial_i\psi\|_{\mathrm{TV}}$. Then $M_\psi \cdot M'_\psi \geq \sqrt{d}/2$.*

Proof of Lemma 4. Fix a coordinate i and let f be the i-th marginal density of ψ. Integrating out the other coordinates, f' is a finite measure with $\|f'\|_{\mathrm{TV}} \leq \|\partial_i\psi\|_{\mathrm{TV}}$, so f has a BV representative vanishing at $\pm\infty$; its total variation is at least the rise and the fall, giving $\|\partial_i\psi\|_{\mathrm{TV}} \geq 2\|f\|_\infty$. Write $M_i := \|f\|_\infty$. By the bathtub principle, among densities with $0 \leq g \leq M_i$ the first absolute moment is minimised by $g = M_i\mathbf{1}_{[-a,a]}$ with $a = 1/(2M_i)$, so

$$E|X_i| = \int|x|f(x)\,dx \;\geq\; M_i\int_{-a}^{a}|x|\,dx = M_i a^2 = \frac{1}{4M_i}.$$

Multiplying the two displays, $E|X_i|\cdot\|\partial_i\psi\|_{\mathrm{TV}} \geq \frac{1}{2}$, hence $E|X_i| \geq 1/(2M'_\psi)$ for every i. Finally $\|y\| \geq \|y\|_1/\sqrt{d}$ pointwise gives

$$M_\psi = E\|X\| \;\geq\; \frac{1}{\sqrt{d}}\sum_{i=1}^{d}E|X_i| \;\geq\; \frac{d}{2\sqrt{d}\,M'_\psi} = \frac{\sqrt{d}}{2M'_\psi}. \qquad \square$$

The bound is scale invariant, as it must be: dilating ψ multiplies M_ψ by λ and M'_ψ by $1/\lambda$. Its consequence for Theorem 1 is immediate. Since $C_0 = 2L_R M_\psi$ and $C_{\mathrm{op}} \geq \frac{\lambda_{\max}}{2}c_\chi d(1+2RM'_\psi)$,

$$C_{\mathrm{op}}C_0 \;\geq\; \lambda_{\max}c_\chi L_R\,d\,M_\psi(1+2RM'_\psi) \;\geq\; 2R\,\lambda_{\max}c_\chi L_R\,d\,M_\psi M'_\psi \;\geq\; R\,\lambda_{\max}c_\chi L_R\,d^{3/2},$$

so $C_d = (32C_{\mathrm{op}}C_0)^{-1} = O(d^{-3/2})$ for *every* admissible mollifier, while the Gaussian and any compactly supported product family attain $M_\psi M'_\psi = \Theta(\sqrt{d})$ and hence the order. This proves the scaling $C_d = \Theta(d^{-3/2})$ stated in Equation (4). Improving it requires removing the first-order mollification error itself – the same term Theorem 3 removes under a regularity hypothesis – not a cleverer mollifier. Table 12 measures the trade-off against this floor for three families.

Cutoff sensitivity. The cutoff enters in exactly two places: the global Lipschitz and sup bounds $L_R = R(4+15C_1/2)$ and $B_R = 15R^2/2$ of Step 0, and the constant c_χ of Lemma 3. Collecting that proof's terms with $\|x\| \leq 3R$, $|u| \leq 3R^2$, $\|\nabla\chi\|_\infty \leq C_1/R$ and $\|D^2\chi\|_{\mathrm{op}} \leq C_2/R^2$,

$$\|\mathrm{Hess}(\tilde{\phi}^*_\Lambda)\|_{\mathrm{tr}} \leq \underbrace{d + 6C_1 + \tfrac{9}{2}C_2 d}_{\text{quadratic part}} + \underbrace{2RM'_\psi d\Lambda + 2C_1 + 3C_2 d}_{\text{convex part}} \leq c_\chi d(1+2RM'_\psi\Lambda)$$

with $c_\chi = 1 + 8C_1 + \frac{15}{2}C_2$, using $d \geq 1$ to absorb the constant terms. For the Hermite cubic ramp $\chi(r) = 1 - 3t^2 + 2t^3$, $t = (r-2R)/R$ on the annulus $[2R, 3R]$, one has $\sup|\chi'| = 3/(2R)$ and $\sup|\chi''| = 6/R^2$, i.e. $C_1 = 3/2$, $C_2 = 6$, hence $c_\chi \leq 58$ and $L_R = 15.25R$. Two points deserve emphasis. First, all of these are dimension-free, so the cutoff profile moves the constant by an $O(1)$ factor and cannot change the $d^{3/2}$ scaling; alternate localisations (weighted potentials, or test functions tailored to the exponential decay of $p_G - p_{G'}$) can reduce L_R but not that scaling. Second, only $\|\nabla\chi\|_\infty$ and $\|D^2\chi\|_{\mathrm{op}}$ are used, so $\chi \in C_c^{1,1}$ suffices – the Hermite cubic is admissible even though it is not C^∞ – because the mollification in Step 1 supplies all the smoothness the distributional inversion identity needs. Finally, $\chi \equiv 1$ on $B(0,2R) \supset \mathrm{supp}(P - P')$, so the cutoff contributes *no* error to the pairing at all: its entire effect is on constants.

Unlocalised algebraic benchmark. For isotropic $\Sigma(c) = (1 + L_\Sigma c)\mathrm{Id}$ and single-Dirac mixing measures, the canonical uncut potential is globally quadratic and has $D^2\phi^* = \mathrm{Id}$. Formally, or rigorously after a cutoff approximation using the Laplace density's finite second moment, the unlocalised cross term is therefore $\frac{d}{2}|L_\Sigma(c - c')|$. This checks the algebra of the operator mismatch only. The actual proof uses $\chi\phi^*$, whose Hessian is not identically Id in the cutoff annulus, so the benchmark is not an exact test of the localised bound or of its worst-case dimensional dependence.

Remark 4 (Scope of the output-space corollaries). A density lower bound does not by itself yield output Wasserstein separation; that implication is false in general (Section C). The moment and sign-mollification results in Section 3.3 are conditional replacements with explicit coverage assumptions; they do not cover every condition-dependent shared-scale family.

C The Output-Metric Bridges

C.1 Why the naive implication fails

For general densities p, q, $\|p - q\|_1$ large does *not* imply $W_2(\mu_p, \mu_q)$ large. Take p uniform on $[0, 1]$ and $q = p + h$ with $h(x) = \varepsilon \sin(2\pi N x)$. Then $\|p - q\|_1 = \varepsilon \int_0^1 |\sin(2\pi N x)| dx = 2\varepsilon/\pi$, independent of N, while the excess mass of q need only be moved onto the adjacent deficit half-period, a distance $O(1/N)$. So $W_2 \to 0$ with N while $\|p - q\|_1$ stays fixed (Figure 3). Without control on the frequency content of $p - q$ no such implication can hold; the question is whether the mixture class supplies that control.

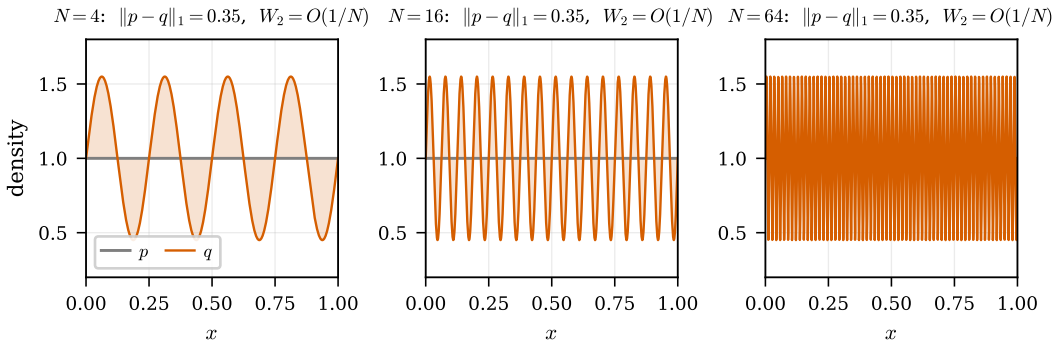

Figure 3: Why L^1 large $\not\Rightarrow$ W_2 large. The L^1 distance is independent of N, while the mass need only travel $O(1/N)$, so the transport cost vanishes while the density discrepancy does not.

C.2 Proofs of Propositions 1 and 2

Proof of Proposition 1. Put $v = m(\mu) - m(\mu')$ and take the 1-Lipschitz test function $h(x) = v \cdot x/\|v\|$ (the claim is trivial if $v = 0$). Since h is *linear*, $h(x+y) = h(x) + h(y)$ exactly, so $\int h\, d(P * \nu) = h(m(P)) + h(m(\nu))$ and identically for $P' * \nu'$, giving $\int h\, d\mu - \int h\, d\mu' = \|m(\mu) - m(\mu')\|$ with $m(\mu) = m(P) + m(\nu)$. Kantorovich–Rubinstein duality gives the W_1 bound. When all four input measures have finite second moments, $W_2 \geq W_1$. If $\nu = \nu'$, or if both kernels are centred, the kernel means cancel and this is $\|m(P) - m(P')\|$. □

This needs no bounded support and no dimension-dependent constant, and it never passes through L^1 or PDE inversion, so it avoids the dimension-dependence issue of Theorems 1 and 4 entirely. Stated with two kernels it also covers the condition-dependent shared-scale family $\nu_c = f(\cdot, \Sigma(c))$, which is centred for every c, so the displacement is exactly $\|m(P_c) - m(P_{c'})\|$ there.

Lemma 5 (Gelbrich bound). *Let $\tilde{\mu}, \tilde{\mu}'$ be centred measures in $\mathcal{P}_2(\mathbb{R}^d)$ with covariances A, B. Then $W_2(\tilde{\mu}, \tilde{\mu}')^2 \geq \mathfrak{B}^2(A, B)$, with equality when both are Gaussian.*

Proof. For any coupling π put $C_\pi = E_\pi[\tilde{X}\tilde{Y}^\top]$. Then $E_\pi\|\tilde{X} - \tilde{Y}\|^2 = \mathrm{tr}A + \mathrm{tr}B - 2\mathrm{tr}C_\pi$, and the block matrix $\left(\begin{smallmatrix} A & C_\pi \\ C_\pi^\top & B \end{smallmatrix}\right)$ is the covariance of $(\tilde{X}, \tilde{Y})$, hence positive semidefinite. Maximising $\mathrm{tr}C$

over C subject to that block constraint is a classical problem whose value is $\mathrm{tr}[(A^{1/2}BA^{1/2})^{1/2}]$ (Olkin & Pukelsheim, 1982; Gelbrich, 1990); the maximiser $C_0 = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$ is realised by the Gaussian coupling, which is exactly the Dowson–Landau optimal map between the two Gaussians (Dowson & Landau, 1982). Therefore $W_2^2(\tilde{\mu}, \tilde{\mu}') \geq \mathrm{tr}A + \mathrm{tr}B - 2\mathrm{tr}[(A^{1/2}BA^{1/2})^{1/2}]$, with equality in the Gaussian case because the bound is then attained. □

Proof of Proposition 2. Centre at the means: $\tilde{X} = X - m_\mu$, $\tilde{Y} = Y - m_{\mu'}$. For *any* coupling, $\|X - Y\|^2 = \|\tilde{X} - \tilde{Y}\|^2 + 2(\tilde{X} - \tilde{Y}) \cdot (m_\mu - m_{\mu'}) + \|m_\mu - m_{\mu'}\|^2$, and $E_\pi[\tilde{X}] = E_\pi[\tilde{Y}] = 0$ under every coupling, so the cross term vanishes in expectation and $W_2(\mu, \mu')^2 = \|m_\mu - m_{\mu'}\|^2 + W_2(\tilde{\mu}, \tilde{\mu}')^2$. Convolution adds independent summands, so $m_\mu = m(P) + m(\nu)$ and $\Sigma_\mu = \Sigma_P + \Sigma_\nu$, and likewise for μ'. Applying Lemma 5 to the centred parts with $A = \Sigma_\mu$, $B = \Sigma_{\mu'}$ gives the claim; equality for Gaussian μ, μ' is the equality case of Lemma 5 together with the exact mean split. □

Proof of Corollary 1. By Cauchy–Schwarz for Schatten norms, $\mathrm{tr}[(A^{1/2}BA^{1/2})^{1/2}] = \|A^{1/2}B^{1/2}\|_1 \leq \|A^{1/2}\|_2 \|B^{1/2}\|_2 = \sqrt{\mathrm{tr}A}\sqrt{\mathrm{tr}B}$, so $\mathfrak{B}^2(A, B) \geq \mathrm{tr}A + \mathrm{tr}B - 2\sqrt{\mathrm{tr}A\,\mathrm{tr}B} = (\sqrt{\mathrm{tr}A} - \sqrt{\mathrm{tr}B})^2$. □

The trace form is what a direct Cauchy–Schwarz on $E_\pi[\tilde{X} \cdot \tilde{Y}]$ produces. Example 1 shows the loss is not cosmetic: it can be the whole certificate. Writing $S := \mathrm{tr}\Sigma_P + \mathrm{tr}\Sigma_\nu$ and $S' := \mathrm{tr}\Sigma_{P'} + \mathrm{tr}\Sigma_{\nu'}$, the trace difference equals $(S - S')/(\sqrt{S} + \sqrt{S'})$ when $S + S' > 0$; the square-root form also covers $S = S' = 0$ without a $0/0$ expression. For a condition-dependent shared-scale family, $\Sigma_\mu - \Sigma_{\mu'}$ contains both the mixing and the kernel covariance change, so scale drift contributes to the certificate – and, unlike in the trace form, drift that rotates principal axes without changing total variance is still detected.

For the symmetric split $P = \delta_0 \to P' = \frac{1}{2}\delta_{-\epsilon} + \frac{1}{2}\delta_\epsilon$ we have $\Sigma_P = 0$, $\Sigma_{P'} = \epsilon^2$, and with $\sigma^2 := \mathrm{tr}\Sigma_\nu$ fixed this gives $W_2(\mu, \mu') \gtrsim \epsilon^2/(2\sigma)$ – order ϵ^2, exactly the rate measured in Table 9. Note the kernel variance Σ_ν is essential: dropping it would predict order ϵ and contradict the measurement.

C.3 The non-degeneracy hypothesis

Assumption 6 (Spectral non-degeneracy)**.** With $\mu_c = P_c * \nu_c$ and $\nu_c = f(\cdot, \Sigma(c))$ with covariance Σ_{ν_c} ($= \Sigma(c)$ for the Laplace kernel), the quantity

$$\rho(c, c') := \Big(\|m(\mu_c) - m(\mu_{c'})\|^2 + \mathfrak{B}^2\big(\Sigma_{P_c} + \Sigma_{\nu_c},\ \Sigma_{P_{c'}} + \Sigma_{\nu_{c'}}\big)\Big)^{1/2}$$

given by the square root of the right-hand side of Proposition 2 satisfies $\rho(c, c') \geq \alpha'' d_{\mathcal{C}}(c, c') - \kappa''$.

Because Proposition 2 is stated for possibly distinct kernels, this assumption and Corollary 3 apply to a condition-dependent shared-scale model; for a fixed centred kernel the kernel terms drop out of the mean and enter Σ_{μ_c} identically, so ρ reduces to a statement about P_c alone. Replacing the trace form by $\mathfrak{B}$ strictly enlarges the set of families the assumption covers, by Corollary 1, and strictly so on anisotropic families (Example 1).

Corollary 3 (Linear coarse lower bound)**.** *Under Assumption 6, $c \mapsto \mu_c$ satisfies the lower half of Equation* (1) *with $a = \alpha''$, $b = \kappa''$, and no other constants.*

Assumption 2 does *not* imply Assumption 6: the moment-matched family has $W_1(P, P') = \Theta(\epsilon)$ – a healthy quasi-isometric margin – while matching mean *and* covariance exactly, so $\rho \equiv 0$. Matching the full covariance, not merely its trace, is a strictly stronger requirement on the adversary, which is the sense in which the Bures certificate is harder to defeat. It must still be stated explicitly rather than derived. It is satisfied for encodings whose mean-or-covariance displacement is itself quasi-isometric in the condition metric, for example a nondegenerate translation, a scale path, or a rotation of anisotropy; it need not hold for a generic small-mass transfer or split/merge parametrisation and must be checked in each family.

C.4 Proof of Proposition 3

The structural fact the proof exploits is that $g = (P - P') * f(\cdot, \Sigma)$ inherits its regularity from the kernel alone. In particular, for any unit e, the distributional derivative satisfies $\partial^2_{ee} g = (P - P') * \partial^2_{ee} f$, so

$$\left\|\partial^2_{ee} g\right\|_{\mathrm{TV}} \;\le\; \|P - P'\|_{\mathrm{TV}} \left\|\partial^2_{ee} f\right\|_{\mathrm{TV}} \;\le\; 2K_2(\Sigma, d), \tag{7}$$

regardless of how rough P, P' are – only $\|P - P'\|_{\mathrm{TV}} \le 2$ is used, so P, P' may be atomic with infinitely many atoms, as in the Dirichlet-process setting.

Lemma 6 (Mollification error at second order)**.** *Let ψ be a nonnegative even probability density with $m_2(\psi) = \int \|u\|^2 \psi(u) du < \infty$, and let $K_2(\Sigma, d) < \infty$. Then $\|g * \psi_\Lambda - g\|_1 \le m_2(\psi) K_2(\Sigma, d)\, \Lambda^{-2}$.*

Proof. For fixed y, use the central second difference

$$\Delta^2_y g := g(\cdot - y) + g(\cdot + y) - 2g.$$

For $y \neq 0$ put $e = y/\|y\|$. In the sense of distributions,

$$\Delta^2_y g = \int_{-1}^{1} (1 - |t|)\|y\|^2 \partial^2_{ee} g(\cdot - ty)\, dt.$$

The integral is understood weakly against test functions. Applying the identity to a test bounded by one, then taking the supremum, avoids any need for translation continuity in total variation. Since $\Delta^2_y g \in L^1$, its measure total variation equals its L^1 norm, and translation invariance together with Equation (7) gives

$$\|\Delta^2_y g\|_1 \le \|y\|^2\, 2K_2 \int_{-1}^{1} (1 - |t|) dt = 2K_2 \|y\|^2.$$

This identity is valid without assuming that ∇g is an L^1 function. Since ψ_Λ is even,

$$g * \psi_\Lambda - g = \frac{1}{2} \int \Delta^2_y g\, \psi_\Lambda(y)\, dy,$$

and therefore

$$\|g * \psi_\Lambda - g\|_1 \le \frac{1}{2} \int \|\Delta^2_y g\|_1 \psi_\Lambda(y) dy \le K_2 \int \|y\|^2 \psi_\Lambda(y) dy = \frac{m_2(\psi) K_2}{\Lambda^2}. \qquad \square$$

Remark 5 (Why not the Laplacian). It is tempting, within this pointwise Taylor-remainder argument, to write the second-order term as $\frac{m_2(\psi)}{2\Lambda^2}\|\Delta g\|_1$, using radial symmetry to turn $\int y_i y_j \psi_\Lambda(y) dy$ into $(m_2/d\Lambda^2)\delta_{ij}$. That step is *invalid* for $d > 1$: the Taylor remainder evaluates $D^2 g$ at an intermediate point depending on y, so the absolute value must be taken before the radial average, and what survives is a directional second derivative, not the trace. Lemma 6 is the correct form, and the reason K_2 in Equation (5) is defined through directional derivatives. This does not preclude the semigroup proof below from legitimately exploiting the full Laplace generator (the Laplacian in the isotropic case).

Proof of Proposition 3. By Kantorovich–Rubinstein, $W_1(\mu, \mu') = \sup_{\mathrm{Lip}(h) \le 1} \int hg$. If $V = 0$ the claim is trivial, so assume $V > 0$. Put $h_\Lambda := \mathrm{sign}(g) * \psi_\Lambda$. Since $|\mathrm{sign}(g)| \le 1$ *everywhere* – unlike ϕ^* in Section A.2, no cutoff is needed here – moving one derivative onto the mollifier gives $\mathrm{Lip}(h_\Lambda) \le \Lambda M_1(\psi)$ with $M_1(\psi) = \int \|\nabla \psi\| du$, so $\tilde{h}_\Lambda := h_\Lambda / (\Lambda M_1(\psi))$ is admissible and

$$W_1(\mu, \mu') \;\ge\; \frac{1}{\Lambda M_1(\psi)} \int (\mathrm{sign}(g) * \psi_\Lambda)\, g\, dx.$$

As ψ_Λ is even, convolution is self-adjoint in this pairing: $\int (\mathrm{sign}(g) * \psi_\Lambda) g = \int \mathrm{sign}(g) g_\Lambda$ with $g_\Lambda = g * \psi_\Lambda$. Splitting $g_\Lambda = g + (g_\Lambda - g)$ and using $\int \mathrm{sign}(g) g = \|g\|_1 = V$ with Hölder, $\int \mathrm{sign}(g) g_\Lambda \ge V - \|g_\Lambda - g\|_1$, and Lemma 6 bounds the latter by c_1/Λ^2 with $c_1 := m_2(\psi) K_2$. With $c_2 := M_1(\psi)$, $W_1(\mu, \mu') \ge (V - c_1/\Lambda^2)/(c_2 \Lambda)$. Maximising over Λ: the derivative vanishes at $\Lambda^* = \sqrt{3c_1/V}$, where the numerator is $\frac{2}{3}V$, giving $W_1(\mu, \mu') \ge \frac{2}{3c_2\sqrt{3c_1}} V^{3/2}$. $\square$

C.5 When K_2 is finite

Proposition 3 is only as general as Equation (5). For a kernel with $D^2 f \in L^1$ – the Gaussian $N(0, \Sigma)$ density in particular – K_2 is finite in every dimension. Indeed, for a unit vector e the Gaussian formula and the triangle inequality give $\|\partial^2_{ee} f\|_1 \le 2e^\top \Sigma^{-1} e \le 2/\lambda_{\min}$.

For the multivariate Laplace kernel, in $d = 1$ with scalar scale s,

$$f_s'' = \frac{2}{s}(f_s\,dx - \delta_0), \qquad K_2 = \frac{4}{s} \le \frac{4}{\lambda_{\min}},$$

so Proposition 3 applies. The directional-K_2 criterion fails for $d \ge 2$. In standardised isotropic coordinates the kernel is $f(x) \propto \|x\|^{(2-d)/2}\mathcal{K}_{(d-2)/2}(\sqrt{2}\|x\|)$, which behaves like $-\log\|x\|$ near the origin when $d = 2$ and like $\|x\|^{2-d}$ when $d \ge 3$. For any fixed unit e, $|\partial^2_{ee} f|$ is a nonzero angular factor times $\|x\|^{-d}$ on a set of directions of positive surface measure. Consequently

$$\int_{\|x\|<1} |\partial^2_{ee} f(x)|\,dx \;\gtrsim\; \int_0^1 r^{-d}\, r^{d-1}\,dr \;=\; \int_0^1 \frac{dr}{r} \;=\; \infty.$$

An invertible linear change of variables carries a directional derivative for an admissible anisotropic Σ to a nonzero directional derivative in these standardised coordinates and does not turn this non-finite measure into a finite one. Thus $K_2 = \infty$ for every $d \ge 2$. This is a limitation of the directional criterion, not an impossibility result: the Laplace inversion operator itself provides the alternative heat-semigroup bridge below.

Proof of Proposition 4. Write $Q = P - P'$, $g = p_G - p_{G'}$, and $L_\Sigma = \Sigma : \mathrm{Hess}$. Laplace inversion gives $L_\Sigma g = 2(g\,dx - Q)$ as a finite signed measure, whence

$$\|L_\Sigma g\|_{\mathrm{TV}} \le 2(\|g\|_1 + \|Q\|_{\mathrm{TV}}) \le 8.$$

Let $(H_t)_{t\ge 0} = \exp(tL_\Sigma)$ be the Gaussian heat semigroup with covariance $2t\Sigma$. Its contraction on finite signed measures and the semigroup identity imply

$$\|(H_t g - g)\,dx\|_{\mathrm{TV}} \le \int_0^t \|H_s(L_\Sigma g)\|_{\mathrm{TV}} ds \le 8t.$$

The measure on the left is absolutely continuous, so this is precisely the L^1 error. For $h_t = H_t \operatorname{sign}(g)$ and a unit vector e, the Gaussian heat kernel q_t satisfies

$$\|\partial_e q_t\|_1 = \sqrt{\frac{e^\top \Sigma^{-1} e}{\pi t}} \le \frac{1}{\sqrt{\pi \lambda_{\min} t}}.$$

Taking the supremum over e therefore gives $\mathrm{Lip}(h_t) \le (\pi\lambda_{\min} t)^{-1/2}$. Self-adjointness of H_t gives $\int h_t g = \int \operatorname{sign}(g) H_t g \ge V - 8t$. Normalising h_t in the Kantorovich–Rubinstein dual and choosing $t = V/24$ (the case $V = 0$ is trivial) yields

$$W_1(\mu, \mu') \ge (V - 8t)\sqrt{\pi\lambda_{\min} t} = \frac{\sqrt{\pi\lambda_{\min}}}{3\sqrt{6}} V^{3/2}.$$

□

Remark 6 (Where the exponent comes from). The $3/2$ balances a numerator losing accuracy at Λ^{-2} (mollifying g, whose directional second derivatives are uniformly bounded finite measures) against a Lipschitz constant growing at Λ^{+1} (mollifying the discontinuous sign, which cannot be improved past first order without further information). This is a different balance from Theorem 1, where the mollified object was itself only Lipschitz. The resulting powers belong to different inequalities and should not be compared as alternative rates for a common pair of quantities.

Remark 7 (The kernel's cusp, and why it is not fatal in $d = 1$). The Laplace density has a cusp: its distributional second derivative carries a Dirac mass at the origin, so $\|\Delta f(\cdot, \Sigma)\|_{L^1}$ is not well-defined. This is not an obstruction, for two reasons. First, as Remark 5 explains, the Laplacian is not the relevant quantity in the pointwise Taylor-remainder proof; the directional quantity Equation (5) is. Second, once the constant is measured in *total variation* rather than L^1, the Dirac is not an obstruction but simply a contribution: in $d = 1$, $f'' = 2f - 2\delta_0$ and $K_2 = 4$ at unit scale (and

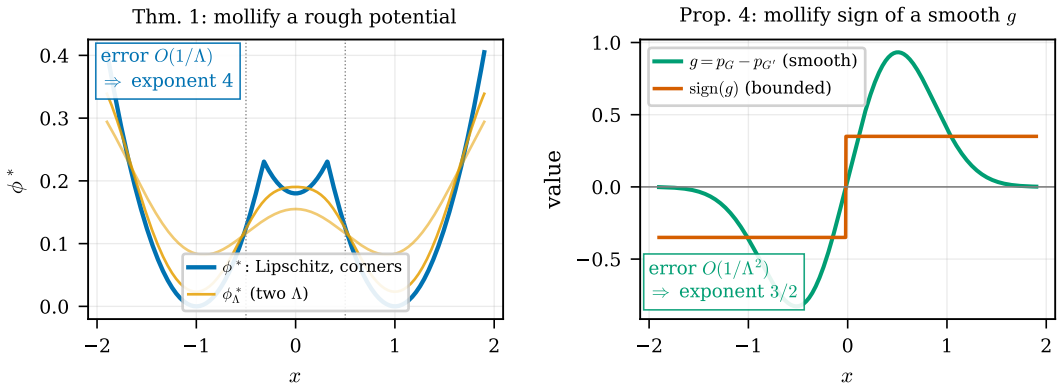

Figure 4: The two mollification strategies side by side. Theorem 1 smooths a potential that has corners and is Lipschitz only, so its error is first order in $1/\Lambda$. Proposition 3 instead smooths $\mathrm{sign}(g)$ while exploiting that g itself is regular, so the error is second order. The two powers describe different links: $V \gtrsim W_2(P, P')^4$ for latent inversion and $W_1(\mu, \mu') \gtrsim V^{3/2}$ for the output bridge; they are not competing rates for the same quantities.

$K_2 = 4/s$ at scalar scale s), so Lemma 6 applies directly and no separate argument about kinks is needed.

Note also what Equation (7) does *not* require. The bound $\|\partial^2_{ee} g\|_{\mathrm{TV}} \le 2K_2$ uses only $\|P - P'\|_{\mathrm{TV}} \le 2$; it is uniform over all pairs of mixing measures and in particular does not care how many atoms they have. No restriction to finitely many kinks of g is needed – such a restriction would exclude the Dirichlet-process mixtures this framework targets – because the non-smoothness is placed in the kernel rather than in g. The real restriction is that the directional K_2 criterion is finite only for $d = 1$ in the Laplace case. Proposition 4 bypasses that criterion for a fixed Laplace kernel in every dimension.

Remark 8 (A second, independent dimension dependence). For a Gaussian mollifier $M_1(\psi) = E_{U\sim N(0,\mathrm{Id})}\|U\| \le \sqrt{E\|U\|^2} = \sqrt{d}$, so C_{sign} in Proposition 3 carries an explicit $\sqrt{d}$ from the mollifier's own gradient scaling – independent of, and additional to, K_2's own dimension dependence and Proposition 7's $O(d)$. A fully dimension-quantified version would need all three tracked; we do not claim any of the three is $O(1)$.

C.6 Different kernels: proof of Proposition 5

Step (i) of Proposition 4 uses that *both* densities are mixtures of the *same* f_Σ: for $\Sigma \ne \Sigma'$, $L_\Sigma p_{G'} \ne 2P'$ and the generator identity fails. We therefore interpolate in the scale and pay for it once, additively. Two matrix lemmas do the work.

Lemma 7 (Bures perturbation). *For $A, B \succeq \lambda_{\min}\mathrm{Id}$, $\mathfrak{B}(A, B) \le \|A - B\|_F/(2\sqrt{\lambda_{\min}})$.*

Proof. Write $X = A^{1/2}$, $Y = B^{1/2}$. Since $\mathrm{tr}[(A^{1/2}BA^{1/2})^{1/2}] = \|XY\|_1 = \mathrm{tr}|XY| \ge \mathrm{tr}(XY)$, we get $\mathfrak{B}^2(A, B) \le \mathrm{tr}A + \mathrm{tr}B - 2\mathrm{tr}(XY) = \|X - Y\|_F^2$. Let $\{e_i\}, \{f_j\}$ be orthonormal eigenbases of X, Y with eigenvalues $x_i, y_j \ge \sqrt{\lambda_{\min}}$. From $A - B = X(X - Y) + (X - Y)Y$, $\langle e_i, (A - B)f_j\rangle = (x_i + y_j)\langle e_i, (X - Y)f_j\rangle$, so $\|X - Y\|_F^2 = \sum_{i,j} \frac{|\langle e_i, (A-B)f_j\rangle|^2}{(x_i+y_j)^2} \le \frac{\|A-B\|_F^2}{4\lambda_{\min}}$. □

Lemma 8 (L^1 cost of a scale change). *Let $\Sigma, \Sigma' \in \mathcal{S}_d^+(\lambda_{\min}, \lambda_{\max})$, $\Delta = \Sigma' - \Sigma$, $\kappa_\Sigma = \lambda_{\max}/\lambda_{\min}$. Then $\delta := \|f_\Sigma - f_{\Sigma'}\|_1 \le C_\delta \|\Delta\|_F$ with $C_\delta = \sqrt{\gamma_{\kappa_\Sigma}}/\lambda_{\min}$ and $\gamma_\kappa = \sup_{s\in[1/\kappa,\kappa]}(s - 1 - \log s)/(s - 1)^2 = (\kappa - \kappa^2 + \kappa^2 \log\kappa)/(\kappa - 1)^2$, with $\gamma_1 = \frac{1}{2}$.*

Proof. Using the normal-variance representation $f_\Sigma = \int N(0, e\Sigma)\, e^{-e} de$ and the triangle inequality, $\delta \le \int \|N(0, e\Sigma) - N(0, e\Sigma')\|_1 e^{-e} de$. Pinsker's inequality gives $\|p - q\|_1 \le \sqrt{2\,\mathrm{KL}(q\|p)}$,

and $\mathrm{KL}(N(0,e\Sigma')\|N(0,e\Sigma)) = \frac{1}{2}[\mathrm{tr}(\Sigma^{-1}\Sigma') - d - \log\det(\Sigma^{-1}\Sigma')]$ does not depend on e, so the integral is that single value. Put $S = \Sigma^{-1/2}\Sigma'\Sigma^{-1/2}$, whose eigenvalues s_i lie in $[1/\kappa_\Sigma, \kappa_\Sigma]$ because $\lambda_{\min}\mathrm{Id} \preceq \Sigma, \Sigma' \preceq \lambda_{\max}\mathrm{Id}$. The map $s \mapsto (s - 1 - \log s)/(s-1)^2$ is decreasing, so its supremum on that interval is at $s = 1/\kappa_\Sigma$, giving γ_{κ_Σ} and

$$2\,\mathrm{KL} = \sum_i (s_i - 1 - \log s_i) \le \gamma_{\kappa_\Sigma}\|S - \mathrm{Id}\|_F^2 = \gamma_{\kappa_\Sigma}\|\Sigma^{-1/2}\Delta\Sigma^{-1/2}\|_F^2 \le \gamma_{\kappa_\Sigma}\frac{\|\Delta\|_F^2}{\lambda_{\min}^2}. \quad \square$$

The one-sided bound $s - 1 - \log s \le \frac{1}{2}(s-1)^2$, valid only for $s \ge 1$, is *not* available here: for $s < 1$ the inequality reverses, which is why the constant carries the condition number κ_Σ.

Proof of Proposition 5. Write $\tilde{V} := \|P * f_\Sigma - P' * f_\Sigma\|_1$. By the triangle inequality for W_1,

$$W_1(\mu,\mu') \;\ge\; W_1(P * f_\Sigma,\, P' * f_\Sigma) - W_1(P' * f_\Sigma,\, P' * f_{\Sigma'}).$$

For the first term, Proposition 4 applies verbatim (a common kernel f_Σ) and gives $c_{\mathrm{heat}}\tilde{V}^{3/2}$. For the relation between $\tilde{V}$ and V, Young's inequality gives $V \le \tilde{V} + \|P' * (f_\Sigma - f_{\Sigma'})\|_1 \le \tilde{V} + \delta$, so $\tilde{V} \ge V - \delta \ge 0$ under the hypothesis. For the second term, convexity of W_1 under a shared mixing measure gives $W_1(P' * f_\Sigma, P' * f_{\Sigma'}) \le \int W_1(f_\Sigma(\cdot - \theta), f_{\Sigma'}(\cdot - \theta))dP'(\theta) = W_1(f_\Sigma, f_{\Sigma'}) \le W_2(f_\Sigma, f_{\Sigma'})$. Coupling the two normal-variance mixtures through their shared $E \sim \mathrm{Exp}(1)$ and using $W_2^2(N(0,e\Sigma), N(0,e\Sigma')) = e\,\mathfrak{B}^2(\Sigma,\Sigma')$,

$$W_2^2(f_\Sigma, f_{\Sigma'}) \le \int e\,\mathfrak{B}^2(\Sigma,\Sigma')e^{-e}de = \mathfrak{B}^2(\Sigma,\Sigma') \le \frac{\|\Delta\|_F^2}{4\lambda_{\min}}$$

by Lemma 7, i.e. the second term is at most $C_2\|\Delta\|_F$. Finally, convexity of $t \mapsto t^{3/2}$ gives $(V-\delta)^{3/2} \ge V^{3/2} - \frac{3}{2}V^{1/2}\delta$, and Lemma 8 bounds $\delta \le C_\delta\|\Delta\|_F$. Collecting terms proves the display. Monotonicity: the derivative of $t \mapsto c_{\mathrm{heat}}t^{3/2} - C_1\|\Delta\|_F t^{1/2}$ is nonnegative once $t \ge C_1\|\Delta\|_F/(3c_{\mathrm{heat}}) = \frac{1}{2}C_\delta\|\Delta\|_F$, which the hypothesis $V \ge C_\delta\|\Delta\|_F$ implies. $\square$

Activation in the condition metric. Under Assumption 3, $\|\Sigma(c) - \Sigma(c')\|_F \le \sqrt{d}\,L_\Sigma d_{\mathcal{C}}(c,c')$, so with v any lower bound on V satisfying $v \ge C_\delta\sqrt{d}L_\Sigma d_{\mathcal{C}}$ – for instance $v = (C_{4,d}W_2(P_c,P_{c'})^4 - L_d L_\Sigma d_{\mathcal{C}})_+$ from Theorem 4, which is the latent statement available when the scales differ –

$$W_1(\mu_c,\mu_{c'}) \;\ge\; c_{\mathrm{heat}}v^{3/2} - \big(C_1 v^{1/2} + C_2\big)\sqrt{d}\,L_\Sigma d_{\mathcal{C}}(c,c').$$

Exactly as in Definition 1, this is a statement with an activation threshold rather than a uniform guarantee: it is non-vacuous only where the density signal exceeds the scale drift, it degrades to Proposition 4 as $L_\Sigma \to 0$, and the slack it pays is $O(\sqrt{d}\,L_\Sigma d_{\mathcal{C}})$, so it is polynomial separation with additive slack – not a linear coarse embedding, and not a heteroscedastic analogue of Equation (2), whose scale term is positive rather than subtracted. What it does supply is the first output-space comparison in this framework between mixtures whose kernels genuinely differ.

C.7 The composed chain

For a common fixed Laplace kernel, composing Proposition 4 with Theorem 1 ($V \ge C_d\,W_2(P,P')^4$) gives Corollary 2:

$$W_2(\mu,\mu') \ge W_1(\mu,\mu') \ge \frac{\sqrt{\pi\lambda_{\min}}}{3\sqrt{6}}\,C_d^{3/2}W_2(P,P')^6.$$

Under Assumption 2, this further gives

$$W_1(\mu_c,\mu_{c'}) \ge \frac{\sqrt{\pi\lambda_{\min}}}{3\sqrt{6}}C_d^{3/2}(\alpha d_{\mathcal{C}}(c,c') - \kappa)_+^6.$$

In $d = 1$, Proposition 3 supplies the same degree with its own constant. For a non-Laplace kernel, the generic sign bridge can be composed only with a separately available inverse theorem for that same kernel. This is polynomial separation, not the linear lower bound of Corollary 3. It holds regardless of moment non-degeneracy but requires the compared conditions to share the fixed kernel. Conversely, Corollary 3 applies only to families satisfying its explicit moment-displacement hypothesis.

D EXPERIMENTAL DETAILS

D.1 PIPELINE STAGES

Six stages run in sequence, each writing atomically before the next starts: (1) pilot validation of the exact discrete-transport and importance-sampling estimators, gated on adaptive sample size, monotonicity of W_1, W_2 in ϵ, and 1D quadrature agreement; (2) the inverse-rate sweep (Table 2); (2b) the worst-case direction/covariance search (Table 3); (3) the condition-dependent shared-scale sweep including the direct I_2 computation (Table 4); (4) the coarse-geometry Sinkhorn sweep (Table 5); (5) aggregation into figures, tables and a machine-checkable claim-verdict report. Every Monte Carlo estimate reports within- and between-seed uncertainty; the pilot gates the whole run and the runner refuses to proceed on failure.

Table 2: Inverse-rate sweep configuration (stage 2).

Parameter	Value
Dimensions	$\{1, 2, 4, 8, 16, 32, 64\}$
ϵ range	20 points, geometric, $[0.02, 0.50]$
Families	translation, split/merge, moment-matched, small-mass transfer
Kernels	Laplace, Gaussian
Seeds per cell	5
Base samples/seed	12,000, adaptively doubled to a cap of 200,000
Relative-CI target	10%
Support radius R	2.0

Table 3: Worst-case direction/covariance search (stage 2b).

Parameter	Value
Dimensions	$\{2, 8, 32, 64\}$
ϵ values	$\{0.02, 0.054, 0.144, 0.316\}$
Random directions per cell	12 (plus axis and dense)
Random anisotropic covariances	6 (plus isotropic), eigenvalues in $[0.25, 4.0]$
Screening samples	3,000 per candidate
Refinement	60,000/seed, 5 seeds, top-3 worst by the auxiliary diagnostic L_1/W_2^3 and the proved-rate ratio L_1/W_2^4

Table 4: Condition-dependent shared-scale sweep and I_2 configuration (stage 3).

Parameter	Value
Dimensions	$\{2, 8, 32\}$
L_Σ values	$\{0.0, 0.05, 0.1, 0.2, 0.5\}$
Condition distance	14 points, geometric, $[0.02, 1.0]$
I_2 outer samples (draws from $p_{G_{c'}}$)	4,000
I_2 inner samples (Stein-identity draws)	64
Mollifier bandwidth	$h = W_2(P_c, P_{c'})$ (diagnostic; not the proof's Λ^*, Section E.3)

D.2 ADAPTIVE PRECISION

Each rate-sweep cell mirrors the pilot's adaptive-doubling gate: the cross-seed relative CI is recomputed after each doubling and the cell stops when it clears the 10% target or hits the cap. This is per-cell, not per-sweep, so easy cells terminate quickly while hard cells (moment-matched at small ϵ and large d, where the signal is a fourth-order cancellation) consume the full budget. Cells still failing at the cap are flagged with a `cell_ci_gate_passed` column rather than silently averaged in.

Table 5: Coarse-geometry Sinkhorn sweep (stage 4).

Parameter	Value
Dimension	8
Conditions	50 evenly spaced on $[0, 1]$ (1,225 unique pairs)
Samples per condition	1,024
Sinkhorn regularisation	$\{0.10, 0.05\}$, chosen by known-correspondence calibration
Sinkhorn iterations	300
Null-floor sample sizes	$\{256, 512, 1024, 2048, 4096\}$, 3 trials each
Slope grid (frontier)	50 points on the originally planned $[0.1, 2.0]$ range

D.3 Worst-case search

Per $(d, \epsilon, \text{family}, \text{kernel})$ cell, candidate directions are the coordinate axis, the dense direction $\mathbf{1}/\sqrt{d}$, and 12 random unit vectors; candidate covariances are the identity and 6 random $Q\,\mathrm{diag}(\lambda_1, \ldots, \lambda_d)\,Q^\top$ with Q random orthogonal (QR of a Gaussian matrix) and eigenvalues log-uniform on $[0.25, 4.0]$. All combinations are screened cheaply; the top-3 worst by each ratio, plus the axis/isotropic baseline (always included, for a precision-matched comparison), are refined at full precision.

D.4 The real PDE cross term

I_2 uses the discrete Kantorovich c-transform potential $\phi(x) = \min_j[\frac{1}{2}\|x - y_j\|^2 - v_j]$, with v the dual solution of the discrete optimal-transport LP between the two mixing measures (POT's exact LP solver), together with the Gaussian-mollifier Stein identity

$$\Sigma : \mathrm{Hess}(\phi_\Lambda)(x) = E_{Z\sim N(0,h^2\mathrm{Id})}\Big[\phi(x - Z) \cdot \frac{\Sigma : (ZZ^\top - h^2\mathrm{Id})}{h^4}\Big],$$

which never differentiates ϕ (only Lipschitz, as throughout) and instead pushes both derivatives analytically onto the smooth mollifier. It follows from $\mathrm{Hess}(f * g) = f * \mathrm{Hess}(g)$ and the closed-form Gaussian Hessian. We validated it against the exact single-Dirac answer (Section E) before trusting it on the sweep.

D.5 Coarse-geometry calibration

Two procedures run before the main sweep. *Known-correspondence calibration* translates one fixed point cloud by known distances and selects the regularisation minimising relative error – a check on estimator bias under a trivial, exactly-known coupling, *not* a measure of real noise. *Null-floor calibration* draws two *independent* samples from the *same* distribution at each sample size and computes their debiased Sinkhorn distance, whose population value is exactly 0; anything measured is pure finite-sample noise. The second is what flags real pairs as at-or-below floor. The sweep additionally uses common random numbers: one shared draw of component assignment and base noise, transformed per condition, rather than independent draws per condition.

D.6 Configurations for the certificate experiments

Two implementation points generalise beyond this paper. First, the multivariate Laplace density must be evaluated through its Bessel closed form $f \propto (q/2)^{\nu/2}\mathcal{K}_\nu(\sqrt{2q}), \nu = (2-d)/2$: evaluating it by Gauss–Laguerre quadrature against the $\mathrm{Exp}(1)$ variance mixture is singular at the kernel centre for $d \geq 2$, whereas the heat flow at $t > 0$ is smooth and quadrature is fine there. Second, importance-sampling proposals must have heavier tails than the integrand; a Gaussian proposal against a Laplace integrand makes the weights diverge, which we observed before switching to a product-Laplace proposal with scale 2.

Table 6: Configurations for the five certificate experiments of Section 4. All integrands are evaluated in closed form; only the entries marked IS/MC are estimated.

Experiment	Setting	Evaluation
Heat chain, $d = 2$ (Section E.8)	5- and 4-atom P, P' in $B(0, 1.5)$, diagonal Σ	exact grid, Bessel density
Heat chain, $d = 4, 8$	same family; $t \in \{0.05, 0.1, 0.2, 0.4\}$, $t^* = \widehat{V}/24$	IS, $N = 1.2 \times 10^5$
Bures vs trace (Section E.9)	$d = 2$, $\varepsilon \in [10^{-3}, 1]$, $\Sigma_\nu \in \{0, 0.5\mathrm{Id}\}$	closed form
Mollifier (Section E.10)	$d \in \{1, \ldots, 64\}$, three families	closed form (M_ψ MC)
MDN (Section E.11)	$K = 5$ Laplace components, shared scale, $2K+1$ outputs	exact 1D quantile
Sharpness III (Section E.4)	six-atom family, $\epsilon \in [0.0012, 0.04]$	exact 1D quantile

E Additional Numerical Verification

E.1 The latent-rate sweep

Figure 5 (left) shows the smallest sampled ratio L_1/W_2^4 is nearly flat over $d = 1, \ldots, 64$ (max/min factor 1.11). This is a finite-family diagnostic, not an estimate of the optimal constant; the plotted $q = 3$ curve is an auxiliary diagnostic only and has no theorem attached to it. A worst-case search over 12 random directions $\times$ 6 anisotropic covariances per cell (right) finds configurations well below the axis-only value at low d (median $0.52\times$, min $0.07\times$ at $d = 2$), so axis-aligned configurations are not worst-case there – but the gap *shrinks* with dimension (median $0.94\times$, min $0.61\times$ at $d = 64$) on this finite search.

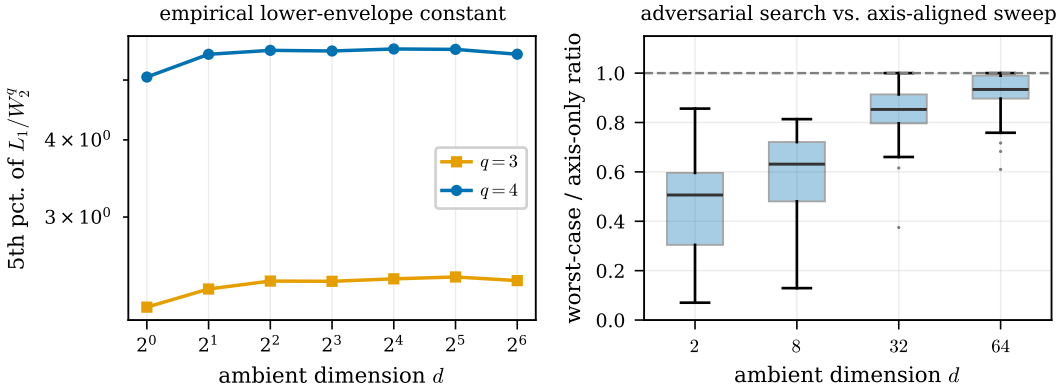

Figure 5: Left: smallest sampled L_1/W_2^q ratios; only $q = 4$ corresponds to Theorem 1, while $q = 3$ is an auxiliary diagnostic. Right: distribution over sampled cells of (worst case found)/(axis-only) for the auxiliary L_1/W_2^3 ratio. Neither panel estimates the optimal theorem constant.

The cross-term benchmark. We compute I_2 from the discrete Kantorovich c-transform potential and a Gaussian-mollifier Stein identity that never differentiates ϕ^* (Section D). Figure 8 shows $|I_2|$ is linear in its claimed driver $L_\Sigma d_{\mathcal{C}}(c, c')$ and scales as $0.054 \to 0.217 \to 0.869$ across $d = 2, 8, 32$ – ratios 4.02 and 4.00 against dimension ratios of exactly 4, matching the uncut single-Dirac identity $\frac{d}{2}|L_\Sigma \Delta c|$. This checks the algebra of I_2 before cutoff localisation; it does *not* establish Theorem 4, the localised $O(d[1 + RM'_\psi(d)])$ bound, or its behaviour for multi-atom potentials.

E.2 Inverse-rate curves and fitted slopes

Figure 6 shows the raw Laplace inverse-rate curves underlying Figure 5; Figure 7 shows the fitted small-scale slopes by family, kernel and dimension. The families separate cleanly by order and the

ordering is stable across dimension, which makes the plotted finite-family minimum in Figure 5 interpretable.

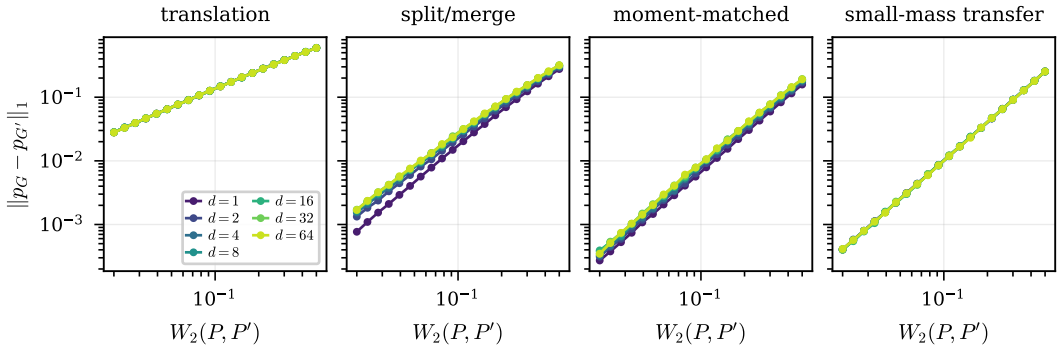

Figure 6: Laplace inverse-rate curves, axis-aligned families, all dimensions.

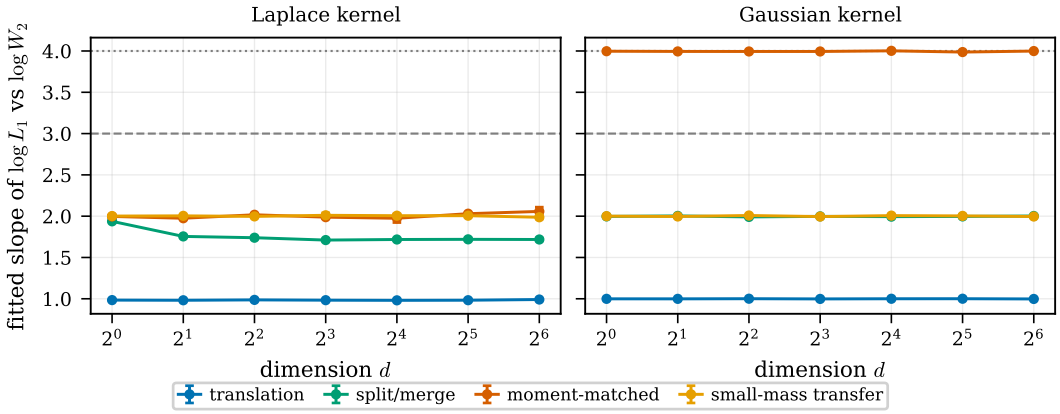

Figure 7: Fitted small-scale slope of $\log L_1$ against $\log W_2$. Dashed and dotted lines mark the auxiliary cubic diagnostic and the proved quartic rate, respectively. Only the line at 4 corresponds to a theorem.

E.3 Bandwidth-independence of I_2

For `shared`-mode cells both mixing measures are single Diracs, so the *uncut* ϕ is globally quadratic and $\mathrm{Hess}(\phi_\Lambda) = \mathrm{Id}$ for every bandwidth (Section B.3). We checked this unlocalised algebraic identity: at $d = 8$ with $L_\Sigma \Delta c = 0.06$ (prediction $-\frac{1}{2}\mathrm{tr}(\Sigma(c) - \Sigma(c')) = 0.240$), Table 7 reports the measured I_2 at four bandwidths.

Table 7: I_2 at different mollifier bandwidths against the exact closed-form prediction 0.240; 4,000 outer $\times$ 64 inner samples.

Bandwidth convention	h	Measured I_2
Reciprocal convention ($h = 1/W_2$)	6.667	0.238
Small-bandwidth diagnostic ($h = W_2$; not the proof optimiser)	0.150	0.266
Arbitrary	1.000	0.238
Small (what the correct convention gives as $W_2 \to 0$)	0.010	7.903

The first three rows are consistent with bandwidth-independence of this unlocalised identity. In particular, whether the bandwidth is taken as the standard deviation $1/\Lambda$ of $\psi_\Lambda(x) = \Lambda^d \psi(\Lambda x)$ or

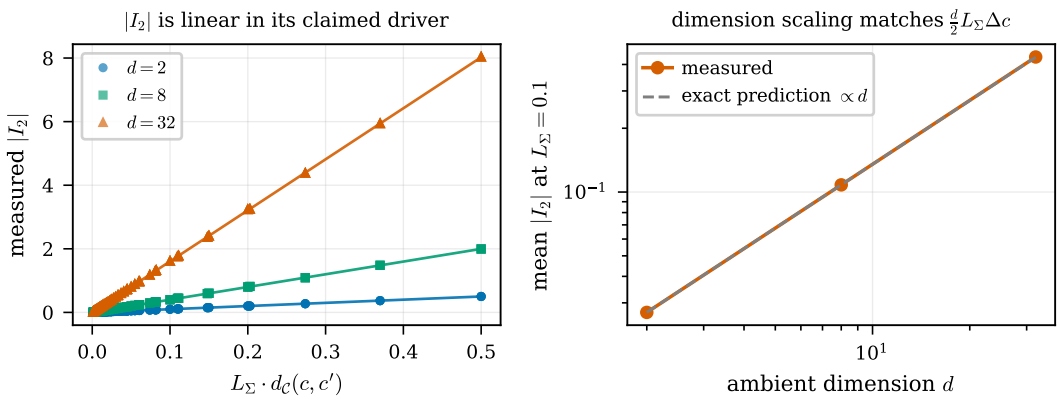

Figure 8: The condition-dependent shared-scale cross term I_2 in the unlocalised isotropic single-Dirac benchmark. Its observed linearity in $L_\Sigma d_C$ and d checks that algebraic configuration only, not the cutoff-localised Theorem 4 or its worst-case constant.

as its reciprocal does not visibly change I_2 in this configuration. Neither $h = W_2$ nor its reciprocal is the optimiser in the quartic proof; that optimiser depends on the observed density discrepancy. The fourth row is a separate finding: the Stein-identity estimator's variance scales as $1/h^4$, so at small bandwidth it overwhelms a signal that is in truth bandwidth-independent. That is the obstacle flagged in the numerical estimator; it does not affect the algebraic benchmark but prevents this experiment from validating the localised multi-atom or anisotropic bound.

E.4 Sharpness probes

Figure 9 generalises the moment-matched family to $2m$-th order stencils matching moments $0, \ldots, 2m-1$. For the Gaussian kernel the L_1 order tracks the stencil order without apparent bound (3.98, 5.95, 7.91, 9.48 for $m = 2, 3, 4, 5$; $R^2 \geq 0.996$, the mild degradation at $m = 5$ being the quadrature floor at $L_1 \sim 10^{-16}$, not statistical). This does not threaten Theorem 1, which concerns the Laplace operator T_Σ specifically and says nothing about the Gaussian kernel; it is consistent with the analytic characteristic function and Hermite structure behind classical strong-identifiability results for Gaussian location mixtures. For the Laplace kernel the order stays pinned at ≈ 2 regardless of stencil order (1.997, 1.991, 1.991, 1.992; $R^2 = 1.0000$), comfortably *above* the quartic floor – so this family is far from tight and cannot decide sharpness either way.

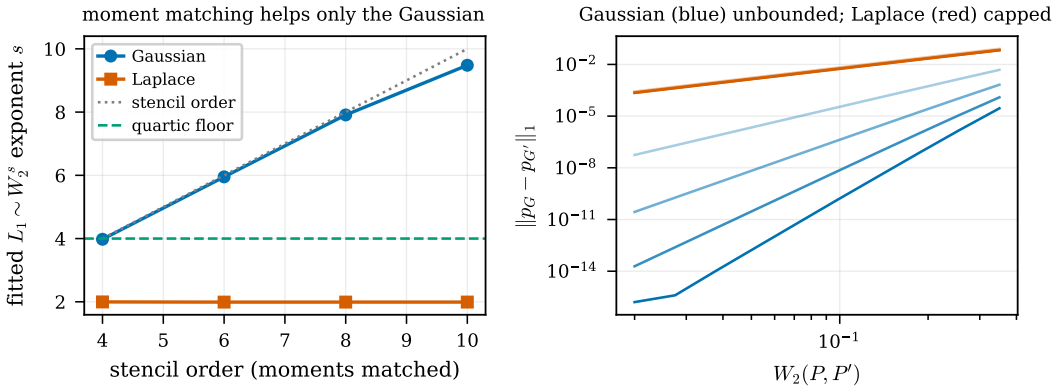

Figure 9: Left: fitted $L_1 \sim W_2^s$ exponent against stencil order. Matching more moments raises the Gaussian order across the tested range but leaves the Laplace order pinned near 2. Right: the underlying curves.

The likely reason is structural: the Laplace characteristic function $(1 + t^\top \Sigma t/2)^{-1}$ decays only polynomially, so moment matching – which controls low-frequency behaviour – does not control the bulk-frequency contribution that dominates for this kernel, unlike the Gaussian's exponential decay.

A second construction, at fixed support radius. The stencil above shrinks the whole configuration toward the origin, so the effective support radius shrinks with ϵ; since Theorem 1's constant depends on R, that is a valid but non-generic way to send $W_2 \to 0$. We therefore also tested fixed atom positions over a fixed radius, perturbing only the *weights* in a moment-matched *direction*. Every one of the six combinations tested gives $W_2 \sim \sqrt{\epsilon}$ and $L_1 \sim W_2^2$ *exactly*, with $R^2 = 1.0000$ and no dependence on order whatsoever. That null result is informative: matching moments of a fixed perturbation *vector* is a purely algebraic property of that vector and gives no reason for $\sum_j \delta_j f(x - \theta_j)$ to be small, unlike moment matching via position shifts, where it is exactly the condition making a Taylor expansion of the shift vanish to high order. Both constructions therefore rule *themselves* out as sharpness probes, for two different and identified reasons, without confirming or denying sharpness.

A third construction, exactly moment-stationary. Both families above match moments only to the order of the stencil. To remove that caveat we use a six-atom family whose mean and variance are stationary *exactly*, for every ϵ, with no truncation:

$$P_0 = \{\pm\tfrac{1}{2} : 0.4,\ \pm 2 : 0.1\}, \qquad P_\epsilon = \{\pm\tfrac{1}{2} : 0.4 + \tfrac{4\epsilon}{3},\ \pm 2 : 0.1 - \tfrac{7\epsilon}{3},\ \pm 3 : \epsilon\}.$$

Symmetry gives mean 0; the weights are chosen so that total mass and second moment are unchanged, $2(\frac{1}{4} \cdot \frac{4\epsilon}{3} - 4 \cdot \frac{7\epsilon}{3} + 9\epsilon) = 0$, and both identities are asserted numerically at every ϵ rather than assumed. All quantities are exact (1D quantile transport, closed-form densities). The measured slopes are $W_1 \sim \epsilon^{1.001}$, $W_2 \sim \epsilon^{0.501}$, $V \sim \epsilon^{1.000}$, so V/W_2^2 is stable at $0.491 \to 0.495$ across $\epsilon \in [0.0012, 0.04]$ while V/W_2^4 grows from 1.5 to 52.2 (Table 8, Figure 10). The family therefore sits at $V \asymp W_2^2$, a factor $\Theta(W_2^{-2})$ *above* the quartic floor of Theorem 1, and cannot reach it. Note that this is consistent with, not contrary to, Theorem 1: a lower bound $V \geq C_d W_2^4$ is not threatened by a family with $V \asymp W_2^2$.

Table 8: The third sharpness probe: mean and variance are stationary exactly. V/W_2^2 is stable while V/W_2^4 diverges, so the quartic floor is not approached.

ϵ	0.04	0.02	0.01	0.005	0.0025	0.0012
$W_1(P_0, P_\epsilon)$	0.2398	0.1199	0.0599	0.0299	0.0149	0.0072
$W_2(P_0, P_\epsilon)$	0.5648	0.3991	0.2820	0.1991	0.1407	0.0974
$V = \|p - p'\|_1$	0.1567	0.0783	0.0392	0.0196	0.0098	0.0047
V/W_2^2	0.491	0.492	0.493	0.494	0.495	0.495
V/W_2^4	1.5	3.1	6.2	12.5	25.0	52.2

E.5 Output scalings of the four perturbation families

Table 9: Fitted log–log slope of output W_2 against ϵ (exact 1D computation). Mixing-measure $W_2 = \Theta(\epsilon)$ throughout, so a slope above 1 means the output distance vanishes strictly faster than the mixing-measure distance.

Family	Mixing-measure motion	Gaussian	Laplace
translation	mean moves, order ϵ	1.000	1.000
small-mass transfer	mean moves, order ϵ^2	1.997	1.997
split/merge	mean fixed, variance moves	1.998	1.984
moment-matched	mean *and* variance fixed	3.990	2.524

These families are indexed by a perturbation scale ϵ and probe the *scaling* of each bridge; they are not condition families equipped with a metric $d_{\mathcal{C}}$, and no claim about Assumption 2 is made for them. For small-mass transfer, for instance, the mixing-measure W_2 is $\Theta(\epsilon)$ while the mean moves only at order ϵ^2, so no single choice of $d_{\mathcal{C}}$ makes both linear.

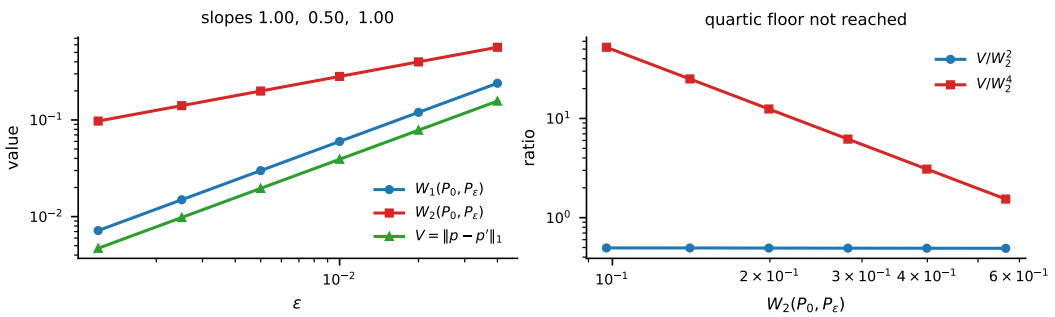

Figure 10: The exactly moment-stationary probe. Left: W_1, W_2 and V against ϵ with fitted slopes 1.00, 0.50, 1.00. Right: V/W_2^2 is flat while V/W_2^4 diverges, so the construction stays far above the quartic floor.

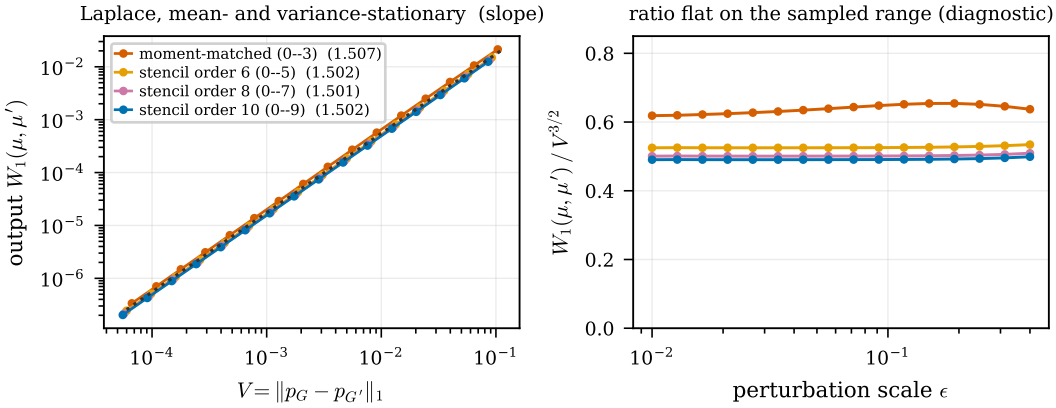

Figure 11: Proposition 3 on sampled mean-and-variance-stationary families. Left: fitted exponents sit near $3/2$. Right: the normalised ratio is flat on the sampled range; this is a diagnostic, not a proof of sharpness.

E.6 Proposition 3: mechanism check

Lemma 6 gives the required rate outright once $K_2 < \infty$, which Section C.5 establishes in closed form for the $d = 1$ Laplace kernel. As an independent check on that chain – and on how large Λ must be before the asymptotic regime is reached – we measured the rate the proof actually needs by explicit numerical Gaussian convolution of g at bandwidths $\sigma = 1/\Lambda$ over $\Lambda \in [1, 400]$, on a grid always at least $5\times$ finer than the smallest bandwidth (Figure 12 and Table 10).

Table 10: Mollification-error rate in the asymptotic regime.

Kernel	Family	Slope (large Λ)	R^2
Gaussian	all four base families	-2.000	1.0000
Laplace	translation	-1.984	1.0000
Laplace	split/merge	-1.989	1.0000
Laplace	moment-matched	-1.989	1.0000
Laplace	small-mass transfer	-1.973	0.99999

At small to moderate Λ the Laplace decay is measurably slower (global-fit slopes as weak as -1.20 to -1.55, worst for moment-matched where atoms cluster most tightly) – a genuine pre-asymptotic effect of the kinks, worth flagging for anyone choosing a finite Λ in practice, since finite-Λ bounds for this kernel are looser than the asymptotic rate suggests. In the asymptotic regime both kernels

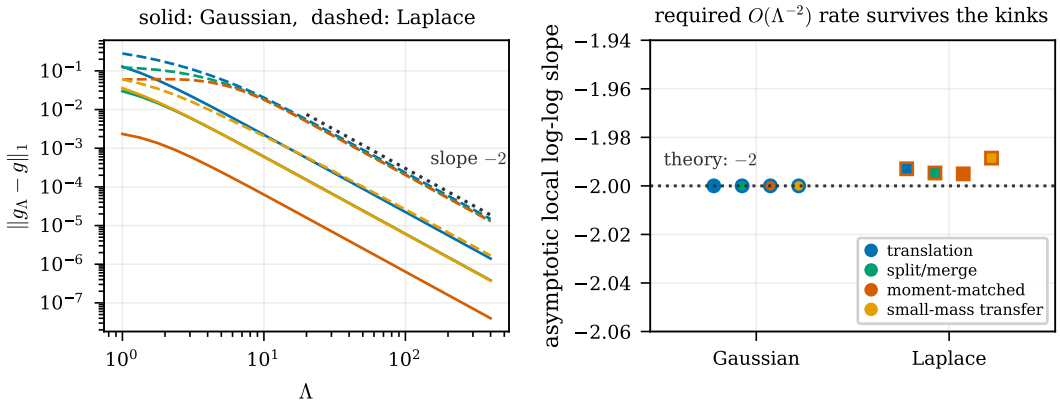

Figure 12: Left: mollification error against Λ; the dotted reference has slope -2. Right: asymptotic local slopes, both kernels, all four families.

converge cleanly to -2, consistent with the rate Lemma 6 provides. Note that the corrected proof does not actually rest on this numerical check: Lemma 6 gives the $O(\Lambda^{-2})$ rate outright whenever $K_2 < \infty$, which Section C.5 verifies for the $d = 1$ Laplace kernel in closed form. The measurement is a consistency check on the constant's magnitude and on the size of the pre-asymptotic regime, not the justification for the rate.

E.7 Proposition 3: sampled exponent diagnostic

Table 11 gives the per-family fits behind Figure 11. The ratio is not merely bounded away from zero but essentially constant over nearly two decades, which is consistent with $3/2$ on these families. We emphasise the scope: four one-dimensional constructions, in the regime where Proposition 3 applies to this kernel at all (Section C.5). It is not a proof of sharpness over mixing measures or dimensions.

Table 11: Proposition 3 for the Laplace kernel on mean- and variance-stationary families ($d = 1$, exact quadrature). Ratio range is $(\min, \max)$ of $W_1(\mu, \mu')/V^{3/2}$ over $\epsilon \in [0.01, 0.4]$.

Family (moments matched)	Fitted slope	Ratio range	Stable?
moment-matched (0–3)	1.507	(0.619, 0.654)	yes
stencil, order 6 (0–5)	1.502	(0.525, 0.534)	yes
stencil, order 8 (0–7)	1.501	(0.501, 0.509)	yes
stencil, order 10 (0–9)	1.502	(0.490, 0.499)	yes

With $K_2 \le 4$ for the standardised $d = 1$ Laplace kernel (Section C.5) and a Gaussian mollifier, Proposition 3's closed form gives $C_{\text{sign}} = \frac{2}{3} M_1(\psi)^{-1}(3 m_2(\psi) K_2)^{-1/2} \approx 0.241$ (using $m_2 = 1$, $M_1 = \sqrt{2/\pi}$, $K_2 \le 4$), comfortably below the fitted values 0.49–0.65: the proved constant is valid but conservative by a factor of about 2 to 2.7 on these families, as one expects from the chain of inequalities in Section C.4. That the proved and measured constants sit within a small factor of one another – rather than differing by orders of magnitude – is a useful consistency check for these examples. It does not quantify the worst-case loss in the proof.

E.8 The heat-bridge chain in $d = 2, 4, 8$

Proposition 4 is proved through four displayed inequalities, and each is a separately measurable quantity. We check all four rather than only the conclusion, because a conclusion that holds while an intermediate step fails would indicate a proof error that a single end-to-end number would hide.

Exact evaluation, $d = 2$**.** P, P' are five- and four-atom measures in $B(0, 1.5)$ with an anisotropic diagonal Σ; the Laplace density is evaluated in the Bessel closed form $f \propto (q/2)^{\nu/2} \mathcal{K}_\nu(\sqrt{2q})$,

$\nu = (2-d)/2$, and the heat flow $H_t g$ for $t > 0$ by Gauss–Laguerre quadrature against the $\mathrm{Exp}(1)$ variance mixture, which is smooth for $t > 0$. Integrals are taken on a fixed grid. The four checks are $\|L_\Sigma g\|_{\mathrm{TV}} = 6.05 \le 8$; $\|H_t g - g\|_1 = 0.171, 0.270, 0.401, 0.574$ against $8t = 0.4, 0.8, 1.6, 3.2$; $\mathrm{Lip}(h_t) = 1.442 \le (\pi\lambda_{\min} t)^{-1/2} = 1.591$ at $t = 0.2$; and $\int \mathrm{sign}(g) H_{t^*} g = 0.996 \ge V - 8t^* = 0.697$ at $t^* = V/24$, giving a witness value 0.292 against the predicted $c_{\mathrm{heat}} V^{3/2} = 0.204$.

Importance sampling, $d = 4, 8$. The integrands remain exact; only the integrals are estimated, with $N = 1.2 \times 10^5$ draws from a product Laplace proposal with scale 2. The proposal must have heavier tails than the integrand: a Gaussian proposal makes the importance weights diverge, which we observed before switching. Results: at $d = 4$ the four semigroup errors are $0.25, 0.43, 0.67, 0.98$ against $0.4, 0.8, 1.6, 3.2$, the pairing is 1.415 ± 0.010 against 0.986, and the witness 0.600 against 0.418; at $d = 8$ the errors are $0.14, 0.29, 0.58, 0.96$, the pairing 1.456 ± 0.066 against 0.890, and the witness 0.459 against 0.280. Uncertainties are one standard error; the pass criteria use three. Because the witness $h_{t^*}/\mathrm{Lip}(h_{t^*})$ is 1-Lipschitz by construction, its pairing against $\mu - \mu'$ is a valid lower bound on $W_1(\mu, \mu')$ up to the Monte Carlo error, so this is a certified high-dimensional value and not only a step check. Two limitations remain: $\widehat{V}$ appears on both sides of the last inequality through t^*, so those errors are correlated, and the estimator is unbiased for the integrals but not for the ratio.

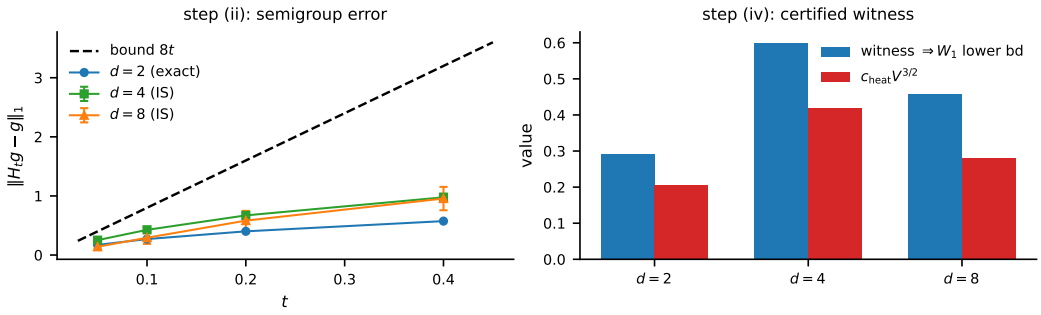

Figure 13: The heat-bridge chain at $d = 2$ (exact grid) and $d = 4, 8$ (importance sampling, exact integrands, 3-SE bars). Left: step (ii) against its bound $8t$. Right: the certified witness against the predicted $c_{\mathrm{heat}} V^{3/2}$ of Proposition 4.

E.9 Bures versus trace certificates

For centred Gaussians with $\Sigma_\mu = \mathrm{diag}(\sigma^2, 0) + \varepsilon^2 \mathrm{Id}$ and $\Sigma_{\mu'} = \mathrm{diag}(0, \sigma^2) + \varepsilon^2 \mathrm{Id}$ at $\sigma = 1$, the two matrices commute, so $\mathfrak{B}^2 = 2(\sigma^2 + 2\varepsilon^2) - 4\varepsilon\sqrt{\sigma^2 + \varepsilon^2}$ in closed form. The measured sweep is $\varepsilon = 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001 \mapsto \mathfrak{B} = 0.586, 1.052, 1.280, 1.372, 1.400, 1.410, 1.413$, against the trace certificate, which is 0 at every ε because the traces agree identically. At $\varepsilon = 10^{-3}$ the Bures certificate matches the true W_2 to all reported digits ($\mathfrak{B}^2 = 1.9960 = W_2^2$), as it must for Gaussians. The wide kernel $\Sigma_\nu = 0.5\,\mathrm{Id}_2$ ($\varepsilon^2 = 0.5$, a point of the same family) gives $\mathfrak{B} = 0.732$ and trace 0. This is a two-parameter family, not a sweep over model fits; its role is to exhibit the gap, which is qualitative (zero versus attained), not to estimate a typical improvement.

E.10 Mollifier constants against the proved floor

M_ψ and M'_ψ are computed in closed form: for the Gaussian, $M_\psi = \sqrt{2}\,\Gamma(\frac{d+1}{2})/\Gamma(\frac{d}{2})$ and $M'_\psi = \sqrt{2/\pi}$; for the unit-ball family $\psi \propto (1 - \|x\|^2)^4$, both constants reduce to ratios of the polynomial moments $J(k) = \int_0^1 (1-r^2)^4 r^k \, dr$, with the angular factor $E|\omega_1| = \Gamma(\frac{d}{2})/(\sqrt{\pi}\,\Gamma(\frac{d+1}{2}))$ evaluated exactly rather than by its large-d surrogate $\sqrt{2/(\pi d)}$ – the surrogate is 0.798 instead of 1 at $d = 1$ and pushes that entry spuriously below the floor of Lemma 4. For the product family $\psi = \prod_i \rho(x_i)$ with $\rho(x) = \frac{15}{16}(1 - x^2)^2$, $M'_\psi = \int |\rho'| = \frac{15}{8}$ exactly and independently of d, while $M_\psi = E\|X\|$ is sampled from $\mathrm{Beta}(3, 3)$ with 4×10^5 draws. Table 12 reports the products.

Table 12: Mollifier constants, computed in closed form (the product family's M_ψ is sampled). The last row is the floor of Lemma 4; no entry falls below it, and no family has both constants $O(1)$.

$M_\psi M'_\psi$	d=1	2	4	8	16	32	64
Gaussian	0.64	1.00	1.50	2.19	3.14	4.48	6.36
product, compact	0.59	0.91	1.35	1.96	2.80	3.99	5.66
unit-ball radial	0.61	0.96	1.44	2.12	3.08	4.42	6.31
floor $\sqrt{d}/2$	0.50	0.71	1.00	1.41	2.00	2.83	4.00

E.11 Certifying a trained mixture density network

The network has one hidden layer of width 64 and outputs $2K + 1$ parameters for $K = 5$ Laplace components – K locations, K logit weights and one shared log-scale – which is the shared-scale model class the theory assumes (per-component scales, i.e. $3K$ outputs, would leave that class). Training is by maximum likelihood, 4,000 samples, Adam. Task A draws $x \mid c \sim \mathrm{Laplace}(3c - 1.5, \sqrt{0.5})$; task B draws $x \mid c \sim P_c * \mathrm{Laplace}(1)$ with P_c the exactly moment-stationary family of Section E.4 at $\epsilon = c$. Certificates and output W_2 are computed from the *learned* parameters by exact 1D quantile integration on an adaptive span $\pm(\max_j |\theta_j| + 14 b_{\max})$ with 1.2×10^5 grid points; a fixed span of ± 8 truncated the tails and produced an apparent 2% certificate violation, which the adaptive span removed. Because the two conditionals may have different learned scales, the W_2 computation uses the two-scale monotone rearrangement rather than a shared-scale shortcut. Fitted log–log slopes against $d_{\mathcal{C}}$ are 0.962 (ρ) against 0.969 (W_2) on task A, and 1.0019 against 1.0019 on task B; the equality on task B holds because the learned conditionals form a location–scale family in c, for which the moment certificate is exact.

E.12 The inconclusive higher-dimensional numerical attempt

$d = 1$ exact quadrature does not scale, so we attempted a Monte Carlo extension: adaptive-doubling L^1 and a debiased-Sinkhorn W_1 estimate (linear, not squared, cost – Proposition 3 bounds W_1). The first attempt calibrated against the exact $d = 1$ answer at small ϵ (output $W_1 \sim 5.7 \times 10^{-4}$) with 4,000–8,000 samples and gave relative errors of 1200%–6700%: the empirical W_1 between two independent size-n samples is already $O(n^{-1/2})$, which swamps a signal that small, and growing n hits an $O(n^2)$ memory wall in the Sinkhorn cost matrix well before it would help. Restricting to larger ϵ (calibrated to 4.8–9.4% error at $n = 12{,}000$) fixed calibration but not tractability: a reduced 8-cell grid ($d \in \{1, 2, 4, 8\}$, $\epsilon \in \{0.35, 0.5\}$, 3 seeds) cost 5–10 GPU-minutes per cell, and the one cell that completed still failed its own 25% relative-CI gate ($\mathrm{relCI} = 0.739$). Its point estimate (0.68) sits close to the exact $d = 1$ range (0.62–0.65), which is encouraging but, at that confidence width, is weak non-conclusive corroboration. We stopped rather than spend substantially more compute on a result that would likely need many more seeds, a larger point cloud than an $O(n^2)$ cost matrix allows on a 16GB card, or a fundamentally different non-empirical-OT estimator. This numerical failure does not leave the fixed-Laplace bridge open: Proposition 4 proves it analytically in every dimension, and Section E.8 verifies every step of that proof numerically at $d = 2, 4, 8$ by evaluating the proof's own witness instead of estimating a transport plan. The lesson we draw is that the obstruction was the estimator, not the dimension: any check routed through empirical optimal transport inherits an $n^{-1/d}$ noise floor, whereas the dual witness the proof constructs is available in closed form.

E.13 The coarse-geometry negative result

Figure 14 records the outcome that motivated the bridge lemmas in the first place. Estimating output W_2 directly by empirical optimal transport at $d = 8$ is defeated by the curse of dimensionality: the measured noise floor between two independent draws of the *same* distribution decays only like $n^{-1/8}$, and every one of the 1,225 condition pairs sits at or below it, even with common random numbers, a genuine independent-draw floor and four times the samples. The same finite-sample optimal-transport floor also blocks the Monte Carlo attempt of the preceding subsection, which

is why the bridges of Section 3.3 are proved analytically rather than verified numerically in high dimension.

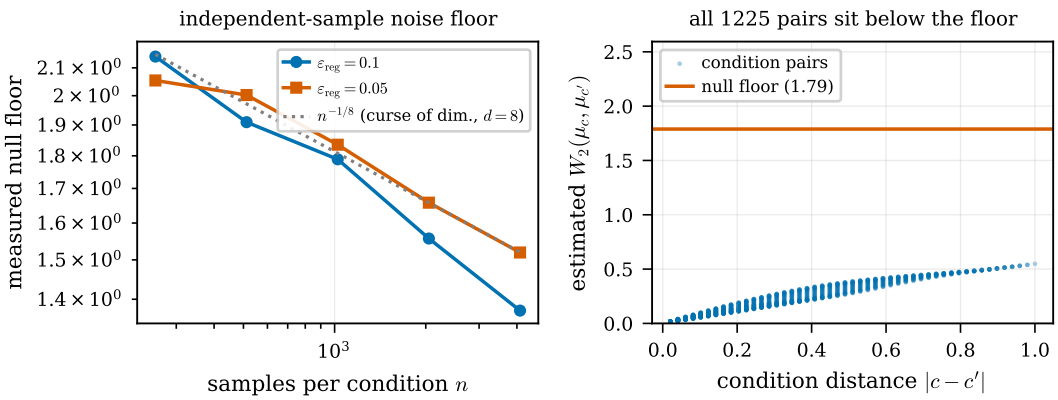

Figure 14: The negative result behind Section 3.3. Left: the independent-sample noise floor falls only at the rate the curse of dimensionality predicts for $d = 8$. Right: all 1,225 condition pairs lie below that floor, so no signal is recoverable at this budget.