Minimax Wasserstein deconvolution under vanishing and singular Gaussian noise

Scott Schubr
Luminary Oddities

26 November 2026

Abstract

We determine the minimax rates for recovering a log-concave density in quadratic Wasserstein distance from observations corrupted by known Gaussian noise of variance s that may vanish or be singular. On the line, with two-sided potential curvature, the rate is $\sqrt{\log\log n/n} + (s/\log n)^{3/2}$ uniformly over bounded s; in the plane it is $\log(n)/\sqrt{n} + (s/\log(2+ns^3))^{3/2}$; in dimension $d \geq 3$ with k bounded potential derivatives it is $n^{-(k+1)/(2k+d)} + (s/\log(2+ns^{k+d/2}))^{(k+1)/2}$, and arbitrary real smoothness $\alpha \geq 2$ is covered in the critical dimensions one and two. At zero noise these are the minimax quadratic Wasserstein rates of density estimation over strongly log-concave classes on the whole space, without a density floor. This answers, for order two and these classes, the question recorded by Chewi, Niles-Weed and Rigollet of extending minimax Wasserstein density estimation beyond the compactly supported Besov classes with a density floor. For deterministic inversion, the latent modulus with any positive-rank covariance is of order $\log(1/\varepsilon)^{-3/2}$ uniformly in the rank; along a Gaussian heat ray, scaled scores increase in convex order, which gives exact linear subregimes. Finally we compute the exact Wasserstein moduli at log-concave and superquadratic tails, with leading constant $\Gamma(r+1)\zeta(r)$ times the direct tail scale.

Keywords. Wasserstein deconvolution, density estimation, minimax rates, log-concave densities, Gaussian noise, optimal transport.
MSC 2020. Primary 62G07; Secondary 62C20, 62G05, 49Q22.

Contents

1 Introduction **3**

2 The minimax transition for Wasserstein deconvolution with vanishing Gaussian noise **4**
2.1 The experiment and the minimax rate 4
2.2 Curvature and weighted distribution-function bounds 5
2.2.1 A Fourier cutoff for distribution functions 6
2.3 Upper bound for small noise 7
2.4 Upper bound for the remaining variances 8
2.5 Minimax lower bounds 9
2.5.1 The tail contribution for all variances 9
2.5.2 The inverse contribution 10
2.6 An interval of possible noise variances 10
2.7 Relation to statistical deconvolution 11

3 The planar minimax transition for Gaussian deconvolution in Wasserstein distance 11
3.1 The experiment and rate . . . 11
3.2 Curvature and integrated heat flux . . . 12
3.3 Estimating the flux for small noise . . . 13
3.3.1 Fourier bounds . . . 13
3.3.2 The estimator and its error . . . 14
3.4 The remaining variances and covariance completion . . . 15
3.5 Minimax lower bounds . . . 16
3.5.1 The direct-sampling lower bound . . . 16
3.5.2 The inverse lower bound . . . 17
3.6 Relation to empirical matching and deconvolution . . . 18

4 Sharp Wasserstein deconvolution in dimensions three and higher 18
4.1 The minimax rate . . . 18
4.2 Weighted field comparison . . . 19
4.2.1 High-pass kernels and cancellation . . . 19
4.2.2 From field errors to transport . . . 20
4.3 The adaptive estimator for small noise . . . 20
4.4 The remaining noise range . . . 22
4.5 Matching lower bounds . . . 23
4.5.1 Disjoint perturbations for the direct lower bound . . . 23
4.5.2 Noisy Gaussian packets . . . 24
4.6 Application to Brenier maps from two noisy samples . . . 24
4.7 Relation to smooth estimation and transport stability . . . 25

5 Smoothness in the critical dimensions for Wasserstein deconvolution 25
5.1 The two transition laws . . . 26
5.2 Local smoothness and transport comparisons . . . 26
5.2.1 Weighted field fitting . . . 27
5.3 Small-noise estimates at the two dimensions . . . 27
5.3.1 The localized distribution function on the line . . . 28
5.3.2 Coarse heat and fine correction in the plane . . . 28
5.3.3 Probability estimates from the fields . . . 30
5.4 Larger noise and covariance completion . . . 30
5.5 The direct and inverse lower bounds . . . 30
5.5.1 Direct lower bounds for both dimensions . . . 31
5.5.2 Gaussian oscillations and the inverse terms . . . 32
5.6 Relation to distribution and matching problems . . . 32

6 Sharp deterministic Gaussian deconvolution with singular noise 33
6.1 Model and results . . . 33
6.2 A conditional transport inequality . . . 34
6.3 Sharpness with exact marginal and moment constraints . . . 35
6.4 Identification and small eigenvalues . . . 36
6.5 A uniform range for shrinking noise . . . 37

7 Reverse heat displacement and vanishing-noise Gaussian inverse bounds 38
7.1 Observations and results . . . 38
7.2 Convex order of the heat score and displacement . . . 39
7.3 Score inverse bounds as noise vanishes . . . 41
7.4 The one-dimensional output Wasserstein modulus . . . 42

8 Exact Wasserstein moduli at log-concave tails 43
8.1 Statement . . . 43
8.2 Reference tail estimates . . . 44
8.3 Two density anchors and the survival deficit . . . 44
8.4 Integration of the quantile discrepancy . . . 45
8.5 The tangent and cutoff construction . . . 46
8.6 Gaussian smoothing . . . 47
8.6.1 Fourier comparison in the central range . . . 48
8.6.2 The complementary bounded-variance range . . . 50
8.6.3 Matching lower bounds . . . 50
8.6.4 Localized oscillations for a general reference . . . 51
8.7 Radial laws in higher dimensions . . . 52

9 Sharp Gaussian deconvolution for superquadratic potentials 52
9.1 Statement . . . 53
9.2 Moment recovery from observed transport . . . 53
9.3 From a Fourier interval to cubic transport control . . . 54
9.4 Moment cancellation supported away from the center . . . 56

1 Introduction

Recovering a distribution from Gaussian-contaminated observations is ill-posed: the Gaussian kernel damps high frequencies exponentially, and the logarithmic rates of Wasserstein deconvolution reflect this. The Luminary Research Archive article *Sharp Wasserstein stability for Gaussian deconvolution of log-concave laws* [1] proved a sharp logarithmic inverse bound under fixed noise. This article treats noise that may vanish or be singular, where the problem interpolates between direct density estimation and deconvolution. Sections 2–5 determine the minimax rates on the line, in the plane, in dimension three and higher, and at arbitrary smoothness in the critical dimensions. Sections 6–7 give the deterministic moduli for singular covariances and along heat rays. Sections 8–9 compute the exact tail moduli.

At zero noise the four minimax theorems are statements about density estimation from direct samples. Chewi, Niles-Weed and Rigollet record that, for Wasserstein distances of order $p > 1$, the minimax results of Niles-Weed and Berthet [8] are confined to densities in Besov classes, with a density bounded below needed for the smooth rate, and that extending the arguments of Niles-Weed and Berthet to other classes of densities is an open question [2, §2.9, p. 84]. The following consequence gives such an extension for $p = 2$.

Corollary 1.1 (Direct sampling)**.** *Fix $0 < m < M < \infty$, an integer $k \geq 2$ and the bounds M_j, $3 \leq j \leq k$, of Section 4 (for $d \geq 3$), or the smoothness $\alpha \geq 2$ and its bounds of Section 5. Take zero noise, $s = 0$ or $\Sigma = 0$, so that the observations are n independent draws from the unknown density f, and take the infimum over measurable probability estimators with finite second moment.*

For $q \in \{1,2\}$ and all sufficiently large n, with constants depending only on these parameters and on d, the minimax risks of Theorems 2.1, 3.1 and 4.1 are of order

$$\sqrt{\frac{\log\log n}{n}} \quad (d=1), \qquad \frac{\log n}{\sqrt{n}} \quad (d=2), \qquad n^{-(k+1)/(2k+d)} \quad (d \geq 3),$$

over the respective classes of densities $f = e^{-V}$ on $\mathbb{R}^d$ with $mI \preceq D^2V \preceq MI$, and, when $d \geq 3$, bounded derivatives of V of orders $3, \ldots, k$. By Theorem 5.1, the same orders hold in dimensions one and two over the classes of every real potential smoothness $\alpha \geq 2$.

Proof. Each of the four theorems holds uniformly over noise levels that include zero: $0 \leq s \leq \Lambda$ in Theorem 2.1 and every covariance $\Sigma \succeq 0$ with $\|\Sigma\|_{\mathrm{op}} \leq \Lambda$ in the other three. At $s = 0$ every inverse term vanishes, since each is a positive power of s divided by a logarithm that is at least $\log 2$. The direct-sampling orders remain. □

Corollary 1.1 answers the question of [2, §2.9] for the quadratic Wasserstein distance and the strongly log-concave classes above. These classes are smooth, so the new feature relative to the compact theory is the setting: the whole space, without a density floor, where Corollary 1.1 determines the minimax order with matching lower bounds. In dimensions one and two the order is set by the Gaussian-type tails and does not improve with smoothness; in dimension $d \geq 3$ the exponent $(k+1)/(2k+d)$ of the compact theory with a density floor persists. This second edition adds Corollary 1.1 and this answer, and states in Section 6 the one fixed-noise inverse bound that its singular-noise upper bounds use; the other results and proofs of the first edition are unchanged.

2 The minimax transition for Wasserstein deconvolution with vanishing Gaussian noise

We estimate a one-dimensional distribution in quadratic Wasserstein distance from independent observations with known additive Gaussian noise. Over densities with fixed strict upper and lower log-curvature bounds, the minimax risk is of order $\sqrt{\log\log n/n} + (s/\log n)^{3/2}$, uniformly over bounded noise variance s, including zero. A family of tail perturbations forces the iterated logarithm even for direct samples. For small noise, the estimator adds a spatially localized deconvolution correction to the empirical distribution function, then projects onto the curvature class using the candidate density as a weight. This preserves the direct-sampling variance order while attaining the inverse-problem term. If the variance is known only to lie in an interval, its width adds an identification term.

2.1 The experiment and the minimax rate

Fix $0 < m < M < \infty$ and $0 < \Lambda < \infty$. Let $\mathcal{F}$ be the class of probability densities $f = e^{-V}$ on $\mathbb{R}$ with $V \in C^2(\mathbb{R})$ and $m \leq V'' \leq M$. The class includes all translations. We observe $Y_i = X_i + \sqrt{s}\, Z_i$, with independent X_i of density f and iid standard normals Z_i independent of the latent sample. The known variance is $s \in [0, \Lambda]$. Write γ_s for the centered Gaussian law of variance s, and $p_s = f * \gamma_s$ for the observed density, with $p_0 = f$. An estimator is a measurable probability law with finite second moment. For $q \in \{1,2\}$ define the risks

$$\mathcal{R}_{n,q}(s) = \inf_{\widehat{\mu}} \sup_{f \in \mathcal{F}} \bigl(\mathbb{E}_{f,s} W_2(\widehat{\mu}, f)^q\bigr)^{1/q}, \qquad r_n = \sqrt{\frac{\log\log n}{n}}.$$

Here f also denotes its probability measure in W_2.

Theorem 2.1. *There are positive constants c, C, n_0 depending only on m, M, Λ such that for $n \geq n_0$ and $0 \leq s \leq \Lambda$,*

$$c\left[r_n + \left(\frac{s}{\log n}\right)^{3/2}\right] \leq \mathcal{R}_{n,1}(s) \leq \mathcal{R}_{n,2}(s) \leq C\left[r_n + \left(\frac{s}{\log n}\right)^{3/2}\right].$$

In particular the transition between the two terms is at noise variance of order $\log n\,(\log\log n/n)^{1/3}$.

The fixed-positive-noise order for this class follows from the statistical theorem in [1, Theorem 2.5]. The Fourier attenuation argument for the inverse lower bound below is a scalar specialization of that work. At $s = 0$, the Gaussian empirical W_2 loss has the iterated logarithm [4, Corollary 6.14]; showing that it is unavoidable for arbitrary estimators on the curvature class requires a minimax lower bound. On the unit interval with a positive density floor, the one-dimensional upper bound for the Besov classes of [8, Theorem 1] has parametric order. Here the tail contribution and the Gaussian inversion contribution can both be retained uniformly in one experiment.

2.2 Curvature and weighted distribution-function bounds

We use constants that may depend on m, M, Λ. Write m_f for the mean of f, F for its distribution function, P_s for that of p_s, and $\tau_F(x) = \min(F(x), 1 - F(x))$. The same notation for a different density uses its own mean and distribution function.

Lemma 2.2 (Uniform envelopes)**.** *The following estimates hold for both f and p_s. With $u = x - m_f$ and q equal to either density, there are uniform constants such that*

$$c_0 e^{-C_0 u^2} \leq q(x) \leq C_0 e^{-c_0 u^2}, \tag{1}$$

$$|(\log q)'(x)| \leq C_0(1 + |u|), \qquad |q''(x)| \leq C_0(1 + u^2)q(x), \tag{2}$$

$$q(x) \asymp (1 + |u|)\tau_Q(x), \tag{3}$$

where Q is the distribution function of q. If $x \geq m_f + C_0$, then $\int_x^\infty (y-x)^2 q(y)\,dy \leq C_0 q(x)/(1+|u|)^3$. The reflected bound holds to the left of $m_f - C_0$. Also $p_s(x)/f(x)$ is bounded below uniformly and is bounded above uniformly whenever $s(1 + u^2)$ is bounded.

Proof. For $s > 0$, the conditional potential of X given $Y = x$ has curvature between $m + s^{-1}$ and $M + s^{-1}$. If U has a density with potential A and mean $\bar{u}$, integration by parts gives $\mathbb{E}[(U - \bar{u})(A'(U) - A'(\bar{u}))] = 1$. When $a \leq A'' \leq b$, this gives $b^{-1} \leq \mathrm{Var}(U) \leq a^{-1}$. Differentiating the Gaussian convolution gives

$$-(\log p_s)''(x) = s^{-1} - s^{-2}\,\mathrm{Var}(X \mid Y = x),$$

so the curvature of p_s lies between $m/(1 + ms)$ and $M/(1 + Ms)$. Both densities have mean m_f.

For a density with curvature in a fixed positive compact interval, the potential is bounded by two quadratics about its minimum. Normalization bounds the density at that minimum above and below and makes the difference between the mode and mean uniformly bounded. The two quadratics give (1); the curvature bound and vanishing derivative at the mode give (2). To the right of the mode by a distance bounded below, the potential slope is comparable to that distance. Integrating the exponential of the tangent line bounds the right tail above by the density divided by that slope. Integrating over a further interval of length comparable to the reciprocal distance gives a matching lower bound by the upper curvature bound. The same tangent line bound with a squared displacement gives the stated second-moment estimate. Reflection gives the left estimates,

and the compact central region satisfies (3) by (1). Finally, the quadratic Taylor bounds on $\log f(x+v)-\log f(x)$ have form $(\log f)'(x)v-av^2/2$ for $a=m,M$. Their exponentials have integrals against γ_s equal to

$$(1+as)^{-1/2}\exp\left(\frac{s((\log f)'(x))^2}{2(1+as)}\right).$$

This proves the ratio assertions, including $s=0$. □

If $I=[a,b]$, denote by μ^I the image of a law under clipping to I. The distribution function of μ^I on the interior of I is that of μ.

Lemma 2.3 (Clipped transport)**.** *Let μ,ν be laws supported on I, possibly with atoms at the endpoints, with distribution functions F,G. If μ has positive interior density f, then*

$$W_2^2(\mu,\nu)\le 4\int_I \frac{(F-G)^2}{f}\,dx.$$

If ν also has interior density g, then

$$W_2^2(\mu,\nu)\le \pi^2\int_I \frac{(F-G)^2}{f+g}\,dx.$$

Proof. Use the continuity equation for $\rho_t=(1-\theta(t))\mu+\theta(t)\nu$ with flux $-\theta'(t)(G-F)\,dx$ on I, extended by zero outside. The distributional derivative of the extended difference $G-F$ includes the changes in endpoint atoms, so the continuity equation holds on $\mathbb{R}$. For $\theta=2t-t^2$, the interior density is at least $(1-\theta)f$ and $\theta'^2/(1-\theta)=4$. For $\theta=\sin^2(\pi t/2)$, use the lower bound $\min(\theta,1-\theta)(f+g)$ and $\theta'^2/\min(\theta,1-\theta)\le\pi^2$. The flux is absolutely continuous with respect to ρ_t where it is nonzero for $0<t<1$. The dynamic transport bound [3] gives both inequalities. □

2.2.1 A Fourier cutoff for distribution functions

Use $\widehat{K}(v)=\int e^{-ivx}K(x)\,dx$, with $\widehat{K}$ real, even, smooth, compactly supported on $[-1,1]$ and equal to one near zero. Thus K is Schwartz, of mass one and satisfying $\int xK(x)\,dx=\int x^2K(x)\,dx=0$. Set $K_h(x)=h^{-1}K(x/h)$. On a window $|x-m_f|\le L$, put $w=(1+L)^{-1}$. For $0<h\le w$, Taylor expansion to degree two, with splitting at $|v|=w$, gives for every fixed integer $N>0$

$$|(K_h*F)(x)-F(x)|\le Ch^3(1+|x-m_f|^2)f(x)+C_N(h/w)^N. \tag{4}$$

Indeed, on the short shifts, the values of f are comparable to $f(x)$ by (2), and the third derivative of F is f''. The first two Taylor terms integrate to zero over the full kernel. On $|v|>w$ bound F and the subtracted quadratic Taylor polynomial by $C(1+|v|^2)$ and use arbitrary Schwartz decay. This gives (4) without any logarithmic loss in its first term.

Let L_A be the inverse Fourier transform of $\widehat{K}(v)e^{Av^2/2}$, where $A=s/h^2$, and set

$$D_{s,h}(u)=\int_{-\infty}^{u/h} L_A(v)\,dv-\mathbf{1}_{\{u\ge 0\}}.$$

The associated deconvolution density kernel is $h^{-1}L_A(\cdot/h)$. Its convolution with p_s equals K_h*f, by the Gaussian Fourier multiplier. Integrating this identity from $-\infty$ and using Fubini gives

$$P_s(x)+\int D_{s,h}(x-y)p_s(y)\,dy=(K_h*F)(x). \tag{5}$$

The Fourier transform of the unscaled primitive difference is $(\widehat{K}(v)e^{Av^2/2}-1)/(iv)$, interpreted at zero by continuity. Plancherel, using the first-order bound in $A|v|$ near zero and $|v|^{-1}$ outside the support, gives

$$\int |D_{s,h}(u)|^2\,du \le Ch(1+A)^2e^{CA}. \tag{6}$$

For $|u|/h \ge 1$, repeated Fourier integration by parts on the smooth compact multiplier and integration of the resulting kernel tail give, for every fixed $N>0$,

$$|D_{s,h}(u)| \le C_N(1+A)^{N+2}e^{CA}(1+|u|/h)^{-N}. \tag{7}$$

The constant C in the exponential may be chosen independent of N, since the multiplier support is fixed.

2.3 Upper bound for small noise

Fix a small $\eta>0$, to be chosen in Section 2.4. In this section $0\le s\le n^{-\eta}$. Choose

$$h=\max\left(n^{-1/6},\sqrt{\frac{s}{\gamma\log n}}\right), \tag{8}$$

with a sufficiently small fixed $\gamma>0$ depending on η. Thus $A=s/h^2\le\gamma\log n$, and $h\le n^{-d}$ for some fixed $d>0$ and all large n.

Split the sample into two parts of size comparable to n. Let $\widehat{m}$ be the mean of the first part and put

$$L_n=B\sqrt{\log n},\qquad I=[\widehat{m}-L_n,\widehat{m}+L_n],\qquad w_n=(2+L_n)^{-1}.$$

Choose the constant B large enough that the squared cost of clipping any candidate centered within one of $\widehat{m}$ to I is $O(n^{-2})$, by (1). On the event $\mathcal{A}=\{|\widehat{m}-m_f|<1/2\}$, this also applies to f. The complement has exponentially small probability, uniformly over f,s, by the Gaussian envelope in Lemma 2.2 and the Chernoff bound for a sample mean. Indeed a centered variable with this envelope has a moment generating function bounded by e^{Ct^2} for $|t|\le1$; the proof follows by Taylor expansion under the envelope and uses only a fixed nonzero t for the displayed event.

With Y_i from the second part of size N_n, set

$$\widehat{F}(x)=\frac{1}{N_n}\sum_i\left[\mathbf{1}_{\{Y_i\le x\}}+D_{s,h}(x-Y_i)\mathbf{1}_{\{|x-Y_i|\le w_n\}}\right]. \tag{9}$$

It is a signed distribution-function estimate. The physical truncation of the correction allows the variance to be measured against the local density. On $\mathcal{A}$, I lies within $\{|x-m_f|\le L_n+1\}$, where $f\ge n^{-C_B}$. On this window and its w_n buffer, $p_s(x)/f(x)\asymp1$ by Lemma 2.2, since $s\log n\to0$. For $|v|\le w_n$ the score bound gives $p_s(x-v)\le Cf(x)$.

We bound the conditional weighted mean square error against the true f, given the first part on $\mathcal{A}$. By (7), $h\le n^{-d}$, and $A\le\gamma\log n$, the truncation changes the expectation in (5) by at most n^{-J} uniformly in x, for any prescribed J and all large n, by choosing the fixed decay power large enough. By (4) the square bias integrated against $1/f$ on I is $O(h^6+n^{-2})$. In that estimate, the first term in (4) reduces to a bounded fourth moment, and the remainder is negligible by the polynomial lower density bound and arbitrary Schwartz decay. For the raw empirical term the integrated variance is bounded by

$$\frac{C}{n}\int_I\frac{P_s(x)(1-P_s(x))}{f(x)}\,dx\le\frac{C}{n}\int_{-L_n-1}^{L_n+1}\frac{dv}{1+|v|}\le C\frac{\log\log n}{n}.$$

This uses the output hazard and the density comparison. The integral of the correction variance is bounded by

$$\frac{C|I|}{n}\int |D_{s,h}(u)|^2\,du \le \frac{C|I|h}{n}(1+A)^2e^{CA} = o(n^{-1}).$$

For the last equality choose γ with $C\gamma < d/2$; for example any $d < \min(1/6, \eta/2)$ is available for all large n. The variance of the sum is at most twice the sum of the two variance bounds. Thus

$$\mathbb{E}\left[\int_I \frac{(\widehat{F}-F)^2}{f}\,dx \;\middle|\; \text{first part}\right] \le C(h^6 + r_n^2) \quad \text{on } \mathcal{A}. \tag{10}$$

To obtain a probability estimator, let $\mathcal{D}$ be a fixed countable dense subset of $\mathcal{F}$ in the joint topology of W_2 and locally uniform convergence of densities. Such a subset exists because both ambient spaces are separable and metrizable with countable bases. Among the candidates $g \in \mathcal{D}$ with $|m_g - \widehat{m}| < 1$, choose the first g whose loss

$$\ell(g) = \int_I \frac{(\widehat{F}-G)^2}{g(x)}\,dx$$

is within n^{-2} of the infimum. Use a strict tolerance in the selection so that an index always exists. The feasible subset is nonempty by density and availability of translations, and the losses are measurable by integration. On $\mathcal{A}$, the true f is strictly feasible in the mean constraint. Approximating f in the stated topology shows that the infimum is at most $\ell(f)$: on the compact interval, the positive densities converge uniformly and their distribution functions converge uniformly, by weak convergence to the continuous F. Therefore

$$\int_I \frac{(F-G)^2}{f+g}\,dx \le 2\ell(f) + 2\ell(g) \le 4\ell(f) + 2n^{-2}.$$

Use Lemma 2.3 for the two clipped laws and then the tail clipping bounds. With (10) this gives squared risk $O(h^6 + r_n^2)$ on $\mathcal{A}$. On the complement, the feasible mean constraint and uniformly bounded centered second moments give $W_2(f,g)^2 \le C(1+|\widehat{m}-m_f|^2)$. The complement contributes $O(n^{-1})$, using its probability and the unconditional variance of $\widehat{m}$. Consequently the estimator gives the desired upper bound for $s \le n^{-\eta}$.

2.4 Upper bound for the remaining variances

A direct two-density comparison suffices when $s \ge n^{-\eta}$. Let $f, g \in \mathcal{F}$, and write $\epsilon = W_1(f * \gamma_s, g * \gamma_s) \le 1$. Their means differ by at most ϵ, since convolution preserves means and the identity is a 1-Lipschitz test. Choose a window I_h centered at m_f of radius $L_h = B_0\sqrt{\log(1/h)}$, for $h \in (0, h_0)$ sufficiently small. By choosing B_0 large, (1) gives squared tail clipping costs $O(h^6)$ for both densities. Throughout I_h, each density is bounded below by h^P for some fixed P. With the window enlarged by a constant to encompass the possible mean difference, (4) gives

$$\left(\int_{I_h} \frac{(F - K_h * F)^2}{f}\right)^{1/2} + \left(\int_{I_h} \frac{(G - K_h * G)^2}{g}\right)^{1/2} \le Ch^3.$$

For example the short-shift term is paid by the fourth moment, and $(h(2+L_h))^N$ pays the remaining weighted norm for a fixed N sufficiently large relative to P.

For the smoothed difference, its Fourier multiplier is the output characteristic-function difference times $\widehat{K}(hv)e^{sv^2/2}/(iv)$. To justify the primitive identity, $F - G \in L^1(\mathbb{R})$ by the one-dimensional

W_1 formula and the signed derivative is $f - g$. Coupling the outputs gives the absolute bound $\epsilon|v|$ on their characteristic-function difference. Fourier inversion over $|v| \le h^{-1}$ hence yields

$$\|K_h * (F - G)\|_\infty \le C\epsilon h^{-1} e^{s/(2h^2)}.$$

Use Lemma 2.3 on I_h, putting the two bias terms over their own densities and the smoothed term over $f + g$. Restoring the tails gives

$$W_2(f, g) \le C\left(h^3 + \epsilon h^{-1-P/2} \log^{1/4}(1/h) e^{s/(2h^2)}\right). \tag{11}$$

To apply this estimate, let $\widehat{Q}_n$ be the empirical observed law and $e_n = W_1(\widehat{Q}_n, f * \gamma_s)$. The scalar CDF formula and (1) give

$$\mathbb{E}e_n \le n^{-1/2} \int_{\mathbb{R}} \sqrt{P_s(x)(1 - P_s(x))}\, dx \le Cn^{-1/2}.$$

Use a countable W_2-dense subset of the mean-zero shapes in $\mathcal{F}$, translating each shape to have mean $\overline{Y}$, the sample mean. Select the first candidate whose noisy law fits $\widehat{Q}_n$ in W_1 to within n^{-1} of the countable infimum. This is measurable and the fit can use the true shape translated to $\overline{Y}$, by continuity and density. The translation costs at most e_n. Thus the noisy W_1 distance from the selected candidate to the true law is at most $3e_n + n^{-1}$. On $\{e_n \le n^{-1/8}\}$, put

$$h = \max\left(n^{-a}, \sqrt{\frac{s}{\gamma_0 \log n}}\right)$$

in (11). Choose $a, \gamma_0 > 0$ so small that $a(4 + P/2) + \gamma_0/2 < 1/8$. Then the second term of (11) is $O(h^3)$ uniformly. If $0 < \eta < \min(a, 1/8)$ and $s \ge n^{-\eta}$, the second term in this definition of h dominates the first for large n, so $h^3 \asymp (s/\log n)^{3/2}$. The exceptional event has probability $O(n^{-3/8})$ by Markov's inequality. Mean anchoring and uniform centered variances bound its contribution to squared risk by $O(n^{-3/8} + n^{-1})$, which is $o((s/\log n)^3)$ with the same η. This completes the uniform upper bound.

2.5 Minimax lower bounds

2.5.1 The tail contribution for all variances

Let g be a centered Gaussian with curvature strictly between m and M. Fix a nonzero smooth function ψ supported in $(0, 1)$ with integral zero. For $j \ge 1$, let $x_j = \sqrt{j}$, $w_j = (\sqrt{j+1} - \sqrt{j})/2$ and $I_j = [x_j, x_j + w_j]$. These intervals are disjoint. On I_j, g is comparable to $g_j = g(x_j)$ by boundedness of $x_j w_j$. For $J = \lfloor b \log n \rfloor$, where $b > 0$ is small, and for signs $\omega_j \in \{-1, 1\}$, set

$$f_\omega(x) = g(x) + \sum_{j=1}^{J} \omega_j a_j \psi((x - x_j)/w_j), \qquad a_j = d\sqrt{\frac{g_j}{nw_j}},$$

with $d > 0$ fixed and small. Each bump has zero mass. Relative to g, its derivatives of order $k = 0, 1, 2$ are bounded by $Ca_j/(g_j w_j^k)$, using $x_j w_j = O(1)$ to differentiate the Gaussian ratio. The upper bound for $k = 2$ is at most $Cd(ng_j w_j^5)^{-1/2}$, which tends to zero uniformly over $j \le J$ for b sufficiently small. Hence, for all large n, the densities are between $g/2$ and $3g/2$ and the log-curvature perturbation is smaller than the Gaussian curvature margin. All f_ω lie in $\mathcal{F}$.

For adjacent sign vectors, only one interval contributes to the squared density difference. The one-sample chi-square divergence in either direction is bounded by $Ca_j^2 w_j/g_j = Cd^2/n$. Adding

independent Gaussian noise decreases chi-square divergence, by conditional Jensen, including the identity at $s = 0$. The product chi-square formula and a fixed small d make total variation between the corresponding noisy sample laws at most $1/2$, uniformly over n and s.

All densities have the same mass and the same endpoint CDF on each I_j. Their monotone transport therefore couples mass within each interval, with the remaining identical mass coupled to itself. On a differing interval the two local measures of mass $m_j \asymp g_j w_j$ have first-power transport cost

$$2a_j w_j^2 \int_0^1 \left| \int_0^v \psi(u)\, du \right| dv.$$

This is the integral of the absolute CDF difference on the interval. Cauchy–Schwarz under the mass-m_j transport plan gives a squared W_2 contribution at least $ca_j^2 w_j^3/g_j \asymp w_j^2/n$, with constants allowed to depend on the fixed d, ψ. Adding the disjoint contributions,

$$W_2(f_\omega, f_{\omega'})^2 \geq c_1 \sum_{j:\omega_j \neq \omega'_j} w_j^2/n. \tag{12}$$

For any probability estimator, decode it to a nearest vertex in W_2, with a fixed tie order. The triangle inequality bounds its loss below by half the distance from the true to the decoded vertex. Under uniform random signs, the probability of a decoding error for each bit is at least $1/4$, by testing adjacent pairs and their total variation bound. Let Z be the sum of w_j^2/n over incorrectly decoded bits and let $S = \sum_j w_j^2/n$. Then $0 \leq Z \leq S$, $\mathbb{E}Z \geq S/4$, and $\mathbb{E}\sqrt{Z} \geq \mathbb{E}Z/\sqrt{S}$. Since $w_j^2 \asymp 1/j$ and $J \asymp \log n$, (12) gives a lower bound cr_n even for expected W_2 loss.

2.5.2 The inverse contribution

Use a centered Gaussian g of variance σ_0^2 with curvature in the same interior and let

$$f_\xi(x) = g(x)\left[1 + \frac{\kappa}{\xi^2}\left(\cos(\xi x) - e^{-\sigma_0^2\xi^2/2}\right)\right].$$

This has mass one. For ξ sufficiently large and a fixed small $\kappa > 0$, the first two log derivatives of the bracket change curvature by at most $C\kappa$, so $f_\xi \in \mathcal{F}$. Testing with the 1-Lipschitz function $x \mapsto \cos(\xi x)/\xi$ gives $W_2(f_\xi, g) \geq W_1(f_\xi, g) \geq c\xi^{-3}$, since $\mathrm{Var}_g(\cos(\xi X))$ stays bounded below.

Under the Gaussian reference and noise variance $s > 0$, the conditional variance of X given Y is $v_s = \sigma_0^2 s/(\sigma_0^2 + s)$. The conditional cosine has absolute value at most $e^{-v_s\xi^2/2}$ and is centered by its unconditional Gaussian expectation in the perturbation. Therefore

$$\chi^2(f_\xi * \gamma_s \mid g * \gamma_s) \leq \frac{\kappa^2}{\xi^4} e^{-v_s\xi^2}.$$

For $s \leq \Lambda$, $v_s \geq c_\Lambda s$. Choose $\xi^2 = A_0 \log n/s$ with a fixed large A_0. The product chi-square divergence tends to zero uniformly over $0 < s \leq \Lambda$. Testing these two alternatives then bounds expected W_2 risk below by $c(s/\log n)^{3/2}$. Indeed, decode an estimate to the nearer of the two laws and apply their testing lower bound and the triangle inequality. The maximum of this bound and the tail bound is comparable to their sum, completing the proof of Theorem 2.1.

2.6 An interval of possible noise variances

When the variance is unknown but known to lie in an interval, the allowed width gives a separate identification term. Write $\mathcal{R}_{n,q}([s_0, s_0 + \Delta])$ for the same risk with an additional supremum over

that interval and an estimator depending only on the interval. Matias [9] studies an unknown Gaussian variance under a no-Gaussian-component assumption on the signal. The curvature class here permits Gaussian components.

Corollary 2.4. *Uniformly over $0 \le s_0 \le s_0 + \Delta \le \Lambda$ and $q = 1, 2$, the minimax risk for all sufficiently large n has order*

$$r_n + \left(\frac{s_0}{\log n}\right)^{3/2} + \Delta.$$

Proof. If s is the actual variance, write an observation as a sample from $f * \gamma_{s-s_0}$ plus independent Gaussian noise of variance s_0. Lemma 2.2 puts this smoothed latent density in the two-curvature class with parameters $m/(1 + m\Lambda)$ and M. The heat path $f_t = f * \gamma_t$ has velocity $-(\log f_t)'/2$ in the continuity equation. Its squared velocity has uniformly bounded integral against f_t by (1)–(2). Rescaling this path to unit time in the dynamic transport bound gives $W_2(f, f * \gamma_{s-s_0}) \le C(s - s_0)$. Apply Theorem 2.1 at the known variance s_0 to the smoothed class and use the triangle inequality.

For the lower bound fix $s = s_0$ to obtain the two terms of Theorem 2.1. For the interval term, take two centered latent Gaussians of variances v and $v - \Delta'$ in the interior of $[1/M, 1/m]$, with $\Delta' = \min(\Delta, \Delta_*)$ for a fixed sufficiently small $\Delta_* > 0$. At noise variances s_0 and $s_0 + \Delta'$, respectively, the observed laws coincide. The latent W_2 distance is $\sqrt{v} - \sqrt{v - \Delta'} \asymp \Delta'$. Since $\Delta \le \Lambda$, this is bounded below by a constant times Δ. Identical data and the triangle inequality give the claim in either loss. □

2.7 Relation to statistical deconvolution

Wasserstein minimax deconvolution with fixed supersmooth error was studied in [6] over moment classes. Distribution-function and quantile deconvolution estimators are studied in [7]. Delaigle [5] treats the joint limit of growing sample size and shrinking errors in density estimation. Park and Kim [10] report ordinary- and supersmooth density-risk rates under log-concavity in their conference abstract. Here curvature controls a weighted transport norm for the latent density, and the weighted variance of the empirical CDF must be retained. Only the short-range correction in (9) carries the Gaussian inverse multiplier. The tail hypercube shows why a parametric direct-sampling bound fails for the full translated curvature class, even with known zero Gaussian variance. The result is a rate comparison, with constants uniform in s on the stated bounded interval.

3 The planar minimax transition for Gaussian deconvolution in Wasserstein distance

We determine the minimax quadratic Wasserstein rate for recovering a planar density with two-sided potential-Hessian bounds from independent samples contaminated by known Gaussian noise. Uniformly over bounded, possibly singular noise covariances with largest eigenvalue s, the rate is $(\log n)/\sqrt{n} + (s/\log(2 + ns^3))^{3/2}$. A multiscale wavelet construction over a growing Gaussian window forces the direct-sampling term. Gaussian packets of width $\sqrt{s}$ give the effective inverse logarithm. For small noise, we estimate a localized Fourier approximation to the integrated heat flux and use it to transport an independent part of the sample. The two terms balance at variance of order $n^{-1/3}(\log n)^{2/3} \log\log n$.

3.1 The experiment and rate

Fix $0 < m < M < \infty$ and $0 < \Lambda < \infty$. Let $\mathcal{F}$ be the probability densities $f = e^{-V}$ on $\mathbb{R}^2$ with $V \in C^2(\mathbb{R}^2)$ and $mI \preceq D^2V \preceq MI$. Means and covariances of f are unknown. We observe independent $Y_i = X_i + Z_i$,

where the X_i have density f and the Z_i are centered Gaussian with known common covariance $\Sigma \succeq 0$, independent of the latent sample. Write $\gamma_\Sigma = N(0,\Sigma)$ and $\gamma_t = N(0,tI_2)$. Put $s = \|\Sigma\|_{\mathrm{op}}$ and suppose $s \le \Lambda$. For $q = 1, 2$, write

$$\mathcal{R}_{n,q}(\Sigma) = \inf_{\widehat{\mu}} \sup_{f\in\mathcal{F}} \big(\mathbb{E}_{f,\Sigma} W_2(\widehat{\mu}, f)^q\big)^{1/q}$$

over measurable probability estimators with finite second moment, and put $r_n = (\log n)/\sqrt{n}$ and $B_n(s) = \log(2 + ns^3)$.

Theorem 3.1. *There are $c, C > 0$ and n_0 depending only on m, M, Λ such that for all $n \ge n_0$ and all the indicated covariances,*

$$c\left[r_n + \left(\frac{s}{B_n(s)}\right)^{3/2}\right] \le \mathcal{R}_{n,1}(\Sigma) \le \mathcal{R}_{n,2}(\Sigma) \le C\left[r_n + \left(\frac{s}{B_n(s)}\right)^{3/2}\right].$$

The transition in the largest noise variance occurs at order $n^{-1/3}(\log n)^{2/3}\log\log n$.

The fixed-positive-noise statistical theorem in [1, Theorem 2.5] gives order $(\log n)^{-3/2}$ in the stated class, with constants allowed to depend on a fixed covariance. The logarithm in Theorem 3.1 retains information about a covariance tending to zero. In the noiseless case the empirical distribution has squared risk $O((\log n)^2/n)$ by planar Gaussian matching [15, 13] and Gaussian contraction. The lower construction shows that no probability estimator improves this order uniformly over the curvature class.

3.2 Curvature and integrated heat flux

Let m_f be the mean of f. We write $p_t = f * \gamma_t$ for the density after isotropic Gaussian noise of variance t, and $p_0 = f$. From the two Hessian bounds, these laws have uniformly Gaussian upper and lower envelopes about m_f:

$$ce^{-C|x-m_f|^2} \le p_t(x) \le Ce^{-c|x-m_f|^2}. \tag{13}$$

Their score and density Hessian satisfy

$$|\nabla \log p_t(x)| \le C(1 + |x - m_f|), \qquad \|D^2 p_t(x)\| \le C(1 + |x - m_f|^2)p_t(x), \tag{14}$$

and their scores are C-Lipschitz, uniformly for $t \in [0, \Lambda]$. Here and below the constants can depend on m, M, Λ. To verify these facts for $t > 0$, the conditional potential of X given $X + \sqrt{t}Z = x$ has Hessian between $(m + t^{-1})I$ and $(M + t^{-1})I$. The Brascamp–Lieb inequality [11] gives the upper covariance bound $(m + t^{-1})^{-1}I$. For the lower bound, fix a unit vector v. If A is the conditional potential, then integration by parts gives $\mathbb{E}[(v \cdot (X - \mathbb{E}X))(v \cdot \nabla A)] = 1$ and $\mathbb{E}[(v \cdot \nabla A)^2] = \mathbb{E}[v^T D^2 A\, v]$. Cauchy–Schwarz yields the lower covariance bound $(M + t^{-1})^{-1}I$. Differentiating the convolution gives

$$-D^2 \log p_t = t^{-1}I - t^{-2}\operatorname{Cov}(X \mid X + \sqrt{t}Z = x),$$

so its eigenvalues lie between $m/(1 + mt)$ and $M/(1 + Mt)$. Normalization and quadratic bounds around the mode give (13); the mean stays within bounded distance of the mode by these same quadratic bounds. Bounds on the score and the density Hessian then give (14).

We will also use, for $0 \le t \le s \le \Lambda$,

$$p_t(x) \le Cp_s(x), \qquad |p_s(x) - p_t(x)| \le C(s - t)(1 + |x - m_f|^2)p_s(x). \tag{15}$$

The first bound follows by integrating the lower quadratic Taylor bound on $\log p_t(x + z) - \log p_t(x)$ against γ_{s-t}. Its integral is at least $1/(1+M(s-t))$ in two dimensions. The second bound then follows by integrating $\partial_t p_t = \Delta p_t/2$ and using (14). The formula extends to $t = 0$ by continuity of the classical heat evolution for these C^2 densities.

Define the flux and the associated map

$$J_s(x) = \tfrac{1}{2}\int_0^s \nabla p_t(x)\, dt, \qquad U_s(x) = x + J_s(x)/p_s(x). \tag{16}$$

The flux satisfies $\operatorname{div} J_s = p_s - f$ and $|J_s/p_s|(x) \le Cs(1 + |x - m_f|)$.

Lemma 3.2 (Transport by the flux)**.** *There is an exact map U_s^* sending p_s to f such that*

$$\left\|U_s^* - U_s\right\|_{L^2(p_s)} \le Cs^2.$$

Proof. The vector field $-\frac{1}{2}\nabla \log p_t$ is uniformly Lipschitz in space with at most linear growth. Its flow X_t started from f has marginal p_t, by the continuity equation and uniqueness along the flow. The flow maps are invertible and the inverse from time s to zero is U_s^*. The mean square norm of the velocity is bounded uniformly by (14) and (13), whence $\|X_s - X_t\|_2 \le C(s-t)$. Using the Lipschitz bound on the score, replace $\nabla \log p_t(X_t)$ by $\nabla \log p_t(X_s)$ in $X_0 - X_s = \frac{1}{2}\int_0^s \nabla \log p_t(X_t)\,dt$. The L^2 error is at most Cs^2. Next multiply the latter score by $p_t(X_s)/p_s(X_s)$. The integrated error is at most Cs^2 by (15) and the bounded sixth moment of $X_s - m_f$. The resulting expression is $J_s(X_s)/p_s(X_s)$. □

For later use, if $X_1, \dots, X_N$ are iid f,

$$\mathbb{E}W_2^2(N^{-1}\textstyle\sum_i \delta_{X_i}, f) \le C(\log N)^2/N. \tag{17}$$

Indeed Caffarelli's contraction theorem [12] gives an $m^{-1/2}$-Lipschitz map from a standard Gaussian to f. Apply the map simultaneously to the Gaussian sample and reference measure, and use the planar Gaussian empirical matching bound [15, 13].

3.3 Estimating the flux for small noise

We first assume $\Sigma = sI$ and $s \le n^{-\eta}$, for a fixed $\eta > 0$ chosen below. For $s = 0$ the empirical estimate (17) suffices. For $s > 0$ take

$$h = \max\left(n^{-1/6}(\log n)^{1/3}, \sqrt{\frac{s}{\gamma B_n(s)}}\right), \qquad A = s/h^2, \tag{18}$$

where $\gamma > 0$ is a sufficiently small constant. In this range h is bounded above by a negative power of n, and $A \le \gamma B_n(s)$.

3.3.1 Fourier bounds

Use $\widehat{u}(\xi) = \int e^{-ix\cdot\xi}u(x)\,dx$. Let $\widehat{K}$ be a real even smooth cutoff supported in the unit ball and equal to one near zero. Write K_h for the density scaling of its inverse transform. By (16),

$$\widehat{J}_s(\xi) = \frac{i\xi}{|\xi|^2}(\widehat{f} - \widehat{p}_s)(\xi);$$

at zero the expression is interpreted by the integrated heat formula. The regularized flux $K_h * J_s$ is therefore a convolution of p_s with the kernel

$$L_{s,h}(x) = h^{-1}\mathcal{F}^{-1}\left[\widehat{K}(v)\frac{iv}{|v|^2}(e^{A|v|^2/2} - 1)\right](x/h).$$

The symbol is smooth at zero by the integral representation

$$\frac{e^{A|v|^2/2} - 1}{|v|^2} = \frac{A}{2}\int_0^1 e^{tA|v|^2/2}\,dt.$$

Plancherel and integration by parts give

$$\int |L_{s,h}(x)|^2\,dx \le C(1+A)^2 e^{CA}, \tag{19}$$

$$|L_{s,h}(x)| \le C_N h^{-1}(1+A)^{C_N}e^{CA}(1+|x|/h)^{-N} \tag{20}$$

for every fixed $N > 0$. The exponential constant is independent of N because the Fourier support is fixed. The power of h in (19) is zero in two dimensions.

To bound the bias, let Q be the inverse transform of $-iv(1-\widehat{K}(v))/|v|^4$. This symbol vanishes near zero, is of order -3 at infinity, and is integrable together with every derivative. Thus Q is bounded with arbitrarily fast decay at infinity. Set $Q_h(x) = h^{-2}Q(x/h)$. We have

$$J_s - K_h * J_s = h^3 Q_h * \Delta(f - p_s). \tag{21}$$

This follows by multiplying the Fourier symbols. For all uses below, if $|x - m_f| \le L$ put $w = (2+L)^{-1}$. On short shifts of size at most w, (14) makes $p_s(x-u)$ comparable to $p_s(x)$. Equations (14) and (15) then bound the contribution from short shifts in (21) by $Ch^3(1+|x-m_f|^2)p_s(x)$. On the remaining shifts, the global density Hessian bounds and decay of Q give $C_N h^3 (h/w)^N$, for any fixed N. The same short-shift estimate, with p_s replaced by f or any $g \in \mathcal{F}$, applies to $h^3 Q_h * \Delta f$ or $h^3 Q_h * \Delta g$.

3.3.2 The estimator and its error

Split the sample into four independent parts of sizes comparable to n. From the first part compute its mean $\widehat{m}$, and use the ball $I = \{x : |x - \widehat{m}| \le L_n\}$ with $L_n = D\sqrt{\log n}$ for sufficiently large D. Put $w_n = (3+L_n)^{-1}$. Let $\mathcal{A} = \{|\widehat{m} - m_f| \le 1\}$. Its complement has exponentially small probability, by (13) and the Chernoff bound on each coordinate of the sample mean. On $\mathcal{A}$, the density p_s is at least n^{-P} on I for a sufficiently large fixed P, and $p_s(x-u) \le C p_s(x)$ whenever $x \in I,\ |u| \le w_n$. The bound on mean and all estimates are invariant under translations of the experiment.

Average $L_{s,h}(x - Y_i)\mathbf{1}_{\{|x-Y_i| \le w_n\}}$ over the second part to obtain $\widehat{J}(x)$. By (20) the truncation changes its expectation by $O(n^{-D_0})$ uniformly in x, for every prescribed $D_0 > 0$. Here h^{-1} is bounded by a fixed power of n from (18), and h/w_n is bounded by a negative power of n, so one fixed large N suffices. The quadratic bias from (21) after division by p_s and integration on I is $O(h^6)$, using the Gaussian moments for the short shifts and an arbitrarily large decay power for the rest. For the variance, use $p_s(x-u) \le C p_s(x)$ on the retained shifts, then integrate (19) over the $O(\log n)$ area of I. Thus, conditionally on the first part on $\mathcal{A}$,

$$\mathbb{E}\int_I \frac{|\widehat{J} - J_s|^2}{p_s}\,dx \le C\left(h^6 + \frac{\log n}{n}(1+A)^2 e^{CA}\right). \tag{22}$$

The second term is $O(h^6 + r_n^2)$. Indeed let $X = 2 + ns^3$. Enlarge a fixed constant C_1 so that $(1+A)^2 e^{CA} \le CX^{C_1\gamma}$. If $X \le (\log n)^3$, choose $C_1\gamma < 1/3$ and use $r_n^2 = (\log n)^2/n$. If $X > (\log n)^3$, then $\log n \le X^{1/3}$ and $X^{1/3 + C_1\gamma} \le C(X-2)/(\log X)^3$, which is paid by the second entry of (18) to the sixth power.

Estimate p_s using the third part by a compactly supported nonnegative symmetric probability kernel of bandwidth $b = n^{-1/6}(\log n)^{1/3}$. Its estimate $\widehat{p}$ has pointwise squared bias at most $Cb^4(1+|x-m_f|^2)^2 p_s(x)^2$ on I, by the vanishing first moment, (14) and short-shift density comparison. Its variance is bounded by $Cp_s(x)/(nb^2)$ there. The radius b is smaller than w_n for large n. Consequently, on the same good first-part event,

$$\mathbb{E}\int_I (1+|x-\widehat{m}|)^2 \frac{|\widehat{p} - p_s|^2}{p_s}\,dx \le C\left(b^4 + \frac{(\log n)^2}{nb^2}\right) = O(n^{-2/3}(\log n)^{4/3}). \tag{23}$$

Inside I, define $\widehat{d}(x)$ by projecting $\widehat{J}(x)/\max(\widehat{p}(x), n^{-P})$ onto the Euclidean ball of radius $C_* s(1+|x-\widehat{m}|)$ centered at zero. Outside I put $\widehat{d} = 0$. Choose C_* so that $d(x) = J_s(x)/p_s(x)$ is in the projection ball on $\mathcal{A}$. Writing $p^\dagger = \max(\widehat{p}, n^{-P})$, we have $|p^\dagger - p_s| \le |\widehat{p} - p_s|$ on I on $\mathcal{A}$. Where $p^\dagger \ge p_s/2$, nonexpansiveness of projection and the ratio formula give

$$|\widehat{d} - d|^2 \le C\frac{|\widehat{J} - J_s|^2 + |d|^2|\widehat{p} - p_s|^2}{p_s^2}.$$

Where $p^\dagger < p_s/2$, the two displacements have difference at most twice the radius, which is at most four times that radius times $|p^\dagger - p_s|/p_s$. Integrating both bounds gives

$$\mathbb{E}\|\widehat{d} - d\|^2_{L^2(p_s)} \le C\big(h^6 + r_n^2 + s^2 n^{-2/3}(\log n)^{4/3}\big).$$

Here the exterior of I on $\mathcal{A}$ contributes $O(n^{-2})$ with large enough D, by the Gaussian envelope and the linear bound on d. On $\mathcal{A}^c$, the clipping bounds the integrated loss by $Cs^2(1+|\widehat{m} - m_f|^2)$; Gaussian moments

and the exponentially small probability make its expected contribution negligible. The square root of the denominator contribution is lower order: for $s \le n^{-1/3}$ compare it with r_n; for $s > n^{-1/3}$ compare it with $s^{3/2}/B_n(s)^{3/2}$ using $B_n(s) = O(\log n)$. Also $s^2 = o(s^{3/2}/B_n(s)^{3/2})$ uniformly when $0 < s \le n^{-\eta}$.

For the final part of the sample, output the empirical law of $Y_i + \widehat{d}(Y_i)$. Conditional on training, $U_s^*(Y_i)$ are independent latent draws of density f. Pair the two empirical laws by their sample indices, and apply the triangle inequality, Lemma 3.2 and (17). This gives root mean square loss $O(h^3 + r_n)$, the asserted rate for this range. The paired comparison uses the exact heat flow only in the proof.

3.4 The remaining variances and covariance completion

For two densities $f, g \in \mathcal{F}$, let $\epsilon = W_1(f * \gamma_s, g * \gamma_s) \le 1$ and $\rho = f + g$. The means of f and g differ by at most ϵ. We first show that for all sufficiently small $h > 0$,

$$W_2(f, g) \le C \left(h^3 + \epsilon h^{-2-P/2} \sqrt{\log(1/h)}\, e^{s/(2h^2)} \right) \tag{24}$$

with $P > 0$ depending only on the class and a fixed cutoff. This comparison is used for the two-curvature class on both sides.

The mixture ρ, of mass two, has a uniform Poincaré inequality for functions of zero ρ-mean. Indeed each density has Poincaré constant at most m^{-1}, by Brascamp–Lieb. A fixed unit ball B about m_f has a uniform positive density floor for both f and g. The squared difference between the average of a function on B and its average under either density is bounded by a constant times the corresponding variance, by Jensen and that floor. Using this common ball average as centering shows

$$\int |\phi - \bar{\phi}_\rho|^2 \rho \le C \int |\nabla \phi|^2 \rho, \qquad \bar{\phi}_\rho = \tfrac{1}{2} \int \phi \rho.$$

Write $d = f - g$ and let $L = D_0 \sqrt{\log(1/h)}$ with D_0 sufficiently large. For a test function ϕ, subtract its ρ-average and then split it by a smooth cutoff χ that is one on $B_L(m_f)$, supported on $B_{L+1}(m_f)$, with bounded gradient. The exterior bound is

$$\left| \int d(1 - \chi)(\phi - \bar{\phi}_\rho) \right| \le C h^3 \left(\int |\nabla \phi|^2 \rho \right)^{1/2},$$

by the Gaussian envelope and mixture Poincaré. In the interior use a flux for d with symbol $-i\xi \widehat{d}(\xi)/|\xi|^2$, split by $\widehat{K}(h\xi)$. The high-frequency component is, up to sign, $h^3 Q_h * \Delta(f - g)$. On B_{L+1} its $L^2(1/\rho)$ norm is $O(h^3)$, by the short- and long-shift estimates of (21), applying the short-shift bounds to each density separately. The far remainders are negligible since $\rho \ge h^P$ there for a fixed P. For the low-frequency component, the difference of observed characteristic functions is bounded by $\epsilon|\xi|$ by coupling. Removing the known Gaussian characteristic function and integrating over $|\xi| \le h^{-1}$ gives the uniform pointwise bound $C\epsilon h^{-2} e^{s/(2h^2)}$. Multiplying by the square root of the area and the inverse density floor gives the second term in (24) for its $L^2(1/\rho)$ norm on B_{L+1}.

Integrate the flux against $\nabla[\chi(\phi - \bar{\phi}_\rho)]$; its $L^2(\rho)$ norm is at most $C\|\nabla\phi\|_{L^2(\rho)}$ by Poincaré. Thus the full linear form $\phi \mapsto -\int d\phi$ is bounded by the right side of (24) times $\|\nabla\phi\|_{L^2(\rho)}$. By Hilbert space representation on the closure of these gradients there is an exact flux J with $\operatorname{div} J = d$ and weighted norm $\|J/\sqrt{\rho}\|_2$ bounded by the same expression. Interpolate between f and g by $(1-\theta)f + \theta g$ with $\theta(t) = \sin^2(\pi t/2)$ and flux $\theta'(t)J$. Its action is at most $\pi^2 \|J/\sqrt{\rho}\|_2^2$, since the interpolating density is at least $\min(\theta, 1-\theta)\rho$. The dynamic transport bound [3] proves (24).

For the statistical estimator in this range, let e_n be W_1 between the empirical observed law and its population law. By the same contraction reasoning as (17), now for p_s, $\mathbb{E}e_n \le Cr_n$. Fit a countable W_2-dense set of mean-zero shapes in $\mathcal{F}$, shifted to have the observed sample mean, to the empirical observed law in Gaussian-convolved W_1. Specifically choose the first whose convolved discrepancy is strictly less than its countable infimum plus n^{-1}. The fit is measurable and has convolved distance to the truth at most $3e_n + n^{-1}$, because the true distribution translated to the sample mean is available in the limit and its translation costs at most e_n. On $e_n \le n^{-1/8}$, use

$$h = \max\left(n^{-a}, \sqrt{\frac{s}{\gamma_0 \log n}} \right)$$

in (24). Choose $a, \gamma_0 > 0$ small enough that $a(5 + P/2) + \gamma_0/2 < 1/8$. Then the second term there is $O(h^3)$. For $0 < \eta < \min(a, 1/8)$ and $s \geq n^{-\eta}$, the second bandwidth entry dominates, and $B_n(s) \asymp \log n$. The excluded event has probability $O((\log n)n^{-3/8})$. Mean anchoring and uniform centered second moments bound its contribution to squared loss by $O((\log n)n^{-3/8} + n^{-1})$. This is negligible relative to $(s/B_n(s))^3$ for this choice of η. It completes the isotropic upper bound.

For arbitrary known $\Sigma \succeq 0$, add independent Gaussian noise of covariance $sI - \Sigma$ to each observation and use the isotropic procedure. This has the same risk as for sI. For a deterministic probability estimator in the original data, output the mixture of those conditional probability outputs over the artificially added randomness. The mixture has finite second moment almost surely, and squared W_2 is convex in its first probability argument, by mixing couplings. Consequently its squared risk is no larger. This covers all covariances in the theorem, including singular ones.

3.5 Minimax lower bounds

We use two families of alternatives. For clarity, consider a finite sign family f_ω for which adjacent sample laws have total variation at most $1/2$, and suppose

$$W_2(f_\omega, f_{\omega'})^2 \geq c_0 \sum_{j:\omega_j \neq \omega'_j} v_j, \qquad v_j \geq 0. \tag{25}$$

Decoding an arbitrary probability estimate to its nearest vertex loses a factor of at most two in the metric by the triangle inequality. Under uniform signs each bit error has probability at least $1/4$ by testing adjacent laws. If Z is the total erroneous weight and $S = \sum_j v_j$, then $\mathbb{E}Z \geq S/4$ and $\sqrt{Z} \geq Z/\sqrt{S}$. Hence (25) gives an expected W_2 lower bound of order $\sqrt{S}$. Independently adding noise only decreases chi-square divergence, so direct-data adjacency estimates apply at every covariance.

We also use a simple coupling inequality. If a smooth test function φ has bounded Hessian, then Taylor expansion in any coupling of f_ω and $f_{\omega'}$ gives

$$\int \varphi(f_\omega - f_{\omega'}) \leq \|\nabla\varphi\|_{L^2(f_\omega)} W_2(f_\omega, f_{\omega'}) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(f_\omega, f_{\omega'})^2. \tag{26}$$

This follows by applying Cauchy–Schwarz under an optimal coupling. In the two constructions we find a nonnegative quantity E for which the left side is at least cE, the gradient norm is at most $C\sqrt{E}$, and the Hessian is uniformly bounded. Solving the quadratic inequality gives $W_2^2 \geq c'E$.

3.5.1 The direct-sampling lower bound

Let g be a centered isotropic Gaussian with curvature in the interior of $[m, M]$. Pack disjoint axis-aligned squares B of side $\ell = (\log n)^{-1/2}$ in a square about zero of side $c_1\sqrt{\log n}$, with $c_1 > 0$ sufficiently small. Their total area is comparable to $\log n$. On each square $g \asymp g_B$ where g_B is its value at the center, and $g_B \geq n^{-\beta}$ for some fixed $0 < \beta < 1$. Both statements follow from the Gaussian exponent and $\sup_{x\in B} |x|\ell = O(1)$.

Use one orientation of a compactly supported C^2 orthonormal tensor wavelet family [14], dilated and translated to each square. At a level of physical length $r = \ell 2^{-j}$ retain the wavelets supported strictly inside B. Starting at a sufficiently large fixed j, there are comparable to $(\ell/r)^2$ such orthonormal functions at that level. Stop at $j = \lfloor \kappa \log_2 n \rfloor$ with $\kappa > 0$ small, so the smallest physical length is comparable to $\ell n^{-\kappa}$. The support condition makes functions from different squares orthogonal. At each level at most a bounded number overlap at a point. Each wavelet ψ_i has zero integral, L^2 norm one, and derivatives of order $a = 0, 1, 2$ bounded in sup norm by Cr_i^{-1-a}, where r_i is its physical length.

Perturb the Gaussian by independently signed wavelets,

$$f_\omega = g + \sum_i \omega_i a_{B(i)} \psi_i, \qquad a_B = d_1\sqrt{g_B/n}$$

with a fixed small $d_1 > 0$. The relative perturbations through derivative order two are bounded by a constant times $\max_B(a_B/g_B)r_{\min}^{-3}$, using finite overlap and summing the geometric physical scales. Derivatives of $1/g$ through order two cost at most $C\ell^{-2}/g_B$ or the corresponding first-order bound on B, and $r_i \leq \ell$. Choose

$\kappa < (1-\beta)/6$, so this bound tends to zero. Each density has mass one, lies between $g/2$ and $3g/2$, and satisfies the curvature bounds for all sufficiently large n. For adjacent vertices the one-sample chi-square distance is at most $Ca_B^2/g_B = Cd_1^2/n$, by orthonormal normalization. The product chi-square formula makes sample total variation at most $1/2$ for small enough d_1, with or without noise.

For a pair of vertices write $c_i = (\omega_i - \omega_i')a_{B(i)}$. Use the test $\varphi = \sum_i (r_i^2/g_{B(i)})c_i\psi_i$. By orthonormality the pairing is exactly

$$E = \sum_i (r_i^2/g_{B(i)})c_i^2.$$

The gradient norm squared under f_ω is at most CE. Here is an elementary justification of the needed wavelet bound. Within one square, for a single level sum u_j, bounded overlap gives $\|\Delta u_j\|_2 \le Cr_j^{-2}\|u_j\|_2$; the latter norm is the Euclidean norm of its coefficients. For $r_j \ge r_k$, integration by parts gives

$$|\langle \nabla u_j, \nabla u_k\rangle| \le C\frac{r_k}{r_j}\left(\frac{\|u_j\|_2}{r_j}\right)\left(\frac{\|u_k\|_2}{r_k}\right).$$

The geometric relative scale factors are summable, giving $\|\sum_j \nabla u_j\|_2^2 \le C\sum_j r_j^{-2}\|u_j\|_2^2$. Since $f_\omega \asymp g_B$ in each square, the weighted gradient estimate follows by summing over squares. The Hessian of the test is bounded by $C\max_B(a_B/g_B)r_{\min}^{-1}$ using finite overlap and summability, hence uniformly bounded. Equation (26) proves (25) with weights comparable to r_i^2/n.

Each square contributes $\asymp \ell^2/n$ per level to the sum of these weights. There are order $\log n$ levels and total square area of order $\log n$, so $S \asymp (\log n)^2/n$. The sign-family argument gives a lower bound cr_n at every covariance.

3.5.2 The inverse lower bound

Rotate the axes so that the noise covariance is $\mathrm{diag}(s,t)$ with $0 \le t \le s$. First suppose $0 < s \le s_0$ with a sufficiently small fixed $s_0 > 0$. Use a centered isotropic Gaussian reference g of curvature strictly between m and M. Take grid centers z_j in a fixed square about zero with spacing $D\sqrt{s}$, for a large fixed D. By reducing s_0 after choosing D, the number of centers is comparable to $1/s$. Put $T^2 = A_0B_n(s)/s$, for a large constant A_0, and

$$\psi_j(x) = e^{-|x-z_j|^2/(2s)}\left[\cos(T(x_1 - z_{j,1})) - e^{-sT^2/2}\right], \qquad f_\omega = g + a\sum_j \omega_j\psi_j, \qquad a = \kappa_0/T^2.$$

The subtraction sets the integral of each packet to zero. The sum of the relative derivatives through order $k \le 2$ is bounded in absolute value by CT^k before multiplication by a. Indeed each derivative is bounded by CT^k times a weaker Gaussian envelope $e^{-c|x-z_j|^2/s}$ after division by g: the bandwidth derivative factor $s^{-1/2}$ is at most CT, polynomials are absorbed in the weaker envelope, and the center stays in a fixed square. The relative derivatives of $1/g$ are handled by the same envelope for small s_0. The sum of the envelopes on the separated grid is bounded uniformly. Thus, with small fixed κ_0, the densities have mass one, are between $g/2$ and $3g/2$, and belong to $\mathcal{F}$.

Convolving a packet with the known noise gives exactly

$$(\psi_j * \gamma_\Sigma)(y) = \frac{s}{\sqrt{2s(s+t)}}\, e^{-(y_1-z_{j,1})^2/(4s) - (y_2-z_{j,2})^2/(2(s+t))}\left[e^{-sT^2/4}\cos\!\left(\tfrac{T}{2}(y_1 - z_{j,1})\right) - e^{-sT^2/2}\right].$$

For the first coordinate this is the conditional characteristic function of a Gaussian of variance s given noise of variance s, centered at $z_{j,1}$; the second coordinate is ordinary Gaussian convolution and allows $t = 0$. The noisy reference $g * \gamma_\Sigma$ has a uniform Gaussian lower envelope. Dividing the square of the displayed packet by that reference and integrating gives Cse^{-csT^2}, since s_0 is small and the centers are in a fixed square. The noisy vertex densities stay above half the noisy reference. Consequently adjacent noisy one-sample chi-square divergence is at most $Ca^2se^{-csT^2}$. Writing $X = ns^3$, its n-fold testing parameter is bounded by

$$C\kappa_0^2 X\,(A_0\log(2+X))^{-2}(2+X)^{-cA_0}.$$

For a fixed sufficiently large A_0 and small κ_0, the corresponding sample laws have total variation at most $1/2$ uniformly over the range.

For the metric separation, write $c_j = (\omega_j - \omega'_j)a$ and use $\varphi = T^{-2}\sum_j c_j\psi_j$. The L^2 Gram matrix of the packets has diagonal at least $c_1 s$ when $T\sqrt{s}$ is sufficiently large: rescale by $\sqrt{s}$ and integrate the cosine square against the Gaussian, using the exponentially small subtraction. Its off-diagonal entries are bounded by $Cs\exp(-c|z_j - z_k|^2/s)$ by overlap of the envelopes. Taking D large makes each off-diagonal row sum less than half the lower diagonal bound. Hence

$$\int \varphi(f_\omega - f_{\omega'}) \geq c_2 E, \qquad E = T^{-2}s\sum_j c_j^2.$$

The gradient Gram matrix in unweighted L^2 has row sums in absolute value at most $CT^2 s$, by the derivative envelope. Since f_ω is bounded above, this gives $\|\nabla\varphi\|^2_{L^2(f_\omega)} \leq CE$. The Hessian sup norm is bounded by Ca by the pointwise derivative and overlap estimates. Applying (26) gives $W_2(f_\omega, f_{\omega'})^2 \geq cE$. Thus all differing bits have weights comparable to $T^{-6}s$. With order $1/s$ bits, the sign-testing argument gives the expected W_2 lower bound cT^{-3}, that is, $c(s/B_n(s))^{3/2}$.

For $s_0 \leq s \leq \Lambda$, perturb the same Gaussian by the single relative factor $1+\kappa_0 T^{-2}(\cos(Tx_1) - \mathbb{E}_g\cos(TX_1))$ with $T^2 = C\log n/s$. Its curvature remains in the class by the same derivative bound. Testing with $\cos(Tx_1)/T$ gives a latent W_1, and hence W_2, separation at least cT^{-3}. Conditional Gaussian attenuation in the first coordinate gives noisy one-sample chi-square divergence at most $CT^{-4}e^{-csT^2}$, uniformly in the second eigenvalue and for bounded s. The product total variation tends to zero for large C. This two-point test gives the same order since $B_n(s) \asymp \log n$ on this range. The direct lower bound also holds at $s = 0$.

Taking the maximum of the two lower bounds, which is equivalent to their sum up to constants, completes the proof of Theorem 3.1. Equating their squares gives $X/\log^3(2+X) \asymp (\log n)^2$, where $X = ns^3$; this yields the stated transition scale.

3.6 Relation to empirical matching and deconvolution

Gaussian empirical matching [15, 13] supplies the sampling upper bound through Gaussian contraction. In the compact positive-density setting, related Wasserstein and optimal-map risks are studied in [8, 16]. The direct minimax lower bound above uses both a growing tail area and order $\log n$ orthogonal scales per square. For the noisy problem, [6] considers fixed supersmooth errors over moment classes, and [5] studies jointly shrinking noise and growing sample size under density loss. Park and Kim [10] report density-risk rates under log-concavity in their conference abstract. The fixed-covariance theorem of [1] has a different role from the uniform transition: the packet width $\sqrt{s}$ and the local variance of the flux kernel are needed to determine which logarithm enters near the direct-sampling regime. The theorem concerns rates uniformly over the stated class, without claiming sharp leading constants.

4 Sharp Wasserstein deconvolution in dimensions three and higher

We determine the minimax quadratic Wasserstein rate for estimating a smooth strongly log-concave law on $\mathbb{R}^d$, $d \geq 3$, from independent observations with known additive Gaussian noise. If the potential has two-sided Hessian bounds and bounded derivatives of orders three through k, for an integer $k \geq 2$, the rate is $n^{-(k+1)/(2k+d)} + (s/\log(2 + ns^{k+d/2}))^{(k+1)/2}$, uniformly over bounded, possibly singular noise covariances of largest eigenvalue s. The estimator fits a coarse density and a localized divergence field. Its bandwidth increases in the tails, making the weighted variance integrable without logarithmic losses. Gaussian packets and disjoint smooth perturbations give matching lower bounds. The result also gives the sharp rate for estimating a Brenier map from two independent noisy samples.

4.1 The minimax rate

Fix integers $d \geq 3$, $k \geq 2$, bounds $0 < m < M < \infty$, and, if $k > 2$, constants $M_j > 0$ for $3 \leq j \leq k$. Let $\mathcal{F}_k$ consist of densities $f = e^{-V}$ on $\mathbb{R}^d$ with $V \in C^k(\mathbb{R}^d)$, $mI \preceq D^2V \preceq MI$, and $\|D^jV\|_\infty \leq M_j$ for those j.

The class allows all translations. Observe independent $Y_i = X_i + Z_i$, with iid X_i of density f and iid centered Gaussian Z_i independent of the latent sample. Their known covariance is $\Sigma \succeq 0$ with $s = \|\Sigma\|_{\mathrm{op}} \le \Lambda$, for fixed $\Lambda < \infty$. Write $\gamma_\Sigma = N(0, \Sigma)$ and $\gamma_t = N(0, tI_d)$. For $q \in \{1, 2\}$, use

$$\mathcal{R}_{n,q}(\Sigma) = \inf_{\widehat{\mu}} \sup_{f \in \mathcal{F}_k} \left(\mathbb{E}_{f,\Sigma} W_2(\widehat{\mu}, f)^q\right)^{1/q}$$

with the infimum over measurable probability estimators of finite second moment. Set $a_n = n^{-(k+1)/(2k+d)}$ and $B_n(s) = \log(2 + ns^{k+d/2})$.

Theorem 4.1. *There are $c, C, n_0 > 0$, depending only on the fixed parameters, such that for $n \ge n_0$ and every indicated covariance,*

$$c\left[a_n + \left(\frac{s}{B_n(s)}\right)^{(k+1)/2}\right] \le \mathcal{R}_{n,1}(\Sigma) \le \mathcal{R}_{n,2}(\Sigma) \le C\left[a_n + \left(\frac{s}{B_n(s)}\right)^{(k+1)/2}\right].$$

In particular the two orders balance at $s \asymp n^{-2/(2k+d)}$.

On a fixed cube with a uniform positive density floor, the direct exponent is the one in the smooth Wasserstein minimax theory of [8]. That work also treats an unbounded sub-Gaussian Hölder class of order less than one. Here the density floor deteriorates in the tails, and the result retains the exact direct order for the unbounded curvature class. The fixed-noise result of [1, Theorem 2.5] gives order $(\log n)^{-3/2}$ for $k = 2$. Theorem 4.1 resolves the vanishing-noise range and tracks higher bounded derivatives.

4.2 Weighted field comparison

We collect the deterministic estimates for the upper bound. Write $m_f = \int x f(x)\, dx$. By Taylor's theorem at the mode, the curvature bounds give Gaussian upper and lower envelopes, uniformly after translation by m_f:

$$ce^{-C|x-m_f|^2} \le f(x) \le Ce^{-c|x-m_f|^2}. \tag{27}$$

Indeed normalization bounds the density at the mode, and its distance to the mean is bounded by the resulting Gaussian tail bound. Also $|\nabla V(x)| \le C(1 + |x - m_f|)$, so the product and chain rules give, for multiindices $|\alpha| \le k$,

$$|D^\alpha f(x)| \le C(1 + |x - m_f|^{|\alpha|}) f(x). \tag{28}$$

These include global sup norm bounds for the density and each of these derivatives. If $p_s = f * \gamma_s$, integrating the upper quadratic bound for $\log f(x + z) - \log f(x)$ yields

$$p_s(x) \le Cf(x) \exp(Cs(1 + |x - m_f|^2)). \tag{29}$$

For example the exact integral is bounded by $\exp(s|\nabla V(x)|^2/2)$ using the tangent plane, or one can retain the lower Hessian term in that Gaussian integral. The observed laws have uniformly Gaussian tails about the same mean since $s \le \Lambda$.

4.2.1 High-pass kernels and cancellation

Let $\widehat{K}$ be real, even, smooth, supported in the unit ball and equal to one on $\{|\xi| \le c_K\}$ for some $c_K > 0$. We use the convention $\widehat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\, dx$. Write $K_b(x) = b^{-d} K(x/b)$, where K is its inverse Fourier transform. Let Q be the inverse transform of

$$q(\xi) = -\frac{i\xi}{|\xi|^2}(1 - \widehat{K}(\xi)).$$

This is a spatially integrable vector kernel with all absolute moments finite and every signed monomial moment zero. To see this, split the symbol away from zero into smooth dyadic annuli $|\xi| \asymp 2^j$ with $j \ge 0$, allowing a fixed finite low-frequency transition. The spatial kernel of each annulus is bounded

by $C_N 2^{(d-1)j}(1+2^j|x|)^{-N}$ for any N, by rescaling and integration by parts. Its L^1 norm is $O(2^{-j})$. Summing gives the integrable kernel and arbitrary integrated tail decay, hence all absolute moments. The signed moments vanish by differentiating its transform at the origin, where q vanishes on a neighborhood.

Put $Q_b(x) = b^{-d}Q(x/b)$ and define

$$U_b(f) = K_b * f, \qquad J_b(f) = bQ_b * f.$$

The identity $f = U_b(f) + \operatorname{div} J_b(f)$ holds in distributions. For $|x - m_f| \le L$ and $w = (2+L)^{-1}$, the moments of Q and Taylor's formula through degree $k-1$ give for $b \le w$ and each $N > 0$

$$|J_b(f)(x)| \le Cb^{k+1}(1 + |x - m_f|^k)f(x) + C_N b(b/w)^N. \tag{30}$$

For shifts $|u| \le w$, the density along the Taylor segment is comparable to $f(x)$ by the score bound, and (28) controls the order-k remainder. Farther shifts, including the subtracted Taylor polynomial, are paid by the global derivative bounds and the arbitrary weighted tail decay of Q. In particular, this argument needs integrability of Q in physical space. It does not require the order--1 symbol to be integrable in the frequency variable.

4.2.2 From field errors to transport

Let $f, g \in \mathcal{F}_k$ have their means within a fixed distance. Put $\rho = f + g$. The mixture has a uniform Poincaré inequality

$$\int |\phi - \bar{\phi}_\rho|^2 \rho \le C \int |\nabla \phi|^2 \rho, \qquad \bar{\phi}_\rho = \tfrac{1}{2}\int \phi\rho. \tag{31}$$

Each summand has a Poincaré inequality by Brascamp–Lieb [11]. A unit ball about m_f has a positive density floor for both densities by (27). The squared difference between either component mean of ϕ and its ordinary average over the ball is bounded by a constant times that component variance. Use that ball average as a common centering and then minimize over constants to prove (31).

Suppose I is a ball of radius $L \ge 2$, centered within a fixed distance of both means. With I^- its concentric ball of radius $L-1$, for every $b > 0$ we have

$$\begin{aligned} W_2(f,g) \le C \Bigg[&\left(\int_{\mathbb{R}^d \setminus I^-} \rho \right)^{1/2} \\ &+ \left(\int_I \frac{|U_b(f) - U_b(g)|^2 + |J_b(f) - J_b(g)|^2}{\rho} \right)^{1/2} \Bigg]. \end{aligned} \tag{32}$$

Indeed split $\phi - \bar{\phi}_\rho$ by a smooth cutoff equal to one on I^-, supported in I, with uniformly bounded gradient. In the exterior, $|f - g| \le \rho$ and (31) pay the first term times $\|\nabla\phi\|_{L^2(\rho)}$. In the interior insert $f - g = U_b(f) - U_b(g) + \operatorname{div}(J_b(f) - J_b(g))$. Cauchy–Schwarz and (31) bound the cutoff-regularized test and its gradient in $L^2(\rho)$ by that same gradient norm. Consequently the functional $\nabla\phi \mapsto -\int (f-g)\phi$ is bounded on the closure of smooth compactly supported gradients in $L^2(\rho)$. The representing vector u gives an exact flux $J = \rho u$ with $\operatorname{div} J = f - g$ and norm $\|J/\sqrt{\rho}\|_2 = \|u\|_{L^2(\rho)}$ bounded by the displayed right-hand side. The density curve $(1-\theta)f + \theta g$ with flux $\theta' J$ and $\theta(t) = \sin^2(\pi t/2)$ solves the continuity equation. Its action is at most $\pi^2 \|J/\sqrt{\rho}\|_2^2$, using $(1-\theta)f + \theta g \ge \min(\theta, 1-\theta)\rho$. The dynamic W_2 inequality [3] proves (32).

4.3 The adaptive estimator for small noise

For now let $\Sigma = sI_d$ and $0 \le s \le n^{-\eta}$, where the fixed $\eta > 0$ is chosen sufficiently small. Split the sample into two comparable parts. Let $\widehat{m}$ be the first-part mean, and $I = B_{L_n}(\widehat{m})$ with $L_n = D\sqrt{\log n}$ for a sufficiently large fixed D. Put $w_n = (3 + L_n)^{-1}$. The event $\mathcal{A} = \{|\widehat{m} - m_f| < 1/2\}$ has complement of exponentially small probability, by the uniform Gaussian envelope of the observed law and the Chernoff bound on each coordinate. On $\mathcal{A}$, $f \ge n^{-P}$ on I with P fixed, and for $x \in I$, $|u| \le w_n$,

$$p_s(x - u) \le Cf(x). \tag{33}$$

This follows from (29), since $s \log n = o(1)$, and bounded variation of $\log f$ on such a shift by the score bound.

Set, for $x \in I$,

$$b = n^{-\eta/8}, \qquad h_0 = \max\left(n^{-1/(2k+d)}, \sqrt{\frac{s}{\gamma B_n(s)}}\right), \qquad h(x) = h_0 e^{a|x-\widehat{m}|^2}. \tag{34}$$

Choose η small enough that $\eta/2 < 1/(2k+d)$. Then $h_0 \le Cn^{-\eta/2}$ and $h_0 \ge n^{-1/(2k+d)}$. Choose $a > 0$ small so that $aD^2 < \eta/8$ and the exponent $(2k+2)a$ is less than half of a valid Gaussian upper-envelope exponent in (27). It follows that $h(x) \ll b$ uniformly on I and $A(x) = s/h(x)^2 \le s/h_0^2 \le \gamma B_n(s)$. The constant $\gamma > 0$ is fixed and small.

For each x, form a scalar and a vector sample average using kernels with Fourier multipliers

$$\widehat{K}(b\xi)e^{s|\xi|^2/2}, \qquad -\frac{i\xi}{|\xi|^2}\big(\widehat{K}(h(x)\xi) - \widehat{K}(b\xi)\big)e^{s|\xi|^2/2}, \tag{35}$$

respectively, applied to the second-part observations. In both, the real-space kernel is set to zero outside $|u| \le w_n$ before evaluation at $u = x - Y_i$. Denote the results by $\widehat{U}(x)$ and $\widehat{J}(x)$. The varying bandwidth is used only at the evaluation point. The ensuing field comparison will use the constant-bandwidth operators U_b, J_b on each candidate density.

The scalar multiplier without physical truncation gives population field $U_b(f)$. For the vector multiplier that field is $(K_{h(x)} * J_b(f))(x)$, because $\widehat{K}(h(x)\xi)\widehat{K}(b\xi) = \widehat{K}(b\xi)$ for all $x \in I$ when n is large. The same plateau identity shows that

$$J_b(f)(x) - (K_{h(x)} * J_b(f))(x) = J_{h(x)}(f)(x).$$

In this identity $h(x)$ is held fixed in the convolution. For the mean errors caused by physical truncation, both kernels are smooth band-limited before truncation, and the vector multiplier vanishes on $|\xi| \le c_K/b$. We give the quantitative bound. Write $h = h(x)$, $A = s/h^2$. On the vector support $c_K/b \le |\xi| \le 1/h$, each derivative of its rational factor of order j is bounded by $Cb^j|\xi|^{-1}$. The cutoffs and the exponential cost at most b^j times a fixed polynomial in $1 + A$ per total derivative order, in addition to $e^{A/2}$. Thus the L^1 norm of derivatives of total order N is at most

$$C_N b^N h^{1-d}(1 + A)^{C_N} e^{CA}.$$

The same argument for the scalar multiplier gives $C_N b^{N-d}(1+s/b^2)^{C_N} e^{Cs/b^2}$. Directional Fourier integration by parts bounds each kernel uniformly outside $|u| \le w_n$ by the corresponding bound times w_n^{-N}. Since b is polynomially small, h is bounded below by a fixed power of n, and $A = O(\log n)$, for every prescribed $F > 0$ the truncation changes both expectations by $O(n^{-F})$ uniformly on I, by taking a sufficiently large fixed N.

The remaining vector bias is bounded by (30). The integral of its square weighted by $1/f$ is $O(h_0^{2k+2})$ on the good first-part event. Indeed the short-shift term gives

$$C \int_I h(x)^{2k+2}(1 + |x - m_f|^{2k}) f(x)\, dx \le C h_0^{2k+2}$$

by the choice of a. The far term is negligible by the polynomial density floor and arbitrary integrated kernel tail. There is no remaining scalar bias.

For the variances, (33) means that the integral of the squared truncated kernel against the data density, divided by $f(x)$, is at most a constant times its unweighted L^2 norm squared. The scalar kernel has norm squared at most $Cb^{-d}e^{Cs/b^2}$ by Plancherel. Thus its integrated variance is $O(n^{-1}(\log n)^{d/2}b^{-d})$, since $s/b^2 \to 0$. We choose

$$\eta d/8 < 1 - (2k+2)/(2k+d). \tag{36}$$

Then the scalar variance is $o(n^{-(2k+2)/(2k+d)})$. For the vector norm, integrate $|\xi|^{-2}e^{s|\xi|^2}$ over $|\xi| \le 1/h$. Since $d > 2$, its norm squared is at most $Ch^{2-d}e^{Cs/h^2}$. The integrated variance is at most

$$\frac{C}{n}\int_I h(x)^{2-d} e^{Cs/h(x)^2}\, dx \le \frac{C}{n} h_0^{2-d} e^{Cs/h_0^2}.$$

Here $h(x)^{2-d} = h_0^{2-d} e^{-a(d-2)|x-\widehat{m}|^2}$ removes the growing volume factor. To compare the bound with h_0^{2k+2}, put $X = ns^{k+d/2}$, $q_0 = k + d/2$. We have

$$nh_0^{2k+d} \geq \max\left(1, \frac{X}{(\gamma \log(2+X))^{q_0}}\right), \qquad e^{Cs/h_0^2} \leq (2+X)^{C\gamma}.$$

For $X \leq 1$ the exponential is bounded, and for $X \geq 1$ it is at most a constant times $X/\log^{q_0}(2+X)$ if $C\gamma < 1$. Combining the bias and variance estimates,

$$\mathbb{E}\left[\int_I \frac{|\widehat{U} - U_b(f)|^2 + |\widehat{J} - J_b(f)|^2}{f}\right] \leq C h_0^{2k+2} \tag{37}$$

conditionally on any first part in $\mathcal{A}$.

Choose a fixed countable subset of $\mathcal{F}_k$ dense simultaneously in W_2 and locally uniform density convergence. It exists by second countability of the product of those spaces. Among its candidates g with $|m_g - \widehat{m}| < 1$ select the first with loss

$$\ell(g) = \int_I \frac{|\widehat{U} - U_b(g)|^2 + |\widehat{J} - J_b(g)|^2}{g}$$

strictly less than the infimum plus n^{-2}. The feasible set is always nonempty by translations and density. The selection is measurable by countability and measurability of the loss integrals. On $\mathcal{A}$ the true f is strictly feasible in the mean constraint. Its loss is approximated by the dense set: the fields converge uniformly on compact sets when the densities converge locally uniformly, since K_b, Q_b are integrable and the densities have a uniform global bound. The denominator is bounded below on the compact ball in the limit. Thus $\ell(g) \leq \ell(f) + n^{-2}$. The paired field discrepancy in (32) with denominator $f + g$ is at most $2\ell(f) + 2\ell(g)$. The exterior mass in that bound is $O(n^{-2})$ by taking D sufficiently large. On $\mathcal{A}^c$, the mean constraint and uniform centered second moments give $W_2^2(f,g) \leq C(1 + |\widehat{m} - m_f|^2)$. Its expected bad-event contribution is $O(n^{-1})$, which is smaller than the target squared order since $d > 2$. Together (37) and (32) give the assertion of Theorem 4.1 for $0 \leq s \leq n^{-\eta}$.

4.4 The remaining noise range

For isotropic noise and $f, g \in \mathcal{F}_k$, write $\epsilon = W_1(f * \gamma_s, g * \gamma_s) \leq 1$. Their means differ by at most ϵ. Use (32) on a ball of radius $L = D_0\sqrt{\log(1/h)}$ about m_f, with $b = h$. Choose D_0 so that the exterior term is $O(h^{k+1})$, by (27). On the ball the density floor is at least h^P for some fixed P. By (30) the weighted vector discrepancy of $J_h(f)$ and $J_h(g)$ against $1/(f+g)$ has norm $O(h^{k+1})$: use each density for its own short-shift bound and arbitrary decay for the remaining term. The observed characteristic functions differ by at most $\epsilon|\xi|$ by coupling. Removing Gaussian smoothing on the support of $\widehat{K}(h\xi)$ and integrating shows

$$\|U_h(f) - U_h(g)\|_\infty \leq C\epsilon h^{-d-1} e^{s/(2h^2)}.$$

Therefore, for all sufficiently small $h > 0$,

$$W_2(f,g) \leq C\left(h^{k+1} + \epsilon h^{-d-1-P/2} \log^{d/4}(1/h) e^{s/(2h^2)}\right). \tag{38}$$

Only a polynomial prefactor in h^{-1} is needed here.

For an empirical observed law $\widehat{P}_n$, let $e_n = W_1(\widehat{P}_n, f * \gamma_s)$. Uniformly, $\mathbb{E}e_n \leq Cn^{-\beta}$, where $\beta = 1/[4(d+1)]$. For a proof translate the population mean to zero for the estimate, truncate to a box of radius n^β by moving exterior points to zero, then quantize with mesh $n^{-\beta}$. The first two operations cost $O(n^{-\beta})$ in expectation by the second moment bound and the mesh size. The grid has $O(n^{2d\beta})$ points. The expected histogram total variation is at most half the square root of the number of points divided by n, by summing Bernoulli variances. Its transport cost is at most the diameter times this bound, which is $O(n^{-1/4})$ with the stated β.

Fit, in convolved W_1, a fixed countable W_2-dense set of mean-zero shapes in $\mathcal{F}_k$ translated to have the observed mean $\overline{Y}$. Choose the first within n^{-1} of the countable infimum, with strict inequality. The resulting

convolved discrepancy from the truth is at most $3e_n + n^{-1}$, since translating the truth to the sample mean costs at most e_n and can be approximated in the set of shapes. On $e_n \le n^{-\beta/2}$, take in (38)

$$h = \max\left(n^{-a_0}, \sqrt{\frac{s}{\gamma_0 \log n}}\right)$$

with $a_0, \gamma_0 > 0$ small enough that $a_0(k + 2 + d + P/2) + \gamma_0/2 < \beta/2$. This makes the second term in (38) $O(h^{k+1})$. When $s \ge n^{-\eta}$ and $\eta < a_0$, the second bandwidth dominates and $B_n(s) \asymp \log n$. We choose η to satisfy this, (36), the earlier condition $\eta/2 < 1/(2k + d)$, and $\eta(k + 1) < \beta/2$. The exceptional event has probability $O(n^{-\beta/2})$. Since the estimate has mean $\overline{Y}$, coupling centered variables gives squared loss at most $C + |\overline{Y} - m_f|^2$. The exceptional contribution to squared risk is $O(n^{-\beta/2} + n^{-1}) = o((s/\log n)^{k+1})$. This proves the isotropic upper bound.

For general Σ, independently add Gaussian noise of covariance $sI - \Sigma$ to the observed sample and use the isotropic estimator. To remove auxiliary randomness, output its conditional mixture given the original observations. The mixture has finite second moment because the isotropic estimates are in the curvature class, with uniformly bounded centered variance and means within a fixed distance of a Gaussian-perturbed sample mean. Convexity of squared W_2 in its first measure gives no larger squared risk. In either the isotropic or the general case, we can keep the output in a fixed countable subset of $\mathcal{F}_k$ by projecting the probability estimate in W_2 onto that dense subset, with tolerance n^{-1}. By density and the triangle inequality, its loss is at most twice the previous loss plus n^{-1}. This last version is used for the map estimate below.

4.5 Matching lower bounds

We use the following sign-testing argument. Suppose a finite sign family of latent densities f_ω has pairwise squared distances bounded below by $c_0 \sum_{j:\omega_j \ne \omega'_j} v_j$ for weights $v_j \ge 0$, and the sample laws for adjacent sign vectors have total variation at most $1/2$. Decoding any estimate to its nearest vertex loses at most a factor two in metric distance. For uniform signs the decoded error for each bit has probability at least $1/4$. If $S = \sum_j v_j$ and Z is the random erroneous weight, then $\mathbb{E}Z \ge S/4$ and $\mathbb{E}\sqrt{Z} \ge \mathbb{E}Z/\sqrt{S}$ since $0 \le Z \le S$. It gives a lower bound $c\sqrt{S}$ for expected loss. The chi-square divergence between adjacent one-sample laws is a convenient testing bound, since $1 + \chi^2$ tensorizes and convolution decreases χ^2 by conditional Jensen.

A transport estimate used below is the following Taylor bound, for a smooth test function φ and each pair of latent densities $f_\omega, f_{\omega'}$:

$$\int \varphi(f_\omega - f_{\omega'}) \le \|\nabla\varphi\|_{L^2(f_\omega)} W_2(f_\omega, f_{\omega'}) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(f_\omega, f_{\omega'})^2. \tag{39}$$

Integrate the Taylor inequality against an optimal coupling and use Cauchy–Schwarz. In both constructions the left side is at least cE for a nonnegative E, with gradient norm at most $C\sqrt{E}$ and Hessian norm uniformly bounded. It follows that $W_2^2(f_\omega, f_{\omega'}) \ge c'E$.

4.5.1 Disjoint perturbations for the direct lower bound

Fix a centered isotropic Gaussian g whose curvature lies strictly in (m, M). Let ψ be a nonzero smooth function of compact support in a unit cube and integral zero. Put $h = n^{-1/(2k+d)}$. Pack order h^{-d} disjoint translated cubes of side h in a fixed cube, with centers z_j, and use $\psi_j(x) = \psi((x - z_j)/h)$ supported in those cells after placing the reference cube around zero. Set

$$f_\omega = g + a\sum_j \omega_j \psi_j, \qquad a = \kappa h^k.$$

On the fixed cube, g and its reciprocal are smooth and bounded, with all derivatives of fixed finite order bounded. Derivatives through order $j \le k$ of the relative perturbation are therefore bounded by $C\kappa h^{k-j}$. For small fixed κ, mass is unchanged and the log-potential derivative perturbations respect the Hessian margin and every M_j. The densities lie between $g/2$ and $3g/2$. The adjacent direct one-sample chi-square bound is $Ca^2h^d = C\kappa^2/n$, which pays the sample total variation bound with or without noise.

For two vertices set $c_j = (\omega_j - \omega'_j)a$ and $\varphi = h^2 \sum_j c_j \psi_j$. The pairing is $E = h^{d+2} \|\psi\|_2^2 \sum_j c_j^2$ by disjointness. A gradient adds h^{-1} and integration over a cell contributes h^d, so its weighted squared norm is bounded by CE. The Hessian sup norm is bounded by $C \max_j |c_j|$. By (39) the weights per bit have order h^{2k+d+2}. Summing order h^{-d} such weights gives $S \asymp h^{2k+2}$ and the lower bound ca_n.

4.5.2 Noisy Gaussian packets

Diagonalize Σ, with eigenvalues $s = t_1 \geq t_j \geq 0$. First suppose $0 < s \leq s_0$ for a sufficiently small positive constant s_0. In a fixed central cube, choose grid centers z_j spaced by $D\sqrt{s}$, for a large fixed D. After choosing s_0 small enough there are order $s^{-d/2}$ centers. Put $T^2 = A_0 B_n(s)/s$ and

$$\psi_j(x) = e^{-|x-z_j|^2/(2s)} \left[\cos(T(x_1 - z_{j,1})) - e^{-sT^2/2}\right], \qquad f_\omega = g + a \sum_j \omega_j \psi_j, \qquad a = \kappa T^{-k}.$$

Each packet has integral zero. For any derivative order $0 \leq j \leq k$, the sum of relative derivatives before the factor a is bounded by CT^j. Indeed $s^{-1/2} \leq CT$, and derivatives of the envelope and of $1/g$ can be absorbed into a weaker Gaussian envelope $e^{-c|x-z_j|^2/s}$, using small s_0 and bounded centers. The sum of these envelopes is uniformly bounded by grid separation. With small κ, all perturbed densities are in $\mathcal{F}_k$ and between $g/2$ and $3g/2$.

The exact convolution of a packet is

$$(\psi_j * \gamma_\Sigma)(y) = \prod_{i=1}^d \sqrt{\frac{s}{s+t_i}}\, e^{-\sum_{i=1}^d (y_i - z_{j,i})^2/[2(s+t_i)]} \left[e^{-sT^2/4} \cos\!\left(\tfrac{T}{2}(y_1 - z_{j,1})\right) - e^{-sT^2/2}\right].$$

The oscillatory factor follows from conditional Gaussian integration in coordinate one, where both variances equal s. The formula in the other coordinates includes $t_i = 0$. The noisy reference has a uniform Gaussian lower envelope. Dividing the squared packet by it and integrating gives at most $Cs^{d/2}e^{-csT^2}$ for small enough s_0. The noisy vertices are bounded below by half that reference. The adjacent n-sample chi-square testing parameter is thus controlled by the bound on n times the one-sample divergence,

$$C\kappa^2 X (A_0 \log(2+X))^{-k} (2+X)^{-cA_0}, \qquad X = ns^{k+d/2}.$$

A sufficiently large A_0 and small κ give the required total variation.

To separate the vertices in W_2, write $c_j = (\omega_j - \omega'_j)a$ and $\varphi = T^{-2} \sum_j c_j \psi_j$. The L^2 Gram matrix of the packets has diagonal at least $cs^{d/2}$, by rescaling and integrating the cosine square for $T\sqrt{s}$ large. Its off-diagonal entries are bounded by $Cs^{d/2}e^{-c|z_j - z_i|^2/s}$. Choosing D large enough makes each off-diagonal row sum less than half of the lower diagonal bound. Hence the pairing is at least cE with $E = T^{-2}s^{d/2} \sum_j c_j^2$. The gradient Gram row sums are at most $CT^2 s^{d/2}$ by overlap of the derivative envelopes. Since f_ω is bounded above, the gradient norm squared is at most CE. The Hessian of the test has sup norm at most Ca, by the pointwise overlap bound. This verifies (39) and gives weight $T^{-2k-2}s^{d/2}$ up to a constant per bit. The sum over order $s^{-d/2}$ bits gives the expected W_2 lower bound cT^{-k-1}.

For $s_0 \leq s \leq \Lambda$, a single relative cosine perturbation of the same Gaussian, of amplitude κT^{-k} and with its Gaussian mean removed, suffices at $T^2 = C \log n/s$. It stays in $\mathcal{F}_k$ by the same derivative bound. Testing with $\cos(Tx_1)/T$ gives $W_2 \geq W_1 \geq cT^{-k-1}$. Gaussian attenuation gives one-sample chi-square at most $CT^{-2k}e^{-csT^2}$ uniformly, so the product total variation is at most $1/2$ for all large n. This matches because $B_n(s) \asymp \log n$ on that range. Combined with the direct lower bound, this proves Theorem 4.1. The two terms balance when $X/\log^{k+d/2}(2+X)$ has order one, giving $s \asymp n^{-2/(2k+d)}$.

4.6 Application to Brenier maps from two noisy samples

Suppose two independent samples observe latent densities $f, g \in \mathcal{F}_k$ with sample sizes n_1, n_2 and known Gaussian covariances Σ_1, Σ_2. Let $T_{f,g}$ be the Brenier map from f to g. For an estimate $\widehat{T}$ use risk

$$\left(\mathbb{E} \int |\widehat{T}(x) - T_{f,g}(x)|^2 f(x)\, dx\right)^{1/2}.$$

Write $R(n,s) = n^{-(k+1)/(2k+d)} + (s/\log(2+ns^{k+d/2}))^{(k+1)/2}$.

Theorem 4.2. *For n_1, n_2 sufficiently large, the two-sample minimax risk over the stated class is of order $R(n_1, \|\Sigma_1\|_{\mathrm{op}}) + R(n_2, \|\Sigma_2\|_{\mathrm{op}})$, uniformly over bounded known noise covariances.*

Proof. Use the class-valued versions of the two estimates $\widehat{f}, \widehat{g}$ and take their Brenier map. They can be chosen from countable dense subsets of the class, so a fixed representative for each pair of maps gives a measurable map estimate. By the two Hessian bounds, Caffarelli's contraction theorem [12] gives a uniform Lipschitz bound for all maps between these densities and for their inverses. Equivalently, the true Brenier potential is uniformly smooth and strongly convex. The stability inequality of Balakrishnan and Manole [17, Theorem 3] gives

$$\|\widehat{T} - T_{f,g}\|_{L^2(\widehat{f})} \le C\big(W_2(f,\widehat{f}) + W_2(g,\widehat{g})\big).$$

Couple f and $\widehat{f}$ in W_2 and use uniform Lipschitz bounds for both maps to replace $L^2(\widehat{f})$ by $L^2(f)$, at cost at most $CW_2(f,\widehat{f})$. Theorem 4.1 and Minkowski give the upper bound.

For the lower bound for the target sample, fix the source at the reference Gaussian g_0 and vary the target over the families in the proof of Theorem 4.1. For two such target densities $g_\omega, g_{\omega'}$, the two maps on the common Gaussian source give a coupling of them. Hence the $L^2(g_0)$ distance between the maps is at least $W_2(g_\omega, g_{\omega'})$. The adjacent testing distances are unchanged by the fixed first sample.

For the lower bound for the source sample, fix the target at g_0 and vary the source in the same families, denoted by f_ω. The source densities are comparable to g_0 uniformly, so their L^2 losses are comparable to the common $L^2(g_0)$ loss. Writing $T_\omega = T_{f_\omega, g_0}$, the inverse maps give a coupling of the two source densities. Use the Lipschitz constant of $T_{\omega'}^{-1}$, and change variables by T_ω, to obtain

$$\begin{aligned} W_2(f_\omega, f_{\omega'}) &\le \|T_\omega^{-1} - T_{\omega'}^{-1}\|_{L^2(g_0)} \\ &\le C\|T_\omega - T_{\omega'}\|_{L^2(f_\omega)} \le C'\|T_\omega - T_{\omega'}\|_{L^2(g_0)}. \end{aligned}$$

Hence the same sign-testing and two-point bounds apply to the maps under a common loss, and the fixed target sample cannot improve the testing distances. The argument lower-bounds even the mean of the L^2 loss, and therefore the stated root-mean-square risk. The maximum of the two sample lower bounds proves the claim. □

4.7 Relation to smooth estimation and transport stability

The direct exponent on compact classes with a density floor is classical [8]. In the present class no density floor extends over the sample region as n grows. The bandwidth in (34) is used only for a pointwise estimate of the constant-coarse-scale field. Its spatial integrability in variance and the Gaussian integrability of its bias are both needed. The noisy lower bound uses the total local information $ns^{k+d/2}$ before the Gaussian attenuation, which determines the inverse logarithm through the crossover. Fixed supersmooth Wasserstein deconvolution over moment classes is studied in [6], and Delaigle [5] considers shrinking noise for density loss. Park and Kim [10] report shape-constrained density-risk results for log-concave deconvolution in their conference abstract. For direct sampling in the whole-space two-Hessian class, Divol, Niles-Weed, and Pooladian [18, Corollary 1] use a wavelet semi-dual estimator. Balakrishnan and Manole [17, Corollary 6] use nearest neighbors and empirical transport. Theorem 4.2 uses smooth density recovery to attain the minimax order with or without Gaussian noise.

5 Smoothness in the critical dimensions for Wasserstein deconvolution

We obtain sharp quadratic Wasserstein minimax rates in dimensions one and two for strongly log-concave densities of any fixed real potential smoothness $\alpha \ge 2$, observed with known Gaussian contamination. The direct-sampling terms are $\sqrt{\log\log n/n}$ on the line and $(\log n)/\sqrt{n}$ in the plane. The inverse terms are $(s/\log n)^{(\alpha+1)/2}$ and $(s/\log(2+ns^{\alpha+1}))^{(\alpha+1)/2}$, respectively, where s is the largest noise variance. The planar upper bound uses a weighted fit to a coarse heat

density and a flux. An exact difference of Gaussian flux kernels retains the direct-sampling variance order while a localized inverse kernel resolves the higher frequencies.

5.1 The two transition laws

Let $d = 1$ or 2, $0 < m < M < \infty$, and $\alpha = k + \theta \geq 2$, where k is an integer, $0 \leq \theta < 1$. Consider densities $f = e^{-V}$ on $\mathbb{R}^d$ with $V \in C^k$, $mI \preceq D^2V \preceq MI$, and fixed positive sup norm bounds on potential derivatives of orders $3, \ldots, k$. If $\theta > 0$, also impose a fixed positive bound on the global θ-Hölder seminorm of D^kV. These classes allow all translations. Observe n independent $Y_i = X_i + Z_i$, with independent latent observations of density f and independent centered Gaussian noise of known covariance $\Sigma \succeq 0$. Let $s = \|\Sigma\|_{\mathrm{op}} \leq \Lambda$ for a fixed bound Λ, and write $\gamma_\Sigma = N(0, \Sigma)$ and $\gamma_t = N(0, tI_d)$. For $q = 1, 2$, let $\mathcal{R}_{n,q}(\Sigma)$ be the infimum, over probability estimators, of the supremum of $(\mathbb{E}W_2(\widehat{f}, f)^q)^{1/q}$ over the stated class.

Theorem 5.1. *Uniformly over bounded known noise covariances, for either loss and all sufficiently large n,*

$$\mathcal{R}_{n,q}(\Sigma) \asymp \begin{cases} \sqrt{\dfrac{\log\log n}{n}} + (s/\log n)^{(\alpha+1)/2}, & d = 1, \\ \dfrac{\log n}{\sqrt{n}} + (s/\log(2 + ns^{\alpha+1}))^{(\alpha+1)/2}, & d = 2. \end{cases}$$

The transition variances are of orders

$$(\log n)(\log\log n/n)^{1/(\alpha+1)} \quad \textit{and} \quad n^{-1/(\alpha+1)}(\log n)^{2/(\alpha+1)} \log\log n,$$

respectively.

The direct orders coincide with the Gaussian empirical transport orders [4, 13]. The lower bounds show how the tail accumulation persists for arbitrary estimators despite the additional potential regularity. The fixed-covariance theorem of [1, Theorem 2.5] covers the curvature case $\alpha = 2$ at fixed positive noise. Here the higher regularity changes the inverse exponent, and the planar inverse logarithm changes its order near the direct regime.

5.2 Local smoothness and transport comparisons

Write $m_f = \mathbb{E}_f X$. All densities in the class satisfy uniformly

$$ce^{-C|x-m_f|^2} \leq f(x) \leq Ce^{-c|x-m_f|^2}, \qquad |\nabla V(x)| \leq C(1 + |x - m_f|). \tag{40}$$

These follow from quadratic bounds about the mode, normalization and the bounded distance between mean and mode. The product and chain rules show

$$|D^j f(x)| \leq C(1 + |x - m_f|^j) f(x), \qquad 0 \leq j \leq k.$$

If $\theta > 0$, on $|x - m_f| \leq L$, $|u| \leq (2 + L)^{-1}$ we also have

$$|D^k f(x + u) - D^k f(x)| \leq C|u|^\theta (1 + |x - m_f|^{k+1}) f(x). \tag{41}$$

Indeed $D^k f$ is f times a polynomial in potential derivatives. The D^kV term has the assumed Hölder bound, all lower derivative differences follow from their next derivative, and $f(x + u)/f(x)$ is uniformly bounded on these shifts. The convolution $p_s = f * \gamma_s$ has uniform centered Gaussian tails. By the tangent-plane inequality,

$$\frac{p_s(x - u)}{f(x)} \leq \exp\left(\nabla V(x) \cdot u + \tfrac{s}{2}|\nabla V(x)|^2\right). \tag{42}$$

In particular this ratio is bounded on a growing window of radius $L \asymp \sqrt{\log n}$ and short shifts $|u| \leq (2+L)^{-1}$ when $s \leq n^{-\eta}$, for any fixed $\eta > 0$.

Let $\widehat{K}$ be a real even smooth Fourier cutoff supported in the unit ball and identically one near zero, using $\widehat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\,dx$. Write $K_h(x) = h^{-d}K(x/h)$ for its spatial kernel. The inverse transform Q of

$$\frac{i\xi}{|\xi|^2}(1 - \widehat{K}(\xi))$$

is integrable in space, has arbitrary integrated tail decay and all monomial moments zero. Indeed the symbol vanishes near zero and has order -1 at infinity. A smooth dyadic annulus at frequency 2^j gives kernel bounded by $C_N 2^{(d-1)j}(1 + 2^j|x|)^{-N}$, by rescaling and Fourier integration by parts. Its L^1 norm is $O(2^{-j})$. All absolute moments follow by summation and all signed moments vanish by the flat region of the symbol at zero.

Set $Q_h(x) = h^{-d}Q(x/h)$. On a window $|x - m_f| \le L$, with $w = (2 + L)^{-1}$,

$$|hQ_h * f(x)| \le Ch^{\alpha+1}(1 + |x - m_f|^{k+1})f(x) + C_N h(h/w)^N. \tag{43}$$

Subtract the Taylor polynomial of f to degree $k-1$ if $\theta = 0$ and k otherwise, using respectively the bounded derivative or (41) for the short-shift remainder. Off the short-shift region use arbitrary tail decay and uniform bounds on the derivatives. In dimension one, if F is the distribution function of f, the same argument for $K_h * F - F$ gives

$$|K_h * F(x) - F(x)| \le Ch^{\alpha+1}(1 + |x - m_f|^{k+1})f(x) + C_N(h/w)^N. \tag{44}$$

The Taylor degrees are k if $\theta = 0$ and $k + 1$ otherwise, since $F^{(k+1)} = D^k f$.

5.2.1 Weighted field fitting

For use below, let two densities f, g in the class have their means within fixed distance, put $\rho = f + g$, and let I be a ball of radius $L \ge 2$ centered within fixed distance of both means. If $f - g = U - \operatorname{div} J$ in distributions, then

$$W_2(f, g) \le C\left[\left(\int_{\mathbb{R}^d \setminus I^-} \rho\right)^{1/2} + \left(\int_I \frac{|U|^2 + |J|^2}{\rho}\right)^{1/2}\right], \tag{45}$$

where I^- has radius $L - 1$. To verify this, the mixture ρ has a uniform Poincaré inequality. Each density has one by Brascamp–Lieb [11]. Compare both component means of a test function to its Lebesgue average on a fixed unit ball about m_f, where both densities are bounded below, to get the mixture inequality. Subtract the mixture mean from a smooth test function ϕ. In the integral of $(f - g)\phi$, use a smooth cutoff to I equal to one on I^-. The exterior is bounded by the first term of (45) times $C\|\nabla\phi\|_{L^2(\rho)}$. In the interior, insert $U - \operatorname{div} J$; the cutoff test and its gradient have $L^2(\rho)$ norms at most $C\|\nabla\phi\|_{L^2(\rho)}$ by the mixture Poincaré inequality. Riesz representation on the closed $L^2(\rho)$ space of test gradients gives a flux with divergence $f - g$ and $L^2(1/\rho)$ norm bounded by the asserted right-hand side. Interpolate from f to g with densities $(1 - \theta_0)f + \theta_0 g$, $\theta_0 = \sin^2(\pi t/2)$. The flux is multiplied by θ_0'; its action is at most π^2 times its squared $L^2(1/\rho)$ norm since $\theta_0'^2/\min(\theta_0, 1 - \theta_0) \le \pi^2$. The dynamic transport bound [3] proves (45).

5.3 Small-noise estimates at the two dimensions

Assume isotropic noise for now, $0 \le s \le n^{-\eta}$ with a fixed sufficiently small $\eta > 0$. Split the sample into two comparable parts. Use the first to form its mean $\widehat{m}$, and set

$$I = B_{L_n}(\widehat{m}), \quad L_n = D\sqrt{\log n}, \qquad w_n = (3 + L_n)^{-1}, \qquad \mathcal{A} = \{|\widehat{m} - m_f| < 1/2\}.$$

Uniform Gaussian tails give exponentially small probability of $\mathcal{A}^c$ by Chernoff. On $\mathcal{A}$, the density f is at least n^{-P} on I, and (42) gives a comparison $p_s(x - u) \le Cf(x)$ for $x \in I$, $|u| \le w_n$. We take D large enough that the exterior term in (45) is $O(n^{-1})$ for any two fit candidates whose means remain within one of $\widehat{m}$.

5.3.1 The localized distribution function on the line

Here use

$$h = \max\bigl(n^{-1/(2(\alpha+1))}, \sqrt{s/(\gamma \log n)}\bigr), \qquad A = s/h^2$$

with $\gamma > 0$ small. Thus h is bounded above by a fixed negative power of n, and $A \le \gamma \log n$. Let L_A be the inverse Fourier transform of $\widehat{K}(v)e^{Av^2/2}$, and put

$$D_{s,h}(u) = \int_{-\infty}^{u/h} L_A(v)\, dv - \mathbf{1}_{\{u \ge 0\}}.$$

Use the second-part empirical distribution function plus the sample averages of $D_{s,h}(x - Y_i)$ restricted to $|x - Y_i| \le w_n$. Denote their sum by $\widehat{F}(x)$. Without truncation its population value is $K_h * F$ by Gaussian Fourier inversion and integration of the resulting density from $-\infty$. The unscaled primitive difference has Fourier transform $(\widehat{K}(v)e^{Av^2/2} - 1)/(iv)$; near zero the numerator vanishes to second order, and outside the Fourier support the integrand is $O(1/|v|)$. Plancherel gives

$$\int |D_{s,h}(u)|^2\, du \le Ch(1 + A)^2 e^{CA}.$$

For $|u|/h \ge 1$, integrating the Fourier kernel tail gives

$$|D_{s,h}(u)| \le C_N(1 + A)^{C_N} e^{CA}(1 + |u|/h)^{-N}$$

for every fixed $N > 0$. The exponential constant can be fixed independently of N by compact Fourier support. It follows that physical truncation changes the expectation by $O(n^{-J})$ for every prescribed $J > 0$ when N is sufficiently large. Equation (44), the polynomial density floor and Gaussian moments now give squared weighted bias $O(h^{2\alpha+2} + n^{-2})$.

For the raw empirical distribution term, let P_s be the distribution function of p_s. In dimension one the output potential has curvature bounded above and below uniformly. One way to see this is the posterior formula

$$-(\log p_s)'' = s^{-1} - s^{-2} \operatorname{Var}(X \mid Y).$$

For a conditional variable of potential A_0, integration by parts gives $\operatorname{Cov}(X, A_0'(X)) = 1$. The derivative slopes between $m + s^{-1}$ and $M + s^{-1}$ bound this variance between $(M + s^{-1})^{-1}$ and $(m + s^{-1})^{-1}$. The formula at zero noise is unnecessary. For either strongly log-concave density the smaller tail probability is bounded by a constant times its density divided by $1 + |x - m_f|$: far from the mode use the tangent plane and integrate its exponential, and on the remaining fixed interval use the uniform density floor. Thus conditionally on the first part on $\mathcal{A}$,

$$\frac{C}{n} \int_I \frac{P_s(x)(1 - P_s(x))}{f(x)}\, dx \le C \frac{\log\log n}{n}$$

by the density ratio and integration of $1/(1 + |x - m_f|)$. For the correction, the localized ratio and the squared kernel norm give integrated weighted variance at most $C|I|h(1 + A)^2 e^{CA}/n$, which is $o(n^{-1})$ by choosing γ small enough. Consequently

$$\mathbb{E}\left[\int_I \frac{|\widehat{F} - F|^2}{f} \,\middle|\, \text{first part}\right] \le C\left(h^{2\alpha+2} + \frac{\log\log n}{n}\right) \quad \text{on } \mathcal{A}. \tag{46}$$

We will use F as the candidate field, taking $J = G - F$ and $U = 0$ in (45).

5.3.2 Coarse heat and fine correction in the plane

Write $p_t = f * \gamma_t$, and use the candidate fields

$$U_b(f) = p_{b^2}, \qquad J_b(f) = \tfrac{1}{2} \int_0^{b^2} \nabla p_t\, dt, \qquad b^2 = (\log n)^{-1}.$$

In distributions $f = U_b(f) - \operatorname{div} J_b(f)$. Put $v = b^2 - s$, which is positive and comparable to b^2 for large n. Estimate $U_b(f)$ by averaging $\gamma_v(x - Y_i)$, with the Gaussian density understood as a kernel. Its population value is exact.

For the flux choose

$$h = \max\left((\log n/\sqrt{n})^{1/(\alpha+1)}, \sqrt{\frac{s}{\gamma \log(2 + n s^{\alpha+1})}} \right), \qquad A = s/h^2. \tag{47}$$

This is polynomially small in the present range and $h^2 \ll v$. We aim to estimate the population flux with symbol

$$\frac{i\xi}{|\xi|^2}\big(\widehat{K}(h\xi) - e^{-b^2|\xi|^2/2}\big)\widehat{f}(\xi).$$

After multiplying by the inverse Gaussian characteristic function to act on the observations, the kernel splits into two multipliers

$$\frac{i\xi}{|\xi|^2}\big(\widehat{K}(h\xi)e^{s|\xi|^2/2} - e^{-h^2|\xi|^2/2}\big), \tag{48}$$

$$\frac{i\xi}{|\xi|^2}\big(e^{-h^2|\xi|^2/2} - e^{-v|\xi|^2/2}\big). \tag{49}$$

In (48), after scaling $u = h\xi$, the numerator is equal near zero to the difference of two radial analytic functions equal at zero. The spatial kernel therefore has the form $h^{-1}L_A(x/h)$ for a Schwartz profile with seminorms bounded by $C_N(1 + A)^{C_N}e^{CA}$. This follows by differentiating the smooth scaled symbol, which is the sum of a compactly supported part and a rapidly decaying part away from zero. Plancherel bounds its squared L^2 norm by $C(1 + A)^C e^{CA}$ in two dimensions. In the sample average physically truncate this fine kernel at radius w_n. Its mean changes by $O(n^{-J})$ for arbitrary J, using the Schwartz seminorm and polynomial smallness of h. Its integrated weighted variance is bounded by

$$\frac{C \log n}{n}(1 + A)^C e^{CA}$$

by short-shift density comparison. Writing $X = 2 + ns^{\alpha+1}$, choose γ with $(1 + A)^C e^{CA} \le CX^c$ for a fixed $c < 1/3$. If $X \le (\log n)^3$ the displayed variance is $O((\log n)^2/n)$. Otherwise $\log n \le X^{1/3}$ and $X^{1/3+c} \le C(X - 2)/(\log X)^{\alpha+1}$, so the variance is $O(h^{2\alpha+2})$.

The kernel in (49) is $R_v - R_{h^2}$, where in two dimensions the integrated heat kernel is explicitly

$$R_a(u) = \tfrac{1}{2}\int_0^a \nabla\gamma_t(u)\,dt = -\frac{u}{2\pi|u|^2}\,e^{-|u|^2/(2a)}, \qquad u \neq 0.$$

This follows by integrating $t^{-2}e^{-|u|^2/(2t)}$. Estimate this part without spatial truncation. Since $h^2 \le v$, the difference has magnitude at most

$$C \min(|u|/h^2, 1/|u|)\, e^{-c|u|^2/v}.$$

For $x \in I$, the full density ratio in (42) is bounded by $Ce^{CL_n|u|}$. As vL_n^2 is bounded, this factor is absorbed into a weaker Gaussian in the squared kernel bound by completing the square. Polar integration of the resulting weighted kernel square gives $C(1 + \log(\sqrt{v}/h)) = O(\log n)$. For γ_v the same comparison gives a squared weighted kernel bound $C/v = O(\log n)$. Thus both coarse averages have integrated weighted variance $O((\log n)^2/n)$, since I has area $O(\log n)$.

Let $\widehat{J}$ be the sum of the fine and coarse estimates and $\widehat{U}$ the heat density estimate. Before physical truncation, the difference of the population flux from $J_b(f)$ is exactly $-hQ_h * f$. This follows by subtracting the two Fourier symbols, which cancels the coarse Gaussian multiplier. Thus (43) bounds its squared weighted bias by $Ch^{2\alpha+2}$, without differentiating a Gaussian-smoothed density. The weighted variance of the sum is at most twice the sum of the two flux variance bounds. We have proved

$$\mathbb{E}\left[\int_I \frac{|\widehat{U} - U_b(f)|^2 + |\widehat{J} - J_b(f)|^2}{f} \,\middle|\, \text{first part} \right] \le C\left(h^{2\alpha+2} + \frac{(\log n)^2}{n} \right) \quad \text{on } \mathcal{A}. \tag{50}$$

5.3.3 Probability estimates from the fields

For either dimension choose a fixed countable dense set in the class for the joint W_2 and locally uniform density topologies. Among g with $|m_g - \widehat{m}| < 1$, select the first whose loss is within the infimum plus n^{-2}, with strict inequality. In dimension one the loss is $\int_I |\widehat{F} - G|^2/g$. In dimension two it is the integral in (50) with each candidate field evaluated at g and with denominator g. This is measurable by countability. The mean constraint is always feasible by translation and density of the set. On $\mathcal{A}$ the infimum is at most the loss of f by approximation: CDFs converge uniformly to the continuous true CDF, and Gaussian and integrated-heat kernels are integrable so their fields converge locally uniformly for locally uniform bounded densities. The limiting f has a positive floor on compact sets.

With denominator $f + g$, each paired squared field discrepancy is bounded by twice the sum of the corresponding two one-sided loss terms. Apply (45), taking $G - F$ as the flux in dimension one. The exterior contributes $O(n^{-1})$ in distance by the choice of D. On $\mathcal{A}^c$, uniform second moments about the means and the constraint on g give $W_2^2(f, g) \le C(1 + |\widehat{m} - m_f|^2)$; its expected contribution is $O(n^{-1})$ using the exceptional probability and unconditional variance. Equations (46) and (50) give the required upper bounds for small noise.

5.4 Larger noise and covariance completion

For f, g in the class with isotropic convolved discrepancy $\epsilon = W_1(f * \gamma_s, g * \gamma_s) \le 1$, apply (45) with

$$U = K_h * (f - g), \qquad J = hQ_h * (f - g)$$

on the ball of radius $L = B\sqrt{\log(1/h)}$ about m_f. The means differ by at most ϵ. Choose B so the exterior term is $O(h^{\alpha+1})$. Both densities are at least h^P on this ball for some fixed P, and (43) bounds the weighted norm of J by $Ch^{\alpha+1}$, paying each local bias with its own density. In the scalar field U, remove Gaussian smoothing on the cutoff support; the observed characteristic functions differ by at most $\epsilon|\xi|$ by coupling. Fourier inversion and the core density floor yield

$$W_2(f, g) \le C\left(h^{\alpha+1} + \epsilon h^{-d-1-P/2} \log^{d/4}(1/h) e^{s/(2h^2)}\right). \tag{51}$$

An empirical observed law has expected W_1 error $O(n^{-\beta})$, $\beta = 1/[4(d + 1)]$, uniformly. Indeed center at the theoretical mean, truncate outside a box of radius n^{β} at cost $O(n^{-\beta})$ by the second moment bound, and quantize at mesh $n^{-\beta}$ at the same cost. The grid has $O(n^{2d\beta})$ points, and the empirical histogram total variation error is at most the square root of the number of sites divided by n, up to a constant. Multiplying by the grid diameter gives $O(n^{-1/4})$.

Fit a countable dense set of mean-zero class shapes translated to the sample mean, minimizing convolved W_1 from the empirical law up to n^{-1}. This is a measurable choice by a fixed tie order with strict tolerance. Its convolved distance from the truth is at most $3e_n + n^{-1}$ when e_n denotes the empirical error, using the true shape translated to the sample mean as a comparator. On $e_n \le n^{-\beta/2}$, use $h = \max(n^{-a}, \sqrt{s/(\gamma_0 \log n)})$ in (51) with $a, \gamma_0 > 0$ so small that $a(\alpha + 2 + d + P/2) + \gamma_0/2 < \beta/2$. Then both terms are bounded by $Ch^{\alpha+1}$. For $s \ge n^{-\eta}$ with $\eta < \min(a, 1/(\alpha + 1))$, the Gaussian entry of h dominates and the two inverse logarithms in the theorem are both comparable to $\log n$. The bad event has probability $O(n^{-\beta/2})$. On it the squared loss is bounded by $C + |\overline{Y} - m_f|^2$ by mean anchoring and centered moments. Its expected contribution is $O(n^{-\beta/2} + n^{-1})$, negligible compared with the stated squared inverse rate if $\eta(\alpha + 1) < \beta/2$. This completes the isotropic upper bound.

For general known Σ, add independent Gaussian contamination of covariance $sI - \Sigma$ and run the isotropic estimator. The conditional mixture of its probabilities, given the original data, is an estimator without auxiliary randomization. It has finite second moment by the uniform centered variances and the mean anchoring of the isotropic procedures. Its squared W_2 risk is no larger by convexity in the first argument.

5.5 The direct and inverse lower bounds

For a finite sign family f_ω, suppose adjacent sample laws have total variation at most $1/2$, and squared latent distances dominate a weighted sum of differing signs with total available weight S. A nearest-vertex

decoder in W_2 loses at most a factor two. Each decoded sign has error probability at least $1/4$ under uniform signs, by testing each adjacent pair. If Z is the total mistaken weight, then $\mathbb{E}Z \geq S/4$ and $\mathbb{E}\sqrt{Z} \geq \mathbb{E}Z/\sqrt{S}$. This gives a lower bound of order $\sqrt{S}$ for expected W_2. Adjacent chi-square divergence contracts under convolution, and $1+\chi^2$ tensorizes on the sample. We use these facts throughout.

5.5.1 Direct lower bounds for both dimensions

Use a centered isotropic Gaussian g with potential curvature strictly inside (m, M). For the line, put $x_j = \sqrt{j}$, $w_j = (\sqrt{j+1} - \sqrt{j})/2$ and let $g_j = g(x_j)$. The intervals $[x_j, x_j + w_j]$ are disjoint and have density comparable to g_j. Fix a nonzero C^∞ bump ψ supported in $(0,1)$ with integral zero. For $j \leq J = \lfloor c \log n \rfloor$, add a signed bump

$$\omega_j a_j \psi((x - x_j)/w_j), \qquad a_j = \kappa \sqrt{\frac{g_j}{n w_j}}$$

to g. Relative to g, all derivatives of orders up to $k+1$ tend uniformly to zero as $n \to \infty$, by taking c small. The highest bound is at most $C\kappa(n g_j w_j^{2k+3})^{-1/2}$, using $x_j w_j = O(1)$. These bounds imply the required log-Hessian and higher-derivative constraints and the Hölder constraint when $\theta > 0$. For the latter, one may bound a derivative difference by the minimum of twice its sup norm and its Lipschitz bound times the separation. All perturbed densities are comparable to g. The adjacent n-fold testing bound follows from one-sample chi-square at most $C a_j^2 w_j / g_j = C\kappa^2/n$, with small fixed κ.

The endpoint CDFs are identical on every bump interval, so monotone transport couples the mass of each interval within itself. The local first-power cost on a changed interval is

$$2 a_j w_j^2 \int_0^1 \left| \int_0^v \psi(u)\, du \right| dv.$$

Cauchy–Schwarz, dividing its square by the interval mass of order $g_j w_j$, gives squared quadratic cost at least $c w_j^2/n$. The total available weight is $\asymp n^{-1} \sum_{j \leq J} j^{-1} \asymp \log\log n / n$.

For the plane, pack disjoint squares B of side $\ell = (\log n)^{-1/2}$ in a centered square of side $c\sqrt{\log n}$. Their total area is order $\log n$. In each, g is comparable to its central value g_B, with $g_B \geq n^{-\delta}$ for a fixed $0 < \delta < 1$ if c is small. Place compactly supported orthonormal tensor wavelets of regularity C^{k+1} [14] within each square, retaining those whose supports lie inside the square. Use one orientation. At level $r = \ell 2^{-j}$ there are order $(\ell/r)^2$ translates in a square, starting at some fixed j. Stop at a small positive multiple of $\log n$ levels, so the smallest length is $r_{\min} \asymp \ell n^{-\nu}$ for small fixed $\nu > 0$. The L^2-normalized functions ψ_i, of respective length r_i, have derivatives bounded by $C r_i^{-1-a}$ at order $a \leq k+1$ and bounded overlap at each level.

Add independent signs times $a_B = \kappa\sqrt{g_B/n}$ times these wavelets to g. The relative derivative sup norms through $k+1$ are bounded by $C \max_B (a_B/g_B) r_{\min}^{-k-2}$, using finite overlap and summing the geometric scales; derivatives of $1/g$ cost powers of ℓ^{-1} on the squares. Take $\nu < (1-\delta)/(2(k+2))$. Then all these bounds tend to zero, so the global Hölder condition follows as on the line, and the densities have mass one and belong to the class. The adjacent chi-square bound is $C a_B^2/g_B$ by L^2 normalization.

For pairwise separation write the signed difference as $\sum_i c_i \psi_i$. Use the test function

$$\varphi = \sum_i (r_i^2/g_{B(i)}) c_i \psi_i.$$

Orthonormality gives the pairing

$$E = \sum_i r_i^2 c_i^2 / g_{B(i)}.$$

Its gradient has squared $L^2(f_\omega)$ norm at most CE. Indeed for one level sum u_j the bound $\|\Delta u_j\|_2 \leq C r_j^{-2} \|u_j\|_2$ follows by normalized derivatives and overlap. For $r_j \geq r_k$, integration by parts then gives

$$|\langle \nabla u_j, \nabla u_k \rangle| \leq C(r_k/r_j)(\|u_j\|_2/r_j)(\|u_k\|_2/r_k).$$

Summing the geometric ratios and using comparability to g_B on each square proves the estimate. The test's Hessian sup norm is bounded by $C \max_B (a_B/g_B) r_{\min}^{-1}$, hence uniformly bounded. Taylor expansion in a coupling bounds the pairing above by $C\sqrt{E}\, W_2 + C W_2^2$, so $W_2^2 \geq cE$. Each square contributes order ℓ^2/n in weights per level. There are order $\log n$ levels and area of order $\log n$, so $S \asymp (\log n)^2/n$. Both direct lower bounds persist after contamination by data processing.

5.5.2 Gaussian oscillations and the inverse terms

A centered Gaussian g of strictly interior curvature can be multiplied by the factor

$$1 + \kappa T^{-\alpha}(\cos(Tx_1) - \mathbb{E}_g \cos(TX_1)).$$

This preserves mass. For $T \geq 1$ and small fixed κ, derivatives of its logarithm at orders $j \leq k+1$ are bounded by $C\kappa T^{j-\alpha}$. In the fractional case $0 < \theta < 1$, the difference of the k-th derivative at two points is therefore bounded by $C\kappa \min(T^{-\theta}, T^{1-\theta}|x-y|)$, which is at most $C\kappa|x-y|^\theta$. This gives all the potential constraints, including integer α without imposing any extra derivative bound. Testing with $\cos(Tx_1)/T$ gives a latent W_2 separation at least $cT^{-\alpha-1}$, by the W_1 test and the Gaussian cosine variance. When the first coordinate is an eigenvector of largest noise variance $s > 0$, conditional Gaussian integration gives one-sample chi-square at most $CT^{-2\alpha}e^{-csT^2}$ for bounded s. Indeed the conditional variance in that coordinate is between positive constant multiples of s, and centering the conditional cosine only decreases its second moment. Taking $T^2 = A_0 \log n/s$ with fixed large A_0 makes the product testing distance small uniformly. This proves the line inverse lower bound for all $s > 0$ and the plane inverse lower bound for s bounded below, where the effective logarithm is comparable to $\log n$.

For the plane and $0 < s \leq s_0$ with a small fixed s_0, rotate axes to $\Sigma = \mathrm{diag}(s,t)$, $0 \leq t \leq s$. Place $N \asymp 1/s$ centers z_j spaced $D\sqrt{s}$ apart in a fixed square, with large fixed D. Put $T^2 = A_0 \log(2 + ns^{\alpha+1})/s$ and

$$\psi_j(x) = e^{-|x-z_j|^2/(2s)} \left(\cos(T(x_1 - z_{j,1})) - e^{-sT^2/2}\right), \qquad f_\omega = g + a\sum_j \omega_j \psi_j, \qquad a = \kappa T^{-\alpha}.$$

Each packet has zero mass. Before multiplication by a, relative derivatives through order $j \leq k+1$ are bounded by CT^j after summing the envelopes over the separated centers. This uses $s^{-1/2} \leq CT$ and a weaker Gaussian envelope $e^{-c|x-z_j|^2/s}$ to absorb $1/g$ and the derivatives. Consequently the logarithmic perturbation obeys the same derivative and Hölder bounds as the single cosine. All densities lie in the class and between $g/2$ and $3g/2$.

The noisy packet is exactly

$$\frac{s}{\sqrt{2s(s+t)}}\, e^{-(y_1-z_{j,1})^2/(4s) - (y_2-z_{j,2})^2/(2(s+t))} \left(e^{-sT^2/4}\cos\left(\tfrac{T}{2}(y_1 - z_{j,1})\right) - e^{-sT^2/2}\right).$$

This includes $t = 0$. It follows by convolving the two Gaussian envelope factors and using the conditional cosine formula in the first coordinate. Its squared integral divided by the noisy reference is at most Cse^{-csT^2} by the Gaussian envelope and smallness of s_0. For $X = ns^{\alpha+1}$, the n-fold testing bound between adjacent vertices follows from n times the one-sample chi-square being at most

$$C\kappa^2 X(A_0 \log(2+X))^{-\alpha}(2+X)^{-cA_0}.$$

Choose A_0 large and κ small.

For metric separation test the difference $\sum_j c_j\psi_j$ against $T^{-2}\sum_j c_j\psi_j$, where $c_j = (\omega_j - \omega'_j)a$. The packet Gram matrix in L^2 has diagonals at least cs since $T\sqrt{s}$ is large. Off-diagonal row sums are bounded by $Cs\sum_{l\in\mathbb{Z}^2\setminus\{0\}} e^{-cD^2|l|^2}$ and made smaller than half the diagonal lower bound. The pairing is at least cE, $E = T^{-2}s\sum_j c_j^2$. The gradient Gram has row sums at most $CT^2 s$ by the same envelope overlap, hence the test gradient has squared norm at most CE under each uniformly bounded vertex density. The test Hessian has uniformly bounded sup norm by pointwise overlap and the derivative bounds. Taylor expansion in an optimal coupling gives $W_2^2 \geq cE$ as for the wavelets. Each differing bit has weight of order $T^{-2\alpha-2}s$, so $N \asymp 1/s$ bits yield the lower bound $T^{-\alpha-1}$. This proves the claimed planar inverse term for $0 < s \leq s_0$. Together with the direct terms and the larger-noise argument, it completes the proof of Theorem 5.1. Equate the two summands for the line. In the plane, equating their squares gives $X/\log^{\alpha+1}(2+X) \asymp (\log n)^2$, which yields the displayed planar transition variance.

5.6 Relation to distribution and matching problems

Classical distribution-function deconvolution [7] uses the integrated deconvolution kernel. The role of the raw empirical CDF here is to isolate its weighted tail variance from the inverse correction. Vanishing-error

density estimation is studied in [5]. Durot and Mukherjee [19] treat shrinking supersmooth noise in W_1 over moment classes. The rates here concern latent W_2 with quantitative potential regularity. In the plane, Gaussian empirical matching [13] exhibits the same direct scale for clean samples. The coarse heat fit shows how to retain it while inverting the Gaussian blur. Its bias is a single flat-top high pass applied to the latent density, which makes the real-order Taylor estimate sufficient.

6 Sharp deterministic Gaussian deconvolution with singular noise

For known Gaussian noise of arbitrary positive rank, we give the sharp deterministic inverse modulus in quadratic Wasserstein distance over log-concave inputs around a reference with two-sided potential curvature bounds. The output Wasserstein modulus has order $(\log(1/\epsilon))^{-3/2}$. For Bayes regret, the corresponding observation combines the noiseless marginal with covariance-preconditioned regret. The constants are uniform over rank and orientation when the largest noise eigenvalue stays in a fixed positive interval, even if other nonzero eigenvalues approach zero. We characterize the exact zero-regret ambiguity and show that both marginal identification and the covariance preconditioning are necessary for these uniform statements. We also give matching joint bounds in a range where the largest noise eigenvalue tends to zero.

6.1 Model and results

Let g be a probability density e^{-V}/Z on $\mathbb{R}^d$, with $V \in C^2$ and $cI \le \nabla^2 V \le CI$ for fixed $0 < c \le C$. Let f range over all log-concave probability densities on $\mathbb{R}^d$. We observe $Y = X + Z$ where $X \sim f$ and $Z \sim \mathcal{N}(0, \Sigma)$ are independent, and Σ is positive semidefinite and known. Put $U = \ker \Sigma$ and $E = U^\perp$. The first and second coordinates below refer to this orthogonal splitting. Denote the two output laws by p and q, the two noiseless marginal laws by f_U and g_U, the posterior means by m_f, m_g, and the Moore–Penrose inverse of Σ by Σ^+. The ordinary excess risk is

$$\mathrm{Reg}_{\Sigma,g}(f) = \mathbb{E}_f\|X - m_g(Y)\|^2 - \mathbb{E}_f\|X - m_f(Y)\|^2 = \mathbb{E}_p\|m_f(Y) - m_g(Y)\|^2.$$

We use the two observations

$$\begin{aligned} D_W(f) &= W_1(p, q), \\ D_S(f) &= W_1(f_U, g_U) + \left(\mathbb{E}_p\|\Sigma^+(m_f(Y) - m_g(Y))\|^2\right)^{1/2}. \end{aligned}$$

Theorem 6.1. *Fix $0 < \lambda_0 \le \Lambda < \infty$. For either $D = D_W$ or $D = D_S$ there exist $a, b > 0$ and $\epsilon_0 \in (0, 1)$ depending only on $d, c, C, \lambda_0, \Lambda$ such that, for every g as above and every Σ with $\lambda_0 \le \lambda_{\max}(\Sigma) \le \Lambda$,*

$$a[\log(1/\epsilon)]^{-3/2} \le \sup_{f:\, D(f) \le \epsilon} W_2(f, g) \le b[\log(1/\epsilon)]^{-3/2}, \qquad 0 < \epsilon \le \epsilon_0.$$

The upper bound uses only d, c, C, Λ and remains valid for all $0 \le \Sigma \le \Lambda I$. Zero observation gives $f = g$. The lower alternatives have $f_U = g_U$, exactly match the reference's mean and covariance, and have potential Hessian between $(c/2)I$ and $(C + c/2)I$.

For every fixed positive-rank Σ, the same result holds with D_S replaced by $W_1(f_U, g_U) + \sqrt{\mathrm{Reg}_{\Sigma,g}(f)}$. Uniform upper constants for that version are available under an additional positive lower bound on the nonzero eigenvalues of Σ.

Proposition 6.2. *The ordinary regret is zero precisely when f has the form $f(u, v) = a(u)g(u, v)$ for a nonnegative measurable a with $\int a\, dg_U = 1$. If $\dim U > 0$, any finite list of linear constraints $\int \phi_j f = \int \phi_j g$, with $\phi_j \in L^1(g)$, still allows distinct such f with two-sided positive potential curvature. This includes the exact mean and covariance constraints. When $\dim U = 0$, ordinary regret is identifying.*

Proposition 6.3. *Assume $d \geq 2$ and permit noise covariances with largest eigenvalue fixed at a positive number but with some nonzero eigenvalues arbitrarily small. Replacing Σ^+ in D_S by the pseudoinverse power Σ^{-r} on E for any $r < 1$ rules out a uniform inverse modulus tending to zero. This remains true with exact matching of the kernel marginal, mean and covariance. More generally, for a fixed positive spectral multiplier $b(s)$ on active eigenvalues $s \in (0, \Lambda]$, uniform stability holds if $\inf_{0<s\leq\Lambda} sb(s) > 0$. On the indicated class it fails whenever $sb(s) \to 0$ along a sequence of permitted eigenvalues for which a second eigenvalue can be held fixed at a positive value.*

In [1, Corollary 2.2], the deterministic theorem uses scalar positive noise. The positive-rank result there concerns estimation from repeated observations [1, Theorem 2.5]. Its proof already uses scalar noise completion and an anisotropic oscillation at a Gaussian reference. Here those methods supply the deterministic W_1 result at every stipulated reference. The conditional regret criterion, uniform preconditioning and identifiability claims address a different observation problem. Sugasawa–Zhao [21, §2] derive the scalar conditional score–regret identity in a model with covariates. Nguyen [20, Theorem 2] and Dedecker–Michel [6] give the Gaussian mixture and supersmooth estimation context. The upper bounds below use one prior result, the fixed scalar-noise inverse bound stated next. The lower bounds use an oscillatory construction in one active direction, which Section 6.3 carries out in full.

Scalar inverse bound ([1, Corollary 2.2 and Lemma 10.1])**.** *Fix $d \geq 1$, $\sigma > 0$ and $0 < c \leq C < \infty$. There are $K, \epsilon_* > 0$, depending only on d, σ, c, C, with the following property. Let $g = Z^{-1}e^{-V}$ be a probability density on $\mathbb{R}^d$ with $V \in C^2$ and $cI \leq \nabla^2 V \leq CI$, with arbitrary mean and covariance, and let f be any log-concave probability density on $\mathbb{R}^d$. If $\epsilon = W_1(f * \mathcal{N}(0, \sigma^2 I),\, g * \mathcal{N}(0, \sigma^2 I)) \leq \epsilon_*$, then*

$$W_2(f, g) \leq K[\log(1/\epsilon)]^{-3/2},$$

and $\epsilon = 0$ forces $f = g$.

This is the analytic core of the fixed-noise theory. Its proof (a midpoint flux bound, Fourier estimates on a growing core, convex trace estimates, a weighted high-frequency flux and an exterior correction, followed by moment recovery and alignment) is given in [1, §§4–8 and §10]. Theorem 6.4 below also uses the scale dependence of that proof.

6.2 A conditional transport inequality

Let $D = \Sigma|_E$ when E is nonzero, so D is positive definite. For u in the support of the noiseless marginal, write p_u and q_u for the conditional output densities on E. Explicitly, $p_u = f(\cdot \mid u) * \mathcal{N}(0, D)$ and $q_u = g(\cdot \mid u) * \mathcal{N}(0, D)$. Differentiating the Gaussian convolution in the noisy coordinates gives conditional versions of Tweedie's identity [23]

$$m_f(u, y) = (u, y + D\nabla_y \log p_u(y)), \qquad m_g(u, y) = (u, y + D\nabla_y \log q_u(y)). \tag{52}$$

Differentiation only requires the Gaussian smoothing within each fiber. The difference of the two posterior means lies in E, and

$$J_E(p \mid q) := \int f_U(du) \int \|\nabla_y \log(p_u/q_u)(y)\|^2 p_u(y)\, dy = \mathbb{E}_p \|\Sigma^+(m_f - m_g)\|^2. \tag{53}$$

Finite moments of log-concave densities and the linear growth of the reference output score ensure that the ordinary regret is well defined. The zero-dimensional score and J_E are defined to be zero.

A density with potential Hessian at least cI, convolved with any Gaussian of covariance at most ΛI, has potential Hessian at least κI, where $\kappa = c/(1 + c\Lambda)$. One proof first uses positive definite noise and writes $V(x) = c\|x\|^2/2 + W(x)$ with W convex. Completing the quadratic square in the convolution factors out the Gaussian with covariance $c^{-1}I + \Sigma$. The other factor is the integral of a jointly log-concave function of output and integration variable, after a linear change of integration variable. Prékopa's theorem [22] shows it is log-concave. Positive semidefinite noise follows by continuity. When $\nabla^2 V \leq CI$ as well, differentiating with respect to output via the translated prior gives $\nabla^2(-\log q) = \mathbb{E}[\nabla^2 V \mid Y] - \mathrm{Cov}(\nabla V \mid Y) \leq CI$. The same curvature inequalities apply to the conditional reference in the noisy directions, uniformly in u.

If ν has curvature at least κI, the strongly log-concave transport–information estimate is

$$W_2(\mu,\nu) \le \kappa^{-1}\sqrt{J(\mu \mid \nu)}. \tag{54}$$

This follows by combining the log-Sobolev inequality with the T_2 transport inequality, or directly by the entropy interpolation inequality [24]. Thus

$$\int W_1(p_u,q_u)\, f_U(du) \le \int W_2(p_u,q_u)\, f_U(du) \le \kappa^{-1}\sqrt{J_E(p \mid q)}.$$

We also have a Lipschitz bound on the conditional reference outputs,

$$W_1(q_u,q_{u'}) \le (C/c)\|u-u'\|. \tag{55}$$

For the conditional inputs, the difference of their v-scores is bounded in norm by $C\|u-u'\|$ from the cross Hessian bound, and their curvature is at least c. Equation (54) gives the bound for their W_2 distance. Adding the same noise proves (55).

Couple the noisy conditional output at u first to q_u, then replace the marginal f_U by g_U using (55). Conditional couplings may be chosen measurably by the usual transport disintegration. Equivalently, both steps follow by applying the Kantorovich dual formula in each fiber and using the Lipschitz constant of the integrated test in u. We obtain

$$W_1(p,q) \le (1+C/c)W_1(f_U,g_U) + (1/c+\Lambda)\sqrt{J_E(p \mid q)}. \tag{56}$$

Also $W_1(f_U,g_U) \le W_1(p,q)$ by projection. Equation (52) and the identities above cover the full-rank case by an empty U, and (56) reduces to the identity $W_1(f,g) = W_1(f_U,g_U)$ when $\Sigma = 0$.

To finish the inverse bound, add the same independent Gaussian noise of covariance $(\Lambda+1)I - \Sigma$ to each output. The new covariance of the noise is the fixed scalar $(\Lambda+1)I$, and W_1 contracts under adding the same noise. Apply the scalar inverse bound with $\sigma^2 = \Lambda + 1$; its constants depend only on d, c, C, Λ. This proves both upper moduli and the zero cases in Theorem 6.1. For a positive lower bound $\lambda^+_{\min}(\Sigma)$ on the active spectrum, (53) gives $\sqrt{J_E} \le (\lambda^+_{\min})^{-1}\sqrt{\mathrm{Reg}_{\Sigma,g}(f)}$. The reverse inequality $\sqrt{\mathrm{Reg}_{\Sigma,g}(f)} \le \Lambda\sqrt{J_E}$ follows from (52).

6.3 Sharpness with exact marginal and moment constraints

Translate the minimizer of V to zero and rotate the coordinates so that $e_1 \in E$ is an eigenvector of Σ for $s = \lambda_{\max}(\Sigma) > 0$. The reference envelopes, including $g(x) \ge k\exp(-C\|x\|^2/2)$, are uniform in d, c, C. Fix $\tau > 0$ with $\tau^{-2} > C$ and write ψ for the density of $\mathcal{N}(0,\tau^2 I)$. We use the following mass- and covariance-cancelling profile, a one-dimensional Gaussian-profile oscillation of the kind used at fixed noise in [1, §9]; every property needed here is verified below,

$$h_T(t) = \cos(Tt) - B_T\cos t + C_T, \quad B_T = T^2 e^{-\tau^2(T^2-1)/2}, \quad C_T = (T^2-1)e^{-\tau^2T^2/2}.$$

For large real T, let $f_T = g + \delta_T\psi(x)h_T(x_1)$, with $\delta_T = a_0/T^2$. If $Z_1 \sim \mathcal{N}(0,\tau^2)$, direct differentiation of $\mathbb{E}\cos(tZ_1) = e^{-\tau^2t^2/2}$ gives $\mathbb{E}h_T(Z_1) = \mathbb{E}Z_1^2h_T(Z_1) = 0$. Parity and independence give the other mean and covariance cancellations. In addition, integrating first in x_1 shows that the entire U marginal is unchanged. In fact any orthogonal projection onto a subspace of $e_1^\perp$ is unchanged.

The reference Hessian bounds show that $w = \psi/g$ and its first two derivatives are bounded uniformly: the derivative formulas have polynomial factors of degree at most two times w, and the stricter Gaussian envelope of ψ controls these factors. Applying the product and logarithm rules to $f_T = g(1+\delta_T w h_T)$ shows that with fixed sufficiently small $a_0 > 0$, the ratio f_T/g lies in $[1/2, 3/2]$ and the Hessian of $-\log(f_T/g)$ has norm at most $c/2$. Here the two derivatives of h_T cost at most a multiple of T^2, paid by δ_T. This proves the claimed curvature bounds. Testing with the 1-Lipschitz map $\cos(Tx_1)/T$ gives

$$W_2(f_T,g) \ge W_1(f_T,g) \ge a_0/(3T^3) \tag{57}$$

for large T, because $\mathbb{E}[h_T(Z_1)\cos(TZ_1)] \ge 1/2 - B_T \ge 1/3$.

Write ψ_Σ for the density of $\mathcal{N}(0, \tau^2 I + \Sigma)$,

$$b_s = \frac{\tau^2}{\tau^2 + s}, \qquad \alpha_s = \frac{\tau^2 s}{2(\tau^2 + s)}.$$

The noisy perturbation is exactly $p_T = q + \delta_T b_T$, where

$$b_T(y) = \psi_\Sigma(y)[e^{-\alpha_s T^2} \cos(b_s T y_1) - B_T e^{-\alpha_s} \cos(b_s y_1) + C_T].$$

This is conditional Gaussian integration in the eigenvector direction, including when some other noise eigenvalues are zero. Uniformly for $0 < s \le \Lambda$, the bracket has absolute bound $C_1 e^{-\alpha_s T^2}$, since $\tau^2/2 - \alpha_s \ge \tau^4/[2(\tau^2 + \Lambda)] > 0$. Its gradient is bounded by $C_1 T e^{-\alpha_s T^2}$. The first-moment test for a signed perturbation now gives

$$W_1(p_T, q) \le C_2 T^{-2} e^{-\alpha_s T^2}.$$

For the score observation, let ϕ^C_Σ be the density of $\mathcal{N}(0, C^{-1} I + \Sigma)$. The reference lower envelope gives $q \ge k_1 \phi^C_\Sigma$. For every fixed polynomial growth weight P, the integrals $\int |P(y)| \psi_\Sigma(y)^2 / q(y)\, dy$ are uniformly finite, since $\tau^2 < C^{-1}$ and the spectra of both Gaussian covariances are in fixed compact positive intervals. The two-sided curvature of q established above gives $\|\nabla \log q(y)\| \le C_3(1 + \|y\|)$ uniformly. For instance the bound at zero follows by integrating $\nabla \log q(0) - \nabla \log q(Y)$ under q, since the score has mean zero and the first moment of q is uniformly bounded. Also $1/2 \le p_T/q \le 3/2$ by conditional averaging. The Fisher expression and the derivative bound yield

$$\int p_T \|\nabla \log(p_T/q)\|^2 = \delta_T^2 \int \frac{\|\nabla b_T - b_T \nabla \log q\|^2}{q(1 + \delta_T b_T/q)} \le C_4 T^{-2} e^{-2\alpha_s T^2}.$$

The conditional noisy-coordinate Fisher integrand in (53) uses just the E projection of this score, because conditioning removes a factor depending on u alone. Thus $D_S(f_T) \le C_5 T^{-1} e^{-\alpha_s T^2}$. When $s \ge \lambda_0$, choose $T^2 = \alpha_s^{-1} \log(C_6/\epsilon)$ and use (57). Constants are uniform over directions, ranks and all other eigenvalues, giving Theorem 6.1. The same choice proves the lower bound for the ordinary-regret observation using $\sqrt{\mathrm{Reg}} \le \Lambda \sqrt{J_E}$.

6.4 Identification and small eigenvalues

For u with positive f_U density and $\dim E > 0$, the conditional output densities are positive and smooth on E. If regret is zero, (52) implies their scores agree p_u-almost everywhere for f_U-almost all u. By continuity they agree everywhere in those fibers, so the densities are equal after normalization. Gaussian convolution is injective on each fiber by its nonvanishing Fourier multiplier. It follows that $f(\cdot \mid u) = g(\cdot \mid u)$ there, exactly the multiplier assertion of Proposition 6.2. The converse is immediate in (52). For $E = \{0\}$ both posterior means are the identity and the same characterization describes every density.

For $\dim U > 0$, write $a(u) = 1 + \epsilon h(u)$ with $h \in C_c^\infty(U)$. Requiring $\int h\, dg_U = 0$ and $\int \phi_j(u, v) h(u) g(u, v)\, du\, dv = 0$ imposes finitely many homogeneous linear constraints on h. There exists a nonzero such h, for example by taking more disjoint bumps than constraints and a nonzero vector in the kernel of the constraint matrix. Small ϵ gives positivity and changes the Hessian by at most $c/2$ via the two bounded derivatives of $\log(1 + \epsilon h)$. Since g_U is everywhere positive and $h \ne 0$, $f \ne g$. This proves Proposition 6.2.

For Proposition 6.3 take a product Gaussian reference and a fixed smooth log-concave alternative differing only on one axis. Apply the above oscillatory construction in dimension one at any sufficiently large fixed T to choose alternative and reference with identical mean and variance. Keep a second axis with noise eigenvalue $\lambda_0 > 0$, and let the first axis have positive eigenvalue $s \downarrow 0$ with all other axes unchanged. On the differing axis, the score discrepancy of the noisy densities has bounded Fisher integral uniformly in s, by the above estimate or uniform curvature and first two moments. Hence the covariance-inverse power observation on this axis is at most $C s^{1-r}$ by (52). The kernel marginal and all latent mean and covariance entries match, and latent W_2 is bounded away from zero by the differing one-dimensional projection. This proves the claimed failure for every $r < 1$. With multiplier $b(s)$ the same observation along this axis is bounded by $C s b(s)$. Conversely a positive lower bound on $s b(s)$ bounds $\sqrt{J_E}$ by a fixed multiple of the multiplier observation, so Theorem 6.1 applies.

A threshold alternative using unweighted regret. For $a > 0$, write E_a for the span of noise eigenvectors with eigenvalue at least a, and $L = E_a^\perp$. Observing the output marginal on L and the ordinary regret yields

$$W_1(p,q) \le (1 + C/\kappa) W_1(p_L, q_L) + \kappa^{-1} a^{-1} \sqrt{\mathrm{Reg}_{\Sigma,g}(f)}.$$

To see this, disintegrate the *outputs* along L, using the two-sided curvature of q for the conditional transport and Lipschitz estimates. Their conditional score difference in E_a has squared integral at most $a^{-2}\,\mathrm{Reg}$ by projecting (52). This is (56)'s argument applied directly to the outputs. The scalar completion still gives the inverse upper bound. Since $W_1(p_L, q_L) \le W_1(p,q)$, the lower construction gives a matching modulus for every fixed a. If E_a contains the largest-eigenvalue axis, the construction even leaves this projected output marginal unchanged. Here even a weakly noisy direction can be handled as a marginal when its eigenvalue lies below the selected threshold.

Other noiseless marginal distances. The bounded-Lipschitz distance can replace $W_1(f_U, g_U)$ with the same logarithmic exponent. Small bounded-Lipschitz distance places at least $3/4$ of both probabilities in a fixed ball after common translation, by use of a Lipschitz cutoff and the uniform reference tails. A log-concave law with this much mass in a fixed ball has uniform exponential norm tails. For example if K is the ball and $t > 1$, the combination $(1-s)K + s(\mathbb{R}^{\dim U} \setminus tK)$ lies outside K at $s = 2/(t+1)$, so log-concavity bounds the exterior mass of tK by $(1-\mu(K))^{1/s}\mu(K)^{-(1-s)/s}$. Truncating unit-Lipschitz tests at size $O(\log(1/\delta))$ shows $W_1 \le C\delta\log(1/\delta)$ for bounded-Lipschitz discrepancy δ small. This suffices in the upper bound, and the lower construction keeps the marginal identical. Total variation and stronger discrepancies such as Hellinger distance similarly suffice by domination of bounded-Lipschitz distance up to fixed constants.

6.5 A uniform range for shrinking noise

The dependence on the largest noise eigenvalue can be retained in part of the joint small-noise and small-observation range.

Theorem 6.4. *There are $\alpha > 0$ and $a, b, \epsilon_0 > 0$, depending only on d, c, C, Λ, such that for either observation D_W, D_S, all $0 < \epsilon \le \epsilon_0$, and all Σ with $0 < s = \lambda_{\max}(\Sigma) \le \Lambda$,*

$$a\left(\frac{s}{\log(1/\epsilon)}\right)^{3/2} \le \sup_{f:\, D(f)\le\epsilon} W_2(f,g) \le b\left[\epsilon^\alpha + \left(\frac{s}{\log(1/\epsilon)}\right)^{3/2}\right].$$

The upper bound includes $s = 0$. In particular the two bounds match uniformly when $s \ge \epsilon^\beta$ for any fixed $0 < \beta < 2\alpha/3$ and all sufficiently small ϵ.

Proof. The lower construction used above is uniform for $0 < s \le \Lambda$, with α_s bounded above and below by constant multiples of s. With $T^2 = \alpha_s^{-1}\log(C_6/\epsilon)$, inequality (57) gives exactly the claimed lower bound.

For the upper bound, either observation gives $\delta = W_1(p,q) \le K\epsilon$ by (56). Recenter the common reference at its mean. As in [1, §10], first moments of the output give $\|\mathbb{E}_f X\| \le \delta$, and uniform exponential input tails follow from the first-moment bound and log-concavity. The tail and first-moment bounds here are uniform for bounded noise covariances. A smooth cutoff of each centered output second-moment test at radius L gives a discrepancy $O(L\delta + L^{-2})$ using the uniform fourth moments. The known common noise cancels in the difference of covariances. Thus $\|C_f - C_g\| = O(\epsilon^{2/3})$, where C_f, C_g are the latent covariance matrices. The reference eigenvalues lie in $[C^{-1}, c^{-1}]$ by the two-sided curvature, and C_f lies in a fixed positive spectral interval for small ϵ. Center the unknown input and transport it by $C_g^{1/2}C_f^{-1/2}$ to match the reference covariance. The latent W_2 cost of this operation, including centering, is $O(\epsilon^{2/3})$ by the matrix square-root Lipschitz bound on a fixed positive spectral interval. The same-noise coupling bounds the output W_1 change by that amount. Whitening both aligned laws by $C_g^{-1/2}$ gives isotropic inputs, a reference with two-sided curvature bounds $c/C, C/c$, and output discrepancy at most $K'\epsilon^{2/3}$. Their noise covariance lies between zero and CsI. Add common noise to reach exactly CsI, keeping the discrepancy bounded by $K'\epsilon^{2/3}$. For $s = 0$ this step adds nothing.

We now track the scale in the actual scalar upper proof [1, §§4–8]. Use its fixed frequency kernels and let B be the common support radius of their Fourier cutoffs. Its Lemma 4.2 bounds the smoothed discrepancy through order two, for any frequency cutoff R, by

$$K''\epsilon^{2/3}R^{d+3}\exp(CsB^2R^2/2).$$

The squared low-frequency flux before the density weight is at most $K''\epsilon^{4/3}R^d\exp(CsB^2R^2)$. The proof uses the first bound at most R^{-N_0}, with a fixed N_0 depending on the standardized curvature and dimension. In its core-energy estimate the second bound is multiplied by at most $C'R^P$ for a fixed P, and must be at most a constant times R^{-6}. Once these two requirements hold, the core comparison, convex curvature measure, weighted high-frequency flux and exterior correction prove a squared transport bound $O(R^{-6})$. These remaining estimates depend on R and the standardized latent reference class, with no use of the observation covariance. In particular the reference envelopes and all lower thresholds are uniform by the argument of [1, Lemma 10.1].

Take

$$R=\min\left\{\epsilon^{-\theta},\left(\frac{b_0\log(1/\epsilon)}{s}\right)^{1/2}\right\}$$

with second entry interpreted as infinity when $s=0$. Choose b_0 so small that $CsB^2R^2\le\frac{1}{6}\log(1/\epsilon)$, then choose fixed $\theta>0$ sufficiently small in terms of d,N_0,P and with $3\theta<2/3$. The exponential factors are paid by the displayed powers of ϵ, and both frequency requirements follow from $R\le\epsilon^{-\theta}$. Also $R\to\infty$ uniformly over bounded s. This gives $W_2\le C_1R^{-3}$ in the whitened comparison. Undoing the alignment and whitening adds $O(\epsilon^{2/3})$ and bounded multiplicative constants. Taking $\alpha=3\theta$ proves the upper bound. The matching range follows because $\epsilon^\alpha=o((\epsilon^\beta/\log(1/\epsilon))^{3/2})$ under the indicated strict inequality on β. □

7 Reverse heat displacement and vanishing-noise Gaussian inverse bounds

The scaled score of a log-concave law increases in convex order along a possibly singular Gaussian heat flow. One coupling then bounds every convex displacement cost by a scaled endpoint score, with sharp constants for all finite Wasserstein orders. The estimate gives a rank-uniform linear inverse modulus from the active output score and the noiseless marginal when the largest noise eigenvalue is below the observation tolerance. For inputs with a bounded potential Hessian, a relative-score interpolation extends the linear regime to the two-thirds power of that tolerance. In one dimension we also determine the main logarithmic power in the inverse modulus from output W_1 at very small noise, with a remaining iterated-logarithm gap.

7.1 Observations and results

Let Σ be a positive semidefinite covariance matrix on $\mathbb{R}^d$. Write $U=\ker\Sigma$, $E=U^\perp$, and $s=\lambda_{\max}(\Sigma)$. Let $p=f*\mathcal{N}(0,\Sigma)$ and $q=g*\mathcal{N}(0,\Sigma)$, with f a log-concave probability density and $g=e^{-V}/Z$ a reference with $V\in C^2$ and $cI\le\nabla^2V\le CI$. The known factor Σ can be singular, and a zero-dimensional marginal distance is zero. For U coordinates u, let p_u,q_u denote the conditional output densities on E. Define

$$J_E(p\mid q)=\int f_U(du)\int_E\|\nabla_E\log(p_u/q_u)\|^2p_u\,dy,\qquad D_2=W_2(f_U,g_U)+\sqrt{J_E(p\mid q)}.$$

Equivalently J_E is the mean squared difference of the posterior means preconditioned by the Moore–Penrose inverse Σ^+, by the conditional Tweedie identity [21] and Section 6. Define $D_W=W_1(p,q)$. The observations here preserve full freedom of latent means and covariances.

Theorem 7.1. *Let μ be a log-concave probability measure and let $\mu_t=\mu*\mathcal{N}(0,t\Sigma)$ with $\Sigma\ge0$. Write active scores by disintegrating over the unchanged U marginal. For any fixed linear map A on E, the law of*

$\sqrt{t}\,A\nabla_E \log \mu_t$ *under* μ_t *increases in convex order for* $t > 0$. *For each* $b > 0$ *there is one coupling* $(X_t)_{0\le t\le b}$ *of the heat marginals such that for every* $0 \le a < b$,

$$X_b - X_a \ \preceq_{\mathrm{cx}}\ -(b - \sqrt{ab})\Sigma\nabla_E \log \mu_b(X_b).$$

In particular for $1 \le p < \infty$,

$$W_p(\mu, \mu_1) \le \left(\int \|\Sigma\nabla_E \log \mu_1\|^p \, d\mu_1\right)^{1/p}.$$

The constant 1 *is sharp. At rank zero all the asserted scores and displacements are zero.*

Theorem 7.2. *Fix* d, c, C *and an upper bound for* s. *Uniformly over ranks and orientations, for* $D_2 \le \epsilon \le 1$,

$$W_2(f, g) \le C_1(\epsilon + s).$$

If the unknown log-concave potential is C^2 *with Hessian bounded above by* C' *as well, then*

$$W_2(f, g) \le C_2(\epsilon + s^{3/2}).$$

Thus the supremum is of order ϵ *for* $0 < s \lesssim \epsilon$ *over all log-concave inputs, and for* $0 < s \lesssim \epsilon^{2/3}$ *over the subclass with bounded potential Hessian that contains translations of* g. *Using the curvature-preserving oscillatory alternatives of Section 6.3, that smoother-class supremum is bounded below for* $s > 0$ *by a constant times*

$$\max\left\{\epsilon,\ \left[\frac{s}{1 + \log_+(\sqrt{s}/\epsilon)}\right]^{3/2}\right\}.$$

Here $\log_+ x = \max(0, \log x)$. *Both upper statements apply to the full-rank preconditioned score observation itself and also to singular noise with exactly matched noiseless marginals.*

Theorem 7.3. *In dimension one, with scalar noise variance* $0 \le s \le \epsilon$, *put* $L = \log(1/\epsilon)$. *Uniformly for small* $\epsilon > 0$,

$$c_3 \frac{\sqrt{\epsilon}}{L^{1/4}\sqrt{\log L}} \le \sup_{f:\, D_W \le \epsilon} W_2(f, g) \le C_3 \frac{\sqrt{\epsilon}}{L^{1/4}}.$$

The lower construction is admissible under every additional common Gaussian convolution. In particular the power $-1/4$ *on* L *is sharp when an iterated-logarithm factor is allowed.*

At fixed positive noise the sharp inverse around a two-sided-curvature reference is logarithmic under the observations of [1] and Section 6. The present theorems concern the joint range where the noise approaches zero. For singular noise with an unknown noiseless marginal, the score theorem assumes W_2 control of that marginal. The one-dimensional result uses W_1 as its sole observation. The distinction also appears with score observation and zero noise, since the active score disappears at rank zero.

7.2 Convex order of the heat score and displacement

Work first on a single active subspace fiber. Let Σ_E be its positive covariance, p_t its convolved density, $q_t = \nabla_E \log p_t$ its score, and $H_t = -\nabla_E^2 \log p_t$. Convolution preserves log-concavity, and differentiation of the Gaussian kernel gives

$$H_t = (t\Sigma_E)^{-1} - (t\Sigma_E)^{-1}\,\mathrm{Cov}(X \mid Y_t)(t\Sigma_E)^{-1}.$$

Consequently $0 \le H_t \le (t\Sigma_E)^{-1}$ and $H_t\Sigma_E H_t \le H_t/t$, by conjugation and squaring the bounded positive matrix $t\Sigma_E^{1/2} H_t \Sigma_E^{1/2}$. The posterior identity includes priors supported on a smaller affine subspace.

For a fixed linear map A, differentiate the expectation of a smooth test function of Aq_t along $\partial_t p_t = \frac{1}{2}\Sigma_E : \nabla_E^2 p_t$. If A is held fixed in this differentiation, the result is

$$\frac{d}{dt}\int \Phi(Aq_t)p_t = -\frac{1}{2}\int \mathrm{tr}\big[\nabla^2\Phi(Aq_t)AH_t\Sigma_E H_t A^\top\big]p_t.$$

Indeed, $\partial_t q_t = \frac{1}{2}\Sigma_E : \nabla_E^2 q_t + (\nabla_E q_t)\Sigma_E q_t$. Integration by parts in the heat term of the score generates the squared Jacobian contribution and one copy of the drift term with coefficient $-1/2$. The derivative of the density gives its other copy with coefficient $-1/2$. This is the convex-test version of the ordinary Fisher dissipation identity [25, §7]. Differentiation and integrations by parts may be justified on $t \in [\tau, b]$ for $\tau > 0$ by smooth cutoffs. Every convolved density derivative divided by p_t is a conditional expectation of a Gaussian Hermite polynomial multiplied by a coefficient bounded on that interval. Jensen supplies moments of every fixed order for the ratios, independent of the noiseless fiber and prior, so the cutoff errors vanish. The same estimates control scaled scores.

Now put $Z_t = \sqrt{t}\, Aq_t$. In the scaled differentiation one adds $(2t)^{-1}\mathbb{E}[\nabla\Phi(Z_t)\cdot Z_t]$, which by one more integration by parts equals $\frac{1}{2}\mathbb{E}\,\mathrm{tr}(\nabla^2\Phi(Z_t)AH_tA^\top)$. Thus

$$\frac{d}{dt}\mathbb{E}\Phi(Z_t) = \frac{1}{2}\mathbb{E}\,\mathrm{tr}\left[\nabla^2\Phi(Z_t)A(H_t - tH_t\Sigma_E H_t)A^\top\right] \geq 0 \tag{58}$$

for convex Φ. Start with smooth convex tests whose derivatives have polynomial growth. General convex tests follow by increasing maxima of finitely many affine supporting functions, smoothing each such maximum and then taking limits. Each score has mean zero. Regular conditional laws of a log-concave measure on affine slices are log-concave on their relative supports almost everywhere, by disintegrating its density on the affine hull. Therefore the preceding argument holds for almost every active fiber. Averaging over their unchanged noiseless marginal proves the stated convex order in the full law.

For $d = 1$, the normalized score is a constant multiple of the negative posterior mean of a standard Gaussian noise given its sum with $X/\sqrt{t}$. Its convex-order monotonicity also follows from Mizuno's comparison of scalar additive signals [27, Proposition 1 and Remarks 1, 3]: the scaled independent inputs are ordered by dispersive spread, and the Gaussian posterior regression in a log-concave additive-noise model is increasing. The scalar quadratic case follows as well from the conditional variance bound and the derivative of the Gaussian minimum mean square error [26, Proposition 9]. Formula (58) applies to all projected convex tests in the multivariate setting.

To obtain the single coupling, start at $X_b \sim \mu_b$ and solve the deterministic heat-flow equation backward with velocity $-\frac{1}{2}\Sigma\nabla_E \log \mu_t$ on each active fiber, keeping the noiseless coordinate fixed. On each closed positive time interval the field is globally Lipschitz in the active variable by the Hessian bound. Its magnitude has at most linear growth there, its time dependence is smooth, and its dependence on the fixed fiber is measurable. These flows transport the fiber laws by the continuity equation, since the velocity times density is $-\frac{1}{2}\Sigma_E\nabla_E p_t$; this may also be checked directly by change of variables along characteristics. The trajectories form consistent couplings on intervals approaching zero. By the convex order with $A = \Sigma_E$, their expected speed is integrable down to zero, bounded by $\mathbb{E}\|\sqrt{b}\,\Sigma\nabla_E \log \mu_b\|/(2\sqrt{t})$. Hence trajectories converge almost surely as $t \downarrow 0$, and the time-zero marginal is μ since convolution by variance $t\Sigma$ tends to the identity. This constructs the coupling at all $t \in [0, b]$.

For $0 \leq a < b$ write

$$X_b - X_a = (\sqrt{b} - \sqrt{a}) \int_a^b \left[\, -\sqrt{t}\,\Sigma\nabla_E \log \mu_t(X_t)\right] \frac{dt}{2\sqrt{t}\,(\sqrt{b} - \sqrt{a})}.$$

The measure in the integral is a probability. Jensen's inequality and the convex order of the reflected normalized score prove $X_b - X_a \preceq_{\mathrm{cx}} -(b - \sqrt{ab})\Sigma\nabla_E \log \mu_b(X_b)$. At a point-mass input the trajectory adds the same Gaussian direction with amplitude $\sqrt{t}$, which gives equality in distribution for all these increments. Applying convex costs $\|\cdot\|^p$ gives the Wasserstein bounds, including $p = 2$ for the sequel. For a centered Gaussian input with covariance tending to zero, these bounds converge to equality, so sharpness does not require a singular input.

The order is along $t\Sigma$ with a fixed covariance shape. Arbitrary increasing covariance matrices need not order their normalized posterior noise in convex order. For example, take a centered two-dimensional Gaussian input with covariance C with $C_{12} \neq 0$, and noise $\Sigma = \mathrm{diag}(a, 1)$. The posterior mean of the standardized noise has covariance $\Sigma^{1/2}(C + \Sigma)^{-1}\Sigma^{1/2}$. Its second diagonal entry decreases strictly as a increases, which prevents convex-order increase for the quadratic test on the second coordinate.

7.3 Score inverse bounds as noise vanishes

The conditional reference outputs q_u have potential Hessian at least κI with $\kappa = c/(1+cs)$, by convolution curvature. Their inputs have curvature at least cI, and their E-scores at two noiseless coordinates differ by at most $C\|u-u'\|$. The transport–information inequality for a κ-strongly log-concave law is $W_2 \le \kappa^{-1}\sqrt{J}$, from the log-Sobolev and quadratic transport inequalities [24]. Applying it within each fiber and using the reference's conditional Lipschitz bound gives

$$W_2(p,q) \le (1+C/c)W_2(f_U,g_U) + \kappa^{-1}\sqrt{J_E(p \mid q)}. \tag{59}$$

Indeed, in the first step keep u fixed and couple p_u to q_u, incurring root mean square cost at most $\kappa^{-1}\sqrt{J_E}$. In the second step optimally couple the noiseless marginals and use the Lipschitz conditional bound for $q_u, q_{u'}$, obtained for the conditional inputs by $T_2 I$ at curvature c and then adding the same Gaussian. The two steps combine by the triangle inequality.

A full reference output has potential Hessian between κI and CI. The lower bound follows by factoring out the Gaussian of covariance $c^{-1}I + \Sigma$ and using Prékopa's theorem on the remaining kernel. The upper bound follows by differentiating the translated prior in the convolution: the potential Hessian is the conditional mean of $\nabla^2 V$, minus the conditional covariance of ∇V. Positive semidefinite noise follows with the same bounds by continuity, or directly by conditioning. Center the reference at its mode. The score grows at most linearly with a uniform bound at the origin, since its expectation under q is zero, it is uniformly Lipschitz, and the reference output has a bounded first moment. Thus (59) for $D_2 \le \epsilon$ gives uniform second moments for p, and

$$\left(\int \|\Sigma \nabla_E \log p\|^2 \, dp\right)^{1/2} \le s\left(\sqrt{J_E(p \mid q)} + \|\nabla_E \log q\|_{L^2(p)}\right) \le Cs.$$

Conditional scores are the projections of full scores where full derivatives exist, and the identity between the reference conditional score and the active part of its full score always holds. The same absolute endpoint bound holds for q. Theorem 7.1 and the triangle inequality with (59) yield $W_2(f,g) \le C(\epsilon + s)$.

For the smoother class we need a relative-score interpolation. If two smooth densities μ, ν have uniformly Lipschitz scores and bounded second moments, set $D = W_2(\mu,\nu)$ and $a = \nabla_E \log(\mu/\nu)$. Then, for D in a fixed bounded range,

$$\|a\|_{L^2(\mu)} \le CD^{1/3}. \tag{60}$$

To prove it, mollify the globally Lipschitz active vector field a at scale $\delta \le 1$ to a_δ. The uniform score and moment bounds give

$$\|a - a_\delta\|_\infty \le C\delta, \qquad \|\nabla a_\delta\|_\infty \le C, \qquad \|\nabla^2 a_\delta\|_\infty \le C/\delta, \qquad \|a_\delta(x)\| \le C(1+\|x\|).$$

The bound of each score at zero again follows by its zero mean and its Lipschitz bound. Write $\psi = -\operatorname{div}_E a_\delta - a_\delta \cdot \nabla_E \log \nu$. Integration by parts gives $\int a \cdot a_\delta \, d\mu = \int \psi \, d(\mu - \nu)$. Since $\|\nabla\psi(x)\| \le C(1 + \delta^{-1} + \|x\|)$, an optimal W_2 coupling and Cauchy–Schwarz bound this integral by $CD(1+\delta^{-1})$. The mollifier error costs $C\delta\|a\|_{L^2(\mu)}$. Absorbing half of the square and choosing $\delta = D^{1/3}$ for $D \le 1$ gives (60). The bounded larger-D case follows from score growth, and $D = 0$ is direct.

Now take $p_t = f * \mathcal{N}(0, t\Sigma)$ and $q_t = g * \mathcal{N}(0, t\Sigma)$ for $0 \le t \le 1$. Their scores are uniformly Lipschitz by log-concavity and the Hessian upper bound preserved by convolution, and their second moments are uniformly bounded. Let $D_0 = W_2(f,g)$. Shared-noise coupling gives $W_2(p_t, q_t) \le D_0$ at all t. The deterministic heat transport velocities

$$-\tfrac{1}{2}\Sigma\nabla \log p_t, \qquad -\tfrac{1}{2}\Sigma\nabla \log q_t$$

are uniformly Lipschitz, with time-continuous flows also at $t = 0$ since the input potentials are C^2. These flows preserve the noiseless coordinate. Pull an optimal coupling of p, q backward through the two flows. The reference velocity has Lipschitz constant at most Cs. The difference of velocities evaluated on the first coupled point has root mean square at most $(s/2)\|\nabla_E \log(p_t/q_t)\|_{L^2(p_t)}$. By Grönwall and (60), this yields

$$D_0 \le C\epsilon + Cs\int_0^1 \|\nabla_E \log(p_t/q_t)\|_{L^2(p_t)} \, dt \le C\epsilon + C' s D_0^{1/3}.$$

If $D_0 > 2C\epsilon$ this forces $D_0 \le C'' s^{3/2}$. This proves the smoother upper bound, including singular covariances.

For the linear lower bound at $s > 0$ translate g by a vector $v \in E$. This leaves the noiseless marginal unchanged and gives latent distance $\|v\|$. Its output is the corresponding translation of q, and the Lipschitz score gives $\sqrt{J_E} \le C\|v\|$. At rank zero, $D_2 = W_2(f, g)$, so the exact inverse is immediate. For the additional lower bound at positive rank, the mass- and moment-preserving alternatives f_T of Section 6.3, in the highest noise eigendirection and with Gaussian profile, leave the noiseless marginal unchanged. The anisotropic calculation of Section 6 gives score discrepancy at most $CT^{-1}e^{-c_1 sT^2}$ and latent distance at least $c_2 T^{-3}$ for all sufficiently large real T, uniformly in $0 < s \le \Lambda$. These estimates use no lower bound on the remaining eigenvalues. Setting $T^2 = A[1 + \log_+(\sqrt{s}/\epsilon)]/s$, with A sufficiently large, puts the discrepancy below ϵ and yields the second lower bound. The curvature bound for these alternatives can be taken as $C + c/2$, or any prescribed upper bound strictly larger than C after reducing the perturbation size. This proves Theorem 7.2.

7.4 The one-dimensional output Wasserstein modulus

We first prove the upper bound without noise. Center g at its mode and set $\delta = W_1(f, g)$ and $L_\delta = \log(1/\delta)$, with $0 < \delta$ small. Write Q_f, Q_g for the quantile functions. In dimension one their common-quantile coupling is optimal in both orders [4], so

$$\delta = \int_0^1 |Q_f(v) - Q_g(v)|\, dv, \qquad W_2^2(f, g) = \int_0^1 |Q_f(v) - Q_g(v)|^2\, dv.$$

Take a fixed sufficiently small $a > 0$. There exists an upper-tail probability $p_+ \in [\delta^a, 2\delta^a]$ for which

$$|Q_f(1 - p_+) - Q_g(1 - p_+)| \le \delta^{1-a}.$$

This follows by averaging the W_1 integrand over that interval. There is an analogous lower probability p_- with the same bound. The reference quantile at either endpoint is bounded in absolute value by $C\sqrt{L_\delta}$ by its Gaussian upper tail. The medians are uniformly bounded since the first moments are, by the observation and the reference moments.

For a log-concave density, the negative logarithm of the survival function is convex [4, Appendix B]. Its secant between the median and the upper selected quantile has slope at least $c_a\sqrt{L_\delta}$, since the survival logarithms differ by $\log(1/p_+) - \log 2$ and the points are at distance at most $C\sqrt{L_\delta}$. Beyond the selected quantile the same lower slope holds by convexity, including for bounded support. Thus $t \mapsto Q_f(1 - p_+e^{-t})$ for $t \ge 0$ is $C_a/\sqrt{L_\delta}$-Lipschitz. The same holds for g and for the lower tail by reflection. On the upper tail let $d(t) = |Q_f(1 - p_+e^{-t}) - Q_g(1 - p_+e^{-t})|$. Integration by parts gives

$$\int_0^\infty d(t)^2 e^{-t}\, dt \le d(0)^2 + \frac{C_a}{\sqrt{L_\delta}} \int_0^\infty d(t) e^{-t}\, dt.$$

Indeed the boundary at infinity vanishes by Lipschitz growth. Multiplying by p_+, the boundary term is $O(\delta^{2-a})$ and the last integral is bounded by the corresponding W_1 integral times $C_a/\sqrt{L_\delta}$. Both tails therefore cost at most $C\delta/\sqrt{L_\delta}$.

On the central probability interval $[p_-, 1 - p_+]$, the reference quantile is Lipschitz with constant $K \le C\delta^{-C'a}$ even after enlarging that interval by $\delta^a/2$ on both ends. This follows from its positive Gaussian lower density bound on a quantile band of radius $C\sqrt{aL_\delta}$. If the pointwise quantile discrepancy has size D, monotonicity of Q_f and the Lipschitz bound on Q_g force discrepancy at least $D/2$ on a one-sided interval of length at least $\min(\delta^a/2, D/(2K))$, choosing the side according to its sign. The W_1 integral then gives $D \le C \max(\delta^{1-a}, \sqrt{\delta K})$. Choose a small, so this is $o(L_\delta^{-1/2})$. The central squared transport cost is bounded by this supremum times δ. We have proved

$$W_2^2(f, g) \le C\, \delta/\sqrt{\log(1/\delta)}. \tag{61}$$

For the matching main logarithmic power choose a large $S > 0$, let $\ell = V'(S) \asymp S$, and construct the following log-concave density. Keep g on $x \le S$, use its tangent exponential $g(S)e^{-\ell(x-S)}$ up to $T = S + A/\ell$, and take zero to the right of T. Choose A to preserve the total right-tail mass beyond S. Explicitly,

$$e^{-A} = 1 - \ell \Pr_g(X > S)/g(S) = 1 - \int_0^\infty e^{-z - r_S(z)}\, dz, \qquad r_S(z) = V(S + z/\ell) - V(S) - z.$$

The convexity of V and its two Hessian bounds give $cz^2/(2\ell^2) \le r_S(z) \le Cz^2/(2\ell^2)$. Thus $e^{-A} \asymp \ell^{-2}$ and $A \asymp \log \ell$. The slope is unchanged at S, constant after it, and the hard cutoff preserves convexity of the extended potential. On (S,T) the new density is at least g, and above T it is zero. Since the tail masses match, the new law is stochastically dominated by g and its W_1 difference equals the mean difference. The scaled tail first moments are

$$\int_0^\infty z e^{-z-r_S(z)}\,dz = 1 - O(\ell^{-2}), \qquad \int_0^A z e^{-z}\,dz = 1 - (1+A)e^{-A}.$$

Their difference is comparable to Ae^{-A}, so the observation scale is $\delta \asymp g(S)(\log \ell)/\ell^4$. Every coupling has squared cost at least $\mathbb{E}_g(X-T)_+^2 \ge c_4 g(T)/\ell^3$, by using $x - T$ of order ℓ^{-1} and the upper Hessian bound. Since $g(T) \asymp g(S)e^{-A}$, the cost is at least $c_5 g(S)/\ell^5$. The Gaussian envelopes yield $\ell \asymp \sqrt{\log(1/\delta)}$, which is the lower bound in Theorem 7.3. The continuous construction parameter S gives every sufficiently small target δ, since the positive continuous W_1 observation tends to zero. Common convolution contracts W_1, so the construction stays admissible in the noisy problem.

To transfer the upper bound to $0 \le s \le \epsilon$, first use independent noise coupling to get $W_1(f,g) \le \epsilon + 2\sqrt{s} = o(1)$. This places at least $3/4$ of the mass of f in a fixed bounded interval, uniformly after centering g. The log-concavity of its survival tails then gives uniform exponential tails by extrapolation from fixed quantile levels. Using cutoff second-moment tests yields convergence of $\mathrm{Var}(f)$ to $\mathrm{Var}(g)$ uniformly, with a positive lower bound. A one-dimensional log-concave density satisfies $\|f\|_\infty \le 1/\sqrt{\mathrm{Var}(f)}$ [4, Appendix B]. It is unimodal, so $\|Df\|_{\mathrm{TV}} \le 2\|f\|_\infty = O(1)$ even for a hard cutoff. If F_t is the distribution function under scalar Gaussian variance t, then $\partial_t F_t = \frac{1}{2}\partial_x f_t$ for $t > 0$. The L^1 expression for W_1 and $\|\partial_x f_t\|_1 \le \|Df\|_{\mathrm{TV}}$ give

$$W_1(f, f * \mathcal{N}(0,s)) \le (s/2)\|Df\|_{\mathrm{TV}}.$$

The same holds for g. Now $W_1(f,g) \le C(\epsilon + s) \le C'\epsilon$, so (61) gives the asserted upper bound. This completes the proof of Theorem 7.3.

8 Exact Wasserstein moduli at log-concave tails

We determine the exact small-W_1 modulus of W_r^r, for each $1 < r < \infty$, around a one-dimensional log-concave reference with two power-type potential tails. The modulus contains an iterated logarithm, and the constant is $\Gamma(r+1)\zeta(r)$ after normalization by the reference tail slope. Convexity forces the extremal tail discrepancy to persist across a long interval, and a tangent-exponential cutoff attains the bound. Under common Gaussian smoothing the same leading constant holds below an explicit crossover variance. Additive upper and lower bounds identify that crossover, including its logarithmic factors. For a Gaussian reference, they also match in order throughout every bounded variance range.

8.1 Statement

Let $g = e^{-V}$ be a positive probability density on $\mathbb{R}$, where $V \in C^2$ is convex. On each side, use the reflected potential on the left and the original potential on the right. Assume that for all sufficiently large positive x these potentials satisfy

$$V_\pm'(x) \asymp x^{\beta_\pm - 1}, \qquad V_\pm''(x) \asymp x^{\beta_\pm - 2}, \qquad \beta_\pm > 1.$$

The comparison constants are fixed. In particular the assumptions cover a reference with two positive finite Hessian bounds, with $\beta_+ = \beta_- = 2$. For probability mass p let $Q_g^+(p)$ be the upper tail quantile, and $Q_g^-(p)$ the upper tail quantile after reflection. Write

$$H_\pm(\varepsilon) = V_\pm'\big(Q_g^\pm(\varepsilon)\big), \qquad H_*(\varepsilon) = \min\{H_+(\varepsilon), H_-(\varepsilon)\}, \qquad L_\varepsilon = \log\log(1/\varepsilon).$$

Theorem 8.1. *For every fixed $1 < r < \infty$,*

$$\sup_{\substack{f \text{ log-concave} \\ W_1(f,g) \le \varepsilon}} W_r(f,g)^r \sim \Gamma(r+1)\zeta(r)\,\frac{\varepsilon}{H_*(\varepsilon)^{r-1} L_\varepsilon} \qquad (\varepsilon \downarrow 0).$$

For standard Gaussian g, the squared W_2 modulus is asymptotic to

$$\frac{\pi^2}{3\sqrt{2}}\,\frac{\varepsilon}{\sqrt{\log(1/\varepsilon)}\,\log\log(1/\varepsilon)}.$$

Theorem 8.2. *Replace the discrepancy constraint by $W_1(f*\gamma_s, g*\gamma_s) \le \varepsilon$, with $\gamma_s = N(0,s)$. Put $t = \log(1/\varepsilon)$ and set*

$$A_\varepsilon = \frac{\varepsilon}{H_*(\varepsilon)^{r-1}L_\varepsilon}, \qquad s_c = tA_\varepsilon^{2/(3r)}.$$

For each fixed $\Lambda < \infty$, uniformly for $0 \le s \le \Lambda$, the noisy modulus is at most

$$\big(\Gamma(r+1)\zeta(r) + o(1)\big)A_\varepsilon + C(s/t)^{3r/2}.$$

There are fixed $b > 0$ and $c_r > 0$ such that the noisy modulus is at least

$$\big(\Gamma(r+1)\zeta(r) - o(1)\big)A_\varepsilon + c_r\big(\min(s, b/t)/t\big)^{3r/2}.$$

Consequently it has the direct equivalent if and only if $s = o(s_c)$. For a standard Gaussian reference, the same lower bound holds with s in place of $\min(s, b/t)$.

The monotone-quantile and CDF identities used below are classical [4]. General Wasserstein and negative Sobolev inequalities [28] provide analytic comparisons under different norms. The fixed-positive Gaussian inverse modulus around a two-curvature reference is studied in [1]. The small-variance and noiseless behavior here is controlled by a tail constraint from log-concavity, in addition to the transport identities.

8.2 Reference tail estimates

Write $S_u(x) = \int_x^\infty u$ and $Q_u(p)$ for its inverse upper quantile. We analyze the right side; reflection gives the same statements for the left. Put $t = \log(1/\varepsilon)$, $L = \log t$, $H = H_+(\varepsilon)$, and $m_0 = \varepsilon H/L$. The constants below may depend on the reference and r. If $p \downarrow 0$ and $x = Q_g(p)$, the curvature assumptions give

$$\log(1/p) \asymp x^{\beta_+}, \qquad p = \frac{g(x)}{V'(x)}\left(1 + O\big((\log(1/p))^{-1}\big)\right).$$

For the second formula write the survival integral with $u = V'(x)(y-x)$. Convexity bounds the integrand above by e^{-u}. For $y \in [x, x+x/2]$, the quadratic Taylor correction to the exponent is between fixed multiples of $x^{\beta_+-2}(y-x)^2$, hence is $O(u^2/x^{\beta_+})$. The rest of the integral under the exponential upper envelope is exponentially small in x^{β_+}. Integrating proves the second formula, while integrating the potential derivative proves the first.

For any fixed $B > 0$, on probabilities $p \in [\varepsilon t^{-B}, \varepsilon t^B]$ the potential slope and the hazard $g(Q_g(p))/p$ are both $H(1+o(1))$, uniformly. In fact the derivative of $\log V'(Q_g(e^{-y}))$ with respect to y is $V''(Q_g)e^{-y}/(V'(Q_g)g(Q_g)) = O(1/y)$. Here $|y-t| \le B\log t$. Throughout this band $V'' \le CH^2/t$, and the band has length $O(L/H)$. In particular for fixed $0 < a < b < \infty$,

$$Q_g(ap) - Q_g(bp) = H^{-1}\log(b/a) + o(H^{-1}) \tag{62}$$

uniformly for p of order m_0, and uniformly for a, b in compact positive intervals. We use these estimates also on bands enlarged by a fixed power of t.

8.3 Two density anchors and the survival deficit

Write $D = S_f - S_g$ on the right, where $\int_{\mathbb{R}} |D| \le \varepsilon$. The following elementary anchoring argument also works for log-concave densities with a finite support endpoint. Take $p_A = \varepsilon t^B$, $p_T = \varepsilon t^{-B}$, $x_A = Q_g(p_A)$, $x_T = Q_g(p_T)$, for sufficiently large fixed B. In a fixed number of intervals of lengths and separations of order $1/H$ near x_A, the reference density is comparable to Hp_A. For any such interval, choose a point in each of its two outer quarters where $|D| \le C\varepsilon H$; these exist by averaging. The interval between the two points

has f-mass differing from its g-mass by $O(\varepsilon H)$, and its g-mass is comparable to p_A. Thus the ratio of the two masses is $1 + O(\varepsilon H/p_A)$. Choose two separated intervals just to the left of x_A, with another interval on each side to ensure by positivity of their f-masses that both inner intervals lie inside the support of f. The weighted mean value property gives points $x_1 < x_2 \le x_A$, separated by order $1/H$, with

$$\left|\log \frac{f(x_i)}{g(x_i)}\right| \le C\varepsilon H/p_A.$$

Where f is positive, the log ratio $\log(f/g)$ has distributional second derivative at most V''. Subtracting a quadratic of second derivative CH^2/t therefore gives a concave function on the whole relevant tail strip. Extrapolating the secant through the two anchors shows

$$f(x) \le (1 + \delta_t)g(x), \qquad x_A \le x \le x_T, \qquad \delta_t = O(L^2/t). \tag{63}$$

Indeed the error from the two log values is $O(L\varepsilon H/p_A)$ and the quadratic remainder is $O(L^2/t)$. The first is smaller for B large. The inequality is automatic if $f(x) = 0$. The secant of $\log f$ itself is at most $-(1-o(1))H$, from the corresponding reference slope and the small anchor errors. By concavity f decays at least at that rate after x_2 as long as it is positive. In particular $S_f(x_T) \le Cp_T$ by (63) and the reference hazard estimate. Integrating the density upper bound back from x_T yields

$$D(x) \le \delta_t S_g(x) + Cp_T \qquad (x_A \le x \le x_T). \tag{64}$$

At every subsequent point in the support the hazard of f is at least cH by the same exponential bound on its density.

Fix $0 < b_0 < 1$, and then $c_0 > 0$ with $b_0 + 2c_0 < 1$. If $p = S_g(x) \in [p_T, m_0 t^{b_0}]$, the preceding interval of length $c_0 L/H$ stays to the right of x_A for large B. Throughout that interval $S_g(y) \le m_0 t^{b_0+2c_0}$ by the reference hazard bound. For any y in the interval,

$$D(y) = D(x) + \int_y^x (f - g) \le D(x) + \delta_t S_g(y) = D(x) + o(m_0).$$

If $D(x) < 0$, this inequality and the W_1 budget force $-D(x) \le Cm_0$. Together with (64),

$$|D(x)| \le Cm_0 \qquad (p_T \le S_g(x) \le m_0 t^{b_0}). \tag{65}$$

For the exact coefficient restrict now to a fixed window $S_g(x)/m_0 \in [\delta_0, K_0]$ with $0 < \delta_0 < K_0 < \infty$, and let $0 < \eta < 1$ be fixed. The preceding interval can be enlarged to length $(1-\eta)L/H$, because its g-tail mass is at most $Cm_0 t^{1-\eta/2}$. Thus the positive excess accumulated on it is still $o(m_0)$. Write $\varepsilon_+ = \int_0^\infty |D(x)|\,dx$. The whole preceding interval is eventually in this half-line, so the same argument gives

$$-\frac{\varepsilon_+ H}{(1-\eta)L} - o(m_0) \le D(x) \le o(m_0) \tag{66}$$

uniformly on this fixed window. In particular the constant is determined by the length over which the deficit persists, rather than by the proportional width of a single tail cell.

8.4 Integration of the quantile discrepancy

Let Q_u always refer to an upper tail quantile in the current side's coordinates. Monotone coupling gives the r-cost as the integral of $|Q_f(p) - Q_g(p)|^r$ over $0 < p < 1$ [4]. Equation (66) and quantile bracketing on fixed scaled windows give

$$|Q_f(p) - Q_g(p)| \le H^{-1} \log\left(1 + \frac{\varepsilon_+/\varepsilon}{(1-\eta)(p/m_0)}\right) + o(H^{-1}).$$

To see this, at reference mass $p + cm_0$ inequality (66) gives $S_f \ge p$ if $c = \varepsilon_+/[(1-\eta)\varepsilon] + o(1)$, with the error taken on its upper side. At mass $p - o(m_0)$ it gives $S_f \le p$. The reference quantile bracket then follows by (62). The errors are uniform in $\varepsilon_+/\varepsilon \in [0, 1]$.

We check the complement of the fixed windows uniformly. For $Km_0 \le p \le m_0 t^{b_0}/2$, with K large, (65) brackets the quantile by reference masses $p \pm Cm_0$. The hazard bound gives $|Q_f(p) - Q_g(p)| \le Cm_0/(Hp)$. Its cost is bounded by $C(m_0/H^r)K^{1-r}$. At the other end use one anchor mass which is a sufficiently large fixed multiple of m_0. The quantile displacement there is $O(H^{-1})$, and both densities beyond it have hazard bounded below by cH. Consequently for $p \le \delta m_0$ the displacement is at most $CH^{-1}(1 + |\log(p/m_0)|)$. Its integrated r-cost is $o_\delta(m_0/H^r)$, where the coefficient tends to zero as $\delta \downarrow 0$.

Apply the same tail arguments on the left. It remains to control the central quantile range, bounded on each side by the corresponding mass $m_0 t^{b_0}/2$. Enlarge the central range by halving both these endpoint tail masses. Throughout the enlarged range the reference density at its quantile is bounded below by

$$g_{\min} \ge c\,\varepsilon H_*(\varepsilon)^2 t^{b_0}/L,$$

using the tail hazard bound and unimodality between the two boundary quantiles. Let $p_{\min}$ be the smaller of the two original boundary tail masses, of order $\varepsilon H_* t^{b_0}/L$. Monotonicity of Q_f, and the derivative bound $|Q_g'| \le 1/g_{\min}$ on the enlarged range, imply

$$\sup_{\text{central}} |Q_f - Q_g| \le C\left(\sqrt{\varepsilon/g_{\min}} + \varepsilon/p_{\min}\right).$$

Indeed a displacement of size u at a quantile persists with size at least $u/2$ on an appropriate adjacent probability interval of length $c\min(ug_{\min}, p_{\min})$. Its integrated absolute displacement is at most ε. The displayed bound is polynomially smaller in t than $H_*^{-1}L^{-1/(r-1)}$. Multiplying its $(r-1)$st power by the W_1 budget makes the central r-cost negligible.

We can now integrate the exact fixed-window bound, send the window endpoints to zero and infinity using the uniform complement estimates, and finally send $\eta \downarrow 0$. On the right the resulting bound, with additive error $o(\varepsilon/(LH_*^{r-1}))$, is

$$\frac{\varepsilon_+}{LH_+^{r-1}} \int_0^\infty \log^r(1 + 1/u)\,du.$$

For example at positive η the integral with factor $(\varepsilon_+/\varepsilon)/(1-\eta)$ inside its logarithm scales by that factor, and the complement errors are uniform in $\varepsilon_+/\varepsilon$. The corresponding left budget ε_- satisfies $\varepsilon_+ + \varepsilon_- \le \varepsilon$. Their sum is therefore at most the claimed upper equivalent, since H_* is the smaller of the two slopes. The substitution $v = \log(1 + 1/u)$ gives

$$\int_0^\infty \log^r(1 + 1/u)\,du = \int_0^\infty v^r \frac{e^v}{(e^v - 1)^2}\,dv = \Gamma(r+1)\zeta(r),$$

by the nonnegative expansion $e^v/(e^v-1)^2 = \sum_{k\ge1} ke^{-kv}$. The integral is finite exactly for $r > 1$.

8.5 The tangent and cutoff construction

Keep g unchanged up to a large right-tail location S. Write $H_S = V'(S)$, $p_S = S_g(S)$, and $t_S = \log(1/p_S)$. Replace the density to its right by the tangent exponential $E_S(x) = g(S)e^{-H_S(x-S)}$ until a cutoff T, then set it to zero. The cutoff is chosen so that the tail mass remains p_S. This produces a log-concave probability density f_S, since the potential continues linearly from its tangent and then becomes infinite.

The exponential tangent has excess mass $b_S = g(S)/H_S - p_S \asymp p_S/t_S$. The upper bound follows from the tail integral calculation above. The matching lower follows by taking scaled shifts $H_S(x - S) \in [1, 2]$ in the same calculation and using the positive lower curvature bound on that window. The removed tangent mass is exactly b_S, so

$$T - S = \frac{\log t_S + O(1)}{H_S}, \qquad g(T) = E_S(T)(1 + o(1)), \qquad S_g(T) \sim b_S.$$

For the density comparison the accumulated curvature error on that interval is $O((\log t_S)^2/t_S) = o(1)$; the last assertion then follows from the hazard estimate.

On $[S, T]$ the survival difference is $-\int_S^x (E_S - g)$, because the total tail masses agree at S. It reaches $-(1 - o(1))b_S$ after a distance $o(\log t_S)/H_S$. Indeed at scaled distance $K_S \to \infty$ with $K_S = o(\log t_S)$, the

remaining integral of $E_S - g$ is $o(b_S)$. This follows by the bound $C(g(S)/H_S)t_S^{-1}\int_{K_S}^{\infty} u^2 e^{-u}\,du$ on local shifts and the exponentially small tail of the tangent beyond $x - S = S/2$. The contribution beyond T to the W_1 integral is $O(b_S/H_S)$. We obtain

$$W_1(f_S, g) \sim b_S \log t_S / H_S. \tag{67}$$

At any fixed $u > 0$, the f_S-quantile of tail mass ub_S is exactly $T - H_S^{-1}\log(1+u)$ for sufficiently large S, by integration of the truncated exponential. The reference quantile there is $T - H_S^{-1}\log u + o(H_S^{-1})$. The error is uniform on compact positive u-intervals. Integrating first on such intervals and then letting the intervals exhaust $0 < u < \infty$ proves

$$\liminf_{S\to\infty} \frac{H_S^r W_r(f_S, g)^r}{b_S} \geq \Gamma(r+1)\zeta(r).$$

The function $S \mapsto W_1(f_S, g)$ is positive and continuous for large S, and tends to zero, so it attains every sufficiently small ε. All choices attaining ε in this range tend to infinity as $\varepsilon \downarrow 0$. Equation (67) gives $\log(1/\varepsilon) = t_S + O(\log t_S)$, and hence $H_S/H_+(\varepsilon) \to 1$ by the slope derivative bound. Reflection gives the construction on the other side, and we choose the smaller slope. This proves Theorem 8.1. A finite support endpoint also shows that the analogous W_∞ modulus is infinite.

8.6 Gaussian smoothing

We first work in the window $s \leq \varepsilon^{\kappa_0}$ for a fixed positive κ_0. We recover the tail and central anchors directly from the observed budget. Put $D_s = \gamma_s * D$, $b_s = \sqrt{s}\,t$, and $\delta_s = 2e^{-t^2/2}$. An independent centered Gaussian shift of variance s exceeds b_s in magnitude with probability at most δ_s. On any cell of length w where g is comparable to some value $g_0 > 0$, choose endpoints $l < r$ in its two inset outer quarters with $|D_s(l)|, |D_s(r)| \leq C\varepsilon/w$. Then the observed mass in that interval differs from its reference observed mass by $O(\varepsilon/w)$. The f-mass on $[l + b_s, r - b_s]$ is at most the observed mass on $[l, r]$ plus δ_s, and the f-mass on $[l - b_s, r + b_s]$ is at least that observed mass minus δ_s. The analogous reference inequalities and local comparability of g show that the resulting upper and lower relative mass errors are at most

$$C\left(\frac{\varepsilon}{g_0 w^2} + \frac{b_s}{w} + \frac{\delta_s}{g_0 w}\right). \tag{68}$$

The endpoints and their shifts remain in the cell for the polynomial s range. Choosing bracketing cells with positive f-mass places the desired cells inside its support. In each desired cell, one point has ratio f/g at most one plus this error and one at least one minus it. Continuity gives the requisite anchor point. At $s = 0$ the same proof uses zero buffer and error probability.

In the tail anchor cells of Section 3, $w = H^{-1}$ and $g_0 \asymp Hp_A$. The last two terms of (68) are negligible compared with any fixed negative power of t, since s is bounded by a positive power of ε. Thus the same extrapolation gives (63), (64), and the downstream hazard bounds. The density ratio estimate also holds up to a distance $1/H$ past the displayed strip endpoints, by moving the initial anchors slightly to the left and extending the secant estimate.

On the intervals preceding the evaluation point in the medium and fixed-window persistence estimates, $\sup_x |D_s(x) - D(x)| = o(m_0)$. For $|z| \leq 1/H$ the two densities at $x - z$ are bounded by $CHS_g(x)$ by the ratio and local reference bounds. The contribution of such shifts to the difference is at most $C\sqrt{s}\,HS_g(x)$. The remaining Gaussian probability is $O(e^{-c/(sH^2)})$, and $|D| \leq 1$. The total is $o(m_0)$ uniformly on the intervals, since their survival masses are at most ε times a power of t. We can therefore use the observed integral $\int |D_s|$ for (65), and its right and left portions for the sharp persistence bound (66). The tail quantile integration and its complement estimates are unchanged.

For the central range take $b_0 = 1/4$ as in the direct proof. Extend the interval twice, first halving its endpoint tail masses and then decreasing them to one eighth of their initial values. The spatial collars have widths comparable to the reciprocals of the respective tail slopes. At each point of this enlarged interval choose the local scale $w \asymp (1 + |V'(x)|)^{-1}$. The right and left endpoints have densities of order $\varepsilon H_\pm^2 t^{b_0}/L$, so $g(x)w^2 \gtrsim \varepsilon t^{b_0}/L$ there and throughout the central interval. This follows by monotonicity outward of g

and of $|V'|$ in each tail, with a positive fixed floor in the center. Thus all terms in (68) are $o(1)$ uniformly. Taking two anchors on each side of a point, log concavity and its secants give

$$c\,g(x) \le f(x) \le Cg(x), \qquad |(\log f)'(x)| \le C/w \le CH^{\max}, \quad H^{\max} = \max(H_+, H_-)$$

where the slope exists. The lower density bound follows from the chord through two surrounding anchors. These conclusions hold on an interval slightly larger than the positions of all the latent central quantiles: the endpoint brackets from (65) put the quantiles between reference positions of masses differing by at most Cm_0, which is $o(m_0 t^{b_0})$.

Write $f_s = f * \gamma_s$ and $g_s = g * \gamma_s$. On the smaller collar, $g_s \ge cg$ since the convolution samples comparable reference density on $|z| \le b_s$. Use the larger collar for the anchors and the smaller collar for the estimates below.

All smoothed central quantiles stay inside this collar. For either density, expand its CDF to first order in z and integrate against the centered Gaussian on $|z| \le b_s$. The linear term integrates to zero. Local slope bounds give a remainder at most $C|z|^2 H^{\max} f(x)$ for f, and the analogous bound holds for g. Thus

$$|F_{f_s}(x) - F_f(x)| \le CsH^{\max} f(x) + \delta_s$$

and likewise for g. With $g_{\min}$ the central reference density floor from Section 4, shifts of size $C(sH^{\max} + \delta_s/g_{\min})$ bracket the inverse CDF. Across such shifts the densities are comparable to their initial values, so the CDF increment exceeds the smoothing error. These shifts are negligible compared with the collars. In particular the estimate of Section 4 with the same enlarged probability range, density floor $cg_{\min}$, and observed budget ε gives

$$\int_{\text{central}} |Q_{f_s}(p) - Q_{g_s}(p)|^r \, dp = o\left(\frac{\varepsilon}{LH_*^{r-1}}\right). \tag{69}$$

8.6.1 Fourier comparison in the central range

Choose $2/(3r) < \kappa_1 < 1$. In the window $s \le \varepsilon^{\kappa_0}$, take

$$h^2 = \max\{\varepsilon^{\kappa_1}, s/(\gamma t)\}, \qquad a = \varepsilon^{1/2},$$

with a sufficiently small fixed $\gamma > 0$. For a local comparison of the CDFs we use a positive-kernel curvature argument, given in full below; a fixed-noise version appears in [1]. Here and below Ct^C denotes a fixed polynomial bound. If K is Schwartz with smooth compactly supported Fourier transform, and $K_h(z) = h^{-1}K(z/h)$, then

$$\|\partial^j (D * K_h)\|_1 + h\|\partial^j (D * K_h)\|_\infty \le C_{K,j}\varepsilon^{1-c}h^{-j}. \tag{70}$$

Indeed D_s has L^1 norm at most ε. The rescaled multiplier for recovering $\partial^j(D * K_h)$ from D_s is $(i\xi)^j e^{s\xi^2/(2h^2)}\widehat{K}(\xi)$, with the factor h^{-j}. The L^∞ bound on its rescaled inverse Fourier transform follows by integrating over the fixed frequency support. For its L^1 bound use two integrations by parts away from a unit interval in position space. Both norms are bounded by a polynomial in s/h^2 times $C_{K,j}\exp(C_K s/h^2)$. For the two fixed kernels and derivative orders used below, γ can therefore be chosen so that $c > 0$ is as small as needed. We take it to satisfy

$$1 - c - \tfrac{1}{2} - \tfrac{\kappa_1}{2} > 0, \qquad 1 - c + (r-1)(1 - c - \tfrac{1}{2} - \tfrac{\kappa_1}{2}) > 1. \tag{71}$$

At a central point x, choose a spatial radius w of order $(1 + |V'(x)|)^{-1}$, with a sufficiently small constant so that $[x - 2w, x + 2w]$ lies in the larger central interval. The ratios $g(y)/g(x)$ there are bounded above and below. The same statements hold at points on a slightly widened central quantile interval. Locally $D^{(3)} = -(f - g)''$ is a signed measure in the distributional sense, because $D = S_f - S_g$. Its negative part is bounded by $C(1+|x|)^C g(x)\,dy$, using $(f'')_+ \le f[(\log f)']^2\,dy$ and the bound on g''. The first inequality follows by differentiating $f' = f(\log f)'$ in distributions on the positive interval, since the log slope is decreasing. The local bound $|(\log f)'| \le C(1 + |V'|)$ and the power-tail hypotheses justify this polynomial dependence on x.

Take a nonnegative Schwartz kernel P with compact smooth Fourier support and $P \ge c > 0$ on $[-1, 1]$; a rescaled square of an inverse Fourier bump suffices. At any $y \in [x - w, x + w]$, localize $D^{(3)} * P_h(y)$ to the

interval $[y-w/2, y+w/2]$ by a smooth nonnegative cutoff equal to one on $[y-w/4, y+w/4]$. The omitted integral is $O_M(h^M)$ for every fixed M. To see this for a distribution, integrate by parts three times against D, using $|D| \le 1$, Schwartz decay off a distance $w/4$, and polynomial bounds for the cutoff derivatives. Set $E = D^{(3)} * P_h$ and fix a sufficiently large M. The bound on the local negative part and the positivity of P_h give

$$|D^{(3)}|([y-h, y+h]) \le Ch\big((1+|x|)^C g(x) + |E(y)| + h^M\big).$$

Here the cutoff is one throughout $[y-h, y+h]$ and $P_h(y-\cdot) \ge c/h$ there. The h^M term accounts for localization and h lies between fixed powers of ε.

Choose now a Schwartz kernel K with $\widehat{K} = 1$ near zero and smooth compact Fourier support. Its mass is one and its first and second moments vanish. At a point of differentiability of D', Taylor expansion through order two with measure remainder gives error at most

$$C|z|^2 |D^{(3)}|([x-|z|, x+|z|])$$

when expanding $D(x-z)$ about x for $|z| \le w/2$. Cover the segment by intervals of radius h centered at $x+kh$, where $|k| \le 2+|z|/h$, and use the last cell estimate. After integrating against $|K_h(z)|$, the coefficient of $|E(x+kh)|$ is at most

$$Ch^3 \int_{|u| \ge (|k|-2)_+} u^2 |K(u)|\, du \le C_N h^3 (1+|k|)^{-N}.$$

The $g(x)$ term sums to $Ch^3(1+|x|)^C g(x)$. The contribution from $|z| > w/2$ is $O_M(h^M)$ by Schwartz decay, since D and the local Taylor coefficients are polynomially bounded. Increasing the orders of decay absorbs all errors of this form into the same $O_M(h^M)$. Using the vanishing moments and then continuity gives

$$|D(x)| \le Ch^3(1+|x|)^C g(x) + R_h(x) + C_M h^M, \tag{72}$$

where for any fixed $N > 1$ we can take

$$R_h(x) = |(D * K_h)(x)| + C_N h^3 \sum_{k \in \mathbb{Z}} (1+|k|)^{-N} |E(x+kh)|.$$

The sum has the same L^1 and L^∞ bounds as $|E|$, up to a fixed factor. Applying (70) at $j = 0$ for K and $j = 3$ for P gives

$$\|R_h\|_1 + h\|R_h\|_\infty \le C\varepsilon^{1-c}. \tag{73}$$

On the part of the central interval with $x = Q_g(p)$ and $g(x) \ge a$, the normalized bound $|D(x)|/g(x)$ is smaller than any reciprocal power of t, by (71) and the choice of h. Since f is comparable to g on the interval of radius w, bracketing its inverse CDF gives $|Q_f(p) - Q_g(p)| \le C|D(x)|/g(x)$. The leading term in (72) then contributes $O(h^{3r})$ by the finite polynomial moments of g. For R_h, change variables $dp = g(x)\, dx$; its contribution is at most

$$C\|R_h\|_1 (\|R_h\|_\infty / a)^{r-1} \le C\varepsilon^{1-c+(r-1)(1-c-1/2-\kappa_1/2)} = o(A_\varepsilon).$$

The h^M term is negligible by choosing M large. Consequently this part of the r-cost is $o(A_\varepsilon) + O((s/t)^{3r/2})$.

The rest of the central probability interval, where $g(Q_g(p)) < a$, has mass $O(a)$ by the eventual positive lower bound on each reference tail slope. Use the observed displacement from (69) and the two individual quantile smoothing shifts there. The extra cost is at most $Ct^C a s^r + o(A_\varepsilon)$. For $s \le s_c$, the first term is $o(A_\varepsilon)$, since $s_c^r = \varepsilon^{2/3}$ times a factor bounded by powers of t and $a = \varepsilon^{1/2}$. For $s > s_c$, the ratio $Ct^C a s^r/(s/t)^{3r/2}$ is at most $Ct^C a s_c^{-r/2} = o(1)$.

The tail contributions are charged to the observed budgets on the two sides. Together with the central estimate they give the stated additive upper bound in the polynomial window.

8.6.2 The complementary bounded-variance range

Fix $B > 3r$ and choose the exponent $\kappa_0 > 0$ small enough that $\kappa_0 < 2/(3r)$ and $\kappa_0(B+4)/2 < 1-c$, with the same fixed c as in (70). For $\varepsilon^{\kappa_0} \le s \le \Lambda$, take $h^2 = s/(\gamma t)$. Use the positive kernel P_h from Section 6.1 normalized to have mass one. The deconvolution bound at $j = 0$ gives $\|D * P_h\|_1 \le C\varepsilon^{1-c}$. The function $D * P_h$ is the survival discrepancy for $f * P_h, g * P_h$.

Take the central interval between the two reference quantiles of tail mass h^B and extend it to endpoint masses $h^B/8$. Throughout this interval $g \gtrsim h^B$; at any point x the anchor scale $w \asymp (1 + |V'(x)|)^{-1}$ is bounded below by a reciprocal power of $\log(1/h)$. Use an inner collar at endpoint masses $h^B/2$. Apply the observed mass anchor argument of (68) with P_h in place of Gaussian smoothing. Use buffer $b = h^{1/2}$ and error $O_N(h^N)$ for the probability of exceeding it, available for any fixed N by Schwartz decay. The first term of the anchor error is now $C\varepsilon^{1-c}/(g(x)w^2)$. It tends to zero uniformly because $\varepsilon^{1-c} = o(h^{B+4})$ under the choice of κ_0. The other two terms are also $o(1)$ for large N. With four anchors and log-concave secants we recover the same local comparability $f \asymp g$ and log-slope bound C/w on a collar of the central interval.

The Fourier estimate (72) now applies on this interval, with the same polynomial factor in x. The localization and Taylor arguments only need the local density and slope bounds just established and w/h tending to infinity at least as a power of $1/h$. By (73), $\|R_h\|_\infty \le C\varepsilon^{1-c}/h = o(h^{B+3})$. The $O_M(h^M)$ term is absorbed at $g \gtrsim h^B$. Quantile bracketing and integration against the moments of g as before therefore give r-cost $O(h^{3r})$ from the central probability interval.

At either endpoint, the quantile bracket puts the f-quantile of mass $p_0 = h^B$ within $Ch^3(1 + |x|)^C = o(1/H)$ of the reference position x, where H is the outward reference potential slope at x. Using the two upstream density anchors at separations of order $1/H$ gives both tails hazard at least cH downstream of these quantile positions, including for an alternative with a cutoff. Indeed the log-ratio errors are $o(1)$, while the reference log slope is $-H(1 + o(1))$ in outward coordinates. The resulting chord bound on the log density extends outward by concavity. Thus, on each side, the displacement for $p < p_0$ is bounded by $CH^{-1}(1+\log(p_0/p))$. Its r-cost is $O(p_0/H^r) = o(h^{3r})$. This proves $O((s/t)^{3r/2})$ in the complementary range and completes the upper bound.

8.6.3 Matching lower bounds

The tangent-cutoff alternatives satisfy a stochastic order relative to the reference, since their survival difference has one sign. This order persists under a common convolution, and the mean difference is unchanged by centered Gaussian noise. In one dimension W_1 for stochastically ordered integrable laws equals the absolute mean difference. Thus the lower construction has exactly the same observed W_1 at every variance. For a standard Gaussian g, combine the tangent cutoff with a Gaussian oscillation, as in Section 6.3 (compare [1]), here requiring only mass normalization. Let

$$f_T(x) = g(x)\big[1 + cT^{-2}(\cos(Tx) - e^{-T^2/2})\big].$$

For a fixed sufficiently small $c > 0$ and $T \ge 1$ the bracket is positive and the absolute second derivative of its logarithm is at most $Cc < 1/2$. Thus $V_T = -\log f_T$ has curvature between $1/2$ and $3/2$, with $|V_T'(x) - x| \le C/T$. The conditional Gaussian law of X given $X + Z = y$ has mean $y/(1 + s)$ and variance $s/(1 + s)$. Consequently

$$(f_T)_s(y) - g_s(y) = cT^{-2}g_s(y)\left[e^{-sT^2/(2(1+s))}\cos\left(\frac{Ty}{1+s}\right) - e^{-T^2/2}\right].$$

In the W_1 dual formula, subtract the value of each test at zero and bound it by $|y|$. This gives $W_1((f_T)_s, g_s) \le C_\Lambda T^{-2}e^{-sT^2/(2(1+\Lambda))}$ for $0 < s \le \Lambda$. Choose $T^2 = C_0 t/s$ with a fixed large C_0, so that this observed discrepancy is at most ε^2.

Apply Section 5 to V_T on the right tail, obtaining a tangent-cutoff density $\widetilde{f}_T$ with $W_1(\widetilde{f}_T, f_T) = \varepsilon - \varepsilon^2$. The comparisons in that construction are uniform in T: the curvature lies between two fixed positive constants, the density ratio to g is $1 + O(T^{-2})$, and the difference of the slopes is $O(T^{-1})$. In particular the cutoff is constructed at an initial position tending to infinity. The budget is unchanged by convolution by the stochastic order just proved, so the observed discrepancy of $\widetilde{f}_T$ from g is at most ε by the triangle inequality.

For every fixed compact window $p/b_S \in [\delta, K]$, the reference f_T and g quantiles there differ by $O(T^{-2}/H_*)$. Indeed their tail probabilities have ratio $1 + O(T^{-2})$ and the Gaussian hazard is comparable to H_* on that window, since b_S differs from ε by at most polynomial factors in t. The tangent-cutoff profile relative to f_T therefore has the same limit relative to g, with the same local units. Integrating the profile on such windows and letting $\delta \downarrow 0, K \uparrow \infty$ gives a tail cost of at least $(\Gamma(r+1)\zeta(r) - o(1))A_\varepsilon$.

For a disjoint lower bound, take a fixed smooth nonnegative nonzero bump ϕ of compact support and the test $\phi(x)\cos(Tx)/T$, with uniformly bounded Lipschitz norm. The difference of its expectations under f_T and g is bounded below by a positive constant times T^{-3}. This follows from $\int \phi g \cos^2(Tx)\,dx \to \frac{1}{2}\int \phi g\,dx > 0$ and $e^{-T^2/2} \to 0$. The test is zero outside its support. In the monotone coupling the expectation difference comes entirely from probabilities in a fixed central interval, since f_T/g tends uniformly to one. The tangent cutoff preserves f_T-quantiles on that interval. The Lipschitz inequality and Hölder's inequality then give r-cost at least $c_r T^{-3r} = c_r'(s/t)^{3r/2}$ on this central interval. It is disjoint from all the fixed scaled tail windows for small ε, which proves the additive lower bound. At $s = 0$ just use the original tangent-cutoff construction. This proves the Gaussian lower bound in Theorem 8.2.

8.6.4 Localized oscillations for a general reference

For the general reference choose a compact interval J on which V'' has a positive lower bound, available in either tail. Let χ be a smooth cutoff supported in the interior of J, with $0 \le \chi \le 1$ and $\chi = 1$ on a subinterval. Choose a smaller interval J_0 inside the latter, with positive separation from its boundary. Let ψ be smooth, nonnegative and of mass one, supported in the interior of J.

For $s > 0$ put $v = \min(s, b/t)$. Place centers z_i in J_0 at spacing $\sqrt{v}$, with their number of order $v^{-1/2}$. Set

$$U_v(x) = \sum_i e^{-(x-z_i)^2/(2v)}, \quad T^2 = C_0 t/v, \qquad m_T = \int \chi(x) U_v(x)\cos(Tx)\,dx.$$

Define the mass-preserving perturbation

$$q_T = cT^{-2}\big(\chi U_v \cos(T\,\cdot) - m_T\psi\big).$$

For $j = 0, 1, 2$ the derivatives of U_v have sup norm at most $C_j v^{-j/2}$, by summing the separated Gaussian tails. Thus q_T and its derivatives of order at most two are bounded by Cc. Since g has a positive floor on J, choosing $c > 0$ small keeps $g + q_T$ positive and the curvature of $-\log(g + q_T)$ nonnegative. Outside J this potential equals V.

The integral of each uncut envelope against $\cos(Tx)$ has magnitude at most $\sqrt{2\pi v}\, e^{-vT^2/2}$. The omitted part $(1-\chi)U_v$ has weighted L^1 norm with weight $1 + |x|$ at most $Ce^{-c_0/v}$, since the centers have fixed positive separation from the omitted region. These estimates give $|m_T| \le C(e^{-vT^2/2} + e^{-c_0/v})$. Convolving each uncut envelope times $\cos(Tx)$ with γ_s produces an envelope of variance $v + s$ and the attenuation factor $e^{-vsT^2/(2(v+s))}$. This follows from the same conditional Gaussian calculation as above, with the individual means z_i and variance v. The new envelopes have total mass $O(1)$ and uniformly bounded weighted L^1 norms for $s \le \Lambda$. Convolution of the omitted part and normalization term costs at most a constant times their weighted L^1 norms, by the first absolute moment of γ_s. The dual W_1 bound for the observed perturbation is therefore

$$C_\Lambda T^{-2}\big(e^{-vT^2/4} + e^{-c_0/v}\big).$$

Taking C_0 large and b small makes this at most ε^2.

A fixed compactly supported smooth cutoff equal to one on J, multiplied by $\cos(Tx)/T$, is uniformly Lipschitz. Its expectation under $g + q_T$ differs from that under g by at least a constant times T^{-3}. Indeed each uncut envelope has cosine-square integral $\sqrt{2\pi v}(1 + e^{-2vT^2}\cos(2Tz_i))/2$. Summing gives a positive lower bound, while the omitted parts and m_T tend to zero. The expectation difference is carried by a fixed central probability interval containing the support of the test in the monotone coupling, since $g + q_T \to g$ uniformly in distribution. This gives central r-cost $\gtrsim T^{-3r}$ by the Lipschitz and Hölder inequalities.

The perturbation has zero mass and is supported in J. Choose the tangent cutoff on the less steep side of the reference, outside J, with budget $\varepsilon - \varepsilon^2$. On that tail the density and probabilities of $g + q_T$ agree exactly with those of g before the cutoff is applied. The cutoff preserves the central quantiles and has the

same convolved W_1 cost by stochastic order. The combined alternative is log-concave and lies within the observed budget. Its tail and central costs add as in the Gaussian construction. This proves the general lower bound. Since $s_c = o(t^{-1})$, it also proves the necessity of $s = o(s_c)$ for the direct equivalent.

8.7 Radial laws in higher dimensions

Theorem 8.3. *Let $g(x)$ on $\mathbb{R}^d$ be proportional to $e^{-V(|x|)}$, with V finite, convex and nondecreasing on $[0,\infty)$, and smooth with the same power-tail assumptions at infinity. For the class of radial log-concave alternatives f, the equivalent in Theorem 8.1 holds with H_* replaced by the right-tail slope H_R of the reference law of the radius. In particular H_R is the slope of $U(r) = V(r) - (d-1)\log r$ at its upper tail quantile of mass ε.*

Proof. For radial laws the Euclidean W_q distance equals the W_q distance of the radius laws, for every finite $q \geq 1$. The lower inequality is the contraction under the radius map. For the upper inequality couple the radii monotonically and take a common independent uniform direction. The reference radius density is proportional to $r^{d-1}e^{-V(r)}$. The alternative radius density has the power factor r^{d-1} times the radial profile of f. Both radius densities are log-concave on $r > 0$. The effective reference potential U satisfies the same power-tail assumptions. Thus the right-tail upper proof applies unchanged. Its whole budget is at most ε.

There is no second heavy tail. At the left endpoint, the reference radius quantile in lower probability coordinates has order $p^{1/d}$ as $p \downarrow 0$, since V is finite and continuous there. Choose $d/(d+r) < a < 1$, and put $p_0 = \varepsilon^a$. For $p \leq p_0$, monotonicity and the W_1 budget give $Q_f^{\mathrm{low}}(p) \leq Q_g^{\mathrm{low}}(2p_0) + \varepsilon/p_0$: if the inequality failed, integration over $[p_0, 2p_0]$ would exceed the budget. Therefore the cost from this range is $O(p_0^{1+r/d} + \varepsilon^r p_0^{1-r})$, which is negligible at the stated rate. On the remaining central range the lower endpoint density is bounded below by a constant times $p_0^{(d-1)/d}$, larger than the right-endpoint density at mass $m_0 t^{b_0}/2$. After extending the central range by halving both boundary tail masses, the reference density floor is still paid by the right endpoint, since the radius density is unimodal. The central quantile argument of Section 4 applies, with H_R in place of H_*.

For the lower bound keep the ambient potential V up to radius S, continue it with its tangent, and truncate at T to preserve the radius tail mass. This continuation is convex and nondecreasing in radius, so its radial law is log-concave in $\mathbb{R}^d$. The radius tangent density is

$$g_R(S)(x/S)^{d-1}e^{-V'(S)(x-S)}, \qquad x \geq S.$$

On scaled shifts $V'(S)(x-S)$ of bounded size, the extra factor $(x/S)^{d-1}$ tends uniformly to one. The excess tangent mass is of order p_S/t_S using the original V'' on shifts of this size and the same exponential estimates as in Section 5. Indeed that excess is the tangent integral weighted by $1 - e^{-[V(x)-V(S)-V'(S)(x-S)]}$, so the factor $(x/S)^{d-1}$ only changes the comparison constants. The cutoff satisfies $V'(S)(T-S) = \log t_S + O(1)$, and the W_1 deficit has the same plateau and equivalent as in (67), since the remaining excess past a diverging scaled distance is again $o(b_S)$. At bounded scaled distances from T, the radius tangent density has the same limiting exponential profile and its excess removal mass b_S is asymptotic to the reference survival at T. Thus the same limiting quantile displacement $\log(1 + 1/u)$ in units $1/V'(S)$ gives the constant. Finally $V'(S)/H_R(\varepsilon) \to 1$ since $U'(x) = V'(x) - (d-1)/x$ and the reference tail masses differ only by polynomial factors in t_S. The cutoff radius depends continuously on S by strict monotonicity of the tangent tail integral in the cutoff. The resulting W_1 discrepancy is continuous, positive, and tends to zero. The construction therefore realizes every sufficiently small prescribed ε. This completes the radial proof. □

9 Sharp Gaussian deconvolution for superquadratic potentials

We determine the inverse scale for Gaussian smoothing near a one-dimensional log-concave reference with superquadratic potential tails of a common power $\beta > 2$. For fixed positive variance and observed W_1 tolerance ε, put $t = \log(1/\varepsilon)$. The sharp latent W_r modulus has order $(\log t/t)^{3(1-1/\beta)}$ for every fixed $r > 1$ in the stated reference class. Clipped Hermite tests recover moments beyond the resolution of the direct Gaussian Fourier bound. A polynomial construction on the two tails cancels these moments while keeping the perturbation log-concave, even when the reference potential is flat in its interior. We also obtain the variance-dependent inverse scale and add its lower bound to the exact direct tail cost.

9.1 Statement

Let $g = e^{-V}$ be a positive probability density on $\mathbb{R}$, where $V \in C^2$ is convex. Write $V_+(x) = V(x)$ and $V_-(x) = V(-x)$ for $x > 0$. Assume that for a common $\beta > 2$, on both sides as $x \to \infty$,

$$V'_\pm(x) \asymp x^{\beta-1}, \qquad V''_\pm(x) \asymp x^{\beta-2}.$$

All comparison constants are fixed. Set

$$t = \log(1/\varepsilon), \qquad q = 1 - \frac{1}{\beta}, \qquad \alpha = 1 - \frac{2}{\beta}, \qquad \ell_s = \log(e + st^\alpha).$$

For fixed $r > 1$ and $\Lambda < \infty$, consider the modulus

$$\mathcal{M}_r(\varepsilon, s) = \sup_{\substack{f \text{ log-concave} \\ W_1(f*\gamma_s, g*\gamma_s) \le \varepsilon}} W_r(f, g)^r, \qquad \gamma_s = N(0, s), \quad 0 \le s \le \Lambda.$$

Here f is a probability law and we identify it with its density when it has one. Put $H_*(\varepsilon)$ equal to the smaller outward reference potential slope at the two tail quantiles of mass ε, and define

$$A_\varepsilon = \frac{\varepsilon}{H_*(\varepsilon)^{r-1} \log t}, \qquad C_r = \Gamma(r+1)\zeta(r), \qquad B_{\varepsilon,s} = \min\Big\{(s/t)^{3r/2}, (\ell_s/t)^{3rq}\Big\}.$$

Theorem 9.1. *Uniformly for $0 \le s \le \Lambda$,*

$$\mathcal{M}_r(\varepsilon, s) \le (C_r + o(1))A_\varepsilon + CB_{\varepsilon,s}.$$

The matching lower bound is

$$\mathcal{M}_r(\varepsilon, s) \ge (C_r - o(1))A_\varepsilon + cB_{\varepsilon,s}$$

in the same range. In particular at any fixed $s > 0$, the modulus for W_r has order $(\log t/t)^{3q}$.

The hypotheses include $g(x) \propto e^{-|x|^\beta}$ and $g(x) \propto e^{-x^2-|x|^\beta}$. For st^α bounded, the inverse term has order $(s/t)^{3r/2}$. When $st^\alpha \to \infty$, the moment scale $(\ell_s/t)^{3rq}$ governs. The direct term and the first inverse upper term follow from the companion tail theorem Section 8. The new input is the moment upper bound and a single construction giving the entire inverse lower term.

General fixed-noise Wasserstein deconvolution over moment classes has logarithmic rates [6]. Rigollet and Weed [29, Section 2.3 and Lemma 4.1] use recovered moments and compact moment matching for bounded latent support. The Gaussian inverse modulus in [1] uses a two-sided Hessian bound on the reference. The present model combines unbounded super-Gaussian reference tails with log-concavity of the alternative. The local shape constraint gives a cubic gain at the recovered spatial resolution.

9.2 Moment recovery from observed transport

We give the argument uniformly for $st^\alpha \ge 1$. The observed reference law has tail bound

$$\Pr_{g*\gamma_s}(|Y| \ge y) \le C\exp(-cy^\beta) + C\exp(-cy^2/s).$$

This follows by splitting the event between X and the independent Gaussian. For a large fixed K take $L = K(t^{1/\beta} + \sqrt{st})$. Integration of the bound gives $\mathbb{E}_{g*\gamma_s}(|Y| - L)_+ \le Ce^{-2t}$. The observed W_1 constraint and the fact that $(|y| - L)_+$ is one-Lipschitz thus give $\Pr_{f*\gamma_s}(|Y| \ge 2L) \le C\varepsilon/L$.

Both observed laws are log-concave. Their first absolute moments are uniformly bounded by the observed W_1 constraint, the reference first moment and $s \le \Lambda$. Hence, for a fixed sufficiently large M, each outward survival function at the signed position corresponding to $-M$ in outward coordinates is at least $1/2$. A log-concave survival function has concave logarithm. Using its chord from $-M$ to $2L$ therefore gives, for either observed law,

$$\Pr(|Y| \ge 2L + u) \le Ce^{-t/2}e^{-ctu/L}, \qquad u \ge 0. \tag{74}$$

The constants in this estimate are independent of s.

Define the monic Hermite polynomial $P_{k,s}$ by the generating function

$$e^{zy-sz^2/2} = \sum_{k=0}^{\infty} P_{k,s}(y) z^k/k!.$$

It satisfies $P'_{k,s} = kP_{k-1,s}$ and, for a centered Gaussian of variance s independent of X, $\mathbb{E}[P_{k,s}(X+Z) \mid X] = X^k$. The representation $P_{k,s}(y) = \mathbb{E}(y + i\sqrt{s}G)^k$ with $G \sim N(0,1)$ yields

$$|P_{k,s}(y)| \le (|y| + C\sqrt{sk})^k.$$

All latent moments exist by log-concavity. Clip the argument of $P_{k,s}$ at $\pm 2L$. The clipped function has Lipschitz constant at most $k(CL)^{k-1}$ for $k \le t$. The contribution of the clipping error to either expectation is at most $Ce^{-t/2}(CL)^k$. For this last assertion integrate the derivative of $(CL+u)^k$ against the survival bound (74). Up to its factor $Ce^{-t/2}$, the estimate is the kth moment of CL plus an exponential variable of rate ct/L, bounded by $(C'L)^k$ for $k \le t$. Consequently for a fixed $d > 0$,

$$\left| \int x^k (f-g)(dx) \right| \le e^{-dt}(CL)^k, \qquad 1 \le k \le t, \tag{75}$$

after decreasing d to absorb fixed and polynomial factors.

Take an even integer n of order at/ℓ_s for a small fixed $a > 0$, with $n \le at/\ell_s$, and put $R_n = n^{1/\beta}$. For sufficiently small ε uniformly in s,

$$L/R_n \le C(\ell_s/a)^{1/\beta}(1 + \sqrt{st^\alpha}).$$

The logarithm of the right side is at most $C(1 + \log(1/a) + \ell_s)$. Taking a small, each moment in (75), divided by R_n^k, is bounded by $e^{-dt/2}$ for $k \le n$. In particular the order-n absolute moment of f is at most $(CR_n)^n$, because n is even and the reference power tails give this bound for g. The case $k = 2$ also gives bounded second moments uniformly in ε, s.

Let $D = S_f - S_g$ be the survival difference. It is integrable since both laws have finite first moment. For $\xi \ne 0$, $|\widehat{D}(\xi)| = |\widehat{f}(\xi) - \widehat{g}(\xi)|/|\xi|$. Taylor expand the Fourier exponential through degree $n-1$. The normalized moment difference bounds its polynomial part after division by $|\xi|$ by

$$R_n e^{-dt/2} e^{|\xi| R_n}.$$

The integral form of the Taylor remainder on the imaginary axis bounds the remaining quotient by $2(CR_n)^n |\xi|^{n-1}/n!$. For a sufficiently small fixed $\eta > 0$ we conclude that

$$\sup_{|\xi| \le \eta n/R_n} |\widehat{D}(\xi)| \le Ce^{-cn}, \tag{76}$$

including $\xi = 0$ by continuity. Here $n \le at$, and $n! \ge (n/e)^n$.

9.3 From a Fourier interval to cubic transport control

We include the local shape estimate used to convert (76) into transport. Its positive-kernel step is the one of Section 8, proved in full in the lemma below (compare [1]).

Lemma 9.2. *Assume f is log-concave and f, g have bounded second moments. Suppose, for $h \downarrow 0$, that $\sup_{|\xi| \le 1/h} |\widehat{D}(\xi)| = O_N(h^N)$ for every fixed N. Then $W_r(f,g)^r \le Ch^{3r}$.*

Proof. Choose a nonnegative Schwartz kernel P of mass one with smooth Fourier transform supported in $[-1,1]$ and $P \ge c > 0$ on $[-b,b]$ for a fixed $b > 0$. One can take a suitably normalized square of an inverse Fourier bump. Define $P_h(x) = h^{-1}P(x/h)$. For any fixed Schwartz kernel K with smooth Fourier support in $[-1,1]$, the hypothesis gives $\|(D * K_h)^{(j)}\|_\infty = O_N(h^N)$ for each fixed j, N. Also $\|D * P_h\|_1 = O_N(h^N)$.

Indeed $D * P_h$ is the survival difference of $f * P_h, g * P_h$. Integrate its sup bound on $[-h^{-N}, h^{-N}]$. The tail integral is $O(h^N)$ by the uniform second moments of these two convolved laws.

Fix $B > 3r$. Consider the interval between the two reference quantiles of tail mass h^B, with a collar extended to endpoint tail masses $h^B/8$. There $g \gtrsim h^B$; at each point x the local scale $w \asymp (1 + |V'(x)|)^{-1}$ is at least a reciprocal power of $\log(1/h)$. On local cells of this width the reference density is comparable to $g(x)$. For a cell, choose $l < r$ in inset outer quarters with $|(D * P_h)(l)| + |(D * P_h)(r)| = O_N(h^N/w)$ by averaging. The mass discrepancy of the two smoothed laws on $[l, r]$ has this bound. A shift drawn from P_h exceeds $b_h = \sqrt{h}$ in absolute value with probability $O_N(h^N)$. The latent f-mass on $[l + b_h, r - b_h]$ is bounded above by the smoothed mass plus that probability. The latent mass on $[l - b_h, r + b_h]$ is bounded below by the smoothed mass minus that probability. Comparing the reference masses on these intervals gives relative errors $O_N(h^N/(g(x)w^2)) + O(b_h/w) = o(1)$. Choose six bracketing cells, three on each side, with widths and spacings of order w. Positive mass in the outermost cells ensures positivity and continuity in the four inner cells. The mass bounds yield a ratio anchor $\log(f/g) = o(1)$ in each of those four cells, using one mass lower bound and one mass upper bound on the two inset intervals. Applying concavity chords and secants to the anchors gives $f \asymp g$ and $|(\log f)'| \le C/w$ on an inner collar with endpoint tail masses $h^B/2$. All cells fit by choosing the constant in w small.

On this collar $D^{(3)} = -(f - g)''$ is a signed measure. Its negative part on a local cell about x is at most $C(1+|x|)^C g(x)\, dy$. Indeed the positive part of f'' is at most $f[(\log f)']^2\, dy$, by differentiating in distributions and using the decreasing log slope. The reference g'' and the local log slopes have the same stated polynomial control. Localize $(D * P_h)^{(3)}$ with a nonnegative cutoff supported within the cell and equal to one on the half cell. The omitted term is $O_N(h^N)$ for every fixed N by integrating D against three derivatives of the remote kernel and cutoff. Here $|D| \le 1$, w is at least a reciprocal power of $\log(1/h)$, and P is Schwartz. Since the localized kernel is at least c/h on an interval of radius bh, the convolved bound and the bound on the negative part imply

$$|D^{(3)}|([y - bh, y + bh]) \le Ch(1 + |x|)^C g(x)$$

at any y in the inner part of the cell. The residual $O_N(h^N)$ is absorbed using $g(x) \gtrsim h^B$. Covering by these intervals bounds variation along a shift z of length at most a fixed small multiple of w by $C(|z| + h)(1 + |x|)^C g(x)$.

Take a signed Schwartz K with $\widehat{K} = 1$ near zero and smooth support in $[-1, 1]$, so its first two moments vanish and its mass is one. At a point of differentiability of D', expand D through degree two. The measure remainder is bounded by $C|z|^2|D^{(3)}|([x - |z|, x + |z|])$. Integrate it against $|K_h(z)|$. The local contribution is $O(h^3(1 + |x|)^C g(x))$; remote shifts give $O_N(h^N)$ by Schwartz decay, $|D| \le 1$, and the local bounds for the Taylor coefficients. Using the vanishing moments, the small low-frequency term $D * K_h$, and continuity gives

$$|D(x)| \le Ch^3(1 + |x|)^C g(x)$$

on the original interval. At $x = Q_g(p)$, comparability of f, g on a local cell brackets $Q_f(p)$ within $Ch^3(1+|x|)^C$ of x. Integrating over central probabilities gives $O(h^{3r})$ since g has finite polynomial moments.

At either boundary quantile of mass $p_0 = h^B$ the bracket has width $o(1/H)$, where H is the local outward reference slope. Two upstream mass anchors at spacing of order $1/H$ give an outward log-density secant at most $-cH$ for the alternative. The reference has this slope as well. Concavity extends the slope bound outward and implies survival hazards at least cH, also when f has a cutoff. Therefore the two quantiles differ by at most $CH^{-1}(1 + \log(p_0/p))$ at tail mass $p < p_0$. Their combined r-cost is $O(h^B/H^r) = o(h^{3r})$ with the appropriate slope on each side. □

With $h = R_n/(\eta n)$, (76) satisfies the lemma. This proves

$$\mathcal{M}_r(\varepsilon, s) \le C(\ell_s/t)^{3rq} \qquad (st^\alpha \ge 1).$$

Combine it with the companion bound $(C_r + o(1))A_\varepsilon + C(s/t)^{3r/2}$ from Section 8. For $st^\alpha < 1$, one has $(s/t)^{3r/2} \le t^{-3rq} \le (\ell_s/t)^{3rq}$. This proves the asserted uniform upper bound.

9.4 Moment cancellation supported away from the center

For $s \le C_*/t$ with any fixed $C_* > 0$, the general lower bound of Section 8 already pays the claimed cost. Indeed $\min(s, b/t) \ge cs$ on that range and the first term in the minimum defining $B_{\varepsilon,s}$ governs. We take C_* sufficiently large below and work in the remaining range.

Fix a constant $A > 0$ beyond which the two tails have strictly positive curvature. Fix a right interval J with closure in $(A+1, A+3)$ and a smaller interval J_0 with closure inside J. Take an integer $m \asymp K_0 t/\ell_s$ with K_0 a large fixed constant, and put

$$R = a_0 \min\{\sqrt{sm},\, m^{1/\beta}\}, \qquad h = R/m$$

with $a_0 > 0$ a small fixed constant, chosen after K_0 and before C_*. For $s \ge C_*/t$, taking C_* sufficiently large ensures $R \ge 2(A+4)$ for small ε. Indeed if $st^\alpha \le 1$ then $\ell_s \asymp 1$ and $sm \gtrsim C_*$; otherwise $sm \gtrsim t^{2/\beta}/\log t \to \infty$. Also $R^2/m \le a_0^2 m^{-\alpha} \to 0$ uniformly.

For $A < x < R$ let $0 < \theta(x) < \pi$ be defined by

$$x^2 = A^2 + (R^2 - A^2)\sin^2(\theta(x)/2).$$

Let θ_i be a grid of spacing $m^{-1/2}$ in $\theta(J)$, with endpoints omitted when necessary. For sufficiently large m, this gives distance $O(m^{-1/2})$ from every point of $\theta(J_0)$ to a grid point, since $|\theta(J)| \asymp R^{-1}$ and $R/\sqrt{m} \to 0$. Define the even trigonometric polynomial

$$U_m(\theta) = \sum_i \left[\cos^{2m}\left(\frac{\theta - \theta_i}{2}\right) + \cos^{2m}\left(\frac{\theta + \theta_i}{2}\right)\right].$$

Its degree is at most m. Fix an integer $L_0 \ge 4$ and take $N = 4m$. Define an even signed function u by setting, for $A < x < R$,

$$u(x) = c_0 h^2 x\,(R\sin\theta(x))^{2L_0-1} U_m(\theta(x))\cos(N\theta(x)), \tag{77}$$

and setting it equal to zero for $0 \le x \le A$ and $x \ge R$. Here $c_0 > 0$ is a small fixed constant.

All moments of order less than m vanish. Odd moments vanish by evenness. For the even moment x^{2j}, the Jacobian is $dx = (R^2 - A^2)\sin\theta\, d\theta/(4x)$. Thus the integral is proportional to

$$\int_0^\pi \big(A^2 + (R^2 - A^2)\sin^2(\theta/2)\big)^j \sin^{2L_0}\theta\, U_m(\theta)\cos(N\theta)\, d\theta.$$

The factors before the last cosine form an even trigonometric polynomial of degree at most $j + 2L_0 + m$. For $2j < m$ this degree is below N for large m. Orthogonality of the cosine modes on $[0,\pi]$ proves the cancellation, including $\int u = 0$.

We next check the curvature constraint. For $j = 0, 1, 2$, bounds on the derivatives of $\cos^{2m}(\varphi/2)$ on the circle give

$$\big|(\partial_\theta)^j (U_m(\theta)\cos(N\theta))\big| \le C_j m^j \exp\big(-cm\ \mathrm{dist}(\theta, \theta(J))^2\big) \quad (0 \le \theta \le \pi).$$

To see this, the individual shifted factors and their first two derivatives are bounded by $C_j m^j$ times $\exp(-cm d_{\mathrm{circ}}(\theta, \pm\theta_i)^2)$, where d_{circ} is circle distance. The two grids on the circle each have spacing at least a constant times $m^{-1/2}$. Sum the Gaussian envelopes, first retaining half of the distance exponent and then using the spacing for the other half. For $\theta, \theta_i \in [0,\pi]$ the reflected circle distance is at least $|\theta - \theta_i|$. Also $U_m \ge c > 0$ on $\theta(J_0)$ by the grid proximity.

Write $W(x) = R\sin\theta(x)$. Direct differentiation gives

$$W(x) \le C(1 + x), \qquad |\theta'(x)| = \frac{2x}{\sqrt{(x^2 - A^2)(R^2 - x^2)}} \ge c/R.$$

Using $\sqrt{(x^2 - A^2)(R^2 - x^2)} \asymp RW(x)$, the upper bounds are

$$|\theta'| \le \frac{Cx}{RW}, \qquad |\theta''| \le \frac{C}{RW} + \frac{Cx^2}{RW^3}.$$

The derivatives of W follow from $W = R\sin\theta$. In (77), the factor W^{2L_0-1} absorbs the negative powers of W from any two derivatives. Combining the elementary product bounds and $|\theta(x)-\theta(z)| \geq c|x-z|/R$, for $x, z \in (A, R)$, gives

$$|u^{(j)}(x)| \leq Cc_0 h^{2-j}(1+x)^C \exp\left(-c\frac{m}{R^2}\operatorname{dist}(x,J)^2\right), \qquad 0 \leq j \leq 2. \tag{78}$$

The first two derivatives and the value have zero traces at both endpoints: W is of order a square root of endpoint distance for each fixed R, and $2L_0-1 \geq 7$. Thus u is C^2 across the boundaries and has the reflected bound on the left.

At sufficiently large x on the support, $\operatorname{dist}(x, J) \geq cx$. Since $R^\beta/m \leq a_0^\beta$, the exponential in (78), for sufficiently small a_0, absorbs $e^{V(\pm x)}$ and the polynomial factors from its first two derivatives. More precisely the exponent $-cmx^2/R^2$ dominates Cx^β with enough slack to retain $-c'mx^2/R^2$ for $x \leq R$. On the compact part of this support, g and the curvature have positive minima on the two closed regions outside $(-A, A)$. The derivative bounds therefore imply both uniform $O(c_0)$ bounds and $O(c_0V'')$ bounds on u/g and its first two derivatives. These statements also hold on the tails by the curvature lower bound there and the retained exponential decay. In the central interval the perturbation is zero. Choosing c_0 small ensures positive $g+u$ and $|(\log(1+u/g))''| \leq V''/2$, using both bounds in the squared first derivative term. Hence $g+u$ is log-concave. The envelopes also give $\|u\|_1 \leq Ch^2$.

For the observed budget use the Gaussian derivative bound in the weighted norm $\|v\|_{1,1} = \int(1+|y|)|v(y)|\,dy$. For $|z| \leq R$, Hermite orthogonality and Cauchy–Schwarz give

$$\left\|\gamma_s^{(m)}(\,\cdot - z)\right\|_{1,1} \leq C(1+R+\sqrt{s})\, s^{-m/2}\sqrt{m!}.$$

Indeed the integrand becomes $s^{-m/2}(1+|z+\sqrt{s}G|)|\mathrm{He}_m(G)|$ under a standard Gaussian G. Taylor-expand $\gamma_s(y-x)$ in x around zero through order $m-1$. The moments vanish and the integral remainder is bounded by the preceding estimate times $R^m/m!$. Consequently,

$$W_1((g+u)*\gamma_s, g*\gamma_s) \leq C(1+R+\sqrt{s})\,\|u\|_1\left(\frac{CR}{\sqrt{sm}}\right)^m.$$

Here the signed measure has zero mass, so a W_1 dual test centered at zero is bounded by $|y|$. The parenthesized factor is $Ca_0\min\{1, (sm^\alpha)^{-1/2}\}$. For large st^α, $\log(sm^\alpha) \geq c\ell_s$ because $m \asymp K_0 t/\ell_s$. On the complementary bounded range $\ell_s \asymp 1$. Take K_0 large and a_0 small. The observed discrepancy is then at most ε^2, absorbing the polynomial prefactor in t.

To witness the latent cost take a smooth nonnegative nonzero ϕ supported in J_0 and the test $h\phi(x)\cos(N\theta(x))$, extended by zero. It is uniformly Lipschitz by the local bound $|\theta'| \leq C/R$ there. Its pairing with (77) is at least a positive multiple of h^3. Indeed U_m, W, and x are bounded below on J_0, and $\int\phi(x)\cos^2(N\theta(x))\,dx$ tends to $\frac{1}{2}\int\phi(x)\,dx$. For the latter claim integrate the double-frequency cosine by parts, using $\theta' \asymp R^{-1}$ and $|\theta''| \leq C/R$ on the fixed test support to obtain error $O(R/N)$. The expectation difference comes entirely from a fixed central probability interval in the monotone coupling, since $\|u\|_1 \to 0$. The Lipschitz inequality and Hölder's inequality give r-cost at least ch^{3r} there.

With $m \asymp t/\ell_s$ the chosen h is, up to constant factors, $\min\{\sqrt{s\ell_s/t}, (\ell_s/t)^q\}$. For bounded st^α one has $\ell_s \asymp 1$. For large st^α, the ratio $\sqrt{s/t}/(\ell_s/t)^q = \sqrt{st^\alpha}/\ell_s^q$ tends to infinity, so both minima choose the moment scale. Thus $h^{3r} \asymp B_{\varepsilon,s}$ uniformly.

Finally choose the tangent cutoff from Section 8 on the less steep reference side with discrepancy $\varepsilon - \varepsilon^2$. Its initial position has magnitude of order $t^{1/\beta}$. As $R \leq a_0 m^{1/\beta}$ with $m = O(t)$, making a_0 smaller puts both perturbation intervals strictly inside both possible cutoff positions. Because u has zero mass, the density and tail mass of $g+u$ at the cutoff agree with those of g. The cutoff preserves its fixed central quantiles. Stochastic order preserves the cutoff's W_1 cost under the common convolution, so the total observed budget is at most ε. The sharp tail displacement relative to g occurs at probabilities disjoint from the central interval, and gives $(C_r - o(1))A_\varepsilon$ by the tangent profile of Section 8. Adding the two costs proves the lower bound.

Statements and Declarations

Funding. This research was self-funded.

Competing interests. The author owns Luminary Oddities and has no other relevant competing interests.

AI assistance. OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of each part. Anthropic Claude edited, consolidated and cross-referenced the parts and checked computations. Anthropic Claude identified the answered question, edited this edition and checked the argument. Scott Schubr directed the research and is responsible for the content.

References

[1] S. Schubr, *Sharp Wasserstein stability for Gaussian deconvolution of log-concave laws*, edition 4. Luminary Research Archive, 2026. `https://research.luminaryoddities.org/a-sharp-logarithmic-inverse-bound-for-gaussian-estimation/v/4`; `https://doi.org/10.5281/zenodo.24472633`.

[2] S. Chewi, J. Niles-Weed and P. Rigollet, *Statistical Optimal Transport*, Lecture Notes in Mathematics, Springer, Cham, 2025. DOI. Section and page numbers refer to arXiv:2407.18163v2 (lecture notes of the École d'Été de Probabilités de Saint-Flour XLIX, 2019).

[3] 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.

[4] S. G. Bobkov and M. Ledoux, *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*, Mem. Amer. Math. Soc. **261** (2019), no. 1259. DOI.

[5] A. Delaigle, *An alternative view of the deconvolution problem*, Statist. Sinica **18** (2008), 1025–1045. Journal copy.

[6] 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. arXiv:1302.6103.

[7] P. Hall and S. N. Lahiri, *Estimation of distributions, moments and quantiles in deconvolution problems*, Ann. Statist. **36** (2008), 2110–2134. arXiv:0810.4821.

[8] J. Niles-Weed and Q. Berthet, *Minimax estimation of smooth densities in Wasserstein distance*, Ann. Statist. **50** (2022), 1519–1540. DOI.

[9] C. Matias, *Semiparametric deconvolution with unknown noise variance*, ESAIM Probab. Stat. **6** (2002), 271–292. DOI.

[10] J. Park and A. K. H. Kim, *Minimax rate-optimal estimation in log-concave deconvolution*, Joint Statistical Meetings, 2026, conference abstract. Program abstract.

[11] 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.

[12] L. A. Caffarelli, *Monotonicity properties of optimal transportation and the FKG and related inequalities*, Comm. Math. Phys. **214** (2000), 547–563. DOI.

[13] E. Caglioti and F. Pieroni, *Random matching in 2D with exponent 2 for Gaussian densities*, J. Stat. Phys. **191** (2024), article 62. DOI.

[14] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math. **41** (1988), 909–996. DOI.

[15] M. Ledoux, *On optimal matching of Gaussian samples II*, 2018. Author manuscript.

[16] T. Manole, S. Balakrishnan, J. Niles-Weed, and L. Wasserman, *Plugin estimation of smooth optimal transport maps*, Ann. Statist. **52** (2024), 966–998. arXiv:2107.12364.

[17] S. Balakrishnan and T. Manole, *Stability bounds for smooth optimal transport maps and their statistical implications*, Electron. J. Stat. **20** (2026), 3078–3105. DOI.

[18] V. Divol, J. Niles-Weed, and A.-A. Pooladian, *Optimal transport map estimation in general function spaces*, Ann. Statist. **53** (2025), 963–988. DOI.

[19] C. Durot and D. Mukherjee, *Minimax optimal rates of convergence in monotone shuffled and unlinked regression models under vanishing noise*, Bernoulli **32** (2026), 2244–2270. DOI; arXiv:2404.09306.

[20] X. Nguyen, *Convergence of latent mixing measures in finite and infinite mixture models*, Ann. Statist. 41 (2013), 370–400.

[21] S. Sugasawa and Z. Zhao, *Beyond Tweedie's formula: Conditional score modeling for empirical Bayes inference*, arXiv:2609.11136v1, 2026.

[22] A. Prékopa, *On logarithmic concave measures and functions*, Acta Sci. Math. (Szeged) 34 (1973), 335–343.

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

[24] 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.

[25] M. Ledoux, *Heat flow derivatives and minimum mean-square error in Gaussian noise*, IEEE Trans. Inform. Theory 62 (2016), 3401–3409.

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

[27] T. Mizuno, *A relation between positive dependence of signal and the variability of conditional expectation given signal*, J. Appl. Probab. 43 (2006), 1181–1185.

[28] R. Peyre, *Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance*, ESAIM Control Optim. Calc. Var. **24** (2018), 1489–1501. DOI.

[29] P. Rigollet and J. Weed, *Uncoupled isotonic regression via minimum Wasserstein deconvolution*, Inf. Inference **8** (2019), 691–717. DOI.