Wasserstein testing and confidence sets under Gaussian noise

Scott Schubr
Luminary Oddities

November 2026

Abstract

We determine the minimax separation rates for testing log-concave laws in latent quadratic Wasserstein distance when observations carry known, partly known or unknown Gaussian noise. For goodness of fit against a fixed reference, and for two-sample testing with unknown latent laws, the separation equals in order the sum of a direct radius, with explicit logarithmic corrections in dimensions two to four, and an inverse radius driven by the noise level. With unequal sample sizes and different noise covariances the rate is the maximum of the group-specific rates. Honest adaptive confidence sets have sharp diameters, with a logarithmic loss of permissible variance at the exact adaptation boundary in high dimension. Independently calibrated noise adds $S/\sqrt{k}$ to every rate, a single calibration-free test attains the known-covariance rate under shared noise, and tests adapt to unknown smoothness and unknown covariance without loss in dimensions one to four.

Contents

1 Introduction **4**

2 Wasserstein goodness-of-fit testing with vanishing Gaussian noise **4**
2.1 The testing problem 4
2.2 Deterministic transport and Fourier fields 5
2.3 The test for small noise 7
2.3.1 Screening the exterior and the small-occupancy cells 7
2.3.2 Gaussian coarse fields and the inverse field 8
2.3.3 Covariance of the observed fields 9
2.3.4 The quadratic test 11
2.4 Larger noise and covariance completion 11
2.5 Lower bounds from mixtures 12
2.5.1 The direct polynomial lower bound 12
2.5.2 The direct critical lower bound 13
2.5.3 The inverse mixture 14
2.6 Testing below dimension four 15
2.6.1 A spatially varying coarse heat time 15
2.6.2 Covariance at a variable heat time 16
2.6.3 Matching lower bounds 16
2.7 An honest confidence diameter at a specified reference 17
2.8 Relation to direct and indirect testing 17

3 Two-sample Wasserstein testing with unequal sizes and Gaussian noise 18
3.1 Experiment and separation . . . 18
3.2 Density bounds and a weighted transport comparison . . . 19
3.2.1 The reverse transport comparison and the heat fields . . . 20
3.3 Effective scale and large-noise groups . . . 21
3.4 Estimating the common weights from inward probes . . . 22
3.5 Matched latent fields from different observation kernels . . . 23
3.5.1 Conditional covariance . . . 23
3.6 The unequal-sample statistic . . . 25
3.7 Lower bounds from mixtures . . . 25
3.7.1 The direct polynomial lower bound . . . 26
3.7.2 The direct critical lower bound . . . 26
3.7.3 The inverse mixture . . . 27
3.7.4 Lower bounds in dimensions one through three . . . 28
3.8 Relation to direct and indirect two-sample tests . . . 28

4 Two-sample Wasserstein testing with unknown latent distributions and Gaussian noise 28
4.1 Experiment and separation . . . 29
4.2 Deterministic transport and Fourier fields . . . 29
4.3 Estimating the common weights from inward probes . . . 30
4.4 A held-out quadratic test . . . 31
4.4.1 Conditional covariance . . . 32
4.4.2 Rejection rule . . . 33
4.5 Larger noise and covariance completion . . . 34
4.6 Lower bounds from mixtures . . . 34
4.6.1 The direct polynomial lower bound . . . 35
4.6.2 The direct critical lower bound . . . 35
4.6.3 The inverse mixture . . . 36
4.6.4 Lower bounds in dimensions one through three . . . 37
4.7 Relation to direct and indirect two-sample tests . . . 37

5 Honest adaptive Wasserstein confidence sets under Gaussian contamination 37
5.1 Statement of the adaptation theorem . . . 38
5.2 Density bounds and a weighted transport comparison . . . 39
5.2.1 The reverse transport comparison and the heat fields . . . 40
5.3 Estimating the common weights from inward probes . . . 41
5.4 The validation fields . . . 42
5.4.1 Conditional covariance . . . 43
5.4.2 Estimating the validation loss . . . 44
5.5 Point estimation at both smoothness levels . . . 44
5.6 Construction and coverage . . . 45
5.7 Lower bounds for confidence diameter . . . 46
5.7.1 Estimation families . . . 46
5.7.2 Testing families at the Gaussian . . . 48
5.8 The adaptation window . . . 51

6 Wasserstein inference with independently calibrated Gaussian noise 51
6.1 Experiments and rates . . . 51
6.2 Gaussian smoothing in the curvature class . . . 52
6.3 A tolerant known-covariance comparison . . . 53
6.4 Calibration reduction and upper bounds . . . 53
6.5 Lower bounds . . . 54
6.6 Deterministic covariance uncertainty . . . 55
6.7 Allocation of error measurements . . . 55

7 Calibration-free Wasserstein testing with shared Gaussian noise 55
7.1 Experiment and separation . . . 55
7.2 Deterministic transport and Fourier fields . . . 56
7.3 Estimating the common weights from inward probes . . . 57
7.4 A differentiable family of inverse fields . . . 58
7.4.1 Conditional covariance . . . 59
7.5 The covariance and resolution scan . . . 61
7.5.1 Coverage of the actual noise scale . . . 61
7.6 A common empirical test at larger noise . . . 62
7.7 Lower bounds from mixtures . . . 62
7.7.1 The direct polynomial lower bound . . . 63
7.7.2 The direct critical lower bound . . . 63
7.7.3 The inverse mixture . . . 64
7.7.4 Lower bounds in dimensions one through three . . . 64
7.8 Common noise with repeated measurements . . . 65
7.9 Comparison with measurement-error tests . . . 65

8 Joint adaptation to smoothness and Gaussian noise in Wasserstein testing 65
8.1 Separation with unknown smoothness . . . 66
8.2 Density bounds and a weighted transport comparison . . . 66
8.2.1 The reverse transport comparison and the heat fields . . . 68
8.3 Gaussian smoothing in the curvature class . . . 68
8.3.1 A reverse bound for small common heat smoothing . . . 69
8.4 A weighted feature at one trial . . . 70
8.5 The conditional sign test . . . 71
8.6 Choosing the trials . . . 72
8.6.1 Known covariance . . . 72
8.6.2 Unknown covariance . . . 72
8.6.3 Large noise . . . 73
8.7 Mixtures over smoothness and noise levels . . . 73
8.7.1 The clean adaptation lower bound . . . 74
8.7.2 A shared observed null at different noise levels . . . 74
8.7.3 The inverse lower bound with supplied covariance . . . 75
8.7.4 The extra inverse loss with both parameters unknown . . . 76

9 Adaptation without loss in low-dimensional Gaussian Wasserstein testing 76
9.1 Experiment and separation . . . 76
9.2 Density bounds and a weighted transport comparison . . . 77
9.2.1 The reverse transport comparison and the heat fields . . . 78

9.3 Gaussian smoothing in the curvature class . . . 79
9.3.1 A reverse bound for small common heat smoothing . . . 79
9.4 Estimating the common weights from inward probes . . . 80
9.5 One common field and a family of corrections . . . 81
9.5.1 Population comparison at a nearby matrix . . . 82
9.5.2 Conditional covariance . . . 82
9.6 The conditional sign test . . . 84
9.7 Choosing the correction trials . . . 84
9.8 A common empirical test at larger noise . . . 85
9.9 Lower bounds from mixtures . . . 85
9.9.1 The direct critical lower bound . . . 86
9.9.2 The inverse mixture in dimension four . . . 86
9.9.3 Lower bounds in dimensions one through three . . . 87

1 Introduction

Testing in Wasserstein distance asks for the smallest latent separation at which two hypotheses can be distinguished from contaminated data. Without noise, Tran and Schreuder [2] showed that a sliced Wasserstein permutation test attains the minimax separation $n^{-1/2}$; our direct radius agrees with this rate in dimension one and identifies the logarithmic corrections in dimensions two to four. Under Gaussian noise the answer combines the direct testing radius, which improves on the estimation rate, and an inverse radius set by the deconvolution modulus of [1]. Section 2 treats goodness of fit, Sections 3–4 two-sample testing, and Section 5 honest adaptive confidence sets. Sections 6–9 remove knowledge of the noise and of the smoothness. For estimation, the Archive article *Sharp transport identification with unknown Gaussian covariance* [3] showed that m pure-noise observations reduce the covariance ambiguity to $\min(\text{uncertainty}, m^{-1/2})$ and that independent-error pairs recover the known-noise order. Section 6 proves the testing and two-sample counterparts: independent calibration adds $S/\sqrt{k}$ to the known-noise separation, and paired repeats replace n by $n + k$, so identification and testing pay the same calibration price.

2 Wasserstein goodness-of-fit testing with vanishing Gaussian noise

We determine the minimax separation rate for testing a known latent density in quadratic Wasserstein distance from independently sampled Gaussian-contaminated observations in a two-sided curvature class of potential smoothness $\alpha \geq 2$. The direct separation orders in dimensions one through four are $n^{-1/2}$ times, respectively, 1, $(\log\log n)^{1/4}$, $(\log n)^{1/4}$, and $(\log n)^{3/4}$. In dimension $d > 4$ the direct order is $n^{-2(\alpha+1)/(4\alpha+d)}$. The inverse term is $(s/\log(2 + ns^{\alpha+d/4}))^{(\alpha+1)/2}$ for $d \geq 4$, with $\log n$ in the denominator for $d < 4$, where s is the largest eigenvalue of the known, possibly singular noise covariance. The test uses a quadratic statistic of localized transport fields and screens for large discrepancies in the tails. Gaussian packet mixtures give the inverse lower bound at small noise when $d \geq 4$; a single packet suffices in the remaining regimes. The result holds about each known reference strictly inside the curvature and smoothness bounds.

2.1 The testing problem

Fix an integer $d \geq 1$ and a real smoothness order $\alpha = k + \theta \geq 2$, with integer k and $0 \leq \theta < 1$. Let $\mathcal{F}_\alpha$ consist of probability densities $f = e^{-V}$ on $\mathbb{R}^d$ with $V \in C^k$, $mI \preceq D^2V \preceq MI$ for $0 < m < M < \infty$, and fixed positive bounds on derivatives D^jV of orders $3, \ldots, k$ in the sup norm.

If $\theta > 0$, additionally bound the global θ-Hölder seminorm of $D^k V$ by a fixed positive constant. Translations are unrestricted. The known reference density g satisfies each bound with positive slack. We observe $Y_i = X_i + Z_i$, with independent X_i of density f and independent centered Gaussian Z_i of known covariance $\Sigma \succeq 0$, independent of the latent sample. Put $s = \|\Sigma\|_{\mathrm{op}} \le \Lambda$. Use $\gamma_\Sigma = N(0, \Sigma)$ and $\gamma_s = N(0, sI_d)$. At separation r, the risk of a test $\phi \in [0,1]$ is

$$\mathfrak{T}_n(r, \Sigma; \phi) = \mathbb{E}_g \phi + \sup_{\substack{f \in \mathcal{F}_\alpha \\ W_2(f,g) \ge r}} \mathbb{E}_f(1 - \phi).$$

Here expectations are for the distribution of the observations.

Theorem 2.1. *Define the direct orders by*

$$a_{n,d} = \begin{cases} n^{-1/2}, & d = 1, \\ (\log\log n)^{1/4}/\sqrt{n}, & d = 2, \\ (\log n)^{1/4}/\sqrt{n}, & d = 3, \\ (\log n)^{3/4}/\sqrt{n}, & d = 4, \\ n^{-2(\alpha+1)/(4\alpha+d)}, & d > 4, \end{cases}$$

Set $r_n(s) = a_{n,d} + (s/B_{n,d}(s))^{(\alpha+1)/2}$, where $B_{n,d}(s) = \log n$ if $d < 4$, and $B_{n,d}(s) = \log(2 + ns^{\alpha+d/4})$ if $d \ge 4$. For every fixed $0 < \varepsilon < 1/2$, there are positive $c_\varepsilon, C_\varepsilon, n_0$, depending only on the fixed class parameters, the reference slack and ε, such that uniformly over $n \ge n_0$ and the indicated covariances,

$$\inf_\phi \mathfrak{T}_n(c_\varepsilon r_n(s), \Sigma; \phi) \ge 1 - \varepsilon, \qquad \inf_\phi \mathfrak{T}_n(C_\varepsilon r_n(s), \Sigma; \phi) \le \varepsilon.$$

The transition variance is of order $n^{-4/(4\alpha+d)}$ for $d > 4$, and $n^{-1/(\alpha+1)}(\log n)^{3/[2(\alpha+1)]} \log\log n$ for $d = 4$. For $d < 4$ it is of order $(\log n) a_{n,d}^{2/(\alpha+1)}$.

Fixed-reference latent Wasserstein testing at fixed positive scalar noise was obtained in [12, Theorem 2.4] at $\alpha = 2$, with a broader log-concave alternative class subject to moment normalization. The tests here allow unnormalized smooth alternatives and treat the direct and shrinking-noise regimes. Quadratic deconvolution tests for density L^2 loss go back to [6]. Smooth direct Wasserstein testing has also been reported in [11]. The separation here uses a random-sign mixture and a weighted quadratic transport statistic. In $d \ge 4$, the effective noise parameter is $ns^{\alpha+d/4}$.

2.2 Deterministic transport and Fourier fields

Uniformly over the class, f has Gaussian upper and lower envelopes about its mean m_f, the score of f is $O(1 + |x - m_f|)$, and

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

Indeed quadratic bounds at the mode give the envelopes and bound the distance between the mode and the mean. Differentiating e^{-V} gives the derivative bounds. For $\theta > 0$, on $|x - m_f| \le L$ and $|u| \le (2 + L)^{-1}$, the same product rule and the Hölder bound on $D^k V$ give

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

The local density ratio is bounded by the score, derivatives of lower order in the product are Lipschitz, and the highest one uses the assumed Hölder seminorm.

For $p_f = f * \gamma_s$, isotropic $0 \le s \le \Lambda$, Gaussian integration of the two quadratic Taylor bounds gives

$$cf(x) \le p_f(x) \le f(x) e^{Cs(1+|x-m_f|^2)}. \tag{1}$$

The potential of p_f also has Hessian in a fixed positive interval. For $s > 0$, its Hessian is $s^{-1} I - s^{-2} \operatorname{Cov}(X \mid Y)$. The conditional potential has Hessian between $(m + s^{-1})I$ and $(M + s^{-1})I$. The upper covariance bound follows from Brascamp–Lieb [5]. The lower bound follows, in each unit direction v, from Cauchy–Schwarz and the conditional identities

$$\mathbb{E}[(v \cdot (X - \mathbb{E}X))(v \cdot \nabla A(X))] = 1, \qquad \mathbb{E}[(v \cdot \nabla A(X))^2] = \mathbb{E}[v^T D^2 A(X) v],$$

where A is the conditional potential. Consequently the scores of p_f are $O(1 + |x - m_f|)$. The observed law always has a uniformly Gaussian upper envelope about m_f.

Use $\widehat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\, dx$. Let $\widehat{K}$ be a real even smooth Fourier cutoff that equals one near zero and is supported in the unit ball. Use $K_b = b^{-d} K(\cdot/b)$ for its spatial kernel, and define constant-bandwidth fields

$$U_b(f) = K_b * f, \qquad \widehat{J_b(f)}(\xi) = -\frac{i\xi}{|\xi|^2}(1 - \widehat{K}(b\xi))\widehat{f}(\xi).$$

Then $f = U_b(f) + \operatorname{div} J_b(f)$. The inverse transform Q of $-i\xi(1 - \widehat{K}(\xi))/|\xi|^2$ is integrable with all absolute moments finite and all signed moments zero. In fact a dyadic partition into shells 2^j away from the origin gives inverse transforms bounded by $C_N 2^{(d-1)j}(1 + 2^j|x|)^{-N}$. Summability of their L^1 norms $O(2^{-j})$ gives the assertion, with moment cancellation since the symbol vanishes near zero. For $|x - m_f| \le L$, $w = (2 + L)^{-1}$, and $h \le w$, subtract the local Taylor polynomial of f to degree $k - 1$ in the integer case and k otherwise. This gives

$$|J_h(f)(x)| \le Ch^{\alpha+1}(1 + |x - m_f|^{k+1}) f(x) + C_N h (h/w)^N. \tag{2}$$

Indeed $J_h(f) = hQ_h * f$, the short-shift remainder has order α, and the farther shifts use uniform absolute derivative bounds and the arbitrary integrated tails.

Translate the known reference so its mode is zero. If the means of f, g are within a fixed distance, put $\rho = f + g$. For a test φ, set $u = \varphi - \frac{1}{2}\int \varphi\rho$. Then

$$\int (1 + |x|^2)|u|^2 \rho \le C \int |\nabla\varphi|^2 \rho. \tag{3}$$

Both components satisfy ordinary Poincaré by Brascamp–Lieb. Comparing their component means to a common Lebesgue average on a fixed ball with a positive density floor gives the mixture inequality. For either potential V, $x \cdot \nabla V(x) \ge m|x|^2/2 - C$, since its mode stays close to zero. Integration by parts gives

$$\int u^2 x \cdot \nabla V f = d \int u^2 f + 2 \int ux \cdot \nabla\varphi f.$$

Use Cauchy–Schwarz to absorb half of the resulting quadratic term, add the two estimates, and apply mixture Poincaré. Gaussian tails justify integration by parts for compactly supported tests minus their mean.

Let $I_0 \subset I$ admit a C^1 cutoff supported in I, equal to one on I_0, with $|\nabla\chi| \le C(1 + |x|)$. For any identity $f - g = U + \operatorname{div} J$ in distributions,

$$W_2(f, g) \le C \left[\left(\int_{I_0^c} \frac{\rho}{1 + |x|^2} \right)^{1/2} + \left(\int_I \frac{|U|^2/(1 + |x|^2) + |J|^2}{\rho} \right)^{1/2} \right]. \tag{4}$$

To see this, split $\int (f-g)u$ with the cutoff and use Cauchy–Schwarz with (3) in the exterior. In the interior insert the field identity and integrate by parts. Approximation by smooth tests justifies this for a C^1 cutoff. The $L^2(\rho)$ norms of $(1+|x|^2)^{1/2}\chi u$ and $\nabla(\chi u)$ are at most $C\|\nabla\varphi\|_{L^2(\rho)}$. Riesz representation produces an exact flux of divergence $f-g$ whose squared action norm $\int |J|^2/\rho$ is bounded by the square of the displayed right-hand side. Interpolate between the densities with $(1-\lambda)f+\lambda g$, $\lambda(t)=\sin^2(\pi t/2)$, and flux $\lambda'(t)J$. Its action is at most $\pi^2\int |J|^2/\rho$, by the lower bound $\min(\lambda,1-\lambda)\rho$. The dynamic transport inequality [4] gives (4).

2.3 The test for small noise

Assume $d \geq 4$ and isotropic noise sI, with $s \leq n^{-\eta}$ for a sufficiently small fixed $\eta > 0$. Put $a = \log n$, $B_n = \log(2 + ns^{\alpha+d/4})$, and define

$$h_0 = \max\left(a_{n,d}^{1/(\alpha+1)}, \sqrt{\frac{s}{\gamma B_n}}\right), \qquad r = h_0^{\alpha+1}.$$

Here $\gamma > 0$ will be sufficiently small. The separation r is comparable to $r_n(s)$, and h_0 lies between fixed negative powers of n in the present range. Split the sample into a screening part and an independent statistic part, of sizes comparable to n.

2.3.1 Screening the exterior and the small-occupancy cells

Write $\tau(x) = -\log g(x) + \log g(0)$, with minimum zero. For $u \geq 1$,

$$\frac{d}{du}\,\mathrm{Vol}\{\tau \leq u\} \asymp u^{d/2-1}, \qquad g\{\tau > u\} \asymp e^{-u}u^{d/2-1}. \tag{5}$$

Along each ray from zero the radial coordinate of level u is comparable to $\sqrt{u}$, and the radial derivative of τ there is also comparable to $\sqrt{u}$. The polar-coordinate change of variables proves the first estimate, and integrating e^{-u} against it proves the second.

Take $u \asymp a$ such that $e^{-u}u^{d/2-1} = ar^2$, using the decreasing branch for large u. Let $I_0 = \{\tau \leq u\}$ and $I = \{\tau \leq u+1\}$. A cutoff in the scalar level has gradient bounded by $C(1+|x|)$, so (4) applies. By (1) and $|x|^2 \leq C\tau(x)$ away from the mode, $p_g\{\tau > u\} \leq Car^2$: integrate $e^{-(1-Cs)t}t^{d/2-1}$ for $t > u$, and use $su = o(1)$. Let $w = a^{-1/2}$. Tile with cubes of side w and include all full cubes that meet I. Their number is $O(a^d)$. If q_j is the null observed probability of a cube, then for x in it,

$$q_j \asymp w^d e^{-\tau(x)}, \qquad nq_j \gtrsim nr^2a^{2-d}.$$

This uses local comparability of g on the cube, the local bound $p_g \asymp g$, and the fact that τ increases by at most a constant on the extra cube layer. The lower bound is polynomial in n for $d > 4$. For $d = 4$ it is at least $ca^{-1/2}$.

First reject if the sample mean differs from m_g by more than one or if the exterior empirical mass on I_0^c exceeds a sufficiently large fixed multiple of ar^2. Under the null both probabilities are $o(1)$ by concentration. If the population mean differs by more than two, or if the population exterior probability exceeds twice the chosen threshold, the screening rejects with probability $1-o(1)$. In the tail test the mean count at threshold diverges, since $nar^2 \geq a^{5/2}$ at $d = 4$ and is polynomial when $d > 4$. This uses Gaussian tail concentration only for the sample mean and binomial concentration for the indicator count.

Also screen the cube counts. Write $\lambda_j = n_1 q_j$, where n_1 is the screening sample size. For $\lambda_j \geq C_0 \log a$, reject if its count exceeds $2\lambda_j$. For $a^{-1/4} \leq \lambda_j < C_0 \log a$, reject if it exceeds $C_1 \log a$.

For $\lambda_j < a^{-1/4}$, reject if it exceeds a fixed integer K_0. Choose C_0 large enough for the union bound on the first tier, C_1 sufficiently larger than C_0 for the middle tier, and $K_0 > 4d+4$ for the last tier. Under the null, the first two tiers have binomial Chernoff probabilities $O(a^{-d-2})$ per cube, and the last has probability $O(a^{-K_0/4})$ by the factorial-moment bound. Thus null screening has error $o(1)$. If a population cube mean count exceeds a sufficiently large multiple of its threshold, that cube rejects with probability at least $1-\varepsilon/4$, uniformly for large n. For a sparse cell this is a fixed-error guarantee obtained by increasing K_0 if necessary.

It remains to control alternatives satisfying the three population bounds: bounded mean difference, exterior mass $\le Car^2$, and all cube mean counts at most a fixed multiple of their thresholds. Call these the screened laws. The observed score bound gives comparability of p_f on each full cube, since $|m_f - m_g|$ is bounded. Therefore

$$p_f(x)/g(x) \le \mathcal{M}(x), \qquad x \in I, \tag{6}$$

where $\mathcal{M}$ is bounded by a constant when $d > 4$. For $d = 4$, set $\lambda(x) = nw^4 e^{-\tau(x)}$. An envelope is a constant plus the following two terms, with fixed constants in the indicator thresholds:

$$C\frac{\log a}{\lambda(x)}\mathbf{1}_{\{ca^{-1/4}\le\lambda(x)\le C\log a\}} \;+\; C\frac{1}{\lambda(x)}\mathbf{1}_{\{\lambda(x)\le Ca^{-1/4}\}}.$$

It has

$$\|\mathcal{M}\|_\infty = O(\sqrt{a}), \quad \int_I \mathcal{M}^2 = O(a^2), \quad \int_I \frac{\mathcal{M}^2}{(1+|x|^2)^2} = O(\log a), \quad \sup_I \frac{\mathcal{M}}{1+|x|^2} = O(1). \tag{7}$$

Indeed the middle tier lies in a level band of width $O(\log a)$, of volume $O(a\log a)$ by (5). Its squared envelope is at most $O(a^{1/2}(\log a)^2)$, giving $o(a^2)$. On the sparse tier, integrate $1/\lambda(x)^2$ over the levels. The integrand grows as $e^{2\tau}$, the level-volume derivative is $O(a)$, and its maximum is $O(a)$ by the lower bound on $\lambda(x)$. Hence its integral is $O(a^2)$. The additional terms are supported at $\tau \ge a - O(\log a)$, so $1+|x|^2 \asymp a$ there. The constant part of the envelope gives $\int_I (1+|x|^2)^{-2} = O(\log a)$ in four dimensions. This proves (7). Finally (1) gives $f \le Cp_f$ globally. In particular every screened law has latent exterior mass $\le Car^2$, and $\rho/(1+|x|^2)$ there has integral $O(r^2)$ since $\tau \ge u \asymp a$.

2.3.2 Gaussian coarse fields and the inverse field

Let $b^2 = 1/a$, $v = b^2 - s \asymp b^2$, and use the exact coarse fields

$$U(f) = f * \gamma_{b^2}, \qquad J(f) = \tfrac{1}{2}\int_0^{b^2} \nabla(f*\gamma_t)\,dt.$$

They satisfy $f = U(f) - \operatorname{div} J(f)$. For $d = 4$ use $h(x) = h_0$; for $d > 4$ use $h(x) = h_0 e^{c_0|x|^2}$ on I. Choose $c_0 > 0$ small enough that $h(x)$ stays polynomially small on I, and $e^{(2\alpha+2)c_0|x|^2} f(x)$ has bounded polynomial moments for all screened laws. This is possible by the polynomial upper bound on h_0, the radius $O(\sqrt{a})$ of I, and the uniform centered Gaussian envelopes. In particular $h(x)^2 \ll v$. When $|x - x'| \le 2w$, $h(x)/h(x')$ lies between fixed positive constants.

For an observation y form a scalar feature $\gamma_v(x-y)$, where the notation denotes a Gaussian density, and two vector features. The coarse vector kernel is $R_v - R_{h(x)^2}$, where $R_t = \frac{1}{2}\int_0^t \nabla\gamma_z\,dz$. The fine vector kernel for a fixed $h = h(x)$ has symbol

$$\frac{i\xi}{|\xi|^2}\left(\widehat{K}(h\xi)e^{s|\xi|^2/2} - e^{-h^2|\xi|^2/2}\right). \tag{8}$$

In the fine feature only, truncate the spatial kernel to $|x-y| \le w$. The scalar and coarse vector features retain their full spatial kernels. Without truncation the scalar expectation is $U(f)$, and the sum of the two vector expectations differs from $J(f)$ by $J_{h(x)}(f)(x)$ as defined in (2). This follows by adding the two symbols: at input p_f, their untruncated sum is obtained by deconvolving the multiplier $i\xi(\widehat{K}(h\xi) - e^{-b^2|\xi|^2/2})/|\xi|^2$. Subtracting the symbol of $J(f)$ leaves exactly the flat-top high-pass error. The bandwidth is held fixed at x in this identity.

The fine kernel has the form $h^{1-d}L_A((x-y)/h)$, with $A = s/h^2$. The scaled symbol is smooth at zero because near zero the difference in (8) is radial analytic and vanishes to second order. Its Schwartz seminorms after inversion are at most $C_N(1+A)^{C_N}e^{CA}$, using the Fourier cutoff and the Gaussian far part. The truncation changes the mean by $O(n^{-J_0})$ uniformly on I, for any prescribed $J_0 > 0$, since h is polynomially small, $h \ge h_0$, and $A = O(\log n)$.

In the vector expectation the remaining weighted norm of the high-pass bias with denominator g is $O(r)$. Indeed (2) reduces its squared short-shift contribution to

$$C \int_I h(x)^{2\alpha+2}(1 + |x - m_f|^{2k+2})\, f(x)^2/g(x)\, dx.$$

When $\mathcal{M}$ is bounded, use $f^2/g \le Cf$ and the Gaussian moments with the chosen c_0. In the critical dimension $h = h_0$, the same argument applies where the extra envelope terms vanish. Their support lies at $|x| \gtrsim \sqrt{a}$, where $f^2/g \le C\sqrt{a}f$ and the Gaussian envelope pays this factor and every fixed polynomial moment. The far bias and the truncation error are negligible by $g(x) \ge n^{-P}$ on I and arbitrary decay powers. These bounds also hold at the null.

2.3.3 Covariance of the observed fields

Write $D(x) = 1 + |x|^2$ and normalize the scalar feature by $\sqrt{D(x)g(x)}$ and the vector feature by $\sqrt{g(x)}$. They define a feature F_y in the direct sum $L^2(I) \oplus L^2(I; \mathbb{R}^d)$, with the vector feature the sum of its coarse and fine parts. Write C_f for its centered covariance operator and $m = \mathbb{E}_f F_Y - \mathbb{E}_g F_Y$. From (4), the screened exterior bound and the bias just proved,

$$W_2(f, g) \le C(r + \|m\|). \tag{9}$$

Here the denominator $\rho = f + g$ in (4) is at least g.

To bound covariance start with the uncentered positive second-moment operators of the three separate features. The combined centered covariance is bounded in positive operator order by three times the sum of these, embedding the operators in the direct sum, by the scalar Cauchy–Schwarz inequality for inner products with any test vector. Both the operator and Hilbert–Schmidt norms are monotone for positive operators in this order. We give kernel estimates for these three operators.

For a screened law, the tangent-plane inequality for $\log p_f$ at $x \in I$, its uniformly controlled score, and (6) give the global inequality

$$p_f(y)/g(x) \le \mathcal{M}(x) e^{C\sqrt{a}\,|x-y|}.$$

Taking the geometric mean at x, x' bounds the normalized density in a covariance kernel by

$$\sqrt{\mathcal{M}(x)\mathcal{M}(x')}\, e^{C\sqrt{a}(|x-y|+|x'-y|)/2},$$

with $\mathcal{M}(x)/D(x)$ in place of $\mathcal{M}(x)$ for the scalar feature. The exponential factors can be absorbed into the Gaussians of variance $v \asymp 1/a$ when present, by completing a square. They are bounded on the two fine-kernel supports.

For the scalar kernel this gives the bound

$$Cv^{-d/2}e^{-c|x-x'|^2/v}\sqrt{\frac{\mathcal{M}(x)\mathcal{M}(x')}{D(x)D(x')}}.$$

Schur's inequality and Cauchy–Schwarz in x give operator norm $O(\|\mathcal{M}/D\|_\infty)$ and squared Hilbert–Schmidt norm $O(v^{-d/2}\|\mathcal{M}/D\|_2^2)$. In dimension four these are $O(1)$ and $O(a^2\log a)$ by (7). For $d>4$ they are $O(1)$ and at most a fixed power of a.

For the coarse vector kernel, integrate the gradient between h^2 and v to obtain

$$|R_v(u)-R_{h^2}(u)|\le C(h+|u|)^{1-d}e^{-c|u|^2/v}.$$

For $|u|\le h$, directly integrating the time powers gives $C|u|h^{-d}$ with the displayed Gaussian factor. For larger u, keep half the Gaussian at time v and extend the remaining time integral to infinity; it is $O(|u|^{1-d})$. After absorbing the two score tilts, the convolution of the absolute kernels at x,x' is bounded, for $|x-x'|\le 2w$, by

$$C(h(x)+|x-x'|)^{2-d}e^{-c|x-x'|^2/v}.$$

For $|x-x'|>2w$, the unregularized factor $C|x-x'|^{2-d}$ suffices with the same Gaussian. Indeed the bandwidths are comparable on the near region. The elementary convolution inequality follows by splitting near each center and its complement: at separation z above the regularization, the near contribution is $C|z|^{1-d}\int_0^{|z|}dr$, and the far one is $C\int_{|z|}^\infty r^{1-d}dr$. When $|z|\le h$, Cauchy–Schwarz gives Ch^{2-d}. A Gaussian in z can be extracted since $|y-x|^2+|y-x'|^2\ge|x-x'|^2/2$.

The convolution bound is multiplied by $\sqrt{\mathcal{M}(x)\mathcal{M}(x')}$. Its operator norm is $O(v\|\mathcal{M}\|_\infty)$ by integrating the unregularized kernel. If $d=4$, h is constant and the square integral of the convolution kernel in z is $O(1+\log(\sqrt{v}/h))=O(a)$. Then Cauchy–Schwarz for the two envelope factors and (7) give squared Hilbert–Schmidt norm $O(a^3)$. If $d>4$, the near contribution to that squared norm is $O(\int_I h(x)^{4-d}\,dx)$ because $\mathcal{M}$ is bounded, and the far contribution is bounded by a fixed power of a. The adaptive $h(x)$ gives

$$\int_I h(x)^{4-d}\,dx\le Ch_0^{4-d}.$$

The remaining polylogarithmic term is smaller than h_0^{4-d}, since h_0 is polynomially small.

For the fine kernel, put $A_0=s/h_0^2$ and let $E=C(1+A_0)^Ce^{CA_0}$ bound the finite collection of Schwartz seminorms used here. Its covariance kernel vanishes for $|x-x'|>2w$. On the near region, the absolute convolution after scaling and using comparable bandwidths is at most

$$CE^2h(x)^{2-d}(1+|x-x'|/h(x))^{-N}\sqrt{\mathcal{M}(x)\mathcal{M}(x')}.$$

For every fixed N needed in the integrals this follows from convolution of two profiles decaying at a larger fixed power. Hence its operator norm is at most $CE^2\sup_I h(x)^2\,\|\mathcal{M}\|_\infty$. The squared Hilbert–Schmidt norm is $O(E^4h_0^{4-d})$ for $d>4$ by the adaptive integral above. For $d=4$, h is constant, the squared integral in the difference variable has no power of h, and Cauchy–Schwarz gives $O(E^4\|\mathcal{M}\|_2^2)=O(E^4a^2)$.

Combining the three bounds, uniformly at the null and for screened laws,

$$\|C_f\|_{\mathrm{HS}}/n\le Cr^2,\qquad \|C_f\|_{\mathrm{op}}/n=o(r^2).\tag{10}$$

Here are the scale comparisons. If $d > 4$, the coarse polylogarithms divided by n are negligible relative to $a_{n,d}^2$. With $X = ns^{\alpha+d/4}$ and $A_0 \le \gamma \log(2+X)$, a sufficiently small γ gives

$$E^2 \le Cnh_0^{2\alpha+d/2},$$

since the right-hand side dominates $\max(1, X/(\gamma\log(2+X))^{\alpha+d/4})$ and $E^2 \le C(2+X)^c$ for some $c < 1$. Thus $E^2 h_0^{2-d/2}/n \le Cr^2$, and the fine operator term is smaller by $O(h_0^{d/2-2}\sup_I h^2)$. If $d = 4$, the coarse Hilbert–Schmidt term divided by n is $O(a^{3/2}/n)$, and the coarse operator term is $O(n^{-1})$. The fine term is $O(E^2 a/n)$. For $X' = 2 + ns^{\alpha+1} \le a^2$, take $E^2 \le C(X')^c$ with $c < 1/4$ to pay it by $a^{3/2}/n$. For $X' > a^2$, use $a \le \sqrt{X'}$ and $(X')^{1/2+c} \le C(X'-2)/\log^{\alpha+1} X'$ to pay it by $h_0^{2\alpha+2}$. The fine operator term divided by n is smaller than $E^2 a/n$ by $O(h_0^2/\sqrt{a})$, using $\|\mathcal{M}\|_\infty = O(\sqrt{a})$. This proves (10).

2.3.4 The quadratic test

Center each field F_y by its exact null mean μ_g. On the main sample of size N, use

$$\mathcal{U} = \frac{1}{N(N-1)} \sum_{i \ne j} \langle F_{Y_i} - \mu_g, F_{Y_j} - \mu_g \rangle.$$

Its mean is $\|m\|^2$. Expanding the quadratic average into its linear and centered degenerate parts gives

$$\mathrm{Var}_f(\mathcal{U}) = \frac{4}{N}\langle m, C_f m\rangle + \frac{2}{N(N-1)}\|C_f\|_{\mathrm{HS}}^2.$$

Reject, in addition to the screening, if $\mathcal{U} > A_\varepsilon r^2$. Under the null the first variance term vanishes and (10) gives type-I error at most $\varepsilon/4$ with A_ε large, leaving room for the null screening error. For screened alternatives at a sufficiently large multiple of r, (9) gives $\|m\|^2 \ge 2A_\varepsilon r^2$, with the separation constant increased further as needed. Chebyshev's inequality using (10) gives type-II error at most $\varepsilon/4$, by making the signal constant large and then n large. All other alternatives are rejected by the screening with probability at least $1 - \varepsilon/4$. This proves the small-noise upper bound.

2.4 Larger noise and covariance completion

If $\delta = W_1(p_f, p_g) \le 1$, their means are within one. Apply (4) with $U = U_h(f) - U_h(g)$, $J = J_h(f) - J_h(g)$ on a ball of radius $D\sqrt{\log(1/h)}$. Take D large enough that the Gaussian exterior contributes $O(h^{\alpha+1})$. On this ball there is a uniform density floor h^P for some fixed P, and (2) gives an $O(h^{\alpha+1})$ weighted bound on the vector field, paying each local bias in its own density. Coupling the output laws bounds their characteristic function difference by $\delta|\xi|$. Fourier inversion on the cutoff support then gives $\|U\|_\infty \le C\delta h^{-d-1} e^{s/(2h^2)}$. Consequently

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

Uniformly over the class, the empirical observed W_1 error e_n has expectation $O(n^{-\beta})$ with $\beta = 1/[4(d+1)]$. For an elementary proof translate to the theoretical mean, truncate at distance n^β using the uniform second moment and then quantize at mesh $n^{-\beta}$. Both costs are $O(n^{-\beta})$. There are $O(n^{2d\beta})$ sites, so Cauchy–Schwarz for the Bernoulli counts bounds expected total variation by $Cn^{d\beta-1/2}$. Multiplying by the grid diameter gives $O(n^{-1/4})$. Under the null reject if the empirical

observed law has distance more than $2n^{-\beta/2}$ from the known p_g. Its type-I error tends to zero by Markov's inequality. For any alternative whose observed discrepancy is $\delta \le 3n^{-\beta/2}$, use

$$h = \max(n^{-c_1}, \sqrt{s/(\gamma_1 \log n)})$$

in (11), with constants chosen so that $c_1(\alpha + 2 + d + P/2) + \gamma_1/2 < \beta/2$. The two terms in (11) are then $O(h^{\alpha+1})$. If $s \ge n^{-\eta}$ with $\eta > 0$ sufficiently small, the second entry of h dominates and $\log(2 + ns^{\alpha+d/4}) \asymp \log n$. Thus for a sufficiently large multiple of $r_n(s)$ in latent distance, the observed discrepancy exceeds $3n^{-\beta/2}$. On $e_n \le n^{-\beta/2}$ the test then rejects. This proves the upper bound in the remaining range.

For general known $\Sigma \succeq 0$, add independent Gaussian noise of covariance $sI - \Sigma$ to each observation, and apply the isotropic test. Its conditional rejection probability given the original observations is an admissible test and has the same two error probabilities. A deterministic test, if desired, compares that conditional probability to one half, doubling at most either error. The upper proof can use a smaller error target before that step.

2.5 Lower bounds from mixtures

We will use finite families $f_\omega = g + \sum_{i=1}^N \omega_i q_i$ where ω_i are independent signs and each q_i has integral zero. Let G be the Gram matrix of the noisy score perturbations in $L^2(p_g)$:

$$G_{ij} = \int \frac{(q_i * \gamma_\Sigma)(q_j * \gamma_\Sigma)}{g * \gamma_\Sigma}.$$

For the uniform mixture of n-sample laws over signs, one plus the chi-square divergence to the null equals $\mathbb{E}_{\omega,\omega'}(1 + \omega^T G \omega')^n$. The factor inside the power is nonnegative as the inner product of two likelihood ratios. Bounding its power by $\exp(n\omega^T G\omega')$ and integrating one sign vector gives $\mathbb{E}_\omega \exp(n^2\|G\omega\|^2/2)$ as an upper bound, using $\cosh(t) \le e^{t^2/2}$. Introduce a standard Gaussian vector to linearize the square and apply the same sign bound again. The result is at most

$$\det(I - n^2G^2)^{-1/2} \le \exp(Cn^2\|G\|_{\mathrm{HS}}^2) \qquad \text{if } n\|G\|_{\mathrm{op}} \le 1/2.$$

Thus a small bound on $n^2\|G\|_{\mathrm{HS}}^2$ makes the mixture total variation arbitrarily small. If every mixture member has $W_2(f_\omega, g) \ge r$, the sum of testing errors at separation r is at least one minus that total variation.

For transport separation we use a smooth test φ. Taylor expansion under an optimal coupling gives

$$\int \varphi(f_\omega - g) \le \|\nabla\varphi\|_{L^2(f_\omega)} W_2(f_\omega, g) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(f_\omega, g)^2. \tag{12}$$

If the pairing dominates cE, the gradient norm is at most $C\sqrt{E}$, and the Hessian norm is bounded, it follows that $W_2^2(f_\omega, g) \ge c'E$.

2.5.1 The direct polynomial lower bound

For $d > 4$, put $h = n^{-2/(4\alpha+d)}$. Pack order h^{-d} disjoint cubes of side h in a fixed cube about the reference mode. Let ψ be a nonzero smooth bump supported inside the reference unit cube with integral zero, and use $q_i(x) = a_0\psi((x - z_i)/h)$, $a_0 = \kappa h^\alpha$. The relative perturbations q_i/g have uniformly bounded C^α norm of order κ on the combined support, by rescaling and the fixed positive density floor there. The vanishing near edges gives the same bound across cubes. The logarithmic

derivative bounds follow by composition with $\log(1+\cdot)$; for sufficiently small κ the reference slack places every alternative in the class.

The clean score Gram matrix is diagonal with entries bounded by $Ca_0^2h^d$. Observed score Gram matrices are bounded above by their clean counterpart in positive operator order: the observed score is the conditional expectation of the clean score under the reference joint law, so this follows by conditional Cauchy–Schwarz for any linear combination. Thus $n^2\|G\|_{\mathrm{HS}}^2 \le C\kappa^4 n^2 h^{4\alpha+d} = C\kappa^4$. For separation take $\varphi = h^2\sum_i \omega_i q_i$. Its pairing with $f_\omega - g$ is of order $E = h^{d+2}\sum_i a_0^2 \asymp h^{2\alpha+2}$, by disjointness. The gradient's squared L^2 norm is $O(E)$, and its Hessian is uniformly bounded. Equation (12) therefore gives the direct lower bound for every sign.

2.5.2 The direct critical lower bound

When $d = 4$, pack disjoint cubes of side $\ell = (\log n)^{-1/2}$ inside a cube of side $c\sqrt{\log n}$ about the mode, with $c > 0$ small. Their total volume is order $(\log n)^2$. The reference g is comparable to its value g_B at the center of each cube B, and $g_B \ge n^{-\beta_0}$ for a fixed $\beta_0 < 1$. In the rescaled unit cube the ratio g_B/g has bounded C^α norm uniformly: on length ℓ, the score bound is paid by $\ell\sqrt{\log n} = 1$, and all higher potential derivatives are bounded with the corresponding scale factors. The top Hölder bound scales by ℓ^α.

Use a fixed orientation of compactly supported orthonormal tensor wavelets [7] of regularity C^{k+1}, supported strictly inside each cube. At physical length $r_i = \ell 2^{-j}$ their number in a cube is comparable to $(\ell/r_i)^4$, for j from a sufficiently large fixed index through $\lfloor \nu \log_2 n \rfloor$, with $\nu > 0$ small. Write ψ_i for these L^2-normalized wavelets. In particular $S_4 = \sum_i r_i^4 \asymp (\log n)^3$: each level contributes order the total volume. Set

$$b_i = \frac{\kappa^2 r_i^2}{n\sqrt{S_4}}, \qquad a_i = \sqrt{g_{B(i)} b_i}, \qquad q_i = a_i \psi_i.$$

Each wavelet has zero integral. At physical scale r, the relative perturbation in cube B has C^α norm at most $C(a_i/g_B)r^{-2-\alpha}$, by wavelet scaling, bounded overlap at the scale, and the rescaled reciprocal density bound. Summing over scales gives at most

$$C\kappa n^{-(1-\beta_0)/2} S_4^{-1/4} r_{\min}^{-1-\alpha} = o(1)$$

if $\nu < (1-\beta_0)/(2(1+\alpha))$. For fractional order this controls the global seminorm: on a pair of points in distinct cubes one may use the zero values of the top derivatives on the boundary of each cube and distances to the first exit points on the segment. Thus all perturbations stay within the reference margins by logarithmic composition.

The clean Gram matrix is $D_0 H D_0$ with $D_0 = \mathrm{diag}(\sqrt{b_i})$, and $H \preceq CI$. This uses orthonormality and the uniform bounds on g_B/g within each cube. The noisy Gram shrinks in positive operator order as before, so $n^2\|G\|_{\mathrm{HS}}^2 \le Cn^2 \sum_i b_i^2 = C\kappa^4$. For the metric test take $\varphi = \sum_i (r_i^2/g_{B(i)})\omega_i a_i \psi_i$. Its pairing is exactly $E = \sum_i r_i^2 b_i = \kappa^2\sqrt{S_4}/n$. The gradient squared norm in a cube is controlled by the sum of squared coefficient norms weighted by r_i^{-2}, up to a fixed factor. For clarity, within one level sum u_j, $\|\Delta u_j\|_2 \le C r_j^{-2}\|u_j\|_2$ by overlap. When $r_j \ge r_\ell$, integration by parts yields

$$|\langle \nabla u_j, \nabla u_\ell \rangle| \le C(r_\ell/r_j)(\|u_j\|_2/r_j)(\|u_\ell\|_2/r_\ell).$$

Summing the geometric ratios and multiplying by the comparable density in each cube gives squared gradient norm $O(E)$ under f_ω. The test Hessian is uniformly bounded, since its per cube scale sum is bounded by $C\kappa(ng_B)^{-1/2} S_4^{-1/4} r_{\min}^{-1}$. Thus (12) gives separation at least $c(\log n)^{3/4}/\sqrt{n}$. This lower bound holds under every known Gaussian covariance, by the same mixture contraction.

2.5.3 The inverse mixture

Rotate axes so $\Sigma = \mathrm{diag}(s, t_2, \dots, t_d)$ with $0 \le t_i \le s$, and suppose $0 < s \le s_0$ with $s_0 > 0$ small. Place $N \asymp s^{-d/2}$ centers z_i spaced $D\sqrt{s}$ apart in a fixed cube. Here D is sufficiently large, and s_0 small after choosing D. Put $T^2 = A_0 \log(2 + ns^{\alpha+d/4})/s$, and use

$$q_i(x) = a_0\psi_i(x), \qquad a_0 = \kappa T^{-\alpha}, \qquad \psi_i(x) = e^{-|x-z_i|^2/(2s)}\left(\cos(T(x_1 - z_{i,1})) - e^{-sT^2/2}\right).$$

Each packet has zero mass. The relative perturbation before the factor a_0 has derivatives through order $j \le k$ bounded by CT^j. Indeed derivatives of the packets are bounded by $CT^j e^{-c|x-z_i|^2/s}$ because $T\sqrt{s}$ is large, and division by g costs only Gaussian and polynomial factors absorbed in a weaker envelope for small s_0. Separated grid envelopes have uniformly bounded sum. For $0 < \theta < 1$, the k-th derivative has θ-Hölder seminorm at most CT^α. For shifts longer than $1/T$ use the sup norm. For a shorter shift expand the product-rule terms of $D^k(\psi_i/g)$. A difference of a smooth packet derivative of order j costs $CT^{j+1}|u|$ with the same weakened envelope. A difference of a reciprocal derivative of order $k - j$ costs $C|u|^\theta$ times polynomial factors and the maximum of the two reciprocal densities, by the product rules and the reference Hölder condition on $D^k V_g$. The nearby packet envelopes still absorb those factors since $|u| \le 1/T \le C\sqrt{s}$. Summing the bounds gives $CT^\alpha|u|^\theta$. The logarithmic perturbation thus costs $O(\kappa)$ in the stipulated norms and preserves the class for small enough κ.

Convolution with the noise changes the packet envelope to variances $s + t_i$, with $t_1 = s$, multiplying its amplitude by $\prod_i \sqrt{s/(s + t_i)}$. Its bracket becomes

$$e^{-sT^2/4}\cos\Big(\tfrac{T}{2}(y_1 - z_{i,1})\Big) - e^{-sT^2/2}.$$

These formulas follow by conditional Gaussian integration and include zero eigenvalues in the other coordinates. The convolved reference has a lower Gaussian envelope, obtained by convolving one for g. Combining the two narrow packet envelopes and dividing by this lower envelope gives

$$|G_{ij}| \le Ca_0^2 s^{d/2} e^{-csT^2} e^{-c|z_i - z_j|^2/s}.$$

Indeed the common center lies in a fixed cube, the variances of the two envelopes are between s and $2s$, and choosing s_0 small absorbs the reciprocal reference into a weaker Gaussian. Thus

$$n\|G\|_{\mathrm{HS}} \le Cna_0^2 s^{d/4} e^{-csT^2}.$$

With $X = ns^{\alpha+d/4}$, the right-hand side is at most $C\kappa^2 X(A_0 \log(2 + X))^{-\alpha}(2 + X)^{-cA_0}$. It is arbitrarily small by choosing A_0 large and κ small.

For transport separation take $\varphi = T^{-2}\sum_i \omega_i q_i$. The spatial L^2 Gram matrix of the ψ_i has diagonal at least $cs^{d/2}$ by integrating the squared cosine in its Gaussian envelope and taking $T\sqrt{s}$ large. The absolute off-diagonal row sums are at most $Cs^{d/2}\sum_{j\in\mathbb{Z}^d\setminus\{0\}} e^{-cD^2|j|^2}$, less than half the diagonal bound for large D. The pairing with $f_\omega - g$ is at least cE for $E = T^{-2}a_0^2 s^{d/2} N$. The gradient Gram has absolute row sums at most $CT^2 s^{d/2}$ by the same overlap estimate, so the test gradient has squared norm $O(E)$ under the uniformly bounded vertex densities. The pointwise derivative envelope bounds the test Hessian by Ca_0. Equation (12) gives $W_2(f_\omega, g) \ge cT^{-\alpha-1}$ for every sign.

For $s \ge s_0$, use instead one additive packet throughout, replacing its envelope variance by a fixed τ_0^2 smaller than the variance in the Gaussian lower envelope for g, with its cosine mean removed. Use amplitude $\kappa T^{-\alpha}$ and $T^2 = C\log n/s$. The derivative and Hölder argument above applies to this single packet. Its convolved Gaussian envelope squared, divided by the convolved reference lower

envelope, is integrable uniformly in Σ by the variance choice. Its first-coordinate cosine attenuates by $\exp(-\tau_0^2 s T^2/[2(\tau_0^2+s)])$, so the n-sample chi-square tends to zero uniformly for sufficiently large C. Testing the latent difference with $\cos(Tx_1)/T$ gives separation of order $T^{-\alpha-1}$. This matches the inverse term since its logarithm is comparable to $\log n$ on this range. Taking the maximum gives the lower bounds for $d \geq 4$. The transition statements in those dimensions follow by equating the two terms: at $d = 4$ the equation is $X/\log^{\alpha+1}(2+X) \asymp (\log n)^{3/2}$, with $X = ns^{\alpha+1}$.

2.6 Testing below dimension four

We now prove the assertions of Theorem 2.1 for $d = 1, 2, 3$.

2.6.1 A spatially varying coarse heat time

For small isotropic noise take

$$h = \max(a_{n,d}^{1/(\alpha+1)}, \sqrt{s/(\gamma \log n)}), \qquad r = h^{\alpha+1}, \qquad a = \log n.$$

Use the sublevel cutoff and screening above with this r and cube side $w = a^{-1/2}$. The same computations give $nar^2 \to \infty$ and $\lambda_{\min} \gtrsim nr^2 a^{2-d}$ for $\lambda(x) = nw^d e^{-\tau(x)}$. This last lower bound is at least of orders $a, \sqrt{\log a}, a^{-1/2}$ in dimensions one, two and three. The binomial bounds use only $O(a^d)$ cells, so null screening has vanishing error and violation of a relaxed population criterion is rejected with error at most $\varepsilon/4$. As before, screened laws obey $p_f(x)/g(x) \leq \mathcal{M}(x)$ on I, with the same tier formula in terms of $\lambda(x)$. In dimension one all cells are in the dense tier.

Put $b(x)^2 = (1+|x|^2)^{-1}$ and $v_x = b(x)^2 - s$. Then $v_x \asymp b(x)^2$ on I as $s \leq n^{-\eta}$. We use

$$\sup_I \mathcal{M} b^2 = O(1), \qquad \int_I \mathcal{M}^2 b^{4-d} \leq C \begin{cases} 1, & d = 1, \\ \log a, & d = 2, \\ a, & d = 3. \end{cases} \tag{13}$$

The constant part of the envelope gives the claimed integral by radial integration up to $O(\sqrt{a})$. All additional tiers lie at $\tau \geq a - O(\log a)$, where $b \asymp a^{-1/2}$. For $d = 2$ there is no sparse tier for large n. The middle tier has volume $O(\log a)$ by (5), and $\mathcal{M}^2 = O(\log a)$ by the lower bound on $\lambda_{\min}$. Its weighted contribution is $O((\log a)^2/a)$. For $d = 3$ the middle-tier volume is $O(a^{1/2} \log a)$ and $\mathcal{M}^2 = O(a^{1/2}(\log a)^2)$, so multiplying by $b = O(a^{-1/2})$ gives $o(a)$. The sparse-tier integral of $\lambda(x)^{-2}$ is $O(a^{3/2})$: integrate the exponentially growing factor along levels, using the maximum $O(a)$ and level-volume derivative $O(a^{1/2})$. Multiplying by b gives $O(a)$. These pointwise bounds on the outer tiers also prove the sup bound, and give $\mathcal{M}(x) \leq C(1+|x|)$ on I.

For $p_t = f * \gamma_t$, take the exact fields

$$J(f)(x) = \tfrac{1}{2} \int_0^{b(x)^2} \nabla p_t(x)\, dt, \qquad U(f)(x) = p_{b(x)^2}(x) + \tfrac{1}{2} \nabla b(x)^2 \cdot \nabla p_{b(x)^2}(x).$$

The spatial gradients of p_t are evaluated at fixed heat time. Differentiating the upper limit in the divergence gives $f = U(f) - \operatorname{div} J(f)$. The scalar kernel at row x is

$$G_x(u) = \gamma_{v_x}(u) + \tfrac{1}{2} \nabla b(x)^2 \cdot \nabla \gamma_{v_x}(u).$$

It is unbiased for $U(f)$, and is bounded by $Cb(x)^{-d} e^{-c|u|^2/b(x)^2}$ in absolute value, since $|\nabla b(x)^2| \leq Cb(x)^3$. For the vector field use $R_{v_x} - R_{h^2}$ together with the fine kernel (8) at the constant h, with

the same physical truncation. The pointwise Fourier calculation gives exactly the high-pass error $J_h(f)(x)$ before truncation; $b(x)$ is held fixed in this computation. The bias with normalization $g^{-1/2}$ is $O(r)$ by (2): on the outer tiers the Gaussian envelope pays the polynomial factors in $\mathcal{M}$ as above. The screened exterior also contributes $O(r)$ in distance. Use the same scalar weight $(Dg)^{-1/2}$, $D = 1 + |x|^2$, and vector weight $g^{-1/2}$ as before.

2.6.2 Covariance at a variable heat time

The score tilt at row x is $e^{C(1+|x|)|u|}$, so it can be absorbed into a Gaussian at bandwidth $b(x)$, since $b(x)(1+|x|)$ is bounded. For the scalar feature, the covariance kernel is bounded by the convolution of the resulting absolute Gaussian profiles times $\sqrt{\mathcal{M}(x)\mathcal{M}(x')/(D(x)D(x'))}$. For the coarse vector feature, the corresponding profiles are bounded by $\Phi_{b(x)}(u) = C|u|^{1-d}e^{-c|u|^2/b(x)^2}$, with envelope $\sqrt{\mathcal{M}(x)\mathcal{M}(x')}$. Here we dropped the small-scale regularization in the bound on $R_{v_x} - R_{h^2}$. Since $d < 4$, this profile has $L^{4/3}$ norm $O(b(x)^{1-d/4})$ and L^1 norm $O(b(x))$.

Split into pairs with $(1 + |x|)/(1 + |x'|)$ between $1/2$ and 2, and the remaining pairs. In the first region the bandwidths are comparable, so both profiles are dominated by their common-scale bounds at $b(x)$ up to constants. For the scalar profiles the squared L^2 norm of their convolution is $O(b(x)^{-d})$. Young's inequality with $L^{4/3}$ gives the vector bound $O(b(x)^{4-d})$ for the same squared norm. Use $2uv \le u^2 + v^2$ on the squared envelope product and integrate in either row. Both squared Hilbert–Schmidt bounds are $C\int_I \mathcal{M}(x)^2 b(x)^{4-d}\,dx$, using $D = b^{-2}$ for the scalar feature. Their operator norms are $O(1)$ by Schur, using $\sup \mathcal{M}b^2 = O(1)$ and L^1 convolution bounds $O(1)$ and $O(b^2)$ for the two profiles.

In the other region $|x-x'| \gtrsim 1+\max(|x|, |x'|)$. The Gaussian profiles, all of bounded bandwidth, give exponential decay in the separation. For the vector profiles, after extracting that decay their remaining singularities at the two centers are integrable and separated, giving bounded convolution: isolate disjoint neighborhoods of each center and use Gaussian decay on the complement. Scalar prefactors and envelopes grow at most polynomially in the two radii. Thus this region contributes $O(1)$ to both norms. Together with (13) the coarse bounds are

$$\|C_{\mathrm{coarse}}\|_{\mathrm{HS}} \le Cna_{n,d}^2, \qquad \|C_{\mathrm{coarse}}\|_{\mathrm{op}} \le C.$$

For the fine part the constant-bandwidth Schwartz profile argument above gives squared Hilbert–Schmidt norm at most $CE^4h^{4-d}\int_I \mathcal{M}^2$, and operator norm at most $CE^2h^2\|\mathcal{M}\|_\infty$. Both tend to zero: h is polynomially small, the envelope factors are polylogarithmic, and $E^2 \le Cn^{C\gamma}$ with γ sufficiently small. The combined covariance therefore obeys $\|C_f\|_{\mathrm{HS}}/n \le Cr^2$ and $\|C_f\|_{\mathrm{op}}/n \le Cr^2$. The quadratic test argument applies. The operator contribution to variance relative to the squared signal is bounded by $Cr^2/\|m\|^2$, so a sufficiently large separation constant pays it, including in dimension one where the operator part need not be lower order. The large-noise and covariance-completion arguments apply with the stated inverse logarithm.

2.6.3 Matching lower bounds

In dimension one a shift of the reference by δ has W_2 distance $|\delta|$. Its clean one-sample relative entropy to the reference is $O(\delta^2)$, by expanding the potential and using the zero mean score. Since noise contracts entropy, $\delta = c/\sqrt{n}$ gives the desired two-point lower bound.

For $d = 2, 3$, take disjoint annuli at dyadic radii R from a fixed constant to $c\sqrt{\log n}$, for small fixed $c > 0$. Pack cubes of side $\ell_i \asymp R^{-1}$ in each annulus, with order R^{2d} cubes at radius R. The reference values g_i at their centers satisfy $g_i \ge n^{-\beta_0}$, $\beta_0 < 1$, and $g \asymp g_i$ inside each cube. Use one

L^2-normalized smooth mean-zero bump ψ_i per cube, with

$$S_4 = \sum_i \ell_i^4, \qquad b_i = \frac{\kappa^2 \ell_i^2}{n\sqrt{S_4}}, \qquad q_i = \sqrt{g_i b_i}\,\psi_i.$$

The rescaled reciprocal bounds used for the critical construction apply here since $\ell_i(1+|x|) = O(1)$. All relative C^α perturbations tend to zero because $g_i \geq n^{-\beta_0}$ and $\ell_i \gtrsim (\log n)^{-1/2}$. The clean Gram matrix is diagonal with entries $O(b_i)$, and the noisy Gram contracts. Hence $n^2\|G\|_{\mathrm{HS}}^2 = O(\kappa^4)$. The test $\varphi = \sum_i (\ell_i^2/g_i)\omega_i q_i$ has pairing $\sum_i \ell_i^2 b_i$, squared gradient norm at most a constant times that pairing, and bounded Hessian by rescaling on the disjoint cells. Equation (12) gives squared separation of order $\sqrt{S_4}/n$. A ring contributes order R^{2d-4} to S_4; summing gives order $\log\log n$ for $d = 2$ and $\log n$ for $d = 3$.

For the inverse lower bound in all three dimensions, the single additive narrow packet with $T^2 = C\log n/s$ works throughout $0 < s \leq \Lambda$. Its attenuated cosine multiplier is bounded by e^{-csT^2} uniformly for the fixed narrow variance, and the noisy reference denominator is paid as above. It gives latent separation of order $(s/\log n)^{(\alpha+1)/2}$ with small total variation. Together with the direct lower bounds this completes Theorem 2.1.

2.7 An honest confidence diameter at a specified reference

Let $0 < \zeta < 1/4$. Call a random set with measurable membership events and diameter honest if it contains f with probability at least $1-\zeta$ uniformly over $f \in \mathcal{F}_\alpha$, including the reference. Its diameter is measured in W_2.

Corollary 2.2. *For each known interior reference g, the optimal local confidence diameter has order $r_n(s)$. More precisely, for any fixed $0 < \varepsilon < 1/2$ with $2\zeta + \varepsilon < 1$, every honest set has diameter at least $c_\varepsilon r_n(s)$ with probability at least $1 - 2\zeta - \varepsilon$ at g. There is an honest set whose diameter is at most $C_\zeta r_n(s)$ with probability at least $1-\zeta$ at g.*

Proof. For the lower bound use the finite testing alternatives at distance at least $c_\varepsilon r_n(s)$ and with mixture total variation at most ε to the null. The event that the set intersects that finite family has mixture probability at least $1-\zeta$ by honesty, hence null probability at least $1-\zeta-\varepsilon$. The set also contains g with null probability at least $1-\zeta$, proving the asserted diameter bound.

For the upper bound use the test of Theorem 2.1 with each error at most $\zeta/2$ at separation $Cr_n(s)$, taking the deterministic version of the test. On acceptance return the class intersected with the open ball of that radius about g. On rejection return the class intersected with a fixed-radius open W_2 ball about the point mass at the observed sample mean. These sets have measurable diameter, since on an open ball one can compute its intersection diameter using a fixed countable dense subset of the class. The latter ball covers every f with probability at least $1-\zeta/2$ for a large enough fixed radius, because $W_2^2(f, \delta_{\overline{Y}}) = \mathbb{E}_f|X - m_f|^2 + |\overline{Y} - m_f|^2$ and both terms have uniform expectation bounds. If f is near g, acceptance always covers it; if it is farther away, failure due to acceptance is at most $\zeta/2$ by the test. Add the fallback-ball failure probability to get honesty. At g the test accepts with probability at least $1-\zeta/2$, giving diameter at most $2Cr_n(s)$. □

In particular, honesty over one smoothness class imposes its testing order at every interior smooth reference, even if a point estimator over a smaller, smoother class is faster there.

2.8 Relation to direct and indirect testing

Testing at a smoother null against rougher smoothness classes in the bounded-density torus model is used by Deo and Randrianarisoa [8, Section 3.3] to characterize adaptive Wasserstein confidence

sets. Their direct W_2 testing lower bound has the polynomial exponent in the case $d > 4$, with an upper bound up to logarithms. Lee, Kim, and Park [9, Theorem 5] show the critical direct logarithm for compact bounded-density W_1 testing. Tang and Yang [13] develop multiscale testing for smooth discrepancies on compact domains. Related direct smooth Wasserstein tests are reported in [11]. In one dimension, Hur and Liang [10] study direct detection along a specified displacement path. In the present whole-space problem, the four-dimensional direct lower bound uses a spatial volume of order $(\log n)^2$ and order $\log n$ wavelet levels on that volume. The test upper bound accounts for both effects through the screened density envelope and Hilbert–Schmidt norm. The inverse logarithm follows from the signed-mixture testing scale, which is governed by a Gram matrix rather than by discrimination of individual packet signs.

3 Two-sample Wasserstein testing with unequal sizes and Gaussian noise

We determine the minimax separation for testing equality of two unknown strongly log-concave densities in quadratic Wasserstein distance from independent samples with known Gaussian contamination. The covariances may be distinct and singular, and the sample-size ratio is arbitrary. The sharp separation is the worse of the two individual testing rates, each evaluated at its own sample size and noise covariance. An inward log-density extrapolation supplies the common weights in the tails, and Gaussian inverses preserve a common latent feature mean despite unequal observation kernels. When one group has sufficiently large noise, a two-sided transport comparison lets a fitted density take the place of its noisy features.

3.1 Experiment and separation

Fix $d \geq 1$, real $\alpha = k+\theta \geq 2$, integer k, $0 \leq \theta < 1$, and constants $0 < m < M < \infty$. Let $\mathcal{F}_\alpha$ be the translated class of 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 $\theta > 0$, impose a fixed positive bound on the global θ-Hölder seminorm of D^kV. We observe independent samples of sizes n_1, n_2 from $p_f = f * \gamma_{\Sigma_1}$ and $p_g = g * \gamma_{\Sigma_2}$. The two independent Gaussian errors have known covariances $\Sigma_i \succeq 0$ of operator norm at most Λ. Here $\gamma_\Sigma = N(0, \Sigma)$ and $\gamma_s = N(0, sI_d)$. Let $s_i = \|\Sigma_i\|_{\mathrm{op}}$. For a generic pair n, s, define

$$a_{n,d} = \begin{cases} n^{-1/2}, & d = 1, \\ (\log\log n)^{1/4}/\sqrt{n}, & d = 2, \\ (\log n)^{1/4}/\sqrt{n}, & d = 3, \\ (\log n)^{3/4}/\sqrt{n}, & d = 4, \\ n^{-2(\alpha+1)/(4\alpha+d)}, & d > 4. \end{cases} \qquad B_{n,d}(s) = \begin{cases} \log n, & d < 4, \\ \log(2 + ns^{\alpha+d/4}), & d \geq 4. \end{cases}$$

The risk of a test is the sum of its worst type-I error over the composite null $f = g \in \mathcal{F}_\alpha$ and worst type-II error over $f, g \in \mathcal{F}_\alpha$ with $W_2(f, g) \geq r$.

Theorem 3.1. *For arbitrary n_1, n_2 and known Σ_1, Σ_2, the minimax separation at any fixed error budget is of order*

$$r_* = \max_{i=1,2} \left[a_{n_i,d} + (s_i / B_{n_i,d}(s_i))^{(\alpha+1)/2} \right].$$

For each fixed $0 < \varepsilon < 1/2$ there are $c_\varepsilon, C_\varepsilon > 0$ such that the infimum risk at $c_\varepsilon r_*$ is at least $1 - \varepsilon$, and at $C_\varepsilon r_*$ it is at most ε, uniformly as $\min(n_1, n_2)$ tends to infinity.

Fixed-reference latent Wasserstein testing for fixed positive Gaussian noise is treated in [12, Theorem 2.4]. Direct two-sample Wasserstein tests for general metric spaces and compact cubes are studied in [9]; compact smooth adversarial losses are studied in [13], and sliced Wasserstein testing in [15]. Here both reference shape and tails are unknown. The pointwise estimates below use bounded curvature to normalize a composite-null test without imposing a global density floor.

3.2 Density bounds and a weighted transport comparison

Uniformly over the class, f has Gaussian upper and lower envelopes about its mean m_f, the score of f is $O(1+|x-m_f|)$, and

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

Indeed quadratic bounds at the mode give the envelopes and bound the distance between the mode and the mean. Differentiating e^{-V} gives the derivative bounds. For $\theta > 0$, on $|x-m_f| \le L$ and $|u| \le (2+L)^{-1}$, the same product rule and the Hölder bound on $D^k V$ give

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

The local density ratio is bounded by the score, derivatives of lower order in the product are Lipschitz, and the highest one uses the assumed Hölder seminorm.

For $p_f = f * \gamma_s$, isotropic $0 \le s \le \Lambda$, Gaussian integration of the two quadratic Taylor bounds gives

$$cf(x) \le p_f(x) \le f(x)e^{Cs(1+|x-m_f|^2)}. \tag{14}$$

The potential of p_f also has Hessian in a fixed positive interval. For $s > 0$, its Hessian is $s^{-1}I - s^{-2}\operatorname{Cov}(X \mid Y)$. The conditional potential has Hessian between $(m+s^{-1})I$ and $(M+s^{-1})I$. The upper covariance bound follows from Brascamp–Lieb [5]. The lower bound follows, in each unit direction v, from Cauchy–Schwarz and the conditional identities

$$\mathbb{E}[(v\cdot(X-\mathbb{E}X))(v\cdot\nabla A(X))] = 1, \qquad \mathbb{E}[(v\cdot\nabla A(X))^2] = \mathbb{E}[v^T D^2 A(X) v],$$

where A is the conditional potential. Consequently the scores of p_f are $O(1+|x-m_f|)$. The observed law always has a uniformly Gaussian upper envelope about m_f.

Use $\widehat{u}(\xi) = \int e^{-ix\cdot\xi}u(x)\,dx$. Let $\widehat{K}$ be a real even smooth Fourier cutoff that equals one near zero and is supported in the unit ball. Use $K_b = b^{-d}K(\cdot/b)$ for its spatial kernel, and define constant-bandwidth fields

$$U_b(f) = K_b * f, \qquad \widehat{J_b(f)}(\xi) = -\frac{i\xi}{|\xi|^2}(1-\widehat{K}(b\xi))\widehat{f}(\xi).$$

Then $f = U_b(f) + \operatorname{div} J_b(f)$. The inverse transform Q of $-i\xi(1-\widehat{K}(\xi))/|\xi|^2$ is integrable with all absolute moments finite and all signed moments zero. In fact a dyadic partition into shells 2^j away from the origin gives inverse transforms bounded by $C_N 2^{(d-1)j}(1+2^j|x|)^{-N}$. Summability of their L^1 norms $O(2^{-j})$ gives the assertion, with moment cancellation since the symbol vanishes near zero. For $|x-m_f| \le L$, $w = (2+L)^{-1}$, and $h \le w$, subtract the local Taylor polynomial of f to degree $k-1$ in the integer case and k otherwise. This gives

$$|J_h(f)(x)| \le Ch^{\alpha+1}(1+|x-m_f|^{k+1})f(x) + C_N h(h/w)^N. \tag{15}$$

Indeed $J_h(f) = hQ_h * f$, the short-shift remainder has order α, and the farther shifts use uniform absolute derivative bounds and the arbitrary integrated tails.

Choose the origin within a fixed distance of the means of f, g, and put $\rho = f + g$. For a test φ, set $u = \varphi - \frac{1}{2}\int \varphi\rho$. Then

$$\int (1+|x|^2)|u|^2\rho \le C\int |\nabla\varphi|^2\rho. \tag{16}$$

Both components satisfy ordinary Poincaré by Brascamp–Lieb. Comparing their component means to a common Lebesgue average on a fixed ball with a positive density floor gives the mixture inequality. For either potential V, $x\cdot\nabla V(x) \ge m|x|^2/2 - C$, since its mode stays close to zero. Integration by parts gives

$$\int u^2 x\cdot\nabla V f = d\int u^2 f + 2\int ux\cdot\nabla\varphi f.$$

Use Cauchy–Schwarz to absorb half of the resulting quadratic term, add the two estimates, and apply mixture Poincaré. Gaussian tails justify integration by parts for compactly supported tests minus their mean.

Let $I_0 \subset I$ admit a C^1 cutoff supported in I, equal to one on I_0, with $|\nabla\chi| \le C(1+|x|)$. For any identity $f - g = U + \operatorname{div} J$ in distributions,

$$W_2(f,g) \le C\left[\left(\int_{I_0^c} \frac{\rho}{1+|x|^2}\right)^{1/2} + \left(\int_I \frac{|U|^2/(1+|x|^2)+|J|^2}{\rho}\right)^{1/2}\right]. \tag{17}$$

To see this, split $\int (f-g)u$ with the cutoff and use Cauchy–Schwarz with (16) in the exterior. In the interior insert the field identity and integrate by parts. Approximation by smooth tests justifies this for a C^1 cutoff. The $L^2(\rho)$ norms of $(1+|x|^2)^{1/2}\chi u$ and $\nabla(\chi u)$ are at most $C\|\nabla\varphi\|_{L^2(\rho)}$. Riesz representation produces an exact flux of divergence $f-g$ whose squared action norm $\int |J|^2/\rho$ is bounded by the square of the displayed right-hand side. Interpolate between the densities with $(1-\lambda)f + \lambda g$, $\lambda(t) = \sin^2(\pi t/2)$, and flux $\lambda'(t)J$. Its action is at most $\pi^2 \int |J|^2/\rho$, by the lower bound $\min(\lambda, 1-\lambda)\rho$. The dynamic transport inequality [4] gives (17).

3.2.1 The reverse transport comparison and the heat fields

Use $D(x) = 1 + |x|^2$ and suppose the origin is within fixed distance of the two means. Let

$$\mathcal{H}(f,g) = \sup_{\int |\nabla\varphi|^2 \rho \le 1} \left| \int \varphi (f-g) \right|, \qquad \rho = f + g,$$

where the supremum is over smooth compactly supported tests. The Riesz and dynamic argument used for (17) gives $W_2(f,g) \le C\mathcal{H}(f,g)$. We prove a converse for the endpoint mixture weight. For related general comparisons using intermediate or bounded densities, see Peyre [14].

Set $\ell(x) = \epsilon/(1+|x|)$ for fixed small $\epsilon > 0$. For $|x-y| \le c\min(\ell(x), \ell(y))$, the Euclidean ball-average inequality gives

$$|\varphi(x) - \varphi(y)| \le C|x-y|\big(M_{\mathrm{loc}}|\nabla\varphi|(x) + M_{\mathrm{loc}}|\nabla\varphi|(y)\big),$$

where M_{loc} takes centered ball averages with radii at most $C_0\ell$. Indeed telescope ball averages from radius $|x-y|$ to zero at both endpoints using the L^1 ball Poincaré inequality, and compare the two initial averages on a common enlarged ball. The localized maximal operator is bounded on $L^2(\rho)$. To see this, cover with fixed enlargements of a Vitali selection of pairwise disjoint balls of radii $c_0\ell(x)$. The enlarged balls have bounded overlap by radius comparability when they meet and disjointness of the original balls. Choose ϵ small for the enlargement factors. On each enlarged ball all the relevant radii are comparable and ρ is comparable to its center value by the score bound. Each local maximal function is dominated there by the usual maximal function of the input restricted to a further fixed enlargement, and the usual L^2 bound and overlap yield the assertion.

For the other pairs, both $\ell(x)$ and $\ell(y)$ are bounded by $C|x-y|$, since ℓ is Lipschitz and the smaller radius is at most $|x-y|/c$. Subtract the mixture mean of φ, writing the result as u. Then

$$|\varphi(x) - \varphi(y)| \le C|x-y|\big(\sqrt{D(x)}|u(x)| + \sqrt{D(y)}|u(y)|\big).$$

Integrate both inequalities against an optimal coupling and use Cauchy–Schwarz, the local maximal bound and (16). This proves $\mathcal{H}(f,g) \le CW_2(f,g)$. In particular there is a vector flux L with $\operatorname{div} L = f - g$ and $\|L/\sqrt{\rho}\|_2 \le CW_2(f,g)$ by Riesz representation. The flux is integrable by weighted Cauchy–Schwarz.

Take

$$b(x)^2 = \delta/D(x), \qquad R_t = \tfrac{1}{2}\int_0^t \nabla\gamma_z \, dz,$$

with a sufficiently small fixed $\delta > 0$ in every dimension. For $q = f - g$ let $p_t = q * \gamma_t$, and define

$$U(q)(x) = p_{b(x)^2}(x) + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla p_{b(x)^2}(x), \qquad J(q)(x) = (R_{b(x)^2} * q)(x).$$

Gradients on p_t are at fixed t. Direct differentiation gives $q = U(q) - \operatorname{div} J(q)$. We claim

$$\left\| U(q)/\sqrt{D\rho} \right\|_2 + \|J(q)/\sqrt{\rho}\|_2 \le CW_2(f,g). \tag{18}$$

To prove (18), apply the row kernels to $\operatorname{div} L$ and write $L = \sqrt{\rho}\, H$. The conjugation factor is $\sqrt{\rho(y)/\rho(x)}$ at $z = x - y$. Globally

$$|\log \rho(y) - \log \rho(x)| \le C\big((1+|x|)|z| + |z|^2\big),$$

by the mixture score bound along the segment. With δ small this factor can be absorbed into every row Gaussian. The scalar kernel is bounded by $Cb(x)^{-d} e^{-c|z|^2/b(x)^2}$. Here $D^{-1/2}$ pays the first factor b^{-1} from integration by parts, and $|\nabla b^2| \le C_\delta b^3$ pays the derivative correction. Its row and column integrals are uniformly bounded. For the column integral split at $|x-y| \le c_1(1+|y|)$ for fixed small c_1, where $b(x) \asymp b(y)$. On the other part $|x-y|$ controls both $1+|x|$ and $1+|y|$, and the Gaussian pays every polynomial prefactor. Schur's inequality bounds the scalar norm by $C\|H\|_2$.

For the vector term first integrate from $\varepsilon b(x)^2$ to $b(x)^2$, $0 < \varepsilon < 1$, and integrate by parts against L at fixed row. Without conjugation the operator consists of two Gaussian averages, at these two row times, of a fixed matrix of Riesz products applied to H. This follows from the multiplier $\xi_i\xi_j/|\xi|^2$ times the difference of the Gaussian symbols. Its L^2 norm is bounded independently of ε, by the L^2 Fourier bound for the matrix and the maximal bound for Gaussian averaging. Off the kernel diagonal, the absolute derivative of the truncated R-kernel is at most $C|z|^{-d}\exp(-c|z|^2/b(x)^2)$, by integrating Gaussian second derivatives over the heat times. The error from conjugation has an extra factor $\sqrt{\rho(y)/\rho(x)} - 1$. By the log-ratio bound this yields an absolute kernel bounded by

$$Cb(x)^{-1}|z|^{1-d}\exp(-c'|z|^2/b(x)^2).$$

The same column split and direct row integration show a uniform Schur bound. The truncated convolutions tend pointwise to $J(q)$: q is uniformly bounded and $\|R_{\varepsilon b(x)^2}\|_1 \le C\sqrt{\varepsilon} b(x)$. Fatou's lemma gives the required weighted vector bound, proving (18).

3.3 Effective scale and large-noise groups

Let m_* be an integer with $a_{m_*,d} \asymp r_*$ and $m_* \le \min_i n_i$. Such an integer exists for all sufficiently large samples: the clean rate decreases eventually, consecutive values have ratio tending to one, and $r_* \ge a_{\min_i n_i, d}$. Also $m_* \to \infty$ uniformly, since the individual rates tend to zero uniformly for bounded noise. Use a sufficiently small fixed $\eta > 0$. Call a group large-noise when $s_i > m_*^{-\eta}$ and small-noise otherwise. Complete each observation to its own scalar covariance $s_i I$, by adding independent Gaussian noise of covariance $s_i I - \Sigma_i$.

A large-noise group admits a class-valued estimate $\widehat{f}_i$ with W_2 error $O_{\mathbb{P}}(r_*)$ uniformly. We show this directly. For a generic sample of size n_i, write $s = s_i$ and let e_i be its empirical error in observed W_1. The uniform second moment about the population mean gives $\mathbb{E}e_i = O(n_i^{-\beta})$, $\beta = 1/[4(d+1)]$. Truncate to radius n_i^{β} about the mean, quantize at mesh $n_i^{-\beta}$, and compare histogram weights by Cauchy–Schwarz and their binomial variances. The first two costs are $O(n_i^{-\beta})$, and the final cost is at most $Cn_i^{\beta+d\beta-1/2} = Cn_i^{-1/4}$. Fit in observed W_1 among a countable dense set of latent mean-zero shapes translated to the sample mean, choosing the first within n_i^{-1} of the infimum. Its observed W_1 discrepancy ϵ_i from the truth is at most $3e_i + n_i^{-1}$.

To convert this discrepancy, note that the mean gap is at most ϵ_i. When $\epsilon_i \le 1$, use the fields U_h, J_h from Section 2 in the transport comparison (17), on a ball of radius $D_0\sqrt{\log(1/h)}$ about one mean. The exterior and the weighted vector term are $O(h^{\alpha+1})$ in distance by Gaussian tails and (15). The scalar term uses a density floor h^P on this ball and an unweighted sup norm at most $C\epsilon_i h^{-d-1}e^{s/(2h^2)}$: the observed characteristic functions differ by at most $\epsilon_i|\xi|$ by coupling, and one inverts the Gaussian only on the Fourier cutoff. Thus

$$W_2(f_i, \widehat{f}_i) \le C\left(h^{\alpha+1} + \epsilon_i h^{-d-1-P/2}\log^{d/4}(1/h)e^{s/(2h^2)}\right).$$

Here $s > m_*^{-\eta} \ge n_i^{-\eta}$. On $e_i \le n_i^{-\beta/2}$, choose $h = \sqrt{s/(\gamma_0 \log n_i)}$ for small γ_0 and then small η, so the second term is $O(h^{\alpha+1})$. The event has uniform probability $1 - o(1)$. The effective testing logarithm for that group is comparable to $\log n_i$ since η is small. This gives the stated rate for the fitted density.

When both groups are large-noise, compare the two fits in W_2 with a sufficiently large fixed multiple of r_* as threshold. The triangle inequality and the preceding bounds give the desired errors. In a mixed experiment the large-noise group will be represented by its fitted density using validation on the other group.

3.4 Estimating the common weights from inward probes

Here at least one group is small-noise. Denote the pair of latent densities under comparison by f, g, taking the independent large-noise fit in place of that group's true density in a mixed experiment. Set

$$a = \log m_*, \quad w = a^{-1/2}, \quad h_0 = L_0 r_*^{1/(\alpha+1)}, \quad r = h_0^{\alpha+1}, \quad \rho_0 = r^2 a^{2-d/2}$$

with L_0 a sufficiently large fixed constant. Both h_0 and r lie between fixed negative powers of m_*. For each small-noise group split into three parts of sizes comparable to its own sample size. Use the first parts to estimate the two means. In a mixed experiment use the fitted density's mean for its group. Reject when the means differ by more than one. Otherwise take the first small-noise estimated mean as origin. On an event of probability $1 - o(1)$ the individual sample mean errors are smaller than $1/4$. On this event whenever the mean test passes the two means under comparison lie within a fixed distance of the origin. Our subsequent bounds are conditional on these first splits and on the fitted candidate when there is one.

Write $\rho = f + g$. Tile with w-cubes, retaining all cubes meeting the ball of radius $L = D_0\sqrt{a}$. There are $O(a^d)$ cubes. Write x for a cube center. For each small-noise observed density p separately, estimate $p(x)$ at inner rows $|x| \le c_1\sqrt{a}$ using its second sample count in the cube divided by the second sample size times w^d. For outer rows put $\omega = x/|x|$, $t = C_1 \log a/\sqrt{a}$, and count in the w-cubes centered at $x - t\omega$ and $x - 2t\omega$. Each count of at least $A_1 \log a$ yields a density estimate at its probe by the same division. If both counts attain this threshold, set the estimate at x to the square of the first probe estimate divided by the second. Otherwise set it to zero. The small-noise densities use their own second splits. For a fitted candidate, use its exact latent density at each cube center.

Uniformly with probability $1 - o(1)$, every count above the threshold agrees with its mean up to fixed factors and every count of mean at least $2A_1 \log a$ is above the threshold. This follows from binomial Chernoff bounds and a union bound over $O(a^d)$ probes with A_1 large. A probe cube average agrees with its center value up to fixed factors because the observed score on the grid region is $O(\sqrt{a})$. The inner counts have polynomially large means, by the uniform Gaussian lower envelope with c_1 sufficiently small.

On the outer segments, the derivative of $-\log p$ in direction ω is at least $c\sqrt{a}$. Indeed the Hessian is bounded below and the mode is within a fixed distance of the origin. Each inward probe value is therefore at least $a^{cC_1} p(x)$. Since

$$m_* w^d \rho_0 = m_* r^2 a^{2-d} \gtrsim a^{-1/2},$$

every outer center with $p(x) \ge \rho_0$ has two reliable probes if C_1 is large. The Hessian upper bound gives

$$\big|\log p(x) - 2\log p(x - t\omega) + \log p(x - 2t\omega)\big| \le Ct^2 = o(1).$$

Thus every positive estimate $\widehat{p}(x)$ is comparable to $p(x)$, and a zero estimate implies $p(x) < \rho_0$. No independence of counts within a sample is needed.

Let I be the union of the retained cubes for which the sum of the two estimates at the center is above $K_1\rho_0$, with K_1 a large fixed constant, and let $\widehat{\rho}$ be that sum extended constantly on each selected cube. Throughout the successful pilot event,

$$c\rho(x) \le \widehat{\rho}(x) \le C\rho(x) \quad (x \in I), \qquad \{\rho > C_2\rho_0\} \subseteq I. \tag{19}$$

Indeed the ignored terms at a selected center can contribute at most $2\rho_0$ in observed density. Center values and cube values are comparable by the score bounds, and the latent and observed values are comparable there by (14) and $s_i a = o(1)$ for each small-noise group. Every center with large enough density has a positive estimate of the same order. Choose D_0 large so that beyond the tiled ball the Gaussian upper bound is below ρ_0. This establishes the inclusion.

For the analysis take a smooth cutoff as a scalar function of ρ/ρ_0, zero below C_2 and one above $2C_2$. Its support is in I (increase its lower threshold if needed) and its gradient is bounded by $C(1 + |x|)$. For each component density $u = f, g$ integrate on its potential levels about its mode. On a level z above the minimum, the radius and radial derivative are comparable to $\sqrt{z}$. The level-volume derivative is $O(z^{d/2-1})$ and $1 + |x|^2 \gtrsim z$ on large levels. Since $\log(1/\rho_0) \asymp a$, this proves

$$\int_{\{\rho \le 2C_2\rho_0\}} \frac{\rho(x)}{1 + |x|^2}\, dx \le C\rho_0 a^{d/2-2} = Cr^2.$$

Here the domain is contained in the individual low-density domain for each component. Only the analytical cutoff uses the unknown densities.

3.5 Matched latent fields from different observation kernels

Use $b(x)^2 = \delta/D(x)$ as in (18), and let $v_{i,x} = b(x)^2 - s_i \asymp b(x)^2$ on I for each small-noise group i. Use $h(x) = h_0$ if $d \le 4$, and $h(x) = h_0 e^{c_0|x|^2}$ if $d > 4$. Choose $c_0 > 0$ sufficiently small so that h remains polynomially small on I and so that $e^{(2\alpha+2)c_0|x|^2}$ has uniformly bounded polynomial moments under the two latent laws. Then $h(x)^2 \ll v_{i,x}$ and $h(x) \asymp h(x')$ for $|x - x'| \le 2w$.

For an observed y use scalar and vector kernels at row x applied at $x - y$. The scalar kernel is

$$G_x = \gamma_{v_{i,x}} + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla \gamma_{v_{i,x}},$$

divided by $(D(x)\widehat{\rho}(x))^{1/2}$. The vector kernel is the sum of $R_{v_{i,x}} - R_{h(x)^2}$ and the kernel with symbol

$$\frac{i\xi}{|\xi|^2}\left(\widehat{K}(h(x)\xi)e^{s_i|\xi|^2/2} - e^{-h(x)^2|\xi|^2/2}\right). \tag{20}$$

Use the fine kernel on the whole space and divide the vector sum by $\widehat{\rho}(x)^{1/2}$. Write $F_{i,y}$ for the resulting feature in $L^2(I) \oplus L^2(I; \mathbb{R}^d)$.

Conditional on the pilot, the scalar expectation is the normalized $U(f_i)$. Holding bandwidths fixed in each row, the vector mean, before normalization, has the latent multiplier $i\xi(\widehat{K}(h(x)\xi) - e^{-b(x)^2|\xi|^2/2})/|\xi|^2$. Indeed the coarse kernel multiplier supplies the Gaussian part of the expression, and the fine kernel cancels both the coarse regularization at $h(x)$ and the observation noise. Thus the two means are evaluations on their latent densities of the same fields, independent of s_i. In the mixed experiment evaluate this latent field expectation exactly on the fit. The vector mean differs from the normalized $J(f_i)$ by the high-pass term (15), which has weighted norm $O(r)$ by the Gaussian moment choice of c_0, the density floor and $f_i^2/\widehat{\rho} \le Cf_i$.

For a row of the fine kernel, rescaling gives $h^{1-d}L_{A_i(x)}((x-y)/h)$, where $A_i(x) = s_i/h^2$. The scaled symbol is smooth at zero because the bracket is radial analytic there and vanishes to second order. Every finite collection of Schwartz seminorms of $L_{A_i(x)}$ has bound $C(1 + A_i(x))^C e^{CA_i(x)}$. With μ the difference of the mean features, (19), the tail cutoff, the bias bound, (17) and (18) give

$$r + W_2(f, g) \asymp r + \|\mu\|. \tag{21}$$

When both groups are small-noise, their means agree exactly at the null conditional on any pilot. The untruncated fine kernel is important for this conclusion when the noise scales differ.

3.5.1 Conditional covariance

We work throughout on the successful pilot event. It suffices to bound the Hilbert–Schmidt norms of the four uncentered second moment operators, treating the fine vector as the sum of its near part $|x - y| \le w$ and its exterior part. Their sum, multiplied by four, dominates the centered covariance in positive operator order by Cauchy–Schwarz, and the Hilbert–Schmidt norm is monotone on positive operators. Fix either completed input law with density p. Its supporting log-density tangent at each row and (19) bound

$$\frac{p(y)}{\sqrt{\widehat{\rho}(x)\widehat{\rho}(x')}} \le C\exp(C(1 + |x|)|y - x| + C(1 + |x'|)|y - x'|).$$

Each factor is absorbed in the row Gaussian, since $b(x)(1 + |x|) \le C$. It is bounded on the near fine support.

In absolute value the tilted scalar kernel before division by $\sqrt{D(x)}$ is bounded by

$$Cb(x)^{-d}e^{-c|x-y|^2/b(x)^2}.$$

This includes the derivative correction since $|\nabla b(x)^2| \le Cb(x)^3$ in the variable case. The tilted coarse vector has absolute kernel bounded by

$$C(h(x) + |x - y|)^{1-d}e^{-c|x-y|^2/b(x)^2}.$$

For $|x - y| \le h(x)$ integrate the time powers above $h(x)^2$. At larger distances, keep half the Gaussian at scale $b(x)$ and extend the remaining time integral to infinity.

For $d < 4$, at pairs with $1 + |x| \asymp 1 + |x'|$ (ratio between $1/2$ and 2), use common bandwidth $b(x)$. The scalar covariance kernel has squared row integral bounded by $Cb(x)^{-d}D(x)^{-2}$. For the coarse vector use $C|z|^{1-d}e^{-c|z|^2/b(x)^2}$, whose $L^{4/3}$ norm is $O(b(x)^{1-d/4})$; Young's inequality gives the same row bound $Cb(x)^{4-d}$. For noncomparable pairs, the separation is comparable to the larger radius. Extract a Gaussian in separation using both row Gaussians with bounded bandwidth. Scalar prefactors then grow at most polynomially in the radii. For the vector, split into balls of radius one third of the separation about either row and their complement. On each ball the other singularity is bounded by a polynomial in separation and the singularity at the ball center is integrable against its Gaussian. On the complement both singularities are bounded. The two squared kernels are thus integrable on this pair region uniformly in n. Integrating the comparable row bound $Cb(x)^{4-d}$ on I gives $O(1), O(\log a), O(a)$ in dimensions one, two, three respectively.

For $d \geq 4$, the scalar argument on comparable and noncomparable radii above again gives squared norm bounded by $C(1 + \int_I b(x)^{4-d}\,dx)$. This is $O(a^2)$ in dimension four and polylogarithmic above four. For the coarse vector, the convolution bound at $z = x - x'$, $|z| \leq 2w$, is

$$C(h(x) + |z|)^{2-d}e^{-c|z|^2}.$$

The two regularizations are comparable there. For $|z| \leq h(x)$ apply Cauchy–Schwarz to the two profiles. Above that scale, split near either singular center and in the remainder to obtain $C|z|^{2-d}$ before the Gaussian: the two radial integrals are bounded by $|z|^{1-d}\int_0^{|z|} dt$ and $\int_{|z|}^\infty t^{1-d}\,dt$. The same unregularized bound applies at $|z| > 2w$. In dimension four the squared near integral in z is $O(a)$, and the far contribution per row is $O(1+\log(1/w))$. These give $O(a^3)$ after integration in x. For $d > 4$ the squared near bound is $C\int_I h(x)^{4-d}\,dx \leq Ch_0^{4-d}$. The far bound is polylogarithmic and is absorbed by h_0^{4-d}.

For the near fine operator, put $A_i = s_i/h_0^2$ and write $E_i = C(1 + A_i)^C e^{CA_i}$ for the finite collection of seminorms used below. Its kernel vanishes at $|x - x'| > 2w$. For the other pairs the bandwidths are comparable, and convolution of the scaled decaying profiles gives the kernel bound

$$CE_i^2 h(x)^{2-d}(1 + |x - x'|/h(x))^{-N_0}$$

for sufficiently large fixed N_0. This uses the score tilts only on the near supports. Therefore the Hilbert–Schmidt norm is at most $CE_i^2(\int_I h(x)^{4-d}\,dx)^{1/2}$.

For the far fine operator we use its trace. The observation density is uniformly bounded on the whole space, and the inverse pilot weight is at most $C/\rho_0 \leq m_*^P$ on I. The scaled L^2 tail of the fine profile at radius $w/h(x)$ is bounded by $CE_i^2(h(x)/w)^N$ for arbitrarily large fixed N, upon including the corresponding seminorm in E_i. Multiplying by the row scale $h(x)^{2-d}$ and integrating over I gives trace bound $CE_i^2 m_*^{-K}$ for any prescribed fixed K, by the polynomial upper and lower bounds on $h(x)$. This bound applies without imposing a log-m_* bound on A_i. It estimates only covariance, while the full fine kernel supplies the exact mean formula.

Put $X_i = n_i$ if $d < 4$, and $X_i = 2 + n_i s_i^{\alpha+d/4}$ otherwise. By the definition of r_*, $A_i \leq L_0^{-2}\log X_i$ and hence $E_i^2 \leq CX_i^c$ for a fixed $c > 0$ as small as required upon taking L_0 large. Choose the finite tail order and the near seminorms before choosing L_0. Since $X_i \leq Cn_i$, the far bound divided by n_i is $O(m_*^{-K}n_i^{c-1})$.

For $d < 4$ the near bound divided by n_i is at most $Cm_*^{c-1}h_0^{2-d/2}a^{d/4} = o(m_*^{-1})$ for small c. In dimension four it is $CE_i^2 a/n_i$. If $X_i \leq a^2$, take $c < 1/4$ to pay it by $Ca^{3/2}/m_*$. Otherwise $a \leq \sqrt{X_i}$ and

$$E_i^2 a/n_i \leq C\frac{X_i - 2}{n_i \log^{\alpha+1} X_i} = C(s_i/\log X_i)^{\alpha+1} \leq Cr^2.$$

For $d > 4$, the near bound divided by n_i is at most $CE_i^2 h_0^{2-d/2}/n_i$. Indeed $\int_I h(x)^{4-d}\,dx \leq Ch_0^{4-d}$. The inequalities $h_0 \geq cn_i^{-2/(4\alpha+d)}$ and $h_0 \geq L_0\sqrt{s_i/\log X_i}$ show that

$$n_i h_0^{2\alpha+d/2} \geq c' \max(1, (X_i - 2)/(\log X_i)^{\alpha+d/4}) \geq c'' E_i^2.$$

Thus the near bound is at most Cr^2. The scalar and coarse vector estimates are paid by the clean testing floor using $n_i \geq m_*$ and $a_{m_*,d} \asymp r$. In every dimension, conditional on the pilot event, this proves

$$n_i^{-1}\|C_i\|_{\mathrm{HS}} \leq Cr^2 \tag{22}$$

for the covariance C_i of each small-noise group's feature.

3.6 The unequal-sample statistic

When both groups are small-noise, use their independent third splits of sizes $N_i \asymp n_i$. Let $\overline{F}_i$ be their feature means and U_i the average inner product of features over ordered distinct pairs in group i. Define

$$Q = U_1 + U_2 - 2\langle \overline{F}_1, \overline{F}_2 \rangle.$$

Conditional on the pilot its mean is $\|\mu\|^2$. Writing each empirical mean as its expectation plus its centered deviation, the linear contribution to the variance is

$$4 \sum_{i=1}^{2} N_i^{-1} \langle \mu, C_i \mu \rangle.$$

The two within-group degenerate terms and the centered cross term are uncorrelated with the linear term and with each other. Their summed variance is

$$2 \sum_{i=1}^{2} \frac{\|C_i\|_{\mathrm{HS}}^2}{N_i(N_i - 1)} + \frac{4 \operatorname{tr}(C_1 C_2)}{N_1 N_2}.$$

Bound the trace by the product of Hilbert–Schmidt norms. In a mixed experiment subtract the known fitted mean from every small-group feature and take the paired inner product average within that group. The single-group variance formula omits the other within-group term and the cross term. In both cases

$$\operatorname{Var}(Q) \le C(r^4 + r^2 \|\mu\|^2)$$

on the pilot event.

Reject when $Q > A_\varepsilon r^2$, in addition to the preliminary mean rule. The mean rule has $o(1)$ null error by concentration, and in the mixed null also by the fitted density error $O_{\mathbb{P}}(r)$. With both groups small-noise $\mu = 0$ exactly under the null. For a mixed null, the fitted estimate is within Cr of the common latent law with probability $1 - o(1)$ by the large-noise bound above, so $\|\mu\| \le C' r$ by (21). In either case the conditional variance and Chebyshev's inequality give the required error with a sufficiently large A_ε. For alternatives the fitted estimate changes the latent separation by at most Cr on the same large-noise estimation event. When the latent separation is larger than a sufficiently large fixed multiple of r, the mean rule either rejects or (21) gives a sufficiently large $\|\mu\|/r$ on the good event. Chebyshev then gives the required power. The additions used for own-group Gaussian completion define an admissible randomized test on the original observations by averaging over them. A deterministic test can compare that conditional rejection probability to one half after using half the desired error bound.

3.7 Lower bounds from mixtures

Hold one latent law fixed at a centered isotropic Gaussian g of precision strictly between m and M, under the null and under the alternatives. Vary the group which attains r_*, writing n, s, Σ for that group's sample size, largest noise eigenvalue, and original covariance. The unchanged independent group does not alter the likelihood ratio or total variation. It suffices to show the individual testing lower bound.

We will use finite families $f_\omega = g + \sum_{i=1}^{N} \omega_i q_i$ where ω_i are independent signs and each q_i has integral zero. Let G be the Gram matrix of the noisy score perturbations in $L^2(p_g)$:

$$G_{ij} = \int \frac{(q_i * \gamma_\Sigma)(q_j * \gamma_\Sigma)}{g * \gamma_\Sigma}.$$

For the uniform mixture of n-sample laws over signs, one plus the chi-square divergence to the null equals $\mathbb{E}_{\omega,\omega'}(1 + \omega^T G \omega')^n$. The factor inside the power is nonnegative as the inner product of two likelihood ratios. Bounding its power by $\exp(n \omega^T G \omega')$ and integrating one sign vector gives $\mathbb{E}_\omega \exp(n^2 \|G\omega\|^2/2)$ as an upper bound, using $\cosh(t) \le e^{t^2/2}$. Introduce a standard Gaussian vector to linearize the square and apply the same sign bound again. The result is at most

$$\det(I - n^2 G^2)^{-1/2} \le \exp(C n^2 \|G\|_{\mathrm{HS}}^2) \qquad \text{if } n\|G\|_{\mathrm{op}} \le 1/2.$$

Thus a small bound on $n^2\|G\|_{\mathrm{HS}}^2$ makes the mixture total variation arbitrarily small. If every mixture member has $W_2(f_\omega, g) \geq r$, the sum of testing errors at separation r is at least one minus that total variation.

For transport separation we use a smooth test φ. Taylor expansion under an optimal coupling gives

$$\int \varphi(f_\omega - g) \leq \|\nabla\varphi\|_{L^2(f_\omega)} W_2(f_\omega, g) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(f_\omega, g)^2. \tag{23}$$

If the pairing dominates cE, the gradient norm is at most $C\sqrt{E}$, and the Hessian norm is bounded, it follows that $W_2^2(f_\omega, g) \geq c'E$.

3.7.1 The direct polynomial lower bound

For $d > 4$, put $h = n^{-2/(4\alpha+d)}$. Pack order h^{-d} disjoint cubes of side h in a fixed cube about the reference mode. Let ψ be a nonzero smooth bump supported inside the reference unit cube with integral zero, and use $q_i(x) = a_0\psi((x - z_i)/h)$, $a_0 = \kappa h^\alpha$. The relative perturbations q_i/g have uniformly bounded C^α norm of order κ on the combined support, by rescaling and the fixed positive density floor there. The vanishing near edges gives the same bound across cubes. The logarithmic derivative bounds follow by composition with $\log(1 + \cdot)$; for sufficiently small κ the reference slack places every alternative in the class.

The clean score Gram matrix is diagonal with entries bounded by $Ca_0^2h^d$. Observed score Gram matrices are bounded above by their clean counterpart in positive operator order: the observed score is the conditional expectation of the clean score under the reference joint law, so this follows by conditional Cauchy–Schwarz for any linear combination. Thus $n^2\|G\|_{\mathrm{HS}}^2 \leq C\kappa^4 n^2 h^{4\alpha+d} = C\kappa^4$. For separation take $\varphi = h^2\sum_i \omega_i q_i$. Its pairing with $f_\omega - g$ is of order $E = h^{d+2}\sum_i a_0^2 \asymp h^{2\alpha+2}$, by disjointness. The gradient's squared L^2 norm is $O(E)$, and its Hessian is uniformly bounded. Equation (23) therefore gives the direct lower bound for every sign.

3.7.2 The direct critical lower bound

When $d = 4$, pack disjoint cubes of side $\ell = (\log n)^{-1/2}$ inside a cube of side $c\sqrt{\log n}$ about the mode, with $c > 0$ small. Their total volume is order $(\log n)^2$. The reference g is comparable to its value g_B at the center of each cube B, and $g_B \geq n^{-\beta_0}$ for a fixed $\beta_0 < 1$. In the rescaled unit cube the ratio g_B/g has bounded C^α norm uniformly: on length ℓ, the score bound is paid by $\ell\sqrt{\log n} = 1$, and all higher potential derivatives are bounded with the corresponding scale factors. The top Hölder bound scales by ℓ^α.

Use a fixed orientation of compactly supported orthonormal tensor wavelets [7] of regularity C^{k+1}, supported strictly inside each cube. At physical length $r_i = \ell 2^{-j}$ their number in a cube is comparable to $(\ell/r_i)^4$, for j from a sufficiently large fixed index through $\lfloor \nu \log_2 n \rfloor$, with $\nu > 0$ small. Write ψ_i for these L^2-normalized wavelets. In particular $S_4 = \sum_i r_i^4 \asymp (\log n)^3$: each level contributes order the total volume. Set

$$b_i = \frac{\kappa^2 r_i^2}{n\sqrt{S_4}}, \qquad a_i = \sqrt{g_{B(i)} b_i}, \qquad q_i = a_i\psi_i.$$

Each wavelet has zero integral. At physical scale r, the relative perturbation in cube B has C^α norm at most $C(a_i/g_B)r^{-2-\alpha}$, by wavelet scaling, bounded overlap at the scale, and the rescaled reciprocal density bound. Summing over scales gives at most

$$C\kappa n^{-(1-\beta_0)/2} S_4^{-1/4} r_{\min}^{-1-\alpha} = o(1)$$

if $\nu < (1-\beta_0)/(2(1+\alpha))$. For fractional order this controls the global seminorm: on a pair of points in distinct cubes one may use the zero values of the top derivatives on the boundary of each cube and distances to the first exit points on the segment. Thus all perturbations stay within the reference margins by logarithmic composition.

The clean Gram matrix is $D_0 H D_0$ with $D_0 = \mathrm{diag}(\sqrt{b_i})$, and $H \preceq CI$. This uses orthonormality and the uniform bounds on g_B/g within each cube. The noisy Gram shrinks in positive operator order as before, so $n^2\|G\|_{\mathrm{HS}}^2 \leq Cn^2\sum_i b_i^2 = C\kappa^4$. For the metric test take $\varphi = \sum_i (r_i^2/g_{B(i)})\omega_i a_i \psi_i$. Its pairing is exactly $E = \sum_i r_i^2 b_i = \kappa^2\sqrt{S_4}/n$. The gradient squared norm in a cube is controlled by the sum of squared coefficient

norms weighted by r_i^{-2}, up to a fixed factor. For clarity, within one level sum u_j, $\|\Delta u_j\|_2 \le C r_j^{-2}\|u_j\|_2$ by overlap. When $r_j \ge r_\ell$, integration by parts yields

$$|\langle \nabla u_j, \nabla u_\ell\rangle| \le C(r_\ell/r_j)(\|u_j\|_2/r_j)(\|u_\ell\|_2/r_\ell).$$

Summing the geometric ratios and multiplying by the comparable density in each cube gives squared gradient norm $O(E)$ under f_ω. The test Hessian is uniformly bounded, since its per cube scale sum is bounded by $C\kappa(ng_B)^{-1/2}S_4^{-1/4}r_{\min}^{-1}$. Thus (23) gives separation at least $c(\log n)^{3/4}/\sqrt{n}$. This lower bound holds under every known Gaussian covariance, by the same mixture contraction.

3.7.3 The inverse mixture

Rotate axes so $\Sigma = \mathrm{diag}(s, t_2, \dots, t_d)$ with $0 \le t_i \le s$, and suppose $0 < s \le s_0$ with $s_0 > 0$ small. Place $N \asymp s^{-d/2}$ centers z_i spaced $D\sqrt{s}$ apart in a fixed cube. Here D is sufficiently large, and s_0 small after choosing D. Put $T^2 = A_0 \log(2 + ns^{\alpha+d/4})/s$, and use

$$q_i(x) = a_0\psi_i(x), \qquad a_0 = \kappa T^{-\alpha}, \qquad \psi_i(x) = e^{-|x-z_i|^2/(2s)}\left(\cos(T(x_1 - z_{i,1})) - e^{-sT^2/2}\right).$$

Each packet has zero mass. The relative perturbation before the factor a_0 has derivatives through order $j \le k$ bounded by CT^j. Indeed derivatives of the packets are bounded by $CT^j e^{-c|x-z_i|^2/s}$ because $T\sqrt{s}$ is large, and division by g costs only Gaussian and polynomial factors absorbed in a weaker envelope for small s_0. Separated grid envelopes have uniformly bounded sum. For $0 < \theta < 1$, the k-th derivative has θ-Hölder seminorm at most CT^α. For shifts longer than $1/T$ use the sup norm. For a shorter shift expand the product-rule terms of $D^k(\psi_i/g)$. A difference of a smooth packet derivative of order j costs $CT^{j+1}|u|$ with the same weakened envelope. A difference of a reciprocal derivative of order $k-j$ costs $C|u|^\theta$ times polynomial factors and the maximum of the two reciprocal densities, by the product rules and the reference Hölder condition on $D^k V_g$. The nearby packet envelopes still absorb those factors since $|u| \le 1/T \le C\sqrt{s}$. Summing the bounds gives $CT^\alpha|u|^\theta$. The logarithmic perturbation thus costs $O(\kappa)$ in the stipulated norms and preserves the class for small enough κ.

Convolution with the noise changes the packet envelope to variances $s + t_i$, with $t_1 = s$, multiplying its amplitude by $\prod_i \sqrt{s/(s+t_i)}$. Its bracket becomes

$$e^{-sT^2/4}\cos\big(\tfrac{T}{2}(y_1 - z_{i,1})\big) - e^{-sT^2/2}.$$

These formulas follow by conditional Gaussian integration and include zero eigenvalues in the other coordinates. The convolved reference has a lower Gaussian envelope, obtained by convolving one for g. Combining the two narrow packet envelopes and dividing by this lower envelope gives

$$|G_{ij}| \le Ca_0^2 s^{d/2} e^{-csT^2} e^{-c|z_i - z_j|^2/s}.$$

Indeed the common center lies in a fixed cube, the variances of the two envelopes are between s and $2s$, and choosing s_0 small absorbs the reciprocal reference into a weaker Gaussian. Thus

$$n\|G\|_{\mathrm{HS}} \le Cna_0^2 s^{d/4} e^{-csT^2}.$$

With $X = ns^{\alpha+d/4}$, the right-hand side is at most $C\kappa^2 X(A_0\log(2+X))^{-\alpha}(2+X)^{-cA_0}$. It is arbitrarily small by choosing A_0 large and κ small.

For transport separation take $\varphi = T^{-2}\sum_i \omega_i q_i$. The spatial L^2 Gram matrix of the ψ_i has diagonal at least $cs^{d/2}$ by integrating the squared cosine in its Gaussian envelope and taking $T\sqrt{s}$ large. The absolute off-diagonal row sums are at most $Cs^{d/2}\sum_{j\in\mathbb{Z}^d\setminus\{0\}} e^{-cD^2|j|^2}$, less than half the diagonal bound for large D. The pairing with $f_\omega - g$ is at least cE for $E = T^{-2}a_0^2 s^{d/2}N$. The gradient Gram has absolute row sums at most $CT^2 s^{d/2}$ by the same overlap estimate, so the test gradient has squared norm $O(E)$ under the uniformly bounded vertex densities. The pointwise derivative envelope bounds the test Hessian by Ca_0. Equation (23) gives $W_2(f_\omega, g) \ge cT^{-\alpha-1}$ for every sign.

For $s \ge s_0$, use instead one additive packet throughout, replacing its envelope variance by a fixed τ_0^2 smaller than the variance in the Gaussian lower envelope for g, with its cosine mean removed. Use

amplitude $\kappa T^{-\alpha}$ and $T^2 = C \log n/s$. The derivative and Hölder argument above applies to this single packet. Its convolved Gaussian envelope squared, divided by the convolved reference lower envelope, is integrable uniformly in Σ by the variance choice. Its first-coordinate cosine attenuates by $\exp(-\tau_0^2 s T^2/[2(\tau_0^2+s)])$, so the n-sample chi-square tends to zero uniformly for sufficiently large C. Testing the latent difference with $\cos(Tx_1)/T$ gives separation of order $T^{-\alpha-1}$. This matches the inverse term since its logarithm is comparable to $\log n$ on this range.

3.7.4 Lower bounds in dimensions one through three

In dimension one a shift of the reference by δ has W_2 distance $|\delta|$. Its clean one-sample relative entropy to the reference is $O(\delta^2)$, by expanding the potential and using the zero mean score. Since noise contracts entropy, $\delta = c/\sqrt{n}$ gives the desired two-point lower bound.

For $d = 2, 3$, take disjoint annuli at dyadic radii R from a fixed constant to $c\sqrt{\log n}$, for small fixed $c > 0$. Pack cubes of side $\ell_i \asymp R^{-1}$ in each annulus, with order R^{2d} cubes at radius R. The reference values g_i at their centers satisfy $g_i \geq n^{-\beta_0}$, $\beta_0 < 1$, and $g \asymp g_i$ inside each cube. Use one L^2-normalized smooth mean-zero bump ψ_i per cube, with

$$S_4 = \sum_i \ell_i^4, \qquad b_i = \frac{\kappa^2 \ell_i^2}{n\sqrt{S_4}}, \qquad q_i = \sqrt{g_i b_i}\, \psi_i.$$

The rescaled reciprocal bounds used for the critical construction apply here since $\ell_i(1 + |x|) = O(1)$. All relative C^α perturbations tend to zero because $g_i \geq n^{-\beta_0}$ and $\ell_i \gtrsim (\log n)^{-1/2}$. The clean Gram matrix is diagonal with entries $O(b_i)$, and the noisy Gram contracts. Hence $n^2 \|G\|_{\mathrm{HS}}^2 = O(\kappa^4)$. The test $\varphi = \sum_i (\ell_i^2/g_i)\omega_i q_i$ has pairing $\sum_i \ell_i^2 b_i$, squared gradient norm at most a constant times that pairing, and bounded Hessian by rescaling on the disjoint cells. Equation (23) gives squared separation of order $\sqrt{S_4}/n$. A ring contributes order R^{2d-4} to S_4; summing gives order $\log\log n$ for $d = 2$ and $\log n$ for $d = 3$.

For the inverse lower bound in all three dimensions, the single additive narrow packet with $T^2 = C \log n/s$ works throughout $0 < s \leq \Lambda$. Its attenuated cosine multiplier is bounded by e^{-csT^2} uniformly for the fixed narrow variance, and the noisy reference denominator is paid as above. It gives latent separation of order $(s/\log n)^{(\alpha+1)/2}$ with small total variation. Together with the direct lower bounds this completes Theorem 3.1.

3.8 Relation to direct and indirect two-sample tests

Lee, Lahiri, and Sinha [17] test equality of latent distributions under measurement error through deconvolved characteristic functions and bootstrap calibration. Their framework centers on asymptotic calibration and consistency. Theorem 3.1 treats uniform latent W_2 separation and arbitrary sample-size ratios. The direct metric-space results of [9] allow rough distributions, with additional bounds for unbounded exponential tails. Manole [16] reports a direct testing gap without smoothness. Tang and Yang [13] treat smooth adversarial losses on compact domains. The sliced statistic in [15] has a parametric separation in the powered sliced loss in its matching regimes. Bounded potential curvature on the whole space supplies both the inward density extrapolation and the weighted transport comparison here. The low-dimensional logarithms arise from the number and scale of distinguishable perturbations in the tails, with an additional accumulation over resolution at dimension four.

4 Two-sample Wasserstein testing with unknown latent distributions and Gaussian noise

We determine the minimax separation for testing equality of two unknown strongly log-concave densities in quadratic Wasserstein distance from independent samples with known Gaussian contamination. For comparable sample sizes or equal covariances, the result holds in every fixed dimension, for real potential smoothness at least two, and allows singular covariances. The main step estimates both densities down to the edge of the testing region, where a cube may have a

vanishing expected count. We extrapolate log densities from two nearby inward locations with larger expected counts. A quadratic test of transport fields then attains the same separation orders as when one density is known.

4.1 Experiment and separation

Fix $d \geq 1$, real $\alpha = k+\theta \geq 2$, integer k, $0 \leq \theta < 1$, and constants $0 < m < M < \infty$. Let $\mathcal{F}_\alpha$ be the translated class of 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 $\theta > 0$, impose a fixed positive bound on the global θ-Hölder seminorm of D^kV. We observe independent samples of sizes n_1, n_2 from $p_f = f * \gamma_{\Sigma_1}$ and $p_g = g * \gamma_{\Sigma_2}$. The two independent Gaussian errors have known covariances $\Sigma_i \succeq 0$ of operator norm at most Λ. Here $\gamma_\Sigma = N(0, \Sigma)$ and $\gamma_s = N(0, sI_d)$. Let $n = \min(n_1, n_2)$, $s = \max_i \|\Sigma_i\|_{\mathrm{op}}$, and

$$a_{n,d} = \begin{cases} n^{-1/2}, & d = 1, \\ (\log\log n)^{1/4}/\sqrt{n}, & d = 2, \\ (\log n)^{1/4}/\sqrt{n}, & d = 3, \\ (\log n)^{3/4}/\sqrt{n}, & d = 4, \\ n^{-2(\alpha+1)/(4\alpha+d)}, & d > 4. \end{cases} \qquad B_{n,d}(s) = \begin{cases} \log n, & d < 4, \\ \log(2 + ns^{\alpha+d/4}), & d \geq 4. \end{cases}$$

The risk of a test is the sum of its worst type-I error over the composite null $f = g \in \mathcal{F}_\alpha$ and worst type-II error over $f, g \in \mathcal{F}_\alpha$ with $W_2(f, g) \geq r$.

Theorem 4.1. *If the sample sizes are comparable, or if $\Sigma_1 = \Sigma_2$, the minimax separation at any fixed error budget is of order*

$$r_{n,d}(s) = a_{n,d} + (s/B_{n,d}(s))^{(\alpha+1)/2}.$$

Precisely, for any fixed $0 < \varepsilon < 1/2$ there are $c_\varepsilon, C_\varepsilon > 0$ such that the infimum risk at $c_\varepsilon r_{n,d}(s)$ is at least $1 - \varepsilon$, and at $C_\varepsilon r_{n,d}(s)$ it is at most ε, uniformly over the stipulated experiment for sufficiently large n. For comparable sizes the constants can depend on the fixed comparability bound.

Fixed-reference latent Wasserstein testing for fixed positive Gaussian noise is treated in [12, Theorem 2.4]. Direct two-sample Wasserstein tests for general metric spaces and compact cubes are studied in [9]; compact smooth adversarial losses are studied in [13], and sliced Wasserstein testing in [15]. Here both reference shape and tails are unknown. The pointwise estimates below use bounded curvature to normalize a composite-null test without imposing a global density floor.

4.2 Deterministic transport and Fourier fields

Uniformly over the class, f has Gaussian upper and lower envelopes about its mean m_f, the score of f is $O(1 + |x - m_f|)$, and

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

Indeed quadratic bounds at the mode give the envelopes and bound the distance between the mode and the mean. Differentiating e^{-V} gives the derivative bounds. For $\theta > 0$, on $|x - m_f| \leq L$ and $|u| \leq (2 + L)^{-1}$, the same product rule and the Hölder bound on D^kV give

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

The local density ratio is bounded by the score, derivatives of lower order in the product are Lipschitz, and the highest one uses the assumed Hölder seminorm.

For $p_f = f * \gamma_s$, isotropic $0 \leq s \leq \Lambda$, Gaussian integration of the two quadratic Taylor bounds gives

$$cf(x) \leq p_f(x) \leq f(x) e^{Cs(1+|x-m_f|^2)}. \tag{24}$$

The potential of p_f also has Hessian in a fixed positive interval. For $s > 0$, its Hessian is $s^{-1}I - s^{-2}\operatorname{Cov}(X \mid Y)$. The conditional potential has Hessian between $(m + s^{-1})I$ and $(M + s^{-1})I$. The upper covariance bound

follows from Brascamp–Lieb [5]. The lower bound follows, in each unit direction v, from Cauchy–Schwarz and the conditional identities

$$\mathbb{E}[(v\cdot(X-\mathbb{E}X))(v\cdot\nabla A(X))]=1,\qquad \mathbb{E}[(v\cdot\nabla A(X))^2]=\mathbb{E}[v^TD^2A(X)v],$$

where A is the conditional potential. Consequently the scores of p_f are $O(1+|x-m_f|)$. The observed law always has a uniformly Gaussian upper envelope about m_f.

Use $\widehat{u}(\xi)=\int e^{-ix\cdot\xi}u(x)\,dx$. Let $\widehat{K}$ be a real even smooth Fourier cutoff that equals one near zero and is supported in the unit ball. Use $K_b=b^{-d}K(\cdot/b)$ for its spatial kernel, and define constant-bandwidth fields

$$U_b(f)=K_b*f,\qquad \widehat{J_b(f)}(\xi)=-\frac{i\xi}{|\xi|^2}(1-\widehat{K}(b\xi))\widehat{f}(\xi).$$

Then $f=U_b(f)+\operatorname{div}J_b(f)$. The inverse transform Q of $-i\xi(1-\widehat{K}(\xi))/|\xi|^2$ is integrable with all absolute moments finite and all signed moments zero. In fact a dyadic partition into shells 2^j away from the origin gives inverse transforms bounded by $C_N2^{(d-1)j}(1+2^j|x|)^{-N}$. Summability of their L^1 norms $O(2^{-j})$ gives the assertion, with moment cancellation since the symbol vanishes near zero. For $|x-m_f|\le L$, $w=(2+L)^{-1}$, and $h\le w$, subtract the local Taylor polynomial of f to degree $k-1$ in the integer case and k otherwise. This gives

$$|J_h(f)(x)|\le Ch^{\alpha+1}(1+|x-m_f|^{k+1})f(x)+C_Nh(h/w)^N. \tag{25}$$

Indeed $J_h(f)=hQ_h*f$, the short-shift remainder has order α, and the farther shifts use uniform absolute derivative bounds and the arbitrary integrated tails.

Choose the origin within a fixed distance of the means of f,g, and put $\rho=f+g$. For a test φ, set $u=\varphi-\frac12\int\varphi\rho$. Then

$$\int(1+|x|^2)|u|^2\rho\le C\int|\nabla\varphi|^2\rho. \tag{26}$$

Both components satisfy ordinary Poincaré by Brascamp–Lieb. Comparing their component means to a common Lebesgue average on a fixed ball with a positive density floor gives the mixture inequality. For either potential V, $x\cdot\nabla V(x)\ge m|x|^2/2-C$, since its mode stays close to zero. Integration by parts gives

$$\int u^2x\cdot\nabla Vf=d\int u^2f+2\int ux\cdot\nabla\varphi f.$$

Use Cauchy–Schwarz to absorb half of the resulting quadratic term, add the two estimates, and apply mixture Poincaré. Gaussian tails justify integration by parts for compactly supported tests minus their mean.

Let $I_0\subset I$ admit a C^1 cutoff supported in I, equal to one on I_0, with $|\nabla\chi|\le C(1+|x|)$. For any identity $f-g=U+\operatorname{div}J$ in distributions,

$$W_2(f,g)\le C\left[\left(\int_{I_0^c}\frac{\rho}{1+|x|^2}\right)^{1/2}+\left(\int_I\frac{|U|^2/(1+|x|^2)+|J|^2}{\rho}\right)^{1/2}\right]. \tag{27}$$

To see this, split $\int(f-g)u$ with the cutoff and use Cauchy–Schwarz with (26) in the exterior. In the interior insert the field identity and integrate by parts. Approximation by smooth tests justifies this for a C^1 cutoff. The $L^2(\rho)$ norms of $(1+|x|^2)^{1/2}\chi u$ and $\nabla(\chi u)$ are at most $C\|\nabla\varphi\|_{L^2(\rho)}$. Riesz representation produces an exact flux of divergence $f-g$ whose squared action norm $\int|J|^2/\rho$ is bounded by the square of the displayed right-hand side. Interpolate between the densities with $(1-\lambda)f+\lambda g$, $\lambda(t)=\sin^2(\pi t/2)$, and flux $\lambda'(t)J$. Its action is at most $\pi^2\int|J|^2/\rho$, by the lower bound $\min(\lambda,1-\lambda)\rho$. The dynamic transport inequality [4] gives (27).

4.3 Estimating the common weights from inward probes

After completing the covariances by independent Gaussian additions, both observation laws have noise sI. We may use n observations of each law. For small noise assume $0\le s\le n^{-\eta}$ for a sufficiently small fixed $\eta>0$. Set $a=\log n$, $w=a^{-1/2}$,

$$h_0=\max\{a_{n,d}^{1/(\alpha+1)},\sqrt{s/(\gamma B_{n,d}(s))}\},\qquad r=h_0^{\alpha+1},\qquad \rho_0=r^2a^{2-d/2}.$$

The positive constant γ can be chosen sufficiently small. Both r and h_0 are bounded above and below by fixed negative powers of n. Split each sample into three parts of sizes comparable to n. The first parts estimate each mean. Reject if their difference has norm greater than one and, for the remaining calculation, translate by the first estimated mean. Gaussian concentration gives an event of probability $1 - o(1)$, uniformly over the experiment, on which this preliminary test passes under the null and both actual means are within a fixed distance of the new origin whenever the test passes. The following construction is conditional on those first parts.

Write $\rho = f + g$. Tile with w-cubes, retaining all cubes meeting the ball of radius $L = D_0\sqrt{a}$. There are $O(a^d)$ cubes. Write x for a cube center. For each observed density p separately, estimate $p(x)$ at inner rows $|x| \le c_1\sqrt{a}$ using its second sample count in the cube divided by the second sample size times w^d. For outer rows put $\omega = x/|x|$, $t = C_1 \log a/\sqrt{a}$, and count in the w-cubes centered at $x - t\omega$ and $x - 2t\omega$. Each count of at least $A_1 \log a$ yields a density estimate at its probe by the same division. If both counts attain this threshold, set the estimate at x to the square of the first probe estimate divided by the second. Otherwise set it to zero. The estimates for the two densities use their own observations.

Uniformly with probability $1 - o(1)$, every count above the threshold agrees with its mean up to fixed factors and every count of mean at least $2A_1 \log a$ is above the threshold. This follows from binomial Chernoff bounds and a union bound over $O(a^d)$ probes with A_1 large. A probe cube average agrees with its center value up to fixed factors because the observed score on the grid region is $O(\sqrt{a})$. The inner counts have polynomially large means, by the uniform Gaussian lower envelope with c_1 sufficiently small.

On the outer segments, the derivative of $-\log p$ in direction ω is at least $c\sqrt{a}$. Indeed the Hessian is bounded below and the mode is within a fixed distance of the origin. Each inward probe value is therefore at least $a^{cC_1}p(x)$. Since

$$nw^d\rho_0 = nr^2a^{2-d} \gtrsim a^{-1/2},$$

every outer center with $p(x) \ge \rho_0$ has two reliable probes if C_1 is large. The Hessian upper bound gives

$$\big|\log p(x) - 2\log p(x - t\omega) + \log p(x - 2t\omega)\big| \le Ct^2 = o(1).$$

Thus every positive estimate $\widehat{p}(x)$ is comparable to $p(x)$, and a zero estimate implies $p(x) < \rho_0$. No independence of counts within a sample is needed.

Let I be the union of the retained cubes for which the sum of the two estimates at the center is above $K_1\rho_0$, with K_1 a large fixed constant, and let $\widehat{\rho}$ be that sum extended constantly on each selected cube. Throughout the successful pilot event,

$$c\rho(x) \le \widehat{\rho}(x) \le C\rho(x) \quad (x \in I), \qquad \{\rho > C_2\rho_0\} \subseteq I. \tag{28}$$

Indeed the ignored terms at a selected center can contribute at most $2\rho_0$ in observed density. Center values and cube values are comparable by the score bounds, and the latent and observed values are comparable there by (24) and $sa = o(1)$. Every center with large enough density has a positive estimate of the same order. Choose D_0 large so that beyond the tiled ball the Gaussian upper bound is below ρ_0. This establishes the inclusion.

For the analysis take a smooth cutoff as a scalar function of ρ/ρ_0, zero below C_2 and one above $2C_2$. Its support is in I (increase its lower threshold if needed) and its gradient is bounded by $C(1 + |x|)$. For each component density $u = f, g$ integrate on its potential levels about its mode. On a level z above the minimum, the radius and radial derivative are comparable to $\sqrt{z}$. The level-volume derivative is $O(z^{d/2-1})$ and $1 + |x|^2 \gtrsim z$ on large levels. Since $\log(1/\rho_0) \asymp a$, this proves

$$\int_{\{\rho \le 2C_2\rho_0\}} \frac{\rho(x)}{1 + |x|^2}\, dx \le C\rho_0 a^{d/2-2} = Cr^2.$$

Here the domain is contained in the individual low-density domain for each component. Only the analytical cutoff uses the unknown densities.

4.4 A held-out quadratic test

Put $D(x) = 1 + |x|^2$ and take $b(x)^2 = D(x)^{-1}$ if $d < 4$ and $b(x)^2 = 1/a$ otherwise. Let $v_x = b(x)^2 - s \asymp b(x)^2$ on I. Use $h(x) = h_0$ if $d \le 4$, and $h(x) = h_0 e^{c_0|x|^2}$ if $d > 4$. Choose $c_0 > 0$ sufficiently small so that h

remains polynomially small on I and so that $e^{(2\alpha+2)c_0|x|^2}$ has uniformly bounded polynomial moments under the two latent laws. Then $h(x)^2 \ll v_x$ and $h(x) \asymp h(x')$ for $|x - x'| \le 2w$.

For $u = f, g$, let $p_{u,t} = u * \gamma_t$. The exact identity $u = U(u) - \operatorname{div} J(u)$ uses the fields

$$J(u)(x) = \tfrac{1}{2} \int_0^{b(x)^2} \nabla p_{u,t}(x)\, dt, \qquad U(u)(x) = p_{u,b(x)^2}(x) + \tfrac{1}{2} \nabla b(x)^2 \cdot \nabla p_{u,b(x)^2}(x).$$

The gradients of $p_{u,t}$ hold t fixed. The chain rule and the heat equation give the identity. All fields on the pilot region can be viewed as restrictions of fields on a neighborhood of the cutoff.

Define $R_t = \frac{1}{2} \int_0^t \nabla \gamma_z \, dz$. For an observed y use scalar and vector kernels at row x applied at $x - y$. The scalar kernel is

$$G_x = \gamma_{v_x} + \tfrac{1}{2} \nabla b(x)^2 \cdot \nabla \gamma_{v_x},$$

divided by $(D(x)\widehat{\rho}(x))^{1/2}$. The vector kernel is the sum of $R_{v_x} - R_{h(x)^2}$ and the kernel with symbol

$$\frac{i\xi}{|\xi|^2} \left(\widehat{K}(h(x)\xi) e^{s|\xi|^2/2} - e^{-h(x)^2|\xi|^2/2} \right). \tag{29}$$

Truncate the latter kernel to $|x - y| \le w$ and divide the sum by $\widehat{\rho}(x)^{1/2}$. Denote the resulting feature, in $L^2(I) \oplus L^2(I; \mathbb{R}^d)$, by F_y.

Conditional on the pilot, the scalar expectation is the normalized $U(u)$. Holding row bandwidths fixed, the untruncated vector expectation has symbol $i\xi(\widehat{K}(h(x)\xi) - e^{-b(x)^2|\xi|^2/2})/|\xi|^2$ on the latent law, before normalization. It equals $J(u)$ plus the high-pass remainder from (25). That remainder has normalized L^2 norm $O(r)$ by (28), $u^2/\widehat{\rho} \le Cu$, and the Gaussian moment choice of c_0.

For the fine kernel, rescale by $h = h(x)$ to obtain $h^{1-d} L_A((x-y)/h)$, $A = s/h^2$. The symbol after scaling has a smooth extension through zero, since the bracket is radial analytic there and vanishes to second order. For each fixed number of Schwartz seminorms the bound for L_A is $C(1 + A)^C e^{CA}$, by compact support of $\widehat{K}$ and Gaussian decay of the second term. Since h is polynomially small, $A \le \gamma B_{n,d}(s) = O(\log n)$, and $\widehat{\rho} \gtrsim \rho_0$ on I, the truncation costs $O(n^{-T})$ in normalized L^2 for every fixed T. With $\mu = \mathbb{E}_f F_Y - \mathbb{E}_g F_Y$, the cutoff, the transport comparison (27), and these bias bounds give

$$W_2(f, g) \le C(r + \|\mu\|). \tag{30}$$

4.4.1 Conditional covariance

We work throughout on the successful pilot event. It suffices to bound the Hilbert–Schmidt norms of the three uncentered second moment operators (scalar, coarse vector, fine vector). Their sum, multiplied by three, dominates the centered covariance in positive operator order by Cauchy–Schwarz, and the Hilbert–Schmidt norm is monotone on positive operators. Fix either completed input law with density p. Its supporting log-density tangent at each row and (28) bound

$$\frac{p(y)}{\sqrt{\widehat{\rho}(x)\widehat{\rho}(x')}} \le C \exp(C(1 + |x|)|y - x| + C(1 + |x'|)|y - x'|).$$

Each factor is absorbed in the row Gaussian, since $b(x)(1 + |x|) \le C$. It is bounded on the fine support.

In absolute value the tilted scalar kernel before division by $\sqrt{D(x)}$ is bounded by

$$Cb(x)^{-d} e^{-c|x-y|^2/b(x)^2}.$$

This includes the derivative correction since $|\nabla b(x)^2| \le Cb(x)^3$ in the variable case. The tilted coarse vector has absolute kernel bounded by

$$C(h(x) + |x - y|)^{1-d} e^{-c|x-y|^2/b(x)^2}.$$

For $|x - y| \le h(x)$ integrate the time powers above $h(x)^2$. At larger distances, keep half the Gaussian at scale $b(x)$ and extend the remaining time integral to infinity.

For $d < 4$, at pairs with $1 + |x| \asymp 1 + |x'|$ (ratio between $1/2$ and 2), use common bandwidth $b(x)$. The scalar covariance kernel has squared row integral bounded by $Cb(x)^{-d}D(x)^{-2}$. For the coarse vector use $C|z|^{1-d}e^{-c|z|^2/b(x)^2}$, whose $L^{4/3}$ norm is $O(b(x)^{1-d/4})$; Young's inequality gives the same row bound $Cb(x)^{4-d}$. For noncomparable pairs, the separation is comparable to the larger radius. Extract a Gaussian in separation using both row Gaussians with bounded bandwidth. Scalar prefactors then grow at most polynomially in the radii. For the vector, split into balls of radius one third of the separation about either row and their complement. On each ball the other singularity is bounded by a polynomial in separation and the singularity at the ball center is integrable against its Gaussian. On the complement both singularities are bounded. The two squared kernels are thus integrable on this pair region uniformly in n. Integrating the comparable row bound $Cb(x)^{4-d}$ on I gives $O(1), O(\log a), O(a)$ in dimensions one, two, three respectively.

For $d \geq 4$ and the common bandwidth b, the scalar squared norm is bounded by

$$Cb^{-d} \int_I D(x)^{-2}\, dx$$

by Cauchy–Schwarz in the row variables and Gaussian convolution. It is $O(a^2 \log a)$ for $d = 4$, and at most a fixed power of a for $d > 4$. For the coarse vector, the convolution bound at $z = x - x'$, $|z| \leq 2w$, is

$$C(h(x) + |z|)^{2-d} e^{-c|z|^2/b^2}.$$

The two regularizations are comparable there. For $|z| \leq h(x)$ apply Cauchy–Schwarz to the two profiles. Above that scale, split near either singular center and in the remainder to obtain $C|z|^{2-d}$ before the Gaussian: the two radial integrals are bounded by $|z|^{1-d} \int_0^{|z|} dt$ and $\int_{|z|}^{\infty} t^{1-d}\, dt$. The same unregularized bound applies at $|z| > 2w$. In dimension four the squared near integral in z is $O(1 + \log(b/h_0)) = O(a)$, which gives $O(a^3)$ after integration in x, including the far contribution. For $d > 4$ the squared near bound is $C \int_I h(x)^{4-d}\, dx \leq Ch_0^{4-d}$. The far bound is polylogarithmic and is absorbed by h_0^{4-d}.

The fine covariance vanishes at $|x - x'| > 2w$. With $A_0 = s/h_0^2$ and $E = C(1 + A_0)^C e^{CA_0}$, its absolute kernel on the near pairs is bounded by $CE^2 h(x)^{2-d}(1 + |z|/h(x))^{-N_0}$ for fixed arbitrarily large N_0. This follows by convolving the two scaled Schwartz profiles at comparable bandwidths. The squared norm is therefore bounded by $CE^4 \int_I h(x)^{4-d}\, dx$. For $d < 4$ this is $o(1)$ since h_0 is polynomially small and $A_0 \leq \gamma \log n$. For $d = 4$ it is $O(E^4 a^2)$, and above four it is $O(E^4 h_0^{4-d})$.

For $d > 4$, set $X = ns^{\alpha+d/4}$. For sufficiently small γ we have $E^2 \leq C(2 + X)^c$ with $c < 1$, while

$$nh_0^{2\alpha+d/2} \geq \max\left(1, \frac{X}{(\gamma \log(2 + X))^{\alpha+d/4}}\right).$$

This pays the fine term $E^2 h_0^{2-d/2}/n$ by Cr^2. The coarse terms divided by n are smaller than $a_{n,d}^2$. For $d = 4$, set $X = 2 + ns^{\alpha+1}$. Choose γ so $E^2 \leq CX^c$ with $c < 1/4$. If $X \leq a^2$, then $E^2 a/n \leq Ca^{3/2}/n$. Otherwise $E^2 a \leq CX^{1/2+c} \leq C(X - 2)/(\log X)^{\alpha+1}$, so the fine term is at most Cr^2. The coarse term divided by n is $O(a^{3/2}/n)$. The low-dimensional bounds above agree with $na_{n,d}^2$ after taking square roots. Consequently, in every dimension,

$$n^{-1} \| \operatorname{Cov}(F_Y) \|_{\mathrm{HS}} \leq Cr^2 \tag{31}$$

for either input law.

4.4.2 Rejection rule

On the independent third parts, pair observations into $Z_i = F_{Y_i} - F_{Y_i'}$, $1 \leq i \leq N$, where $N \asymp n$. They are independent and identically distributed conditional on the pilot, with mean μ and covariance $C = C_f + C_g$. Use the statistic

$$\mathcal{U} = \frac{1}{N(N-1)} \sum_{i \neq j} \langle Z_i, Z_j \rangle.$$

Its conditional mean is $\|\mu\|^2$ and its variance equals

$$\frac{4}{N} \langle \mu, C\mu \rangle + \frac{2}{N(N-1)} \|C\|_{\mathrm{HS}}^2.$$

Indeed expansion into a linear sum and a centered degenerate quadratic sum gives uncorrelated summands of the two stated kinds. At the composite null $\mu = 0$ exactly, by the covariance completion and independence of the pilot. On the good pilot event the variance is at most $C(r^4 + r^2\|\mu\|^2)$ by (31). Reject when $\mathcal{U} > A_\varepsilon r^2$. Chebyshev bounds the null error by an arbitrarily small constant with A_ε large. When the latent separation exceeds a sufficiently large multiple of r, (30) makes the mean an arbitrarily large multiple of this threshold and the same inequality gives the required type-II bound. Add the $o(1)$ failure probabilities of the pilots, and include preliminary mean rejection, to finish the small-noise upper bound.

4.5 Larger noise and covariance completion

If $\delta = W_1(p_f, p_g) \le 1$, their means are within one. Translate by the mean of g and apply (27) with $U = U_h(f) - U_h(g)$, $J = J_h(f) - J_h(g)$ on a ball of radius $D\sqrt{\log(1/h)}$. Take D large enough that the Gaussian exterior contributes $O(h^{\alpha+1})$. On this ball there is a uniform density floor h^P for some fixed P, and (25) gives an $O(h^{\alpha+1})$ weighted bound on the vector field, paying each local bias in its own density. Coupling the output laws bounds their characteristic function difference by $\delta|\xi|$. Fourier inversion on the cutoff support then gives $\|U\|_\infty \le C\delta h^{-d-1} e^{s/(2h^2)}$. Consequently

$$W_2(f, g) \le C\left[h^{\alpha+1} + \delta h^{-d-1-P/2} \log^{d/4}(1/h) e^{s/(2h^2)}\right]. \tag{32}$$

Uniformly over the class, the empirical observed W_1 error e_n has expectation $O(n^{-\beta})$ with $\beta = 1/[4(d+1)]$. For an elementary proof translate to the theoretical mean, truncate at distance n^β using the uniform second moment and then quantize at mesh $n^{-\beta}$. Both costs are $O(n^{-\beta})$. There are $O(n^{2d\beta})$ sites, so Cauchy–Schwarz for the Bernoulli counts bounds expected total variation by $Cn^{d\beta - 1/2}$. Multiplying by the grid diameter gives $O(n^{-1/4})$. Use the empirical observed W_1 distance between both samples, rejecting above $2n^{-\beta/2}$. Under the null, Markov's inequality and the triangle inequality give error $o(1)$. For any alternative whose observed discrepancy is $\delta \le 5n^{-\beta/2}$, use

$$h = \max(n^{-c_1}, \sqrt{s/(\gamma_1 \log n)})$$

in (32), with constants chosen so that $c_1(\alpha+2+d+P/2)+\gamma_1/2 < \beta/2$. The two terms in (32) are then $O(h^{\alpha+1})$. If $s \ge n^{-\eta}$ with $\eta > 0$ sufficiently small, the second entry of h dominates and $\log(2 + ns^{\alpha+d/4}) \asymp \log n$. Thus for a sufficiently large multiple of $r_n(s)$ in latent distance, the observed discrepancy exceeds $5n^{-\beta/2}$. If each empirical error is at most $n^{-\beta/2}$, the empirical discrepancy is greater than $3n^{-\beta/2}$ and the test rejects. The exceptional probability is $o(1)$ by the two Markov bounds.

The common covariance completion also applies in this range. Its conditional rejection probability given the original observations is an admissible test and has the same two error probabilities. A deterministic test compares that conditional probability to one half, doubling at most either error. The upper proof can use a smaller error target before that step.

4.6 Lower bounds from mixtures

Put one latent law equal to a centered isotropic Gaussian g of precision strictly between m and M, under both the null and the alternatives. If the covariances are equal, vary the group with sample size n. Otherwise vary the group whose covariance has largest eigenvalue s; its size lies between n and a fixed multiple of n. In the latter case the Gram bounds below hold with the actual varied sample size by decreasing the amplitude constants. The other group always has the same law and is independent, so adding it leaves total variation unchanged.

We will use finite families $f_\omega = g + \sum_{i=1}^N \omega_i q_i$ where ω_i are independent signs and each q_i has integral zero. Let G be the Gram matrix of the noisy score perturbations in $L^2(p_g)$:

$$G_{ij} = \int \frac{(q_i * \gamma_\Sigma)(q_j * \gamma_\Sigma)}{g * \gamma_\Sigma}.$$

For the uniform mixture of n-sample laws over signs, one plus the chi-square divergence to the null equals $\mathbb{E}_{\omega,\omega'}(1 + \omega^T G \omega')^n$. The factor inside the power is nonnegative as the inner product of two likelihood ratios.

Bounding its power by $\exp(n\omega^T G\omega')$ and integrating one sign vector gives $\mathbb{E}_\omega \exp(n^2\|G\omega\|^2/2)$ as an upper bound, using $\cosh(t) \le e^{t^2/2}$. Introduce a standard Gaussian vector to linearize the square and apply the same sign bound again. The result is at most

$$\det(I - n^2G^2)^{-1/2} \le \exp(Cn^2\|G\|_{\mathrm{HS}}^2) \qquad \text{if } n\|G\|_{\mathrm{op}} \le 1/2.$$

Thus a small bound on $n^2\|G\|_{\mathrm{HS}}^2$ makes the mixture total variation arbitrarily small. If every mixture member has $W_2(f_\omega, g) \ge r$, the sum of testing errors at separation r is at least one minus that total variation.

For transport separation we use a smooth test φ. Taylor expansion under an optimal coupling gives

$$\int \varphi(f_\omega - g) \le \|\nabla\varphi\|_{L^2(f_\omega)} W_2(f_\omega, g) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(f_\omega, g)^2. \tag{33}$$

If the pairing dominates cE, the gradient norm is at most $C\sqrt{E}$, and the Hessian norm is bounded, it follows that $W_2^2(f_\omega, g) \ge c'E$.

4.6.1 The direct polynomial lower bound

For $d > 4$, put $h = n^{-2/(4\alpha+d)}$. Pack order h^{-d} disjoint cubes of side h in a fixed cube about the reference mode. Let ψ be a nonzero smooth bump supported inside the reference unit cube with integral zero, and use $q_i(x) = a_0\psi((x - z_i)/h)$, $a_0 = \kappa h^\alpha$. The relative perturbations q_i/g have uniformly bounded C^α norm of order κ on the combined support, by rescaling and the fixed positive density floor there. The vanishing near edges gives the same bound across cubes. The logarithmic derivative bounds follow by composition with $\log(1 + \cdot)$; for sufficiently small κ the reference slack places every alternative in the class.

The clean score Gram matrix is diagonal with entries bounded by $Ca_0^2h^d$. Observed score Gram matrices are bounded above by their clean counterpart in positive operator order: the observed score is the conditional expectation of the clean score under the reference joint law, so this follows by conditional Cauchy–Schwarz for any linear combination. Thus $n^2\|G\|_{\mathrm{HS}}^2 \le C\kappa^4 n^2 h^{4\alpha+d} = C\kappa^4$. For separation take $\varphi = h^2\sum_i \omega_i q_i$. Its pairing with $f_\omega - g$ is of order $E = h^{d+2}\sum_i a_0^2 \asymp h^{2\alpha+2}$, by disjointness. The gradient's squared L^2 norm is $O(E)$, and its Hessian is uniformly bounded. Equation (33) therefore gives the direct lower bound for every sign.

4.6.2 The direct critical lower bound

When $d = 4$, pack disjoint cubes of side $\ell = (\log n)^{-1/2}$ inside a cube of side $c\sqrt{\log n}$ about the mode, with $c > 0$ small. Their total volume is order $(\log n)^2$. The reference g is comparable to its value g_B at the center of each cube B, and $g_B \ge n^{-\beta_0}$ for a fixed $\beta_0 < 1$. In the rescaled unit cube the ratio g_B/g has bounded C^α norm uniformly: on length ℓ, the score bound is paid by $\ell\sqrt{\log n} = 1$, and all higher potential derivatives are bounded with the corresponding scale factors. The top Hölder bound scales by ℓ^α.

Use a fixed orientation of compactly supported orthonormal tensor wavelets [7] of regularity C^{k+1}, supported strictly inside each cube. At physical length $r_i = \ell 2^{-j}$ their number in a cube is comparable to $(\ell/r_i)^4$, for j from a sufficiently large fixed index through $\lfloor \nu \log_2 n \rfloor$, with $\nu > 0$ small. Write ψ_i for these L^2-normalized wavelets. In particular $S_4 = \sum_i r_i^4 \asymp (\log n)^3$: each level contributes order the total volume. Set

$$b_i = \frac{\kappa^2 r_i^2}{n\sqrt{S_4}}, \qquad a_i = \sqrt{g_{B(i)} b_i}, \qquad q_i = a_i\psi_i.$$

Each wavelet has zero integral. At physical scale r, the relative perturbation in cube B has C^α norm at most $C(a_i/g_B)r^{-2-\alpha}$, by wavelet scaling, bounded overlap at the scale, and the rescaled reciprocal density bound. Summing over scales gives at most

$$C\kappa n^{-(1-\beta_0)/2} S_4^{-1/4} r_{\min}^{-1-\alpha} = o(1)$$

if $\nu < (1-\beta_0)/(2(1+\alpha))$. For fractional order this controls the global seminorm: on a pair of points in distinct cubes one may use the zero values of the top derivatives on the boundary of each cube and distances to the first exit points on the segment. Thus all perturbations stay within the reference margins by logarithmic composition.

The clean Gram matrix is $D_0 H D_0$ with $D_0 = \mathrm{diag}(\sqrt{b_i})$, and $H \preceq CI$. This uses orthonormality and the uniform bounds on g_B/g within each cube. The noisy Gram shrinks in positive operator order as before, so $n^2\|G\|_{\mathrm{HS}}^2 \le Cn^2 \sum_i b_i^2 = C\kappa^4$. For the metric test take $\varphi = \sum_i (r_i^2/g_{B(i)})\omega_i a_i \psi_i$. Its pairing is exactly $E = \sum_i r_i^2 b_i = \kappa^2\sqrt{S_4}/n$. The gradient squared norm in a cube is controlled by the sum of squared coefficient norms weighted by r_i^{-2}, up to a fixed factor. For clarity, within one level sum u_j, $\|\Delta u_j\|_2 \le C r_j^{-2}\|u_j\|_2$ by overlap. When $r_j \ge r_\ell$, integration by parts yields

$$|\langle \nabla u_j, \nabla u_\ell\rangle| \le C(r_\ell/r_j)(\|u_j\|_2/r_j)(\|u_\ell\|_2/r_\ell).$$

Summing the geometric ratios and multiplying by the comparable density in each cube gives squared gradient norm $O(E)$ under f_ω. The test Hessian is uniformly bounded, since its per cube scale sum is bounded by $C\kappa(n g_B)^{-1/2} S_4^{-1/4} r_{\min}^{-1}$. Thus (33) gives separation at least $c(\log n)^{3/4}/\sqrt{n}$. This lower bound holds under every known Gaussian covariance, by the same mixture contraction.

4.6.3 The inverse mixture

Rotate axes so $\Sigma = \mathrm{diag}(s, t_2, \dots, t_d)$ with $0 \le t_i \le s$, and suppose $0 < s \le s_0$ with $s_0 > 0$ small. Place $N \asymp s^{-d/2}$ centers z_i spaced $D\sqrt{s}$ apart in a fixed cube. Here D is sufficiently large, and s_0 small after choosing D. Put $T^2 = A_0 \log(2 + n s^{\alpha+d/4})/s$, and use

$$q_i(x) = a_0 \psi_i(x), \qquad a_0 = \kappa T^{-\alpha}, \qquad \psi_i(x) = e^{-|x-z_i|^2/(2s)} \left(\cos(T(x_1 - z_{i,1})) - e^{-sT^2/2}\right).$$

Each packet has zero mass. The relative perturbation before the factor a_0 has derivatives through order $j \le k$ bounded by CT^j. Indeed derivatives of the packets are bounded by $CT^j e^{-c|x-z_i|^2/s}$ because $T\sqrt{s}$ is large, and division by g costs only Gaussian and polynomial factors absorbed in a weaker envelope for small s_0. Separated grid envelopes have uniformly bounded sum. For $0 < \theta < 1$, the k-th derivative has θ-Hölder seminorm at most CT^α. For shifts longer than $1/T$ use the sup norm. For a shorter shift expand the product-rule terms of $D^k(\psi_i/g)$. A difference of a smooth packet derivative of order j costs $CT^{j+1}|u|$ with the same weakened envelope. A difference of a reciprocal derivative of order $k-j$ costs $C|u|^\theta$ times polynomial factors and the maximum of the two reciprocal densities, by the product rules and the reference Hölder condition on $D^k V_g$. The nearby packet envelopes still absorb those factors since $|u| \le 1/T \le C\sqrt{s}$. Summing the bounds gives $CT^\alpha |u|^\theta$. The logarithmic perturbation thus costs $O(\kappa)$ in the stipulated norms and preserves the class for small enough κ.

Convolution with the noise changes the packet envelope to variances $s + t_i$, with $t_1 = s$, multiplying its amplitude by $\prod_i \sqrt{s/(s+t_i)}$. Its bracket becomes

$$e^{-sT^2/4}\cos\big(\tfrac{T}{2}(y_1 - z_{i,1})\big) - e^{-sT^2/2}.$$

These formulas follow by conditional Gaussian integration and include zero eigenvalues in the other coordinates. The convolved reference has a lower Gaussian envelope, obtained by convolving one for g. Combining the two narrow packet envelopes and dividing by this lower envelope gives

$$|G_{ij}| \le C a_0^2 s^{d/2} e^{-csT^2} e^{-c|z_i - z_j|^2/s}.$$

Indeed the common center lies in a fixed cube, the variances of the two envelopes are between s and $2s$, and choosing s_0 small absorbs the reciprocal reference into a weaker Gaussian. Thus

$$n\|G\|_{\mathrm{HS}} \le C n a_0^2 s^{d/4} e^{-csT^2}.$$

With $X = n s^{\alpha+d/4}$, the right-hand side is at most $C\kappa^2 X (A_0 \log(2+X))^{-\alpha}(2+X)^{-cA_0}$. It is arbitrarily small by choosing A_0 large and κ small.

For transport separation take $\varphi = T^{-2}\sum_i \omega_i q_i$. The spatial L^2 Gram matrix of the ψ_i has diagonal at least $cs^{d/2}$ by integrating the squared cosine in its Gaussian envelope and taking $T\sqrt{s}$ large. The absolute off-diagonal row sums are at most $Cs^{d/2}\sum_{j \in \mathbb{Z}^d \setminus \{0\}} e^{-cD^2|j|^2}$, less than half the diagonal bound for large D. The pairing with $f_\omega - g$ is at least cE for $E = T^{-2} a_0^2 s^{d/2} N$. The gradient Gram has absolute row sums at

most $CT^2 s^{d/2}$ by the same overlap estimate, so the test gradient has squared norm $O(E)$ under the uniformly bounded vertex densities. The pointwise derivative envelope bounds the test Hessian by Ca_0. Equation (33) gives $W_2(f_\omega, g) \geq cT^{-\alpha-1}$ for every sign.

For $s \geq s_0$, use instead one additive packet throughout, replacing its envelope variance by a fixed τ_0^2 smaller than the variance in the Gaussian lower envelope for g, with its cosine mean removed. Use amplitude $\kappa T^{-\alpha}$ and $T^2 = C \log n/s$. The derivative and Hölder argument above applies to this single packet. Its convolved Gaussian envelope squared, divided by the convolved reference lower envelope, is integrable uniformly in Σ by the variance choice. Its first-coordinate cosine attenuates by $\exp(-\tau_0^2 s T^2/[2(\tau_0^2 + s)])$, so the n-sample chi-square tends to zero uniformly for sufficiently large C. Testing the latent difference with $\cos(Tx_1)/T$ gives separation of order $T^{-\alpha-1}$. This matches the inverse term since its logarithm is comparable to $\log n$ on this range.

4.6.4 Lower bounds in dimensions one through three

In dimension one a shift of the reference by δ has W_2 distance $|\delta|$. Its clean one-sample relative entropy to the reference is $O(\delta^2)$, by expanding the potential and using the zero mean score. Since noise contracts entropy, $\delta = c/\sqrt{n}$ gives the desired two-point lower bound.

For $d = 2, 3$, take disjoint annuli at dyadic radii R from a fixed constant to $c\sqrt{\log n}$, for small fixed $c > 0$. Pack cubes of side $\ell_i \asymp R^{-1}$ in each annulus, with order R^{2d} cubes at radius R. The reference values g_i at their centers satisfy $g_i \geq n^{-\beta_0}$, $\beta_0 < 1$, and $g \asymp g_i$ inside each cube. Use one L^2-normalized smooth mean-zero bump ψ_i per cube, with

$$S_4 = \sum_i \ell_i^4, \qquad b_i = \frac{\kappa^2 \ell_i^2}{n\sqrt{S_4}}, \qquad q_i = \sqrt{g_i b_i}\, \psi_i.$$

The rescaled reciprocal bounds used for the critical construction apply here since $\ell_i(1 + |x|) = O(1)$. All relative C^α perturbations tend to zero because $g_i \geq n^{-\beta_0}$ and $\ell_i \gtrsim (\log n)^{-1/2}$. The clean Gram matrix is diagonal with entries $O(b_i)$, and the noisy Gram contracts. Hence $n^2 \|G\|_{\mathrm{HS}}^2 = O(\kappa^4)$. The test $\varphi = \sum_i (\ell_i^2/g_i)\omega_i q_i$ has pairing $\sum_i \ell_i^2 b_i$, squared gradient norm at most a constant times that pairing, and bounded Hessian by rescaling on the disjoint cells. Equation (33) gives squared separation of order $\sqrt{S_4}/n$. A ring contributes order R^{2d-4} to S_4; summing gives order $\log\log n$ for $d = 2$ and $\log n$ for $d = 3$.

For the inverse lower bound in all three dimensions, the single additive narrow packet with $T^2 = C \log n/s$ works throughout $0 < s \leq \Lambda$. Its attenuated cosine multiplier is bounded by e^{-csT^2} uniformly for the fixed narrow variance, and the noisy reference denominator is paid as above. It gives latent separation of order $(s/\log n)^{(\alpha+1)/2}$ with small total variation. Together with the direct lower bounds this completes Theorem 4.1.

4.7 Relation to direct and indirect two-sample tests

Lee, Lahiri, and Sinha [17] test equality of latent distributions under measurement error through deconvolved characteristic functions and bootstrap calibration. Their framework addresses different guarantees from the uniform Wasserstein separation of Theorem 4.1. The direct metric-space results of [9] allow rough distributions, with additional bounds for unbounded exponential tails. Manole [16] reports a direct testing gap without smoothness. Tang and Yang [13] treat smooth adversarial losses on compact domains. The sliced statistic in [15] has a parametric separation in the powered sliced loss in its matching regimes. Bounded potential curvature on the whole space supplies both the inward density extrapolation and the weighted transport comparison here. The low-dimensional logarithms arise from the number and scale of distinguishable perturbations in the tails, with an additional accumulation over resolution at dimension four.

5 Honest adaptive Wasserstein confidence sets under Gaussian contamination

We characterize adaptation of honest quadratic Wasserstein confidence sets over two smoothness levels for strongly log-concave distributions observed with known, possibly vanishing or singular

Gaussian noise. The sharp diameter on the smoother class is the larger of its estimation rate and a testing rate at the rougher smoothness. On the rougher class the diameter can simultaneously attain its estimation rate. The proof uses a two-sided comparison between Wasserstein distance and locally weighted heat fields. It lets a held-out quadratic statistic estimate the loss of a random fitted density despite Gaussian tails and the inverse problem. We give the exact noise and smoothness conditions for both classes to attain their respective estimation rates.

5.1 Statement of the adaptation theorem

For $\alpha = k + \theta \geq 2$, where k is integer and $0 \leq \theta < 1$, let $\mathcal{F}_\alpha$ be a translated class of probability densities $f = e^{-V}$ on $\mathbb{R}^d$. Require $V \in C^k$, $mI \preceq D^2 V \preceq MI$ for fixed $0 < m < M$, and fixed positive supremum bounds on derivatives of orders $3, \ldots, k$. For $\theta > 0$ also require a fixed positive global θ-Hölder bound on $D^k V$. Fix $2 \leq \alpha_0 < \alpha_1$, with these bounds chosen so that $\mathcal{F}_{\alpha_1} \subseteq \mathcal{F}_{\alpha_0}$. The precision of a common centered Gaussian can be chosen strictly between m and M. The observations are independent with law $f * \gamma_\Sigma$, where $\gamma_\Sigma = N(0, \Sigma)$, $\Sigma \succeq 0$ is known, and $s = \|\Sigma\|_{\mathrm{op}} \leq \Lambda$.

Write the clean estimation and testing scales as follows:

d	$A_{\alpha,d}$	$T_{\alpha,d}$
1	$(\log\log n)^{1/2}/\sqrt{n}$	$n^{-1/2}$
2	$\log n/\sqrt{n}$	$(\log\log n)^{1/4}/\sqrt{n}$
3	$n^{-(\alpha+1)/(2\alpha+3)}$	$(\log n)^{1/4}/\sqrt{n}$
4	$n^{-(\alpha+1)/(2\alpha+4)}$	$(\log n)^{3/4}/\sqrt{n}$
$d > 4$	$n^{-(\alpha+1)/(2\alpha+d)}$	$n^{-2(\alpha+1)/(4\alpha+d)}$

Define $L_E(\alpha, s) = \log n$ when $d = 1$ and $\log(2 + ns^{\alpha+d/2})$ otherwise. Define $L_T(\alpha, s) = \log n$ when $d < 4$ and $\log(2 + ns^{\alpha+d/4})$ otherwise. Put

$$R_j = A_{\alpha_j,d} + (s/L_E(\alpha_j, s))^{(\alpha_j+1)/2}, \qquad r_0 = T_{\alpha_0,d} + (s/L_T(\alpha_0, s))^{(\alpha_0+1)/2}. \tag{34}$$

The inverse terms vanish when $s = 0$. We consider confidence sets in $\mathcal{F}_{\alpha_0}$ with measurable membership and diameter. An empty set is assigned diameter zero. Randomization independent of the observations is allowed.

Theorem 5.1. *Let $0 < \zeta < 1/4$ and $0 < \xi < 1$. For sufficiently large n, there exist confidence sets $\mathcal{C}_n$ with $\inf_{f \in \mathcal{F}_{\alpha_0}} \mathbb{P}_f(f \in \mathcal{C}_n) \geq 1 - \zeta$, and, for a constant $C_{\zeta,\xi}$,*

$$\sup_{f \in \mathcal{F}_{\alpha_0}} \mathbb{P}_f(\mathrm{diam}_{W_2} \mathcal{C}_n > C_{\zeta,\xi} R_0) \leq \xi,$$
$$\sup_{f \in \mathcal{F}_{\alpha_1}} \mathbb{P}_f(\mathrm{diam}_{W_2} \mathcal{C}_n > C_{\zeta,\xi}(R_1 + r_0)) \leq \xi.$$

The same sets can be used for each fixed ξ, by changing only the constants $C_{\zeta,\xi}$. Conversely, for each fixed $\xi < 1/2 - \zeta$ and all sufficiently large n, every honest set has worst-case diameter at least a small constant times R_0 on the rough class and $R_1 + r_0$ on the smooth class, with probability greater than ξ at some member of the respective class. The r_0 lower bound holds at a common interior Gaussian. All bounds are uniform in the allowed known covariances.

Theorem 5.2. *Along a sequence of known covariances, the simultaneous estimation diameters R_0, R_1 are attainable if and only if the clean condition $T_{\alpha_0,d} = O(A_{\alpha_1,d})$ and the noise condition below both hold. The clean condition is automatic for $d \leq 4$. For $d > 4$ it reads*

$$(d-4)\alpha_1 \leq (2d-4)\alpha_0 + d.$$

When this inequality is strict, or $d \leq 4$, the noise condition is

$$s = O\left(A_{\alpha_1,d}^{2/(\alpha_0+1)} \log n\right).$$

At equality for $d > 4$, it is $s = O(n^{-1/(\alpha_0+d/4)})$.

Deo and Randrianarisoa [8] obtain the direct compact adaptation boundary in W_2 for densities bounded below on the torus, and a whole-space theory in W_1. The fixed positive Gaussian inverse problem in the present curvature setting is studied in [12]. Here the diameter calculation requires resolving both the sampling and inverse regimes in W_2 on the whole space. In particular, at the boundary of the clean adaptation window the allowable Gaussian variance loses a logarithmic factor.

5.2 Density bounds and a weighted transport comparison

Uniformly over the class, f has Gaussian upper and lower envelopes about its mean m_f, the score of f is $O(1 + |x - m_f|)$, and

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

Indeed quadratic bounds at the mode give the envelopes and bound the distance between the mode and the mean. Differentiating e^{-V} gives the derivative bounds. For $\theta > 0$, on $|x - m_f| \le L$ and $|u| \le (2 + L)^{-1}$, the same product rule and the Hölder bound on $D^k V$ give

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

The local density ratio is bounded by the score, derivatives of lower order in the product are Lipschitz, and the highest one uses the assumed Hölder seminorm.

For $p_f = f * \gamma_s$, isotropic $0 \le s \le \Lambda$, Gaussian integration of the two quadratic Taylor bounds gives

$$c f(x) \le p_f(x) \le f(x) e^{Cs(1+|x-m_f|^2)}. \tag{35}$$

The potential of p_f also has Hessian in a fixed positive interval. For $s > 0$, its Hessian is $s^{-1} I - s^{-2} \operatorname{Cov}(X \mid Y)$. The conditional potential has Hessian between $(m + s^{-1})I$ and $(M + s^{-1})I$. The upper covariance bound follows from Brascamp–Lieb [5]. The lower bound follows, in each unit direction v, from Cauchy–Schwarz and the conditional identities

$$\mathbb{E}[(v \cdot (X - \mathbb{E}X))(v \cdot \nabla A(X))] = 1, \qquad \mathbb{E}[(v \cdot \nabla A(X))^2] = \mathbb{E}[v^T D^2 A(X) v],$$

where A is the conditional potential. Consequently the scores of p_f are $O(1 + |x - m_f|)$. The observed law always has a uniformly Gaussian upper envelope about m_f.

Use $\widehat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\, dx$. Let $\widehat{K}$ be a real even smooth Fourier cutoff that equals one near zero and is supported in the unit ball. Use $K_b = b^{-d} K(\cdot/b)$ for its spatial kernel, and define constant-bandwidth fields

$$U_b(f) = K_b * f, \qquad \widehat{J_b(f)}(\xi) = -\frac{i\xi}{|\xi|^2}(1 - \widehat{K}(b\xi))\widehat{f}(\xi).$$

Then $f = U_b(f) + \operatorname{div} J_b(f)$. The inverse transform Q of $-i\xi(1 - \widehat{K}(\xi))/|\xi|^2$ is integrable with all absolute moments finite and all signed moments zero. In fact a dyadic partition into shells 2^j away from the origin gives inverse transforms bounded by $C_N 2^{(d-1)j}(1 + 2^j|x|)^{-N}$. Summability of their L^1 norms $O(2^{-j})$ gives the assertion, with moment cancellation since the symbol vanishes near zero. For $|x - m_f| \le L$, $w = (2+L)^{-1}$, and $h \le w$, subtract the local Taylor polynomial of f to degree $k - 1$ in the integer case and k otherwise. This gives

$$|J_h(f)(x)| \le C h^{\alpha+1}(1 + |x - m_f|^{k+1}) f(x) + C_N h (h/w)^N. \tag{36}$$

Indeed $J_h(f) = h Q_h * f$, the short-shift remainder has order α, and the farther shifts use uniform absolute derivative bounds and the arbitrary integrated tails.

Choose the origin within a fixed distance of the means of f, g, and put $\rho = f + g$. For a test φ, set $u = \varphi - \frac{1}{2}\int \varphi\rho$. Then

$$\int (1 + |x|^2)|u|^2 \rho \le C \int |\nabla\varphi|^2 \rho. \tag{37}$$

Both components satisfy ordinary Poincaré by Brascamp–Lieb. Comparing their component means to a common Lebesgue average on a fixed ball with a positive density floor gives the mixture inequality. For either potential V, $x \cdot \nabla V(x) \ge m|x|^2/2 - C$, since its mode stays close to zero. Integration by parts gives

$$\int u^2 x \cdot \nabla V f = d \int u^2 f + 2 \int u x \cdot \nabla\varphi f.$$

Use Cauchy–Schwarz to absorb half of the resulting quadratic term, add the two estimates, and apply mixture Poincaré. Gaussian tails justify integration by parts for compactly supported tests minus their mean.

Let $I_0 \subset I$ admit a C^1 cutoff supported in I, equal to one on I_0, with $|\nabla\chi| \le C(1+|x|)$. For any identity $f - g = U + \operatorname{div} J$ in distributions,

$$W_2(f,g) \le C\left[\left(\int_{I_0^c} \frac{\rho}{1+|x|^2}\right)^{1/2} + \left(\int_I \frac{|U|^2/(1+|x|^2) + |J|^2}{\rho}\right)^{1/2}\right]. \tag{38}$$

To see this, split $\int (f-g)u$ with the cutoff and use Cauchy–Schwarz with (37) in the exterior. In the interior insert the field identity and integrate by parts. Approximation by smooth tests justifies this for a C^1 cutoff. The $L^2(\rho)$ norms of $(1+|x|^2)^{1/2}\chi u$ and $\nabla(\chi u)$ are at most $C\|\nabla\varphi\|_{L^2(\rho)}$. Riesz representation produces an exact flux of divergence $f-g$ whose squared action norm $\int |J|^2/\rho$ is bounded by the square of the displayed right-hand side. Interpolate between the densities with $(1-\lambda)f + \lambda g$, $\lambda(t) = \sin^2(\pi t/2)$, and flux $\lambda'(t)J$. Its action is at most $\pi^2 \int |J|^2/\rho$, by the lower bound $\min(\lambda, 1-\lambda)\rho$. The dynamic transport inequality [4] gives (38).

5.2.1 The reverse transport comparison and the heat fields

Use $D(x) = 1 + |x|^2$ and suppose the origin is within fixed distance of the two means. Let

$$\mathcal{H}(f,g) = \sup_{\int |\nabla\varphi|^2 \rho \le 1} \left| \int \varphi (f-g) \right|, \qquad \rho = f + g,$$

where the supremum is over smooth compactly supported tests. The Riesz and dynamic argument used for (38) gives $W_2(f,g) \le C\mathcal{H}(f,g)$. We prove a converse for the endpoint mixture weight. For related general comparisons using intermediate or bounded densities, see Peyre [14].

Set $\ell(x) = \epsilon/(1+|x|)$ for fixed small $\epsilon > 0$. For $|x-y| \le c\min(\ell(x), \ell(y))$, the Euclidean ball-average inequality gives

$$|\varphi(x) - \varphi(y)| \le C|x-y|\big(M_{\mathrm{loc}}|\nabla\varphi|(x) + M_{\mathrm{loc}}|\nabla\varphi|(y)\big),$$

where M_{loc} takes centered ball averages with radii at most $C_0\ell$. Indeed telescope ball averages from radius $|x-y|$ to zero at both endpoints using the L^1 ball Poincaré inequality, and compare the two initial averages on a common enlarged ball. The localized maximal operator is bounded on $L^2(\rho)$. To see this, cover with fixed enlargements of a Vitali selection of pairwise disjoint balls of radii $c_0\ell(x)$. The enlarged balls have bounded overlap by radius comparability when they meet and disjointness of the original balls. Choose ϵ small for the enlargement factors. On each enlarged ball all the relevant radii are comparable and ρ is comparable to its center value by the score bound. Each local maximal function is dominated there by the usual maximal function of the input restricted to a further fixed enlargement, and the usual L^2 bound and overlap yield the assertion.

For the other pairs, both $\ell(x)$ and $\ell(y)$ are bounded by $C|x-y|$, since ℓ is Lipschitz and the smaller radius is at most $|x-y|/c$. Subtract the mixture mean of φ, writing the result as u. Then

$$|\varphi(x) - \varphi(y)| \le C|x-y|\big(\sqrt{D(x)}|u(x)| + \sqrt{D(y)}|u(y)|\big).$$

Integrate both inequalities against an optimal coupling and use Cauchy–Schwarz, the local maximal bound and (37). This proves $\mathcal{H}(f,g) \le CW_2(f,g)$. In particular there is a vector flux L with $\operatorname{div} L = f - g$ and $\|L/\sqrt{\rho}\|_2 \le CW_2(f,g)$ by Riesz representation. The flux is integrable by weighted Cauchy–Schwarz.

Take

$$b(x)^2 = \delta/D(x), \qquad R_t = \tfrac{1}{2}\int_0^t \nabla\gamma_z \, dz,$$

with a sufficiently small fixed $\delta > 0$ in every dimension. For $q = f - g$ let $p_t = q * \gamma_t$, and define

$$U(q)(x) = p_{b(x)^2}(x) + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla p_{b(x)^2}(x), \qquad J(q)(x) = (R_{b(x)^2} * q)(x).$$

Gradients on p_t are at fixed t. Direct differentiation gives $q = U(q) - \operatorname{div} J(q)$. We claim

$$\left\| U(q)/\sqrt{D\rho} \right\|_2 + \|J(q)/\sqrt{\rho}\|_2 \le C W_2(f,g). \tag{39}$$

To prove (39), apply the row kernels to $\operatorname{div} L$ and write $L = \sqrt{\rho}\, H$. The conjugation factor is $\sqrt{\rho(y)/\rho(x)}$ at $z = x - y$. Globally

$$|\log \rho(y) - \log \rho(x)| \le C\big((1+|x|)|z| + |z|^2\big),$$

by the mixture score bound along the segment. With δ small this factor can be absorbed into every row Gaussian. The scalar kernel is bounded by $Cb(x)^{-d} e^{-c|z|^2/b(x)^2}$. Here $D^{-1/2}$ pays the first factor b^{-1} from integration by parts, and $|\nabla b^2| \le C_\delta b^3$ pays the derivative correction. Its row and column integrals are uniformly bounded. For the column integral split at $|x-y| \le c_1(1+|y|)$ for fixed small c_1, where $b(x) \asymp b(y)$. On the other part $|x-y|$ controls both $1+|x|$ and $1+|y|$, and the Gaussian pays every polynomial prefactor. Schur's inequality bounds the scalar norm by $C\|H\|_2$.

For the vector term first integrate from $\varepsilon b(x)^2$ to $b(x)^2$, $0 < \varepsilon < 1$, and integrate by parts against L at fixed row. Without conjugation the operator consists of two Gaussian averages, at these two row times, of a fixed matrix of Riesz products applied to H. This follows from the multiplier $\xi_i\xi_j/|\xi|^2$ times the difference of the Gaussian symbols. Its L^2 norm is bounded independently of ε, by the L^2 Fourier bound for the matrix and the maximal bound for Gaussian averaging. Off the kernel diagonal, the absolute derivative of the truncated R-kernel is at most $C|z|^{-d}\exp(-c|z|^2/b(x)^2)$, by integrating Gaussian second derivatives over the heat times. The error from conjugation has an extra factor $\sqrt{\rho(y)/\rho(x)} - 1$. By the log-ratio bound this yields an absolute kernel bounded by

$$Cb(x)^{-1}|z|^{1-d}\exp(-c'|z|^2/b(x)^2).$$

The same column split and direct row integration show a uniform Schur bound. The truncated convolutions tend pointwise to $J(q)$: q is uniformly bounded and $\|R_{\varepsilon b(x)^2}\|_1 \le C\sqrt{\varepsilon} b(x)$. Fatou's lemma gives the required weighted vector bound, proving (39).

5.3 Estimating the common weights from inward probes

We analyze the observation laws after completing to covariance sI. For validation fix a candidate $g \in \mathcal{F}_{\alpha_1}$ supplied by independent training. Use the same scale definitions with $\alpha = \alpha_0$ and $r \asymp r_0$:

$$a = \log n, \quad w = a^{-1/2}, \quad h_0 = \max\{T_{\alpha_0,d}^{1/(\alpha_0+1)}, \sqrt{s/(\gamma L_T(\alpha_0,s))}\}, \quad r = h_0^{\alpha_0+1}, \quad \rho_0 = r^2 a^{2-d/2}.$$

We restrict for now to $s \le n^{-\eta}$ with small fixed $\eta > 0$. A first validation split gives a sample mean. If its distance from the mean of g exceeds one, use only the rough confidence ball constructed below. Otherwise take the first validation mean as origin. Except with uniform probability $o(1)$, both true means are then within a fixed distance of the origin. Conditional on the first split and the fitted candidate, use an independent second validation split for the pilot below.

Write $\rho = f + g$ and $p = p_f$. Tile with w-cubes, retaining all cubes meeting the ball of radius $L = D_0\sqrt{a}$. There are $O(a^d)$ cubes. Write x for a cube center. Use the exact value of the candidate observed density p_g at each center, and estimate $p_f(x)$ at inner rows $|x| \le c_1\sqrt{a}$ using its second sample count in the cube divided by the second sample size times w^d. For outer rows put $\omega = x/|x|$, $t = C_1 \log a/\sqrt{a}$, and count in the w-cubes centered at $x - t\omega$ and $x - 2t\omega$. Each count of at least $A_1 \log a$ yields a density estimate at its probe by the same division. If both counts attain this threshold, set the estimate at x to the square of the first probe estimate divided by the second. Otherwise set it to zero.

Uniformly with probability $1 - o(1)$, every count above the threshold agrees with its mean up to fixed factors and every count of mean at least $2A_1 \log a$ is above the threshold. This follows from binomial Chernoff bounds and a union bound over $O(a^d)$ probes with A_1 large. A probe cube average agrees with its center value up to fixed factors because the observed score on the grid region is $O(\sqrt{a})$. The inner counts have polynomially large means, by the uniform Gaussian lower envelope with c_1 sufficiently small.

On the outer segments, the derivative of $-\log p$ in direction ω is at least $c\sqrt{a}$. Indeed the Hessian is bounded below and the mode is within a fixed distance of the origin. Each inward probe value is therefore at least $a^{cC_1} p(x)$. Since

$$nw^d\rho_0 = nr^2a^{2-d} \gtrsim a^{-1/2},$$

every outer center with $p(x) \geq \rho_0$ has two reliable probes if C_1 is large. The Hessian upper bound gives

$$\left| \log p(x) - 2 \log p(x - t\omega) + \log p(x - 2t\omega) \right| \leq Ct^2 = o(1).$$

Thus every positive estimate $\widehat{p}(x)$ is comparable to $p(x)$, and a zero estimate implies $p(x) < \rho_0$. No independence of counts within a sample is needed.

Let I be the union of the retained cubes for which the sum of the pilot estimate and the candidate value at the center is above $K_1\rho_0$, with K_1 a large fixed constant, and let $\widehat{\rho}$ be that sum extended constantly on each selected cube. Throughout the successful pilot event,

$$c\rho(x) \leq \widehat{\rho}(x) \leq C\rho(x) \quad (x \in I), \qquad \{\rho > C_2\rho_0\} \subseteq I. \tag{40}$$

Indeed the ignored terms at a selected center can contribute at most $2\rho_0$ in observed density. Center values and cube values are comparable by the score bounds, and the latent and observed values are comparable there by (35) and $sa = o(1)$. Every center with large enough density has a positive estimate of the same order. Choose D_0 large so that beyond the tiled ball the Gaussian upper bound is below ρ_0. This establishes the inclusion.

For the analysis take a smooth cutoff as a scalar function of ρ/ρ_0, zero below C_2 and one above $2C_2$. Its support is in I (increase its lower threshold if needed) and its gradient is bounded by $C(1 + |x|)$. For each component density $u = f, g$ integrate on its potential levels about its mode. On a level z above the minimum, the radius and radial derivative are comparable to $\sqrt{z}$. The level-volume derivative is $O(z^{d/2-1})$ and $1 + |x|^2 \gtrsim z$ on large levels. Since $\log(1/\rho_0) \asymp a$, this proves

$$\int_{\{\rho \leq 2C_2\rho_0\}} \frac{\rho(x)}{1 + |x|^2}\, dx \leq C\rho_0 a^{d/2-2} = Cr^2.$$

Here the domain is contained in the individual low-density domain for each component. Only the analytical cutoff uses the unknown densities.

5.4 The validation fields

Use $b(x)^2 = \delta/D(x)$ as in (39), and let $v_x = b(x)^2 - s \asymp b(x)^2$ on I. Use $h(x) = h_0$ if $d \leq 4$, and $h(x) = h_0 e^{c_0|x|^2}$ if $d > 4$. Choose $c_0 > 0$ sufficiently small so that h remains polynomially small on I and so that $e^{(2\alpha_0+2)c_0|x|^2}$ has uniformly bounded polynomial moments under the two latent laws. Then $h(x)^2 \ll v_x$ and $h(x) \asymp h(x')$ for $|x - x'| \leq 2w$.

For an observed y use scalar and vector kernels at row x applied at $x - y$. The scalar kernel is

$$G_x = \gamma_{v_x} + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla\gamma_{v_x},$$

divided by $(D(x)\widehat{\rho}(x))^{1/2}$. The vector kernel is the sum of $R_{v_x} - R_{h(x)^2}$ and the kernel with symbol

$$\frac{i\xi}{|\xi|^2}\left(\widehat{K}(h(x)\xi)e^{s|\xi|^2/2} - e^{-h(x)^2|\xi|^2/2}\right). \tag{41}$$

Truncate the latter kernel to $|x - y| \leq w$ and divide the sum by $\widehat{\rho}(x)^{1/2}$. Denote the resulting feature, in $L^2(I) \oplus L^2(I; \mathbb{R}^d)$, by F_y.

Conditional on the pilot, the scalar expectation is the normalized $U(u)$. Holding row bandwidths fixed, the untruncated vector expectation has symbol $i\xi(\widehat{K}(h(x)\xi) - e^{-b(x)^2|\xi|^2/2})/|\xi|^2$ on the latent law, before normalization. It equals $J(u)$ plus the high-pass remainder from (36). That remainder has normalized L^2 norm $O(r)$ by (40), $u^2/\widehat{\rho} \leq Cu$, and the Gaussian moment choice of c_0.

For the fine kernel, rescale by $h = h(x)$ to obtain $h^{1-d}L_A((x-y)/h)$, $A = s/h^2$. The symbol after scaling has a smooth extension through zero, since the bracket is radial analytic there and vanishes to second order. For each fixed number of Schwartz seminorms the bound for L_A is $C(1 + A)^C e^{CA}$, by compact support of $\widehat{K}$ and Gaussian decay of the second term. Since h is polynomially small, $A \leq \gamma L_T(\alpha_0, s) = O(\log n)$, and $\widehat{\rho} \gtrsim \rho_0$ on I, the truncation costs $O(n^{-T})$ in normalized L^2 for every fixed T. With $\mu = \mathbb{E}_f F_Y - \mathbb{E}_g F_Y$, the cutoff, (40), the bias bounds and the two field comparisons (38) and (39) give

$$r + W_2(f, g) \asymp r + \|\mu\|. \tag{42}$$

5.4.1 Conditional covariance

We work throughout on the successful pilot event. It suffices to bound the Hilbert–Schmidt norms of the three uncentered second moment operators (scalar, coarse vector, fine vector). Their sum, multiplied by three, dominates the centered covariance in positive operator order by Cauchy–Schwarz, and the Hilbert–Schmidt norm is monotone on positive operators. Fix either completed input law with density p. Its supporting log-density tangent at each row and (40) bound

$$\frac{p(y)}{\sqrt{\widehat{\rho}(x)\widehat{\rho}(x')}} \le C\exp(C(1+|x|)|y-x| + C(1+|x'|)|y-x'|)\,.$$

Each factor is absorbed in the row Gaussian, since $b(x)(1+|x|) \le C$. It is bounded on the fine support.

In absolute value the tilted scalar kernel before division by $\sqrt{D(x)}$ is bounded by

$$Cb(x)^{-d}e^{-c|x-y|^2/b(x)^2}.$$

This includes the derivative correction since $|\nabla b(x)^2| \le Cb(x)^3$ in the variable case. The tilted coarse vector has absolute kernel bounded by

$$C(h(x)+|x-y|)^{1-d}e^{-c|x-y|^2/b(x)^2}.$$

For $|x-y| \le h(x)$ integrate the time powers above $h(x)^2$. At larger distances, keep half the Gaussian at scale $b(x)$ and extend the remaining time integral to infinity.

For $d<4$, at pairs with $1+|x| \asymp 1+|x'|$ (ratio between $1/2$ and 2), use common bandwidth $b(x)$. The scalar covariance kernel has squared row integral bounded by $Cb(x)^{-d}D(x)^{-2}$. For the coarse vector use $C|z|^{1-d}e^{-c|z|^2/b(x)^2}$, whose $L^{4/3}$ norm is $O(b(x)^{1-d/4})$; Young's inequality gives the same row bound $Cb(x)^{4-d}$. For noncomparable pairs, the separation is comparable to the larger radius. Extract a Gaussian in separation using both row Gaussians with bounded bandwidth. Scalar prefactors then grow at most polynomially in the radii. For the vector, split into balls of radius one third of the separation about either row and their complement. On each ball the other singularity is bounded by a polynomial in separation and the singularity at the ball center is integrable against its Gaussian. On the complement both singularities are bounded. The two squared kernels are thus integrable on this pair region uniformly in n. Integrating the comparable row bound $Cb(x)^{4-d}$ on I gives $O(1), O(\log a), O(a)$ in dimensions one, two, three respectively.

For $d \ge 4$, the scalar argument on comparable and noncomparable radii above again gives squared norm bounded by $C(1+\int_I b(x)^{4-d}\,dx)$. This is $O(a^2)$ in dimension four and polylogarithmic above four. For the coarse vector, the convolution bound at $z=x-x'$, $|z| \le 2w$, is

$$C(h(x)+|z|)^{2-d}e^{-c|z|^2}.$$

The two regularizations are comparable there. For $|z| \le h(x)$ apply Cauchy–Schwarz to the two profiles. Above that scale, split near either singular center and in the remainder to obtain $C|z|^{2-d}$ before the Gaussian: the two radial integrals are bounded by $|z|^{1-d}\int_0^{|z|} dt$ and $\int_{|z|}^\infty t^{1-d}\,dt$. The same unregularized bound applies at $|z|>2w$. In dimension four the squared near integral in z is $O(a)$, and the far contribution per row is $O(1+\log(1/w))$. These give $O(a^3)$ after integration in x. For $d>4$ the squared near bound is $C\int_I h(x)^{4-d}\,dx \le Ch_0^{4-d}$. The far bound is polylogarithmic and is absorbed by h_0^{4-d}.

The fine covariance vanishes at $|x-x'|>2w$. With $A_0 = s/h_0^2$ and $E = C(1+A_0)^C e^{CA_0}$, its absolute kernel on the near pairs is bounded by $CE^2h(x)^{2-d}(1+|z|/h(x))^{-N_0}$ for fixed arbitrarily large N_0. This follows by convolving the two scaled Schwartz profiles at comparable bandwidths. The squared norm is therefore bounded by $CE^4\int_I h(x)^{4-d}\,dx$. For $d<4$ this is $o(1)$ since h_0 is polynomially small and $A_0 \le \gamma\log n$. For $d=4$ it is $O(E^4a^2)$, and above four it is $O(E^4h_0^{4-d})$.

For $d>4$, set $X = ns^{\alpha_0+d/4}$. For sufficiently small γ we have $E^2 \le C(2+X)^c$ with $c<1$, while

$$nh_0^{2\alpha_0+d/2} \ge \max\left(1, \frac{X}{(\gamma\log(2+X))^{\alpha_0+d/4}}\right).$$

This pays the fine term $E^2 h_0^{2-d/2}/n$ by Cr^2. The coarse terms divided by n are smaller than $T_{\alpha_0,d}^2$. For $d = 4$, set $X = 2 + ns^{\alpha_0+1}$. Choose γ so $E^2 \le CX^c$ with $c < 1/4$. If $X \le a^2$, then $E^2 a/n \le Ca^{3/2}/n$. Otherwise $E^2 a \le CX^{1/2+c} \le C(X-2)/(\log X)^{\alpha_0+1}$, so the fine term is at most Cr^2. The coarse term divided by n is $O(a^{3/2}/n)$. The low-dimensional bounds above agree with $nT_{\alpha_0,d}^2$ after taking square roots. Consequently, in every dimension,

$$n^{-1} \| \operatorname{Cov}(F_Y) \|_{\mathrm{HS}} \le Cr^2 \tag{43}$$

for either input law.

5.4.2 Estimating the validation loss

Use the last independent validation split of size $N \asymp n$. Let $Z_i = F_{Y_i} - \mathbb{E}_g F_Y$, where the second term is computed under the known fitted candidate and the completed noise. Use

$$\mathcal{U} = \frac{1}{N(N-1)} \sum_{i \ne j} \langle Z_i, Z_j \rangle.$$

Its conditional mean is $t^2 = \|\mu\|^2$ and, with $C_f = \operatorname{Cov}_f(F_Y)$, its variance is

$$\frac{4}{N} \langle \mu, C_f \mu \rangle + \frac{2}{N(N-1)} \|C_f\|_{\mathrm{HS}}^2.$$

This follows from the orthogonal linear and centered quadratic terms in the expansion. Thus on the pilot event the variance is bounded by $C(r^4 + r^2 t^2)$. For every fixed error probability, Chebyshev's inequality gives $|\mathcal{U} - t^2| \le C(r^2 + rt)$ with at least the complementary probability. Young's inequality and (42) give

$$r + \sqrt{(\mathcal{U})_+} \asymp r + W_2(f, g) \tag{44}$$

on that event, with constants depending on the error probability. The pointwise pilot bound works for every g in the class conditional on an independent training sample. The covariance does not include a candidate-sampling term because the candidate expectation is available exactly.

5.5 Point estimation at both smoothness levels

We include a class-valued estimator with squared risk CR_j^2 at each individual level. Fix one level α. In the small-noise range, take the first-part mean as origin and the ball I of radius $D_0\sqrt{a}$, with D_0 large. Except with exponentially small probability this origin is within $1/2$ of the true mean. On that event $f \ge n^{-P}$ on I, the observed-to-latent density ratio there is bounded, and the exterior Gaussian tail in (38) is $O(n^{-2})$ for any candidate with mean within one of the origin.

Use the same scalar and vector kernels as in the validation, with $b(x)^2 = \delta/D(x)$ and fine truncation at w, except that the bandwidth is

$$h_E = \max\{A_{\alpha,d}^{1/(\alpha+1)}, \sqrt{s/(\gamma L_E(\alpha, s))}\}, \qquad h(x) = \begin{cases} h_E, & d \le 2, \\ h_E e^{c_0|x|^2}, & d > 2. \end{cases}$$

There is no pilot normalization. Form their empirical averages $\widehat{U}, \widehat{J}$ with the other part of the sample. The bias relative to $U(f), J(f)$, in the sum of the squared scalar norm with denominator Df and vector norm with denominator f, is $O(h_E^{2\alpha+2})$. This follows as in the validation by the Fourier identity, (36), and the polynomial density floor. For $d > 2$ choose c_0 small so that the weight $e^{(2\alpha+2)c_0|x|^2}$ times any fixed polynomial is integrable against the Gaussian upper envelope. The same choice keeps $h(x)$ polynomially small.

It remains to bound the corresponding variances of the averages. Use the pointwise score tilt at row x, absorbed at the coarse Gaussian scale $b(x)$ or bounded on the fine support. After division by $f(x)$ the scalar row variance integral is bounded by $Cb(x)^{-d}$ before the extra scalar weight $D(x)^{-1}$. For the coarse vector the row integral is bounded by

$$C \int (h(x) + |z|)^{2-2d} e^{-c|z|^2/b(x)^2} \, dz \le C \begin{cases} b(x), & d = 1, \\ 1 + \log(b(x)/h(x)), & d = 2, \\ h(x)^{2-d}, & d > 2. \end{cases}$$

The fine integral is at most $CE^2h(x)^{2-d}$, where $E = C(1 + s/h_E^2)^C e^{Cs/h_E^2}$. On integration and division by n, the coarse bound is $O(\log a/n)$ in dimension one and $O(a^2/n)$ in dimension two. In the first case both $\int_I b(x)\,dx$ and $\int_I b(x)^{-1}/D(x)\,dx$ are $O(\log a)$. In the second the scalar integral is $O(a)$ and the coarse vector has integral $O(a^2)$. Above dimension two the scalar integral is polylogarithmic and the vector integral is at most Ch_E^{2-d} by the spatial increase of h.

For $d = 1$ the fine term before dividing by n is $O(E^2 h_E \sqrt{a}) = o(1)$ if γ is small. For $d = 2$, put $X = 2 + ns^{\alpha+1}$ and use $E^2 \le CX^c$ with $c < 1/4$. If $X \le a^2$, the term $E^2 a$ is $O(a^2)$. Otherwise it is $O((X-2)/\log^{\alpha+1} X)$, using $a \le \sqrt{X}$. For $d > 2$,

$$nh_E^{2\alpha+d} \ge \max\left(1, \frac{ns^{\alpha+d/2}}{(\gamma \log(2 + ns^{\alpha+d/2}))^{\alpha+d/2}}\right) \ge cE^2.$$

These bounds pay all the variances by $Ch_E^{2\alpha+2}$.

Choose, from a countable subset of $\mathcal{F}_\alpha$ dense for W_2 and locally uniform convergence, a candidate with mean less than one from the origin and loss within n^{-2} of the feasible infimum. Use a fixed enumeration and strict inequality in the tolerance. The loss is

$$\int_I \frac{|\widehat{U} - U(g)|^2/D + |\widehat{J} - J(g)|^2}{g}\,dx.$$

The population density f is strictly feasible on the good first-split event. The dense subset approximates its loss: the derivatives in the scalar heat field and the vector heat kernels are integrable with uniform tails on the fixed ball for each n, and candidate densities are uniformly bounded. On I the denominator converges uniformly with positive limit minimum. Thus the selected loss is at most the loss at f plus n^{-2}. For the field discrepancies with denominator $f + g$, the triangle bound is at most twice the sum of these two losses. Apply (38) on the interior of the ball with a cutoff on a fixed-width layer. On the bad mean event the mean constraint and uniform second moments about the component means bound W_2^2 by a constant times $1 + |\widehat{m} - m_f|^2$, with negligible expected contribution by Gaussian concentration. This proves the small-noise estimation bound.

For $s \ge n^{-\eta}$, fit in observed W_1 among a countable W_2-dense subset of mean-zero shapes in the class, translated to the sample mean. Take the first convolved fit within n^{-1} of the infimum. If e_n is the empirical observed W_1 error, the convolved discrepancy between truth and fit is at most $3e_n + n^{-1}$ by the triangle inequality and the cost of translation to the sample mean. Uniformly $\mathbb{E}e_n = O(n^{-\beta})$ for $\beta = 1/[4(d+1)]$. Indeed truncate to distance n^β from the theoretical mean by the second moment bound and quantize with mesh $n^{-\beta}$. These operations each cost $O(n^{-\beta})$ in expectation. The grid has $O(n^{2d\beta})$ points and the expected histogram total variation is $O(n^{d\beta - 1/2})$ by Cauchy–Schwarz for Bernoulli counts. Multiplying by grid diameter gives $O(n^{-1/4})$.

For a pair with observed W_1 discrepancy $\epsilon \le 1$, their means are within one. Apply (38) with the constant bandwidth fields U_h, J_h, on a ball of radius $D_1\sqrt{\log(1/h)}$ about one mean. The Gaussian exterior is $O(h^{\alpha+1})$ in distance for large D_1, the interior density floor is at least h^P, and the vector bias in (36) is $O(h^{\alpha+1})$ in weighted norm. The observed characteristic functions differ by at most $\epsilon|\xi|$ by coupling. Fourier inversion over $|\xi| \le h^{-1}$ shows

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

Take $h = \sqrt{s/(\gamma \log n)}$ with γ small and then η sufficiently small. On $e_n \le n^{-\beta/2}$, the second term is $O(h^{\alpha+1})$ since $s \ge n^{-\eta}$. The complement has probability $O(n^{-\beta/2})$. There W_2^2 is at most $C(1 + |\bar{Y} - m_f|^2)$ by the mean constraint, and $\mathbb{E}|\bar{Y} - m_f|^2 = O(n^{-1})$. This exceptional contribution is $o(h^{2\alpha+2})$ on making η small. In this range both effective logs are comparable to $\log n$, and the direct rates are smaller than the inverse rates, since η is small. This proves the claimed squared risk in both ranges. The allocations of fixed fractions of n to training and validation change all scales only by constants.

5.6 Construction and coverage

For $s \le n^{-\eta}$, train on separate data a rough-class estimator $\widehat{f_0}$ and a smooth-class estimator $\widehat{f_1}$ using the preceding construction. Take the intersection, inside $\mathcal{F}_{\alpha_0}$, of the open rough ball about $\widehat{f_0}$ of radius $M_0 R_0$

and the open validation ball about $\widehat{f}_1$ of radius $M_1(r + \sqrt{(\mathcal{U})_+})$. For the latter, use the validation splits with $g = \widehat{f}_1$. If the validation mean rule rejects, return the rough ball alone. If no row is selected, the statistic is zero. With M_0, M_1 sufficiently large, Markov's inequality and (44) give simultaneous coverage with error at most ζ uniformly. The latent pilot failure probability is $o(1)$ and is accounted for in the coverage probability. The rough diameter bound is deterministic. On a smooth truth, the mean-rule failure has probability $o(1)$, by uniform consistency of $\widehat{f}_1$, the inequality of mean gap by W_2, and concentration of the validation mean. Now the upper side of (44) and the smooth point-estimation bound show diameter $O_{\mathbb{P}}(R_1 + r_0)$, uniformly on the smooth class. The construction constants depend only on the prescribed coverage probability, and the diameter constants can depend on the separate tail probability. The resulting sets have measurable membership. If $s \geq n^{-\eta}$, use only the rough confidence ball: $R_0 \asymp r_0$ in that range by the effective logarithms and the inverse dominance just proved.

For nonscalar Σ, independent Gaussian additions of covariance $sI - \Sigma$ give the completed experiment for the construction. The joint probability including this independent randomization is the same as for the completed isotropic samples. The lower arguments below use the original covariance.

5.7 Lower bounds for confidence diameter

We first explain the conversions to a diameter lower bound. For a finite sign family with squared pairwise separation at least $c \sum_{i:\, \omega_i \neq \omega'_i} v_i$, put $S = \sum_i v_i$. If all neighboring sample total variations are at most ε, every vertex decoder under a uniform sign prior has expected mistaken weight $\mathbb{E}Z \geq (1 - \varepsilon)S/2$. Since $0 \leq Z \leq S$, for $0 < \lambda < (1 - \varepsilon)/2$ it follows that

$$\mathbb{P}(Z \geq \lambda S) \geq \frac{(1 - \varepsilon)/2 - \lambda}{1 - \lambda}.$$

Select the first member of the finite family in the confidence set, using an arbitrary vertex if there is none. When the true vertex belongs to the set and the diameter is less than $\sqrt{c\lambda S}$, this decoder has $Z < \lambda S$. Thus some vertex has probability at least the displayed bound minus ζ of diameter above this scale, up to a harmless change of the threshold constant. Small ε, λ give every fixed probability strictly below $1/2 - \zeta$.

For the testing contribution, let g be a common interior Gaussian. Suppose the mixture over a finite rough alternative family, each member at distance at least cr_0 from g, has sample total variation at most ε from the null. By honesty the event that the confidence set intersects this family has mixture probability at least $1 - \zeta$. Transfer it to the null and intersect with the event that g belongs. The resulting probability at g is at least $1 - 2\zeta - \varepsilon$. On this event the diameter is at least cr_0. The larger of the estimation and testing contributions gives the lower bound of Theorem 5.1, since $\zeta < 1/4$. The required finite experiments are given next.

5.7.1 Estimation families

When comparing vertices, one-sample chi-square divergence contracts under convolution and $1+\chi^2$ tensorizes. The constants in all constructions can make the product total variation as small as any fixed $\varepsilon > 0$. For transport, a smooth φ satisfies

$$\int \varphi(f_\omega - f_{\omega'}) \leq \|\nabla\varphi\|_{L^2(f_\omega)} W_2(f_\omega, f_{\omega'}) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(f_\omega, f_{\omega'})^2. \tag{45}$$

This is Taylor's inequality integrated against an optimal coupling. A pairing at least cE, gradient norm $O(\sqrt{E})$ and bounded Hessian imply $W_2^2 \geq c'E$. The same calculation applies to one vertex and the Gaussian reference.

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

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

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 $C\kappa(ng_j w_j^{2k+3})^{-1/2}$, 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 $Ca_j^2 w_j/g_j = C\kappa^2/n$, 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

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

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

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 $g_B \ge n^{-\delta}$ for a fixed $0 < \delta < 1$ if c is small. Place compactly supported orthonormal tensor wavelets of regularity C^{k+1} [7] 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 Cr_i^{-1-a} at order $a \le k+1$ 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 Ca_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 $\|\Delta u_j\|_2 \le Cr_j^{-2}\|u_j\|_2$ follows by normalized derivatives and overlap. For $r_j \ge r_k$, integration by parts then gives

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

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 $C\sqrt{E}\, W_2 + CW_2^2$, so $W_2^2 \ge cE$. 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.

For $d > 2$, let ψ be a nonzero smooth function of compact support in a unit cube and integral zero. Put $h = n^{-1/(2\alpha+d)}$. 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

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

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 $C\kappa h^{\alpha-j}$. If $\theta > 0$, the order-k derivative increment is bounded by the minimum of $C\kappa h^{\alpha-k}$ and $C\kappa h^{\alpha-k-1}$ times distance. Its θ-Hölder seminorm is $O(\kappa)$. 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 (45) the weights per bit have order $h^{2\alpha+d+2}$. Summing order h^{-d} such weights gives $S \asymp h^{2\alpha+2}$ and the clean estimation scale $A_{\alpha,d}$.

For the inverse estimation lower bound, diagonalize the covariance with first eigenvalue s. In dimension one and for all d when $s \geq s_0 > 0$, use the single narrow Gaussian cosine packet described below at $T^2 = C \log n/s$. The separation is $cT^{-\alpha-1}$ and the product total variation can be made arbitrarily small. This matches the inverse estimation term in these ranges. This is the two-point family consisting of the reference and its single perturbation. For $d \geq 2$ and $0 < s \leq s_0$, use the packet geometry of the testing construction below, now with

$$X = ns^{\alpha+d/2}, \qquad T^2 = A_0 \log(2+X)/s.$$

The relative derivative and Hölder estimates there apply at this scale. Adjacent vertices differ in one packet. The squared norm of the noisy packet against the inverse reference density is at most $Cs^{d/2}e^{-csT^2}$, by the same Gaussian integration as below. Thus n times the one-sample chi-square is at most $C\kappa^2 X(A_0 \log(2+X))^{-\alpha}(2+X)^{-cA_0}$, uniformly small. For two arbitrary vertices the coefficients c_i of their difference replace the signed coefficients in the Gram argument below. The test $\varphi = T^{-2}\sum_i c_i\psi_i$ gives squared separation at least $cT^{-2}s^{d/2}\sum_i c_i^2$. The sum of the weights over order $s^{-d/2}$ bits is of order $T^{-2\alpha-2}$, as needed for the estimation lower bound. This also establishes the actual minimax meaning of each R_j when combined with the point-estimation upper bound.

5.7.2 Testing families at the Gaussian

Here $\alpha = \alpha_0$, and g is a common interior centered isotropic Gaussian. For each n, s, use the family corresponding to the larger of the direct and inverse rates.

We will use finite families $f_\omega = g + \sum_{i=1}^N \omega_i q_i$ where ω_i are independent signs and each q_i has integral zero. Let G be the Gram matrix of the noisy score perturbations in $L^2(p_g)$:

$$G_{ij} = \int \frac{(q_i * \gamma_\Sigma)(q_j * \gamma_\Sigma)}{g * \gamma_\Sigma}.$$

For the uniform mixture of n-sample laws over signs, one plus the chi-square divergence to the null equals $\mathbb{E}_{\omega,\omega'}(1+\omega^T G\omega')^n$. The factor inside the power is nonnegative as the inner product of two likelihood ratios. Bounding its power by $\exp(n\omega^T G\omega')$ and integrating one sign vector gives $\mathbb{E}_\omega \exp(n^2\|G\omega\|^2/2)$ as an upper bound, using $\cosh(t) \leq e^{t^2/2}$. Introduce a standard Gaussian vector to linearize the square and apply the same sign bound again. The result is at most

$$\det(I - n^2G^2)^{-1/2} \leq \exp(Cn^2\|G\|_{\mathrm{HS}}^2) \qquad \text{if } n\|G\|_{\mathrm{op}} \leq 1/2.$$

Thus a small bound on $n^2\|G\|_{\mathrm{HS}}^2$ makes the mixture total variation arbitrarily small. If every mixture member has $W_2(f_\omega, g) \geq r$, the sum of testing errors at separation r is at least one minus that total variation.

The direct polynomial lower bound. For $d > 4$, put $h = n^{-2/(4\alpha+d)}$. Pack order h^{-d} disjoint cubes of side h in a fixed cube about the reference mode. Let ψ be a nonzero smooth bump supported inside the reference unit cube with integral zero, and use $q_i(x) = a_0\psi((x - z_i)/h)$, $a_0 = \kappa h^\alpha$. The relative perturbations q_i/g have uniformly bounded C^α norm of order κ on the combined support, by rescaling and the fixed positive density floor there. The vanishing near edges gives the same bound across cubes. The logarithmic derivative bounds follow by composition with $\log(1+\cdot)$; for sufficiently small κ the reference slack places every alternative in the class.

The clean score Gram matrix is diagonal with entries bounded by $Ca_0^2h^d$. Observed score Gram matrices are bounded above by their clean counterpart in positive operator order: the observed score is the conditional

expectation of the clean score under the reference joint law, so this follows by conditional Cauchy–Schwarz for any linear combination. Thus $n^2\|G\|_{\mathrm{HS}}^2 \le C\kappa^4 n^2 h^{4\alpha+d} = C\kappa^4$. For separation take $\varphi = h^2\sum_i \omega_i q_i$. Its pairing with $f_\omega - g$ is of order $E = h^{d+2}\sum_i a_0^2 \asymp h^{2\alpha+2}$, by disjointness. The gradient's squared L^2 norm is $O(E)$, and its Hessian is uniformly bounded. Equation (45) therefore gives the direct lower bound for every sign.

The direct critical lower bound. When $d = 4$, pack disjoint cubes of side $\ell = (\log n)^{-1/2}$ inside a cube of side $c\sqrt{\log n}$ about the mode, with $c > 0$ small. Their total volume is order $(\log n)^2$. The reference g is comparable to its value g_B at the center of each cube B, and $g_B \ge n^{-\beta_0}$ for a fixed $\beta_0 < 1$. In the rescaled unit cube the ratio g_B/g has bounded C^α norm uniformly: on length ℓ, the score bound is paid by $\ell\sqrt{\log n} = 1$, and all higher potential derivatives are bounded with the corresponding scale factors. The top Hölder bound scales by ℓ^α.

Use a fixed orientation of compactly supported orthonormal tensor wavelets [7] of regularity C^{k+1}, supported strictly inside each cube. At physical length $r_i = \ell 2^{-j}$ their number in a cube is comparable to $(\ell/r_i)^4$, for j from a sufficiently large fixed index through $\lfloor \nu \log_2 n \rfloor$, with $\nu > 0$ small. Write ψ_i for these L^2-normalized wavelets. In particular $S_4 = \sum_i r_i^4 \asymp (\log n)^3$: each level contributes order the total volume. Set

$$b_i = \frac{\kappa^2 r_i^2}{n\sqrt{S_4}}, \qquad a_i = \sqrt{g_{B(i)} b_i}, \qquad q_i = a_i \psi_i.$$

Each wavelet has zero integral. At physical scale r, the relative perturbation in cube B has C^α norm at most $C(a_i/g_B) r^{-2-\alpha}$, by wavelet scaling, bounded overlap at the scale, and the rescaled reciprocal density bound. Summing over scales gives at most

$$C\kappa n^{-(1-\beta_0)/2} S_4^{-1/4} r_{\min}^{-1-\alpha} = o(1)$$

if $\nu < (1-\beta_0)/(2(1+\alpha))$. For fractional order this controls the global seminorm: on a pair of points in distinct cubes one may use the zero values of the top derivatives on the boundary of each cube and distances to the first exit points on the segment. Thus all perturbations stay within the reference margins by logarithmic composition.

The clean Gram matrix is $D_0 H D_0$ with $D_0 = \mathrm{diag}(\sqrt{b_i})$, and $H \preceq CI$. This uses orthonormality and the uniform bounds on g_B/g within each cube. The noisy Gram shrinks in positive operator order as before, so $n^2\|G\|_{\mathrm{HS}}^2 \le Cn^2 \sum_i b_i^2 = C\kappa^4$. For the metric test take $\varphi = \sum_i (r_i^2/g_{B(i)})\omega_i a_i \psi_i$. Its pairing is exactly $E = \sum_i r_i^2 b_i = \kappa^2\sqrt{S_4}/n$. The gradient squared norm in a cube is controlled by the sum of squared coefficient norms weighted by r_i^{-2}, up to a fixed factor. For clarity, within one level sum u_j, $\|\Delta u_j\|_2 \le C r_j^{-2}\|u_j\|_2$ by overlap. When $r_j \ge r_\ell$, integration by parts yields

$$|\langle \nabla u_j, \nabla u_\ell \rangle| \le C(r_\ell/r_j)(\|u_j\|_2/r_j)(\|u_\ell\|_2/r_\ell).$$

Summing the geometric ratios and multiplying by the comparable density in each cube gives squared gradient norm $O(E)$ under f_ω. The test Hessian is uniformly bounded, since its per cube scale sum is bounded by $C\kappa(ng_B)^{-1/2} S_4^{-1/4} r_{\min}^{-1}$. Thus (45) gives separation at least $c(\log n)^{3/4}/\sqrt{n}$. This lower bound holds under every known Gaussian covariance, by the same mixture contraction.

The inverse mixture. Rotate axes so $\Sigma = \mathrm{diag}(s, t_2, \dots, t_d)$ with $0 \le t_i \le s$, and suppose $0 < s \le s_0$ with $s_0 > 0$ small. Place $N \asymp s^{-d/2}$ centers z_i spaced $D\sqrt{s}$ apart in a fixed cube. Here D is sufficiently large, and s_0 small after choosing D. Put $T^2 = A_0 \log(2 + ns^{\alpha+d/4})/s$, and use

$$q_i(x) = a_0 \psi_i(x), \qquad a_0 = \kappa T^{-\alpha}, \qquad \psi_i(x) = e^{-|x-z_i|^2/(2s)} \left(\cos(T(x_1 - z_{i,1})) - e^{-sT^2/2} \right).$$

Each packet has zero mass. The relative perturbation before the factor a_0 has derivatives through order $j \le k$ bounded by CT^j. Indeed derivatives of the packets are bounded by $CT^j e^{-c|x-z_i|^2/s}$ because $T\sqrt{s}$ is large, and division by g costs only Gaussian and polynomial factors absorbed in a weaker envelope for small s_0. Separated grid envelopes have uniformly bounded sum. For $0 < \theta < 1$, the k-th derivative has θ-Hölder seminorm at most CT^α. For shifts longer than $1/T$ use the sup norm. For a shorter shift expand

the product-rule terms of $D^k(\psi_i/g)$. A difference of a smooth packet derivative of order j costs $CT^{j+1}|u|$ with the same weakened envelope. A difference of a reciprocal derivative of order $k-j$ costs $C|u|^\theta$ times polynomial factors and the maximum of the two reciprocal densities, by the product rules and the reference Hölder condition on D^kV_g. The nearby packet envelopes still absorb those factors since $|u| \le 1/T \le C\sqrt{s}$. Summing the bounds gives $CT^\alpha|u|^\theta$. The logarithmic perturbation thus costs $O(\kappa)$ in the stipulated norms and preserves the class for small enough κ.

Convolution with the noise changes the packet envelope to variances $s+t_i$, with $t_1 = s$, multiplying its amplitude by $\prod_i \sqrt{s/(s+t_i)}$. Its bracket becomes

$$e^{-sT^2/4}\cos\big(\tfrac{T}{2}(y_1 - z_{i,1})\big) - e^{-sT^2/2}.$$

These formulas follow by conditional Gaussian integration and include zero eigenvalues in the other coordinates. The convolved reference has a lower Gaussian envelope, obtained by convolving one for g. Combining the two narrow packet envelopes and dividing by this lower envelope gives

$$|G_{ij}| \le Ca_0^2 s^{d/2} e^{-csT^2} e^{-c|z_i - z_j|^2/s}.$$

Indeed the common center lies in a fixed cube, the variances of the two envelopes are between s and $2s$, and choosing s_0 small absorbs the reciprocal reference into a weaker Gaussian. Thus

$$n\|G\|_{\mathrm{HS}} \le Cna_0^2 s^{d/4} e^{-csT^2}.$$

With $X = ns^{\alpha+d/4}$, the right-hand side is at most $C\kappa^2 X(A_0 \log(2+X))^{-\alpha}(2+X)^{-cA_0}$. It is arbitrarily small by choosing A_0 large and κ small.

For transport separation take $\varphi = T^{-2}\sum_i \omega_i q_i$. The spatial L^2 Gram matrix of the ψ_i has diagonal at least $cs^{d/2}$ by integrating the squared cosine in its Gaussian envelope and taking $T\sqrt{s}$ large. The absolute off-diagonal row sums are at most $Cs^{d/2}\sum_{j\in\mathbb{Z}^d\setminus\{0\}} e^{-cD^2|j|^2}$, less than half the diagonal bound for large D. The pairing with $f_\omega - g$ is at least cE for $E = T^{-2}a_0^2 s^{d/2}N$. The gradient Gram has absolute row sums at most $CT^2 s^{d/2}$ by the same overlap estimate, so the test gradient has squared norm $O(E)$ under the uniformly bounded vertex densities. The pointwise derivative envelope bounds the test Hessian by Ca_0. Equation (45) gives $W_2(f_\omega, g) \ge cT^{-\alpha-1}$ for every sign.

For $s \ge s_0$, use instead one additive packet throughout, replacing its envelope variance by a fixed τ_0^2 smaller than the variance in the Gaussian lower envelope for g, with its cosine mean removed. Use amplitude $\kappa T^{-\alpha}$ and $T^2 = C\log n/s$. The derivative and Hölder argument above applies to this single packet. Its convolved Gaussian envelope squared, divided by the convolved reference lower envelope, is integrable uniformly in Σ by the variance choice. Its first-coordinate cosine attenuates by $\exp(-\tau_0^2 sT^2/[2(\tau_0^2+s)])$, so the n-sample chi-square tends to zero uniformly for sufficiently large C. Testing the latent difference with $\cos(Tx_1)/T$ gives separation of order $T^{-\alpha-1}$. This matches the inverse term since its logarithm is comparable to $\log n$ on this range.

Lower bounds in dimensions one through three. In dimension one a shift of the reference by δ has W_2 distance $|\delta|$. Its clean one-sample relative entropy to the reference is $O(\delta^2)$, by expanding the potential and using the zero mean score. Since noise contracts entropy, $\delta = c/\sqrt{n}$ gives the desired two-point lower bound.

For $d = 2, 3$, take disjoint annuli at dyadic radii R from a fixed constant to $c\sqrt{\log n}$, for small fixed $c > 0$. Pack cubes of side $\ell_i \asymp R^{-1}$ in each annulus, with order R^{2d} cubes at radius R. The reference values g_i at their centers satisfy $g_i \ge n^{-\beta_0}$, $\beta_0 < 1$, and $g \asymp g_i$ inside each cube. Use one L^2-normalized smooth mean-zero bump ψ_i per cube, with

$$S_4 = \sum_i \ell_i^4, \qquad b_i = \frac{\kappa^2 \ell_i^2}{n\sqrt{S_4}}, \qquad q_i = \sqrt{g_i b_i}\,\psi_i.$$

The rescaled reciprocal bounds used for the critical construction apply here since $\ell_i(1+|x|) = O(1)$. All relative C^α perturbations tend to zero because $g_i \ge n^{-\beta_0}$ and $\ell_i \gtrsim (\log n)^{-1/2}$. The clean Gram matrix is diagonal with entries $O(b_i)$, and the noisy Gram contracts. Hence $n^2\|G\|_{\mathrm{HS}}^2 = O(\kappa^4)$. The test $\varphi =$

$\sum_i(\ell_i^2/g_i)\omega_i q_i$ has pairing $\sum_i \ell_i^2 b_i$, squared gradient norm at most a constant times that pairing, and bounded Hessian by rescaling on the disjoint cells. Equation (45) gives squared separation of order $\sqrt{S_4}/n$. A ring contributes order R^{2d-4} to S_4; summing gives order $\log\log n$ for $d = 2$ and $\log n$ for $d = 3$.

For the inverse lower bound in all three dimensions, the single additive narrow packet with $T^2 = C\log n/s$ works throughout $0 < s \le \Lambda$. Its attenuated cosine multiplier is bounded by e^{-csT^2} uniformly for the fixed narrow variance, and the noisy reference denominator is paid as above. It gives latent separation of order $(s/\log n)^{(\alpha+1)/2}$ with small total variation. The direct and inverse testing families prove the claimed confidence lower bound at the Gaussian.

5.8 The adaptation window

Theorem 5.1 implies that adaptation holds precisely when $r_0 = O(R_1)$. The smoother inverse term is uniformly $o(1)$ times the rough testing inverse term for positive s. To see this, first take $s \ge n^{-\eta}$ with η small. The effective logs are all comparable to $\log n$, so their ratio of inverse rates is $O((s/\log n)^{(\alpha_1-\alpha_0)/2})$. For $s < n^{-\eta}$, use the upper bound $C\log n$ on L_T and the fixed positive lower bound on L_E. Now that ratio is bounded by

$$Cn^{-\eta(\alpha_1-\alpha_0)/2}(\log n)^{(\alpha_0+1)/2} = o(1).$$

Hence $r_0 = O(R_1)$ if and only if both terms in r_0 are $O(A_{\alpha_1,d})$. When $s = 0$ the latter equivalence is immediate.

The clean comparison is automatic in dimensions at most four. For $d > 4$, compare the two powers of n in the table:

$$\frac{\alpha_1+1}{2\alpha_1+d} \le \frac{2(\alpha_0+1)}{4\alpha_0+d}.$$

This is the stated clean inequality. For $d < 4$, $L_T = \log n$, giving the asserted noise condition directly. For $d \ge 4$, write $A_{\alpha_1,d} = n^{-e}$, $p = \alpha_0 + d/4$, $q = 2e/(\alpha_0+1)$. The inverse requirement is $s/\log(2+ns^p) = O(n^{-q})$. Necessarily $s = O(n^{-q}\log n)$. When $qp < 1$ this is sufficient: for $s \le n^{-q}$ use the positive lower bound on the log, and for $s > n^{-q}$ use $ns^p \ge n^{1-qp}$. In dimension four $qp = 2e < 1$. In higher dimensions $qp < 1$ is the strict clean inequality. At equality $qp = 1$, set $u = n^{1/p}s$. The inverse requirement becomes $u/\log(2+u^p) = O(1)$, equivalent to $u = O(1)$. This proves Theorem 5.2.

6 Wasserstein inference with independently calibrated Gaussian noise

We give a transfer principle for inference on strongly log-concave distributions under Gaussian contamination when the error covariance is estimated from auxiliary data. The cost of calibrating the covariance from k independent Gaussian error observations is $\|\Sigma\|_{\mathrm{op}}/\sqrt{k}$ in quadratic Wasserstein distance. This adds to the known-covariance estimation and testing rates. The bound is independent of the smallest positive eigenvalue and allows singular or vanishing covariance. The same calibration cost holds when the auxiliary data are paired repeated measurements, with a corresponding increase in the effective number of observations for the latent law.

6.1 Experiments and rates

Let $\mathcal{F}_\alpha$ be the translated class of densities $f = e^{-V}$ on $\mathbb{R}^d$ with two-sided Hessian bounds $mI \preceq D^2V \preceq MI$ for $0 < m < M$. For fixed $\alpha = j + \theta \ge 2$, j integer and $0 \le \theta < 1$, require $V \in C^j$, positive finite bounds on the sup norms of its derivatives of orders $3, \dots, j$, and for $\theta > 0$ a positive finite global θ-Hölder bound on D^jV. Write $s = \|\Sigma\|_{\mathrm{op}} \le S \le \Lambda$ for the unknown Gaussian error covariance norm and its prescribed envelope. In the calibration experiment observe n independent samples from $f * \gamma_\Sigma$ and k independent draws from $\gamma_\Sigma = N(0,\Sigma)$, the two samples being independent. The paired experiment replaces the k pure error draws by k independent pairs $(X_j + \epsilon_j, X_j + \epsilon'_j)$ at distinct individuals, with $X_j \sim f$, independent errors $\epsilon_j, \epsilon'_j \sim \gamma_\Sigma$, and all pairs independent of the main sample. In the two-sample versions there are two independent such groups, which may have different sample sizes and error covariances.

Put $N = n$ for calibration and $N = n + k$ for the paired design. Write $R_E(N, s)$ and $R_T(N, s)$ for the known-covariance estimation and testing scales:

d	$A_E(N)$	$A_T(N)$
1	$(\log\log N)^{1/2}/\sqrt{N}$	$N^{-1/2}$
2	$\log N/\sqrt{N}$	$(\log\log N)^{1/4}/\sqrt{N}$
3	$N^{-(\alpha+1)/(2\alpha+3)}$	$(\log N)^{1/4}/\sqrt{N}$
4	$N^{-(\alpha+1)/(2\alpha+4)}$	$(\log N)^{3/4}/\sqrt{N}$
$d > 4$	$N^{-(\alpha+1)/(2\alpha+d)}$	$N^{-2(\alpha+1)/(4\alpha+d)}$

For R_E add $(s/L_E)^{(\alpha+1)/2}$ to A_E, where $L_E = \log N$ for $d = 1$ and $L_E = \log(2 + Ns^{\alpha+d/2})$ otherwise. For R_T add $(s/L_T)^{(\alpha+1)/2}$ to A_T, where $L_T = \log N$ for $d < 4$ and $L_T = \log(2 + Ns^{\alpha+d/4})$ otherwise. The inverse summands are zero at $s = 0$.

Theorem 6.1. *The fixed-probability minimax radius for estimating f in W_2 is of order $R_E(N, S) + S/\sqrt{k}$. For the two-sample problem of testing $f_1 = f_2$, the minimax separation is of order*

$$\max_{i=1,2}\{R_T(N_i, S_i) + S_i/\sqrt{k_i}\}.$$

The upper bounds also hold with the true noise norms s_i in place of the envelopes, using procedures that do not know those norms. The rates are uniform when the main and calibration sizes tend to infinity, with no sample-size comparability or rank assumption. The fixed-error testing statement uses the usual sum of worst type-I and type-II errors. For the upper oracle version, each alternative is compared to the separation formed from its own covariances.

The proof uses the known-covariance estimation and lower-bound results, and the unequal two-sample testing construction, in the companion proofs Sections 5, 3. We give the needed tolerance property explicitly below. Auxiliary error observations have long been used for nonparametric density deconvolution [18, 19]. Lee, Lahiri, and Sinha [17] use deconvolved characteristic functions for asymptotic two-sample inference with distinct measurement errors and repeated observations. Delaigle, Hall, and Meister [20, Theorem 3.6] obtain pointwise density-oracle equivalence for fixed supersmooth errors under their replicate design. Scricciolo [21] treats univariate ordinary-smooth deconvolution in W_1 with an auxiliary error experiment. For the fixed known Gaussian inverse problem in the present curvature class, see [12]. Here Gaussian covariance order yields the sharp W_2 calibration cost for independently varying group covariances.

6.2 Gaussian smoothing in the curvature class

Lemma 6.2. *For $A \succeq 0$ with $\|A\|_{\mathrm{op}} \le \Lambda$ and $f \in \mathcal{F}_\alpha$, the law $f_A = f * \gamma_A$ belongs to a class of the same form and smoothness α, with uniformly enlarged derivative bounds and uniformly positive curvature. Also $W_2(f_A, f) \le C\|A\|_{\mathrm{op}}$.*

Proof. For positive definite A, differentiation of the posterior integral gives

$$D^2[-\log f_A](x) = A^{-1} - A^{-1}\operatorname{Cov}(X \mid X + Y = x)A^{-1}.$$

The posterior of X has potential Hessian between $mI + A^{-1}$ and $MI + A^{-1}$. Brascamp–Lieb gives covariance at most $(mI + A^{-1})^{-1}$. For the converse bound, integration by parts gives $\mathbb{E}[(X - \mathbb{E}X)\nabla H(X)^T] = I$ and $\mathbb{E}[\nabla H(X)\nabla H(X)^T] = \mathbb{E}D^2H(X) \preceq MI + A^{-1}$ for the posterior potential H. The block covariance inequality then gives $\operatorname{Cov}(X) \succeq (MI + A^{-1})^{-1}$. It follows that the output Hessian is between $m(1 + m\|A\|_{\mathrm{op}})^{-1}I$ and MI. For singular A pass to $A + \epsilon I$ by local convergence of the first two derivatives of the smoothed positive densities.

We show the derivative bounds directly, including for singular A. Use the probability measure ν_x on $\mathbb{R}^d$ proportional to $e^{-V(x+A^{1/2}z)}\, d\gamma_I(z)$. Its potential has Hessian between I and $(1 + M\Lambda)I$; quadratic bounds around its mode z_x give uniformly bounded moments of every fixed order about z_x. Differentiating the log integral at x expresses each order-l derivative of $\log f_A(x)$ as a sum over partitions of the l derivative indices.

A block of size b contributes the corresponding derivative of $-V$ evaluated at $x+A^{1/2}z$, and the blocks enter a joint cumulant under ν_x. This identity follows by differentiating the multivariable log moment generating function of the increment of $-V$. Local differentiation under the integral is justified by the Gaussian factor, the quadratic bounds on V, and polynomial bounds for derivatives of its exponential.

For $l \geq 2$ the single-block term is the expectation of a bounded derivative. In any cumulant with at least two blocks, each score factor can be centered by subtracting its value at $x + A^{1/2}z_x$, without changing the cumulant. The resulting factor is bounded by $C|z - z_x|$, while every block of size at least two is bounded. The moment formula for cumulants gives uniform bounds up to order j.

When $\theta > 0$, the single-block term at order j is uniformly θ-Hölder. Its integrand shifts by $O(|x - x'|^\theta)$, and the posterior changes by $O(|x - x'|)$ in total variation, because its derivative with respect to x is the measure times a centered score with bounded $L^1(\nu_x)$ norm. Integrate that directional derivative along the segment from x to x'. For all multi-block terms the integrands use orders at most $j - 1$, so one further derivative uses only the derivatives already assumed bounded, except for score factors that can be centered. More precisely, to evaluate this further derivative at a fixed x, hold the subtracted score constants at that x fixed during differentiation. The derivatives of the shifted integrands and the centered-score derivative of the posterior then have uniformly bounded mixed moments by the same mode estimate. The finite moment formula for cumulants gives a uniform Lipschitz bound for those terms. For distances above one, the sup norm suffices for the Hölder condition. Thus all the required bounds hold in the enlarged class.

Finally let $f_t = f * \gamma_{tA}$, $0 \leq t \leq 1$. It satisfies the continuity equation with velocity $-A\nabla \log f_t/2$. Its Fisher information $\int |\nabla \log f_t|^2 f_t$ is at most dM by integration by parts and the output Hessian upper bound, which also holds at $t = 0$. The dynamic transport action is at most $dM\|A\|_{\mathrm{op}}^2/4$. This proves the displacement estimate. □

6.3 A tolerant known-covariance comparison

We use the following consequence of the unequal-sample proof in Section 3. For a fixed $K > 0$, suppose the two covariances are known and a prescribed scale $\lambda \to 0$ satisfies

$$\lambda \geq \max_i R_T(N_i, s_i).$$

There is a test with both errors arbitrarily small at fixed levels when the null is enlarged to $W_2(f_1, f_2) \leq K\lambda$ and the alternatives have $W_2(f_1, f_2) \geq C\lambda$. The constant C can depend on K and the error levels.

Indeed the proof works with the effective sample size given by the clean testing scale comparable to λ. This is at most the smaller actual sample size, up to fixed constants. All pilot and covariance bounds in that construction use only $\lambda \geq R_T(N_i, s_i)$, and the two-sided comparison of the mean feature difference μ gives

$$\lambda + \|\mu\| \asymp \lambda + W_2(f_1, f_2)$$

on the pilot event. If a large-noise group is fitted first, the fit changes the latent distance by $O(\lambda)$ with probability $1 - o(1)$, uniformly. The same comparison therefore holds up to constants for the pair under validation. Its squared-norm statistic has conditional mean $\|\mu\|^2$ and variance at most $C(\lambda^4 + \lambda^2\|\mu\|^2)$. A large fixed threshold times λ^2 thus controls the tolerant null, and a sufficiently larger separation controls the alternative. The preliminary mean rule contributes $o(1)$ under this enlarged null since the means are $O(\lambda)$ apart. If both groups have large noise, comparing both fitted laws uses the triangle inequality instead. These arguments prove the tolerance property, including for covariances in a fixed enlarged curvature class.

6.4 Calibration reduction and upper bounds

Let $\widehat{\Sigma} = k^{-1}\sum_{\ell=1}^{k} Z_\ell Z_\ell^T$, where $Z_\ell \sim N(0, \Sigma)$ are the pure error draws. For any fixed probability budget there is a constant C_0 such that, with $\tau = C_0/\sqrt{k}$, the event

$$(1 - \tau)\Sigma \preceq \widehat{\Sigma} \preceq (1 + \tau)\Sigma$$

has probability at least one minus that budget, uniformly. To see this, whiten on the range of Σ, of dimension $q \leq d$. The whitened sample covariance has expected squared Frobenius deviation $q(q + 1)/k$

from the identity by Gaussian moments. Markov's inequality bounds its operator deviation, and the two quadratic form inequalities follow after undoing the whitening. At rank zero the event holds identically.

Set $\Sigma^- = \widehat{\Sigma}/(1+\tau)$. If an empirical norm exceeds 2Λ, specify an arbitrary fallback decision or a fixed Gaussian estimate. This occurs only outside the good event for sufficiently large k. On the good event with $\tau \le 1/2$,

$$D = \Sigma - \Sigma^- \succeq 0, \qquad D \preceq \frac{2\tau}{1+\tau}\Sigma, \qquad \|\Sigma^-\|_{\mathrm{op}} \asymp s.$$

The observations conditional on calibration have precisely the law

$$f * \gamma_\Sigma = (f * \gamma_D) * \gamma_{\Sigma^-}.$$

The effective latent law $f * \gamma_D$ belongs to the uniformly enlarged class of Lemma 6.2, and is within $C\tau s$ of f in W_2. For s' within fixed positive multiplicative factors of s, each displayed estimation or testing rate is comparable uniformly in N. This follows immediately for $\log N$, and for $\log(2 + Ns^p)$ by multiplying Ns^p by a bounded factor. The same comparability holds when both scales are zero.

Use the known-noise estimator on the conditional latent problem. Its error at any fixed probability is bounded by the constant multiple of $R_E(N, \|\Sigma^-\|_{\mathrm{op}})$ given by the known-covariance result. The triangle inequality gives the claimed true-norm estimation bound. The procedure only uses Σ^-, not s. For two samples use the tolerant comparison with

$$\lambda = \max_i \left\{ R_T(N_i, \|\Sigma_i^-\|_{\mathrm{op}}) + \|\widehat{\Sigma}_i\|_{\mathrm{op}}/\sqrt{k_i} \right\}.$$

Conditional on the jointly successful calibration events this is comparable to the asserted true-norm scale. Under the null the two effective latent laws are within $K\lambda$ by the two heat perturbation estimates. Under the alternatives their distance is at least the original separation minus $K\lambda$. Thus the conditional test, with constants depending only on the fixed budgets, gives the claimed errors. A union bound accounts for the calibration events.

For the paired experiment, transform each pair to its average and its difference divided by $\sqrt{2}$. The difference is $N(0, \Sigma)$ and independent of the average, because the two Gaussian noise linear combinations are independent and independent of the latent variable. The average has latent law f and Gaussian covariance $\Sigma/2$. Use either the n main observations or the k independent averages, whichever sample is larger. Use the differences for calibration. If using averages apply the reduction with half the estimated covariance and half its residual. The actual sample size is within a factor two of $N = n + k$, and the effective noise scale is between $s/2$ and s before calibration. All displayed rates are comparable under these changes.

6.5 Lower bounds

First fix each error covariance at a known value of norm S_i with a top eigenvector in the first coordinate. Pure calibration observations then have the same law in all lower-bound experiments. The fixed-covariance estimation and testing lower bounds in Sections 5, 3 apply to the main samples. In the paired experiment, a more informative fixed-covariance experiment gives $n+k$ independent observations at covariance $\Sigma/2$. From the first n one can generate main observations by adding independent $N(0, \Sigma/2)$ variables. The other k serve as the replicate averages, and fresh independent $N(0, \Sigma)$ variables as their differences. The inverse linear transformation then produces the pairs. Lower bounds for the more informative experiment apply to the paired experiment and have the stated N-sample rates, since multiplication by $1/2$ changes the noise terms only by constants. For testing, vary only one group and fix the other at the common interior Gaussian.

It remains to prove the additional calibration lower bound. If $S > 0$, choose a centered isotropic Gaussian with variance v strictly inside the interval of allowed latent Gaussian variances. Let e be the first coordinate vector and

$$\Sigma_0 = (S/2)ee^T, \qquad \Sigma_1 = (S/2 + \Delta)ee^T, \qquad \Delta = cS/\sqrt{k}.$$

Take the two latent covariance matrices vI and $vI - \Delta ee^T$ with their respective errors Σ_0, Σ_1. For sufficiently small c both Gaussian densities are in the class, and both error covariances satisfy the envelope. The main samples have identical Gaussian laws since the covariance changes cancel. The latent W_2 distance is $\sqrt{v} - \sqrt{v - \Delta} \asymp \Delta$. The pure error draws have the same rank-one support and one-dimensional variance

ratio $1+2\Delta/S$. Their k-sample relative entropy is at most $Ck\Delta^2/S^2$ by the scalar Gaussian relative entropy formula. Thus the total variation can be made as small as required by choosing c small.

For paired data the orthogonal difference has the same entropy cost. The independent replicate average has first-coordinate variance $v+S/4$ under the first law and $v+S/4-\Delta/2$ under the second. The remaining coordinates stay fixed. These variances are bounded away from zero. The average laws add relative entropy at most $Ck\Delta^2 \le C_\Lambda k\Delta^2/S^2$. The main sample still has identical laws. Two-point testing therefore gives the calibration term in the paired experiment as well. For estimation the two-point reduction uses the same latent separation. For two samples use this construction in one group, with the other latent law fixed at the common reference Gaussian and its observation model the same. At $S=0$ there is no calibration term. Taking the larger of the known-covariance and calibration lower bounds, then maximizing over the groups for testing, proves Theorem 6.1.

6.6 Deterministic covariance uncertainty

The same reduction covers a known operator-norm neighborhood without auxiliary data. Suppose there is a specified $\Sigma_0 \succeq 0$ of bounded norm and an unknown PSD Σ with $\|\Sigma-\Sigma_0\|_{\mathrm{op}} \le \delta$. Take $\delta \to 0$ in the asymptotic experiment. The estimation rate is $R_E(n, \|\Sigma_0\|_{\mathrm{op}}) + \delta$; for two independent groups the testing rate is

$$\max_i \{R_T(n_i, \|\Sigma_{0,i}\|_{\mathrm{op}}) + \delta_i\}.$$

Indeed add independent $N(0, \delta I)$ to each observation. Then $A = \Sigma - \Sigma_0 + \delta I$ lies between 0 and $2\delta I$ in quadratic form order. The new conditional problem has exact known covariance Σ_0 and latent density $f * \gamma_A$. Lemma 6.2 and the tolerant comparison give the upper bounds. This argument allows noncommuting matrices. The known-covariance lower bounds use $\Sigma = \Sigma_0$. For the uncertainty term take two error covariances $\Sigma_0, \Sigma_0 + c\delta ee^T$ and the same compensated Gaussian latent pair as above. Their observations have exactly the same law, while the latent distance is at least $c'\delta$. This proves the lower bounds.

6.7 Allocation of error measurements

The formulas give the calibration regime without requiring a comparison of sample sizes alone. With pure error observations, calibration has no effect on the estimation order once $k \gtrsim (S/R_E(n,S))^2$. For testing the corresponding groupwise budget is governed by that group's term $S_i/\sqrt{k_i}$, compared to the larger of the known-noise testing terms of the two groups. For a single group with fixed positive noise, the latent nonparametric error is of order $(\log n)^{-(\alpha+1)/2}$; the error-measurement budget then has threshold of order $(\log n)^{\alpha+1}$. In paired designs this comparison uses $n+k$ in the latent rate because the repeats also supply independent averages.

7 Calibration-free Wasserstein testing with shared Gaussian noise

> We study quadratic Wasserstein testing of two strongly log-concave latent densities observed with the same unknown Gaussian covariance. A single procedure, without noise calibration or a supplied noise scale, attains the known-covariance separation at every covariance. The covariance may be singular and may vary with sample size. The procedure scans inverse heat fields over covariance and resolution. Exact null centering, covariance-derivative bounds and a summable resolution budget control that scan without a rate penalty.

7.1 Experiment and separation

Fix $d \ge 1$, real $\alpha = k+\theta \ge 2$, integer k, $0 \le \theta < 1$, and constants $0 < m < M < \infty$. Let $\mathcal{F}_\alpha$ be the translated class of 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 $3, \dots, k$. If $\theta > 0$, impose a fixed positive bound on the global θ-Hölder seminorm of D^kV. We observe independent samples of sizes n_1, n_2 from $p_f = f * \gamma_\Sigma$ and $p_g = g * \gamma_\Sigma$, with

the same unknown covariance $\Sigma \succeq 0$. Write $n = \min(n_1, n_2)$, $s = \|\Sigma\|_{\mathrm{op}} \le \Lambda$ for fixed Λ, $\gamma_\Sigma = N(0, \Sigma)$, and $\gamma_t = N(0, tI_d)$ for scalar t. Define

$$a_{n,d} = \begin{cases} n^{-1/2}, & d = 1, \\ (\log\log n)^{1/4}/\sqrt{n}, & d = 2, \\ (\log n)^{1/4}/\sqrt{n}, & d = 3, \\ (\log n)^{3/4}/\sqrt{n}, & d = 4, \\ n^{-2(\alpha+1)/(4\alpha+d)}, & d > 4. \end{cases} \qquad B_{n,d}(s) = \begin{cases} \log n, & d < 4, \\ \log(2 + ns^{\alpha+d/4}), & d \ge 4. \end{cases}$$

The null is $f = g$, uniformly over the latent class and all allowed covariances. For power we measure separation at the actual covariance of the alternative.

Theorem 7.1. *For any fixed $0 < \varepsilon < 1/2$, one test that does not know s or Σ has worst type-I error at most $\varepsilon/2$ and worst type-II error at most $\varepsilon/2$ on alternatives with $W_2(f, g) \ge C_\varepsilon r_{n,d}(s)$, where*

$$r_{n,d}(s) = a_{n,d} + (s/B_{n,d}(s))^{(\alpha+1)/2}.$$

This holds uniformly as $n \to \infty$ for every sample-size ratio. For every admissible covariance at each sample size, the sum of errors at $c_\varepsilon r_{n,d}(s)$ is at least $1 - \varepsilon$ even if the covariance is revealed, with constants uniform in the matrix.

The known-noise case for a fixed reference at fixed positive Gaussian variance appears in [12, Theorem 2.4]. Lee, Lahiri, and Sinha [17] use deconvolved characteristic functions for asymptotic two-sample inference with measurement error, including separate unknown error laws. Common noise preserves equality of the observed laws under the latent null. We use that structure to obtain a uniform latent separation bound while selecting neither a noise estimate nor a true inverse kernel under the null.

7.2 Deterministic transport and Fourier fields

Uniformly over the class, f has Gaussian upper and lower envelopes about its mean m_f, the score of f is $O(1 + |x - m_f|)$, and

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

Indeed quadratic bounds at the mode give the envelopes and bound the distance between the mode and the mean. Differentiating e^{-V} gives the derivative bounds. For $\theta > 0$, on $|x - m_f| \le L$ and $|u| \le (2 + L)^{-1}$, the same product rule and the Hölder bound on $D^k V$ give

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

The local density ratio is bounded by the score, derivatives of lower order in the product are Lipschitz, and the highest one uses the assumed Hölder seminorm.

For $p_f = f * \gamma_\Sigma$, $0 \preceq \Sigma \preceq \Lambda I$, Gaussian integration of the two quadratic Taylor bounds gives

$$c f(x) \le p_f(x) \le f(x) e^{Cs(1+|x-m_f|^2)}. \tag{46}$$

The potential of p_f also has Hessian in a fixed positive interval. For positive definite Σ its Hessian is $\Sigma^{-1} - \Sigma^{-1} \operatorname{Cov}(X \mid Y) \Sigma^{-1}$. The posterior potential has Hessian between $mI + \Sigma^{-1}$ and $MI + \Sigma^{-1}$. Brascamp–Lieb [5] bounds its covariance from above by the inverse of the first matrix. Integration by parts against the posterior gives cross covariance I between the centered position and the score, whose covariance is at most the second matrix. The block covariance inequality gives the lower posterior covariance bound by its inverse. Substitution yields an output curvature lower bound $m/(1 + m\Lambda)$ and upper bound M. Approximation by positive definite matrices extends this to singular covariance. Consequently the observed scores are uniformly $O(1 + |x - m_f|)$ and quadratic envelopes hold about the latent mean m_f, which is unchanged by convolution.

Use $\widehat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\, dx$. Let $\widehat{K}$ be a real radial smooth Fourier cutoff that equals one near zero and is supported in the unit ball. Use $K_b = b^{-d} K(\cdot/b)$ for its spatial kernel, and define constant-bandwidth fields

$$U_b(f) = K_b * f, \qquad \widehat{J_b(f)}(\xi) = -\frac{i\xi}{|\xi|^2}(1 - \widehat{K}(b\xi))\widehat{f}(\xi).$$

Then $f = U_b(f) + \operatorname{div} J_b(f)$. The inverse transform Q of $-i\xi(1-\widehat{K}(\xi))/|\xi|^2$ is integrable with all absolute moments finite and all signed moments zero. In fact a dyadic partition into shells 2^j away from the origin gives inverse transforms bounded by $C_N 2^{(d-1)j}(1+2^j|x|)^{-N}$. Summability of their L^1 norms $O(2^{-j})$ gives the assertion, with moment cancellation since the symbol vanishes near zero. For $|x-m_f| \le L$, $w = (2+L)^{-1}$, and $h \le w$, subtract the local Taylor polynomial of f to degree $k-1$ in the integer case and k otherwise. This gives

$$|J_h(f)(x)| \le Ch^{\alpha+1}(1+|x-m_f|^{k+1})f(x) + C_N h(h/w)^N. \tag{47}$$

Indeed $J_h(f) = hQ_h * f$, the short-shift remainder has order α, and the farther shifts use uniform absolute derivative bounds and the arbitrary integrated tails.

Choose the origin within a fixed distance of the means of f, g, and put $\rho = f + g$. For a test φ, set $u = \varphi - \frac{1}{2}\int \varphi\rho$. Then

$$\int (1+|x|^2)|u|^2\rho \le C\int |\nabla\varphi|^2\rho. \tag{48}$$

Both components satisfy ordinary Poincaré by Brascamp–Lieb. Comparing their component means to a common Lebesgue average on a fixed ball with a positive density floor gives the mixture inequality. For either potential V, $x\cdot\nabla V(x) \ge m|x|^2/2 - C$, since its mode stays close to zero. Integration by parts gives

$$\int u^2 x\cdot\nabla V f = d\int u^2 f + 2\int ux\cdot\nabla\varphi f.$$

Use Cauchy–Schwarz to absorb half of the resulting quadratic term, add the two estimates, and apply mixture Poincaré. Gaussian tails justify integration by parts for compactly supported tests minus their mean.

Let $I_0 \subset I$ admit a C^1 cutoff supported in I, equal to one on I_0, with $|\nabla\chi| \le C(1+|x|)$. For any identity $f - g = U + \operatorname{div} J$ in distributions,

$$W_2(f,g) \le C\left[\left(\int_{I_0^c} \frac{\rho}{1+|x|^2}\right)^{1/2} + \left(\int_I \frac{|U|^2/(1+|x|^2)+|J|^2}{\rho}\right)^{1/2}\right]. \tag{49}$$

To see this, split $\int (f-g)u$ with the cutoff and use Cauchy–Schwarz with (48) in the exterior. In the interior insert the field identity and integrate by parts. Approximation by smooth tests justifies this for a C^1 cutoff. The $L^2(\rho)$ norms of $(1+|x|^2)^{1/2}\chi u$ and $\nabla(\chi u)$ are at most $C\|\nabla\varphi\|_{L^2(\rho)}$. Riesz representation produces an exact flux of divergence $f-g$ whose squared action norm $\int |J|^2/\rho$ is bounded by the square of the displayed right-hand side. Interpolate between the densities with $(1-\lambda)f + \lambda g$, $\lambda(t) = \sin^2(\pi t/2)$, and flux $\lambda'(t)J$. Its action is at most $\pi^2\int |J|^2/\rho$, by the lower bound $\min(\lambda, 1-\lambda)\rho$. The dynamic transport inequality [4] gives (49).

7.3 Estimating the common weights from inward probes

Use n observations from each group, and split each into three parts of comparable sizes for means, pilot weights and the statistic. Let $a = \log n$, $w = a^{-1/2}$, $h_b = a_{n,d}^{1/(\alpha+1)}$, and $h_j = 2^j h_b$. Include all integers $j \ge 0$ with $h_j \le n^{-\eta/4}$, where $\eta > 0$ is fixed and sufficiently small. Write $r_j = h_j^{\alpha+1}$ and $\rho_j = r_j^2 a^{2-d/2}$. The first splits estimate each mean. Reject when their difference has norm greater than one. Otherwise choose the first estimated mean as origin. Except on an event of probability $o(1)$, uniformly in the experiment, both actual means then lie within a fixed distance of the origin. The mean rule has null error $o(1)$. We condition on the first splits throughout the pilot.

Write $\rho = f + g$ and $\rho^{\mathrm{obs}} = p_f + p_g$. Tile with w-cubes, retaining all cubes meeting the ball of radius $L = D_0\sqrt{a}$. There are $O(a^d)$ cubes. Write x for a cube center. For each observed density p separately, estimate $p(x)$ at inner rows $|x| \le c_1\sqrt{a}$ using its second sample count in the cube divided by the second sample size times w^d. For outer rows put $\omega = x/|x|$, $t = C_1 \log a/\sqrt{a}$, and count in the w-cubes centered at $x - t\omega$ and $x - 2t\omega$. Each count of at least $A_1 \log a$ yields a density estimate at its probe by the same division. If both counts attain this threshold, set the estimate at x to the square of the first probe estimate divided by the second. Otherwise set it to zero. The estimates for the two densities use their own observations.

Uniformly with probability $1 - o(1)$, every count above the threshold agrees with its mean up to fixed factors and every count of mean at least $2A_1 \log a$ is above the threshold. This follows from binomial Chernoff

bounds and a union bound over $O(a^d)$ probes with A_1 large. A probe cube average agrees with its center value up to fixed factors because the observed score on the grid region is $O(\sqrt{a})$. The inner counts have polynomially large means, by the uniform Gaussian lower envelope with c_1 sufficiently small.

On the outer segments, the derivative of $-\log p$ in direction ω is at least $c\sqrt{a}$. Indeed the Hessian is bounded below and the mode is within a fixed distance of the origin. Each inward probe value is therefore at least $a^{cC_1}p(x)$. Since

$$nw^d\rho_0 = nr_0^2 a^{2-d} \gtrsim a^{-1/2},$$

every outer center with $p(x) \geq \rho_0$ has two reliable probes if C_1 is large. The Hessian upper bound gives

$$\big|\log p(x) - 2\log p(x - t\omega) + \log p(x - 2t\omega)\big| \leq Ct^2 = o(1).$$

Thus every positive estimate $\widehat{p}(x)$ is comparable to $p(x)$, and a zero estimate implies $p(x) < \rho_0$. No independence of counts within a sample is needed.

Use this single pilot for all scales. For each j, let I_j be the union of the retained cubes for which the sum of the two estimates is greater than $K_1\rho_j$, and extend that sum constantly on each selected cube as $\widehat{\rho}_j$. On the successful pilot event, uniformly in j,

$$c\rho^{\mathrm{obs}}(x) \leq \widehat{\rho}_j(x) \leq C\rho^{\mathrm{obs}}(x) \quad (x \in I_j).$$

Values within each cube are comparable by the observed score. Any ignored term at a selected center is less than $\rho_0 \leq \rho_j$. Choosing the grid radius $D_0\sqrt{a}$ large also ensures that the observed density sum outside the tiled ball is below ρ_0. This part holds under any admissible covariance, regardless of its size. For power in the range $s \leq n^{-\eta}$, the latent and observed values are comparable throughout the grid by (46). In this range, increasing C_2 if necessary gives

$$c\rho(x) \leq \widehat{\rho}_j(x) \leq C\rho(x) \quad (x \in I_j), \qquad \{\rho > C_2\rho_j\} \subseteq I_j. \tag{50}$$

For the analysis take a smooth cutoff as a scalar function of ρ/ρ_j, zero below C_2 and one above $2C_2$. Its support is in I_j (increase its lower threshold if needed) and its gradient is bounded by $C(1 + |x|)$. For each component density $u = f, g$ integrate on its potential levels about its mode. On a level z above the minimum, the radius and radial derivative are comparable to $\sqrt{z}$. The level-volume derivative is $O(z^{d/2-1})$ and $1 + |x|^2 \gtrsim z$ on large levels. Since $\log(1/\rho_j) \asymp a$, this proves

$$\int_{\{\rho \leq 2C_2\rho_j\}} \frac{\rho(x)}{1 + |x|^2}\, dx \leq C\rho_j a^{d/2-2} = Cr_j^2.$$

Here the domain is contained in the individual low-density domain for each component. Only the analytical cutoff uses the unknown densities.

7.4 A differentiable family of inverse fields

Let $D_d = d(d+1)/2$, and represent symmetric matrices T by their upper triangular entries. Put $\mathcal{T} = [-2, 2]^{D_d}$. Choose an integer $l > D_d/2$ and use all parameter derivatives through order l. At scale j define

$$S_j = \gamma h_j^2 \begin{cases} a, & d < 4, \\ \log(2 + nh_j^{2\alpha+2}/a), & d = 4, \\ \log(2 + nh_j^{2\alpha+d/2}), & d > 4, \end{cases} \qquad M = S_j T.$$

Here $\gamma > 0$ is sufficiently small. These deterministic trial matrices use no noise estimate. Use $b(x)^2 = \delta/(1 + |x|^2)$ for a fixed small $\delta > 0$, $h(x) = h_j$ for $d \leq 4$ and $h(x) = h_j e^{c_0|x|^2}$ otherwise. Take c_0 small so $h(x)$ is polynomially small uniformly on the grid, and $e^{(2\alpha+2)c_0|x|^2}$ has bounded polynomial moments under the latent laws. Let $t_c(x) = h(x)^2 + C_d S_j$ with large fixed C_d. This is below $b(x)^2$ for large n, since $S_j = O(n^{-\eta/2}\log n)$. For every trial and $t_c(x) \leq t \leq b(x)^2$, the covariance $tI - M$ is between fixed positive multiples of tI. Trial matrices themselves may be indefinite.

The scalar kernel at row x is

$$G_{x,T} = \gamma_{b(x)^2 I - M} + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla \gamma_{b(x)^2 I - M}.$$

The coarse vector kernel is $\frac{1}{2}\int_{t_c(x)}^{b(x)^2} \nabla \gamma_{tI-M}\, dt$. The fine vector kernel has symbol, at fixed row parameters,

$$\frac{i\xi}{|\xi|^2}\left(\widehat{K}(h(x)\xi) - e^{-t_c(x)|\xi|^2/2}\right) e^{\xi^T M \xi/2}. \tag{51}$$

Truncate the fine spatial kernel to $|x-y| \le w$ before evaluation at $x-y$. Divide the scalar by $(D(x)\widehat{\rho}_j(x))^{1/2}$, where $D(x) = 1+|x|^2$, and the sum of vector kernels by $\widehat{\rho}_j(x)^{1/2}$. Call the feature $F_{j,T,y}$ in $L^2(I_j)\oplus L^2(I_j;\mathbb{R}^d)$. The pilot and row bandwidths are held fixed as T varies.

At a scale with $S_j \ge s$, the true $M = \Sigma$ is in the trial family. The scalar expectation at the true matrix is the normalized $U(u)$ and the untruncated vector expectation is the normalized $J(u)$ up to (47), where

$$J(u)(x) = \tfrac{1}{2}\int_0^{b(x)^2} \nabla(u * \gamma_t)(x)\, dt, \quad U(u)(x) = (u * \gamma_{b(x)^2})(x) + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla(u * \gamma_{b(x)^2})(x).$$

Gradients in these formulas hold heat time fixed. The heat equation and chain rule give $u = U(u) - \operatorname{div} J(u)$. The coarse and fine vector multipliers sum to $i\xi(\widehat{K}(h(x)\xi) - e^{-b(x)^2|\xi|^2/2})e^{\xi^T M\xi/2}/|\xi|^2$, and at the true matrix the growth factor cancels. The high-pass term costs $O(r_j)$ in weighted norm by (47), the Gaussian moment choice of c_0, and $u^2/\widehat{\rho}_j \le Cu$ on the successful small-noise pilot event.

The scalar kernel before normalization and each of its parameter derivatives is bounded by $Cb(x)^{-d}\exp(-c|x-y|^2/b(x)^2)$. The coarse vector kernel and its parameter derivatives satisfy

$$C(h(x) + |x-y|)^{1-d}\exp(-c|x-y|^2/b(x)^2) \tag{52}$$

as an absolute bound. Each matrix derivative contributes a factor S_j times two spatial derivatives of the Gaussian, and S_j/t is bounded on the integration interval. The parameter derivatives of the gradient kernel are odd in $x-y$. Their Gaussian bounds gain a factor $\min(1, |x-y|/\sqrt{t})$ by the mean value theorem for $|x-y| \le \sqrt{t}$. Integrate over $t \ge h(x)^2$, extracting part of the Gaussian at scale $b(x)$. When $|x-y| \le h(x)$ the bound from the remaining integral is $C|x-y|h(x)^{-d}$. At larger distances it is $C|x-y|^{1-d}$ by splitting at $t = |x-y|^2$. The odd-kernel factor ensures the integral bound also when $d = 1$.

For the fine kernel rescale frequency to $\xi/h(x)$. Then $M/h(x)^2$ is bounded by CA_j, where $A_j = S_j/h_j^2$, and $t_c(x)/h(x)^2$ lies between 1 and $1 + C_d A_j$. The radial bracket in (51) vanishes quadratically and is radial analytic near zero, so dividing by $|\xi|^2$ is smooth. Its rescaled jets on bounded frequencies, including near zero, cost polynomial factors in $1 + A_j$ times e^{CA_j}. For example $(1 - e^{-v|\xi|^2/2})/|\xi|^2 = (v/2)\int_0^1 e^{-uv|\xi|^2/2}\, du$ on the plateau of $\widehat{K}$. Outside the cutoff the Gaussian in the bracket, multiplied by the matrix exponential, has strictly Gaussian decay uniformly by the choice of C_d. The parameter derivatives only add powers of A_j and the rescaled frequency. Therefore the fine spatial kernel and its parameter derivatives are $h(x)^{1-d}$ times rescaled Schwartz profiles at $h(x)$, with the finitely many required seminorms bounded by $E_j = C(1+A_j)^C e^{CA_j}$. Fix these seminorms before choosing γ. Since $A_j = O(\log n)$, $h(x)$ is polynomially small and the pilot weight has a polynomial floor on I_j, the truncation error in the weighted mean norm is smaller than n^{-P} for any required fixed P, using a sufficiently large fixed Schwartz tail order.

For small true noise let $\mu_{j,T}$ be the feature mean difference. At the true trial matrix, the cutoff of Section 3, (49), and these bias bounds give

$$W_2(f,g) \le C(r_j + \|\mu_{j,T}\|). \tag{53}$$

Under the latent null the observed laws agree. The two feature means and all their parameter derivatives are equal at every trial, even when the true covariance is outside that scan.

7.4.1 Conditional covariance

We work on the successful pilot event. The bounds below hold for each parameter derivative through order l and each input law. It suffices to bound the three uncentered second moment operators of the scalar,

coarse vector and fine vector kernels, since their sum multiplied by three dominates the centered covariance in positive operator order. Hilbert–Schmidt norm is monotone in that order. Fix either observed law with density p. Its supporting log-density tangent at each row and the comparison with ρ^{obs} on I_j give

$$\frac{p(y)}{\sqrt{\widehat{\rho}_j(x)\widehat{\rho}_j(x')}} \le C\exp(C(1+|x|)|y-x| + C(1+|x'|)|y-x'|)\,.$$

Each factor is absorbed in the row Gaussian, since $b(x)(1+|x|) \le C$. It is bounded on the near fine support.

The scalar and coarse vector kernels now have the earlier absolute bounds with a possibly smaller Gaussian exponent. The same holds for the differentiated fields.

For $d < 4$, at pairs with $1+|x| \asymp 1+|x'|$ (ratio between $1/2$ and 2), use common bandwidth $b(x)$. The scalar covariance kernel has squared row integral bounded by $Cb(x)^{-d}D(x)^{-2}$. For the coarse vector use $C|z|^{1-d}e^{-c|z|^2/b(x)^2}$, whose $L^{4/3}$ norm is $O(b(x)^{1-d/4})$; Young's inequality gives the same row bound $Cb(x)^{4-d}$. For noncomparable pairs, the separation is comparable to the larger radius. Extract a Gaussian in separation using both row Gaussians with bounded bandwidth. Scalar prefactors then grow at most polynomially in the radii. For the vector, split into balls of radius one third of the separation about either row and their complement. On each ball the other singularity is bounded by a polynomial in separation and the singularity at the ball center is integrable against its Gaussian. On the complement both singularities are bounded. The two squared kernels are thus integrable on this pair region uniformly in n. Integrating the comparable row bound $Cb(x)^{4-d}$ on I_j gives $O(1), O(\log a), O(a)$ in dimensions one, two, three respectively.

For $d \ge 4$, the scalar argument on comparable and noncomparable radii above again gives squared norm bounded by $C(1+\int_{I_j} b(x)^{4-d}\,dx)$. This is $O(a^2)$ in dimension four and polylogarithmic above four. For the coarse vector, the convolution bound at $z = x - x'$, $|z| \le 2w$, is

$$C(h(x)+|z|)^{2-d}e^{-c|z|^2}.$$

The two regularizations are comparable there. For $|z| \le h(x)$ apply Cauchy–Schwarz to the two profiles. Above that scale, split near either singular center and in the remainder to obtain $C|z|^{2-d}$ before the Gaussian: the two radial integrals are bounded by $|z|^{1-d}\int_0^{|z|} dt$ and $\int_{|z|}^{\infty} t^{1-d}\,dt$. The same unregularized bound applies at $|z| > 2w$. In dimension four the squared near integral in z is $O(a)$, and the far contribution per row is $O(1+\log(1/w))$. These give $O(a^3)$ after integration in x. For $d > 4$ the squared near bound is $C\int_{I_j} h(x)^{4-d}\,dx \le Ch_j^{4-d}$. The far bound is polylogarithmic and is absorbed by h_j^{4-d}.

The fine operator has uncentered second-moment kernels supported on pairs at distance at most $2w$. There the h's are comparable. Convolution of the rescaled profiles with the bounded score tilts shows the absolute kernel is at most

$$CE_j^2 h(x)^{2-d}(1+|x-x'|/h(x))^{-N_0}$$

for a sufficiently large fixed N_0. Integrating its square first in x' gives Hilbert–Schmidt norm at most $CE_j^2(\int_{I_j} h(x)^{4-d}\,dx)^{1/2}$. This is $O(E_j^2 h_j^{2-d/2}a^{d/4})$ for $d<4$, $O(E_j^2 a)$ for $d=4$, and $O(E_j^2 h_j^{2-d/2})$ for $d>4$.

For either group, all these bounds imply that the covariance $C_{j,T,\beta}$ of the feature differentiated with respect to T by a multiindex β, $|\beta| \le l$, satisfies

$$n^{-1}\|C_{j,T,\beta}\|_{\mathrm{HS}} \le Cr_j^2 2^{-c_2 j} \tag{54}$$

for a fixed $c_2 > 0$. First, the scalar and coarse vector bounds divided by n are at most $Ca_{n,d}^2 = Cr_0^2$. In $d > 4$, use $h_j^{2-d/2} \le h_b^{2-d/2}$ and $h_b^{2-d/2}/n = r_0^2$. For $d < 4$, choose γ so $E_j^2 \le Cn^c$ for a small $c > 0$. The fine bound divided by nr_j^2 is at most

$$Cn^{c-1}h_j^{-2\alpha-d/2}a^{d/4}.$$

At $j = 0$ this tends to zero for $c < (2-d/2)/(2\alpha+2)$, and the decay in j is geometric. For $d = 4$, put $X_j = 2 + nr_j^2/a$. Then $X_j - 2 \ge a^{1/2}2^{(2\alpha+2)j}$. Here $E_j^2 \le CX_j^c$ for small c, and $E_j^2 a/(nr_j^2) \le CX_j^{c-1}$. For $d > 4$, put $X_j = 2 + nh_j^{2\alpha+d/2}$. Then $X_j - 2 = 2^{(2\alpha+d/2)j}$. Again $E_j^2 \le CX_j^c$, and $E_j^2 h_j^{2-d/2}/(nr_j^2) \le CX_j^{c-1}$. These estimates give (54) uniformly in the trial cube and under every noise covariance, since the covariance calculation used observed density comparability.

7.5 The covariance and resolution scan

Use paired held-out observations from the third splits, of size $N \asymp n$. For each j, T let $Z_i(T) = F_{j,T,Y_i} - F_{j,T,Y_i'}$, and set

$$Q_j(T) = \frac{1}{N(N-1)} \sum_{i \neq k} \langle Z_i(T), Z_k(T) \rangle.$$

Condition on the earlier splits. Each feature is smooth in T on the cube, as follows from the kernel derivatives and finite row integrals with positive pilot weights. The supremum of $Q_j(T)$ is measurable since it can use a countable dense subset of the cube. Under the null all $Z_i(T)$ and all their derivatives have mean zero for every j, T.

For a derivative of Q_j through order l, apply the product rule inside the sum. Each term is a degenerate bilinear statistic with summands $\langle A_i, B_k \rangle$, where A_i, B_i are different derivatives of $Z_i(T)$. They are centered and independent across pairs. In the second moment of the statistic only identical or swapped index pairs survive. For distinct i, k,

$$\mathbb{E}|\langle A_i, B_k \rangle|^2 = \operatorname{tr}(C_A C_B) \leq \|C_A\|_{\mathrm{HS}} \|C_B\|_{\mathrm{HS}},$$

and swapped pairs use the same bound by Cauchy–Schwarz. The covariance of a derivative of a paired difference is the sum of the two group covariances. Thus (54) gives a conditional second moment bound $C r_j^4 2^{-2c_2 j}$ for each derivative of Q_j. Use the fixed-domain Sobolev inequality

$$\sup_{T \in \mathcal{T}} |Q_j(T)|^2 \leq C \sum_{|\beta| \leq l} \int_{\mathcal{T}} |\partial_T^\beta Q_j(T)|^2 \, dT, \qquad l > D_d/2,$$

and take conditional expectation. Reject if for any included j, $\sup_T Q_j(T) > A_\varepsilon r_j^2$. Markov's inequality and summation of $2^{-2c_2 j}$ bound the type-I error by any prescribed constant on the pilot event. The mean rejection and the pilot failure together cost $o(1)$. In particular the scan costs no factor counting the scales.

7.5.1 Coverage of the actual noise scale

Suppose $s \leq n^{-\eta}$. Put

$$h_* = \max\{h_b, \sqrt{s / B_{n,d}(s)}\}.$$

There is an included dyadic bandwidth $h_j \asymp h_*$ with $S_j \geq s$, where the comparison constants can depend on γ, α, d. For $d < 4$, take a sufficiently large fixed multiple of h_* before rounding up. For $d > 4$, let $p = \alpha + d/4$, $x = ns^p$, $B = \log(2 + x)$. At h_*, the logarithm defining S_j is at least $\log(2 + \max(1, x/B^p))$, which is comparable to B. Indeed $x/B^p \geq x^{1/2}$ for all sufficiently large x, and bounded x costs a fixed constant. For $d = 4$, let $p = \alpha + 1$ and use the same x, B. Now the logarithm is at least $\log(2 + \max(a^{1/2}, x/(aB^p)))$. For $x \leq a^2$ the $a^{1/2}$ term gives a fixed fraction of B. If $x > a^2$, then $x/(aB^p) \geq x^{1/4}$ for sufficiently large n. In either case the required logarithm is at least cB. Enlarging h_* by a fixed factor pays the constant γ, and rounding to a dyadic point preserves the bound. By taking η small, $h_b \leq n^{-\eta/2}$, and $s \leq n^{-\eta}$ gives $h_* \leq C n^{-\eta/2}$. Thus the selected bandwidth is below the cap for large n. This proves $r_j \asymp r_{n,d}(s)$.

At that trial use $T = \Sigma / S_j$. Conditional on the pilots the statistic has mean $\|\mu\|^2$ and variance

$$\frac{4}{N} \langle \mu, C\mu \rangle + \frac{2}{N(N-1)} \|C\|_{\mathrm{HS}}^2,$$

where C is the centered covariance of one paired feature difference. The formula follows by separating the linear terms and the centered degenerate sum. By (54) the variance is at most $C_1(r_j^4 + r_j^2 \|\mu\|^2)$. If the separation is a large enough constant times $r_{n,d}(s)$, then (53) makes $\|\mu\|^2$ a sufficiently large multiple of the rejection threshold. Chebyshev's inequality gives the type-II bound at this one matrix and scale, so the scan rejects. If the preliminary mean rule rejects first, rejection is already guaranteed. The exceptional pilot probabilities are $o(1)$.

7.6 A common empirical test at larger noise

If $\delta = W_1(p_f, p_g) \le 1$, their means are within one. Translate by the mean of g and apply (49) with $U = U_h(f) - U_h(g)$, $J = J_h(f) - J_h(g)$ on a ball of radius $D\sqrt{\log(1/h)}$. Take D large enough that the Gaussian exterior contributes $O(h^{\alpha+1})$. On this ball there is a uniform density floor h^P for some fixed P, and (47) gives an $O(h^{\alpha+1})$ weighted bound on the vector field, paying each local bias in its own density. For the actual common covariance, the Gaussian characteristic function can be removed on the Fourier cutoff at cost at most $e^{s/(2h^2)}$. Coupling the output laws bounds their characteristic function difference by $\delta|\xi|$. Fourier inversion on the cutoff support then gives $\|U\|_\infty \le C\delta h^{-d-1} e^{s/(2h^2)}$. Consequently

$$W_2(f,g) \le C\left[h^{\alpha+1} + \delta h^{-d-1-P/2} \log^{d/4}(1/h) e^{s/(2h^2)}\right]. \tag{55}$$

Uniformly over the class, the empirical observed W_1 error e_n has expectation $O(n^{-\beta})$ with $\beta = 1/[4(d+1)]$. For an elementary proof translate to the theoretical mean, truncate at distance n^β using the uniform second moment and then quantize at mesh $n^{-\beta}$. Both costs are $O(n^{-\beta})$. There are $O(n^{2d\beta})$ sites, so Cauchy–Schwarz for the Bernoulli counts bounds expected total variation by $Cn^{d\beta-1/2}$. Multiplying by the grid diameter gives $O(n^{-1/4})$. Use the empirical observed W_1 distance between both samples, rejecting above $2n^{-\beta/2}$. Under the null, Markov's inequality and the triangle inequality give error $o(1)$. For any alternative whose observed discrepancy is $\delta \le 5n^{-\beta/2}$, use

$$h = \max(n^{-c_1}, \sqrt{s/(\gamma_1 \log n)})$$

in (55), with constants chosen so that $c_1(\alpha+2+d+P/2)+\gamma_1/2 < \beta/2$. The two terms in (55) are then $O(h^{\alpha+1})$. If $s \ge n^{-\eta}$ with $\eta > 0$ sufficiently small, the second entry of h dominates and $\log(2+ns^{\alpha+d/4}) \asymp \log n$. Thus for a sufficiently large multiple of $r_{n,d}(s)$ in latent distance, the observed discrepancy exceeds $5n^{-\beta/2}$. If each empirical error is at most $n^{-\beta/2}$, the empirical discrepancy is greater than $3n^{-\beta/2}$ and the test rejects. The exceptional probability is $o(1)$ by the two Markov bounds.

Add this empirical test to the scan rejection rule at every n. It has null error $o(1)$ uniformly over the covariance and its rejection rule does not use s. The power argument above only uses s analytically and applies for every $s \ge n^{-\eta}$.

7.7 Lower bounds from mixtures

Fix the common covariance at any allowed matrix. Keep the larger group at a centered isotropic Gaussian g with strictly interior precision, and vary only the smaller group. The unchanged independent group cancels out of likelihood ratios. Write n for the smaller size and s for the largest eigenvalue of this fixed covariance. The direct mixtures hold after arbitrary convolution and the inverse packet can be aligned with a top covariance eigenvector.

We will use finite families $f_\omega = g + \sum_{i=1}^N \omega_i q_i$ where ω_i are independent signs and each q_i has integral zero. Let G be the Gram matrix of the noisy score perturbations in $L^2(p_g)$:

$$G_{ij} = \int \frac{(q_i * \gamma_\Sigma)(q_j * \gamma_\Sigma)}{g * \gamma_\Sigma}.$$

For the uniform mixture of n-sample laws over signs, one plus the chi-square divergence to the null equals $\mathbb{E}_{\omega,\omega'}(1+\omega^T G\omega')^n$. The factor inside the power is nonnegative as the inner product of two likelihood ratios. Bounding its power by $\exp(n\omega^T G\omega')$ and integrating one sign vector gives $\mathbb{E}_\omega \exp(n^2\|G\omega\|^2/2)$ as an upper bound, using $\cosh(t) \le e^{t^2/2}$. Introduce a standard Gaussian vector to linearize the square and apply the same sign bound again. The result is at most

$$\det(I - n^2G^2)^{-1/2} \le \exp(Cn^2\|G\|_{\mathrm{HS}}^2) \qquad \text{if } n\|G\|_{\mathrm{op}} \le 1/2.$$

Thus a small bound on $n^2\|G\|_{\mathrm{HS}}^2$ makes the mixture total variation arbitrarily small. If every mixture member has $W_2(f_\omega, g) \ge r$, the sum of testing errors at separation r is at least one minus that total variation.

For transport separation we use a smooth test φ. Taylor expansion under an optimal coupling gives

$$\int \varphi(f_\omega - g) \le \|\nabla\varphi\|_{L^2(f_\omega)} W_2(f_\omega, g) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(f_\omega, g)^2. \tag{56}$$

If the pairing dominates cE, the gradient norm is at most $C\sqrt{E}$, and the Hessian norm is bounded, it follows that $W_2^2(f_\omega, g) \geq c'E$.

7.7.1 The direct polynomial lower bound

For $d > 4$, put $h = n^{-2/(4\alpha+d)}$. Pack order h^{-d} disjoint cubes of side h in a fixed cube about the reference mode. Let ψ be a nonzero smooth bump supported inside the reference unit cube with integral zero, and use $q_i(x) = a_0\psi((x - z_i)/h)$, $a_0 = \kappa h^\alpha$. The relative perturbations q_i/g have uniformly bounded C^α norm of order κ on the combined support, by rescaling and the fixed positive density floor there. The vanishing near edges gives the same bound across cubes. The logarithmic derivative bounds follow by composition with $\log(1 + \cdot)$; for sufficiently small κ the reference slack places every alternative in the class.

The clean score Gram matrix is diagonal with entries bounded by $Ca_0^2h^d$. Observed score Gram matrices are bounded above by their clean counterpart in positive operator order: the observed score is the conditional expectation of the clean score under the reference joint law, so this follows by conditional Cauchy–Schwarz for any linear combination. Thus $n^2\|G\|_{\mathrm{HS}}^2 \leq C\kappa^4 n^2 h^{4\alpha+d} = C\kappa^4$. For separation take $\varphi = h^2 \sum_i \omega_i q_i$. Its pairing with $f_\omega - g$ is of order $E = h^{d+2} \sum_i a_0^2 \asymp h^{2\alpha+2}$, by disjointness. The gradient's squared L^2 norm is $O(E)$, and its Hessian is uniformly bounded. Equation (56) therefore gives the direct lower bound for every sign.

7.7.2 The direct critical lower bound

When $d = 4$, pack disjoint cubes of side $\ell = (\log n)^{-1/2}$ inside a cube of side $c\sqrt{\log n}$ about the mode, with $c > 0$ small. Their total volume is order $(\log n)^2$. The reference g is comparable to its value g_B at the center of each cube B, and $g_B \geq n^{-\beta_0}$ for a fixed $\beta_0 < 1$. In the rescaled unit cube the ratio g_B/g has bounded C^α norm uniformly: on length ℓ, the score bound is paid by $\ell\sqrt{\log n} = 1$, and all higher potential derivatives are bounded with the corresponding scale factors. The top Hölder bound scales by ℓ^α.

Use a fixed orientation of compactly supported orthonormal tensor wavelets [7] of regularity C^{k+1}, supported strictly inside each cube. At physical length $r_i = \ell 2^{-j}$ their number in a cube is comparable to $(\ell/r_i)^4$, for j from a sufficiently large fixed index through $\lfloor \nu \log_2 n \rfloor$, with $\nu > 0$ small. Write ψ_i for these L^2-normalized wavelets. In particular $S_4 = \sum_i r_i^4 \asymp (\log n)^3$: each level contributes order the total volume. Set

$$b_i = \frac{\kappa^2 r_i^2}{n\sqrt{S_4}}, \qquad a_i = \sqrt{g_{B(i)} b_i}, \qquad q_i = a_i\psi_i.$$

Each wavelet has zero integral. At physical scale r, the relative perturbation in cube B has C^α norm at most $C(a_i/g_B)r^{-2-\alpha}$, by wavelet scaling, bounded overlap at the scale, and the rescaled reciprocal density bound. Summing over scales gives at most

$$C\kappa n^{-(1-\beta_0)/2} S_4^{-1/4} r_{\min}^{-1-\alpha} = o(1)$$

if $\nu < (1-\beta_0)/(2(1+\alpha))$. For fractional order this controls the global seminorm: on a pair of points in distinct cubes one may use the zero values of the top derivatives on the boundary of each cube and distances to the first exit points on the segment. Thus all perturbations stay within the reference margins by logarithmic composition.

The clean Gram matrix is $D_0 H D_0$ with $D_0 = \mathrm{diag}(\sqrt{b_i})$, and $H \preceq CI$. This uses orthonormality and the uniform bounds on g_B/g within each cube. The noisy Gram shrinks in positive operator order as before, so $n^2\|G\|_{\mathrm{HS}}^2 \leq Cn^2 \sum_i b_i^2 = C\kappa^4$. For the metric test take $\varphi = \sum_i (r_i^2/g_{B(i)})\omega_i a_i \psi_i$. Its pairing is exactly $E = \sum_i r_i^2 b_i = \kappa^2\sqrt{S_4}/n$. The gradient squared norm in a cube is controlled by the sum of squared coefficient norms weighted by r_i^{-2}, up to a fixed factor. For clarity, within one level sum u_j, $\|\Delta u_j\|_2 \leq C r_j^{-2}\|u_j\|_2$ by overlap. When $r_j \geq r_\ell$, integration by parts yields

$$|\langle \nabla u_j, \nabla u_\ell \rangle| \leq C(r_\ell/r_j)(\|u_j\|_2/r_j)(\|u_\ell\|_2/r_\ell).$$

Summing the geometric ratios and multiplying by the comparable density in each cube gives squared gradient norm $O(E)$ under f_ω. The test Hessian is uniformly bounded, since its per cube scale sum is bounded by $C\kappa(ng_B)^{-1/2} S_4^{-1/4} r_{\min}^{-1}$. Thus (56) gives separation at least $c(\log n)^{3/4}/\sqrt{n}$. This lower bound holds under every known Gaussian covariance, by the same mixture contraction.

7.7.3 The inverse mixture

Rotate axes so $\Sigma = \mathrm{diag}(s, t_2, \dots, t_d)$ with $0 \le t_i \le s$, and suppose $0 < s \le s_0$ with $s_0 > 0$ small. Place $N \asymp s^{-d/2}$ centers z_i spaced $D\sqrt{s}$ apart in a fixed cube. Here D is sufficiently large, and s_0 small after choosing D. Put $T^2 = A_0 \log(2 + ns^{\alpha+d/4})/s$, and use

$$q_i(x) = a_0 \psi_i(x), \qquad a_0 = \kappa T^{-\alpha}, \qquad \psi_i(x) = e^{-|x-z_i|^2/(2s)} \left(\cos(T(x_1 - z_{i,1})) - e^{-sT^2/2} \right).$$

Each packet has zero mass. The relative perturbation before the factor a_0 has derivatives through order $j \le k$ bounded by CT^j. Indeed derivatives of the packets are bounded by $CT^j e^{-c|x-z_i|^2/s}$ because $T\sqrt{s}$ is large, and division by g costs only Gaussian and polynomial factors absorbed in a weaker envelope for small s_0. Separated grid envelopes have uniformly bounded sum. For $0 < \theta < 1$, the k-th derivative has θ-Hölder seminorm at most CT^α. For shifts longer than $1/T$ use the sup norm. For a shorter shift expand the product-rule terms of $D^k(\psi_i/g)$. A difference of a smooth packet derivative of order j costs $CT^{j+1}|u|$ with the same weakened envelope. A difference of a reciprocal derivative of order $k - j$ costs $C|u|^\theta$ times polynomial factors and the maximum of the two reciprocal densities, by the product rules and the reference Hölder condition on $D^k V_g$. The nearby packet envelopes still absorb those factors since $|u| \le 1/T \le C\sqrt{s}$. Summing the bounds gives $CT^\alpha |u|^\theta$. The logarithmic perturbation thus costs $O(\kappa)$ in the stipulated norms and preserves the class for small enough κ.

Convolution with the noise changes the packet envelope to variances $s + t_i$, with $t_1 = s$, multiplying its amplitude by $\prod_i \sqrt{s/(s+t_i)}$. Its bracket becomes

$$e^{-sT^2/4} \cos\big(\tfrac{T}{2}(y_1 - z_{i,1})\big) - e^{-sT^2/2}.$$

These formulas follow by conditional Gaussian integration and include zero eigenvalues in the other coordinates. The convolved reference has a lower Gaussian envelope, obtained by convolving one for g. Combining the two narrow packet envelopes and dividing by this lower envelope gives

$$|G_{ij}| \le C a_0^2 s^{d/2} e^{-csT^2} e^{-c|z_i - z_j|^2/s}.$$

Indeed the common center lies in a fixed cube, the variances of the two envelopes are between s and $2s$, and choosing s_0 small absorbs the reciprocal reference into a weaker Gaussian. Thus

$$n\|G\|_{\mathrm{HS}} \le Cn a_0^2 s^{d/4} e^{-csT^2}.$$

With $X = ns^{\alpha+d/4}$, the right-hand side is at most $C\kappa^2 X (A_0 \log(2+X))^{-\alpha} (2+X)^{-cA_0}$. It is arbitrarily small by choosing A_0 large and κ small.

For transport separation take $\varphi = T^{-2} \sum_i \omega_i q_i$. The spatial L^2 Gram matrix of the ψ_i has diagonal at least $cs^{d/2}$ by integrating the squared cosine in its Gaussian envelope and taking $T\sqrt{s}$ large. The absolute off-diagonal row sums are at most $Cs^{d/2} \sum_{j \in \mathbb{Z}^d \setminus \{0\}} e^{-cD^2|j|^2}$, less than half the diagonal bound for large D. The pairing with $f_\omega - g$ is at least cE for $E = T^{-2} a_0^2 s^{d/2} N$. The gradient Gram has absolute row sums at most $CT^2 s^{d/2}$ by the same overlap estimate, so the test gradient has squared norm $O(E)$ under the uniformly bounded vertex densities. The pointwise derivative envelope bounds the test Hessian by Ca_0. Equation (56) gives $W_2(f_\omega, g) \ge cT^{-\alpha-1}$ for every sign.

For $s \ge s_0$, use instead one additive packet throughout, replacing its envelope variance by a fixed τ_0^2 smaller than the variance in the Gaussian lower envelope for g, with its cosine mean removed. Use amplitude $\kappa T^{-\alpha}$ and $T^2 = C \log n/s$. The derivative and Hölder argument above applies to this single packet. Its convolved Gaussian envelope squared, divided by the convolved reference lower envelope, is integrable uniformly in Σ by the variance choice. Its first-coordinate cosine attenuates by $\exp(-\tau_0^2 sT^2/[2(\tau_0^2 + s)])$, so the n-sample chi-square tends to zero uniformly for sufficiently large C. Testing the latent difference with $\cos(Tx_1)/T$ gives separation of order $T^{-\alpha-1}$. This matches the inverse term since its logarithm is comparable to $\log n$ on this range.

7.7.4 Lower bounds in dimensions one through three

In dimension one a shift of the reference by δ has W_2 distance $|\delta|$. Its clean one-sample relative entropy to the reference is $O(\delta^2)$, by expanding the potential and using the zero mean score. Since noise contracts entropy, $\delta = c/\sqrt{n}$ gives the desired two-point lower bound.

For $d = 2, 3$, take disjoint annuli at dyadic radii R from a fixed constant to $c\sqrt{\log n}$, for small fixed $c > 0$. Pack cubes of side $\ell_i \asymp R^{-1}$ in each annulus, with order R^{2d} cubes at radius R. The reference values g_i at their centers satisfy $g_i \geq n^{-\beta_0}$, $\beta_0 < 1$, and $g \asymp g_i$ inside each cube. Use one L^2-normalized smooth mean-zero bump ψ_i per cube, with

$$S_4 = \sum_i \ell_i^4, \qquad b_i = \frac{\kappa^2 \ell_i^2}{n\sqrt{S_4}}, \qquad q_i = \sqrt{g_i b_i}\, \psi_i.$$

The rescaled reciprocal bounds used for the critical construction apply here since $\ell_i(1 + |x|) = O(1)$. All relative C^α perturbations tend to zero because $g_i \geq n^{-\beta_0}$ and $\ell_i \gtrsim (\log n)^{-1/2}$. The clean Gram matrix is diagonal with entries $O(b_i)$, and the noisy Gram contracts. Hence $n^2 \|G\|_{\mathrm{HS}}^2 = O(\kappa^4)$. The test $\varphi = \sum_i (\ell_i^2/g_i)\omega_i q_i$ has pairing $\sum_i \ell_i^2 b_i$, squared gradient norm at most a constant times that pairing, and bounded Hessian by rescaling on the disjoint cells. Equation (56) gives squared separation of order $\sqrt{S_4}/n$. A ring contributes order R^{2d-4} to S_4; summing gives order $\log\log n$ for $d = 2$ and $\log n$ for $d = 3$.

For the inverse lower bound in all three dimensions, the single additive narrow packet with $T^2 = C \log n/s$ works throughout $0 < s \leq \Lambda$. Its attenuated cosine multiplier is bounded by e^{-csT^2} uniformly for the fixed narrow variance, and the noisy reference denominator is paid as above. It gives latent separation of order $(s/\log n)^{(\alpha+1)/2}$ with small total variation. Together with the direct lower bounds this completes Theorem 7.1.

7.8 Common noise with repeated measurements

Suppose each group also has k_i independent new individuals, each supplying two measurements with conditionally independent Gaussian errors given the same latent draw and with the same unknown error covariance as the main sample in both groups. The rate then uses $n = \min_i(n_i + k_i)$ and the same actual noise norm s. For the upper bound, keep one measurement per individual. This gives $n_i + k_i$ independent observations at the shared noise. For the lower bound, reveal the covariance and give all $n_i + k_i$ observations at noise covariance $\Sigma/2$. The main sample at Σ can be produced by independent Gaussian additions. The remaining cleaner observations are the replicate averages, and independent Gaussian differences complete the pair experiment. Thus the more informative experiment already has the lower order with $s/2$, which is comparable to the stated order.

7.9 Comparison with measurement-error tests

Lee [23, pp. 9–10] discusses consistency of direct tests when the errors and replicate counts are common, using the nonvanishing common error characteristic function, which makes convolution injective. The deconvolved test in [17] treats a broader range of measurement errors and addresses asymptotic calibration. Broniatowski et al. [22] scan a scalar noise nuisance over simple hypotheses about the latent density. Butucea, Matias, and Pouet [24, Corollary 3] show lossless adaptation for univariate fixed-null L^2 testing with an estimated stable exponent and a Fourier lower bound on the signal. The present separation guarantee is uniform over two unknown densities and a shared PSD covariance of unknown norm. The covariance scan costs no rate factor at any true noise level, including at the direct-to-inverse transition.

8 Joint adaptation to smoothness and Gaussian noise in Wasserstein testing

We study latent quadratic Wasserstein testing in dimensions above four when potential smoothness and a common Gaussian error covariance may both be unknown. If the covariance is known, adaptation to smoothness inflates the clean sampling scale but preserves the known-smoothness inverse logarithm. If both are unknown, the effective sample size $n/\sqrt{\log\log n}$ also enters the inverse logarithm. We give a single test with this rate and matching mixtures on moving noise levels. The latter alternatives have different smoothness and covariance but the same observed null. They show a diverging difference between the procedures with and without the covariance near the clean-to-inverse transition.

8.1 Separation with unknown smoothness

Fix $d > 4$ and $2 \le \alpha_- < \alpha_+ < \infty$. For $\alpha = k + \theta$ in this interval let $\mathcal{F}_\alpha$ be the translated class of probability densities $f = e^{-V}$ on $\mathbb{R}^d$ with $V \in C^k$, $mI \preceq D^2 V \preceq MI$ for fixed $0 < m < M$, and fixed positive upper bounds on the sup norms of derivatives of orders $3, \dots, k$. If $0 < \theta < 1$, impose a fixed positive upper bound on the global θ-Hölder seminorm of $D^k V$. Use uniform finite caps for the finitely many derivative orders and the fractional seminorm. The positive caps are common over the interval. The statements do not require estimating α. We observe independent samples of sizes n_1, n_2 from $p_f = f * \gamma_\Sigma$ and $p_g = g * \gamma_\Sigma$, where $\gamma_\Sigma = N(0, \Sigma)$ and the error covariance $\Sigma \succeq 0$ is common to both groups. The covariance may be singular. Write $n = \min(n_1, n_2)$, $s = \|\Sigma\|_{\mathrm{op}} \le \Lambda$ and $\gamma_t = N(0, tI_d)$ for scalar t. The null is latent equality $f = g$ in the union of the classes.

Put $L = \log\log n$, $N = n/\sqrt{L}$, and $p = \alpha + d/4$. Define the clean adaptive scale and two inverse scales in distance by

$$A(\alpha) = N^{-(\alpha+1)/(2p)}, \qquad I_m(\alpha, s) = \left(\frac{s}{\log(2 + ms^p)}\right)^{(\alpha+1)/2} \qquad (m = n, N).$$

The inverse term is zero at zero noise. For the stated separation rates, errors can be prescribed to be any fixed positive constants; the comparison constants can depend on those errors and the fixed class parameters.

Theorem 8.1 (Upper bounds)**.** *Suppose $f, g \in \mathcal{F}_\alpha$ at an unknown $\alpha \in [\alpha_-, \alpha_+]$. If the common covariance is supplied, there is a test that detects at separation $C(A(\alpha)+I_n(\alpha, s))$. If the common covariance is unknown, there is one test that detects at separation $C(A(\alpha) + I_N(\alpha, s))$ simultaneously at all α, Σ. Both tests have uniformly controlled type-I error over the respective composite nulls, and the power bounds are uniform as $n \to \infty$.*

Theorem 8.2 (Adaptive lower bounds)**.** *Uniformly over the smoothness interval, the order $A(\alpha)$ is necessary in the clean experiment. More generally, the same clean lower bound holds after a fixed common Gaussian convolution, with the lower quantifier taken over the interval.*

For the joint unknown-covariance problem, fix $\delta > 0$ and any sequence $1 \le Y_n \le N^{1-\delta}$. Even if the value Y_n is specified, there are alternatives with $Ns^p = Y_n$ and separation at least $c(A(\alpha) + I_N(\alpha, s))$ for which uniform fixed-error testing is impossible. The lower holds with varying α in any positive-width interior subinterval and isotropic covariances. More precisely, the infimum sum of worst errors at that separation tends to one.

The clean lower and the known-covariance inverse lower are separate obstructions. A lower bound uniform in α does not assert an adaptation penalty at each specified α. The moving-level lower in Theorem 8.2, in contrast, gives a sharp joint inverse order across the specified strata.

Corollary 8.3. *There are moving noise levels and alternatives for which the order with known covariance is smaller than the necessary jointly adaptive order by a factor tending to infinity.*

The classical adaptive penalty is already present in wavelet testing in white noise [29]. Tang and Yang [13] study compact-domain testing for smooth discrepancies. Shekhar, Kim, and Ramdas [28] use a split kernel statistic for two-sample testing, while Schrab et al. [27] aggregate kernel bandwidths under direct L^2 separation. Butucea, Matias, and Pouet [26, 24] study indirect fixed-null L^2 tests with adaptive smoothness and, in a partially known noise model, an identifiable stable exponent. The fixed positive known Gaussian Wasserstein inverse problem in the present curvature class appears in [12]. Here full W_2 requires spatially weighted transport fields, and the lower test mixture exploits the unidentifiability of the shared covariance under a common observed Gaussian null.

8.2 Density bounds and a weighted transport comparison

Uniformly over the class, f has Gaussian upper and lower envelopes about its mean m_f, the score of f is $O(1 + |x - m_f|)$, and

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

Indeed quadratic bounds at the mode give the envelopes and bound the distance between the mode and the mean. Differentiating e^{-V} gives the derivative bounds. For $\theta > 0$, on $|x - m_f| \le L$ and $|u| \le (2 + L)^{-1}$, the same product rule and the Hölder bound on $D^k V$ give

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

The local density ratio is bounded by the score, derivatives of lower order in the product are Lipschitz, and the highest one uses the assumed Hölder seminorm.

For $p_f = f * \gamma_s$, isotropic $0 \le s \le \Lambda$, Gaussian integration of the two quadratic Taylor bounds gives

$$cf(x) \le p_f(x) \le f(x) e^{Cs(1+|x-m_f|^2)}. \tag{57}$$

The potential of p_f also has Hessian in a fixed positive interval. For $s > 0$, its Hessian is $s^{-1} I - s^{-2} \operatorname{Cov}(X \mid Y)$. The conditional potential has Hessian between $(m + s^{-1})I$ and $(M + s^{-1})I$. The upper covariance bound follows from Brascamp–Lieb [5]. The lower bound follows, in each unit direction v, from Cauchy–Schwarz and the conditional identities

$$\mathbb{E}[(v \cdot (X - \mathbb{E}X))(v \cdot \nabla A(X))] = 1, \qquad \mathbb{E}[(v \cdot \nabla A(X))^2] = \mathbb{E}[v^T D^2 A(X) v],$$

where A is the conditional potential. Consequently the scores of p_f are $O(1 + |x - m_f|)$. The observed law always has a uniformly Gaussian upper envelope about m_f.

Use $\widehat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\, dx$. Let $\widehat{K}$ be a real radial smooth Fourier cutoff that equals one near zero and is supported in the unit ball. Use $K_b = b^{-d} K(\cdot/b)$ for its spatial kernel, and define constant-bandwidth fields

$$U_b(f) = K_b * f, \qquad \widehat{J_b(f)}(\xi) = -\frac{i\xi}{|\xi|^2}(1 - \widehat{K}(b\xi))\widehat{f}(\xi).$$

Then $f = U_b(f) + \operatorname{div} J_b(f)$. The inverse transform Q of $-i\xi(1 - \widehat{K}(\xi))/|\xi|^2$ is integrable with all absolute moments finite and all signed moments zero. In fact a dyadic partition into shells 2^j away from the origin gives inverse transforms bounded by $C_N 2^{(d-1)j}(1 + 2^j|x|)^{-N}$. Summability of their L^1 norms $O(2^{-j})$ gives the assertion, with moment cancellation since the symbol vanishes near zero. For $|x - m_f| \le L$, $w = (2+L)^{-1}$, and $h \le w$, subtract the local Taylor polynomial of f to degree $k - 1$ in the integer case and k otherwise. This gives

$$|J_h(f)(x)| \le Ch^{\alpha+1}(1 + |x - m_f|^{k+1}) f(x) + C_N h (h/w)^N. \tag{58}$$

Indeed $J_h(f) = hQ_h * f$, the short-shift remainder has order α, and the farther shifts use uniform absolute derivative bounds and the arbitrary integrated tails.

Choose the origin within a fixed distance of the means of f, g, and put $\rho = f + g$. For a test φ, set $u = \varphi - \frac{1}{2}\int \varphi\rho$. Then

$$\int (1 + |x|^2)|u|^2 \rho \le C \int |\nabla\varphi|^2 \rho. \tag{59}$$

Both components satisfy ordinary Poincaré by Brascamp–Lieb. Comparing their component means to a common Lebesgue average on a fixed ball with a positive density floor gives the mixture inequality. For either potential V, $x \cdot \nabla V(x) \ge m|x|^2/2 - C$, since its mode stays close to zero. Integration by parts gives

$$\int u^2 x \cdot \nabla V f = d \int u^2 f + 2 \int ux \cdot \nabla\varphi f.$$

Use Cauchy–Schwarz to absorb half of the resulting quadratic term, add the two estimates, and apply mixture Poincaré. Gaussian tails justify integration by parts for compactly supported tests minus their mean.

Let $I_0 \subset I$ admit a C^1 cutoff supported in I, equal to one on I_0, with $|\nabla\chi| \le C(1 + |x|)$. For any identity $f - g = U + \operatorname{div} J$ in distributions,

$$W_2(f, g) \le C \left[\left(\int_{I_0^c} \frac{\rho}{1 + |x|^2} \right)^{1/2} + \left(\int_I \frac{|U|^2/(1 + |x|^2) + |J|^2}{\rho} \right)^{1/2} \right]. \tag{60}$$

To see this, split $\int (f-g)u$ with the cutoff and use Cauchy–Schwarz with (59) in the exterior. In the interior insert the field identity and integrate by parts. Approximation by smooth tests justifies this for a C^1 cutoff. The $L^2(\rho)$ norms of $(1+|x|^2)^{1/2}\chi u$ and $\nabla(\chi u)$ are at most $C\|\nabla\varphi\|_{L^2(\rho)}$. Riesz representation produces an exact flux of divergence $f-g$ whose squared action norm $\int |J|^2/\rho$ is bounded by the square of the displayed right-hand side. Interpolate between the densities with $(1-\lambda)f+\lambda g$, $\lambda(t)=\sin^2(\pi t/2)$, and flux $\lambda'(t)J$. Its action is at most $\pi^2\int |J|^2/\rho$, by the lower bound $\min(\lambda,1-\lambda)\rho$. The dynamic transport inequality [4] gives (60).

8.2.1 The reverse transport comparison and the heat fields

Use $D(x)=1+|x|^2$ and suppose the origin is within fixed distance of the two means. Let

$$\mathcal{H}(f,g)=\sup_{\int |\nabla\varphi|^2\rho\le 1}\left|\int \varphi(f-g)\right|,\qquad \rho=f+g,$$

where the supremum is over smooth compactly supported tests. The Riesz and dynamic argument used for (60) gives $W_2(f,g)\le C\mathcal{H}(f,g)$. We prove a converse for the endpoint mixture weight. For related general comparisons using intermediate or bounded densities, see Peyre [25].

Set $\ell(x)=\epsilon/(1+|x|)$ for fixed small $\epsilon>0$. For $|x-y|\le c\min(\ell(x),\ell(y))$, the Euclidean ball-average inequality gives

$$|\varphi(x)-\varphi(y)|\le C|x-y|\big(M_{\mathrm{loc}}|\nabla\varphi|(x)+M_{\mathrm{loc}}|\nabla\varphi|(y)\big),$$

where M_{loc} takes centered ball averages with radii at most $C_0\ell$. Indeed telescope ball averages from radius $|x-y|$ to zero at both endpoints using the L^1 ball Poincaré inequality, and compare the two initial averages on a common enlarged ball. The localized maximal operator is bounded on $L^2(\rho)$. To see this, cover with fixed enlargements of a Vitali selection of pairwise disjoint balls of radii $c_0\ell(x)$. The enlarged balls have bounded overlap by radius comparability when they meet and disjointness of the original balls. Choose ϵ small for the enlargement factors. On each enlarged ball all the relevant radii are comparable and ρ is comparable to its center value by the score bound. Each local maximal function is dominated there by the usual maximal function of the input restricted to a further fixed enlargement, and the usual L^2 bound and overlap yield the assertion.

For the other pairs, both $\ell(x)$ and $\ell(y)$ are bounded by $C|x-y|$, since ℓ is Lipschitz and the smaller radius is at most $|x-y|/c$. Subtract the mixture mean of φ, writing the result as u. Then

$$|\varphi(x)-\varphi(y)|\le C|x-y|\big(\sqrt{D(x)}|u(x)|+\sqrt{D(y)}|u(y)|\big).$$

Integrate both inequalities against an optimal coupling and use Cauchy–Schwarz, the local maximal bound and (59). This proves $\mathcal{H}(f,g)\le CW_2(f,g)$. In particular there is a vector flux L with $\operatorname{div} L=f-g$ and $\|L/\sqrt{\rho}\|_2\le CW_2(f,g)$ by Riesz representation. The flux is integrable by weighted Cauchy–Schwarz.

8.3 Gaussian smoothing in the curvature class

Lemma 8.4. *For $A\succeq 0$ with $\|A\|_{\mathrm{op}}\le\Lambda$ and $f\in\mathcal{F}_\alpha$, the law $f_A=f*\gamma_A$ belongs to a class of the same form and smoothness α, with uniformly enlarged derivative bounds and uniformly positive curvature. Also $W_2(f_A,f)\le C\|A\|_{\mathrm{op}}$.*

Proof. For positive definite A, differentiation of the posterior integral gives

$$D^2[-\log f_A](x)=A^{-1}-A^{-1}\operatorname{Cov}(X\mid X+Y=x)A^{-1}.$$

The posterior of X has potential Hessian between $mI+A^{-1}$ and $MI+A^{-1}$. Brascamp–Lieb gives covariance at most $(mI+A^{-1})^{-1}$. For the converse bound, integration by parts gives $\mathbb{E}[(X-\mathbb{E}X)\nabla H(X)^T]=I$ and $\mathbb{E}[\nabla H(X)\nabla H(X)^T]=\mathbb{E}D^2H(X)\preceq MI+A^{-1}$ for the posterior potential H. The block covariance inequality then gives $\operatorname{Cov}(X)\succeq(MI+A^{-1})^{-1}$. It follows that the output Hessian is between $m(1+m\|A\|_{\mathrm{op}})^{-1}I$ and MI. For singular A pass to $A+\epsilon I$ by local convergence of the first two derivatives of the smoothed positive densities.

We show the derivative bounds directly, including for singular A. Use the probability measure ν_x on $\mathbb{R}^d$ proportional to $e^{-V(x+A^{1/2}z)}\,d\gamma_I(z)$. Its potential has Hessian between I and $(1+M\Lambda)I$; quadratic bounds around its mode z_x give uniformly bounded moments of every fixed order about z_x. Differentiating the log integral at x expresses each order-l derivative of $\log f_A(x)$ as a sum over partitions of the l derivative indices. A block of size b contributes the corresponding derivative of $-V$ evaluated at $x+A^{1/2}z$, and the blocks enter a joint cumulant under ν_x. This identity follows by differentiating the multivariable log moment generating function of the increment of $-V$. Local differentiation under the integral is justified by the Gaussian factor, the quadratic bounds on V, and polynomial bounds for derivatives of its exponential.

For $l \geq 2$ the single-block term is the expectation of a bounded derivative. In any cumulant with at least two blocks, each score factor can be centered by subtracting its value at $x + A^{1/2}z_x$, without changing the cumulant. The resulting factor is bounded by $C|z - z_x|$, while every block of size at least two is bounded. The moment formula for cumulants gives uniform bounds up to order k.

When $\theta > 0$, the single-block term at order k is uniformly θ-Hölder. Its integrand shifts by $O(|x - x'|^\theta)$, and the posterior changes by $O(|x - x'|)$ in total variation, because its derivative with respect to x is the measure times a centered score with bounded $L^1(\nu_x)$ norm. Integrate that directional derivative along the segment from x to x'. For all multi-block terms the integrands use orders at most $k-1$, so one further derivative uses only the derivatives already assumed bounded, except for score factors that can be centered. More precisely, to evaluate this further derivative at a fixed x, hold the subtracted score constants at that x fixed during differentiation. The derivatives of the shifted integrands and the centered-score derivative of the posterior then have uniformly bounded mixed moments by the same mode estimate. The finite moment formula for cumulants gives a uniform Lipschitz bound for those terms. For distances above one, the sup norm suffices for the Hölder condition. Thus all the required bounds hold in the enlarged class.

Finally let $f_t = f * \gamma_{tA}$, $0 \leq t \leq 1$. It satisfies the continuity equation with velocity $-A\nabla \log f_t/2$. Its Fisher information $\int |\nabla \log f_t|^2 f_t$ is at most dM by integration by parts and the output Hessian upper bound, which also holds at $t = 0$. The dynamic transport action is at most $dM\|A\|_{\mathrm{op}}^2/4$. This proves the displacement estimate. □

8.3.1 A reverse bound for small common heat smoothing

Let h lie between fixed negative powers of n and let $0 \preceq A_0 \preceq Ch^2 I$. Put $f_A = f * \gamma_{A_0}$, $g_A = g * \gamma_{A_0}$. Uniformly in the smoothness interval,

$$W_2(f, g) \leq C\big(W_2(f_A, g_A) + h^{\alpha+1}\big). \tag{61}$$

For the proof it suffices to restrict the mean gap to a fixed constant. If the mean gap exceeds one, uniform second moments about the means give $W_2(f, g) \leq C|m_f - m_g|$, and the mean gap is unchanged by smoothing. Otherwise center coordinates within a fixed distance of the means.

Write $q = f - g$, $Pq = q * \gamma_{A_0}$, and $\rho_A = f_A + g_A$. The reverse weighted transport comparison gives $Pq = \operatorname{div} H$ with $\|H/\sqrt{\rho_A}\|_2 \leq CW_2(f_A, g_A)$, using the uniform smoothed curvature constants. Choose an integer J with $2(J-1) > \alpha_+ - 1$. The binomial identity gives

$$q = \sum_{j=1}^{J} (-1)^{j+1} \binom{J}{j} P^{j-1}(Pq) + (I - P)^J q.$$

The first sum is the divergence of the same linear combination of $P^{j-1}H$.

On a ball I of radius $R = D_0\sqrt{\log n}$, these fields have $L^2(1/\rho)$ norm at most $C\|H/\sqrt{\rho_A}\|_2$. Indeed the Gaussian ratio bound $\rho_A(x)/\rho(x) \leq e^{C\|A_0\|(1+|x|^2)}$ follows by the upper bound from quadratic Taylor expansion as in (57), also for PSD matrices. For $x \in I$ each smoothed density has supporting log tangent and score $O(R)$, giving $\rho_A(y)/\rho(x) \leq Ce^{CR|x-y|}$. After conjugating by $\sqrt{\rho_A}$ and $1/\sqrt{\rho}$, the absolute kernel for P^{j-1} is bounded by convolution with the Gaussian measure times $Ce^{CR|x-y|}$. This is a finite measure of uniformly bounded mass since $hR = o(1)$. Young's inequality works also for degenerate Gaussian measures, and $j = 1$ is the identity.

The remainder is the divergence of

$$-\tfrac{1}{2}A_0 \int_0^1 P_{tA_0}(I - P)^{J-1} \nabla q \, dt,$$

where P_{tA_0} is convolution by γ_{tA_0}. This follows by integrating the anisotropic heat equation once. The signed measures in the integral annihilate all polynomials of degree below $2(J-1)$, by the zero at the origin of their Fourier transforms. Their absolute moments at scale h are uniformly bounded, and their far tails at distances greater than $w = (2+R)^{-1}$ have arbitrarily fast polynomial decay in n, even when a component measure is degenerate. Subtract the local Taylor polynomial of ∇f and ∇g to degree $\alpha - 2$ for integer α and degree $k-1$ for fractional $\alpha = k + \theta$. The relative derivative bounds from Section 2 give a local remainder of order $h^{\alpha-1}$ times a fixed polynomial in $|x|$ times the component density. After multiplication by A_0 and integration, its $L^2(1/\rho)$ norm on I is $O(h^{\alpha+1})$. The far tails are negligible by the polynomial density floor on the ball. Finally choose D_0 large enough that the exterior term in (60) is $O(h^{\alpha+1})$, and use that inequality with the two fluxes and zero scalar field. This proves (61).

8.4 A weighted feature at one trial

Retain n observations in each group and use equal-sized corresponding splits. Use the first splits to estimate each mean. If their difference exceeds one, reject immediately, and otherwise center all trials at the first mean estimate. The estimates are within $1/4$ of their population means with probability $1 - o(1)$, uniformly, by Gaussian tail bounds. All subsequent calculations condition on these first samples. The null mean rejection error is $o(1)$.

We describe a trial with nominal scalar noise t and bandwidth h, with $t \le \gamma h^2 \log n$ and h between the polynomial scan endpoints used below. The inputs at a trial may have been augmented with independent Gaussian variables in each split. At a correctly specified trial the two observation laws are $u * \gamma_t$, for latent laws $u = f_0, g_0$ in the possibly enlarged smoothness class. We only need the power calculation at such a trial. Under the null the two inputs are equal in law in every trial.

Let $a = \log n$, $w = a^{-1/2}$, $D(x) = 1 + |x|^2$, $b^2 = 1/a$, $v = b^2 - t > 0$, and $h(x) = he^{c_0|x|^2}$, with small fixed $c_0 > 0$. Write $E = C(1 + t/h^2)^{C_0} e^{C_0 t/h^2}$ for a deterministic upper bound on the fine kernel seminorms, where C_0 is fixed sufficiently large. For a prescribed probability exponent ℓ define

$$r_h^2 = \frac{\sqrt{\ell}\, E^2 h^{2-d/2}}{n}, \qquad \rho_h = r_h^2 a^{2-d/2}. \tag{62}$$

In the applications $1 \le \ell \le C \log\log n$; r_h is between fixed negative powers of n and $nw^d\rho_h \ge n^{c_1}$ for some $c_1 > 0$. These follow from $d > 4$, the bandwidth endpoints and sufficiently small γ, as verified below.

Use the independent pilot split to estimate both observed densities on a grid of cubes of side w meeting the ball of radius $D_0\sqrt{a}$. Divide the counts by sample size times cube volume. At any cube center with observed density at least a constant times ρ_h, the expected count is at least a constant times $nw^d\rho_h$. The cube average is comparable to its center by the local score bound. Chernoff inequalities show simultaneously over $O(a^d)$ rows that every accepted count (above a sufficiently large fixed multiple of $nw^d\rho_h$) agrees with its mean up to constants, and every count with sufficiently large mean is accepted. Apply these facts to the pooled count of the two groups, a sum of independent indicators, and use its estimate as $\widehat{\rho}$, constant on each selected cube. Let I be the union of selected cubes. On the resulting event, when the nominal model is correct,

$$\widehat{\rho} \asymp \rho = f_0 + g_0 \ \text{ on } I, \qquad \{\rho > C\rho_h\} \subseteq I,$$

where D_0 is large enough, since $ta = o(1)$ and (57) compares the observed and latent values there. The event has probability $1 - o(1)$, uniformly over desired trials under the stated conditions. A cutoff zero below $C\rho_h$ and one above $2C\rho_h$ has gradient $O(1 + |x|)$, by the score bound. Its exterior term in (60) is $O(r_h)$: integration on potential levels for each component gives weighted tail at density threshold $C'\rho_h$ at most $C\rho_h(\log(1/\rho_h))^{d/2-2}$. Indeed the radial derivative and radius at level u are comparable to $\sqrt{u}$, and the level-volume derivative is $O(u^{d/2-1})$; the factor D^{-1} contributes u^{-1}. The modes stay within a fixed distance of the origin.

For an input y, take scalar kernel $\gamma_v(x-y)$ divided by $\sqrt{D(x)\widehat{\rho}(x)}$. Write $R_u = \frac{1}{2}\int_0^u \nabla\gamma_z\, dz$. The vector kernel is the sum of $R_v - R_{h(x)^2}$ and a fine kernel of symbol

$$\frac{i\xi}{|\xi|^2}\left(\widehat{K}(h(x)\xi)e^{t|\xi|^2/2} - e^{-h(x)^2|\xi|^2/2}\right)$$

with row parameters held fixed. The fine kernel alone is truncated in space to $|x-y| \le w$. Normalize the vector sum by $\sqrt{\widehat{\rho}(x)}$. Denote the combined feature by F_y in $L^2(I) \oplus L^2(I;\mathbb{R}^d)$. If no row is selected take the zero feature.

At a correctly specified trial the scalar expectation is the normalized $u * \gamma_{b^2}$. The vector expectation without truncation has multiplier on the latent law $i\xi(\widehat{K}(h(x)\xi) - e^{-b^2|\xi|^2/2})/|\xi|^2$, so its difference from the exact $R_{b^2} * u$ is a high-pass term as in (58). The scalar and exact heat vector satisfy $u = u * \gamma_{b^2} - \operatorname{div}(R_{b^2} * u)$. The high-pass term has weighted norm $O(h^{\alpha+1})$ by (58), $u^2/\widehat{\rho} \le Cu$, and the Gaussian envelope, taking c_0 small uniformly in α. After rescaling the fine kernel by $h(x)$, its symbol is smooth at the origin: the radial bracket vanishes to second order. Its Schwartz seminorms are at most E, after factoring out $h(x)^{1-d}$ from the spatial kernel. This follows by differentiating the smooth compact Fourier cutoff and the Gaussian term. The polynomial factor in E pays the finitely many derivative orders. The truncation error is negligible by arbitrary fixed Schwartz tail order, $h(x)$ polynomially small, and the pilot density floor. We choose the tail order before γ. Consequently, if μ is the difference of feature means,

$$W_2(f_0, g_0) \le C(r_h + h^{\alpha+1} + \|\mu\|). \tag{63}$$

We record both covariance norms at the correctly specified trial on its pilot event:

$$\|C\|_{\mathrm{HS}} \le C_1 E^2 h^{2-d/2}, \qquad \|C\|_{\mathrm{op}} \le C_1. \tag{64}$$

Here C can denote the covariance of a paired feature difference, since the groups are independent. For either group with input density p, the geometric mean of the log tangent bounds at rows x, x' controls $p(y)/\sqrt{\widehat{\rho}(x)\widehat{\rho}(x')}$ by $C\exp(C\sqrt{a}(|y-x| + |y-x'|))$. The tilt is absorbed by the row Gaussians and is bounded on the fine supports. For the scalar feature, Gaussian convolution gives an absolutely integrable row and column with bounded mass, and squared Hilbert–Schmidt norm at most a power of a. These bounds absorb the scalar factors $D^{-1/2}$.

For the coarse vector the tilted profiles are bounded by $(h(x)+|z|)^{1-d}e^{-ca|z|^2}$, by integrating the gradient heat kernel over $h(x)^2 \le u \le v$. On pairs at distance at most $2w$ the bandwidths are comparable. Convolution of the two profiles gives bound

$$C(h(x) + |x-x'|)^{2-d}e^{-ca|x-x'|^2}.$$

For distances above $2w$, use instead $C|x-x'|^{2-d}e^{-ca|x-x'|^2}$. These bounds follow by splitting near each singular center and the complement, or by Cauchy–Schwarz when the separation is below $h(x)$. The unregularized bound is integrable in both rows and columns, giving bounded operator norm by Schur's inequality. The squared Hilbert–Schmidt norm is at most $C\int_I h(x)^{4-d}\,dx$ plus a power of a; the adaptive bandwidth makes this at most Ch^{4-d}.

For the fine vector, the uncentered second-moment kernel vanishes for $|x-x'| > 2w$; on the remaining pairs the rescaled Schwartz profiles give bound $CE^2h(x)^{2-d}(1 + |x-x'|/h(x))^{-P}$ for large fixed P. The squared norm is $O(E^4h^{4-d})$ by integrating first in the difference variable, and the row and column integrals are bounded by $CE^2\sup_I h(x)^2 = O(1)$. For this last bound choose γ small after c_0, using $t/h^2 \le \gamma a$, and the uniform polynomial upper bound on $h(x)$. Finally the centered covariance is dominated in positive operator order by three times the sum of the three uncentered operators, by Cauchy–Schwarz. The operator and Hilbert–Schmidt bounds therefore give (64).

8.5 The conditional sign test

Reserve two equal-sized paired blocks of order n, independent of the two earlier splits. Each paired observation produces $Z = F_{Y'} - F_{Y''}$. Use the first paired block to form its mean $\widehat{\mu}$. In the other block of size m_v, put $X_i = \langle \widehat{\mu}, Z_i \rangle$. The trial rejects if

$$\sum_{i=1}^{m_v} X_i > \sqrt{2\ell \sum_{i=1}^{m_v} X_i^2}. \tag{65}$$

Under the null, conditional on the earlier splits, trial choices and training block, each paired difference is symmetric, and the validation observations are independent. Conditional additionally on the magnitudes of the X_i, the nonzero signs are independent and fair. The bound $\cosh(u) \le e^{u^2/2}$ for the exponential moment

of the signed sum gives conditional rejection probability at most $e^{-\ell}$. If all magnitudes vanish the strict inequality prevents rejection. This level bound uses neither correct nominal noise nor a successful pilot. In a randomized trial with Gaussian additions, the conditioning is on the earlier blocks and any parameters chosen there, with the independent augmentation of the validation blocks still included in the experiment.

For the power bound at a correct trial let $u = \|\mu\|$, and write C for the paired covariance. With $T = m_v^{-1}\sum_i X_i$ and $W = m_v^{-2}\sum_i X_i^2$, the two independent blocks give

$$\mathbb{E}T = u^2, \qquad \mathrm{Var}(T) \le C_2\big(\|C\|_{\mathrm{op}}u^2/n + \|C\|_{\mathrm{HS}}^2/n^2\big).$$

Indeed the variance of the conditional mean $\langle\widehat{\mu},\mu\rangle$ is $\langle\mu, C\mu\rangle/m_t$, where $m_t \asymp n$ is the training size, and the conditional variance averages to $(\langle\mu, C\mu\rangle + \mathrm{tr}(C^2)/m_t)/m_v$. The second moment of a validation projection gives

$$\mathbb{E}W \le C_2\big(u^4/n + \|C\|_{\mathrm{op}}u^2/n + \|C\|_{\mathrm{HS}}^2/n^2\big).$$

Combining (64) and (62), $\ell\|C\|_{\mathrm{HS}}^2/n^2 \le C_3 r_h^4$. Also $\ell/n = o(r_h^2)$ uniformly, by the polynomial upper bound on h and $d > 4$. For $u \ge Kr_h$, with K a sufficiently large fixed constant, Chebyshev gives $T > u^2/2$ and Markov gives $\sqrt{2\ell W} \le u^2/2$, with probability as close to one as required. This proves the trial power statement. The conclusion holds on the pilot event at the correct trial, with its $o(1)$ exceptional probability added.

8.6 Choosing the trials

Put $p_- = \alpha_- + d/4$. Take bandwidths $h_j = 2^j N^{-1/(2p_-)}$ from $j = 0$ up to the largest with $h_j \le n^{-\eta/4}$, for fixed small $\eta > 0$. Use only positive-nominal-noise trials with $t/h_j^2 \le \gamma\log n$; zero-noise trials also obey this inequality. The scan has $O(\log n)$ bandwidths. Since $d > 4$, the largest h implies $nr_h^2 \ge h^{2-d/2} \ge n^{c_2}$ uniformly in trials. At the smallest h, $h^{2-d/2}/n$ is a fixed negative power of n up to $\log\log n$ factors. Taking γ small makes $E^2 \le Cn^{c_3}$ with c_3 less than half that exponent. Thus (62) has the claimed polynomial bounds throughout the scan. Take D_0 large for the pilot inclusion and c_0 small for the moment and bandwidth bounds, before making γ small enough for the operator estimate.

8.6.1 Known covariance

Complete the common known covariance to sI by independent Gaussian additions. Each trial has $t = s$. Let j_s be the grid index closest to $\sqrt{s}$ on the log-bandwidth scale, clipped at the endpoints, with an arbitrary endpoint at $s = 0$. Set $e^{-\ell_j} = c_\varepsilon(1 + |j - j_s|)^{-2}$ with a small fixed c_ε, so $\ell_j \ge 1$ and the levels sum to a sufficiently small constant. Always $\ell_j = O(L)$.

For $s \le n^{-\eta}$, choose a bandwidth within a fixed factor of a sufficiently large fixed multiple of

$$h_* = \max\{N^{-1/(2p)}, \sqrt{s/\log(2 + ns^p)}\}.$$

Round up in the scan. This is feasible because h_* is below the cap by a fixed polynomial factor for small η; the large multiplier ensures $s/h^2 \le \gamma\log n$. It suffices to show $r_h \le Ch^{\alpha+1}$, equivalently $\sqrt{\ell_j}E^2 \le Cnh^{2p}$. If $h \ge \sqrt{s}$, the amplification is bounded and $nh^{2p} \ge \sqrt{L}$. Otherwise put $X = ns^p$, $B = \log(2 + X)$. The index distance from $\sqrt{s}$ is $O(1 + \log(2 + B))$ since $s/h^2 \le CB$. A sufficiently large bandwidth multiplier gives $E^2 \le C(2 + X)^c$, uniformly for some $c < 1$. For large X, $nh^{2p} \ge c'X/B^p$ pays this factor and the square root of ℓ_j, whose growth in X is at most a fixed logarithmic power. For bounded X the fine factor and ℓ_j are bounded. This proves the needed inequality. The signal comparison and the sign test yield power at the known-covariance rate.

8.6.2 Unknown covariance

For each bandwidth take a deterministic operator-norm $h^2/2$-net of PSD covariance matrices in the ball $\|\Xi\|_{\mathrm{op}} \le \gamma h^2\log n$. In fixed matrix dimension a grid and choice of one PSD representative per intersecting cell gives $(\log n)^{O(d^2)}$ trials over all bandwidths. Write M_n for the total count. Use the same exponent $\ell = \log(C_\varepsilon M_n)$ for each trial, with C_ε large enough. It is $O(L)$, so the conditional sign inequality and a union bound control the null, irrespective of the true covariance.

At a trial add independent $N(0, h^2 I)$ noise to each present observation and then independent $N(0, tI - \Xi)$ noise, where $t = \|\Xi\|_{\mathrm{op}}$. Use the trial feature with nominal scalar noise t. If $\|\Xi - \Sigma\|_{\mathrm{op}} \le h^2/2$, the latent laws in this nominal model are f_A, g_A with common residual covariance $A_0 = \Sigma - \Xi + h^2 I$ between 0 and $2h^2 I$. They are in the uniformly enlarged class by Lemma 8.4. Equality under the null holds at every trial, with or without this approximation. All random additions are independent of the draws in each validation block and of its training block.

For $s \le n^{-\eta}$ select a dyadic bandwidth of order a sufficiently large fixed multiple of

$$h_* = \max\{N^{-1/(2p)}, \sqrt{s/\log(2 + Ns^p)}\}.$$

It lies in the scan range as before. By the bandwidth multiplier the noise norm is below the net radius, and a nearby trial Ξ exists. Write $X = Ns^p$ and $B = \log(2 + X)$. At this trial $t/h^2 \le s/h^2 + 1/2$, so a large enough bandwidth multiplier gives $E^2 \le C(2 + X)^c$ for $c < 1$. Also

$$Nh^{2p} \ge c' \max\{1, X/B^p\}.$$

Hence $r_h^2 \le Ch^{2\alpha+2}$ by $\ell = O(L)$. The true separation satisfies $W_2(f, g) \le C(W_2(f_A, g_A) + h^{\alpha+1})$ by (61). Thus at a sufficiently large multiple of the claimed rate, (63) and (65) give the power bound at this one correct trial. The complete randomized procedure defines a test on the original observations by averaging its rejection probability over the independent Gaussian additions.

8.6.3 Large noise

Both procedures also reject if the empirical observed W_1 distance exceeds $2n^{-\beta/2}$, where $\beta = 1/[4(d + 1)]$. Each empirical law has expected distance $O(n^{-\beta})$ from its observed population law, by truncation at radius n^{β} about its mean, quantization at mesh $n^{-\beta}$, and the bound on expected histogram total variation by the square root of the number of sites divided by n. Thus the additional null error is $o(1)$, uniformly.

For power suppose the actual noise satisfies $s \ge n^{-\eta}$ and the observed population distance is $\epsilon \le 5n^{-\beta/2}$. The means are then within one. The constant-bandwidth fields in Section 2 with h small and a ball of radius $D_1\sqrt{\log(1/h)}$ give

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

Indeed the Gaussian exterior and normalized high-pass field are $O(h^{\alpha+1})$ by (58), the interior has a density floor h^P, and the observed characteristic functions differ by at most $\epsilon|\xi|$. Inverting the common Gaussian on the support $|\xi| \le h^{-1}$ gives the scalar sup bound $C\epsilon h^{-d-1} e^{s/(2h^2)}$. Take $h = \sqrt{s/(\gamma_0 \log n)}$ with γ_0 small and then choose η small uniformly for the smoothness interval. The displayed distance is $O(h^{\alpha+1})$. The two effective logarithms in the inverse rates are comparable to $\log n$ in this range, and the inverse term dominates the clean scale for small η. Therefore an alternative separated by a large enough multiple of the rate has observed population discrepancy above $5n^{-\beta/2}$. When each empirical error is at most $n^{-\beta/2}$ the empirical rule rejects, and that event has probability $1 - o(1)$. This proves Theorem 8.1.

8.7 Mixtures over smoothness and noise levels

Suppose the observed reference is p_0 in both groups. Our alternatives vary only the smaller group; its observed law at level j has density $p_0 + \sum_i \omega_i \widetilde{q}_{ji}$ with independent uniform signs ω_i. Every perturbation integrates to zero. For levels j, k let G_{jk} be the rectangular matrix

$$G_{jk}(i, i') = \int \frac{\widetilde{q}_{ji}\widetilde{q}_{ki'}}{p_0}\, dx.$$

Let L_j be the sign-mixture likelihood ratio of the whole sample at level j. Then $\mathbb{E}_0 L_j L_k = \mathbb{E}_{\omega,\omega'}(1 + \omega^T G_{jk}\omega')^n$. The inner likelihood overlap is nonnegative. Bound its nth power by the exponential, integrate one sign vector using $\cosh(u) \le e^{u^2/2}$, introduce a standard Gaussian vector to linearize the square, and integrate the other signs by the same inequality. This gives

$$\mathbb{E}_0 L_j L_k \le \det(I - n^2 G_{jk} G_{jk}^T)^{-1/2} \le \exp(Cn^2 \|G_{jk}\|_{\mathrm{HS}}^2)$$

whenever $n\|G_{jk}\|_{\mathrm{op}} \le 1/2$. An extra unchanged sample cancels in this calculation. Thus a uniform mixture over J_n levels has total variation $o(1)$ when the diagonal second moments are at most J_n^c with $c < 1$ and the off-diagonal overlaps are bounded by $1 + o(1)$ uniformly.

We will use the following separation bound for any smooth φ, two latent laws u, g, and any optimal coupling:

$$\int \varphi(u - g) \le \|\nabla\varphi\|_{L^2(u)} W_2(u, g) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(u, g)^2. \tag{66}$$

This follows by Taylor expansion along the coupling and Cauchy–Schwarz. A pairing at least cE_0, gradient norm $O(\sqrt{E_0})$ and bounded Hessian imply $W_2(u, g) \ge c'\sqrt{E_0}$.

8.7.1 The clean adaptation lower bound

Use a centered isotropic Gaussian g of strictly interior precision. Let Φ be its coordinatewise probability integral transform, mapping $g(x)\,dx$ to Lebesgue measure on the unit cube. On any fixed interior cube the inverse transform and all fixed derivative bounds are uniform. Put $p_0 = g$ for the clean experiment. At dyadic levels of side h_j, choose compactly supported orthonormal wavelets ψ_{ji} in a fixed interior cube, of regularity exceeding $\alpha_+ + 1$, using one orientation [7]. There are of order h_j^{-d} wavelets per level. Choose the $J_n \asymp \log n$ dyadic levels for which

$$n^2 h_j^{4\alpha_j + d} = L$$

defines α_j in any fixed positive-width interior subinterval. Set

$$q_{ji} = a_j g\,(\psi_{ji} \circ \Phi), \qquad a_j = \kappa h_j^{\alpha_j + d/2}.$$

Their integral vanishes. Within a level, the relative signed sum has derivative bounds $C\kappa h_j^{\alpha_j - r}$ at each integer order $r \le \lfloor \alpha_j \rfloor + 1$, by wavelet scaling, bounded overlap and the transform. Interpolating the derivative increment bound with twice the sup norm gives Hölder seminorm $C\kappa$ at the top fractional order, uniformly across the interval. The logarithmic perturbation respects the curvature and jet margins by small κ; the latent densities are positive and comparable to g.

Take $\varphi = h_j^2 a_j \sum_i \omega_i\,(\psi_{ji} \circ \Phi)$. The pairing with $u - g$ is exactly $E_0 = h_j^2 a_j^2 \#\{i\} \asymp h_j^{2\alpha_j + 2}$, up to the fixed amplitude constant. The gradient norm against u is $O(\sqrt{E_0})$ by bounded transform derivatives, normalized wavelet scaling, and finite overlap. The Hessian is bounded since $a_j h_j^{-d/2} = O(h_j^{\alpha_j})$. Thus (66) proves the separation $cA(\alpha_j)$. The clean Gram matrices are $a_j^2 I$ within a level and zero between levels, by exact orthogonality after the transform. We have $n a_j^2 = o(1)$ uniformly, and $n^2 a_j^4 \#\{i\} \le C\kappa^4 L$. The interval contains $J_n \asymp \log n$ scales, so the mixture comparison proves the clean lower bound by choosing κ small. A fixed common Gaussian convolution is a Markov postprocessing of the whole clean lower-bound experiment and cannot increase total variation.

8.7.2 A shared observed null at different noise levels

Fix Y_n in the range of Theorem 8.2. Choose α_j in the given interior subinterval and let $p_j = \alpha_j + d/4$ and $s_j = (Y_n/N)^{1/p_j}$. As α_j varies, the interval of possible $\log s_j$ has length of order $\log n$ because $Y_n \le N^{1-\delta}$. Choose them equally spaced there with gaps $D_2 \log L$, where $D_2 > 4/d$. There are $J_n \asymp \log n / \log L$ levels with $\log J_n \sim L$, and all the s_j are polynomially small uniformly. Let the shared observation covariance at level j be $s_j I$ and choose an interior Gaussian variance v_0. Use the latent reference $g_j = N(0, (v_0 - s_j)I)$ in both groups at the null. The observed reference in both groups is exactly $p_0 = N(0, v_0 I)$ at every level.

For alternatives at level j hold the other group at g_j and place packet centers z_i in a fixed central cube with spacing $D_3\sqrt{s_j}$, for large fixed D_3. Put $B = \log(2 + Y_n)$, $T_j^2 = A_0 B / s_j$, and $a_j = \kappa T_j^{-\alpha_j}$. The signed perturbations are $q_{ji} = a_j \psi_{ji}$, where

$$\psi_{ji}(x) = e^{-|x - z_i|^2/(2s_j)} \left[\cos(T_j(x_1 - z_{i,1})) - e^{-s_j T_j^2/2}\right].$$

Each integrates to zero. Uniformly in the level, the absolute sum of order-r derivatives after division by g_j, before the factor a_j, is at most CT_j^r. Indeed $s_j^{-1/2} \lesssim T_j$, derivatives of the Gaussian envelope and

its reciprocal reference factors leave a weakened packet envelope, and Gaussian overlap over the separated centers is bounded. For a fractional smoothness the difference of the order-k derivative of the signed sum is bounded by $C\min(T_j^k, T_j^{k+1}|x-y|)$. Its Hölder bound of order θ is at most $CT_j^{k+\theta}$. With small κ the relative perturbation is uniformly small, and taking its logarithm preserves the required Hessian and jet bounds with the interior margin of g_j.

For separation take $\varphi = T_j^{-2} a_j \sum_i \omega_i \psi_{ji}$. The packet L^2 Gram matrix has diagonal of order $s_j^{d/2}$, by integrating the squared cosine against the envelope for $T_j\sqrt{s_j}$ sufficiently large. Its off-diagonal absolute row sums are at most $Cs_j^{d/2}\sum_{k\in\mathbb{Z}^d\setminus\{0\}} e^{-cD_3^2|k|^2}$, less than half the lower diagonal bound. The gradient Gram row sums are $O(T_j^2 s_j^{d/2})$ by the same overlap bounds. It follows that the pairing is at least cE_0 and the squared gradient norm is $O(E_0)$, where $E_0 = T_j^{-2} a_j^2 s_j^{d/2} \#\{i\}$. The Hessian is $O(a_j)$ by the pointwise overlap estimate. Equation (66) gives separation $cT_j^{-\alpha_j-1}$. This is of order $I_N(\alpha_j, s_j)$. Also $I_N(\alpha_j, s_j) \gtrsim A(\alpha_j)$ since $Y^{1/p_j}/\log(2+Y)$ is bounded below uniformly for $Y \geq 1$.

Gaussian convolution changes a packet's envelope variance from s_j to $2s_j$ and its oscillatory amplitude from one to $e^{-s_j T_j^2/4}$. This is the conditional Gaussian formula at the center: the conditional mean of the coordinate is half its observed shift and the variance is $s_j/2$. The subtracted constant has the smaller factor $e^{-s_j T_j^2/2}$. Thus an observed perturbation is bounded by $Cb_j e^{-c|x-z_i|^2/s_j}$, where $b_j = a_j e^{-cs_j T_j^2}$ with a fixed $c > 0$. The denominator p_0^{-1} in any Gram entry is absorbed by weakening the two Gaussian exponents, since the centers are bounded and $\max_j s_j = o(1)$.

For the same level, integration and lattice summation give

$$n\|G_{jj}\|_{\mathrm{HS}} \leq CD_j, \qquad n\|G_{jj}\|_{\mathrm{op}} \leq CD_j s_j^{d/4}, \qquad D_j = nb_j^2 s_j^{d/4}.$$

The operator bound uses absolute row sums, whose size before multiplying by n is $O(b_j^2 s_j^{d/2})$. Since $Ns_j^{p_j} = Y_n$,

$$D_j \leq C\kappa^2\sqrt{L}\, Y_n (A_0 B)^{-\alpha_j} (2+Y_n)^{-c'A_0} \leq C'\kappa^2\sqrt{L}$$

for A_0 large. Thus the operator condition is eventually satisfied, using the uniform polynomial smallness of s_j.

For different levels the Gaussian product integral bounds each cross entry by

$$Cb_j b_k \min(s_j, s_k)^{d/2} \exp\bigl(-c|z_i - z_{i'}|^2/\max(s_j, s_k)\bigr).$$

When summing squared entries, each fine center sees a bounded Gaussian lattice sum over coarse centers. The number of fine centers is at most $C\min(s_j, s_k)^{-d/2}$. Hence

$$n^2\|G_{jk}\|_{\mathrm{HS}}^2 \leq CD_jD_k\left(\frac{\min(s_j,s_k)}{\max(s_j,s_k)}\right)^{d/4} \leq C'\kappa^4 L^{\,1-D_2 d/4} = o(1).$$

For these cross terms the operator condition follows from the Hilbert–Schmidt bound. The uniform cross overlaps are at most $1+o(1)$, and the diagonal overlaps are at most $e^{C\kappa^4 L} \leq J_n^c$ with $c < 1$ for small κ. The uniform mixture is therefore indistinguishable from the common observed null. This proves the moving-level lower bound.

8.7.3 The inverse lower bound with supplied covariance

At any specified α and nonzero covariance, the ordinary inverse term $I_n(\alpha, s)$ is necessary. For small s, use the same single-scale packet construction at width $\sqrt{s}$, with an interior fixed Gaussian reference, frequency $T^2 = A_0\log(2+ns^p)/s$, and the oscillatory coordinate a top eigenvector of the true covariance. The convolution multiplies the oscillation by $e^{-sT^2/4}$ in that coordinate and leaves Gaussian envelope variances between s and $2s$ in the other coordinates, including noiseless ones. Therefore the same within-level Hilbert–Schmidt bound holds with $na^2 s^{d/4} e^{-csT^2} \leq C\kappa^2$, since $a = \kappa T^{-\alpha}$ and $p = \alpha + d/4$. The operator bound also follows from the row sums. The sign mixture can be made arbitrarily close in total variation, and (66) gives distance of order $I_n(\alpha, s)$. For s bounded below use one such packet with a fixed narrow Gaussian envelope

variance and $T^2 = A_0 \log n/s$. Gaussian convolution still attenuates it by at least e^{-csT^2} in the top direction. Its divided squared norm is finite uniformly because the envelope is narrower than the Gaussian reference. The n-sample divergence is small for large A_0, while (66) yields $cT^{-\alpha-1}$. This gives the same order since $\log(2 + ns^p) \asymp \log n$.

8.7.4 The extra inverse loss with both parameters unknown

Take $Y_n = (\log L)^P$ for a fixed $P > \alpha_+ + d/4$. On the alternative strata,

$$\log(2 + ns^p) = \log(2 + \sqrt{L}\, Y_n) \asymp \log L, \qquad \log(2 + Ns^p) \asymp \log\log L.$$

The known-covariance inverse term dominates the adaptive direct scale there, because the ratio of squared bandwidths $Y_n^{1/p}/\log(2 + \sqrt{L}Y_n)$ tends to infinity uniformly. Thus the ratio of the jointly adaptive scale to the scale with known covariance has order $(\log L/\log\log L)^{(\alpha+1)/2}$ along those strata. The lower bound and the known-covariance upper bound give Corollary 8.3. When the true noise is in the fixed-positive or larger-noise regime of Section 6, the two logarithms are comparable and this extra loss does not arise.

9 Adaptation without loss in low-dimensional Gaussian Wasserstein testing

In dimensions one through four we give a single latent quadratic Wasserstein test that adapts simultaneously to unknown smoothness and unknown shared Gaussian covariance. It attains, at every actual smoothness and covariance, the separation rate that would be available if both parameters were supplied. One statistic tests common heat fields of the observed laws. Only the lower-variance corrections for candidate inverse kernels are scanned. This separates covariance and resolution selection from the logarithmic tail penalty in the direct problem.

9.1 Experiment and separation

Fix $1 \le d \le 4$ and $2 \le \alpha_- < \alpha_+ < \infty$. For real $\alpha = k + \theta \in [\alpha_-, \alpha_+]$, k integer, $0 \le \theta < 1$, use the translated class $\mathcal{F}_\alpha$ of densities $f = e^{-V}$ on $\mathbb{R}^d$ with $V \in C^k$, $mI \preceq D^2V \preceq MI$ for fixed $0 < m < M$, and fixed positive sup norm bounds on derivatives of orders $3, \dots, k$. If $\theta > 0$, impose a fixed positive global θ-Hölder bound on D^kV. The constants are common over the interval. We observe independent samples of sizes n_1, n_2 from $p_f = f * \gamma_\Sigma$ and $p_g = g * \gamma_\Sigma$, with the same unknown covariance $\Sigma \succeq 0$. Write $n = \min(n_1, n_2)$, $s = \|\Sigma\|_{\mathrm{op}} \le \Lambda$ for fixed Λ, $\gamma_\Sigma = N(0, \Sigma)$, and $\gamma_t = N(0, tI_d)$ for scalar t. Define

$$a_{n,d} = \begin{cases} n^{-1/2}, & d = 1, \\ (\log\log n)^{1/4}/\sqrt{n}, & d = 2, \\ (\log n)^{1/4}/\sqrt{n}, & d = 3, \\ (\log n)^{3/4}/\sqrt{n}, & d = 4. \end{cases} \qquad B_{n,d}(s) = \begin{cases} \log n, & d < 4, \\ \log(2 + ns^{\alpha+1}), & d = 4. \end{cases}$$

The null is $f = g$, uniformly over the union of the latent classes and all allowed covariances. For power we measure separation at the actual covariance of the alternative.

Theorem 9.1. *For any fixed $0 < \varepsilon < 1/2$, one test that uses neither α, s, nor Σ has worst type-I error at most $\varepsilon/2$ and worst type-II error at most $\varepsilon/2$ on alternatives $f, g \in \mathcal{F}_\alpha$ with $W_2(f, g) \ge C_\varepsilon r_{n,d}(s)$, where*

$$r_{n,d}(s) = a_{n,d} + (s/B_{n,d}(s))^{(\alpha+1)/2}.$$

This holds uniformly as $n \to \infty$ for every sample-size ratio. For every admissible covariance at each sample size, the sum of errors at $c_\varepsilon r_{n,d}(s)$ is at least $1 - \varepsilon$ even if both parameters are revealed, with constants uniform in the matrix and smoothness.

The fixed known Gaussian inverse problem in this curvature setting appears in [12, Theorem 2.4]. Indirect fixed-null testing with adaptive smoothness is studied by Butucea, Matias, and Pouet [26, 24], under different noise and loss assumptions. Here both densities and the shared Gaussian covariance are unknown. Near the direct scale the correction covariance is small enough to pay for scanning even in dimension four. Away from that scale the inverse radius pays the increased variance. A split inner product for kernel tests appears in [28], and paired sign calibration of direct flow witnesses in [30]. Here the sign tests act on weighted heat fields of contaminated laws.

9.2 Density bounds and a weighted transport comparison

Uniformly over the class, f has Gaussian upper and lower envelopes about its mean m_f, the score of f is $O(1 + |x - m_f|)$, and

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

Indeed quadratic bounds at the mode give the envelopes and bound the distance between the mode and the mean. Differentiating e^{-V} gives the derivative bounds. For $\theta > 0$, on $|x - m_f| \le L$ and $|u| \le (2 + L)^{-1}$, the same product rule and the Hölder bound on $D^k V$ give

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

The local density ratio is bounded by the score, derivatives of lower order in the product are Lipschitz, and the highest one uses the assumed Hölder seminorm.

For $p_f = f * \gamma_s$, isotropic $0 \le s \le \Lambda$, Gaussian integration of the two quadratic Taylor bounds gives

$$c f(x) \le p_f(x) \le f(x) e^{Cs(1+|x-m_f|^2)}. \tag{67}$$

The potential of p_f also has Hessian in a fixed positive interval. For $s > 0$, its Hessian is $s^{-1} I - s^{-2} \operatorname{Cov}(X \mid Y)$. The conditional potential has Hessian between $(m + s^{-1})I$ and $(M + s^{-1})I$. The upper covariance bound follows from Brascamp–Lieb [5]. The lower bound follows, in each unit direction v, from Cauchy–Schwarz and the conditional identities

$$\mathbb{E}[(v \cdot (X - \mathbb{E}X))(v \cdot \nabla A(X))] = 1, \qquad \mathbb{E}[(v \cdot \nabla A(X))^2] = \mathbb{E}[v^T D^2 A(X) v],$$

where A is the conditional potential. Consequently the scores of p_f are $O(1 + |x - m_f|)$. The observed law always has a uniformly Gaussian upper envelope about m_f.

Use $\widehat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\, dx$. Let $\widehat{K}$ be a real radial smooth Fourier cutoff that equals one near zero and is supported in the unit ball. Use $K_b = b^{-d} K(\cdot/b)$ for its spatial kernel, and define constant-bandwidth fields

$$U_b(f) = K_b * f, \qquad \widehat{J_b(f)}(\xi) = -\frac{i\xi}{|\xi|^2}(1 - \widehat{K}(b\xi))\widehat{f}(\xi).$$

Then $f = U_b(f) + \operatorname{div} J_b(f)$. The inverse transform Q of $-i\xi(1 - \widehat{K}(\xi))/|\xi|^2$ is integrable with all absolute moments finite and all signed moments zero. In fact a dyadic partition into shells 2^j away from the origin gives inverse transforms bounded by $C_N 2^{(d-1)j}(1 + 2^j|x|)^{-N}$. Summability of their L^1 norms $O(2^{-j})$ gives the assertion, with moment cancellation since the symbol vanishes near zero. For $|x - m_f| \le L$, $w = (2+L)^{-1}$, and $h \le w$, subtract the local Taylor polynomial of f to degree $k - 1$ in the integer case and k otherwise. This gives

$$|J_h(f)(x)| \le C h^{\alpha+1}(1 + |x - m_f|^{k+1}) f(x) + C_N h (h/w)^N. \tag{68}$$

Indeed $J_h(f) = h Q_h * f$, the short-shift remainder has order α, and the farther shifts use uniform absolute derivative bounds and the arbitrary integrated tails.

Choose the origin within a fixed distance of the means of f, g, and put $\rho = f + g$. For a test φ, set $u = \varphi - \frac{1}{2}\int \varphi\rho$. Then

$$\int (1 + |x|^2)|u|^2 \rho \le C \int |\nabla \varphi|^2 \rho. \tag{69}$$

Both components satisfy ordinary Poincaré by Brascamp–Lieb. Comparing their component means to a common Lebesgue average on a fixed ball with a positive density floor gives the mixture inequality. For either potential V, $x \cdot \nabla V(x) \geq m|x|^2/2 - C$, since its mode stays close to zero. Integration by parts gives

$$\int u^2 x \cdot \nabla V f = d \int u^2 f + 2 \int ux \cdot \nabla \varphi f.$$

Use Cauchy–Schwarz to absorb half of the resulting quadratic term, add the two estimates, and apply mixture Poincaré. Gaussian tails justify integration by parts for compactly supported tests minus their mean.

Let $I_0 \subset I$ admit a C^1 cutoff supported in I, equal to one on I_0, with $|\nabla \chi| \leq C(1+|x|)$. For any identity $f - g = U + \operatorname{div} J$ in distributions,

$$W_2(f,g) \leq C \left[\left(\int_{I_0^c} \frac{\rho}{1+|x|^2} \right)^{1/2} + \left(\int_I \frac{|U|^2/(1+|x|^2) + |J|^2}{\rho} \right)^{1/2} \right]. \tag{70}$$

To see this, split $\int (f-g)u$ with the cutoff and use Cauchy–Schwarz with (69) in the exterior. In the interior insert the field identity and integrate by parts. Approximation by smooth tests justifies this for a C^1 cutoff. The $L^2(\rho)$ norms of $(1+|x|^2)^{1/2}\chi u$ and $\nabla(\chi u)$ are at most $C\|\nabla\varphi\|_{L^2(\rho)}$. Riesz representation produces an exact flux of divergence $f-g$ whose squared action norm $\int |J|^2/\rho$ is bounded by the square of the displayed right-hand side. Interpolate between the densities with $(1-\lambda)f + \lambda g$, $\lambda(t) = \sin^2(\pi t/2)$, and flux $\lambda'(t)J$. Its action is at most $\pi^2 \int |J|^2/\rho$, by the lower bound $\min(\lambda, 1-\lambda)\rho$. The dynamic transport inequality [4] gives (70).

9.2.1 The reverse transport comparison and the heat fields

Use $D(x) = 1 + |x|^2$ and suppose the origin is within fixed distance of the two means. Let

$$\mathcal{H}(f,g) = \sup_{\int |\nabla\varphi|^2 \rho \leq 1} \left| \int \varphi (f-g) \right|, \qquad \rho = f + g,$$

where the supremum is over smooth compactly supported tests. The Riesz and dynamic argument used for (70) gives $W_2(f,g) \leq C\mathcal{H}(f,g)$. We prove a converse for the endpoint mixture weight. For related general comparisons using intermediate or bounded densities, see Peyre [25].

Set $\ell(x) = \epsilon/(1+|x|)$ for fixed small $\epsilon > 0$. For $|x-y| \leq c\min(\ell(x), \ell(y))$, the Euclidean ball-average inequality gives

$$|\varphi(x) - \varphi(y)| \leq C|x-y|\big(M_{\mathrm{loc}}|\nabla\varphi|(x) + M_{\mathrm{loc}}|\nabla\varphi|(y)\big),$$

where M_{loc} takes centered ball averages with radii at most $C_0\ell$. Indeed telescope ball averages from radius $|x-y|$ to zero at both endpoints using the L^1 ball Poincaré inequality, and compare the two initial averages on a common enlarged ball. The localized maximal operator is bounded on $L^2(\rho)$. To see this, cover with fixed enlargements of a Vitali selection of pairwise disjoint balls of radii $c_0\ell(x)$. The enlarged balls have bounded overlap by radius comparability when they meet and disjointness of the original balls. Choose ϵ small for the enlargement factors. On each enlarged ball all the relevant radii are comparable and ρ is comparable to its center value by the score bound. Each local maximal function is dominated there by the usual maximal function of the input restricted to a further fixed enlargement, and the usual L^2 bound and overlap yield the assertion.

For the other pairs, both $\ell(x)$ and $\ell(y)$ are bounded by $C|x-y|$, since ℓ is Lipschitz and the smaller radius is at most $|x-y|/c$. Subtract the mixture mean of φ, writing the result as u. Then

$$|\varphi(x) - \varphi(y)| \leq C|x-y|\big(\sqrt{D(x)}|u(x)| + \sqrt{D(y)}|u(y)|\big).$$

Integrate both inequalities against an optimal coupling and use Cauchy–Schwarz, the local maximal bound and (69). This proves $\mathcal{H}(f,g) \leq CW_2(f,g)$. In particular there is a vector flux L with $\operatorname{div} L = f - g$ and $\|L/\sqrt{\rho}\|_2 \leq CW_2(f,g)$ by Riesz representation. The flux is integrable by weighted Cauchy–Schwarz.

9.3 Gaussian smoothing in the curvature class

Lemma 9.2. *For $A \succeq 0$ with $\|A\|_{\mathrm{op}} \le \Lambda$ and $f \in \mathcal{F}_\alpha$, the law $f_A = f * \gamma_A$ belongs to a class of the same form and smoothness α, with uniformly enlarged derivative bounds and uniformly positive curvature. Also $W_2(f_A, f) \le C\|A\|_{\mathrm{op}}$.*

Proof. For positive definite A, differentiation of the posterior integral gives

$$D^2[-\log f_A](x) = A^{-1} - A^{-1}\operatorname{Cov}(X \mid X + Y = x)A^{-1}.$$

The posterior of X has potential Hessian between $mI + A^{-1}$ and $MI + A^{-1}$. Brascamp–Lieb gives covariance at most $(mI + A^{-1})^{-1}$. For the converse bound, integration by parts gives $\mathbb{E}[(X - \mathbb{E}X)\nabla H(X)^T] = I$ and $\mathbb{E}[\nabla H(X)\nabla H(X)^T] = \mathbb{E}D^2H(X) \preceq MI + A^{-1}$ for the posterior potential H. The block covariance inequality then gives $\operatorname{Cov}(X) \succeq (MI + A^{-1})^{-1}$. It follows that the output Hessian is between $m(1 + m\|A\|_{\mathrm{op}})^{-1}I$ and MI. For singular A pass to $A + \epsilon I$ by local convergence of the first two derivatives of the smoothed positive densities.

We show the derivative bounds directly, including for singular A. Use the probability measure ν_x on $\mathbb{R}^d$ proportional to $e^{-V(x+A^{1/2}z)}\, d\gamma_I(z)$. Its potential has Hessian between I and $(1 + M\Lambda)I$; quadratic bounds around its mode z_x give uniformly bounded moments of every fixed order about z_x. Differentiating the log integral at x expresses each order-l derivative of $\log f_A(x)$ as a sum over partitions of the l derivative indices. A block of size b contributes the corresponding derivative of $-V$ evaluated at $x + A^{1/2}z$, and the blocks enter a joint cumulant under ν_x. This identity follows by differentiating the multivariable log moment generating function of the increment of $-V$. Local differentiation under the integral is justified by the Gaussian factor, the quadratic bounds on V, and polynomial bounds for derivatives of its exponential.

For $l \ge 2$ the single-block term is the expectation of a bounded derivative. In any cumulant with at least two blocks, each score factor can be centered by subtracting its value at $x + A^{1/2}z_x$, without changing the cumulant. The resulting factor is bounded by $C|z - z_x|$, while every block of size at least two is bounded. The moment formula for cumulants gives uniform bounds up to order k.

When $\theta > 0$, the single-block term at order k is uniformly θ-Hölder. Its integrand shifts by $O(|x - x'|^\theta)$, and the posterior changes by $O(|x - x'|)$ in total variation, because its derivative with respect to x is the measure times a centered score with bounded $L^1(\nu_x)$ norm. Integrate that directional derivative along the segment from x to x'. For all multi-block terms the integrands use orders at most $k - 1$, so one further derivative uses only the derivatives already assumed bounded, except for score factors that can be centered. More precisely, to evaluate this further derivative at a fixed x, hold the subtracted score constants at that x fixed during differentiation. The derivatives of the shifted integrands and the centered-score derivative of the posterior then have uniformly bounded mixed moments by the same mode estimate. The finite moment formula for cumulants gives a uniform Lipschitz bound for those terms. For distances above one, the sup norm suffices for the Hölder condition. Thus all the required bounds hold in the enlarged class.

Finally let $f_t = f * \gamma_{tA}$, $0 \le t \le 1$. It satisfies the continuity equation with velocity $-A\nabla \log f_t/2$. Its Fisher information $\int |\nabla \log f_t|^2 f_t$ is at most dM by integration by parts and the output Hessian upper bound, which also holds at $t = 0$. The dynamic transport action is at most $dM\|A\|_{\mathrm{op}}^2/4$. This proves the displacement estimate. □

9.3.1 A reverse bound for small common heat smoothing

Let h lie between fixed negative powers of n and let $0 \preceq A_0 \preceq Ch^2 I$. Put $f_A = f * \gamma_{A_0}$, $g_A = g * \gamma_{A_0}$. Uniformly in the smoothness interval,

$$W_2(f, g) \le C\big(W_2(f_A, g_A) + h^{\alpha+1}\big). \tag{71}$$

For the proof it suffices to restrict the mean gap to a fixed constant. If the mean gap exceeds one, uniform second moments about the means give $W_2(f, g) \le C|m_f - m_g|$, and the mean gap is unchanged by smoothing. Otherwise center coordinates within a fixed distance of the means.

Write $q = f - g$, $Pq = q * \gamma_{A_0}$, and $\rho_A = f_A + g_A$. The reverse weighted transport comparison gives $Pq = \operatorname{div} H$ with $\|H/\sqrt{\rho_A}\|_2 \le CW_2(f_A, g_A)$, using the uniform smoothed curvature constants. Choose an

integer J with $2(J-1) > \alpha_+ - 1$. The binomial identity gives

$$q = \sum_{j=1}^{J} (-1)^{j+1} \binom{J}{j} P^{j-1}(Pq) + (I-P)^J q.$$

The first sum is the divergence of the same linear combination of $P^{j-1}H$.

On a ball I of radius $R = D_0\sqrt{\log n}$, these fields have $L^2(1/\rho)$ norm at most $C\|H/\sqrt{\rho_A}\|_2$. Indeed the Gaussian ratio bound $\rho_A(x)/\rho(x) \le e^{C\|A_0\|(1+|x|^2)}$ follows by the upper bound from quadratic Taylor expansion as in (67), also for PSD matrices. For $x \in I$ each smoothed density has supporting log tangent and score $O(R)$, giving $\rho_A(y)/\rho(x) \le Ce^{CR|x-y|}$. After conjugating by $\sqrt{\rho_A}$ and $1/\sqrt{\rho}$, the absolute kernel for P^{j-1} is bounded by convolution with the Gaussian measure times $Ce^{CR|x-y|}$. This is a finite measure of uniformly bounded mass since $hR = o(1)$. Young's inequality works also for degenerate Gaussian measures, and $j = 1$ is the identity.

The remainder is the divergence of

$$-\tfrac{1}{2}A_0 \int_0^1 P_{tA_0}(I-P)^{J-1}\nabla q \, dt,$$

where P_{tA_0} is convolution by γ_{tA_0}. This follows by integrating the anisotropic heat equation once. The signed measures in the integral annihilate all polynomials of degree below $2(J-1)$, by the zero at the origin of their Fourier transforms. Their absolute moments at scale h are uniformly bounded, and their far tails at distances greater than $w = (2+R)^{-1}$ have arbitrarily fast polynomial decay in n, even when a component measure is degenerate. Subtract the local Taylor polynomial of ∇f and ∇g to degree $\alpha - 2$ for integer α and degree $k-1$ for fractional $\alpha = k + \theta$. The relative derivative bounds from Section 2 give a local remainder of order $h^{\alpha-1}$ times a fixed polynomial in $|x|$ times the component density. After multiplication by A_0 and integration, its $L^2(1/\rho)$ norm on I is $O(h^{\alpha+1})$. The far tails are negligible by the polynomial density floor on the ball. Finally choose D_0 large enough that the exterior term in (70) is $O(h^{\alpha+1})$, and use that inequality with the two fluxes and zero scalar field. This proves (71).

9.4 Estimating the common weights from inward probes

Use n observations per group. The first independent splits estimate the means; reject if their difference exceeds one. Otherwise take the first estimated mean as origin. With uniform probability $1 - o(1)$ after this screen the two population means are within a fixed distance of the origin. The screen has null error $o(1)$. Write $a = \log n$, $w = a^{-1/2}$ and $r_0 = a_{n,d}$. Using independent equal-sized pilot splits, set the density floor $\rho_0 = r_0^2 a^{2-d/2}$.

Condition on the first splits throughout the pilot.

Write $\rho = f + g$ and $\rho^{\text{obs}} = p_f + p_g$. Tile with w-cubes, retaining all cubes meeting the ball of radius $L = D_0\sqrt{a}$. There are $O(a^d)$ cubes. Write x for a cube center. For each observed density p separately, estimate $p(x)$ at inner rows $|x| \le c_1\sqrt{a}$ using its second sample count in the cube divided by the second sample size times w^d. For outer rows put $\omega = x/|x|$, $t = C_1 \log a/\sqrt{a}$, and count in the w-cubes centered at $x - t\omega$ and $x - 2t\omega$. Each count of at least $A_1 \log a$ yields a density estimate at its probe by the same division. If both counts attain this threshold, set the estimate at x to the square of the first probe estimate divided by the second. Otherwise set it to zero. The estimates for the two densities use their own observations.

Uniformly with probability $1 - o(1)$, every count above the threshold agrees with its mean up to fixed factors and every count of mean at least $2A_1 \log a$ is above the threshold. This follows from binomial Chernoff bounds and a union bound over $O(a^d)$ probes with A_1 large. A probe cube average agrees with its center value up to fixed factors because the observed score on the grid region is $O(\sqrt{a})$. The inner counts have polynomially large means, by the uniform Gaussian lower envelope with c_1 sufficiently small.

On the outer segments, the derivative of $-\log p$ in direction ω is at least $c\sqrt{a}$. Indeed the Hessian is bounded below and the mode is within a fixed distance of the origin. Each inward probe value is therefore at least $a^{cC_1}p(x)$. Since

$$nw^d\rho_0 = nr_0^2 a^{2-d} \gtrsim a^{-1/2},$$

every outer center with $p(x) \geq \rho_0$ has two reliable probes if C_1 is large. The Hessian upper bound gives

$$\big|\log p(x) - 2\log p(x - t\omega) + \log p(x - 2t\omega)\big| \leq Ct^2 = o(1).$$

Thus every positive estimate $\widehat{p}(x)$ is comparable to $p(x)$, and a zero estimate implies $p(x) < \rho_0$. No independence of counts within a sample is needed.

Use this single pilot in all trials. Write I for the union of retained cubes on which the sum of the two density estimates is greater than $K_1\rho_0$, and $\widehat{\rho}$ for that sum extended constantly on each selected cube. On the successful pilot event,

$$c\rho^{\mathrm{obs}}(x) \leq \widehat{\rho}(x) \leq C\rho^{\mathrm{obs}}(x) \quad (x \in I).$$

Values within each cube are comparable by the observed score. Any ignored term at a selected center is less than ρ_0. Choosing the grid radius $D_0\sqrt{a}$ large also ensures that the observed density sum outside the tiled ball is below ρ_0. This part holds under any admissible covariance, regardless of its size. For power in the range $s \leq n^{-\eta}$, the latent and observed values are comparable throughout the grid by (67). In this range, increasing C_2 if necessary gives

$$c\rho(x) \leq \widehat{\rho}(x) \leq C\rho(x) \quad (x \in I), \qquad \{\rho > C_2\rho_0\} \subseteq I. \tag{72}$$

In the small-noise power argument use either the latent sum or its common smoothing by covariance of size $O(h^2)$ with polynomially small h. The densities at all these stages are uniformly comparable on the grid, by the quadratic Taylor convolution bounds in Section 2 (which apply also to PSD Gaussian matrices). The inclusion in (72) holds for any of these sums after increasing C_2. For the analysis take a smooth cutoff as a scalar function of the relevant density sum divided by ρ_0, zero below C_2 and one above $2C_2$. Its support is in I (increase its lower threshold if needed) and its gradient is bounded by $C(1 + |x|)$. For each component density $u = f, g$ integrate on its potential levels about its mode. On a level z above the minimum, the radius and radial derivative are comparable to $\sqrt{z}$. The level-volume derivative is $O(z^{d/2-1})$ and $1 + |x|^2 \gtrsim z$ on large levels. Since $\log(1/\rho_0) \asymp a$, this proves

$$\int_{\{\rho \leq 2C_2\rho_0\}} \frac{\rho(x)}{1 + |x|^2}\, dx \leq C\rho_0 a^{d/2-2} = Cr_0^2.$$

Here the domain is contained in the individual low-density domain for each component. Only the analytical cutoff uses the unknown densities.

9.5 One common field and a family of corrections

Use $D(x) = 1 + |x|^2$ and take $b(x)^2 = \delta/D(x)$ for $d < 4$ and $b(x)^2 = 1/a$ for $d = 4$, where $\delta > 0$ is fixed. For a density u write

$$J_b(u)(x) = \tfrac{1}{2}\int_0^{b(x)^2} \nabla(u * \gamma_t)(x)\, dt, \qquad U_b(u)(x) = (u * \gamma_{b(x)^2})(x) + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla(u * \gamma_{b(x)^2})(x).$$

In each heat gradient the time is fixed. The heat equation and chain rule give $u = U_b(u) - \operatorname{div} J_b(u)$. Put $R_t = \frac{1}{2}\int_0^t \nabla\gamma_z\, dz$ and choose $\tau = n^{-K}$, where K is a fixed large constant. For the common observed field, use the scalar kernel

$$G_x^0 = \gamma_{b(x)^2} + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla\gamma_{b(x)^2}$$

and vector kernel $R_{b(x)^2} - R_\tau$. Normalize the scalar by $\sqrt{D(x)\widehat{\rho}(x)}$ and the vector by $\sqrt{\widehat{\rho}(x)}$, on I. These features have means given by the same normalized fields on the observed laws, up to the vector $R_\tau * p$. The observed gradient is uniformly bounded, so the vector error is $O(\tau)$ in sup norm. The polynomial density floor and grid volume make its normalized norm negligible at r_0 when K is large.

For a trial bandwidth h and a symmetric trial matrix M of small norm, write $S = \|M\|_{\mathrm{op}}$ and $t_c = h^2 + C_dS$. Take C_d a sufficiently large constant. The bandwidths lie between fixed powers of n^{-1}, and $S/h^2 \leq C + \gamma\log n$. In particular $t_c \ll b(x)^2$ on the grid, and $uI - M$ is positive definite and comparable to uI when $u \geq t_c$. The trial scalar correction kernel is $G_x^M - G_x^0$, where

$$G_x^M = \gamma_{b(x)^2 I - M} + \tfrac{1}{2}\nabla b(x)^2 \cdot \nabla\gamma_{b(x)^2 I - M}.$$

The trial vector kernel is the sum of

$$\tfrac{1}{2}\int_{t_c}^{b(x)^2} \nabla(\gamma_{uI-M} - \gamma_u)\, du$$

and a fine correction with multiplier

$$\frac{i\xi}{|\xi|^2}\left(\widehat{K}(h\xi) - e^{-t_c|\xi|^2/2}\right)\left(e^{\xi^T M\xi/2} - 1\right). \tag{73}$$

Cut off just the fine spatial kernel at distance w. Use the same two weights on I as for the common field. Both types of features are in the same real Hilbert direct sum $L^2(I) \oplus L^2(I;\mathbb{R}^d)$. All trials and the common field use the unaltered observations.

9.5.1 Population comparison at a nearby matrix

Whenever $0 \preceq A_0 = \Sigma - M \preceq 2h^2 I$, write $u = f_A = f * \gamma_{A_0}$ for the trial latent density in one group, and $p = f * \gamma_\Sigma$ for its observation density. Then $\widehat{u} = e^{\xi^T M\xi/2}\widehat{p}$ exactly, including for indefinite M. The scalar correction mean before weighting is $U_b(u - p)$. The untruncated vector correction has multiplier

$$\frac{i\xi}{|\xi|^2}(\widehat{K}(h\xi) - e^{-b(x)^2|\xi|^2/2})(e^{\xi^T M\xi/2} - 1)$$

on p with row parameters held fixed. Thus it equals $J_b(u - p)$ up to the high-pass field $J_h(u - p)$ in (68). At small true noise, p, u are comparable on I by the quadratic likelihood bounds and have the same uniform regularity by Lemma 9.2. Hence the high-pass error has weighted norm $O(h^{\alpha+1})$, applying (68) to each density. Indeed their squared densities divided by $\widehat{\rho}$ are at most constants times the respective densities there. Gaussian moments and the density floor pay the local Taylor and far terms, respectively.

The rescaled fine kernel in (73) is h^{1-d} times a Schwartz profile at h, with any needed finite list of seminorms bounded by $E = C(1 + S/h^2)^{C_0} e^{C_0 S/h^2}$. The radial bracket vanishes to second order at zero, so division by $|\xi|^2$ is smooth even for indefinite M. The exterior Fourier term has strictly Gaussian decay since $t_c I - M$ is positive definite at scale h^2. These facts and differentiation on the compact cutoff give the claimed bound. Choose the necessary fixed tail orders before taking the allowed constant γ small. Since h is polynomially small, $S/h^2 = O(\log n)$ and the pilot weight has a polynomial floor, the spatial truncation costs $o(r_0)$ in the weighted mean norm.

Denote by μ_0 and μ_M the differences of the common mean features and the trial correction mean features. The exact heat identity for p and $u - p$, followed by (70), bounds $W_2(f_A, g_A)$ by a constant times $r_0 + h^{\alpha+1} + \|\mu_0\| + \|\mu_M\|$. Use (71) to conclude

$$W_2(f, g) \le C(r_0 + h^{\alpha+1} + \|\mu_0\| + \|\mu_M\|). \tag{74}$$

Under the null all feature differences have mean zero for every trial because the two observed distributions agree.

9.5.2 Conditional covariance

Consider the successful observed-density pilot event. We first bound the uncentered second moments of the common scalar and vector separately; their sum multiplied by two dominates the centered covariance in positive operator order. Fix either observed law with density p. Its supporting log-density tangent at each row and the comparison with ρ^{obs} on I give

$$\frac{p(y)}{\sqrt{\widehat{\rho}(x)\widehat{\rho}(x')}} \le C\exp(C(1 + |x|)|y - x| + C(1 + |x'|)|y - x'|)\,.$$

Each factor is absorbed in the row Gaussian, since $b(x)(1 + |x|) \le C$.

After the tilt, the scalar kernel before the $D^{-1/2}$ factor is bounded by $Cb(x)^{-d} e^{-c|x-y|^2/b(x)^2}$. For the vector use $|\nabla\gamma_t(z)| \le C|z|t^{-(d+2)/2} e^{-c|z|^2/t}$. Integrating over $t \ge \tau$ while keeping half of the Gaussian at

scale $b(x)$ gives $C(\sqrt{\tau}+|x-y|)^{1-d}e^{-c|x-y|^2/b(x)^2}$. The spatial factor in the gradient also gives this bound in dimension one. For $d<4$, at pairs with $1+|x| \asymp 1+|x'|$ (ratio between $1/2$ and 2), use common bandwidth $b(x)$. The scalar covariance kernel has squared row integral bounded by $Cb(x)^{-d}D(x)^{-2}$. For the coarse vector use $C|z|^{1-d}e^{-c|z|^2/b(x)^2}$, whose $L^{4/3}$ norm is $O(b(x)^{1-d/4})$; Young's inequality gives the same row bound $Cb(x)^{4-d}$. For noncomparable pairs, the separation is comparable to the larger radius. Extract a Gaussian in separation using both row Gaussians with bounded bandwidth. Scalar prefactors then grow at most polynomially in the radii. For the vector, split into balls of radius one third of the separation about either row and their complement. On each ball the other singularity is bounded by a polynomial in separation and the singularity at the ball center is integrable against its Gaussian. On the complement both singularities are bounded. The two squared kernels are thus integrable on this pair region uniformly in n. Integrating the comparable row bound $Cb(x)^{4-d}$ on I gives $O(1), O(\log a), O(a)$ in dimensions one, two, three respectively.

For $d=4$, the bandwidth b is constant. Squared scalar Hilbert–Schmidt norm is at most $Cb^{-d}\int_I D(x)^{-2}\,dx = O(a^2\log a)$ by Gaussian convolution and Cauchy–Schwarz in the row variables. The absolute vector convolution at separation z is bounded by

$$C(\sqrt{\tau}+|z|)^{2-d}e^{-c|z|^2/b^2}.$$

At $|z| \le \sqrt{\tau}$ this follows by Cauchy–Schwarz. Above that distance, split near each center and in the remainder: the near integrals are bounded by $|z|^{1-d}\int_0^{|z|} dt$ and the far integral by $\int_{|z|}^{\infty} t^{1-d}\,dt$, extracting a Gaussian in z. The square integrates in z to $O(1+\log(b/\sqrt{\tau})) = O(a)$. Multiplying by the row volume $O(a^2)$ yields $O(a^3)$. In all dimensions the same row bounds give bounded operator norm by Schur integration. For the variable-bandwidth case, on comparable row radii the scalar envelope integrates to $O(D(x)^{-1})$ and the coarse vector convolution to $O(b(x)^2)$ by its L^1 profile. On noncomparable radii the Gaussian in separation gives uniformly bounded row and column integrals. Consequently the common centered covariance C_0 of a paired feature difference satisfies

$$\|C_0\|_{\mathrm{HS}}/n \le Cr_0^2, \qquad \|C_0\|_{\mathrm{op}} \le C. \tag{75}$$

For the scalar correction, interpolate along covariance $b(x)^2 I - vM$, $0 \le v \le 1$. One covariance derivative introduces $S/b(x)^2$ times a Gaussian of the same spatial scale. The same bound applies to the chain-rule term, as $|\nabla b^2| \le Cb^3$ when b varies. The scalar covariance operator and Hilbert–Schmidt norms are thus polynomially small in n, using $S \le h^2(C+\gamma a)$, $b(x)^{-2} \le Ca$, and the grid volume.

For the coarse vector correction let $\lambda = \sqrt{t_c}$. Interpolating the covariance introduces S times two spatial derivatives of the gradient Gaussian, bounded before tilt by $CSu^{-(d+3)/2}e^{-c|z|^2/u}$ at time u. Retaining half the Gaussian at scale $b(x)$ and integrating over $u \ge t_c$ bounds the spatial kernel by

$$CS(\lambda+|z|)^{-d-1}e^{-c|z|^2/b(x)^2}.$$

After the score tilt, the Gaussian can be discarded, leaving a common profile with L^1 norm $O(\lambda)$ and $L^{4/3}$ norm $O(\lambda^{1-d/4})$, because $S \lesssim \lambda^2$. Young's inequality and the row volume give Hilbert–Schmidt norm $O(\lambda^{2-d/2}a^{d/4})$. The row and column integrals of the convolution are $O(\lambda^2)$, giving the operator bound.

The fine correction is bounded by its rescaled Schwartz profile on $|z| \le w$, where the tilt is bounded. The identical convolution argument gives Hilbert–Schmidt norm $O(E^2h^{2-d/2}a^{d/4})$ and operator norm $O(E^2h^2)$. Positive operator domination by the three uncentered kernels gives the corresponding bound for the centered correction, and also for the sum of two independent group covariances. For every trial, λ is polynomially small; so is E^2h^2 when γ is sufficiently small. Consequently its paired covariance C_M satisfies

$$\|C_M\|_{\mathrm{op}} \le Cn^{-c_3}, \qquad \|C_M\|_{\mathrm{HS}} \le \begin{cases} Cn^{-c_3}, & d<4, \\ CE^2a, & d=4, \end{cases} \tag{76}$$

for some $c_3 > 0$. For $d<4$, use $2-d/2>0$ and choose γ small enough to pay even the $E^2h^{2-d/2}$ factor. These estimates depend on observed density ratios and therefore do not require correct trial noise.

9.6 The conditional sign test

Reserve two equal-sized paired blocks of order n, independent of the two earlier splits. Each paired observation produces $Z = F_Y - F_{Y'}$. Use the first paired block to form its mean $\widehat{\mu}$. In the other block of size m_v, put $X_i = \langle \widehat{\mu}, Z_i \rangle$. The trial rejects if

$$\sum_{i=1}^{m_v} X_i > \sqrt{2\ell \sum_{i=1}^{m_v} X_i^2}. \tag{77}$$

Under the null, conditional on the earlier splits, trial choices and training block, each paired difference is symmetric, and the validation observations are independent. Conditional additionally on the magnitudes of the X_i, the nonzero signs are independent and fair. The bound $\cosh(u) \le e^{u^2/2}$ for the exponential moment of the signed sum gives conditional rejection probability at most $e^{-\ell}$. If all magnitudes vanish the strict inequality prevents rejection. This level bound uses neither correct nominal noise nor a successful pilot. All validation observations are unaltered.

For the power bound for either feature let $u = \|\mu\|$, and write C for the paired covariance. With $T = m_v^{-1} \sum_i X_i$ and $W = m_v^{-2} \sum_i X_i^2$, the two independent blocks give

$$\mathbb{E}T = u^2, \qquad \mathrm{Var}(T) \le C_2\big(\|C\|_{\mathrm{op}} u^2/n + \|C\|_{\mathrm{HS}}^2/n^2\big).$$

Indeed the variance of the conditional mean $\langle \widehat{\mu}, \mu \rangle$ is $\langle \mu, C\mu \rangle / m_t$, where $m_t \asymp n$ is the training size, and the conditional variance averages to $(\langle \mu, C\mu \rangle + \mathrm{tr}(C^2)/m_t)/m_v$. The second moment of a validation projection gives

$$\mathbb{E}W \le C_2\big(u^4/n + \|C\|_{\mathrm{op}} u^2/n + \|C\|_{\mathrm{HS}}^2/n^2\big).$$

For $u > 0$, combining the two moments with Chebyshev and Markov proves power at arbitrarily small fixed error if

$$u^2 \ge K \max\{\ell \|C\|_{\mathrm{op}}/n,\ \sqrt{\ell}\|C\|_{\mathrm{HS}}/n\}$$

for a sufficiently large fixed K, provided $\ell/n = o(1)$. Indeed $T > u^2/2$ and $\sqrt{2\ell W} \le u^2/2$ hold jointly with the required probability. For the common feature take ℓ a large fixed constant depending on the prescribed error. Equation (75) then gives power once $\|\mu_0\| \ge C r_0$, using $r_0^2 \ge 1/n$.

9.7 Choosing the correction trials

Take dyadic bandwidths $h_j = 2^j r_0^{1/(\alpha_- + 1)}$ up to $h_j \le n^{-\eta/4}$ for a fixed small $\eta > 0$. For each h use a deterministic PSD operator-norm $h^2/2$-net Ξ in the ball $\|\Xi\|_{\mathrm{op}} \le \gamma h^2 \log n$, and use $M = \Xi - h^2 I$ in the correction fields. There are at most a fixed power of $\log n$ total trials. With M_n their number, choose $\ell = \log(C_\varepsilon(1 + M_n)) = O(\log\log n)$ for the correction tests. Their total null error is controlled by the sign inequality and the union bound. Add the single common-field test and the preliminary mean rejection, with error budgets chosen so the desired fixed level holds. The observed-density pilot only affects power, since the conditional sign inequality holds for any pilot.

For true $s \le n^{-\eta}$, take h within a fixed factor of a sufficiently large constant multiple of

$$\max\{r_0^{1/(\alpha+1)}, \sqrt{s/B_{n,d}(s)}\}.$$

It lies below the cap by a fixed polynomial factor, for small η, and above the lowest scale uniformly in α. The large fixed multiplier makes s no larger than the net radius, so there is a nearby trial $\|\Xi - \Sigma\|_{\mathrm{op}} \le h^2/2$. Then $A_0 = \Sigma - M$ lies between 0 and $2h^2 I$, and (74) applies. Also $h^{\alpha+1} \asymp r_{n,d}(s)$ and $S \le s + 2h^2$.

For $d < 4$, (76) pays both power terms for the sign test at r_0^2, since the norms are polynomially small while ℓ grows only logarithmically. For $d = 4$, put $X = ns^{\alpha+1}$ and $B = \log(2 + X)$. The chosen bandwidth multiplier gives $E^2 \le C(2 + X)^c$ for any sufficiently small fixed $c > 0$. If $X \le a^2$, then

$$\sqrt{\ell}\|C_M\|_{\mathrm{HS}}/n \le C\sqrt{\log a}\, a^{1+2c}/n \le C' a^{3/2}/n$$

for $c < 1/4$, so the direct scale pays this term. For $X > a^2$, use $a \le \sqrt{X}$ and $\sqrt{\log a} \le (\log X)^{1/2}$ to obtain

$$\sqrt{\ell}\|C_M\|_{\mathrm{HS}}/n \le C X^{1/2+c} (\log X)^{1/2}/n \le C' X/(nB^{\alpha+1}).$$

This is paid by the squared inverse scale. The operator term in either case is negligible by (76). Therefore the correction test has power at a sufficiently large multiple of $r_{n,d}(s)$. Equation (74) implies that a sufficiently separated alternative has a large enough mean in the common test or in this correction test. Since only one trial needs power, we incur only its $o(1)$ pilot error.

9.8 A common empirical test at larger noise

If $\delta = W_1(p_f, p_g) \le 1$, their means are within one. Translate by the mean of g and apply (70) with $U = U_h(f) - U_h(g)$, $J = J_h(f) - J_h(g)$ on a ball of radius $D\sqrt{\log(1/h)}$. Take D large enough that the Gaussian exterior contributes $O(h^{\alpha+1})$. On this ball there is a uniform density floor h^P for some fixed P, and (68) gives an $O(h^{\alpha+1})$ weighted bound on the vector field, paying each local bias in its own density. For the actual common covariance, the Gaussian characteristic function can be removed on the Fourier cutoff at cost at most $e^{s/(2h^2)}$. Coupling the output laws bounds their characteristic function difference by $\delta|\xi|$. Fourier inversion on the cutoff support then gives $\|U\|_\infty \le C\delta h^{-d-1} e^{s/(2h^2)}$. Consequently

$$W_2(f,g) \le C\left[h^{\alpha+1} + \delta h^{-d-1-P/2} \log^{d/4}(1/h) e^{s/(2h^2)}\right]. \tag{78}$$

Uniformly over the class, the empirical observed W_1 error e_n has expectation $O(n^{-\beta})$ with $\beta = 1/[4(d+1)]$. For an elementary proof translate to the theoretical mean, truncate at distance n^β using the uniform second moment and then quantize at mesh $n^{-\beta}$. Both costs are $O(n^{-\beta})$. There are $O(n^{2d\beta})$ sites, so Cauchy–Schwarz for the Bernoulli counts bounds expected total variation by $Cn^{d\beta-1/2}$. Multiplying by the grid diameter gives $O(n^{-1/4})$. Use the empirical observed W_1 distance between both samples, rejecting above $2n^{-\beta/2}$. Under the null, Markov's inequality and the triangle inequality give error $o(1)$. For any alternative whose observed discrepancy is $\delta \le 5n^{-\beta/2}$, use

$$h = \max(n^{-c_1}, \sqrt{s/(\gamma_1 \log n)})$$

in (78), with constants chosen so that $c_1(\alpha + 2 + d + P/2) + \gamma_1/2 < \beta/2$. The two terms in (78) are then $O(h^{\alpha+1})$. If $s \ge n^{-\eta}$ with $\eta > 0$ sufficiently small, the second entry of h dominates and $B_{n,d}(s) \asymp \log n$. Thus for a sufficiently large multiple of $r_{n,d}(s)$ in latent distance, the observed discrepancy exceeds $5n^{-\beta/2}$. If each empirical error is at most $n^{-\beta/2}$, the empirical discrepancy is greater than $3n^{-\beta/2}$ and the test rejects. The exceptional probability is $o(1)$ by the two Markov bounds.

Add this empirical test to the rejection rule at every n. It has null error $o(1)$ uniformly over the covariance and its rejection rule does not use s. The power argument above only uses s analytically and applies for every $s \ge n^{-\eta}$.

9.9 Lower bounds from mixtures

Fix the common covariance at any allowed matrix. Keep the larger group at a centered isotropic Gaussian g with strictly interior precision, and vary only the smaller group. The unchanged independent group cancels out of likelihood ratios. Write n for the smaller size and s for the largest eigenvalue of this fixed covariance. The direct mixtures hold after arbitrary convolution and the inverse packet can be aligned with a top covariance eigenvector.

We will use finite families $f_\omega = g + \sum_{i=1}^N \omega_i q_i$ where ω_i are independent signs and each q_i has integral zero. Let G be the Gram matrix of the noisy score perturbations in $L^2(p_g)$:

$$G_{ij} = \int \frac{(q_i * \gamma_\Sigma)(q_j * \gamma_\Sigma)}{g * \gamma_\Sigma}.$$

For the uniform mixture of n-sample laws over signs, one plus the chi-square divergence to the null equals $\mathbb{E}_{\omega,\omega'}(1 + \omega^T G \omega')^n$. The factor inside the power is nonnegative as the inner product of two likelihood ratios. Bounding its power by $\exp(n\omega^T G\omega')$ and integrating one sign vector gives $\mathbb{E}_\omega \exp(n^2\|G\omega\|^2/2)$ as an upper bound, using $\cosh(t) \le e^{t^2/2}$. Introduce a standard Gaussian vector to linearize the square and apply the same sign bound again. The result is at most

$$\det(I - n^2 G^2)^{-1/2} \le \exp(Cn^2 \|G\|_{\mathrm{HS}}^2) \qquad \text{if } n\|G\|_{\mathrm{op}} \le 1/2.$$

Thus a small bound on $n^2\|G\|_{\mathrm{HS}}^2$ makes the mixture total variation arbitrarily small. If every mixture member has $W_2(f_\omega, g) \geq r$, the sum of testing errors at separation r is at least one minus that total variation.

For transport separation we use a smooth test φ. Taylor expansion under an optimal coupling gives

$$\int \varphi(f_\omega - g) \leq \|\nabla\varphi\|_{L^2(f_\omega)} W_2(f_\omega, g) + \tfrac{1}{2}\|D^2\varphi\|_\infty W_2(f_\omega, g)^2. \tag{79}$$

If the pairing dominates cE, the gradient norm is at most $C\sqrt{E}$, and the Hessian norm is bounded, it follows that $W_2^2(f_\omega, g) \geq c'E$.

9.9.1 The direct critical lower bound

When $d = 4$, pack disjoint cubes of side $\ell = (\log n)^{-1/2}$ inside a cube of side $c\sqrt{\log n}$ about the mode, with $c > 0$ small. Their total volume is order $(\log n)^2$. The reference g is comparable to its value g_B at the center of each cube B, and $g_B \geq n^{-\beta_0}$ for a fixed $\beta_0 < 1$. In the rescaled unit cube the ratio g_B/g has bounded C^α norm uniformly: on length ℓ, the score bound is paid by $\ell\sqrt{\log n} = 1$, and all higher potential derivatives are bounded with the corresponding scale factors. The top Hölder bound scales by ℓ^α.

Use a fixed orientation of compactly supported orthonormal tensor wavelets [7] of regularity C^{k+1}, supported strictly inside each cube. At physical length $r_i = \ell 2^{-j}$ their number in a cube is comparable to $(\ell/r_i)^4$, for j from a sufficiently large fixed index through $\lfloor \nu \log_2 n \rfloor$, with $\nu > 0$ small. Write ψ_i for these L^2-normalized wavelets. In particular $S_4 = \sum_i r_i^4 \asymp (\log n)^3$: each level contributes order the total volume. Set

$$b_i = \frac{\kappa^2 r_i^2}{n\sqrt{S_4}}, \qquad a_i = \sqrt{g_{B(i)} b_i}, \qquad q_i = a_i \psi_i.$$

Each wavelet has zero integral. At physical scale r, the relative perturbation in cube B has C^α norm at most $C(a_i/g_B) r^{-2-\alpha}$, by wavelet scaling, bounded overlap at the scale, and the rescaled reciprocal density bound. Summing over scales gives at most

$$C\kappa n^{-(1-\beta_0)/2} S_4^{-1/4} r_{\min}^{-1-\alpha} = o(1)$$

if $\nu < (1-\beta_0)/(2(1+\alpha))$. For fractional order this controls the global seminorm: on a pair of points in distinct cubes one may use the zero values of the top derivatives on the boundary of each cube and distances to the first exit points on the segment. Thus all perturbations stay within the reference margins by logarithmic composition.

The clean Gram matrix is $D_0 H D_0$ with $D_0 = \mathrm{diag}(\sqrt{b_i})$, and $H \preceq CI$. This uses orthonormality and the uniform bounds on g_B/g within each cube. The noisy Gram shrinks in positive operator order as before, so $n^2\|G\|_{\mathrm{HS}}^2 \leq Cn^2 \sum_i b_i^2 = C\kappa^4$. For the metric test take $\varphi = \sum_i (r_i^2/g_{B(i)})\omega_i a_i \psi_i$. Its pairing is exactly $E = \sum_i r_i^2 b_i = \kappa^2\sqrt{S_4}/n$. The gradient squared norm in a cube is controlled by the sum of squared coefficient norms weighted by r_i^{-2}, up to a fixed factor. For clarity, within one level sum u_j, $\|\Delta u_j\|_2 \leq C r_j^{-2}\|u_j\|_2$ by overlap. When $r_j \geq r_\ell$, integration by parts yields

$$|\langle \nabla u_j, \nabla u_\ell \rangle| \leq C(r_\ell/r_j)(\|u_j\|_2/r_j)(\|u_\ell\|_2/r_\ell).$$

Summing the geometric ratios and multiplying by the comparable density in each cube gives squared gradient norm $O(E)$ under f_ω. The test Hessian is uniformly bounded, since its per cube scale sum is bounded by $C\kappa(ng_B)^{-1/2} S_4^{-1/4} r_{\min}^{-1}$. Thus (79) gives separation at least $c(\log n)^{3/4}/\sqrt{n}$. This lower bound holds under every known Gaussian covariance, by the same mixture contraction.

9.9.2 The inverse mixture in dimension four

Rotate axes so $\Sigma = \mathrm{diag}(s, t_2, \ldots, t_d)$ with $0 \leq t_i \leq s$, and suppose $0 < s \leq s_0$ with $s_0 > 0$ small. Place $N \asymp s^{-d/2}$ centers z_i spaced $D\sqrt{s}$ apart in a fixed cube. Here D is sufficiently large, and s_0 small after choosing D. Put $T^2 = A_0 \log(2 + ns^{\alpha+d/4})/s$, and use

$$q_i(x) = a_0 \psi_i(x), \qquad a_0 = \kappa T^{-\alpha}, \qquad \psi_i(x) = e^{-|x-z_i|^2/(2s)} \left(\cos(T(x_1 - z_{i,1})) - e^{-sT^2/2} \right).$$

Each packet has zero mass. The relative perturbation before the factor a_0 has derivatives through order $j \le k$ bounded by CT^j. Indeed derivatives of the packets are bounded by $CT^j e^{-c|x-z_i|^2/s}$ because $T\sqrt{s}$ is large, and division by g costs only Gaussian and polynomial factors absorbed in a weaker envelope for small s_0. Separated grid envelopes have uniformly bounded sum. For $0 < \theta < 1$, the k-th derivative has θ-Hölder seminorm at most CT^α. For shifts longer than $1/T$ use the sup norm. For a shorter shift expand the product-rule terms of $D^k(\psi_i/g)$. A difference of a smooth packet derivative of order j costs $CT^{j+1}|u|$ with the same weakened envelope. A difference of a reciprocal derivative of order $k-j$ costs $C|u|^\theta$ times polynomial factors and the maximum of the two reciprocal densities, by the product rules and the reference Hölder condition on $D^k V_g$. The nearby packet envelopes still absorb those factors since $|u| \le 1/T \le C\sqrt{s}$. Summing the bounds gives $CT^\alpha|u|^\theta$. The logarithmic perturbation thus costs $O(\kappa)$ in the stipulated norms and preserves the class for small enough κ.

Convolution with the noise changes the packet envelope to variances $s + t_i$, with $t_1 = s$, multiplying its amplitude by $\prod_i \sqrt{s/(s+t_i)}$. Its bracket becomes

$$e^{-sT^2/4}\cos\big(\tfrac{T}{2}(y_1 - z_{i,1})\big) - e^{-sT^2/2}.$$

These formulas follow by conditional Gaussian integration and include zero eigenvalues in the other coordinates. The convolved reference has a lower Gaussian envelope, obtained by convolving one for g. Combining the two narrow packet envelopes and dividing by this lower envelope gives

$$|G_{ij}| \le Ca_0^2 s^{d/2} e^{-csT^2} e^{-c|z_i - z_j|^2/s}.$$

Indeed the common center lies in a fixed cube, the variances of the two envelopes are between s and $2s$, and choosing s_0 small absorbs the reciprocal reference into a weaker Gaussian. Thus

$$n\|G\|_{\mathrm{HS}} \le Cna_0^2 s^{d/4} e^{-csT^2}.$$

With $X = ns^{\alpha+d/4}$, the right-hand side is at most $C\kappa^2 X(A_0 \log(2+X))^{-\alpha}(2+X)^{-cA_0}$. It is arbitrarily small by choosing A_0 large and κ small.

For transport separation take $\varphi = T^{-2}\sum_i \omega_i q_i$. The spatial L^2 Gram matrix of the ψ_i has diagonal at least $cs^{d/2}$ by integrating the squared cosine in its Gaussian envelope and taking $T\sqrt{s}$ large. The absolute off-diagonal row sums are at most $Cs^{d/2}\sum_{j\in\mathbb{Z}^d\setminus\{0\}} e^{-cD^2|j|^2}$, less than half the diagonal bound for large D. The pairing with $f_\omega - g$ is at least cE for $E = T^{-2}a_0^2 s^{d/2} N$. The gradient Gram has absolute row sums at most $CT^2 s^{d/2}$ by the same overlap estimate, so the test gradient has squared norm $O(E)$ under the uniformly bounded vertex densities. The pointwise derivative envelope bounds the test Hessian by Ca_0. Equation (79) gives $W_2(f_\omega, g) \ge cT^{-\alpha-1}$ for every sign.

For $s \ge s_0$, use instead one additive packet throughout, replacing its envelope variance by a fixed τ_0^2 smaller than the variance in the Gaussian lower envelope for g, with its cosine mean removed. Use amplitude $\kappa T^{-\alpha}$ and $T^2 = C\log n/s$. The derivative and Hölder argument above applies to this single packet. Its convolved Gaussian envelope squared, divided by the convolved reference lower envelope, is integrable uniformly in Σ by the variance choice. Its first-coordinate cosine attenuates by $\exp(-\tau_0^2 sT^2/[2(\tau_0^2+s)])$, so the n-sample chi-square tends to zero uniformly for sufficiently large C. Testing the latent difference with $\cos(Tx_1)/T$ gives separation of order $T^{-\alpha-1}$. This matches the inverse term since its logarithm is comparable to $\log n$ on this range.

9.9.3 Lower bounds in dimensions one through three

In dimension one a shift of the reference by δ has W_2 distance $|\delta|$. Its clean one-sample relative entropy to the reference is $O(\delta^2)$, by expanding the potential and using the zero mean score. Since noise contracts entropy, $\delta = c/\sqrt{n}$ gives the desired two-point lower bound.

For $d = 2, 3$, take disjoint annuli at dyadic radii R from a fixed constant to $c\sqrt{\log n}$, for small fixed $c > 0$. Pack cubes of side $\ell_i \asymp R^{-1}$ in each annulus, with order R^{2d} cubes at radius R. The reference values g_i at their centers satisfy $g_i \ge n^{-\beta_0}$, $\beta_0 < 1$, and $g \asymp g_i$ inside each cube. Use one L^2-normalized smooth mean-zero bump ψ_i per cube, with

$$S_4 = \sum_i \ell_i^4, \qquad b_i = \frac{\kappa^2 \ell_i^2}{n\sqrt{S_4}}, \qquad q_i = \sqrt{g_i b_i}\,\psi_i.$$

The rescaled reciprocal bounds used for the critical construction apply here since $\ell_i(1+|x|) = O(1)$. All relative C^α perturbations tend to zero because $g_i \geq n^{-\beta_0}$ and $\ell_i \gtrsim (\log n)^{-1/2}$. The clean Gram matrix is diagonal with entries $O(b_i)$, and the noisy Gram contracts. Hence $n^2\|G\|_{\mathrm{HS}}^2 = O(\kappa^4)$. The test $\varphi = \sum_i(\ell_i^2/g_i)\omega_i q_i$ has pairing $\sum_i \ell_i^2 b_i$, squared gradient norm at most a constant times that pairing, and bounded Hessian by rescaling on the disjoint cells. Equation (79) gives squared separation of order $\sqrt{S_4}/n$. A ring contributes order R^{2d-4} to S_4; summing gives order $\log\log n$ for $d = 2$ and $\log n$ for $d = 3$.

For the inverse lower bound in all three dimensions, the single additive narrow packet with $T^2 = C\log n/s$ works throughout $0 < s \leq \Lambda$. Its attenuated cosine multiplier is bounded by e^{-csT^2} uniformly for the fixed narrow variance, and the noisy reference denominator is paid as above. It gives latent separation of order $(s/\log n)^{(\alpha+1)/2}$ with small total variation. Together with the direct lower bounds this completes Theorem 9.1.

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. 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, Luminary Research Archive, edition 4 (2026). `https://research.luminaryoddities.org/a-sharp-logarithmic-inverse-bound-for-gaussian-estimation`.

[2] T. Tran and N. Schreuder, Minimax-optimal two-sample test with sliced Wasserstein, *Proc. AISTATS* (2026); arXiv:2510.27498. `https://arxiv.org/abs/2510.27498`.

[3] S. Schubr, Sharp transport identification with unknown Gaussian covariance, Luminary Research Archive, edition 3 (2026). `https://research.luminaryoddities.org/sharp-transport-identification-unknown-gaussian-covariance`.

[4] 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.

[5] H. J. Brascamp and E. H. Lieb, *On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation*, J. Funct. Anal. **22** (1976), 366–389. DOI.

[6] C. Butucea, *Goodness-of-fit testing and quadratic functional estimation from indirect observations*, Ann. Statist. **35** (2007), 1907–1930. DOI.

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

[8] N. Deo and T. Randrianarisoa, *On adaptive confidence sets for the Wasserstein distances*, Bernoulli **29** (2023), 2119–2141. DOI.

[9] B. Lee, I. Kim, and C. Park, *Minimax Wasserstein testing under differential privacy*, arXiv:2610.14432v1, 2026.

[10] Y. Hur and T. Liang, *Detecting weak distribution shifts via displacement interpolation*, J. Bus. Econom. Statist. **43** (2025), 178–190. DOI.

[11] T. Manole, S. Balakrishnan, and L. Wasserman, *Minimax goodness-of-fit testing in Wasserstein distance*, CMStatistics 2023, abstract. Conference abstract.

[12] S. Schubr, *Sharp Wasserstein stability for Gaussian deconvolution of log-concave laws*, version 4, Luminary Research Archive, 2026. Archive edition.

[13] R. Tang and Y. Yang, *Minimax nonparametric two-sample test under adversarial losses*, Proc. AISTATS, PMLR **206** (2023), 6132–6165. Proceedings.

[14] 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. Article.

[15] B. T. Tran and N. Schreuder, *Minimax-optimal two-sample test with sliced Wasserstein*, Proc. AISTATS, PMLR **300** (2026), 1927–1935. Proceedings.

[16] T. Manole, *Minimax nonparametric testing in Wasserstein distance*, Göttingen Stochastics Colloquium, 12 June 2024, abstract. Colloquium abstract.

[17] D. H. Lee, S. N. Lahiri, and S. Sinha, *A test of homogeneity of distributions when observations are subject to measurement errors*, Biometrics **76** (2020), 821–833. DOI.

[18] J. Johannes, *Deconvolution with unknown error distribution*, Ann. Statist. **37** (2009), 2301–2323. DOI.

[19] F. Comte and C. Lacour, *Pointwise deconvolution with unknown error distribution*, C. R. Math. Acad. Sci. Paris **348** (2010), 323–326. DOI.

[20] A. Delaigle, P. Hall, and A. Meister, *On deconvolution with repeated measurements*, Ann. Statist. **36** (2008), 665–685. Article.

[21] C. Scricciolo, *Adaptive minimax-optimal Wasserstein deconvolution with unknown error distributions*, Statist. Probab. Lett. **230** (2026), 110589. DOI.

[22] M. Broniatowski, J. Jurečková, A. K. Moses, and E. Miranda, *Composite tests under corrupted data*, Entropy **21** (2019), 63. DOI.

[23] D. H. Lee, *Testing statistical hypotheses for latent variable models and some computational issues*, doctoral dissertation, Texas A&M University, 2018, Chapter 2. Repository copy.

[24] C. Butucea, C. Matias, and C. Pouet, *Adaptivity in convolution models with partially known noise distribution*, Electron. J. Stat. **2** (2008), 897–915. DOI.

[25] 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.

[26] C. Butucea, C. Matias, and C. Pouet, *Adaptive goodness-of-fit testing from indirect observations*, Ann. Inst. Henri Poincaré Probab. Stat. **45** (2009), 352–372. DOI.

[27] A. Schrab, I. Kim, M. Albert, B. Laurent, B. Guedj, and A. Gretton, *MMD aggregated two-sample test*, J. Mach. Learn. Res. **24** (2023), article 194. Article.

[28] S. Shekhar, I. Kim, and A. Ramdas, *A permutation-free kernel two-sample test*, Adv. Neural Inf. Process. Syst. **35** (2022). Preprint.

[29] V. G. Spokoiny, *Adaptive hypothesis testing using wavelets*, Ann. Statist. **24** (1996), 2477–2498. DOI.

[30] Y. Wang, L. Wang, S. Liu, and T. Suzuki, *Zero-flow two-sample tests*, arXiv:2607.21542v2, 2026. Preprint.