A sharp logarithmic inverse bound for Gaussian estimation

Scott Schubr

October 23, 2026

Abstract

Fix a dimension, a positive Gaussian noise level, and a centered isotropic reference density whose twice continuously differentiable potential has Hessian between two positive constant multiples of the identity. We prove that, uniformly over all isotropic log-concave inputs, the input's quadratic Wasserstein distance from the reference has the sharp inverse-logarithmic modulus of order 3/2 from either the W_1 distance of the noisy outputs or the excess quadratic risk of the reference Bayes estimator. An explicit narrow-Gaussian perturbation has exactly zero mass, mean and covariance corrections and supplies the matching lower order at every permitted reference. The Gaussian-reference case yields the corresponding sharp modulus from the total linear-MMSE gap at every fixed positive SNR. The upper proof uses a positive bandlimited core comparison, conditional curvature measures, a convex gradient-oscillation estimate and a weighted divergence correction. Inputs may have compact support or nonsmooth convex potentials. The upper constants are uniform when the reference Hessian lies in a fixed positive curvature interval. This yields sharp minimax W_2 estimation rates of order $(\log n)^{-3/2}$ on isotropic two-sided curvature classes, and fixed-reference goodness-of-fit testing at the same separation order from noisy independent samples. Constants depend on dimension, curvature bounds and noise level; the matching estimation lower bound assumes $0 < c < 1 < C$.

1 The fixed-reference theorem

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

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

Theorem 1.1. *Under (1), there are $c_*, C_*, \varepsilon_* > 0$ such that, for $a = W, R$ and $0 < \varepsilon \le \varepsilon_* < 1$,*

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

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

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

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 $(C + c/2)I_d$.

Corollary 1.2 (Gaussian linear-MMSE gap)**.** *Fix $d \ge 1$ 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 $0 < \varepsilon \le \varepsilon_{d,s}$,

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

At zero gap X is standard Gaussian.

Proof. Take $\varrho = \phi_1$ and $\sigma = s^{-1/2}$ in Theorem 1.1. 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 $\mathrm{Reg}_{\phi_1}(f_X) = \Delta_s(X)$. □

The main theorem concerns exact functionals at a fixed noise level. Lemma 9.1 and the following statistical results give reference-uniform and finite-data consequences. Constants are not dimension-free. Gaussian MMSE extremality and its equality case are classical [12]; the quantitative latent transport modulus is our concern.

Logarithmic deconvolution is established territory. Nguyen [13, 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 [7] prove minimax Wasserstein deconvolution rates under moment hypotheses. Scricciolo [18, Theorem 2] has the power $3/2$ in a smoothness-two statistical W_1 problem with Gaussian errors. Bobkov–Chistyakov–Götze [4] prove logarithmic information-distance stability for Gaussian-regularized summands in Cramér's theorem. Neither Gaussian divisibility nor smoothness of the input is imposed here.

The shape-constrained pointwise deconvolution principle has close ancestry in Eldan–Klartag [10, Section 3]. Weighted negative-Sobolev transport comparisons are classical [16];

the significance of density lower bounds for smooth Wasserstein estimates is discussed in [14]. 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).

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. Our Fourier convention is $\widehat{g}(\xi) = \int e^{-ix\cdot\xi} g(x)\,dx$.

2 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 is standard estimation theory, not an additional inverse result. Tweedie's formula is discussed in [9], which credits Tweedie through Robbins's 1956 work. The score-relative-Fisher form of regret is explicit, for example, in [6, eq. (2)].

Lemma 2.1. *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 [5] 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 $\kappa W_2(p_f, q)^2 \le 2D(p_f \mid q)$ [1, 15]; their Gaussian forms are [11, 19]. Combining and using $W_1 \le W_2$ proves (5). □

For the Gaussian reference, scaling the output in Lemma 2.1 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.

3 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 2.1. All dimensions $d \geq 1$ are included. Write $B_r(x)$ for a Euclidean ball and $B_r = B_r(0)$.

Lemma 3.1 (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 [2] 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 [16]. □

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 [10, eq. (59)], $\|f\|_\infty \leq \exp(\frac{1}{2} d \log d + 6d)$. The Poincaré bound is the classical Kannan–Lovász–Simonovits/Bobkov estimate: [3, 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 [5]. No derivative of U beyond order two enters these estimates.

Fix a real radial compactly supported smooth frequency multiplier χ, equal to one near zero, and write $K = \mathcal{F}^{-1}\chi$. It is a possibly signed Schwartz kernel with

$$\int K = 1, \quad \int y_j K(y)\,dy = 0, \quad \int y_j y_k K(y)\,dy = 0 \quad (1 \leq j,k \leq d). \tag{11}$$

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. Let B contain the frequency supports of $\chi, \widehat{\kappa}$, and use $K_h(x) = h^{-d}K(x/h)$ and the analogous κ_h.

Set

$$R^2 = \frac{\log(1/\varepsilon)}{8\sigma^2 B^2}, \qquad h = R^{-1}, \qquad L = (\Lambda \log R)^{1/2}, \qquad \Lambda = 128/c. \tag{12}$$

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 3.2. *For multiindices $|\alpha| \le 2$,*

$$\|(f-\varrho) * \partial^\alpha \kappa_h\|_\infty \le R^{-N_0} \tag{13}$$

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

$$\|A_0\|_2^2 \le C_d \varepsilon^2 R^d \exp(\sigma^2 B^2 R^2). \tag{14}$$

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

$$|\widehat{f}(\xi) - \widehat{\varrho}(\xi)| \le \varepsilon|\xi| e^{\sigma^2|\xi|^2/2}. \tag{15}$$

Fourier inversion bounds the left side of (13) 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 (14). These bounds do not assume that derivatives of f exist as functions. □

4 Pointwise reference bounds on a growing core

This step is a positive-kernel version of the shape-constrained deconvolution principle in [10, Section 3].

Lemma 4.1. *For all sufficiently large R, the positive support contains B_{L+9}, and*

$$c_*\varrho(x) \le f(x) \le C_*\varrho(x) \quad (|x| \le L+6), \tag{16}$$

$$|p - \nabla U(x)| \le C_* \quad (p \in \partial V(x)) \quad (|x| \le L+5), \tag{17}$$

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

$$c_*\varrho(x) \le \varrho * \kappa_h(x) \le C_*\varrho(x) + C_{*,N}R^{-N}, \tag{18}$$

$$((1+|\cdot|^2)\varrho) * \kappa_h(x) \le C_*(1+|x|^2)\varrho(x) + C_{*,N}R^{-N}. \tag{19}$$

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 (13), (18) 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 (13) and (18), this proves (16). Finally $V = U + O(1)$ on that core; Taylor secants at $x \pm \theta$ and $D^2U \le CI_d$ prove (17). □

5 Conditional trace curvature at small scales

Lemma 5.1 (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

$$\Delta f = |\nabla V|^2 f\, dx - f\, \Delta V. \tag{20}$$

In particular, by Lemma 4.1 we may choose $\Gamma(x) = C_*(1 + |x|^2)\varrho(x)$ so that $\Gamma \ge |\nabla V|^2 f$ on B_{L+5}.

Lemma 5.2. *For $x \in B_{L+1}$, put $P(x) = (1 + |x|^2)\varrho(x)$. The local measure $\nu = \Delta(f - \varrho)$ satisfies*

$$|\nu|(B_r(x)) \le C_* P(x)\, h r^{d-1}, \qquad 0 < r \le h, \tag{21}$$

$$|\nu|(B_r(x)) \le C_* P(x)\, r^d e^{C_*(1+|x|)r + C_* r^2}, \qquad h \le r \le 1. \tag{22}$$

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 (13), (19) 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, and using the core comparison there, 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 5.1, 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 (20), local reference comparability and the density bound on $|\nabla V|^2 f$ prove (21). For (22), 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. □

6 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 3.2. For the high part,

$$A_1 = Q_h * \Delta g, \qquad \widehat{Q}_h(\xi) = i\xi|\xi|^{-4}[1-\chi(h\xi)]. \tag{23}$$

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 6.1. *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. 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. □

Lemma 6.2 (Conditional weighted field bound)**.** *For all sufficiently large R,*

$$|A_1(x)| \le C_*h^3P(x) + C_{*,N}h^N \qquad (|x| \le L) \tag{24}$$

for any prescribed fixed N. Consequently

$$\int_{B_L} \frac{|A(x)|^2}{f(x)+\varrho(x)}\,dx \le C_*h^6. \tag{25}$$

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 $\nu = \Delta g$,

$$\begin{aligned} A_1(x) = &\int \theta_x(y)Q_h(x-y)\,d\nu(y) \\ &+ \int g(y)\Delta_y[(1-\theta_x(y))Q_h(x-y)]\,dy. \end{aligned} \tag{26}$$

The second integral is $O_{d,N}(h^N)$: its kernel arguments have distance at least $1/2$ from zero, and Lemma 6.1 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 (26), 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

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

by (21). 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 (21), 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 (22) 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 (24).

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, (14) 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 (25). □

7 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 7.1. *For $m = (f+\varrho)/2$, under the growing-core estimates,*

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

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 $|\mathbb{E}_{\rho_0} v - \mathbb{E}_f v|^2 \le C_* \operatorname{Var}_f v$, 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 $C_P(f) \le Cd$, $C_P(\varrho) \le 1/c$. The inequality extends to $H^1(m)$ by approximation; m is locally bounded above and below by positive constants. □

Proposition 7.2. *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 (12) for the W_1 observation. 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. Set $\ell = L - 4$. By (17), strong convexity of U and convexity of V, the radial derivative of $V(t\theta)$ is at least $c\ell - C_*$ for $t \geq \ell$ in the support. Thus

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

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

$$\int_{\mathbb{R}^d \setminus B_\ell} (f + \varrho) \leq C_* \ell^{d-2} e^{-a\ell^2} = R^{-32+o(1)}. \tag{27}$$

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 $g^2/m \leq 2(f + \varrho)$, (27) controls the tail part of r in $L^2(m^{-1})$. The annular part is controlled by (25), so

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

On mean-zero $H^1(m)$, Lax–Milgram and Lemma 7.1 solve $\operatorname{div}(m\nabla v) = r$ weakly, with

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

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 $\int |A_*|^2/m \leq C_* h^6$. Lemma 3.1 now yields $W_2(f, \varrho)^2 \leq C_* h^6$, proving the W_1-observation bound.

For the regret observation, Lemma 2.1 gives $\mathcal{O}_W(f) \leq C_*\sqrt{\mathcal{O}_R(f)}$. The preceding estimate and $\log(1/(C_*\sqrt{\varepsilon})) \asymp \log(1/\varepsilon)$ give the claimed regret upper bound. The zero cases were settled in Section 2. These pointwise bounds imply both upper modulus bounds in Theorem 1.1. □

8 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

$$\begin{aligned} w(x) &\leq C_* e^{-(\tau^{-2}-\beta)|x|^2/2}, \\ \nabla w &= w(\nabla U - x/\tau^2), \\ D^2 w &= w\big[(\nabla U - x/\tau^2)(\nabla U - x/\tau^2)^\top + D^2U - \tau^{-2} I_d\big]. \end{aligned} \tag{28}$$

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

$$\begin{aligned} B_T &= T^2 e^{-\tau^2(T^2-1)/2}, \qquad C_T = (T^2-1)e^{-\tau^2 T^2/2}, \\ h_T(t) &= \cos(Tt) - B_T \cos t + C_T, \qquad \delta_T = a_0/T^2, \\ f_T(x) &= \varrho(x) + \delta_T \psi(x) h_T(x_1) = \varrho(x)(1+\delta_T k_T(x)), \qquad k_T(x) = w(x) h_T(x_1). \end{aligned} \tag{29}$$

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

Lemma 8.1. *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

$$W_2(f_T\,dx, \varrho\,dx) \ge \frac{a_0}{3T^3}. \tag{30}$$

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

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

The displayed B_T, C_T give $\mathbb{E}h_T(Z) = \mathbb{E}[Z^2 h_T(Z)] = 0$ exactly. 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 (28) 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 $1/2 - B_T \ge 1/3$: $\mathbb{E}\cos^2(TZ) \ge 1/2$, the C_T term is nonnegative, and $|\mathbb{E}[\cos Z \cos(TZ)]| \le 1$. The dual W_1 inequality and $W_2 \ge W_1$ prove (30). □

Lemma 8.2. *There is $\lambda > 0$ such that*

$$\mathcal{O}_W(f_T) \le C_* T^{-2} e^{-\lambda T^2}, \qquad \mathcal{O}_R(f_T) \le C_* T^{-2} e^{-2\lambda T^2} \tag{31}$$

for all sufficiently large T.

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

$$g_T(y) = \phi_v(y)\big[e^{-\lambda T^2}\cos(bTy_1) - B_T e^{-\lambda}\cos(by_1) + C_T\big]. \tag{32}$$

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

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

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 $\varrho \ge b_3\phi_{1/\sqrt{\beta}}$ for $b_3 > 0$. Thus

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

It follows that the integral of any fixed polynomial times ϕ_v^2/q is finite. By (5), $|\nabla \log q(y)| \le C_*(1+|y|)$. 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 (32) and the preceding Gaussian integrability bound give

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

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 1.1, 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 (31) gives $\mathcal{O}_a(f_T) \le \varepsilon$, while (30) 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.

Corollary 8.3. *At any reference satisfying (1), neither observation admits a uniform inverse bound $W_2(f, \varrho) \le D\mathcal{O}_a(f)^\alpha$ with $D, \alpha > 0$ on $\mathcal{C}_d$. Nor can the exponent $3/2$ in (3) be increased.*

Proof. The matching lower moduli contradict either improvement. □

9 Uniform reference bounds and statistical consequences

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

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

The following uniformity is stronger than fixing a single reference.

Lemma 9.1 (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 (27) 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 7.2 now proves the claim, with a common observation threshold. The zero case follows from the nonvanishing Gaussian Fourier multiplier. □

We next state finite-sample consequences. Observations $Y_1, \dots, Y_n$ 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 $Q_n = n^{-1} \sum_i \delta_{Y_i}$. For clarity these results assert rates up to positive constants, not sharp leading constants or computational efficiency.

Lemma 9.2 (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 $f \in \mathcal{C}_d$,

$$P_f^{\otimes n}\{W_1(Q_n, p_f) > n^{-b/2}\} \le K n^{-b/2}. \tag{33}$$

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 (33). □

Lemma 9.3 (Indistinguishable oscillatory alternatives)**.** *At every reference in* (1)*, the family* (29) *satisfies, for all sufficiently large real* T*,*

$$\chi^2(p_{f_T} \mid q) := \int \frac{(p_{f_T} - q)^2}{q} \le KT^{-4}e^{-2\lambda T^2}, \qquad \lambda = \frac{\tau^2\sigma^2}{2(\tau^2 + \sigma^2)}.$$

Proof. Use the exact formula (32). The lower Gaussian envelope for q has variance $\sigma^2 + 1/\beta > v^2 = \sigma^2 + \tau^2$, so $\int \phi_v^2/q < \infty$. The square bracket in that formula is bounded in absolute value by $e^{-\lambda T^2} + B_T + C_T \le Ke^{-\lambda T^2}$, because $\lambda < \tau^2/2$ absorbs the correction polynomials. Multiply the resulting integral bound by $\delta_T^2 = a_0^2 T^{-4}$. □

Theorem 9.4 (Fixed-reference testing)**.** *Fix a known reference* ρ *satisfying* (1)*. Define*

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

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

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

Proof. Reject when $W_1(Q_n, p_\rho) > 2n^{-b/2}$. Its type-I error tends to zero by Lemma 9.2. Theorem 1.1 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 (33) bounds the type-II error uniformly, proving the first limit.

For the lower bound take $T^2 = \lambda^{-1} \log n$ in Lemma 9.3. 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 8.1 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. □

Theorem 9.5 (Minimax estimation on a curvature class)**.** *Assume* $0 < c < 1 < C < \infty$*. For* $r = 1, 2$ *there are constants* $0 < a_r \le A_r < \infty$*, depending only on* d, σ, c, C*, such that for all sufficiently large* n*,*

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

The infimum is over measurable probability-valued estimators from the noisy observations. The upper estimator can take values in $\mathcal{R}_{c,C}$*.*

Proof. The space of finite-second-moment probability measures with W_2 is separable (finite rationally supported measures form a dense set). Every metric subspace is separable: restrict a countable base and pick one point from each nonempty base element. Thus choose a countable W_2-dense subset $\{f_k\}$ of $\mathcal{R}_{c,C}$. Put $\eta_n = n^{-b/2}$ and choose the first k with

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

This selection is measurable: each displayed distance is a continuous function of the observations, and the infimum is countable. Convolution contracts W_1, and $W_1 \le W_2$, so density of the subset makes the infimum at most $W_1(p_f, Q_n)$ for any true $f \in \mathcal{R}_{c,C}$. The selected estimator therefore satisfies

$$W_1(p_{\widehat{f}_n}, p_f) \le 2W_1(Q_n, p_f) + \eta_n.$$

On the event $W_1(Q_n, p_f) \le \eta_n$, this is at most $3\eta_n$. Lemma 9.1, with either member serving as reference, gives $W_2(\widehat{f}_n, f) \le K(\log n)^{-3/2}$ for large n; the zero-discrepancy case is included by equality. The exceptional probability is $O(n^{-b/2})$ uniformly by (33). Any two centered isotropic laws have $W_2 \le \sqrt{2d}$, by independent coupling. This proves both upper risk bounds.

For the lower bound take the standard Gaussian reference in (29). Its Hessian is I_d, strictly inside the stated curvature interval. In the proof of Lemma 8.1, choose a_0 smaller so that the global norm of $D^2 \log(1 + \delta_T k_T)$ is at most $\min(1 - c, C - 1)/2$. Every sufficiently large T then gives $f_T \in \mathcal{R}_{c,C}$, with unchanged exact moments and W_2 separation. Set $T^2 = \lambda^{-1} \log n$ as before; the two joint laws have total variation tending to zero. If their input separation is D_n, any estimator induces a test by choosing its nearer input. A test error implies transport error at least $D_n/2$, by the triangle inequality. Consequently the larger of the two expected losses is at least $(D_n/2)^r(1 - \mathrm{TV})/2$. Since $D_n \ge K^{-1}(\log n)^{-3/2}$, both lower bounds follow. □

The minimum-distance and two-point steps are standard statistical transfers; the reference-uniform weighted inverse supplies their analytic input. Dedecker–Fischer–Michel [8, Theorem 3.1] give supersmooth W_p bounds under tail conditions, with the generic Gaussian order $(\log n)^{-1/2}$ for W_p. The numerical exponent $3/2$ is already present in the one-dimensional smooth W_1 theory of [18, Assumption S2 and Theorem 2]. Our one-dimensional curvature class satisfies that weighted Sobolev assumption at smoothness two: both ρ'' and $(x\rho)''$ have uniformly bounded L^2 norms by (8) and (9). A W_1 upper rate alone does not give the W_2 upper rate here. The smooth transport comparison in [14] is further ancestry; its compact-domain positive lower density bound is replaced here by a density-weighted argument on the whole space. Park–Kim's conference abstract [17] reports related log-concave deconvolution results in density MSE and MISE. The literature comparison uses the available theorem statements and arguments.

10 Scope

The fixed-reference theorem fixes the dimension, reference and Gaussian noise level. Lemma 9.1 makes the upper constants uniform on fixed two-sided curvature classes. The statistical results concern known fixed Gaussian noise and fixed dimension. The inputs need no symmetry, smoothness, bounded support or positive Hessian lower bound. The reference's two-sided Hessian bounds and C^2 regularity are the stated hypotheses; no higher reference derivatives or Gaussian-relative L^2 premise is used. The scalar Gaussian MMSE statement and the general reference-regret statement belong to this same inverse-deconvolution result.

AI assistance. OpenAI ChatGPT/Codex assisted with literature discovery, proof exploration and drafting. Separate AI-assisted internal checks were performed.

References

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

[2] 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.

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

[4] S. G. Bobkov, G. P. Chistyakov and F. Götze, Stability of Cramer's characterization of normal laws in information distances, preprint (2015), arXiv:1512.03571.

[5] 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.

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

[7] J. Dedecker and B. Michel, Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension, preprint (2013), arXiv:1302.6103.

[8] J. Dedecker, A. Fischer and B. Michel, Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one, *Electron. J. Stat.* **9** (2015), 234–265. arXiv:1404.0646.

[9] B. Efron, Tweedie's formula and selection bias, *J. Amer. Statist. Assoc.* **106** (2011), 1602–1614. Author version.

[10] 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.

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

[12] 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. arXiv:1004.3332.

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

[14] J. Niles-Weed and Q. Berthet, Minimax estimation of smooth densities in Wasserstein distance, preprint (2019; revised 2020), arXiv:1902.01778.

[15] 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.

[16] 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.

[17] J. Park and A. Kim, Minimax rate-optimal estimation in log-concave deconvolution, JSM Annual 2026, contributed presentation, August 6, 2026. Conference abstract.

[18] 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.

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