Sharp Wasserstein Stability for Gaussian Deconvolution of Log-Concave Laws

Scott Schubr

October 23, 2026

Abstract

Gaussian smoothing can conceal large differences between latent probability laws. We determine the sharp inverse modulus in quadratic Wasserstein distance around every reference whose twice continuously differentiable potential has positive lower and finite upper Hessian bounds. In every fixed dimension and at every fixed positive scalar Gaussian noise level, the modulus over all log-concave inputs has order $(\log(1/\varepsilon))^{-3/2}$, for either output W_1 discrepancy or excess risk of the reference Bayes estimator. Inputs may be nonsmooth or compactly supported, and their moments need not be prescribed. The upper proof combines log-concave curvature control at the observable resolution with a density-weighted transport field and an exterior flux correction. Matching oscillatory alternatives preserve the reference's mass, mean and covariance exactly. Uniform upper constants give fixed-reference testing on the stated isotropic class and minimax estimation with unknown moments on two-sided potential-curvature classes. The estimation rate is $(\log n)^{-3/2}$ in expected W_2 loss, also for known Gaussian noise covariance of positive rank; mean anchoring controls exceptional samples. The Gaussian-reference specialization gives the corresponding sharp modulus from the linear-MMSE gap.

Keywords. Gaussian deconvolution; log-concavity; Wasserstein distance; inverse stability; minimax estimation.
2020 Mathematics Subject Classification. 62G07, 62G20, 60E15.

1 Introduction

Gaussian smoothing can make different latent distributions difficult to distinguish. We ask how accurately a log-concave law can be recovered in quadratic Wasserstein distance around a reference with a twice continuously differentiable potential and two-sided positive curvature bounds. The unknown input itself may be nonsmooth or compactly supported.

Both output Wasserstein discrepancy and excess risk of the reference Bayes estimator give the sharp inverse modulus of order $(\log(1/\varepsilon))^{-3/2}$. Matching alternatives preserve the reference's actual mean and covariance. Uniform upper constants yield fixed-reference testing on the stated isotropic class and minimax estimation with unknown means and covariances under known Gaussian noise of positive rank. The Gaussian linear-MMSE gap is a specialization of the regret observation.

Section 2 states the results. We prove the normalized inverse first, construct matching alternatives, and then remove normalization before giving the statistical proofs.

Relation to previous work

Logarithmic deconvolution is established territory. Nguyen [1, Theorem 2] already bounds latent W_2 by discrepancies of Gaussian mixture densities, with inverse-logarithmic power

1/2 for W_2 in the Gaussian case. His proof uses bandlimited mollification and a first-order smoothing transport bound. Dedecker–Michel [2] prove minimax Wasserstein deconvolution rates under moment hypotheses. Scricciolo [3, Theorem 2] has the power 3/2 in a smoothness-two statistical W_1 problem with Gaussian errors. Our one-dimensional normalized curvature class $\mathcal{R}_{c,C}$, defined below, satisfies that weighted smoothness-two assumption: ρ'' and $(x\rho)''$ have uniformly bounded L^2 norms by the Gaussian envelopes and, writing $U = -\log\rho$, the identity $\rho'' = ((U')^2 - U'')\rho$. The different loss remains important: a W_1 upper rate does not supply the W_2 upper here. Bobkov–Chistyakov–Götze [4] prove logarithmic information-distance stability for Gaussian-regularized summands in Cramér's theorem. In the normalized analytic core, the unknown input is any isotropic log-concave density, including nonsmooth and compactly supported cases.

The shape-constrained pointwise deconvolution principle has close ancestry in Eldan–Klartag [5, Section 3]. Weighted negative-Sobolev transport comparisons are classical [6]; the significance of density lower bounds for smooth Wasserstein estimates is discussed in [7]. The present proof extracts a cubic local field estimate from log-concavity and the small observation, retains the density-dependent weight, and corrects the exterior flux. The exact lower family works at every reference in (1).

Park and Kim [8] study univariate log-concave deconvolution under squared L^2 density loss. Their presentation considers a variance floor and, for the sharper class, a bound on the total variation of the density derivative. For supersmooth noise of order β, the latter class has squared-density risk of order $(\log n)^{-3/\beta}$; Gaussian noise has $\beta = 2$. This is a density-risk statement, rather than a latent Wasserstein inverse modulus. Our results concern deterministic W_2 stability in every fixed dimension, observations through output W_1 or reference-estimator regret, and exact moment-preserving lower families at every reference satisfying (1).

The proof mechanism

Gaussian deconvolution resolves frequencies up to $R \asymp \sqrt{\log(1/\varepsilon)}$. The key step is a transport field of size R^{-3} in the density-weighted energy norm. Positive bandlimited smoothing first makes the unknown density comparable to the reference on a growing ball. Log-concavity then bounds the trace curvature on cells of radius $h = R^{-1}$, including smaller subcells. Applying a high-frequency inverse-divergence kernel to this curvature measure gives a field of size h^3. A weighted divergence correction controls its exterior without imposing Gaussian tails on every input. This is the analytic step behind the exponent 3/2. For sharpness, an oscillation of frequency T and amplitude T^{-2} preserves convexity, has transport separation of order T^{-3}, and becomes exponentially small after Gaussian convolution. Thus the transport scale is $R^{-3} \asymp (\log(1/\varepsilon))^{-3/2}$; the exact moment corrections preserve this order.

2 Model and main results

Let γ_d be the standard Gaussian measure on $\mathbb{R}^d$, and let ϕ_σ denote the density of $N(0, \sigma^2 I_d)$. Absolute values denote Euclidean norms or their induced operator norms. We use

$$W_2(\mu,\nu)^2 = \inf\{\mathbb{E}|A-B|^2 :\ \mathcal{L}(A) = \mu,\ \mathcal{L}(B) = \nu\}.$$

Write

$$\mathcal{C}_d = \{f :\ f \text{ is a log-concave probability density on } \mathbb{R}^d,\ \mathbb{E}_f X = 0,\ \mathbb{E}_f XX^\top = I_d\}.$$

Log-concavity allows an extended-real convex potential outside the positive support. We take the upper semicontinuous representative.

Throughout the main theorem fix $d \geq 1$, $\sigma > 0$, and $\varrho \in \mathcal{C}_d$ of the form

$$\varrho = Z^{-1}e^{-U}, \qquad U \in C^2(\mathbb{R}^d), \qquad cI_d \leq D^2U \leq CI_d, \qquad 0 < c \leq C < \infty. \tag{1}$$

No higher derivative of U is assumed. For $f \in \mathcal{C}_d$, observe $Y = X + \sigma G$, with $G \sim \gamma_d$ independent, and put

$$p_f = f * \phi_\sigma, \qquad q = \varrho * \phi_\sigma, \qquad m_f(y) = \mathbb{E}_f[X \mid Y = y].$$

Let m_ϱ be the reference posterior mean. Define the two scalar observations

$$\begin{aligned} \mathcal{O}_W(f) &= W_1(p_f\,dy, q\,dy), \\ \mathcal{O}_R(f) &= \mathrm{Reg}_\varrho(f) := \mathbb{E}_f|X - m_\varrho(Y)|^2 - \mathbb{E}_f|X - m_f(Y)|^2. \end{aligned} \tag{2}$$

The second is the excess quadratic risk of the fixed reference estimator. It is nonnegative by conditional-expectation orthogonality. For $a \in \{W, R\}$ set

$$M_a(\varepsilon) = \sup_{\substack{f \in \mathcal{C}_d \\ \mathcal{O}_a(f) \leq \varepsilon}} W_2(f\,dx, \varrho\,dx).$$

Theorem 2.1 (Normalized analytic core)**.** *Under* (1)*, there are $c_*, C_*, \varepsilon_* > 0$ such that, for $a = W, R$ and $0 < \varepsilon \leq \varepsilon_* < 1$,*

$$c_*[\log(1/\varepsilon)]^{-3/2} \leq M_a(\varepsilon) \leq C_*[\log(1/\varepsilon)]^{-3/2}. \tag{3}$$

Constants may depend on d, σ, ϱ. For a sufficiently small positive $\mathcal{O}_a(f)$, in particular,

$$W_2(f\,dx, \varrho\,dx) \leq C_*[\log(1/\mathcal{O}_a(f))]^{-3/2}.$$

If either observation is zero, then $f = \varrho$ almost everywhere. The lower bound is attained in order by a family of C^2 inputs with potential Hessians uniformly between $cI_d/2$ and $(C + c/2)I_d$.

The normalized theorem is the analytic core. The same moduli hold without prescribing the input moments or normalizing the reference; the observation controls the moments needed for the reduction.

Corollary 2.2 (Unrestricted log-concave inputs)**.** *Fix $d \geq 1$, $\sigma > 0$, and a probability density $g = Z^{-1}e^{-V}$ on $\mathbb{R}^d$ with $V \in C^2$ and $cI \leq D^2V \leq CI$, where $0 < c \leq C < \infty$. Its mean and covariance may be arbitrary. Let $\mathcal{L}_d$ be all log-concave probability densities on $\mathbb{R}^d$, and for $f \in \mathcal{L}_d$ define*

$$\mathcal{O}_W^g(f) = W_1(f * \phi_\sigma, g * \phi_\sigma), \qquad \mathcal{O}_R^g(f) = \mathrm{Reg}_g(f),$$

where the regret is the reference-estimator excess risk in (2) *with reference g. For $a = W, R$, there are positive constants k, K, ε_0 such that*

$$k[\log(1/\varepsilon)]^{-3/2} \leq \sup_{\substack{f \in \mathcal{L}_d \\ \mathcal{O}_a^g(f) \leq \varepsilon}} W_2(f, g) \leq K[\log(1/\varepsilon)]^{-3/2} \quad (0 < \varepsilon \leq \varepsilon_0).$$

Zero observation implies $f = g$. The upper constants may be chosen using only d, σ, c, C. Matching lower families can preserve the reference's actual mean and covariance exactly.

The proof is in Section 10.

Corollary 2.3 (Gaussian linear-MMSE gap). *Fix $d \geq 1$ and $s > 0$. For $f_X \in \mathcal{C}_d$ let*

$$\mathrm{mmse}_X(s) = \mathbb{E}|X - \mathbb{E}[X \mid \sqrt{s}X + G]|^2, \qquad \Delta_s(X) = \frac{d}{1+s} - \mathrm{mmse}_X(s).$$

There are positive constants $c_{d,s}, C_{d,s}, \varepsilon_{d,s}$ such that, for $0 < \varepsilon \leq \varepsilon_{d,s}$,

$$c_{d,s}[\log(1/\varepsilon)]^{-3/2} \leq \sup_{\substack{f_X \in \mathcal{C}_d \\ \Delta_s(X) \leq \varepsilon}} W_2(\mathcal{L}(X), \gamma_d) \leq C_{d,s}[\log(1/\varepsilon)]^{-3/2}.$$

At zero gap X is standard Gaussian.

Proof. Take $\varrho = \phi_1$ and $\sigma = s^{-1/2}$ in Theorem 2.1. The reference posterior mean is $y/(1+\sigma^2)$, whose risk under any isotropic input is $d\sigma^2/(1+\sigma^2) = d/(1+s)$. Rescaling the observation does not change the conditional mean. Thus $\mathrm{Reg}_{\phi_1}(f_X) = \Delta_s(X)$. □

2.1 Uniformity, testing and estimation

For fixed $0 < c \leq C < \infty$ write

$$\mathcal{R}_{c,C} = \{\rho \in \mathcal{C}_d : \ -\log \rho \in C^2(\mathbb{R}^d),\ cI_d \leq D^2(-\log \rho) \leq CI_d\}.$$

For the fixed-reference test, observations $Y_1, \ldots, Y_n$ are independent with density $p_f = f * \phi_\sigma$; d and $\sigma > 0$ are fixed and known. Write $P_f^{\otimes n}$ for their joint law and $Q_n = n^{-1}\sum_i \delta_{Y_i}$.

Lemma 10.1 supplies constants depending only on the fixed dimension, noise level and curvature interval. The following conclusions have different parameter classes: testing permits every input in $\mathcal{C}_d$, while estimation permits arbitrary means and covariances under two-sided potential-curvature bounds.

Theorem 2.4 (Fixed-reference testing). *Fix a known reference ρ satisfying (1). Define*

$$\mathfrak{T}_n(r) = \inf_{\psi} \left[P_\rho^{\otimes n}(\psi = 1) + \sup_{\substack{f \in \mathcal{C}_d \\ W_2(f,\rho) \geq r}} P_f^{\otimes n}(\psi = 0) \right],$$

where $\psi = 1$ means rejection of the reference. There are constants $0 < a_\rho < A_\rho < \infty$ such that

$$\mathfrak{T}_n(A_\rho(\log n)^{-3/2}) \longrightarrow 0, \qquad \mathfrak{T}_n(a_\rho(\log n)^{-3/2}) \longrightarrow 1.$$

For $0 < c < C < \infty$, define the unnormalized curvature class

$$\mathcal{F}_{c,C} = \{f = Z^{-1}e^{-V} : V \in C^2(\mathbb{R}^d),\ cI_d \leq D^2V \leq CI_d\}.$$

The mean and covariance are unknown and unrestricted within this class. For a positive semidefinite matrix Σ, let γ_Σ be the centered Gaussian law of covariance Σ.

Theorem 2.5 (Estimation with unknown moments). *Fix $0 < c < C < \infty$ and a known positive semidefinite covariance Σ of positive rank. From n independent observations with law $f * \gamma_\Sigma$, for $r = 1, 2$ and all sufficiently large n,*

$$a_r(\log n)^{-3r/2} \leq \inf_{\widehat{\mu}_n} \sup_{f \in \mathcal{F}_{c,C}} \mathbb{E}_f W_2(\widehat{\mu}_n, f\,dx)^r \leq A_r(\log n)^{-3r/2},$$

where $0 < a_r \leq A_r < \infty$ depend only on d, c, C, Σ. The infimum is over measurable probability-valued estimators. An upper estimator can take values in $\mathcal{F}_{c,C}$ and have mean exactly equal to the sample mean of the observations. For scalar noise $\Sigma = \sigma^2 I_d$ and $0 < c < 1 < C$, the same order also holds on the normalized subclass $\mathcal{R}_{c,C}$.

Gaussian MMSE extremality and its equality case are classical [9]; the quantitative latent transport modulus is our concern.

All constants below are uniform in f. We use C_* for constants allowed to depend on the fixed d, σ, ϱ and kernels; their values can change. The notation C_d records purely dimensional bounds; the remaining constants may depend on dimension. Our Fourier convention is $\widehat{g}(\xi) = \int e^{-ix\cdot\xi} g(x)\, dx$.

3 Classical posterior-score identities

For smooth positive output densities write

$$J(p \mid q) = \int_{\mathbb{R}^d} |\nabla \log p - \nabla \log q|^2 p.$$

The following classical identities connect reference-estimator risk to output scores. Tweedie's formula is discussed in [10], which credits Tweedie through Robbins's 1956 work. The score-relative-Fisher form of regret is explicit, for example, in [11, eq. (2)].

Lemma 3.1. *For the model above,*

$$\mathrm{Reg}_\varrho(f) = \mathbb{E}_f |m_f(Y) - m_\varrho(Y)|^2 = \sigma^4 J(p_f \mid q). \tag{4}$$

If $\kappa = c/(1 + c\sigma^2)$, *then*

$$\kappa I_d \le D^2(-\log q) \le \sigma^{-2} I_d, \qquad W_1(p_f, q) \le \kappa^{-1} \sqrt{J(p_f \mid q)}. \tag{5}$$

Proof. Gaussian convolution gives $m_f(y) = y + \sigma^2 \nabla \log p_f(y)$, and similarly for the reference. Projection orthogonality proves the first equality in (4); subtracting the two score formulas proves the second. The reference posterior potential has Hessian at least $(c + \sigma^{-2}) I_d$. The Brascamp–Lieb inequality [12] therefore bounds its covariance by $(c + \sigma^{-2})^{-1} I_d$. The differentiated Gaussian-convolution identity

$$D^2(-\log q)(y) = \sigma^{-2} I_d - \sigma^{-4} \operatorname{Cov}_\varrho(X \mid Y = y)$$

proves the Hessian bounds. In particular the reference score has linear growth. The output score of p_f has squared integral at most d/σ^2 by conditional Jensen applied to the Gaussian noise score. All quantities in (4) are thus finite.

The curvature logarithmic Sobolev inequality gives $2\kappa D(p_f \mid q) \le J(p_f \mid q)$, and the corresponding transport inequality gives $\kappa W_2(p_f, q)^2 \le 2D(p_f \mid q)$ [13, 14]; their Gaussian forms are [15, 16]. Combining and using $W_1 \le W_2$ proves (5). □

For the Gaussian reference, scaling the output in Lemma 3.1 gives the familiar exact normalization

$$J\left(\mathcal{L}\left(\frac{\sqrt{s} X + G}{\sqrt{1+s}} \right) \middle| \gamma_d \right) = s(1+s)\Delta_s(X).$$

If $\mathcal{O}_W = 0$, the outputs agree; if $\mathcal{O}_R = 0$, (4) makes their scores agree, so again $p_f = q$. The nonvanishing Gaussian Fourier multiplier then gives $f = \varrho$. It remains to prove both positive-observation upper bounds and the matching lower family.

4 A weighted flux bound and the reference envelopes

We prove the upper estimate from $\varepsilon = \mathcal{O}_W(f) > 0$. The regret upper estimate will then follow from Lemma 3.1. All dimensions $d \geq 1$ are included. Write $B_r(x)$ for a Euclidean ball and $B_r = B_r(0)$.

Lemma 4.1 (Midpoint flux bound)**.** *Let f, g be probability densities with finite second moments, and let a vector field A satisfy $\operatorname{div} A = f - g$ distributionally. If $f + g > 0$ almost everywhere and the following integral is finite, then*

$$W_2(f\,dx, g\,dx)^2 \leq 8 \int_{\mathbb{R}^d} \frac{|A|^2}{f+g}\,dx. \tag{6}$$

Proof. Put $m = (f+g)/2$. The path from m to f given by

$$\rho_t = (1-t)^2 m + (2t - t^2) f, \qquad j_t = -(1-t)A$$

satisfies $\partial_t \rho_t + \operatorname{div} j_t = 0$. Since $\rho_t \geq (1-t)^2 m$, its dynamic transport action [17] is at most $\int |A|^2/m$. The same holds for the path from m to g, with the opposite flux. The triangle inequality gives $W_2(f,g)^2 \leq 4 \int |A|^2/m$, which is (6). The usual approximation and truncation extend the dynamic inequality to finite-action measurable fields and densities. The quadratic time parameter permits zeros in an endpoint density. This is a weighted negative-Sobolev comparison of the type in [6]. □

For isotropic log-concave inputs we use the classical bounds

$$\|f\|_\infty \leq C_d, \qquad C_P(f) \leq Cd. \tag{7}$$

Here C_P is the constant in $\operatorname{Var}_f v \leq C_P(f) \int |\nabla v|^2 f$. One explicit density bound is [5, eq. (59)], $\|f\|_\infty \leq \exp(\frac{1}{2} d \log d + 6d)$. The Poincaré bound is the classical Kannan–Lovász–Simonovits/Bobkov estimate: [18, Theorem 1.2 and eq. (1.8)] gives $C_P(f) \leq 4K_0^2 \mathbb{E}|X - \mathbb{E}X|^2$ for a universal K_0. It allows nonsmooth potentials and support boundaries.

The reference hypotheses give fixed $\beta > C$, $a = c/4$ and positive constants such that

$$b_0 e^{-\beta|x|^2/2} \leq \varrho(x) \leq b_1 e^{-a|x|^2}, \tag{8}$$

$$|\nabla U(x)| \leq C_*(1+|x|), \qquad |\Delta \varrho(x)| \leq C_*(1+|x|^2)\varrho(x), \tag{9}$$

$$\frac{\varrho(x-z)}{\varrho(x)} \leq \exp\{C_*(1+|x|)|z| + C|z|^2/2\}. \tag{10}$$

These follow by Taylor's theorem, completing the square to absorb $U(0), \nabla U(0)$, and $\Delta\varrho = (|\nabla U|^2 - \Delta U)\varrho$. The reference also has $C_P(\varrho) \leq 1/c$ by [12]. No derivative of U beyond order two enters these estimates.

Fix a real radial $\chi \in C_c^\infty(\mathbb{R}^d)$ equal to one near zero. Fix also a nonnegative radial Schwartz kernel κ of mass one, with smooth compactly supported Fourier transform and $\kappa \geq b_2 > 0$ on B_b for some $b > 0$. Such a kernel is obtained by squaring and normalizing a real radial inverse smooth cutoff nonzero at zero. Choose $B > 0$ so that the frequency supports of χ and $\widehat{\kappa}$ lie in the ball B_B, and put $\kappa_h(x) = h^{-d}\kappa(x/h)$.

Set

$$R^2 = \frac{\log(1/\varepsilon)}{8\sigma^2 B^2}, \qquad h = R^{-1}, \qquad L = (\Lambda \log R)^{1/2}, \qquad \Lambda = 128/c. \tag{11}$$

Choose once an integer $N_0 > \beta\Lambda + d + 100$. We take ε sufficiently small that R exceeds all fixed thresholds below, uniformly over f.

Lemma 4.2. *For multiindices $|\alpha| \le 2$,*

$$\|(f-\varrho) * \partial^\alpha \kappa_h\|_\infty \le R^{-N_0} \tag{12}$$

for all sufficiently large R. If $g = f - \varrho$ and $\widehat{A}_0(\xi) = -i\xi|\xi|^{-2}\chi(h\xi)\widehat{g}(\xi)$, then

$$\|A_0\|_2^2 \le C_d \varepsilon^2 R^d \exp(\sigma^2 B^2 R^2). \tag{13}$$

Proof. The characteristic-function coupling inequality and Gaussian deconvolution give

$$|\widehat{f}(\xi) - \widehat{\varrho}(\xi)| \le \varepsilon|\xi| e^{\sigma^2|\xi|^2/2}. \tag{14}$$

Fourier inversion bounds the left side of (12) by $C_d \varepsilon R^{d+|\alpha|+1} e^{\sigma^2 B^2 R^2/2}$. Its exponential factor with ε is $\varepsilon^{15/16}$, which beats any prescribed inverse power of R. Plancherel gives (13). These bounds do not assume that derivatives of f exist as functions. □

5 Pointwise reference bounds on a growing core

This step is a positive-kernel version of the shape-constrained deconvolution principle in [5, Section 3].

Lemma 5.1. *For all sufficiently large R, the positive support contains B_{L+9}, and*

$$c_*\varrho(x) \le f(x) \le C_*\varrho(x) \quad (|x| \le L+6), \tag{15}$$

$$|p - \nabla U(x)| \le C_* \quad (p \in \partial V(x)) \quad (|x| \le L+5), \tag{16}$$

where $V = -\log f$ in the interior of its positive support.

Proof. For $|x| \le L+10$ and every fixed N, (10) and Schwartz decay give

$$c_*\varrho(x) \le \varrho * \kappa_h(x) \le C_*\varrho(x) + C_{*,N}R^{-N}, \tag{17}$$

$$((1+|\cdot|^2)\varrho) * \kappa_h(x) \le C_*(1+|x|^2)\varrho(x) + C_{*,N}R^{-N}. \tag{18}$$

For the lower bound integrate $|z| \le bh$. For the upper, split at $|z| = R^{-1/2}$; the reference ratio is bounded on the inner part and the polynomial reference density is bounded globally for the outer part. The minimum reference density on this core is at least R^{-M_*} eventually, for some fixed $M_* < \beta\Lambda/2 + 10$, by (8).

Fix $|x| \le L+9$ and suppose $f(x) < a_0\varrho(x)$ for a sufficiently small fixed $a_0 > 0$. If the closed convex level set $D = \{f \ge a_0\varrho(x)\}$ is nonempty, separate it from x and let n be the outward unit normal. For $x_1 = x + R^{-1/2}n$, the ball $B_{R^{-1/2}/2}(x_1)$ lies outside D. Thus $f * \kappa_h(x_1) \le a_0\varrho(x) + C_{*,N}R^{-N}$ by (7). Equations (12), (17) give a lower bound $c_*\varrho(x_1) \ge (c_*/2)\varrho(x)$, since the reference ratio over this displacement tends uniformly to one. Take $N = N_0$, absorb the error relative to the core density minimum, and choose $a_0 < c_*/4$ for a contradiction. If D is empty the contradiction is immediate without separation. This proves the reference lower bound on B_{L+9} and places that ball in the positive support.

The density sup bound and reference lower envelope trap V between a fixed lower bound and $C_*(1+L^2)$ on the outer core. Convex unit-distance secants yield $|\partial V| \le A_* \log R$ on B_{L+8}. For $|x| \le L+7$, positive averaging then gives

$$c_* f(x) \le f * \kappa_h(x) \le C_* f(x) + C_{*,N}R^{-N}.$$

For the lower bound use $|z| \le bh$. For the upper, on $|z| \le 1$ use $f(x-z) \le f(x)e^{A_*(\log R)|z|}$. The exponential is bounded for $|z| \le R^{-1/2}$, and on the rest it is at most R^{A_*}, absorbed by fixed sufficiently large Schwartz orders. Outside the unit ball use the density sup bound. Together with (12) and (17), this proves (15). Finally $V = U + O(1)$ on that core; Taylor secants at $x \pm \theta$ and $D^2U \le CI_d$ prove (16). □

6 Conditional trace curvature at small scales

Lemma 6.1 (Trace mass and gradient oscillation)**.** *Let u be convex on $B_{8a}(x)$ and suppose $\Delta u(B_{8a}(x)) \le Ma^d$. For any $p \in \partial u(x)$,*

$$|q-p| \le C_d M a \qquad (y \in B_a(x),\ q \in \partial u(y)).$$

Consequently $\Delta u(B_r(y)) \le C_d M a\, r^{d-1}$ for $y \in B_{a/2}(x)$ and $0 < r \le a/2$.

Proof. Subtract the supporting affine function at x, obtaining $v \ge 0$ and $v(x) = 0$. Along rays, convexity gives $v(x+2a\theta) \le 2a\,\partial_r v(x+t\theta)$ for $4a < t < 8a$ at admissible radii. The spherical average of the derivative is $\Delta v(B_t(x))/(|\mathbb{S}^{d-1}|t^{d-1})$ by the divergence theorem. Hence the average of v on the sphere of radius $2a$ is at most $C_d M a^2$. Its nonnegative subharmonicity and the Poisson majorant in $B_{2a}(x)$ bound v by $C_d M a^2$ on $B_{3a/2}(x)$. Convex secants then bound every subgradient on $B_a(x)$ by $C_d M a$. Subtracting the affine function changes neither the Laplacian nor its flux. Integrating the gradient bound on a sphere gives the last assertion. For nonsmooth u, mollify on the larger ball, choose comparison radii strictly inside $8a$, and pass to the limit. The mollified value at x contributes an additional term tending to $v(x) = 0$, so subtracting the original supporting plane causes no loss in this passage. Almost-everywhere spherical identities suffice; approximation from larger radii gives closed-ball measure bounds. The same argument in dimension one uses the two endpoints of an interval. □

On a positive compact core, convex V is locally Lipschitz and its Hessian is a positive matrix-valued measure. The distributional chain and product rules give

$$\Delta f = |\nabla V|^2 f\, dx - f\, \Delta V. \tag{19}$$

In particular, by Lemma 5.1 we may choose $\Gamma(x) = C_*(1+|x|^2)\varrho(x)$ so that $\Gamma \ge |\nabla V|^2 f$ on B_{L+5}.

Lemma 6.2. *For $x \in B_{L+1}$, put $P(x) = (1+|x|^2)\varrho(x)$. The local measure $\nu = \Delta(f-\varrho)$ satisfies*

$$|\nu|(B_r(x)) \le C_* P(x)\, h r^{d-1}, \qquad 0 < r \le h, \tag{20}$$

$$|\nu|(B_r(x)) \le C_* P(x)\, r^d e^{C_*(1+|x|)r + C_* r^2}, \qquad h \le r \le 1. \tag{21}$$

Only balls contained in the positive core are used here.

Proof. Choose a nonnegative smooth cutoff ζ, equal to one on B_{L+3}, supported in B_{L+4}, with uniformly bounded first two derivatives. The measure $\mu = \zeta(\Gamma\, dx - \Delta f)$ is positive and, where $\zeta = 1$, dominates $f\,\Delta V$. For $x \in B_{L+2}$, replacing $(\zeta\Delta f) * \kappa_h$ by $\Delta f * \kappa_h$ has error $O_{d,N}(R^{-N})$. Indeed, integrate by parts twice against $f \in L^1$; every nonzero derivative term of $(1-\zeta)\kappa_h(x-\cdot)$ is separated from x by at least one and is Schwartz-small. This permits arbitrary remote support boundaries. Equations (12), (18) and $|\Delta\varrho| \le C_*(1+|x|^2)\varrho$ give

$$0 \le \mu * \kappa_h(x) \le C_* P(x) + O(R^{-N_0}) \le C_* P(x).$$

Positivity and the kernel's lower bound on B_b yield $\mu(B_{bh}(x)) \le C_* h^d P(x)$. Covering $B_{32h}(x)$ by finitely many such cells bounds the weighted measure $f\,\Delta V$. On this ball the core comparison bounds f below by a fixed multiple of $\varrho(x)$; dividing by this positive lower bound gives

$$\Delta V(B_{32h}(x)) \le C_* h^d (1+|x|^2).$$

All these balls lie inside the required core for large R. Lemma 6.1, applied with $a = 2h$ using the mass bound on $B_{16h}(x) \subset B_{32h}(x)$, now bounds the oscillation of subgradients of V on $B_{2h}(x)$ by $C_*h(1 + |x|^2)$. Its spherical-flux consequence is $\Delta V(B_r(x)) \le C_*h(1 + |x|^2)r^{d-1}$ for $r \le h$. Equation (19), local reference comparability and the density bound on $|\nabla V|^2 f$ prove (20). For (21), cover by $O_d((r/h)^d)$ cells and use

$$\varrho(y)/\varrho(x) \le e^{C_*(1+|x|)r+Cr^2/2}, \qquad 1 + |y|^2 \le C(1 + |x|^2) \quad (|y - x| \le 1).$$

The fixed multiple of h added by a covering is harmless because $Lh \to 0$. This proves both estimates. □

7 A local divergence field and its weighted bias

Let $g = f - \varrho$. The Newtonian field A is defined in Fourier variables by

$$\widehat{A}(\xi) = -i\xi|\xi|^{-2}\widehat{g}(\xi), \qquad \operatorname{div} A = g.$$

It belongs to L^2: the common centered second moments control the origin, while $g \in L^1 \cap L^2$ by (7). Split it as $A = A_0 + A_1$ using the multiplier $\chi(h\xi)$. The low part was estimated in Lemma 4.2. For the high part,

$$A_1 = Q_h * \Delta g, \qquad \widehat{Q}_h(\xi) = i\xi|\xi|^{-4}[1 - \chi(h\xi)]. \tag{22}$$

The signs agree with our convention: $i\xi \cdot \widehat{A} = \widehat{g}$ and $\widehat{Q}_h(-|\xi|^2)\widehat{g} = (1 - \chi(h\xi))\widehat{A}$.

Lemma 7.1. *The kernel scales as $Q_h(z) = h^{3-d}Q(z/h)$. It and all derivatives decrease faster than every power away from zero. For $0 < |z| \le 1$,*

$$|Q(z)| \le \begin{cases} C_*|z|^{3-d}, & d > 3, \\ C_*(1 + |\log|z||), & d = 3, \\ C_*, & d = 1, 2. \end{cases}$$

Proof. An annular smooth partition of the high-frequency multiplier produces Schwartz kernels bounded by $C_{*,N}2^{j(d-3)}(1 + 2^j|z|)^{-N}$, $j \ge 0$, with finitely many initial transition annuli absorbed in the constants. The same bounds hold uniformly for finite annular partial sums, since we sum their absolute bounds. Summing gives the displayed bounds. Differentiation gives the corresponding annular derivative estimates; on $|z| \ge 1$ arbitrarily large Schwartz orders give rapid decay. The scaling follows by $\eta = h\xi$. In particular Q is locally and globally integrable. □

The next estimate supplies the cubic power of the resolution. The cell bound alone is insufficient near the singularity of the inverse kernel; the extra factor h in the subcell estimate (20) is what makes the field integrable at the required order in every dimension.

Lemma 7.2 (Conditional weighted field bound)**.** *For all sufficiently large R,*

$$|A_1(x)| \le C_*h^3P(x) + C_{*,N}h^N \qquad (|x| \le L) \tag{23}$$

for any prescribed fixed N. Consequently

$$\int_{B_L} \frac{|A(x)|^2}{f(x) + \varrho(x)}\,dx \le C_*h^6. \tag{24}$$

Proof. Fix $x \in B_L$. Choose a smooth θ_x, equal to one on $B_{1/2}(x)$ and supported in $B_1(x)$, with uniform derivative bounds. With the local measure $\nu = \Delta g$,

$$A_1(x) = \int \theta_x(y) Q_h(x-y)\, d\nu(y) + \int g(y) \Delta_y[(1-\theta_x(y)) Q_h(x-y)]\, dy. \tag{25}$$

The second integral is $O_{d,N}(h^N)$: its kernel arguments have distance at least $1/2$ from zero, and Lemma 7.1 bounds each derivative of order at most two there by $C_{*,N} h^N (1+|z|)^{-d-1}$. Since $\|g\|_1 \le 2$, this proves the remote estimate without assuming that Δg is a measure globally. To justify (25), first use finite annular Fourier sums, integrate by parts, and pass to the limit. The local estimates that follow dominate their singular integrals uniformly, and the remote derivative bounds are uniform in the finite sums.

For $|x-y| \le h$ and $d > 3$, a dyadic annulus of radius r contributes at most

$$C_* r^{3-d}[P(x) h r^{d-1}] = C_* P(x) h r^2$$

by (20). Summing $r = 2^{-k}h$ gives $C_* P(x) h^3$. For $d = 3$ the sum is bounded by $C_* P(x) h^3 \sum_{k \ge 0} (1+k) 2^{-2k}$. For $d = 1, 2$ use $|Q_h| \le C_* h^{3-d}$ and the cell mass $C_* h^d P(x)$. Thus the same bound holds in every dimension. This uses the factor h in (20), which a bare cell-mass estimate would not provide at subcell scales.

On annuli $r_j = 2^j h$ with $h \le r_j \le 1$, combine (21) and rapid kernel decay to bound the contribution by

$$C_{*,N} h^3 P(x)\, 2^{j(d-N)} e^{C_*(1+L) r_j + C_*}.$$

For $r_j \le \sqrt{h}$ the exponential is uniformly bounded, and the geometric series is bounded when $N > d$. For $r_j \ge \sqrt{h}$, its sum is at most

$$C_{*,N} h^3 P(x) R^{-(N-d)/2} e^{C_*(1+L)} = h^3 P(x) R^{-(N-d)/2 + o(1)}.$$

Choose any fixed $N > d + 2$. This proves (23).

Since $f + \varrho \ge \varrho$, squaring and integrating the principal term gives $C_* h^6 \int (1+|x|^2)^2 \varrho(x)\, dx$. The error contributes at most $C_{*,N} h^{2N} L^d e^{\beta L^2/2}$, which is bounded by $C_* h^6$ for a sufficiently large fixed $N > \beta\Lambda + d + 100$. For the low part, (13) gives weighted energy at most

$$C_* R^{d+\beta\Lambda/2} \varepsilon^2 \exp(\sigma^2 B^2 R^2) = C_* R^{d+\beta\Lambda/2} \varepsilon^{15/8} = o(h^6).$$

Combining the two parts proves (24). □

8 Global correction and the upper bounds

A Newtonian field can have polynomial tails, so the local estimate must not be integrated globally against an inverse reference density.

Lemma 8.1. *For $m = (f+\varrho)/2$, under the growing-core estimates,*

$$\mathrm{Var}_m\, v \le C_* \int |\nabla v|^2 m.$$

Proof. Let ρ_0 be normalized Lebesgue measure on B_1. The core comparison gives $\rho_0 \le C_* f$ and $\rho_0 \le C_* \varrho$. Hence by Cauchy–Schwarz,

$$|\mathbb{E}_{\rho_0} v - \mathbb{E}_f v|^2 \le \int |v - \mathbb{E}_f v|^2 \rho_0 \le C_* \mathrm{Var}_f\, v,$$

and similarly for the reference. Expand

$$\operatorname{Var}_m v = \tfrac{1}{2}\operatorname{Var}_f v + \tfrac{1}{2}\operatorname{Var}_\varrho v + \tfrac{1}{4}(\mathbb{E}_f v - \mathbb{E}_\varrho v)^2.$$

Bound the last term through ρ_0 and apply $C_P(f) \le Cd$, $C_P(\varrho) \le 1/c$. Here $H^1(m)$ denotes the completion of smooth functions of finite weighted norm in $\int (v^2+|\nabla v|^2)m$. The inequality extends to this space by approximation. The density m is locally bounded above and below by positive constants, so its completion has the usual local weak-gradient interpretation. □

Proposition 8.2. *For sufficiently small $\varepsilon = \mathcal{O}_W(f) > 0$,*

$$W_2(f\,dx, \varrho\,dx) \le C_*[\log(1/\varepsilon)]^{-3/2}.$$

The same conclusion holds with a sufficiently small $\varepsilon = \mathcal{O}_R(f) > 0$.

Proof. First use the scale (11) for the W_1 observation, and set $m = (f+\varrho)/2$. Choose a smooth cutoff η, equal to one on B_{L-4}, supported in B_{L-3}, with bounded gradient. For $g = f - \varrho$ put

$$r = g - \operatorname{div}(\eta A) = (1-\eta)g - \nabla\eta \cdot A.$$

It has integral zero because ηA is compactly supported. The residual has two different sources: the density difference outside the cutoff and the flux across its transition annulus. We estimate the first by tail mass and the second by the local weighted field energy, then solve one global weighted divergence equation. Set $\ell = L-4$. By (16), strong convexity of U and convexity of V, the radial derivative of $V(t\theta)$ is at least $c\ell - C_*$ for $t \ge \ell$ in the support. Thus

$$f(t\theta) \le C_*\varrho(\ell\theta)e^{-(c\ell - C_*)(t-\ell)} \le C_* e^{-a\ell^2} e^{-(c\ell-C_*)(t-\ell)}.$$

Past a finite support endpoint the density vanishes. For large ℓ, integration in polar coordinates and the reference envelope give

$$\int_{\mathbb{R}^d \setminus B_\ell} (f+\varrho) \le C_* \ell^{d-2} e^{-a\ell^2} = R^{-32+o(1)}. \tag{26}$$

Here $a = c/4$ and $\Lambda = 128/c$. The formula includes the scalar ℓ^{-1} tail factor. This is an observation-dependent exterior estimate; a global Gaussian tail bound on the input was not assumed.

Since $g^2/m \le 2(f+\varrho)$, (26) controls the tail part of r in $L^2(m^{-1})$. The annular part is controlled by (24), so

$$\int r^2/m \le C_* h^6.$$

In particular $r/m \in L^2(m)$ and $\int r = 0$. For a test function w,

$$\left|\int rw\right| = \left|\int r(w - \mathbb{E}_m w)\right| \le \left(\int r^2/m\right)^{1/2} \operatorname{Var}_m(w)^{1/2}.$$

Thus the residual defines a bounded functional in the gradient norm. On mean-zero $H^1(m)$, Lax–Milgram and Lemma 8.1 solve $\operatorname{div}(m\nabla v) = r$ weakly, with

$$\int |\nabla v|^2 m \le C_* \int r^2/m.$$

Indeed the weak equation is $\int \nabla v \cdot \nabla w\, m = -\int rw$; Poincaré bounds this functional in the gradient norm. Consequently $A_* = \eta A + m\nabla v$ satisfies $\operatorname{div} A_* = g$ and $\int |A_*|^2/m \le C_* h^6$. Lemma 4.1 now yields $W_2(f,\varrho)^2 \le C_* h^6$, proving the W_1-observation bound.

For the regret observation, Lemma 3.1 gives $\mathcal{O}_W(f) \le C_*\sqrt{\mathcal{O}_R(f)}$. The preceding estimate and $\log(1/(C_*\sqrt{\varepsilon})) \asymp \log(1/\varepsilon)$ give the claimed regret upper bound. The zero cases follow from Lemma 3.1 and Gaussian injectivity. These pointwise bounds imply both upper modulus bounds in Theorem 2.1. □

9 A matching family at every reference

The perturbation will be an absolute narrow-Gaussian profile. This avoids requiring Gaussian Fourier decay of the reference.

Choose $\tau > 0$ so that $\tau^{-2} > \beta + 1$, where β is from (8), and put $\psi = \phi_\tau$ and $w = \psi/\varrho$. Then

$$\begin{aligned} w(x) &\le C_* e^{-(\tau^{-2}-\beta)|x|^2/2}, \\ \nabla w &= w(\nabla U - x/\tau^2), \\ D^2 w &= w\big[(\nabla U - x/\tau^2)(\nabla U - x/\tau^2)^\top + D^2 U - \tau^{-2} I_d\big]. \end{aligned} \tag{27}$$

Hence $w, \nabla w, D^2 w$ are globally bounded: the strict Gaussian envelope absorbs every polynomial factor in these formulas. Only the two assumed derivatives of U have been used.

For real T sufficiently large define

$$\begin{aligned} B_T &= T^2 e^{-\tau^2(T^2-1)/2}, \qquad C_T = (T^2-1)e^{-\tau^2 T^2/2}, \\ h_T(t) &= \cos(Tt) - B_T \cos t + C_T, \qquad \delta_T = a_0/T^2, \\ f_T(x) &= \varrho(x) + \delta_T \psi(x) h_T(x_1) = \varrho(x)(1 + \delta_T k_T(x)), \qquad k_T(x) = w(x) h_T(x_1). \end{aligned} \tag{28}$$

The constant $a_0 > 0$ will be fixed sufficiently small, independently of T.

Lemma 9.1. *For this choice of a_0, every sufficiently large T gives $f_T \in \mathcal{C}_d$, with a C^2 potential satisfying*

$$\tfrac{c}{2} I_d \le D^2(-\log f_T) \le (C + c/2) I_d.$$

Moreover

$$W_2(f_T\,dx, \varrho\,dx) \ge \frac{a_0}{3T^3}. \tag{29}$$

Proof. For $Z \sim N(0, \tau^2)$,

$$\mathbb{E}\cos(tZ) = e^{-\tau^2 t^2/2}, \qquad \mathbb{E}[Z^2 \cos(tZ)] = \tau^2(1 - \tau^2 t^2) e^{-\tau^2 t^2/2}.$$

Put $E_T = e^{-\tau^2 T^2/2}$. Since $B_T e^{-\tau^2/2} = T^2 E_T$ and $C_T = (T^2 - 1)E_T$, the cancellations are

$$\begin{aligned} \mathbb{E} h_T(Z) &= E_T - T^2 E_T + (T^2 - 1)E_T = 0, \\ \tau^{-2}\mathbb{E}[Z^2 h_T(Z)] &= [1 - \tau^2 T^2 - T^2(1 - \tau^2) + T^2 - 1]E_T = 0. \end{aligned}$$

For example, subtracting τ^2 times the mass equation from the second-moment equation cancels the common τ factors and gives the stated B_T. First and mixed moments of $\psi h_T(x_1)$ vanish by parity. For $j > 1$, Gaussian independence gives $\int x_j^2 \psi h_T = \tau^2 \mathbb{E} h_T(Z) = 0$. Thus the additive perturbation has exactly zero mass, mean and covariance correction; f_T inherits mean zero and covariance I_d from the possibly asymmetric reference.

For large T, $B_T + C_T < 1$, $B_T \le 1/6$, and

$$\|h_T\|_\infty \le 2, \quad \|h_T'\|_\infty \le T + 1, \quad \|h_T''\|_\infty \le T^2 + 1.$$

Equation (27) therefore implies

$$\|k_T\|_\infty \le C_*, \qquad \|\nabla k_T\|_\infty \le C_* T, \qquad \|D^2 k_T\|_\infty \le C_* T^2.$$

Choose a_0 small enough that $1 + \delta_T k_T \ge 1/2$ and

$$\|D^2 \log(1 + \delta_T k_T)\|_\infty \le 2\delta_T \|D^2 k_T\|_\infty + 4\delta_T^2 \|\nabla k_T\|_\infty^2 \le c/2.$$

This proves positivity and both Hessian bounds globally.

The test $x \mapsto \cos(Tx_1)/T$ is one-Lipschitz. Its expectation difference under f_T and ϱ is $\delta_T \mathbb{E}[h_T(Z)\cos(TZ)]/T$. The expectation is at least $1/2 - B_T \geq 1/3$: $\mathbb{E}\cos^2(TZ) \geq 1/2$, the C_T term is nonnegative, and $|\mathbb{E}[\cos Z \cos(TZ)]| \leq 1$. The dual W_1 inequality and $W_2 \geq W_1$ prove (29). □

Lemma 9.2. *There is $\lambda > 0$ such that*

$$\mathcal{O}_W(f_T) \leq C_* T^{-2} e^{-\lambda T^2}, \qquad \mathcal{O}_R(f_T) \leq C_* T^{-2} e^{-2\lambda T^2} \tag{30}$$

for all sufficiently large T. The same alternatives satisfy

$$\chi^2(p_{f_T} \mid q) := \int \frac{(p_{f_T} - q)^2}{q} \leq C_* T^{-4} e^{-2\lambda T^2}.$$

Proof. Set

$$v^2 = \tau^2 + \sigma^2, \qquad b = \frac{\tau^2}{\tau^2 + \sigma^2}, \qquad \lambda = \frac{\tau^2 \sigma^2}{2(\tau^2 + \sigma^2)} > 0.$$

Gaussian conditional integration gives the exact formula

$$(\psi \cos(Tx_1)) * \phi_\sigma(y) = \phi_v(y) e^{-\lambda T^2} \cos(bTy_1).$$

Consequently $p_T = f_T * \phi_\sigma = q + \delta_T g_T$, where

$$g_T(y) = \phi_v(y)\big[e^{-\lambda T^2} \cos(bTy_1) - B_T e^{-\lambda} \cos(by_1) + C_T\big]. \tag{31}$$

This signed function has mass zero. Recenter each Lipschitz test to vanish at zero to obtain

$$W_1(p_T, q) \leq \delta_T \int |y|\, |g_T(y)|\, dy \leq C_* T^{-2} e^{-\lambda T^2}.$$

The corrections are absorbed because $\lambda < \tau^2/2$.

For relative Fisher information, the input ratio f_T/ϱ is between $1/2$ and $3/2$, so conditional averaging gives the same bounds on p_T/q. The lower envelope in (8) can be written $\varrho \geq b_3 \phi_{1/\sqrt{\beta}}$ for $b_3 > 0$. Thus

$$q \geq b_3 \phi_{\sqrt{\sigma^2 + 1/\beta}}, \qquad v^2 < \sigma^2 + 1/\beta.$$

It follows that the integral of any fixed polynomial times ϕ_v^2/q is finite. The bracket in (31) is bounded by $C_* e^{-\lambda T^2}$, so

$$\chi^2(p_T \mid q) = \delta_T^2 \int g_T^2/q \leq C_* T^{-4} e^{-2\lambda T^2}.$$

By (5), $|\nabla \log q(y)| \leq C_*(1 + |y|)$. The exact Fisher expression is

$$J(p_T \mid q) = \delta_T^2 \int \frac{|\nabla g_T - g_T \nabla \log q|^2}{q(1 + \delta_T g_T/q)}\, dy.$$

Equation (31) and the preceding Gaussian integrability bound give

$$\begin{aligned} J(p_T \mid q) &\leq C_* \delta_T^2 \big[(T^2 + 1) e^{-2\lambda T^2} + B_T^2 + C_T^2\big] \\ &\leq C_* T^{-2} e^{-2\lambda T^2}. \end{aligned}$$

For the last step use $\tau^2 - 2\lambda = \tau^4/(\tau^2 + \sigma^2) > 0$ to absorb the corrections' polynomial factors. Finally (4) gives the regret bound. □

To finish Theorem 2.1, let $\lambda_W = \lambda$, $\lambda_R = 2\lambda$. For $a \in \{W, R\}$ and every sufficiently small $\varepsilon > 0$, choose the real frequency

$$T^2 = \lambda_a^{-1} \log(C'_a/\varepsilon)$$

with a sufficiently large fixed C'_a. Then (30) gives $\mathcal{O}_a(f_T) \le \varepsilon$, while (29) gives $W_2(f_T, \varrho) \ge c_*[\log(1/\varepsilon)]^{-3/2}$. Taking common constants and the smaller threshold for the two observations proves both lower moduli for every small ε.

The matching lower moduli also rule out a uniform Hölder inverse and any increase of the logarithmic exponent 3/2.

10 Removing moment normalization from the deterministic modulus

The isotropic theorem is the analytic engine. We first track its constants uniformly over a fixed curvature class, then recover unknown moments from the observation.

Lemma 10.1 (Reference-uniform inverse)**.** *There are $K, \varepsilon_0 > 0$, depending only on d, σ, c, C, such that for every $\rho \in \mathcal{R}_{c,C}$ and $f \in \mathcal{C}_d$,*

$$W_2(f, \rho) \le K[\log(1/\varepsilon)]^{-3/2}, \qquad \varepsilon = W_1(p_f, p_\rho) \in (0, \varepsilon_0].$$

At zero output discrepancy the inputs agree. No common modulus of continuity of the reference Hessians is required.

Proof. Normalize $V = -\log \rho$ and put $g = \nabla V(0)$. Integration by parts, justified by the Gaussian tails and linear gradient growth of each reference, gives $\mathbb{E}_\rho \nabla V(X) = 0$. The upper Hessian bound and isotropy therefore imply

$$|g| \le \mathbb{E}|\nabla V(0) - \nabla V(X)| \le C\sqrt{d}.$$

Taylor's inequalities and normalization give

$$I_C(g) \le e^{V(0)} \le I_c(g), \qquad I_a(g) = (2\pi/a)^{d/2} e^{|g|^2/(2a)}.$$

Thus $V(0)$ and g are uniformly bounded. With $\beta = C + 1$, completing squares proves (8)–(10) with common constants depending only on c, C, d. Indeed $|\nabla V(x)| \le C\sqrt{d} + C|x|$, and $\Delta\rho = (|\nabla V|^2 - \Delta V)\rho$; these use only two derivatives. Also $C_P(\rho) \le 1/c$ uniformly.

We now track these constants through the upper proof. Choose the same kernels, $\Lambda = 128/c$, $\beta = C + 1$ and integer $N_0 > \beta\Lambda + d + 100$ for every reference. The frequency estimates are independent of the reference. The core density and slope estimates use only its envelopes and ratios, the bound $D^2 V \le C I_d$, and the dimension-only input density bound. Consequently their error thresholds and constants are common. The conditional curvature argument uses only the common Laplacian envelope and core slopes. The convex trace and high-pass kernel lemmas are dimensional. Their weighted core integral is bounded by a constant times $\int (1 + |x|^2)^2 \rho(x)\, dx$, uniformly finite by the upper Gaussian envelope. The low-field error uses the same $\beta\Lambda$. The exterior estimate (26) uses the common c, envelopes and slopes; its $o(1)$ exponent is uniform. In the mixture Poincaré lemma, both component densities have a common positive lower bound on B_1, while their Poincaré constants are bounded by Cd and $1/c$. The final cutoff, divergence correction and flux action therefore have common constants. Every sufficiently-large-R threshold can be chosen from these same bounds. Proposition 8.2 now proves the claim, with a common observation threshold. The zero case follows from the nonvanishing Gaussian Fourier multiplier. □

We now prove Corollary 2.2 by recovering and aligning the input moments from the known-noise observation.

Proof of Corollary 2.2. First consider output W_1 and write $\delta = \mathcal{O}_W^g(f) \le 1$. Translate both laws by the mean of g. This changes neither the noise model nor the transport distance. The centered reference covariance C_g satisfies

$$C^{-1}I \le C_g \le c^{-1}I.$$

The upper bound is Brascamp–Lieb [12]. For the lower, integration by parts gives $\mathbb{E}_g[X\nabla V(X)^\top] = I$ and $\mathbb{E}_g[\nabla V \nabla V^\top] = \mathbb{E}_g D^2 V \le CI$; directional Cauchy–Schwarz yields the claim.

Moment bounds inferred from the observation. For $Y = X + \sigma G$ with $X \sim f$,

$$\mathbb{E}_f|X| \le \mathbb{E}|Y| + \sigma\mathbb{E}|G| \le \mathbb{E}_g|X + \sigma G| + \delta + \sigma\mathbb{E}|G| \le M,$$

where M depends only on d, σ, c. Increase M to at least one. For any scalar marginal Z of f, $\mathbb{E}|Z| \le M$. Markov's inequality at $a = 4M$ gives $\Pr(Z \ge a) \le 1/4$ and $\Pr(Z \ge -a) \ge 3/4$. Its survival function S is log-concave. If $S(a) > 0$, concavity of $\log S$ gives, for $t \ge a$,

$$S(t) \le S(a) \exp\left[-\frac{\log 3}{2a}(t-a)\right].$$

If $S(a) = 0$ the right tail vanishes. Reflection treats the left tail. Integrating these tails and using $|X|^4 \le d\sum_j |X_j|^4$ gives a uniform fourth-moment bound for f. Centering and Gaussian addition preserve this bound.

Let $m = \mathbb{E}_f X$ and let f_0 be the law of $X - m$. Linear Kantorovich tests give $|m| \le \delta$, so

$$W_1(f_0 * \phi_\sigma, g * \phi_\sigma) \le 2\delta.$$

Apply the following truncated tests to the centered output laws $f_0 * \phi_\sigma$ and $g * \phi_\sigma$, whose W_1 distance is at most 2δ and whose fourth moments have the obtained common bound. Choose a smooth cutoff χ equal to one on $[0,1]$ and zero on $[2,\infty)$, and test against

$$F_{ij,R}(x) = x_i x_j \chi(|x|/R).$$

Then $\mathrm{Lip}(F_{ij,R}) \le KR$ and $|x_i x_j - F_{ij,R}(x)| \le K|x|^4/R^2$. Thus the covariance discrepancy is bounded by $K(R\delta + R^{-2})$. Choosing $R = \delta^{-1/3}$ proves

$$\|C_f - C_g\|_{\mathrm{F}} \le K\delta^{2/3}, \qquad C_f = \mathrm{Cov}(f).$$

The common known noise cancels in the covariance difference. For sufficiently small δ, both covariances lie in a fixed positive spectral interval.

Alignment and positive noise completion. Put $L = C_g^{1/2} C_f^{-1/2}$ and $a = L_\# f_0$. Matrix square roots and inverse square roots are Lipschitz on that spectral interval. For completeness, the square-root estimate follows from

$$A^{1/2}X + XD^{1/2} = A - D, \qquad X = A^{1/2} - D^{1/2},$$

by dividing each entry in separate left and right eigenbases by a uniformly positive sum of square roots. The inverse identity handles inverse square roots. The same-input coupling therefore gives

$$\mathrm{Cov}(a) = C_g, \qquad W_2(f_0, a) \le K\delta^{2/3}.$$

Replacing only the latent law leaves the Gaussian noise unchanged, hence

$$W_1(a * \phi_\sigma, g * \phi_\sigma) \le 2\delta + K\delta^{2/3}.$$

Apply $C_g^{-1/2}$ to both outputs. Both latent laws are now isotropic and the first remains log-concave. The reference potential becomes $V(C_g^{1/2} z)$, with Hessian

$$C_g^{1/2} D^2 V(C_g^{1/2} z) C_g^{1/2},$$

so its curvature bounds are c/C and C/c. The common noise covariance is $\sigma^2 C_g^{-1} \le \sigma^2 C I$. Add the same Gaussian of positive covariance

$$(\sigma^2 C + 1) I - \sigma^2 C_g^{-1} \ge I.$$

Convolution contracts W_1 and produces a fixed scalar noise covariance. Theorem 2.1 with the uniformity of Lemma 10.1 applies to these exactly normalized laws. Undoing the bounded affine map and paying the centering and alignment costs gives

$$W_2(f, g) \le K \left\{ [\log(1/(K\delta^{2/3}))]^{-3/2} + \delta^{2/3} + \delta \right\} \le K' [\log(1/\delta)]^{-3/2}.$$

If a completed discrepancy vanishes, Gaussian injectivity supplies equality at that comparison. At original discrepancy zero, injectivity gives $f = g$ directly.

Regret and matching alternatives. The score and projection identity in Lemma 3.1 requires finite moments and reference curvature, rather than exact normalization. Log-concave probability densities have finite moments, and the reference posterior mean has linear growth. Consequently, for $p = f * \phi_\sigma$ and $q = g * \phi_\sigma$,

$$\mathrm{Reg}_g(f) = \sigma^4 \int p |\nabla \log p - \nabla \log q|^2, \qquad W_1(p, q) \le \kappa^{-1} \sigma^{-2} \sqrt{\mathrm{Reg}_g(f)}, \quad \kappa = \frac{c}{1 + c\sigma^2}.$$

The second estimate is the reference-output argument in (5). The new W_1 upper therefore proves the regret upper as well.

For the lower bounds, use the explicit additive Gaussian-profile construction in Lemmas 9.1 and 9.2, now around the centered g. Its perturbation has zero mass, every first moment, and every second-moment entry independently of the reference moments. It therefore preserves the actual mean and covariance of g. The centered reference has Gaussian envelopes and linear score growth by its curvature bounds. All bounded-ratio derivative, positivity, curvature, transport-separation and output-convolution calculations in that construction apply with these envelopes. They give, for all sufficiently large real T,

$$W_2(f_T, g) \ge k T^{-3}, \quad \mathcal{O}_W^g(f_T) \le K T^{-2} e^{-bT^2}, \quad \mathcal{O}_R^g(f_T) \le K T^{-2} e^{-2bT^2}.$$

Translation restores the original reference mean. Choosing T of order $\sqrt{\log(1/\varepsilon)}$ proves both lower moduli. □

The fixed-reference test and Gaussian linear-MMSE corollary use their stated isotropic classes. Theorem 2.5 uses the uniform normalization-free inverse to estimate curvature-bounded laws with unknown means and covariances.

11 Statistical recovery from noisy samples

We use the reference-uniform inverse and moment-recovery reduction of Section 10. The empirical estimate below supplies a polynomial observation error; logarithmic inversion fixes the final estimation order.

We prove the statistical results stated in Section 2, using its observation experiment and empirical law Q_n. For clarity these results assert rates up to positive constants, not sharp leading constants or computational efficiency.

Lemma 11.1 (A sufficient empirical bound)**.** *If a law Q has second moment at most M, then for $b = 1/[4(d+1)]$,*

$$\mathbb{E}W_1(Q_n, Q) \le K_{d,M} n^{-b}.$$

In particular, uniformly over $f \in \mathcal{C}_d$,

$$P_f^{\otimes n}\{W_1(Q_n, p_f) > n^{-b/2}\} \le K n^{-b/2}. \tag{32}$$

Proof. Map points outside $[-A, A]^d$ to zero and quantize points inside to a grid of mesh η, including zero as a site. The expected transport cost is at most $K_d\eta + M/A$, since $\mathbb{E}[|Y|\mathbf{1}_{|Y|>A}] \le M/A$. The grid has at most $K_d(1 + A/\eta)^d$ sites and diameter at most $K_d A$. The expected total variation of its empirical histogram from its law is at most half the square root of the number of sites divided by n, by Cauchy–Schwarz and Bernoulli variances. Couple the unmatched histogram mass at cost at most the diameter. The triangle inequality, also accounting for empirical quantization, gives

$$\mathbb{E}W_1(Q_n, Q) \le K_{d,M}[\eta + A^{-1} + A\sqrt{(1 + A/\eta)^d/n}].$$

Take $A = n^b$ and $\eta = n^{-b}$. The last term is $O(n^{-1/4})$, and the other terms are $O(n^{-b})$. For $Q = p_f$, the second moment is $d(1 + \sigma^2)$; Markov's inequality proves (32). □

11.1 Proof of Theorem 2.4

Proof of Theorem 2.4. Reject when $W_1(Q_n, p_\rho) > 2n^{-b/2}$. Its type-I error tends to zero by Lemma 11.1. Theorem 2.1 implies that $W_2(f, \rho) \ge A(\log n)^{-3/2}$ forces $W_1(p_f, p_\rho) \ge n^{-\theta}$, where $\theta = (K_\rho/A)^{2/3}$, for sufficiently large n. If the output discrepancy exceeds the fixed inverse threshold, this lower bound is automatic; otherwise it follows by rearranging the inverse inequality. Choose A large enough that $\theta < b/4$. Then $n^{-\theta} > 4n^{-b/2}$ eventually, and acceptance forces $W_1(Q_n, p_f) > 2n^{-b/2}$. Equation (32) bounds the type-II error uniformly, proving the first limit.

For the lower bound take $T^2 = \lambda^{-1} \log n$ in Lemma 9.2. Then $n\chi^2(p_{f_T} \mid p_\rho) \to 0$. The product identity

$$1 + \chi^2(P_{f_T}^{\otimes n} \mid P_\rho^{\otimes n}) = [1 + \chi^2(p_{f_T} \mid p_\rho)]^n$$

makes the total variation of the two joint laws tend to zero. Lemma 9.1 gives $W_2(f_T, \rho) \ge (a_0\lambda^{3/2}/3)(\log n)^{-3/2}$. For any test its two errors on these two laws sum to at least one minus their total variation: integrate the test against the difference of the laws. The infimum risk is also at most one by a constant test. Choose $a_\rho \le a_0\lambda^{3/2}/3$ to obtain the second limit. □

11.2 Proof of Theorem 2.5

Proof. Write $p_f^\Sigma = f * \gamma_\Sigma$, let Q_n be the empirical law of these observations, let $m_f = \mathbb{E}_f X$, and put $M = d/c + \operatorname{tr}\Sigma$. Brascamp–Lieb gives $\operatorname{Cov}(f) \le c^{-1} I_d$, so $\mathbb{E}|Y - m_f|^2 \le M$ uniformly

over $\mathcal{F}_{c,C}$. Translation invariance and Lemma 11.1, applied after the theoretical translation by m_f, give, for $b = 1/[4(d+1)]$,

$$\sup_f \mathbb{E}e_n \le Kn^{-b}, \qquad e_n = W_1(Q_n, p_f^\Sigma).$$

This translation is used only for the bound; the estimator below uses the observed sample mean $\overline{Y}$.

Mean-anchored selection. The finite-second-moment probability laws form a separable W_2 space. A countable base restricts to a countable base of every metric subspace; choosing a point in each nonempty base element gives a countable W_2-dense subset $\{f_j\}$ of $\mathcal{F}_{c,C}$. Put $\eta_n = n^{-b/2}$ and select the first j satisfying

$$W_1(p_{f_j}^\Sigma, Q_n) < \inf_k W_1(p_{f_k}^\Sigma, Q_n) + \eta_n.$$

Each distance is continuous in the observations and the infimum is countable, so the selected density f_* is measurable. Gaussian convolution contracts W_1; hence density of this subset gives $W_1(p_{f_*}^\Sigma, Q_n) \le e_n + \eta_n$. Translate f_* by $a_n = \overline{Y} - m_{f_*}$ to obtain $\widehat{f}_n = (x \mapsto x + a_n)_\# f_*$. Translation preserves $\mathcal{F}_{c,C}$. On each measurable selection set the translation depends continuously on $\overline{Y}$, so $\widehat{f}_n$ is a Borel probability-valued estimator in W_2. Linear Kantorovich tests give

$$|a_n| \le W_1(p_{f_*}^\Sigma, Q_n), \qquad W_1(p_{\widehat{f}_n}^\Sigma, p_f^\Sigma) \le 3e_n + 2\eta_n.$$

The resulting estimator has mean $\overline{Y}$ exactly.

Uniform inversion and exceptional samples. Choose $s > \|\Sigma\|$ and add the same Gaussian of covariance $sI_d - \Sigma$ to both output laws. This contracts W_1 and makes the common noise scalar. Corollary 2.2, with f as reference and its uniform upper constants, therefore bounds $W_2(\widehat{f}_n, f)$ by $K(\log n)^{-3/2}$ on $G_n = \{e_n \le \eta_n\}$ for large n. At zero completed discrepancy, Gaussian injectivity gives equality. Markov's inequality gives $\Pr(G_n^c) \le Kn^{-b/2}$ uniformly.

There is no bounded diameter for the full location class. Instead, independently couple the centered versions of each realized pair $(\widehat{f}_n, f)$ to obtain

$$W_2(\widehat{f}_n, f)^2 \le |\overline{Y} - m_f|^2 + 2d/c.$$

Consequently

$$\mathbb{E}[W_2(\widehat{f}_n, f)^2 \mathbf{1}_{G_n^c}] \le M/n + (2d/c)\Pr(G_n^c) = O(n^{-b/2}),$$

uniformly over every mean. For unsquared loss use $W_2 \le |\overline{Y} - m_f| + \sqrt{2d/c}$ and Cauchy–Schwarz; its exceptional contribution is also polynomially small. Both upper risks follow. This argument pays the exceptional event using only the uniform centered second moment.

Matching alternatives in a noisy direction. Choose $\alpha = (c+C)/2$ and the Gaussian reference $g = N(0, \alpha^{-1} I_d)$. Rotate coordinates so $\Sigma = \mathrm{diag}(s_1, \dots, s_d)$ with $s_1 > 0$. Apply the additive construction (28) at this reference, with $\psi = \phi_\tau$, $\tau^2 < \alpha^{-1}$ sufficiently small, and the same exact-moment corrections B_T, C_T. The Gaussian-ratio derivative bounds in (27) apply directly to this reference. Choose a_0 so the global Hessian perturbation is at most $\min(\alpha - c, C - \alpha)/2$. Thus every sufficiently large T gives $f_T \in \mathcal{F}_{c,C}$ and $W_2(f_T, g) \ge kT^{-3}$.

Put $D = \tau^2 I_d + \Sigma$, $q = g * \gamma_\Sigma$, $\lambda_1 = \tau^2 s_1/[2(\tau^2 + s_1)] > 0$, and $b_1 = \tau^2/(\tau^2 + s_1)$. If φ_D is the Gaussian density with covariance D, conditioning gives

$$(\psi \cos(Tx_1)) * \gamma_\Sigma(y) = \varphi_D(y) e^{-\lambda_1 T^2} \cos(b_1 T y_1).$$

The corrected difference has the same formula as (31), with these quantities. The reference covariance $\alpha^{-1}I_d + \Sigma$ is strictly larger than D, so $\int \varphi_D^2/q < \infty$. Since $\lambda_1 < \tau^2/2$, the correction polynomials are exponentially smaller. Hence

$$\chi^2(p^{\Sigma}_{f_T} \mid p^{\Sigma}_{g}) \le KT^{-4}e^{-2\lambda_1 T^2}.$$

Take $T^2 = \lambda_1^{-1}\log n$. Product total variation tends to zero. If the input separation is D_n, choosing the nearer input turns any estimator into a test and gives lower risk at least $(D_n/2)^r(1-\mathrm{TV})/2$. This proves the lower bounds for both losses and every strict curvature interval $0 < c < C$.

Finally, on $\mathcal{R}_{c,C}$ with $0 < c < 1 < C$ and scalar noise, repeat the countable selection over $\mathcal{R}_{c,C}$, whose W_2 diameter is at most $\sqrt{2d}$. The normalized uniform inverse supplies the upper. The standard Gaussian and the same sufficiently small perturbation give the matching normalized lower. This proves the stated subclass conclusion as well. □

The statistical transfers are standard; the uniform weighted inverse and exact-moment alternatives determine the rate.

AI assistance. OpenAI ChatGPT/Codex assisted with literature discovery, proof exploration and drafting. The author reviewed the resulting manuscript and takes responsibility for its mathematical statements and references.

References

[1] X. Nguyen, Convergence of latent mixing measures in finite and infinite mixture models, *Ann. Statist.* **41** (2013), 370–400. doi:10.1214/12-AOS1065.

[2] J. Dedecker and B. Michel, Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension, *J. Multivariate Anal.* **122** (2013), 278–291. doi:10.1016/j.jmva.2013.08.009.

[3] C. Scricciolo, Minimax rates for Wasserstein–Kantorovich distribution deconvolution under known ordinary smooth errors and dependent signal processes, *Stat. Methods Appl.* (2026). doi:10.1007/s10260-026-00858-9.

[4] S. G. Bobkov, G. P. Chistyakov and F. Götze, Stability of Cramer's characterization of normal laws in information distances, in *High Dimensional Probability VII: The Cargèse Volume*, Progress in Probability **71** (2016), 3–35. doi:10.1007/978-3-319-40519-3_1.

[5] R. Eldan and B. Klartag, Pointwise estimates for marginals of convex bodies, *J. Funct. Anal.* **254** (2008), 2275–2293. doi:10.1016/j.jfa.2007.08.014.

[6] R. Peyré, Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localisation of Wasserstein distance, *ESAIM Control Optim. Calc. Var.* **24** (2018), 1489–1501. doi:10.1051/cocv/2017050.

[7] J. Niles-Weed and Q. Berthet, Minimax estimation of smooth densities in Wasserstein distance, *Ann. Statist.* **50** (2022), 1519–1540. doi:10.1214/21-AOS2161.

[8] J. Park and A. K. H. Kim, Minimax rate-optimal estimation in log-concave deconvolution, JSM Annual 2026, contributed presentation, August 6, 2026. Conference abstract; slides, "Minimax theory for log-concave deconvolution," pp. 7, 10–13.

[9] D. Guo, Y. Wu, S. Shamai (Shitz) and S. Verdú, Estimation in Gaussian noise: Properties of the minimum mean-square error, *IEEE Trans. Inform. Theory* **57** (2011), 2371–2385. doi:10.1109/TIT.2011.2111010.

[10] B. Efron, Tweedie's formula and selection bias, *J. Amer. Statist. Assoc.* **106** (2011), 1602–1614. doi:10.1198/jasa.2011.tm11181.

[11] J. Chen and Y. Wu, Sharp regret–Hellinger bounds for Gaussian empirical Bayes via polynomial approximation, preprint (2026), arXiv:2605.02070v2.

[12] H. J. Brascamp and E. H. Lieb, On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation, *J. Funct. Anal.* **22** (1976), 366–389. doi:10.1016/0022-1236(76)90004-5.

[13] D. Bakry, I. Gentil and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Springer, 2014. doi:10.1007/978-3-319-00227-9.

[14] F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, *J. Funct. Anal.* **173** (2000), 361–400. doi:10.1006/jfan.1999.3557.

[15] L. Gross, Logarithmic Sobolev inequalities, *Amer. J. Math.* **97** (1975), 1061–1083. doi:10.2307/2373688.

[16] M. Talagrand, Transportation cost for Gaussian and other product measures, *Geom. Funct. Anal.* **6** (1996), 587–600. doi:10.1007/BF02249265.

[17] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, *Numer. Math.* **84** (2000), 375–393. doi:10.1007/s002110050002.

[18] S. G. Bobkov, Isoperimetric and analytic inequalities for log-concave probability measures, *Ann. Probab.* **27** (1999), 1903–1921. doi:10.1214/aop/1022874820.