Recovering small additive noise from boundaries, vertices and corners

Scott Schubr
Luminary Oddities

November 2026

Abstract

When a known latent law has a sharp boundary, small additive noise blurs only a thin layer near it, and the observations near the boundary carry the leading-order information about the noise. We determine this information exactly. At a uniform box and at any full-dimensional polytope, the first unmatched normal projection of the noise at codimension k gives the exact b^k squared Hellinger coefficient, and at critical sample sizes the local experiments converge to marked Poisson processes, with independent Gaussian limits for covariance shifts. Vertex profile windows control negative Sobolev norms of the noise, and curved convex bodies give a weighted negative Sobolev form of order $(d+1)/2$ with relative error $O(b^2)$, leading to sharp minimax Wasserstein estimation and testing rates. At power-law boundaries and in one dimension the effective sample size is Nb^r, with critical logarithmic factors at $r = 1/2$ and $r = 3/4$. Product densities give mixed negative Sobolev probes from a single corner, sharp rates under mixed smoothness with logarithmic multiplicities from tied faces, exact information scales for mixed noise moments, and Gaussian, compensated Poisson or degenerate sampling limits for minimum-variance moment probes.

Contents

1 Introduction **3**

2 Normal and Poisson limits for noise at a box **3**
2.1 Results . . . 3
2.2 Localization by matching coordinate marginals . . . 5
2.3 The critical Poisson experiment . . . 6
2.4 Moment bounds for covariance derivatives . . . 6
2.5 The two local normal scales . . . 7
2.6 A Gaussian–Poisson product at the edge scale . . . 8

3 Tangent-cone experiments for additive noise on polytopes **9**
3.1 Results . . . 10
3.2 Localization by the active facets . . . 11
3.3 Profiles and the critical experiment . . . 12
3.4 Cone injectivity and examples at every order . . . 13
3.5 Gaussian covariance at the ridges . . . 14
3.5.1 The ridge Fisher form . . . 14
3.6 Local covariance experiments on a polytope . . . 15

4 Recovering small additive noise from polytope vertices **16**
4.1 Model and rates . . . 17
4.2 Exact vertex identity . . . 17
4.3 Compact vertex windows without fixed projections . . . 19
4.3.1 How many vertices are needed for a cross-polytope? . . . 20
4.4 Observation comparison and upper bounds . . . 21
4.5 Localized alternatives and lower bounds . . . 22

5 Recovering small additive noise from curved convex bodies **23**
5.1 Observation and forward norm . . . 23
5.2 Statistical recovery . . . 25
5.3 Sharp inverse modulus . . . 27
5.4 Positively curved reference bodies . . . 27

6 Noise recovery at power-law boundaries **28**
6.1 Model and conclusions . . . 29
6.2 Uniform shell inversion . . . 30
6.3 Information bounds for the density model . . . 31
6.4 Matching lower bounds . . . 32
6.5 Three Hellinger scales and the Poisson experiment . . . 34
6.6 Estimating a fixed noise moment . . . 35
6.7 The mixed experiment at the logarithmic threshold . . . 36
6.8 Smooth positively curved reference supports . . . 37

7 Critical Wasserstein rates for one-dimensional endpoint noise **39**
7.1 Model and rates . . . 39
7.2 One-sided probes and a weighted information bound . . . 40
7.3 Upper bounds and the noncritical lower bounds . . . 41
7.4 Critical lower bounds across resolutions . . . 42
7.4.1 Estimation at degree 1/2 . . . 43
7.4.2 Testing at degree 3/4 . . . 43
7.5 Fixed-scale gamma errors . . . 44
7.6 The critical lower bound for fixed ordinary-smooth errors . . . 44

8 Recovering joint noise from corners of product densities **45**
8.1 Model and results . . . 46
8.2 Tensorized observation bounds . . . 47
8.3 Upper bounds . . . 48
8.4 Single-resolution lower bounds . . . 48
8.5 Critical testing across scales . . . 49
8.6 Recovering a marginal from the full experiment . . . 50
8.7 The corner Poisson experiment . . . 50

9 Sharp Wasserstein rates under mixed smoothness at product corners **52**
9.1 Model and rate formulas . . . 53
9.2 Observation tools and the dyadic sums . . . 54
9.3 Upper bounds . . . 55
9.4 Multiresolution lower bounds . . . 55

9.4.1 Estimation lower bound . . . 56
9.4.2 Testing lower bound . . . 57
9.4.3 Modulus lower bound . . . 58

10 Information scales for mixed noise moments 58
10.1 Model and scalar information scales . . . 58
10.2 Uniform scalar comparison and its leading form . . . 59
10.3 Tensor information and mixed moments . . . 61
10.4 Variance constants for unbiased moment probes . . . 63
10.5 Gaussian blocks together with corner Poisson data . . . 64
10.6 Why the weighted norm uses an interior buffer . . . 66

11 Sampling transitions for minimum-variance mixed moment probes 66
11.1 The observation and the canonical probes . . . 66
11.2 Limit laws by depth and sample size . . . 67
11.3 Localization and the proof of Theorem 11.1 . . . 68
11.4 The logarithmic depth cutoff . . . 69
11.5 Empirical variance and studentization . . . 70
11.6 Joint limits of finitely many probes . . . 71

1 Introduction

Deconvolution usually treats the noise as known and the latent law as unknown. Here the roles are reversed: the latent law is known and has a boundary, and the problem is to recover a small unknown noise distribution from the observations near that boundary. Compact support is known to make deconvolution robust to an unknown error law: Meister [3] showed that only the error characteristic function near the origin matters for compactly supported signals, and Aarts, Groeneboom and Jongbloed [2] estimated the upper support point in a boxcar deconvolution model. The results here go in the opposite direction, using a known boundary to recover the noise itself with sharp constants and local limits. The Gaussian inverse problem for log-concave laws is treated in [1]. Sections 2–4 treat boxes and polytopes, Sections 5–7 curved and power-law boundaries and the one-dimensional critical rates, and Sections 8–11 product densities and their corners.

2 Normal and Poisson limits for noise at a box

We observe a known uniform box after small additive noise. When competing noise laws agree on all coordinate marginals of order below k and differ on some marginal of order k, the first Hellinger discrepancy comes from codimension-k faces and has order b^k, where b is the noise scale. At Nb^k of constant order the finite-parameter experiment converges to a point process with signed survival profiles as intensities. For Gaussian noise, coordinate variances and off-diagonal covariances have different local normal rates. At the off-diagonal critical scale, the limit is a product of a Gaussian shift for the local variances and a Poisson experiment for the correlations.

2.1 Results

Let $K = \prod_{i=1}^{d}[\ell_i, r_i] \subset \mathbb{R}^d$ be a fixed box with positive side lengths, and U uniform on K. For a noise law P on $\mathbb{R}^d$, let $q_{b,P}$ be the density of $U + bZ$ for independent $Z \sim P$. The noise may be singular. We use $H^2(p, q) = \int(\sqrt{p} - \sqrt{q})^2$. When $P = N(0, A)$, write $q_{b,A}$.

A codimension-k face F fixes a set of coordinates $J(F)$ of size k, each with outward sign $s_i = -1$ at a lower endpoint and $s_i = 1$ at an upper endpoint. In the corresponding order of coordinates put

$$S_{P,F}(t) = \mathbb{P}_P(s_i Z_i \geq t_i \text{ for all } i \in J(F)), \qquad t \in \mathbb{R}^k.$$

Write $|F|$ for its $(d-k)$-dimensional area, counting vertices when $k = d$.

Theorem 2.1. *Fix $1 \leq k \leq d$. Suppose P, Q have finite dth moments and the same joint coordinate marginal on every subset of fewer than k coordinates. Then*

$$\lim_{b \downarrow 0} b^{-k} H^2(q_{b,P}, q_{b,Q}) = \sum_F \frac{|F|}{|K|} \int_{\mathbb{R}^k} (\sqrt{S_{P,F}} - \sqrt{S_{Q,F}})^2 \, dt.$$

The coefficient is finite, and is positive exactly when some k-coordinate marginal differs.

For a finite parameter family $\{P_\theta\}$ with the same lower-coordinate marginals and finite dth moments, suppose $Nb^k \to \lambda \in (0, \infty)$. The experiments of N independent observations converge in Le Cam distance to observation of independent Poisson point processes, one per codimension-k face, with intensity

$$\frac{\lambda}{|K|} S_{P_\theta,F}(t) \, dS(z) \, dt \quad \text{on } F \times \mathbb{R}^k.$$

These processes are locally finite. The face positions z can be kept as marks or discarded as ancillary.

To describe the Gaussian results, let $\mathcal{I}_D(A)$ and $\mathcal{I}_O(A)$ be the limiting diagonal and off-diagonal Fisher matrices of [4]. Here the coordinates of a symmetric matrix use one entry for each pair $i < j$, so its direction has equal ij and ji entries. Both limiting matrices are positive definite and diagonal in the respective coordinate bases. For a diagonal shift h write $\operatorname{diag} h$, and for off-diagonal coordinates u let $O(u)$ be the corresponding symmetric matrix of zero diagonal.

Theorem 2.2. *Fix a positive definite covariance A. If $Nb^2 \to \infty$ as $b \downarrow 0$, the experiment is locally asymptotically normal under the local shifts*

$$A_{N,h,u} = A + \frac{\operatorname{diag} h}{\sqrt{Nb}} + \frac{O(u)}{\sqrt{Nb^2}}.$$

At the reference the limiting log likelihood ratios are

$$h^\top Z_D + u^\top Z_O - \tfrac{1}{2} h^\top \mathcal{I}_D(A) h - \tfrac{1}{2} u^\top \mathcal{I}_O(A) u,$$

where Z_D, Z_O are independent centered Gaussian vectors of covariances $\mathcal{I}_D(A), \mathcal{I}_O(A)$. The convergence holds jointly for every finite set of local shifts.

Theorem 2.3. *Fix a positive diagonal and a finite family of positive definite matrices A_θ with that diagonal. Suppose $Nb^2 \to \lambda \in (0, \infty)$. For any finite set of parameters (h, θ), the experiments with covariance*

$$A_\theta + \frac{\operatorname{diag} h}{\sqrt{Nb}}$$

converge in Le Cam distance to the independent product of the following observations. The first is Gaussian with mean $\mathcal{I}_D h$ and covariance $\mathcal{I}_D$, where $\mathcal{I}_D$ depends only on the common diagonal. The second is the codimension-two Poisson experiment of Theorem 2.1 with $P_\theta = N(0, A_\theta)$.

The scalar and pair profile coefficients for Gaussian noise were obtained in [4], which also compares the local boundary analysis of Brosse and Dalalyan [5]. The all-order marginal comparison here comes from the product structure of a box. The Poisson limits concern rare observations close to its boundary strata. Related nonregular Poisson limit experiments arise in regression with endpoint errors [6]. The normal limits use quadratic mean expansions, as in [7], with separate scales for the two covariance blocks.

2.2 Localization by matching coordinate marginals

Write $a_i(y) = \mathbf{1}_{y_i \in [\ell_i, r_i]}$ and

$$\xi_i(y, Z) = \mathbf{1}_{y_i - bZ_i \in [\ell_i, r_i]} - a_i(y).$$

Then

$$|K| q_{b,P}(y) = \sum_{J \subseteq \{1,\dots,d\}} \left(\prod_{i \notin J} a_i(y) \right) \mathbb{E}_P \prod_{i \in J} \xi_i(y, Z).$$

In the difference between P and Q, all terms with $|J| < k$ cancel. For a fixed displacement, the support of ξ_i in y_i has length at most $2b|Z_i|$, and is within that coordinate's $b|Z_i|$ neighborhood of its endpoints. Thus a term with $|J| > k$ has integrated absolute size $O(b^{|J|})$, by the product length bound and the assumed moments. This is $O(b^{k+1})$ after summing those sets.

For $|J| = k$, every unaffected coordinate has to lie in its interval. If an affected point is farther than Tb from the corresponding union of codimension-k faces, then at least one affected $|Z_i|$ exceeds $T/\sqrt{k}$. The integrated absolute value of that term outside the tube is bounded by

$$Cb^k \mathbb{E}\left[\prod_{i \in J} |Z_i| \mathbf{1}_{\max_{i \in J} |Z_i| > T/\sqrt{k}} \right] = b^k o_T(1).$$

Here $o_T(1)$ tends to zero as T tends to infinity, uniformly over any fixed finite family, and the bound holds for either noise law. Since H^2 at any point is bounded by the absolute density difference, these estimates localize Hellinger discrepancy to the codimension-k tubes. The same localization holds uniformly for a family with uniformly integrable moments of order d.

In the tube at F write

$$y = z + b \sum_{i \in J(F)} s_i t_i e_i, \qquad z \in F.$$

For z in the relative interior of F and fixed t, the latent box constraints apart from the selected endpoints disappear in the limit. Thus

$$q_{b,P}(y) \longrightarrow |K|^{-1} S_{P,F}(t)$$

for almost every t. The only threshold exceptions are coordinate values at atoms, which form a null set in $\mathbb{R}^k$. The Jacobian for this parameterization is b^k. For bounded t, tubes based on distinct face interiors can be made disjoint by omitting tangential strips of width $O(Tb)$ near their boundaries. The portion of the Euclidean neighborhood missed by these trimmed rectangles, and the overlap among untrimmed rectangles, have volume $O_T(b^{k+1})$. In fact at least $k+1$ distinct coordinates must then be within $O(Tb)$ of a corresponding endpoint. When $k = d$, the vertex rectangles are disjoint for small b and have no spill. All densities are bounded by $|K|^{-1}$. Dominated convergence on bounded tubes and the tail localization give the Hellinger limit in Theorem 2.1. The intensity profile is nonnegative, and its survival function determines the signed coordinate law, including in the presence of atoms. More precisely almost-everywhere equality implies equality at all continuity thresholds and hence equality of the laws. Therefore the coefficient is positive exactly under the stated marginal condition.

Every order k can arise. For $k \geq 2$, take k independent uniform sign coordinates in one law and condition their product to be one in the other. Every proper subset of the k coordinates remains uniform. Take any other coordinates independent and common. The full k-coordinate laws differ, and adding independent common Gaussian noise to the sign coordinates also gives continuous examples: Gaussian convolution is injective by its nonzero characteristic function. At order one, a change in one coordinate distribution already gives such an example.

2.3 The critical Poisson experiment

We give the argument in total variation, using Markov kernels in both directions. The Le Cam distance between experiments is the larger of the two directed infima, over such kernels, of the maximum total variation error across parameters. Our parameter set in this section is finite.

For fixed T retain only the observations in the disjoint trimmed rectangles at the codimension-k faces, and express them in the marks (F, z, t), with $t \in [-T, T]^k$. The total selection probability is $p = O_T(b^k)$ under every parameter. The retained count is binomial $\mathrm{Bin}(N, p)$; given the count, marks are independent with the selected density. Replacing the count by $\mathrm{Pois}(Np)$ costs at most $CNp^2 = o(1)$ in total variation by the elementary Poisson approximation for Bernoulli sums. The marks then form a Poisson process with intensity equal to N times the retained mark density. Its intensity converges in L^1 on the bounded mark region to the profile intensity stated in Theorem 2.1, by dominated convergence and $Nb^k \to \lambda$. Poisson processes of finite intensities can be coupled at total variation cost at most the L^1 intensity difference, by sharing their common intensity and adding independent surplus points. This gives convergence of the cropped experiments.

For recovery of discarded observations, fix one reference parameter θ_0. Conditionally on the cropped count and marks, the unretained observations are independent from the original density restricted and normalized on the complement, and the original order is uniformly random. Draw those unretained observations instead from the normalized complement at θ_0, and restore a uniformly random order. If the two complement masses are at least one half, the squared Hellinger distance between the normalized restrictions is bounded by a fixed constant times the squared difference of their unnormalized square roots. Indeed, the same difference controls the difference of their L^2 norms, and normalization then follows by the triangle inequality. The localization bound and the spill estimate make the unnormalized integral $b^k o_T(1) + O_T(b^{k+1})$. The product Hellinger bound for at most N unretained points makes the recovery error $o_T(1) + o_b(1)$ in total variation. A rare Poisson count above N can be mapped arbitrarily, since the count approximation already controls its probability.

It remains to remove cropping from the limit. For a face and $z \sim P_\theta$, write

$$\eta_i(t, z) = \mathbf{1}_{s_i z_i \geq t_i} - \mathbf{1}_{t_i < 0}.$$

In the difference of the two survival profiles only the full product of the k factors η_i survives the same marginal cancellation. Each absolute factor is supported on an interval of length $|z_i|$. Outside the coordinate mark cube $[-T, T]^k$, any nonzero product requires $\max_i |z_i| > T$. Hence the integrated absolute difference of the two survival profiles outside the cube tends to zero uniformly over the finite parameter family. The same holds for the squared difference of square roots.

The limiting Poisson intensity is locally finite, though its total mass may be infinite. To recover the uncropped process from its crop, add independent reference-intensity points in the omitted region. For Poisson laws on a finite window, their Hellinger affinity is $\exp(-\frac{1}{2}\int(\sqrt{\mu} - \sqrt{\nu})^2)$, where μ, ν are the intensity densities. This follows by summing the Poisson likelihood series, or by splitting into their counts and locations. Finite-window bounds extend to a countable exhaustion and give total variation cost $o_T(1)$ here. Taking $b \downarrow 0$ at fixed T and then $T \to \infty$ proves the experiment limit. The positions along each face are uniform and common conditional on its other marks, so they can be dropped or generated by a parameter-independent kernel.

2.4 Moment bounds for covariance derivatives

Restrict Gaussian covariance parameters to a compact subset of the positive definite cone. Let $D^\alpha q$ be a derivative of positive order $r = |\alpha|$ in fixed covariance entries. By the Gaussian Fourier

multiplier it is, up to a constant, b^{2r} times a spatial derivative of order $2r$, with the two index occurrences for each covariance entry. Let j be the number of distinct spatial indices that occur. Then for each fixed $p \geq 1$,

$$\int |D^{\alpha} q|^{p} / q^{p-1} \leq C_{\alpha,p} b^{j}. \tag{1}$$

The bound is uniform over the covariance set. Its restriction away from the Tb tube of the corresponding codimension-j faces has the extra factor $o_T(1)$.

To see this, move one derivative in each distinct spatial index onto the box indicator. The result is a finite signed sum of measures on codimension-j faces convolved with Gaussian derivatives of order $2r - j$. After the b^{2r} factor, the integrand for a point z of a face is bounded in absolute value by $b^{j}\gamma_{b,A}(y - z)$ times a fixed polynomial in $1 + |(y - z)/b|$. For that face use the inner collar of length cb in each of its j inward directions, as in the pair estimate of [4]. Its kernel $K_z(y)$ satisfies

$$q(y) \geq |K|^{-1} \int_F K_z(y)\, dS(z), \qquad K_z(y) \geq c b^{j} \gamma_{b,A}(y - z)/(1 + |(y - z)/b|)^{j}.$$

The last bound follows by integrating only over inward displacements of size $cb/(1 + |(y - z)/b|)$ in each coordinate. Hölder's inequality on the face bounds the pth power of its convolution divided by q^{p-1} by a constant times $\int_F |a_z|^{p}/K_z^{p-1}\, dS(z)$, where a_z is the scaled Gaussian derivative. This is at most b^{j} times the integral of a polynomial times $\gamma_{b,A}(y - z)$. Integrating over y proves (1), and its tail version by the Gaussian decay.

A single diagonal score has moment of order b for each fixed p, while a single off-diagonal score has moment of order b^2. The same bound $O(b^2)$ holds for a second density derivative ratio with at least one off-diagonal coordinate, since it touches at least two distinct indices. For the square-root densities, use

$$\partial_{ij}\sqrt{q} = \frac{\partial_{ij} q}{2\sqrt{q}} - \frac{\partial_i q\, \partial_j q}{4 q^{3/2}},$$

where the derivatives here are in covariance parameters. The bounds with $p = 2, 4$ and Cauchy–Schwarz give

$$\|\partial_{ij}\sqrt{q}\|_2 \leq \begin{cases} C b^{1/2}, & i, j \text{ both diagonal coordinates,} \\ C b^{3/4}, & \text{one diagonal and one off-diagonal,} \\ C b, & i, j \text{ both off-diagonal.} \end{cases} \tag{2}$$

The first and second parameter derivatives can be integrated along a covariance segment using these bounds. Pointwise they are smooth, since the Gaussian covariance stays positive definite.

2.5 The two local normal scales

Let S_D, S_O denote the single-observation score vectors at the reference covariance, in the two coordinate blocks. Their rescaled Fisher blocks converge to $\mathcal{I}_D(A)$ and $\mathcal{I}_O(A)$ by the face and edge limits in [4]. The cross block, divided by $b^{3/2}$, tends to zero. For a fixed Tb edge tube the diagonal score energy is $O_T(b^2)$: the derivative is bounded by the Gaussian polynomial formula, the density is bounded below there, and the tube volume is $O_T(b^2)$. On that tube pair a diagonal score with an off-diagonal score using Cauchy–Schwarz. This gives $O_T(b^2)$. On the complement the off-diagonal energy is $b^2 o_T(1)$ by (1), and the total diagonal energy is $O(b)$. Divide their Cauchy–Schwarz bound by $b^{3/2}$ and send T to infinity.

The central sequence is

$$\frac{1}{\sqrt{N}} \sum_{m=1}^{N} \left(\frac{S_D(Y_m)}{\sqrt{b}}, \frac{S_O(Y_m)}{b} \right).$$

The scores have zero means by differentiation of the total mass. The covariance limits have just been identified. By (1) with $p = 3$, the Lyapunov quantity for each fixed linear combination of the displayed vector is $O(N^{-1/2}b^{-1}) = o(1)$. Thus the limit is the Gaussian pair asserted in Theorem 2.2.

We verify the likelihood expansion because the observation density changes with b. For the stated shift let $r_N(y) = \sqrt{q_{\mathrm{alt}}(y)/q(y)} - 1$. Under the reference, its first-order approximation is

$$v_N(y) = \tfrac{1}{2}\left(\frac{h^\top S_D(y)}{\sqrt{Nb}} + \frac{u^\top S_O(y)}{\sqrt{Nb^2}}\right).$$

The three second-order terms in the square-root Taylor bound (2) give

$$\sqrt{N\mathbb{E}[(r_N - v_N)^2]} \le C_{h,u}\left((Nb)^{-1/2} + (Nb^{3/2})^{-1/2} + (Nb^2)^{-1/2}\right) = o(1).$$

In addition, $N\mathbb{E}v_N^4 = o(1)$ by the fourth moment bounds: the diagonal block contributes $O((Nb)^{-1})$ and the off-diagonal block $O((Nb^2)^{-1})$. It follows that the maximum absolute r_N over the sample tends to zero in probability, the sum of their squares differs from $N\mathbb{E}r_N^2$ by $o_P(1)$, and the sum of centered remainders $r_N - v_N$ is $o_P(1)$. For the sum of squares first use the fourth moment bound for v_N, then Cauchy–Schwarz and the squared remainder bound for the replacement by r_N. Normalization gives $2\mathbb{E}r_N = -\mathbb{E}r_N^2$. Hence the log likelihood ratio is

$$2\sum_{m=1}^{N}\log(1 + r_N(Y_m)) = 2\sum_{m=1}^{N} v_N(Y_m) - 2N\mathbb{E}r_N^2 + o_P(1).$$

The remainder gives $N\mathbb{E}r_N^2 = N\mathbb{E}v_N^2 + o(1)$. The Fisher limits now identify the quadratic term, and the central sequence gives the random linear term. This proves the joint finite-shift statement of Theorem 2.2.

2.6 A Gaussian–Poisson product at the edge scale

Fix T and split the sample into observations in the trimmed edge tubes and those outside. The split is equivalent to the original ordered sample after an independent uniform restoration of the ordering. For covariance $A_\theta + \operatorname{diag} h/\sqrt{Nb}$, the edge count has bounded mean because the selection probability is $O_T(b^2)$. The rare mark experiment converges to its cropped Poisson limit with intensity depending on θ alone. The proof from Theorem 2.1 applies: the varying diagonal shift tends to zero, so at every fixed relative interior point and bounded mark the limiting Gaussian quadrant profile is the one for A_θ. The densities are uniformly bounded, and the finitely many shifted shapes give the same intensity convergence and Poisson approximation.

For the unretained observations fix a reference off-diagonal index θ_0 and retain the actual local diagonal shift h. The complement density at (h, θ) can be replaced, conditionally on the retained count, by the normalized complement density at (h, θ_0). For each fixed h the two covariance matrices have the same diagonal. Their Gaussian noises therefore have the same one-coordinate marginals. The order-two localization in the proof of Theorem 2.1 is uniform over these varying matrices, whose moments stay uniformly integrable. Off the trimmed tube the squared Hellinger discrepancy is at most $b^2 o_T(1) + O_T(b^3)$. Normalize the restrictions as before and apply the product bound over at most N discarded observations. Since Nb^2 is bounded, this replacement has total variation cost $o_T(1) + o_b(1)$. The replaced conditional experiment depends only on h.

Now consider the probability law of a single unretained observation at (h, θ_0), with its normalization on the complement. Its face-scale local normal expansion is the same as without the

restriction. Indeed, the probability of the deleted edge tube is $O_T(b^2)$, and its first and second diagonal covariance derivatives are $O_T(b^2)$ by the uniform Gaussian polynomial bounds on the tube. A conditional score differs from the unrestricted score by the derivative of the log normalizer, which is $O_T(b^2)$. Deleting the tube changes the unrestricted diagonal score energy by only $O_T(b^2)$, because its density is bounded below there and its probability derivatives are uniformly bounded. The conditional score moments and second square-root derivatives thus retain the orders $O(b)$ and $O(\sqrt{b})$ used above for the diagonal block. Apply the diagonal part of the likelihood proof, now requiring only $Nb \to \infty$, which follows from $Nb^2 \to \lambda > 0$. The information limit is $\mathcal{I}_D$, whose face formula depends only on the diagonal of the covariance.

Conditionally on the number M of retained observations, the replaced sample outside consists of $N - M$ independent draws from this conditional face-scale law, independent of the retained marks. Append M draws from the fixed conditional reference at $h = 0$, on the same complement. Using draws at the true h instead would give N independent face observations with a law independent of the retained experiment. Changing the appended draws has squared Hellinger cost at most $C_h M/N$. To see the per-draw estimate, integrate the square-root derivative along the local diagonal shift and use its $O(\sqrt{b})$ norm. Since $\mathbb{E}M = O_T(1)$, the replacement changes the joint experiments by $o(1)$ in total variation. In the other direction, given the rare mark experiment and N independent conditional face observations, discard the required M face observations before reconstructing the ordered sample. Thus for fixed T we have separated the two experiments, up to the stated tail error.

For completeness, the diagonal likelihood expansion gives convergence of these finite parameter experiments in Le Cam distance. Under the conditional reference, the finite vector of likelihood ratios converges jointly to the Gaussian shift likelihood vector. Its entries have mean one in the limit, as they do before the limit, so they are uniformly integrable. The likelihood vector is a sufficient statistic for a family dominated by a common reference. Quantize it on a large bounded coordinate box into a finite partition with limit-null boundaries. Under every parameter the bin probabilities converge by weak convergence and uniform integrability. The experiment of observing the bins approximates the original vector experiment uniformly: regenerate the vector inside its bin using the reference conditional law, and bound the change in each alternative by the within-bin variation of its likelihood ratio. This variation is small on the bounded box by fine quantization, and the omitted contribution is uniformly small by uniform integrability. This gives the finite-parameter experiment convergence by letting the quantization error tend to zero.

The rare mark experiment converges to the cropped edge Poisson experiment independently of this Gaussian shift. Remove the crop by adding independent reference Poisson points as in the proof of Theorem 2.1. The tail error vanishes as $T \to \infty$. The result is the independent product claimed in Theorem 2.3.

3 Tangent-cone experiments for additive noise on polytopes

For a known uniform polytope observed with small additive noise, the first Hellinger discrepancy is carried by the lowest-codimension faces at which the normal projections of the noise laws differ. At the critical sample size the experiment converges to Poisson processes with tangent-cone intensities. The result includes nonsimple polytopes and singular noise. Every codimension can arise for smooth strongly log-concave noise laws of equal covariance. For Gaussian covariance changes only facets and ridges are needed, and the exact ridge Fisher information is given by bivariate profiles weighted by the angles between adjacent facet normals.

3.1 Results

Let K be a full-dimensional compact polytope in $\mathbb{R}^d$, $d \geq 2$, and let U be uniform on K. The observation is $U + bZ$, with independent noise $Z \sim P$ and scale $b > 0$. Write $q_{b,P}$ for its density and use $H^2(p,q) = \int(\sqrt{p} - \sqrt{q})^2$. The noise law can be singular.

For a face F let L_F be its linear normal space, the orthogonal complement of its tangent directions, and let π_F denote orthogonal projection to L_F. If F has codimension k, L_F has dimension k. The inward cone in that space is

$$C_F = \{t \in L_F : \ n_i \cdot t \leq 0 \text{ for every facet } i \text{ containing } F\},$$

where n_i is an outward unit normal. It is pointed and full-dimensional in L_F. Define the profile

$$S_{P,F}(t) = \mathbb{P}_P(t - \pi_F Z \in C_F), \qquad t \in L_F.$$

Let $|F|$ be relative area, with counting measure at the vertices.

Theorem 3.1. *Fix $1 \leq k \leq d$. Suppose P, Q have equal projections to L_F at every face of codimension less than k, and their signed difference $\mu = P - Q$ satisfies $\int |z|^k \, d|\mu|(z) < \infty$. Then*

$$\lim_{b \downarrow 0} b^{-k} H^2(q_{b,P}, q_{b,Q}) = \sum_{\operatorname{codim} F = k} \frac{|F|}{|K|} \int_{L_F} (\sqrt{S_{P,F}} - \sqrt{S_{Q,F}})^2 \, dt.$$

The coefficient is finite and is positive exactly when some codimension-k normal projection differs.

For a finite family of noise laws satisfying the same lower-codimension agreements and this moment condition pairwise, if $Nb^k \to \lambda \in (0, \infty)$, the sample experiments converge in Le Cam distance to independent Poisson point processes on the faces of codimension k, with intensities

$$\frac{\lambda}{|K|} S_{P,F}(t) \, dS(z) \, dt \quad \text{on } F \times L_F.$$

Proposition 3.2. *For every K and each $1 \leq k \leq d$ there are two smooth strongly log-concave noise laws of mean zero and covariance identity whose first differing normal projections occur at codimension k.*

For Gaussian noise write $q_{b,A} = q_{b,N(0,A)}$. A ridge F is a face of codimension two and has exactly two incident facets, with outward normals n_i, n_j. Set $s_F = (1 - (n_i \cdot n_j)^2)^{1/2}$ and let A_F be the two-by-two covariance of $(n_i \cdot Z, n_j \cdot Z)$ under $Z \sim N(0, A)$. For a positive definite two-by-two matrix M, write φ_M for its Gaussian density and $S_M(t) = \mathbb{P}(Z_1 \geq t_1, Z_2 \geq t_2)$. Let $I_b(A; E)$ denote the single-observation Fisher information for the scalar covariance direction E.

Theorem 3.3. *If A, C are distinct positive semidefinite matrices, the leading squared Hellinger order for $q_{b,A}, q_{b,C}$ is b when some facet normal variance differs and b^2 otherwise. When the facet variances agree, the coefficient of b^2 is given by the ridge formula in Theorem 3.1; in the two outward threshold coordinates its summand is*

$$\frac{|F|}{|K| s_F} \int_{\mathbb{R}^2} (\sqrt{S_{A_F}} - \sqrt{S_{C_F}})^2 \, dt,$$

using the survival definition for possibly singular pairs.

For positive definite A, if the symmetric direction E obeys $n_i^\top E n_i = 0$ on every facet, then

$$\lim_{b \downarrow 0} b^{-2} I_b(A; E) = \sum_{F \text{ ridge}} \frac{|F|}{|K| s_F} (n_i^\top E n_j)^2 \int_{\mathbb{R}^2} \frac{\varphi_{A_F}(t)^2}{S_{A_F}(t)} \, dt.$$

This form is positive for every nonzero such direction.

The box case uses coordinate-marginal comparisons and the two Gaussian scales of [4] and Section 2. For a general polytope the active inequalities at a face need not be independent. The order of their localized contribution is the codimension of the face they meet, controlled by the polyhedral error bound of Hoffman [8]. Brosse and Dalalyan [5] describe local tangent-cone profiles for scalar Gaussian smoothing. The present results compare integrated discrepancies and experiments between different noise laws.

3.2 Localization by the active facets

Translate the origin into the interior of K. Use an irredundant facet description

$$K = \{y : n_i \cdot y \le h_i,\ 1 \le i \le M\}$$

with unit normals and positive offsets. The normal span at a face is spanned by the normals of its active facets. Thus the inward cone C_F is pointed, and the projection of a direction pointing from the face into the interior gives a strictly interior point of that cone.

Write $a_i(y) = \mathbf{1}_{n_i \cdot y \le h_i}$ and

$$\xi_i(y, z) = \mathbf{1}_{n_i \cdot (y - bz) \le h_i} - a_i(y).$$

Expand the product of $a_i + \xi_i$ over the facets and integrate against the signed difference $\mu = P - Q$. For a subset J of affected facets let F_J be their intersection within K, with $F_\varnothing = K$. For a nonempty F_J, the term depends on the noise only through its projection to L_{F_J}. Consequently terms with $\operatorname{codim} F_J < k$ cancel exactly.

Here is the geometric bound for the remaining terms. For a given displacement put $\rho = b|z|$. If the corresponding product is nonzero at y, the unaffected inequalities hold, and each affected signed slack has absolute value at most ρ. If $F_J \neq \varnothing$, the polyhedral error bound gives

$$\operatorname{dist}(y, F_J) \le C_J \rho.$$

This is Hoffman's bound [8] for the inequalities of K together with the equalities of J, regarded as paired inequalities. One can see the bound directly from the closest point x to y in that feasible polyhedron. The displacement $y - x$ is a nonnegative linear combination of linearly independent active constraint normals, by the normal cone description and elimination of dependent generators. The sum of the coefficients is at most $C|y - x|$, since there are only finitely many independent subsets. Take the inner product with $y - x$. Each constraint gives a contribution at most its positive violation at y, which is $O(\rho)$. Division by $|y - x|$ proves the bound.

If $F_J = \varnothing$, the simultaneous constraints cannot hold for arbitrarily small ρ. Indeed all the nearly feasible points lie in $\{y : n_i \cdot y \le h_i + \rho\}$, which is bounded for bounded ρ because it is contained in $(1 + \rho/\min_i h_i)K$. Otherwise a limiting point would belong to F_J. Fix a sufficiently small $\epsilon > 0$ so that empty-intersection terms vanish at displacements of length at most ϵ. Work first inside the bounded inflated polytope $K_\epsilon = \{y : n_i \cdot y \le h_i + \epsilon\}$. Perform the projection cancellations in the full signed integrals before splitting the surviving terms at $b|z| = \epsilon$. The large-displacement part of every surviving term on K_ϵ has integrated absolute size at most $|K_\epsilon|\,|\mu|(\{|z| > \epsilon/b\}) = o(b^k)$. Outside K_ϵ use the original signed mixture of translated uniform laws. A shift of size at most ϵ cannot contribute, so the integrated absolute difference is at most $|\mu|(\{|z| > \epsilon/b\}) = o(b^k)$ there as well.

For small $\rho = b|z|$, the $C_J\rho$ neighborhood of a fixed face of codimension j has volume $O(\rho^j)$. For $j > k$, dominated convergence against $|z|^k\, d|\mu|(z)$ shows that

$$b^k \int |z|^k (b|z|)^{j-k} \mathbf{1}_{b|z| \le \epsilon}\, d|\mu|(z) = o(b^k).$$

For $j = k$, the support bound on points at distance more than Tb from the union of codimension-k faces requires $|z| > T/C_J$. Its contribution is at most

$$Cb^k \int_{|z|>T/C_J} |z|^k \, d|\mu|(z) + o(b^k).$$

This localizes the absolute density difference to those face tubes, with error $b^k o_T(1) + o(b^k)$. The same bound holds for squared Hellinger distance, since $(\sqrt{p} - \sqrt{q})^2 \le |p - q|$. The bounds are uniform for a fixed finite noise family and apply uniformly to varying laws whenever their absolute signed kth moments are uniformly integrable.

3.3 Profiles and the critical experiment

For each face F of codimension k, use the orthogonal coordinates

$$y = z + bt, \qquad z \in F, \quad t \in L_F, \quad |t| \le T.$$

The volume element is $b^k \, dS(z) \, dt$. For fixed T these tubes can be made disjoint by discarding a tangential neighborhood of their boundaries of width $O(Tb)$. The discarded and overlapping regions have volume $O_T(b^{k+1})$, including any part of the Euclidean tube not covered by the normal coordinate tubes. In fact a point whose orthogonal projection to the affine face misses that face is near its relative boundary if the point is within Tb of the face. If two different faces intersect, any point within Tb of both is within $O(Tb)$ of their intersection by the same linear error bound applied to the two face systems. The intersection has codimension at least $k + 1$. Disjoint faces have a fixed positive separation. For vertices the tubes are disjoint without tangential trimming.

At a fixed relative interior point $z \in F$, the inequalities not active at z have positive slack and disappear in the scaling. Thus for almost every $t \in L_F$,

$$q_{b,P}(z + bt) \longrightarrow |K|^{-1} S_{P,F}(t).$$

Each active threshold at an atom of its one-dimensional projection is a null hyperplane in L_F. Outside these hyperplanes the claimed convergence follows by bounded convergence under the fixed noise law. The densities are all bounded by $|K|^{-1}$. Dominated convergence on bounded tubes, the negligible trimming error, and the localization from the previous section prove the asserted Hellinger coefficient.

When $Nb^k \to \lambda$, retain only the observations in the disjoint trimmed tubes and rescale their marks. The selection probability is $O_T(b^k)$. Replacing the binomial count by its Poisson approximation costs at most $CNb^{2k} = o(1)$ in total variation. Conditional on the count the marks are independent, so this gives an intensity equal to N times the retained mark density. By the profile convergence that intensity converges in L^1 on the bounded tubes to the stated intensity. A common-intensity coupling gives total variation convergence of the cropped point process experiments.

To reconstruct discarded observations, draw from the normalized tube complement of one fixed reference law and restore the sample order uniformly. The squared Hellinger distance between the normalized complement laws is bounded by a constant times the unnormalized squared square-root discrepancy, since their masses tend to one. The localization bound is $b^k o_T(1) + o_b(b^k)$, and the product bound for at most N draws gives total variation cost $o_T(1) + o_b(1)$. Thus the cropped experiment and the full data are asymptotically equivalent after letting T grow, as in the box case Section 2.

To uncrop the Poisson limit, the profile tails must likewise have small discrepancy. One can read this directly from the cone rules. In L_F, expand the active indicators around $\mathbf{1}_{n_i \cdot t \le 0}$. For each

subset J of the active facets, its intersection $G = F_J$ contains F. The term cancels if G has smaller codimension. Otherwise $G = F$, and

$$\{t \in C_F : n_i \cdot t = 0 \ \forall i \in J\} = \{0\}.$$

Indeed a nonzero t there would give, for small $\epsilon > 0$, a point $z + \epsilon t$ in $G = F$, contradicting $t \in L_F$. The homogeneous error bound now places the support of the product within $|t| \le C|\pi_F Z|$. Thus the integrated absolute difference of profiles past $|t| > T$ tends to zero, bounded by a fixed multiple of the tail kth moment of $|\mu|$, since $|\pi_F \mu| \le \pi_F |\mu|$. Adding independent reference-intensity Poisson points outside the truncated process has total variation cost tending to zero by the Poisson square-root-intensity affinity, first on finite windows and then by exhaustion. This proves the Le Cam limit in Theorem 3.1.

3.4 Cone injectivity and examples at every order

Let C be a pointed full-dimensional polyhedral cone in a finite-dimensional Euclidean space. Convolution with $\mathbf{1}_C$ is injective on finite signed measures, in the sense that $\mathbf{1}_C * \mu = 0$ almost everywhere implies $\mu = 0$.

To prove this, triangulate a compact section of C by a hyperplane strictly transverse to all its rays. This gives a finite decomposition into full-dimensional simplicial cones with disjoint interiors. Let $v_1, \dots, v_m$ list the generating rays used. On a simplicial cone with generators $v_{i_1}, \dots, v_{i_k}$, differentiating its indicator in these k directions gives $|\det(v_{i_1}, \dots, v_{i_k})|\delta_0$. This follows by changing to the coordinates of the generating basis and integrating by parts in each positive coordinate. Thus applying the product of derivatives $D_{v_1} \cdots D_{v_m}$ to $\mathbf{1}_C$ gives $P(D)\delta_0$, where P is a nonzero homogeneous polynomial. To see nonvanishing, evaluate its symbol at a vector strictly positive on every generating ray. Every term then has positive sign and a positive determinant coefficient. This derivative identity is the distributional form of the usual simplicial-cone transform calculation [9].

The derivatives commute with convolution against a finite measure, by applying them to a compactly supported test and integrating translations. Hence $\mathbf{1}_C * \mu = 0$ implies $P(D)\mu = 0$. Taking the Fourier transform gives $P(i\xi)\widehat{\mu}(\xi) = 0$. The transform of a finite measure is continuous, and the nonzero polynomial has zero set with empty interior. Therefore $\widehat{\mu} = 0$ everywhere and $\mu = 0$, proving the claim.

Apply the claim in L_F to $\pi_F P - \pi_F Q$. Any differing normal projection gives a nonzero profile discrepancy on a set of positive measure. Conversely equal projections give equal profiles. This proves the positivity assertion in Theorem 3.1.

The use of a polynomial multiplier to construct laws with matching projections has a precedent in the sharp Cramér–Wold theorem of Cuesta-Albertos, Fraiman and Ransford [10]. Here the smooth perturbation also preserves strong log-concavity and the first two moments. For Proposition 3.2, fix one codimension-k face F. For each normal space L_G of strictly lower dimension choose a linear functional $\ell_G(\xi)$ that vanishes on L_G but is nonzero as a functional on L_F. This is possible because L_F is not contained in L_G. Include the space $\{0\}$ of $G = K$ in the list. The product polynomial

$$P_0(\xi) = \prod_G \ell_G(\xi)$$

vanishes on all the lower normal spaces and is not identically zero on L_F. Choose a real compactly supported smooth bump ψ with nonzero integral. Apply the corresponding real directional derivatives to $\Delta^2\psi$, obtaining a real compactly supported smooth function h. Its Fourier transform is, up to a constant phase, $P_0(\xi)|\xi|^4\widehat{\psi}(\xi)$. This vanishes on all lower normal spaces, and is nonzero at some point of L_F near the origin. It also vanishes through ordinary moment degree two at

the origin. Thus its signed measure has zero mass, first and second moments, zero lower normal projections, and a nonzero projection to L_F.

Let γ be the standard Gaussian density. For small positive ϵ, the densities $\gamma \pm \epsilon h$ are positive and normalized with mean zero and covariance identity. Their negative log densities have Hessian at least $I_d/2$, since h/γ is smooth and compactly supported and the C^2 norm of $\log(1 \pm \epsilon h/\gamma)$ is small. They realize the asserted first differing codimension, proving Proposition 3.2.

3.5 Gaussian covariance at the ridges

Suppose A, C have matching facet normal variances. Write $E = C - A$. If their projected covariances also agree at every ridge, then $n_i^\top E n_j = 0$ across every adjacent facet pair. We show $E = 0$. The vertices of the polar polytope are the vectors n_i/h_i. Its edges are dual to the ridges of K. The quadratic $x^\top E x$ vanishes on all such edges by the endpoint and bilinear identities. A polynomial of degree at most two that vanishes on every edge of a convex polygon vanishes on its affine plane, since at least three distinct affine linear edge equations must divide it on that plane. It therefore vanishes on every two-dimensional face of the polar. Inductively restrict the polynomial to a higher-dimensional face. It vanishes on the relative interior of each of its facets, hence on their affine hyperplanes. These hyperplane equations are distinct, and there are at least three, so again their product cannot divide a nonzero quadratic. At the full-dimensional polar this implies that the polynomial is identically zero. Therefore $E = 0$.

A Gaussian normal projection is determined by its covariance, even when singular. Thus Theorem 3.1 has first order $k = 1$ if any facet variance differs, and $k = 2$ for every remaining distinct pair. At a ridge with normals n_i, n_j, pass from orthonormal coordinates of L_F to $t = (n_i \cdot u, n_j \cdot u)$. The Jacobian determinant in absolute value is s_F. The cone profile becomes the bivariate Gaussian survival function in these threshold variables, and $du = dt/s_F$. This proves the Hellinger assertions of Theorem 3.3.

3.5.1 The ridge Fisher form

For a symmetric direction E with zero facet quadratic forms, the distributional derivative of the box indicator calculation extends as follows:

$$E : \nabla^2 \mathbf{1}_K = \sum_{F \text{ ridge}} \frac{2 n_i^\top E n_j}{s_F} \, dS_F, \tag{3}$$

where dS_F is relative area measure. Indeed the distributional gradient of $\mathbf{1}_K$ is the sum of $-n_i \, dS_i$ over the facets. Testing the second derivative against a smooth function ψ gives $\sum_i \int_{F_i} (E n_i) \cdot \nabla \psi \, dS_i$. The vector $E n_i$ is tangent to F_i by the zero-normal assumption. Apply the tangential divergence theorem on that facet. At a ridge with incident normals n_i, n_j, its outward conormal in F_i is $(n_j - (n_i \cdot n_j) n_i)/s_F$. This yields $n_i^\top E n_j / s_F$ from each incident facet, proving (3).

For the needed localization, an inner collar can be constructed at any fixed face F of codimension j. Choose j interior points $a_1, \dots, a_j$ whose normal projections, relative to the affine plane of F, are linearly independent in L_F. The projections of the interior contain an open set in L_F, so such a choice is possible. The map

$$(z, u) \longmapsto (1 - \textstyle\sum_i u_i) z + \sum_i u_i a_i, \qquad z \in F, \quad 0 \le u_i \le cb$$

lies in K by convexity for small b. The normal coordinates recover u uniquely, and then the tangential coordinates recover z. Its Jacobian is bounded below by a positive constant, from the

independent normal projections and the tangential factor $(1-\sum_i u_i)^{d-j}$. Hence the Gaussian smoothed density dominates a constant times the integral over F of the collar kernels. The kernel at z is bounded below by

$$cb^j \gamma_{b,A}(y-z)/(1+|(y-z)/b|)^j,$$

by restricting the u_i to the shorter intervals $[0, cb/(1+|(y-z)/b|)]$. The shifts in the Gaussian exponent are then bounded uniformly on compact positive definite covariance sets.

Now the covariance derivative of $q_{b,A}$ in direction E is $\frac{1}{2}b^2(E:\nabla^2 q_{b,A})$. Use (3) and the collar with $j=2$. Cauchy–Schwarz after division by $q_{b,A}$ bounds its squared norm in the Fisher integral by $O(b^2)$, and its contribution outside the Tb ridge tube by $b^2 o_T(1)$. Explicitly, a ridge Gaussian convolution squared over the density is bounded by

$$Cb^{-2}\int_F (1+|(y-z)/b|)^2 \gamma_{b,A}(y-z)\, dS_F(z),$$

before the squared factor b^4 from the covariance derivative. Integrate in y; on the omitted region the scaled Gaussian distance exceeds T.

At a fixed relative interior point of the ridge let $y = z + bu$ with $u \in L_F$, and set $t = (n_i \cdot u, n_j \cdot u)$. As before $q_{b,A}(y) \to |K|^{-1} S_{A_F}(t)$. The Gaussian integral of the ridge measure, multiplied by b^2, tends to the planar normal Gaussian density in orthonormal coordinates. In the threshold variables this density is $s_F \varphi_{A_F}(t)$. Multiplying the ridge coefficient in (3) and the covariance derivative factor one half gives the local derivative limit

$$|K|^{-1}(n_i^\top E n_j)\varphi_{A_F}(t).$$

All other ridges have positive distance from each fixed relative interior point. On bounded tubes the density stays bounded below, by a fixed inward ball scaled by b and uniform ellipticity. The derivative is bounded by the usual Gaussian polynomial formula. Dominated convergence applies, and overlaps and tangential trimming have lower-order volume as above. Finally use $du = dt/s_F$ and let the tube size grow by the Fisher localization. This proves the ridge Fisher formula. It is positive for a nonzero zero-normal direction by the polar argument above.

3.6 Local covariance experiments on a polytope

Let V_0 denote the linear space of facet-invisible directions, those symmetric matrices with $n_i^\top E n_i = 0$ at every facet. Choose any complementary subspace V_1 of the real symmetric matrices. The face Fisher form of [4] is positive definite on V_1 and the ridge form in Theorem 3.3 is positive definite on V_0. The following consequences use bases fixed in the two subspaces.

Corollary 3.4. *Let $b = b_N \downarrow 0$ in both assertions. For positive definite A and $Nb^2 \to \infty$, the covariance shifts*

$$A + \frac{E_1}{\sqrt{Nb}} + \frac{E_0}{\sqrt{Nb^2}}, \qquad E_i \in V_i,$$

are locally asymptotically normal with independent face and ridge Gaussian blocks, having the two information forms just described.

If $Nb^2 \to \lambda \in (0,\infty)$, let A_θ be a finite family of positive definite covariances with common facet variances. For any finite parameter set (E_1, θ), the experiments with covariance $A_\theta + E_1/\sqrt{Nb}$ converge in Le Cam distance to the product of a Gaussian shift with the face information and the ridge Poisson experiment in Theorem 3.1.

Proof. We give the derivative estimates which permit the two likelihood arguments of Section 2 for this geometry. A covariance derivative of order $r \geq 1$ has b^{2r} times a spatial derivative of order $2r$, summed over finitely many terms. If it contains a derivative in an invisible direction, first use that entire second-order spatial operator on the indicator in (3). The remaining spatial derivatives act on the Gaussian. The resulting Gaussian derivatives have order $2r-2$ and are summed over ridges. Otherwise move one spatial derivative in each term to the indicator, with measures on facets and Gaussian derivative order $2r-1$. The numerator kernel on each face is therefore bounded by $b^j \gamma_{b,A}$ times a fixed polynomial in the scaled distance, with $j = 2$ or $j = 1$ respectively. The collar lower bound and Hölder on the face, in place of Cauchy–Schwarz, give for each fixed $p \geq 2$

$$\int |D^\alpha q|^p / q^{p-1} \leq C b^j.$$

The bound gains a factor tending to zero off the corresponding Tb tubes. All constants are uniform locally in the positive definite cone. In particular the square-root second-derivative norms are $O(\sqrt{b})$, $O(b^{3/4})$, and $O(b)$ for face–face, face–ridge, and ridge–ridge directions, by the ratio bounds for powers two and four.

The normalized cross score covariance tends to zero. Inside fixed ridge tubes the face score energy is $O_T(b^2)$ by uniform boundedness of the derivative, the lower bound on the density and the tube volume. The ridge energy over the whole space is $O(b^2)$. Outside those tubes its energy is $b^2 o_T(1)$ while face energy is $O(b)$. Cauchy–Schwarz and division by $b^{3/2}$ give the claim. The third moments give Lyapunov for the scaled scores, and the second square-root bounds give an $o(N^{-1/2})$ remainder on the stated local shifts. The triangular quadratic mean likelihood expansion of Section 2 then gives the first assertion with the information limits already established.

At the critical ridge scale, split at fixed-width rescaled ridge tubes. Their count is tight and their profile limit depends only on θ, since $E_1/\sqrt{Nb} \to 0$. Outside those tubes, fixing a reference θ_0 at the same local shift produces squared Hellinger cost $b^2 o_T(1) + o(b^2)$ per discarded observation, by the common-facet localization and normalized restriction bound in Theorem 3.1. The conditional outside law retains the same face local normal experiment: restriction removes only $O_T(b^2)$ probability and face score energy, and differentiating the normalization changes the face score by $O_T(b^2)$. The moment and quadratic mean bounds remain valid. Conditionally on the rare count, pad the outside sample to length N with the fixed reference conditional law at zero local shift. The conditional squared Hellinger cost of padding is $O(1/N)$ per added point, from the face square-root derivative bound. Thus padding has vanishing error and produces the product experiment of the rare process and an independent face sample, up to vanishing errors at fixed T and the localization tail error. Conversely discard the corresponding number of face observations. The finite-parameter face likelihood limits give the Gaussian shift in Le Cam distance by likelihood-vector quantization, as in Section 2. Uncrop the ridge process using the reference Poisson complement in Theorem 3.1 and send T to infinity. This proves the second assertion. □

4 Recovering small additive noise from polytope vertices

We observe a known uniform polytope plus small independent noise. When the lower-dimensional normal projections of the noise are fixed, the difference of observation densities is exactly a collection of rescaled vertex-cone profiles. The combined profiles give a uniform negative Sobolev bound of order equal to the dimension, even at nonsimple vertices. This yields matching Wasserstein estimation and testing rates and a sharp population inverse modulus. A compact collection of vertex observations gives the same bounds with unknown projections. For the cross-polytope, the sharp inverse degree is computed for every choice of observed vertices.

4.1 Model and rates

Let K be a full-dimensional compact polytope in $\mathbb{R}^d$, $d \geq 2$. The observation density with noise density f and scale $b > 0$ is

$$q_{b,f}(y) = T_b f(y) = \frac{1}{|K|} \int \mathbf{1}_K(y - bz) f(z)\, dz.$$

For a face F write L_F for the orthogonal complement of its tangent directions and π_F for projection onto L_F. Include $F = K$, with $L_K = \{0\}$. We call u a contrast if it is real, belongs to L^2, is supported on the unit ball B, and $\pi_{F\#}(u\, dz) = 0$ for every positive-dimensional face F.

Put $0 < a < 1$. Fix a reference density f_0 on B with $0 < c < \operatorname{ess\,inf}_B f_0 \leq \operatorname{ess\,sup}_B f_0 < C$. For $s \geq 0$ let $\mathcal{F}_s$ be the densities $f = f_0 + u$ with mass-zero u supported on B_a, $c \leq f \leq C$ on B, and $\|u\|_{H^s(\mathbb{T}^d)} \leq L$ for some fixed $L > 0$. The torus is $\mathbb{R}^d/(8\mathbb{Z})^d$; its Sobolev norms are defined by Fourier series. The ambient Euclidean metric is used for transport. Let $M = Nb^d$, where N is the sample size.

Theorem 4.1. *Fix $1 \leq p \leq 2$ and $s \geq 0$. Uniformly for all sufficiently small $b > 0$, the minimax expected W_p error over $\mathcal{F}_s$ has order*

$$\min\{1, M^{-(s+1)/(2s+3d)}\}.$$

The minimax separation radius for testing $f = f_0$ against $W_p(f, f_0) \geq \rho$ within $\mathcal{F}_s$, with both errors at most $1/4$, has order $M^{-(s+1)/(2s+5d/2)}$ as $M \to \infty$. The deterministic modulus over pairs from this class at observation Hellinger distance at most $b^{d/2}\delta$ has order $\delta^{(s+1)/(s+d)}$ as $\delta \downarrow 0$, uniformly in b.

All three orders still hold on the subclass in which every positive-dimensional face projection is prescribed to equal the corresponding projection of f_0.

All lower bounds continue to hold when the first and second moments of f_0 are prescribed. If f_0 is a Gaussian normalized on B and $s \geq 2$, they also hold in the strongly log-concave subclass.

In the bivariate boxcar model, Benešová, van Es and Tegelaar [11] derive inversion formulas using shifted cumulative distributions and mixed derivatives. The uniform-box limit experiments in Section 2 identify codimension-k rare events when the first unmatched marginal has dimension k. The cone experiments in Section 3 extend this comparison to arbitrary polytopes. Here the entire vertex-stratum discrepancy is used to recover an unknown density at positive b. For smooth positively curved references, Section 5 obtains an order-$(d+1)/2$ inverse with effective sample size Nb. The known fixed Gaussian kernel in the published Archive result [14] instead yields a logarithmic Wasserstein modulus under curvature hypotheses.

4.2 Exact vertex identity

Write C_v for the inward tangent cone at vertex v. For a contrast u define

$$F_v(t) = (\mathbf{1}_{C_v} * u)(t).$$

The following identity includes nonsimple vertices.

Lemma 4.2. *There is a fixed radius R, depending only on K, such that all the F_v vanish outside the ball of radius R. For all sufficiently small b, $T_b u$ is supported in the disjoint balls $v + bB_R$, where*

$$|K| T_b u(v + bt) = F_v(t), \qquad |t| \leq R.$$

In particular,

$$\|T_b u\|_2^2 = \frac{b^d}{|K|^2} \sum_v \|F_v\|_2^2.$$

Proof. Use the facet inequalities $n_i \cdot y \le h_i$. Write the shifted indicator of each facet halfspace as its unshifted indicator plus the difference triggered by bz, and expand their product. For any affected facet set J whose intersection within K is a positive-dimensional face F_J, the multiplier of $u(z)$ depends on z only through $\pi_{F_J} z$; the term cancels. This includes the empty affected set. Whenever another term is nonzero, all unaffected inequalities hold and the affected slacks have absolute values at most $b|z|$. An empty face intersection gives no such points for all sufficiently small b, since these points lie in a fixed bounded outward enlargement and their accumulation points would solve the exact system. For a nonempty intersection, Hoffman's error bound [8] places the point at distance $O(b|z|)$ from that face, using its equalities and the remaining inequalities. Only vertices can survive. The result is uniform on $|z| \le 1$; this is the same finite-inequality expansion used in Section 3. In any fixed bounded window $y = v + bt$, all inequalities not incident to v hold automatically for $|z| \le 1$ and b sufficiently small. The others are exactly $t - z \in C_v$. This proves the local identity, and using it for each fixed $|t| > R$ and smaller b shows $F_v = 0$ there. The norm identity is change of variables on the disjoint balls. □

Use the convention $\widehat{u}(\xi) = \int e^{-i\xi \cdot x} u(x)\, dx$. Triangulate each C_v into simplicial cones using only its extreme rays, for example by triangulating a compact transverse section without adding vertices. For a simplex of rays $e_1, \dots, e_d$ let V_σ have those columns. Away from the finite union of hyperplanes orthogonal to edges,

$$\widehat{F_v}(\xi) = m_v(\xi)\widehat{u}(\xi), \qquad m_v(\xi) = \sum_\sigma \frac{|\det V_\sigma|}{\prod_{j=1}^d (i\xi \cdot e_j)}.$$

The formula follows from $D_{e_1} \cdots D_{e_d} \mathbf{1}_{C_\sigma} = |\det V_\sigma| \delta_0$, which is integration by parts over positive coordinates in the generating basis. One may apply the product of all distinct edge-direction derivatives to the cone before taking Fourier transforms, then divide away from the hyperplanes. This avoids using products of singular Fourier distributions. Polyhedral cone transforms in this form are classical [12, 9].

Lemma 4.3. *For almost every $\xi \ne 0$,*

$$\sum_v |m_v(\xi)|^2 \ge c_K |\xi|^{-2d}.$$

Consequently

$$\|u\|_{H^{-d}(\mathbb{R}^d)} \le C_K b^{-d/2} \|T_b u\|_2$$

for contrasts. The same bound holds in the torus H^{-d} norm.

Proof. Outside the edge-normal hyperplanes a linear functional $\xi \cdot x$ has a unique maximizing vertex v, because a positive-dimensional exposed face contains an edge. Every inward edge ray at v has $\xi \cdot e_j < 0$. Thus the summands of $m_v(\xi)$ have the same complex phase. Taking the ray directions to have unit length makes any one summand at least $|\det V_\sigma|\, |\xi|^{-d}$ in absolute value. The finite triangulations give a uniform positive constant. The Euclidean inequality follows by Plancherel and Lemma 4.2. For the torus inequality multiply a torus test by a smooth Euclidean cutoff equal to one on B; its Euclidean H^d norm is bounded by the original test norm. □

The vertex formulas give an explicit inverse on contrasts: off the null set of edge-normal hyperplanes, recover $\widehat{u}$ as $\sum_v \overline{m_v}\widehat{F}_v / \sum_v |m_v|^2$. For an axis-aligned box, each of the vertex profiles is a signed copy of the mixed cumulative primitive with the coordinate signs induced by the vertex. Because each coordinate integral of u vanishes, its upper and lower primitives in that coordinate differ by a sign. The exact squared norm is therefore $2^d b^d |K|^{-2}$ times the squared norm of the mixed cumulative primitive.

4.3 Compact vertex windows without fixed projections

The contrast condition gave an exact global identity, but uniform inversion requires only bounded windows at the vertices. In this section $F_v = \mathbf{1}_{C_v} * u$ for arbitrary real $u \in L^2$ supported on B.

Theorem 4.4. *There are constants $R, C > 0$ depending only on K such that*

$$\|u\|_{H^{-d}(\mathbb{R}^d)} \le C \left(\sum_v \|F_v\|^2_{L^2(B_R)} \right)^{1/2} \le C^2 b^{-d/2} \|T_b u\|_2$$

for all sufficiently small $b > 0$. The same holds with the torus norm on the left.

Proof. First bound the weighted full profiles by their compact windows. Fix one tangent cone $C = C_v$ with unit facet normals and inequalities $n_i \cdot t \le 0$ and choose w with $w \cdot t < 0$ on $C \setminus \{0\}$, for example the sum of outward unit normals. For each fixed $\tau > 0$, we claim

$$\|e^{\tau w \cdot t} F_v(t)\|_{L^2(\mathbb{R}^d)} \le C_{K,\tau} \|F_v\|_{L^2(B_R)}.$$

The profile is zero off $C + B$, where $w \cdot t \le -c_1|t| + c_2$. For $t \in C + B$, let $J = \{i : n_i \cdot t \ge -1\}$ and set $G = C \cap \{n_i \cdot z = 0 : i \in J\}$. This is a face containing the origin, and $\operatorname{dist}(t, G) \le C_1$ by Hoffman's bound, since the violations of these inequalities and equalities are all at most one. Thus $x = \pi_{(\operatorname{span} G)^\perp} t$ has norm at most C_1. Every facet not containing G is outside J, and so is automatically satisfied after subtracting any point of B. The remaining inequalities depend only on x. Write $S_G(x)$ for the resulting profile using these inequalities.

For each G choose a fixed bounded relative open patch W_G in its linear span, of positive relative volume, centered sufficiently far along a relative-interior direction of G. Every facet not containing G has slack less than -1 at $x+y$ for $|x| \le C_1$, $y \in W_G$. Thus $F_v(x+y) = S_G(x)$ on that product. For the zero-dimensional face take its singleton with unit counting measure. Choose R to contain these finitely many products, for all vertices. Their squared norm integrals control $\int_{|x| \le C_1} |S_G(x)|^2 \, dx$. For each face assignment, integrate the weighted full profile in the same orthogonal coordinates. The integral in $y \in \operatorname{span} G$ of the squared weight restricted to the support is bounded uniformly in x, by exponential decay on $C + B$. Summing over faces proves the claim.

For the Fourier comparison, put $u_v^\tau(z) = e^{\tau w_v \cdot z} u(z)$. The Fourier transform of the weighted profile is $m_v^\tau(\xi) \widehat{u}_v^\tau(\xi)$, where

$$m_v^\tau(\xi) = \sum_\sigma \frac{|\det V_\sigma|}{\prod_{e \in \sigma} (-\tau w_v \cdot e + i\xi \cdot e)}.$$

This follows by integrating the exponentially decaying kernel on each simplicial cone in its generating coordinates. For a small fixed $\delta > 0$, discard all strips $|\xi \cdot e| < \delta$ for unit edge directions of K. Their total mass in $\langle \xi \rangle^{-2d} |\widehat{u}(\xi)|^2 \, d\xi$ is at most $C_2 \delta \|u\|^2_{H^{-d}}$, where $\langle \xi \rangle = (1 + |\xi|^2)^{1/2}$. Indeed, for each edge direction apply the one-dimensional $H^1([-1,1])$ supremum bound in the coordinate $\xi \cdot e$ to the weighted Fourier transform, and integrate in the orthogonal coordinates. The derivative of

the weight is bounded by a constant times the weight. The differentiated Fourier transform is the transform of $-i(e \cdot z)u$, whose H^{-d} norm is bounded by that of u since the spatial multiplier can be smoothly cut off on B. Choose δ so that the removed mass is less than half.

On the remaining frequencies select the maximizing vertex v. Every inward edge there has $\xi \cdot e \le -\delta$. For sufficiently small τ compared with δ, every summand of m_v^τ lies in a fixed sector of opening at most $\pi/2$ about their common limiting phase. Each denominator factor is at most $C\langle\xi\rangle$ in magnitude, so $|m_v^\tau(\xi)| \ge c_3\langle\xi\rangle^{-d}$ on the region assigned to v. On this region bound the weighted transform of u_v^τ by the transform of the weighted profile. To replace u_v^τ by u, use $\|u_v^\tau - u\|_{H^{-d}} \le C_3\tau\|u\|_{H^{-d}}$, again by smooth multiplication on the fixed support. Summing over the finitely many winner regions and choosing τ smaller if needed absorbs this error. Plancherel and the weighted profile claim prove the first inequality. On the fixed finite windows, the exact local identity $|K|T_b u(v + bt) = F_v(t)$ holds for all sufficiently small b, without a projection constraint. Change of variables proves the second inequality. The test cutoff argument from Lemma 4.3 gives the torus bound. □

4.3.1 How many vertices are needed for a cross-polytope?

For the d-dimensional cross-polytope, let the observed vertices lie on k distinct coordinate axes, with at least one sign observed on each chosen axis. The least order $q \ge 0$ for a uniform H^{-q} estimate in terms of the bounded windows at those vertices is

$$q_k = \max\{d, 2(d-k)\}.$$

In particular, $\lceil d/2 \rceil$ coordinate axes suffice for degree d, and a single axis has degree $2d-2$ when $d \ge 3$.

To prove the upper bound, the cone at a vertex on axis j, after orienting that coordinate inward, is $C = \{(x,t) : t \ge \sum_{\ell=1}^{d-1} |x_\ell|\}$. Its Laplace transform at (α, σ) equals

$$\frac{2^{d-1}\sigma^{d-2}}{\prod_{\ell=1}^{d-1}(\sigma^2 - \alpha_\ell^2)} \quad \text{when } \Re\sigma > \max_\ell |\Re\alpha_\ell|.$$

This follows by integrating first in t and then in each x_ℓ. Let $\sigma = \tau + i\xi_t$, $\alpha_\ell = i\xi_\ell$, with fixed small $\tau > 0$. At any frequency choose j to maximize $|\xi_j|$ among the k observed axes, and set $A = 1 + |\xi_j|$, $R_\xi = \langle\xi\rangle$. The numerator is bounded below by $c_\tau A^{d-2}$. The $k-1$ other observed axes contribute factors bounded above by $C_\tau A^2$ each, and the unobserved axes contribute factors bounded above by $C_\tau R_\xi^2$. Thus the weighted multiplier is bounded below by

$$c_\tau A^{d-2k} R_\xi^{-2(d-k)} \ge c'_\tau R_\xi^{-q_k}.$$

Use the compact-window bound from the proof of Theorem 4.4 on the region of frequencies assigned to each j. Removing the small exponential weight on the supported input costs at most $C\tau\|u\|_{H^{-q_k}}$, as in that proof, and this term is absorbed after summing.

For sharpness, let $\mathcal{P}$ be the edge-direction differential product of the cross-polytope, optionally including Δ^2, and choose a fixed nonzero smooth compact interior bump ψ. If $2(d-k) > d$, take a direction ω whose observed coordinates are all zero and whose unobserved coordinates are generic and nonzero. Otherwise take a generic direction transverse to all edge hyperplanes. Let r_ω be the degree in T of the polynomial symbol of $\mathcal{P}$ at $T\omega + \xi$, viewed as a polynomial in T with polynomial coefficients in ξ. Set $u_T = T^{-r_\omega}\mathcal{P}[\psi(z)\cos(T\omega \cdot z)]$. It is a contrast, since the symbol vanishes on all positive-dimensional face normal spaces. For every $q \ge 0$, restricting the Fourier integral to a

fixed ball about $T\omega$ gives $\|u_T\|_{H^{-q}} \geq c_q T^{-q}$: the scaled symbol converges to a nonzero polynomial times a nonzero constant, the translated transform of ψ does not vanish on its complement, and the opposite packet has rapid decay.

Each observed cone multiplier times the symbol of $\mathcal{P}$ is a polynomial. Its denominator before cancellation has $2(d-1)$ linear factors, which are distinct directions among the edges incident to the vertex. For the first choice of ω, exactly $2(d-k)$ of these factors depend nontrivially on T at $T\omega + \xi$, and the numerator direction, the observed axis itself, is orthogonal to ω. The polynomial degree in T thus drops by $2(d-k)$. For the generic second choice it drops by $2(d-1)-(d-2) = d$. Applying the resulting constant-coefficient operator to the modulated bump gives L^2 profile norm $O(T^{-q_k})$ at every observed vertex. This precludes any $q < q_k$. The argument includes both signs of an observed axis.

4.4 Observation comparison and upper bounds

Write $H(p,q)^2 = \int(\sqrt{p} - \sqrt{q})^2$. For $f, g \in \mathcal{F}_s$ with $f - g$ a contrast, uniformly in small b,

$$H(q_{b,f}, q_{b,g}) \asymp \|T_b(f-g)\|_2. \tag{4}$$

The upper bound on the observation densities is $1/|K|$. For the lower density bound where the difference can be nonzero, take a vertex mark $t \in C_v + B_a$ with $|t| \leq R$. Choose $z_0 \in B_a$ with $t - z_0 \in C_v$, and choose a unit vector w_v in the interior of C_v. Then $z_0 - \eta w_v$ has a whole ball of fixed positive radius around it inside B and inside $t - C_v$, for fixed sufficiently small $\eta > 0$. The observation is exactly the cone profile there, which is uniformly bounded below by the noise density floor times the volume of this ball. This gives (4) by the identity for the difference of square roots. The entire complement has a common density for such contrasts.

A density difference u on B has $W_2(f,g) \leq C\|u\|_{H^{-1}(\mathbb{T}^d)}$ if $f, g \geq c$ on B. Indeed the mean-zero Neumann solution gives a flux of L^2 norm at most $C\|u\|_{H^{-1}}$ by Poincaré and extension of $H^1(B)$ functions to the torus. Using the flux with the linear density path in the Benamou–Brenier formula gives the bound. Interpolation at H^{-1} between H^{-d} and H^s, together with Theorem 4.4 and the inequality $\|T_b(f-g)\|_2 \leq 2|K|^{-1/2} H(q_{b,f}, q_{b,g})$, proves the modulus upper on the full class.

For the statistical bounds, near each fixed frequency cutoff apply Riesz representation jointly as follows. On the space of compactly supported outputs $T_b u$, a real finite trigonometric test ϕ defines the functional $T_b u \mapsto \int \phi u$ of norm at most $Cb^{-d/2}\|\phi\|_{H^d}$ by Theorem 4.4. Extend to the closure of the output space in L^2. Its canonical Riesz representative a_ϕ, viewed in the ambient L^2, is linear in ϕ. The centered observation $a_\phi(Y) - \int a_\phi q_{b,f_0}$ estimates $\int \phi u$ unbiasedly and has variance at most $Cb^{-d}\|\phi\|_{H^d}^2$ under each f, since $q_{b,f}$ is uniformly bounded. Fix measurable representatives for the finitely many basis probes used in a given experiment. For a cutoff $m \geq 1$, the unbiased empirical Fourier reconstruction has expected squared H^{-1} error

$$C\{m^{-2(s+1)} + M^{-1} m^{3d-2}\}.$$

The first term is Sobolev truncation, and the second sums variance with weights $(1+|j|^2)^{-1}$ over $|j| \leq m$. Choosing $m \asymp M^{1/(2s+3d)}$ and projecting to a countable H^{-1}-dense subset of admissible densities proves the estimation upper. For this measurable projection, take the first candidate within m^{-s-1} of the distance infimum. The triangle inequality adds at most twice the reconstruction error and the tolerance. A fixed estimate suffices for $M < 1$.

For testing, denote by X_i the vector of unbiased centered probes through frequency m, multiplied entrywise by the square-root H^{-1} weights. Write $\theta = \mathbb{E}_f X_i$ for the truncated weighted coefficient vector and Σ for its covariance matrix. The joint probe bound gives

$$\Sigma \preceq Cb^{-d}\,\mathrm{diag}\{(1+|j|)^{2d-2}\}.$$

Use the quadratic statistic

$$A_N = \frac{1}{N(N-1)} \sum_{i \neq j} \langle X_i, X_j \rangle.$$

Its expectation is $|\theta|^2$. Expansion by writing $X_i = \theta + (X_i - \theta)$ gives variance $4\theta^\top \Sigma \theta / N + 2\,\mathrm{tr}(\Sigma^2)/(N(N-1))$. Thus its standard deviation is at most

$$C\{M^{-1} m^{5d/2-2} + M^{-1/2} m^{d-1} |\theta|\}.$$

This follows by summing the squared diagonal bounds for the degenerate term, and using the largest diagonal bound for the linear term. Take $m \asymp M^{1/(2s+5d/2)}$. The null expectation vanishes and the null standard deviation is $O(m^{-2s-2})$. Under alternatives separated by a sufficiently large constant times m^{-s-1} in W_p, the transport bound and the Fourier tail bound give $|\theta| \geq B_1 m^{-s-1}$ with B_1 as large as needed. The linear standard deviation divided by $|\theta|^2$ is at most $C B_1^{-1} m^{-d/4}$. Chebyshev's inequality with a threshold equal to a large constant times m^{-2s-2} gives both error bounds. This uses the usual quadratic testing construction in a polynomially ill-posed inverse problem [13], applied to the exact finite-scale density experiment.

A simplicial vertex gives a simpler single-vertex probe. For a smooth test ϕ on the noise support choose a smooth compactly supported extension, using multiplication by a fixed cutoff for torus tests. In the bounded support of the extension, $|K| T_b u(v + bt) = \mathbf{1}_{C_v} * u(t)$ for small b, even for arbitrary mass-zero u. Integrate by parts using the d edge derivatives of the single cone. The resulting probe has L^2 norm bounded by $C b^{-d/2} \|\phi\|_{H^d}$ on the observation scale, by change of variables. This gives the same linear and joint statistical upper bounds. The corresponding H^{-d} population upper follows by duality and by $\|T_b u\|_2 \leq 2|K|^{-1/2} H(q_{b,f}, q_{b,g})$, which only uses the uniform upper bound on densities.

4.5 Localized alternatives and lower bounds

Choose one unit direction e for each distinct unoriented edge direction of K and form the constant-coefficient differential operator $\mathcal{P} = \Delta^2 \prod_e D_e$, of order D. Let $p = \mathcal{P}\psi$ for a nonzero real smooth ψ in a ball of radius $1/4$. Then $p \neq 0$, since the symbol is a nonzero polynomial and $\widehat{\psi}$ is continuous and nonzero somewhere off its zero set. The projection to each positive-dimensional face normal space vanishes: that face contains an edge, so the Fourier symbol vanishes on its normal space. The first two moments vanish by the real differential factors of total order at least four.

For each vertex v, applying the d generating derivatives of every simplicial summand of C_v cancels its cone kernel. In fact

$$\mathbf{1}_{C_v} * p = \mathcal{Q}_v \psi,$$

where $\mathcal{Q}_v$ is a linear combination of real constant-coefficient differential operators of order $D - d$. Signs from the orientation of chosen directions enter its coefficients. Each cone has a basis with distinct unoriented directions, so the required factors occur in $\mathcal{P}$. In particular the profiles remain supported in the original bump support.

Take $J \asymp h^{-d}$ centers x_j in $B_{a/2}$ spaced at least h apart for small h, and put $p_j(z) = h^{-d/2} p((z - x_j)/h)$. By cone scaling their vertex profiles are $h^d h^{-d/2} (\mathcal{Q}_v \psi)((t - x_j)/h)$, with disjoint supports for different j, on every vertex. The densities

$$f_\omega = f_0 + \tau \sum_{j=1}^{J} \omega_j p_j, \qquad \tau = \varepsilon h^{s+d/2}, \qquad \omega_j \in \{-1, 1\},$$

are admissible for fixed sufficiently small $\varepsilon > 0$. For integer Sobolev orders the bounds follow by disjointness and scaling of derivatives; interpolation gives the bound at s. The pointwise size is $O(\varepsilon h^s)$. For the restricted Gaussian baseline and $s \geq 2$, the C^2 perturbation is uniformly small, giving strong log-concavity on the ball. The moments remain fixed.

The single-observation Hellinger distance squared between neighboring vertices is at most $Cb^d\tau^2 h^{2d}$ by (4) and the exact vertex norm. A sum of the signed tests $h^{1+d/2}p_j$ has uniformly bounded Lipschitz constant by disjoint supports. It gives W_1 separation at least $c_1\tau h^{1+d/2}$ times Hamming distance. Taking h a sufficiently small constant times $\min\{1, M^{-1/(2s+3d)}\}$ bounds the neighboring sample total variations away from one. Mapping any estimate to a closest hypercube vertex in W_1 loses at most a factor two. Averaging coordinate errors over neighbor pairs gives a fixed error probability for each of the J coordinates. This proves the estimation lower.

For testing, let P_0 be the one-observation reference and mix the N-fold f_ω experiments uniformly over the signs. Put

$$g_j = \int \frac{(T_b p_j)^2}{q_{b,f_0}}, \qquad 0/0 := 0.$$

The cross terms vanish by disjoint supports of the physical vertex images, and $g_j \leq Cb^d h^{2d}$ by the same uniform positive denominator used in (4). For signs ω, ω' the integrated product of the two single-observation likelihoods is $1 + \tau^2 \sum_j g_j \omega_j \omega'_j$. Taking the Nth power and averaging, using $1 + x \leq e^x$ and $\cosh z \leq e^{z^2/2}$, bounds the mixture chi-square divergence by

$$\exp\{C\varepsilon^4 M^2 h^{4s+5d}\} - 1.$$

Choose $h \asymp M^{-1/(2s+5d/2)}$ and take ε small enough that the total variation is less than $1/2$. The same signed Lipschitz test gives separation from f_0 at least $c_2\tau h^{1+d/2} J \asymp h^{s+1}$ for every sign vector. Thus no test has both errors at most $1/4$ at this order with a sufficiently small separation constant.

Finally the population Hellinger distance between each signed alternative and f_0 is at most $Cb^{d/2}\tau h^d \sqrt{J} \leq C' b^{d/2} h^{s+d}$, again by disjoint vertex images. The same W_1 lower bound is ch^{s+1}. Take h of order $\delta^{1/(s+d)}$ with sufficiently small constant to obtain the matching modulus. Every lower bound uses W_1 and therefore applies to $1 \leq p \leq 2$.

5 Recovering small additive noise from curved convex bodies

We observe a known uniform positively curved convex body plus small independent noise and estimate the unknown noise law. On compactly supported zero-mass signals, the squared norm of the exact observation operator, divided by the noise scale, is uniformly equivalent to a negative Sobolev norm of order $(d+1)/2$, with relative error $O(b^2)$ in the leading formula, with a directional weight given by inverse curvature. On bounded positive densities with interior Sobolev perturbations of order s, the minimax expected Wasserstein loss for $1 \leq p \leq 2$ has order $(Nb)^{-(s+1)/(2s+2d+1)}$ when $Nb \to \infty$. We also obtain the sharp deterministic inverse modulus in observation Hellinger distance.

5.1 Observation and forward norm

Let $B \subset \mathbb{R}^d$ be the unit ball, let $v_d = |B|$, and put $v_0 = 1$. For independent U uniform on B and Z of density f, the observation $Y = U + bZ$ has density $T_b f$, where

$$T_b u(y) = v_d^{-1} \int \mathbf{1}_B(y - bz) u(z)\, dz.$$

The unknown f is the law of the unscaled noise Z. We use $\widehat{u}(\xi) = \int e^{-i\xi\cdot x} u(x)\,dx$ and

$$D_r(u)^2 = \int |\widehat{u}(\xi)|^2 |\xi|^{-2r}\,d\xi, \qquad r = (d+1)/2.$$

For $u \in L^2$ supported on B and of mass zero, this norm is finite since $\widehat{u}(\xi) = O(|\xi|)$ at the origin.

Theorem 5.1. *Uniformly for real $u \in L^2$ supported on B with $\int u = 0$, as $b \downarrow 0$,*

$$\left| b^{-1}\|T_b u\|_2^2 - \frac{D_r(u)^2}{\pi v_d^2} \right| \le C_d b^2 D_r(u)^2.$$

For $d = 1$ the leading identity holds exactly for $0 < b \le 1$. For $d \ge 2$ the relative error $O(b^2)$ is sharp.

Proof. For $|h| \le 2$, the ball overlap is twice a spherical cap volume. Slicing along h gives

$$|B \cap (B+h)| = v_d - v_{d-1} \int_0^{|h|} (1 - t^2/4)^{(d-1)/2}\,dt.$$

Fubini writes $\|T_b u\|_2^2$ as the quadratic form with kernel $v_d^{-2}|B \cap (B + b(x-z))|$. Its constant term vanishes. The leading kernel after division by b is $-v_{d-1}|x-z|/v_d^2$. For $w \in \mathbb{R}^d$,

$$\int \frac{1 - \cos(\xi \cdot w)}{|\xi|^{d+1}}\,d\xi = \pi v_{d-1}|w|.$$

Rotate w to the first axis and integrate in the other $d-1$ variables, obtaining v_{d-1}/ξ_1^2, and then use the one-dimensional cosine integral. Since u has compact support, Fubini with $1 - \cos$ is valid even for the signed product $u(x)u(z)\,dx\,dz$. The constant part vanishes again, giving the leading coefficient in the theorem.

For sufficiently small b, Taylor expansion of the slicing integral gives a remainder $b^2|w|^3 A(b^2|w|^2)$ for $|w| \le 2$, where A is smooth in a fixed neighborhood of zero and includes the dimensional factor. Multiply this remainder kernel, after removing b^2, by a smooth radial cutoff equal to one on $|w| \le 2$ and zero for $|w| \ge 3$. Its derivatives of order $j \le d+1$ are bounded near zero by $C_d|w|^{3-j}$ and are integrable, uniformly in b. These are weak derivatives: integration by parts outside a ball of radius η makes each boundary term tend to zero as $\eta \downarrow 0$. Its Fourier transform is therefore bounded by $C_d(1 + |\xi|)^{-(d+1)}$, by integration by parts in a coordinate with $|\xi_j| \ge |\xi|/\sqrt{d}$. Plancherel gives the error bound. In one dimension the overlap is linear on its support.

In $d \ge 2$ the cubic coefficient is $(d-1)v_{d-1}/(24v_d^2)$. The next term gives uniform relative error $O(b^4)$ by the same argument with $|w|^5$. For a difference of two unit-mass bumps supported near distinct interior points x_0, x_1, the cubic energy tends to $-2|x_0 - x_1|^3$ as the bump width goes to zero. Fixing a sufficiently narrow pair proves sharpness. □

For the comparison to tomography, put $S_u(n,t) = \int_{n\cdot z \ge t} u(z)\,dz$. Fourier transformation in t of its derivative and polar integration give

$$\int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} |S_u(n,t)|^2\,dt\,dn = \pi^{-1} D_r(u)^2.$$

Here dn is surface measure (counting measure in dimension one). This is the cumulative version of the Fourier slice identity. For the ball, the corresponding halfspace discrepancy kernel was computed explicitly in [16, Theorem 4.3]. Park and Slepčev [22, Theorem 5.2] obtain the same Sobolev

order for sliced W_2 under density bounds. Cavalier [20] studies estimation from samples whose density is proportional to a Radon transform. Theorem 5.1 controls the actual translated-ball observation at positive b. Its Fourier multiplier has Bessel zeros. Real-line uniform deconvolution at characteristic-function zeros has established density-loss theory [21, 19]. In the Gaussian model of [14], an unknown latent law instead undergoes known fixed Gaussian smoothing. The exponential attenuation there gives a logarithmic transport modulus under curvature hypotheses. The polynomial order here follows from the first term of the ball overlap and applies uniformly over compactly supported signals.

5.2 Statistical recovery

Fix $0 < a < 1$ and a flat torus $\mathbb{T}^d = \mathbb{R}^d/(8\mathbb{Z})^d$ with B inside its fundamental cube. Define Sobolev norms on this torus by Fourier series. Let f_0 be a positive bounded density on B, extended by zero outside B. Fix constants $0 < c < \inf_B f_0 \le \sup_B f_0 < C$, interpreting essential bounds, and $L > 0$. Let $\mathcal{F}_s$ consist of densities $f = f_0 + u$ satisfying

$$\operatorname{supp} u \subseteq B_a, \quad \int u = 0, \quad \|u\|_{H^s(\mathbb{T}^d)} \le L, \quad c \le f \le C \text{ on } B, \qquad s \ge 0.$$

One may additionally require the mean and covariance of f_0. All Wasserstein distances below use the Euclidean metric.

Theorem 5.2. *For $1 \le p \le 2$, sufficiently small $b > 0$, and $N \ge 1$, the minimax expected loss from N independent draws with density $T_b f$ satisfies*

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{F}_s} \mathbb{E}_f W_p(\widehat{f}, f) \asymp \min\{1, (Nb)^{-(s+1)/(2s+2d+1)}\}.$$

Both bounds hold with the fixed moments. If f_0 is a Gaussian restricted and normalized on B, and $s \ge 2$, both bounds also hold on the subclass of strongly log-concave densities with the fixed moments.

Lemma 5.3. *Uniformly on $\mathcal{F}_s$ and for all sufficiently small b,*

$$H(T_b f, T_b g) \asymp \sqrt{b}\, D_r(f - g), \qquad H(p, q)^2 = \int (\sqrt{p} - \sqrt{q})^2.$$

Proof. The densities are bounded above by $1/v_d$. Their difference vanishes outside $\{y : ||y| - 1| \le ab\}$: inside its inner radius every point of B_a is counted, and outside its outer radius none is. On this shell both densities are bounded below uniformly. Indeed write $y = (1 + bt)n$, $|t| \le a$. A point $z \in B$ contributes if

$$n \cdot z \ge t + \tfrac{b}{2}|tn - z|^2.$$

For small b, a fixed cap $n \cdot z \ge (a + 1)/2$ suffices. It has fixed positive volume and the densities on it are at least c. Thus squared Hellinger and squared L^2 discrepancy are comparable, and Theorem 5.1 applies. □

Proof of Theorem 5.2. Write $M = Nb$, and first let $M \ge 1$. For a Fourier coefficient indexed by $j \in \mathbb{Z}^d$, choose a smooth cutoff equal to one on B and multiply it by the Fourier test exponential. The homogeneous $H^r(\mathbb{R}^d)$ norm of the resulting test is $O((1+|j|)^r)$. By Fourier duality and Theorem 5.1, the functional $b^{-1/2} T_b u \mapsto \widehat{u}_{\mathbb{T}^d}(j)$ on the image of compactly supported zero-mass signals has norm $O((1 + |j|)^r)$. Extend it continuously to the closure and use its L^2 Riesz representative $a_{b,j}$, with

a conjugate convention chosen so the functional is $\int a_{b,j}(b^{-1/2}T_b u)$. Real and imaginary parts may be treated separately. Then

$$\frac{1}{\sqrt{b}}\left\{\frac{1}{N}\sum_{i=1}^{N} a_{b,j}(Y_i) - \int a_{b,j} T_b f_0\right\}$$

estimates the coefficient without bias, with variance at most $CM^{-1}(1+|j|)^{2r}$ by the uniform observation density bound. Measurable representatives may be fixed in advance. Use the zero mass coefficient exactly and use conjugate coefficients for a real truncated field. Adding f_0 to the reconstruction up to $|j| \le m$ gives squared expected error in $H^{-1}(\mathbb{T}^d)$ bounded by

$$C\{m^{-2(s+1)} + M^{-1}m^{2d-1}\}.$$

Here the variance sum uses $2r = d+1$, including $d = 1$. Take m of order $M^{1/(2s+2d+1)}$.

To obtain a density estimator, use an approximate nearest member of a fixed countable H^{-1}-dense subset of the class, choosing the first member at distance within m^{-s-1} of the infimum. The rule is measurable and increases the distance to the true density by at most a factor two plus that tolerance. The same construction works with each moment or shape restriction. For densities on B bounded below by c,

$$W_2(f,g) \le C\|f-g\|_{H^{-1}(\mathbb{T}^d)}.$$

To see this, solve the mean-zero Neumann variational problem on B with source $f-g$. The gradient field has $L^2(B)$ norm bounded by the right side using Poincaré and a bounded H^1 extension from B to the torus. Linear interpolation of the densities and the fixed Neumann current give the displayed bound by the Benamou–Brenier formula [18]. Since $W_p \le W_2$, this proves the upper rate. When $M < 1$, a fixed density on B has bounded risk.

For the lower bound take a real nonzero smooth ψ supported in a ball of radius $1/4$ with zero ordinary moments through order two, for example the bi-Laplacian of a bump. Place $J \asymp h^{-d}$ spaced centers x_j in B_a for small h, with mutual distances at least h, and set

$$f_\theta = f_0 + \varepsilon h^s \sum_{j=1}^{J} \theta_j \psi((\cdot - x_j)/h), \qquad \theta_j \in \{-1, 1\}.$$

For any integer $k \ge 0$, the H^k norm of the sum with the factor h^s is $O(h^{s-k})$ by disjointness. Interpolation at the adjacent integers gives the uniform H^s bound. Fixed sufficiently small ε gives class membership and preserves the moments. When $s \ge 2$ the perturbations have uniformly small C^2 norm, so the negative log density retains a positive Hessian on B for the restricted Gaussian baseline; convexity of the support gives strong log-concavity.

For neighboring vertices, Lemma 5.3 and scaling of the single bump give

$$H^2(T_b f_\theta, T_b f_{\theta'}) \le C b\, h^{2s+2d+1}.$$

Choose $h = \eta \min\{1, M^{-1/(2s+2d+1)}\}$ for sufficiently small fixed η. The neighboring N-sample total variation distances are then bounded by a constant strictly below one using the product Hellinger bound. The Wasserstein distance $W_1(f_\theta, f_{\theta'})$ is at least $c_1 h^{s+d+1}$ times the Hamming distance. For a witness, sum the corresponding signed translates of $h\psi((\cdot - x_j)/h)$. The gradients are uniformly bounded because the supports are disjoint, and every differing coordinate contributes an integral proportional to $h^{s+d+1}\int \psi^2$. The Hamming-risk form of Assouad's argument now gives expected W_1 loss at least $c_2 J h^{s+d+1}$. In detail, map any estimate to its closest hypercube vertex in W_1, losing at most a factor two by the triangle inequality, and average each coordinate error over neighboring pairs. The total variation bound forces a positive error probability for each coordinate. This proves the lower bound for W_1, and $W_p \ge W_1$ gives the remaining claims. □

5.3 Sharp inverse modulus

Proposition 5.4. *For $s \geq 0$, $1 \leq p \leq 2$, and all sufficiently small $b > 0$, the modulus over $f, g \in \mathcal{F}_s$ at observation distance $H(T_b f, T_b g) \leq \sqrt{b}\,\delta$ has order $\delta^{(s+1)/(s+r)}$ as $\delta \downarrow 0$, with constants uniform in b. The moment and log-concavity restrictions of Theorem 5.2 preserve the order.*

Proof. For $u = f - g$, compact support gives $\|u\|_{H^{-r}(\mathbb{T}^d)} \leq C D_r(u)$: multiply a torus test by a smooth cutoff one on B, extend it to $\mathbb{R}^d$, and apply Fourier duality and the Sobolev norm bound for this localization. Interpolate H^{-r} and H^s at H^{-1} and use Lemma 5.3 and the transport bound in the theorem. This includes $r = 1$, where interpolation is unnecessary.

For the reverse bound, choose a nonzero real smooth ψ supported in the interior of B_a and let

$$u_m = m^{-s-4} \partial_1^4 [\psi(x) \cos(m x_1)].$$

These perturbations preserve the first two moments exactly. Their H^s norms are bounded and their pointwise size is $O(m^{-s})$, by differentiating and shifting the frequency. For $s \geq 2$ their C^2 norms are bounded as well. Moreover $D_r(u_m) \leq C m^{-s-r}$. In the integral defining D_r, on $|\xi| \geq m/2$ use the L^2 bound $O(m^{-s})$. On $|\xi| < m/2$ use the rapid decay of both translates of $\widehat{\psi}$ and the factor $m^{-s-4}\xi_1^4$. The integral at zero is finite since in radial variables the resulting power is at least 6. Finally the test $m^{-1}\psi(x)\cos(m x_1)$ has bounded Lipschitz norm. Its integral against u_m is

$$m^{-s-1} \left(\int \psi(x)^2 \cos^2(m x_1)\, dx + O(m^{-1}) \right),$$

with positive limiting factor $\frac{1}{2}\int \psi^2$. Thus a fixed small multiple added to f_0 is admissible and has Wasserstein distance at least $c_3 m^{-s-1}$ from f_0. Take an integer m of order $\delta^{-1/(s+r)}$, with the constant large enough to satisfy the observation constraint by Lemma 5.3. □

5.4 Positively curved reference bodies

Let $d \geq 2$ and let K be a compact convex body with C^∞ boundary and strictly positive Gauss curvature. Keep the noise support B and define $T_{b,K}$ by replacing B in the observation kernel by K and normalizing by $|K|$. Let $G(n)$ be the reciprocal Gauss curvature at the boundary point of outward normal n.

Theorem 5.5. *For real $u \in L^2$ supported on B with zero mass, uniformly as $b \downarrow 0$,*

$$b^{-1}\|T_{b,K}u\|_2^2 = \frac{1}{2\pi|K|^2} \int_{\mathbb{R}^d} |\widehat{u}(\xi)|^2 |\xi|^{-d-1} [G(\widehat{\xi}) + G(-\widehat{\xi})]\, d\xi + O_K(b^2) D_r(u)^2, \qquad \widehat{\xi} = \xi/|\xi|.$$

Theorem 5.2 and Proposition 5.4 hold for this observation model with the same exponents.

Proof. Write a displacement as te with e unit and $t \geq 0$. Let $P_e K$ be the orthogonal projection onto $e^\perp$ and $\ell_e(y)$ the length of the chord through y parallel to e. Fubini on the chords gives the covariogram identity

$$|K \cap (K + te)| = |K| - t|P_e K| + \int_{\{\ell_e(y) < t\}} (t - \ell_e(y))\, dy,$$

where the last integral is over $P_e K$. Directional first variations of covariograms are treated more generally by Galerne [17]. The directional cubic expansion in the positively curved planar case is recorded in [15, Remark 2.2(b)].

We show that the integral is $t^3 A(e, t^2)$, with uniform derivative bounds in e and t^2 near zero. The support function of $P_e K$ is the restriction of that of K. Its spherical Hessian plus the support function times the metric is positive definite, being the restriction of the corresponding form for K. Thus its boundary is smooth and positively curved in dimension at least two (and consists of two endpoints when $d = 2$). For an outward unit normal $\nu \in e^{\perp}$, let $x(\nu)$ be the point of K with that normal. Its projection z is the corresponding boundary point of $P_e K$. Inward normal coordinates in the projected body are $y = z - \delta\nu$, $\delta \geq 0$. Near $x(\nu)$ the boundary in the plane spanned by e, ν is $x(\nu) + ae - F_{e,\nu}(a)\nu$, with $F(0) = F'(0) = 0$ and $F''(0) > 0$. The uniform positive curvature guarantees that $v = a\sqrt{F(a)/a^2}$ has a smooth local inverse $a = A_{e,\nu}(v)$. The two chord endpoints have parameters $v = \pm\sqrt{\delta}$. Their separation is $\sqrt{\delta}\, L(e, \nu, \delta)$, with L smooth positive, by taking the odd part of the inverse in v. These constructions and their derivative bounds are uniform on the compact set of orthogonal unit pairs (e, ν). For chords with sufficiently small length the projected points are in this collar, by compactness and the positive chord length at every interior projected point. The coordinate ℓ_e^2 is a smooth change from δ, with positive derivative at zero. In the inward tube volume element, this changes the radial factor to $2\ell_e J(e, \nu, \ell_e^2)\, d\ell_e$, where J is smooth. Now substitute $\ell_e = tu$ in the overlap correction and integrate $2t^3(1-u)uJ(e, \nu, t^2u^2)$ for $0 \leq u \leq 1$ over the projected boundary. Local rotations in e identify the projected unit spheres smoothly, so the resulting $A(e, t^2)$ has the stated regularity.

The linear term is

$$|w|\,|P_{\widehat{w}}K| = \frac{1}{2}\int_{\partial K} |n(x) \cdot w|\, dS(x),$$

by projection of the two sides of the boundary. For each n the one-dimensional cosine identity applied to the projected signed measure gives

$$-\iint |n \cdot (x - z)|\, u(x)u(z)\, dx\, dz = \frac{1}{\pi}\int_{\mathbb{R}} |\widehat{u}(\rho n)|^2 \rho^{-2}\, d\rho.$$

Change normal variables by the Gauss map and then use polar integration to obtain the leading form. For the remainder, multiply $|w|^3 A(\widehat{w}, b^2|w|^2)$ by a smooth cutoff equal to one on the support differences. Its weak derivatives of order $j \leq d+1$ are $O(|w|^{3-j})$ near zero uniformly by the angular derivative bounds just established. The same Fourier estimate as in Theorem 5.1 gives the relative error. The continuous positive function G is bounded above and below on the sphere, giving the uniform two-sided norm comparison.

The output density is at most $1/|K|$. Differences vanish outside the shell within distance ab of the boundary of K. In this shell write $y = x + btn(x)$ in signed normal coordinates, $|t| \leq a$. For a point $z \in B$, Taylor expansion of the signed distance near x gives $b(t - n(x) \cdot z) + O_K(b^2)$ at $y - bz$. The cap $n(x) \cdot z \geq (a+1)/2$ therefore contributes to the density for sufficiently small b, establishing the same uniform positive denominator as in Lemma 5.3. All statistical arguments and both modulus bounds then use the same Sobolev comparisons and go through with constants depending on K. □

6 Noise recovery at power-law boundaries

We observe a known compact density with a power-law boundary plus small independent noise. Its boundary exponent governs the statistical information about the unknown noise law. For two fixed noise laws, the first differing moment determines whether the leading Hellinger distance comes from the interior, a logarithmic range of depths, or a thin boundary shell. For nonparametric recovery the shell is uniformly informative: the effective sample size is $Nb^{\alpha+1}$ and the inverse has negative Sobolev order $(d+1)/2+\alpha$. We obtain sharp Wasserstein estimation

and testing rates and the population inverse modulus, without knowing the noise moments. At the logarithmic threshold we give efficient finite-moment scales and a mixed Gaussian–Poisson experiment.

6.1 Model and conclusions

Let B be the Euclidean unit ball in $\mathbb{R}^d$, $d \geq 2$. Fix $\alpha > -1$ and set

$$g(x) = c_\alpha (1 - |x|^2)_+^\alpha, \qquad c_\alpha = \frac{\Gamma(1 + \alpha + d/2)}{\pi^{d/2}\Gamma(1+\alpha)}.$$

Powers with the plus subscript are zero when their argument is nonpositive, including $\alpha \leq 0$. The observation is $X + bZ$ for X of density g, independent of the unknown noise Z. Write $T_b f(y) = \int g(y - bz) f(z)\, dz$ and $H^2(p,q) = \int (\sqrt{p} - \sqrt{q})^2$.

Let $0 < a < 1$. Fix a known density f_0 on B, and $0 < c < \operatorname{ess\,inf}_B f_0 \leq \operatorname{ess\,sup}_B f_0 < C$. For $s \geq 0$, let $\mathcal{F}_s$ be the densities $f = f_0 + u$, with mass-zero u supported on B_a, $c \leq f \leq C$ on B, and $\|u\|_{H^s(\mathbb{T}^d)} \leq L$ for fixed $L > 0$. The torus is $\mathbb{R}^d/(8\mathbb{Z})^d$. Put $r = (d+1)/2 + \alpha$ and $M = Nb^{\alpha+1}$.

Theorem 6.1. *For $1 \leq p \leq 2$ and sufficiently small $b > 0$, uniformly in N, the minimax expected W_p error from independent observations has order*

$$\min\{1, M^{-(s+1)/(2s+d+2r)}\}.$$

Testing $f = f_0$ against $W_p(f, f_0) \geq \rho$ in this class, with both errors at most $1/4$, has minimax separation order $M^{-(s+1)/(2s+d/2+2r)}$ as $M \to \infty$. The modulus over pairs in the class at observation Hellinger distance at most $b^{(\alpha+1)/2}\delta$ has order $\delta^{\min\{1,(s+1)/(s+r)\}}$ as $\delta \downarrow 0$, uniformly for small b. The lower bounds hold even if any fixed finite number of noise moments are prescribed to equal those of f_0.

To describe fixed pairs of noises, let P, Q have distinct bounded densities on B, and let $j \geq 1$ be the first order at which their moment tensors differ. This order exists by compact support. Put $\mu = P - Q$,

$$M_j(n) = \int (n \cdot z)^j \, d\mu(z), \quad A_j(y) = \frac{(-1)^j}{j!} \int D^j g(y)[z, \ldots, z]\, d\mu(z) \quad (|y| < 1),$$

and set $S_P(n,t) = 2^\alpha c_\alpha \int (n \cdot z - t)_+^\alpha \, dP(z)$.

Theorem 6.2. *As $b \downarrow 0$, the leading squared Hellinger discrepancy of the two observations has the following three forms.*

$$\begin{aligned}
2j < \alpha + 1: &\quad b^{2j} \left[\tfrac{1}{4} \int_B \frac{A_j(y)^2}{g(y)}\, dy + o(1) \right], \\
2j = \alpha + 1: &\quad b^{2j} \log(1/b) \left[2^{\alpha-2} c_\alpha \tbinom{\alpha}{j}^2 \int_{\mathbb{S}^{d-1}} M_j(n)^2 \, dn + o(1) \right], \\
2j > \alpha + 1: &\quad b^{\alpha+1} \left[\int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} (\sqrt{S_P(n,t)} - \sqrt{S_Q(n,t)})^2 \, dt\, dn + o(1) \right].
\end{aligned}$$

Each displayed coefficient in its regime is finite and positive. For a finite family sharing moments through $\lfloor (\alpha+1)/2 \rfloor$, at $Nb^{\alpha+1} \to \lambda \in (0, \infty)$ the experiments converge in Le Cam distance to observation of a Poisson process with intensity $\lambda S_P(n,t)\, dn\, dt$.

The power $\alpha = 0$ corresponds to a uniform ball, where a covariogram proof gives the exact leading forward norm with a uniform quadratic remainder Section 5. The key issue for other powers is that bulk information about a few moments can be much larger than boundary information. The shell nevertheless suffices to recover the full density. The regular, logarithmic and nonregular limits in Smith's moving-endpoint model [25] give a classical precedent at first order. Here the transition depends on the first differing moment tensor of the noise. The beta characteristic function uses a classical Bessel identity [27]. Carrasco and Florens [26] discuss the Epanechnikov density as an example with isolated characteristic-function zeros. Their fixed-scale spectral approach treats characteristic-function zeros in a different observation regime. The published Gaussian result [14] has a logarithmic inverse modulus at fixed Gaussian smoothing. Here the operator is a power-law inverse and the shrinking sample information is determined by the boundary.

6.2 Uniform shell inversion

For u supported on B, let $D_r(u)^2 = \int |\widehat{u}(\xi)|^2 |\xi|^{-2r}\, d\xi$, with $\widehat{u}(\xi) = \int e^{-i\xi\cdot z} u(z)\, dz$. Choose integer $K = \alpha$ if α is a nonnegative integer, and otherwise choose any integer $K \geq 0$ with $K > \alpha$.

Lemma 6.3. *Uniformly for L^2 signals supported on B whose moments through order K vanish,*

$$b^{-(2\alpha+1)} \|T_b u\|_2^2 = \left[\frac{2^{2\nu}\Gamma(1+\nu)^2}{\pi(2\pi)^d} + o(1) \right] D_r(u)^2, \qquad \nu = d/2 + \alpha.$$

Here D_r and the inhomogeneous $H^{-r}(\mathbb{R}^d)$ norm are equivalent on this moment kernel.

Proof. The one-coordinate marginal of g is proportional to $(1-t^2)^{\nu-1/2}$ on $[-1,1]$. Integrating its Fourier transform by the Bessel integral gives $\widehat{g}(\xi) = 2^\nu \Gamma(1+\nu) |\xi|^{-\nu} J_\nu(|\xi|)$. After Plancherel and rescaling, the normalized squared norm is $2^{2\nu}\Gamma(1+\nu)^2(2\pi)^{-d}$ times

$$\int_{\mathbb{S}^{d-1}} \int_0^\infty W(\rho/b) |\widehat{u}(\rho n)|^2 \rho^{-2\alpha-2}\, d\rho\, dn, \qquad W(x) = x J_\nu(x)^2.$$

The Bessel asymptotic gives bounded W and $W(x) = \pi^{-1}[1 + \cos(2x - \pi\nu - \pi/2)] + O(x^{-1})$ for $x \geq 1$. Taylor subtraction through degree K in the compactly cut off Fourier test shows

$$|\widehat{u}(\rho n)| \leq C\rho^{K+1} \|u\|_{H^{-r}}, \qquad 0 < \rho \leq 1.$$

This follows by duality because all the spatial derivatives of the cutoff remainder through a fixed integer larger than r are $O(\rho^{K+1})$. The low radial part $\rho < \eta$ therefore has mass at most $C\eta^{2(K-\alpha)+1}\|u\|_{H^{-r}}^2$. The norm equivalence follows from this bound and comparison of the Fourier weights away from zero.

On $\rho \geq \eta/2$, use a smooth cutoff vanishing below $\eta/2$ and equal to one above η. The radial derivative of the cut off integrand $|\widehat{u}(\rho n)|^2 \rho^{-2\alpha-2}$ has $L^1(d\rho\, dn)$ norm bounded by $C_\eta \|u\|_{H^{-r}}^2$. Indeed $\partial_\rho \widehat{u}(\rho n)$ is the transform of $-i(n\cdot z)u$ evaluated at ρn. Use Cauchy–Schwarz, the equivalence of the Fourier weights away from zero and bounded multiplication by each coordinate smoothly cut off on B. Integration by parts makes the oscillatory integral $O_\eta(b)\|u\|_{H^{-r}}^2$; the Bessel remainder is bounded the same way. Send b to zero and then η to zero to obtain the uniform leading form. □

Write $\mathcal{A}_T(b) = \{y : ||y| - 1| \leq Tb\}$. The following local bound is valid with no moment constraints.

Lemma 6.4. *For fixed sufficiently large T and sufficiently small b, every $u \in L^2$ supported on B satisfies*

$$\|u\|_{H^{-r}(\mathbb{R}^d)} \le Cb^{-\alpha-1/2}\|T_b u\|_{L^2(\mathcal{A}_T(b))}.$$

The same inequality holds for the torus norm.

Proof. First take u on the moment kernel. Write $h = 1 - |y|$ inside B. For $h \ge T_0 b$, Taylor in z through order K, and use duality of H^r and H^{-r} on the compact noise support, to get

$$|T_b u(y)| \le Cb^{K+1}h^{\alpha-K-1}\|u\|_{H^{-r}}. \tag{1}$$

For detail, insert a fixed smooth cutoff equal to one on B and supported in $2B$. On every Taylor segment the distance to the boundary is comparable to h when $T_0 \ge 4$. Spatial derivatives of g of order ℓ there are $O(h^{\alpha-\ell})$. Derivatives in z introduce factors of b/h on the integral remainder, or differentiate its bounded polynomial factors. This bounds the cutoff remainder in any fixed integer Sobolev norm, and hence in H^r, by the right coefficient of (1). Its polynomial terms vanish. Integrating (1) outside $\mathcal{A}_{T_0}(b)$ gives a squared bound $Cb^{2\alpha+1}T_0^{2\alpha-2K-1}\|u\|^2_{H^{-r}}$; on the exterior side the signal vanishes for $|y| > 1 + b$. Choose T_0 large and apply Lemma 6.3 to get the shell lower bound on the kernel.

For arbitrary u, extract finitely many moments from a second inner collar. Put $y = (1 - bR\sigma)n$ with $1 \le \sigma \le 2$ and $R > T_0$ large but fixed. On this collar the ratio $b^{-\alpha}g((1 - bR\sigma)n - bz)$ equals $2^\alpha c_\alpha(R\sigma + n \cdot z)^\alpha$ up to $O_R(b)$ in the H^r norm of the cutoff test, uniformly in n, σ. This is an ordinary smooth expansion, since $R\sigma + n \cdot z$ is uniformly positive there. If α is an integer, its polynomial in z has degree $K = \alpha$. Otherwise its expansion through degree K has cutoff H^r remainder $O(R^{\alpha-K-1})$, by the same derivative argument. Accordingly the collar profile is a sum of the functions $\sigma^{\alpha-j}$ for $0 \le j \le K$ with coefficients proportional to $R^{\alpha-j}\int (n \cdot z)^j u(z)\, dz$, plus a remainder bounded by $C(R^{\alpha-K-1} + C_R b)\|u\|_{H^{-r}}$. The binomial coefficients in this range are nonzero. Linear independence of the functions of σ bounds the sum of squared coefficient norms on the sphere by the squared collar norm. The norm of a homogeneous polynomial on the unit sphere controls its coefficients. If $\mathfrak{m}(u)$ denotes the Euclidean vector of all moments through order K, then

$$|\mathfrak{m}(u)| \le C_R b^{-\alpha-1/2}\|T_b u\|_{\mathcal{A}_{2R}(b)} + (C/R + C'_R b)\|u\|_{H^{-r}}.$$

In the integer case the C/R term is absent. The collar volume element in (n, σ) is $bR(1 - bR\sigma)^{d-1} dn\, d\sigma$, giving the factor displayed.

Subtract a fixed smooth compactly supported linear lift u_{lo} of $\mathfrak{m}(u)$. Such a lift exists by the positive definite Gram matrix of monomials weighted by a smooth positive bump on an interior ball. It has H^{-r} and L^∞ norms at most $C|\mathfrak{m}(u)|$. For any fixed shell of width $T_0 b$, $\|T_b u_{\mathrm{lo}}\|_{\mathcal{A}_{T_0}(b)} \le C_{T_0} b^{\alpha+1/2}|\mathfrak{m}(u)|$. For $\alpha \ge 0$ this follows by the pointwise kernel bound. For $-1 < \alpha < 0$, integrate in $n \cdot z$ first. On $|t| \le T_0$, $y = (1 + bt)n$, the quantity $b^{-1}(1 - |y - bz|^2) = 2(n \cdot z - t) - b|tn - z|^2$ is bounded and its derivative in $n \cdot z$ is bounded below by one for small b, so the integral of its power on the positive set is uniformly bounded even for the negative exponent. Apply the first part to $u - u_{\mathrm{lo}}$, and use the moment bound. Choosing R large and then b small absorbs the residual $\|u\|_{H^{-r}}$ term. Finally, cutting off torus tests smoothly on B proves the torus norm bound by duality. □

6.3 Information bounds for the density model

On any fixed shell $\mathcal{A}_T(b)$ the observation densities are at most $C_T b^\alpha$. This uses exactly the kernel bound in the preceding proof and the uniform noise density upper bound. Consequently Lemma 6.4 gives, for all $f, g \in \mathcal{F}_s$,

$$\|f - g\|_{H^{-r}(\mathbb{T}^d)} \le Cb^{-(\alpha+1)/2}H(T_b f, T_b g). \tag{2}$$

Indeed, on the shell use $|T_b f - T_b g| \le 2\sqrt{C_T b^\alpha}|\sqrt{T_b f} - \sqrt{T_b g}|$.

A companion upper bound on information, used for lower risks, holds for u supported on B_a in the moment kernel:

$$\int \frac{(T_b u)^2}{T_b f_0} \le C b^{\alpha+1} \|u\|^2_{H^{-r}(\mathbb{R}^d)}, \qquad 0/0 := 0. \tag{3}$$

On the portion of $\mathcal{A}_T(b)$ where the signal can be nonzero, $t \le a$ in the normal parameterization. The noise cap $n \cdot z \ge (a+1)/2$ is counted and its evaluation of $1 - |y - bz|^2$ is between two positive constants times b, uniformly for $-T \le t \le a$. Thus $T_b f_0 \ge c' b^\alpha$, and Lemma 6.3 bounds that part of (3). On the inner region of depth $h \ge Tb$, $T_b f_0 \ge c' h^\alpha$, since all latent arguments stay at depths comparable to h. Equation (1) bounds this part by $C b^{2K+2} \int_{Tb}^1 h^{\alpha-2K-2} dh \, \|u\|^2_{H^{-r}}$, of the claimed order. The signal is zero beyond the outer shell.

For the upper estimation bound, apply Riesz representation on the closure of the restricted output $(T_b u)|_{\mathcal{A}_T(b)}$. A fixed trigonometric test ϕ defines the functional $\int \phi u$ of norm at most $C b^{-\alpha-1/2} \|\phi\|_{H^r}$ by Lemma 6.4. The canonical representative is linear in ϕ and may be extended by zero off the shell. Center its evaluation at $T_b f_0$. It is unbiased for the coefficient of u, and the density upper bound on the shell gives variance at most $C b^{-\alpha-1} \|\phi\|^2_{H^r}$. For a real Fourier basis and cutoff m, reconstruct from the sample means. Its mean squared H^{-1} error is at most

$$C\{m^{-2(s+1)} + M^{-1} m^{d+2r-2}\}.$$

Since $d + 2r - 2 > 0$, choose $m \asymp M^{1/(2s+d+2r)}$ for $M \ge 1$. Turn this estimate into an admissible density using the first near-minimizer within m^{-s-1} in a fixed countable H^{-1}-dense subclass. This is measurable and costs at most twice the reconstruction error plus the tolerance. The Neumann flux bound $W_2(f, g) \le C \|f - g\|_{H^{-1}(\mathbb{T}^d)}$ holds because both densities are bounded below on the ball: use a mean-zero Neumann solution for $f - g$. Poincaré and bounded H^1 extension from the ball to the torus bound the solution gradient in L^2 by the torus dual norm. Use that gradient with linear interpolation of densities in the dynamic formulation of transport, as in Section 5. A fixed estimate suffices for $M < 1$.

For testing, use the vector of centered real Fourier probes through m weighted by $(1 + |j|^2)^{-1/2}$, and denote its mean by θ and covariance by Σ. Joint linearity gives $\Sigma \preceq C b^{-\alpha-1} \operatorname{diag}((1 + |j|)^{2r-2})$. Following the quadratic statistic for indirect goodness-of-fit in [24], average the inner product over ordered distinct observations. This average has mean $|\theta|^2$ and variance $4\theta^\top \Sigma \theta / N + 2 \operatorname{tr}(\Sigma^2)/(N(N-1))$. The first part of its standard deviation is at most $C M^{-1/2} m^{(r-1)_+} |\theta|$. The other part is at most $C M^{-1} m^{d/2+2r-2}$, since $d + 4r - 4 > 0$ for $d \ge 2, \alpha > -1$. Choose $m \asymp M^{1/(2s+d/2+2r)}$. The null standard deviation and the squared Sobolev tail are both $O(m^{-2s-2})$. For alternatives separated by a large constant times m^{-s-1} in W_p, the weighted mean norm is at least that order by the transport upper bound and the tail estimate. The ratio of the linear standard deviation to $|\theta|^2$ tends uniformly to zero: the exponent of m is $-d/4$ when $r \ge 1$, and is $1 - r - d/4 < 0$ otherwise. Chebyshev's inequality gives the testing upper bound. For the population upper, use (2) and interpolate H^{-1} between H^{-r} and H^s if $r > 1$; for $r \le 1$ use the direct norm comparison. This proves all upper claims.

6.4 Matching lower bounds

Fix an integer $q > r$ and choose a nonzero real smooth bump p supported in a ball of radius $1/4$ with all moments through at least $\max\{q-1, K\}$ zero, as well as any prescribed finite moment orders. A sufficiently high derivative of a compact smooth bump has these properties. For separated centers x_ℓ in $B_{a/2}$, at mutual distance at least h, let $p_\ell(z) = h^{-d/2} p((z - x_\ell)/h)$. Take $J \asymp h^{-d}$ such centers

and put

$$f_\omega = f_0 + \gamma \sum_{\ell=1}^{J} \omega_\ell p_\ell, \qquad \gamma = \varepsilon h^{s+d/2}, \qquad \omega_\ell \in \{-1, 1\}.$$

For small fixed $\varepsilon > 0$ these are admissible. Disjointness gives the pointwise bound $C\varepsilon h^s$ and the integer Sobolev bounds $C\varepsilon h^{s-k}$ at each order k, and interpolation gives the required H^s bound.

For any real coefficients z_ℓ,

$$\left\| \sum_\ell z_\ell p_\ell \right\|_{H^{-r}(\mathbb{R}^d)} \le C h^r |z|. \tag{4}$$

In fact the L^2 bound is $C|z|$ by disjointness. For the H^{-q} bound, against a smooth test subtract its L^2 polynomial projection of degree $q-1$ in each disjoint cell containing one bump. The error on a cell is bounded by Ch^q times the L^2 norm of the derivatives of order q there, by the polynomial Poincaré inequality on a fixed ball or cube and rescaling. This inequality follows by choosing a polynomial with the same mean values of derivatives through order $q-1$ and applying ordinary Poincaré successively to the differences. The moments kill the polynomial parts. Sum by Cauchy–Schwarz and then interpolate H^{-q} with L^2 to get (4).

The signed sum of the appropriate $h^{1+d/2} p_\ell$ has bounded Lipschitz constant independently of the signs and centers, since its supports are separated. It gives Wasserstein distance $W_1(f_\omega, f_{\omega'})$ at least $c_1 \gamma h^{1+d/2}$ times their Hamming distance. For neighbors, (3)–(4) and $T_b f_\omega \ge c' T_b f_0$ give squared single-observation Hellinger distance at most $Cb^{\alpha+1}\gamma^2 h^{2r}$. Choose h a sufficiently small constant times $\min\{1, M^{-1/(2s+d+2r)}\}$. Then neighboring product total variations are bounded strictly below one. Project any estimator to a nearest hypercube vertex in W_1, losing at most a factor two, and average coordinate error over each neighboring pair. The resulting expected number of errors is at least a constant times J, giving the estimation lower bound $c_2 h^{s+1}$.

For the testing lower, use the uniform mixture over the N-observation laws of the signs. Let G be the real matrix with entries

$$G_{\ell k} = \int \frac{T_b p_\ell(y)\, T_b p_k(y)}{T_b f_0(y)}\, dy.$$

Equations (3)–(4) give $0 \preceq G \preceq Cb^{\alpha+1} h^{2r} I$. The product likelihood inner product for signs ω, ω' is $(1+\gamma^2 \omega^\top G \omega')^N$. The mixture chi-square divergence is therefore at most $\mathbb{E} e^{N\gamma^2 \omega^\top G \omega'} - 1$. For a real symmetric matrix A with $\|A\|_{\mathrm{op}} < 1$, conditioning on ω and using $\cosh t \le e^{t^2/2}$ bounds $\mathbb{E} e^{\omega^\top A \omega'}$ by $\mathbb{E} e^{\|A\omega\|^2/2}$. Write this as a Gaussian integral, apply the same cosine bound to the remaining signs, and integrate the resulting Gaussian quadratic form. This gives $\det(I - A^2)^{-1/2}$. With $A = N\gamma^2 G$ and $\|A\|_{\mathrm{op}} \le 1/2$, the latter is at most $\exp(C \operatorname{tr} A^2)$. Here $\operatorname{tr} A^2 \le C\varepsilon^4 M^2 h^{4s+d+4r}$. Choose $h \asymp M^{-1/(2s+d/2+2r)}$ and ε sufficiently small. Then the mixture total variation from the null is less than $1/2$. Every mixture term has $W_1(f_\omega, f_0) \ge c_3 h^{s+1}$ by the signed tests. This gives the testing lower bound.

For the population lower with $r > 1$, (3)–(4) give

$$H(T_b f_\omega, T_b f_0) \le C b^{(\alpha+1)/2} h^{s+r}.$$

Take h a sufficiently small constant times $\delta^{1/(s+r)}$ and use the same W_1 separation. For $r \le 1$, multiply a fixed nonzero smooth compact bump in the moment kernel by a small constant times δ. It gives the required linear lower bound, since its W_1 distance from f_0 is a positive constant times the amplitude. The lower arguments apply to all $p \ge 1$ because $W_p \ge W_1$. This completes Theorem 6.1.

6.5 Three Hellinger scales and the Poisson experiment

We prove Theorem 6.2. Let $h = 1 - |y|$ in the latent ball. For $h \geq Tb$, with T a large fixed constant, the common relative approximation is

$$T_b P(y), T_b Q(y) = g(y)(1 + O(b/h)).$$

Here $T_b P$ denotes the observation density for noise law P. Taylor expansion through the first differing moment gives

$$T_b P(y) - T_b Q(y) = b^j A_j(y) + O(b^{j+1} h^{\alpha-j-1}), \qquad |A_j(y)| \leq C h^{\alpha-j}.$$

The estimates are uniform by the interior derivative bounds used for (1). On the complement of the interior with $h \geq Tb$, the squared Hellinger cost is $O_T(b^{\alpha+1})$ from the fixed-shell bounds on the two densities.

If $2j < \alpha + 1$, the upper bound by $Cb^{2j}h^{\alpha-2j}$ is integrable after scaling. Dominated convergence in the interior and the negligible shell give $b^{-2j}H^2 \to \frac{1}{4}\int A_j^2/g$. For $y = (1-h)n$ and $h \downarrow 0$,

$$g(y) = 2^\alpha c_\alpha h^\alpha (1 + O(h)), \qquad A_j(y) = 2^\alpha c_\alpha \binom{\alpha}{j} h^{\alpha-j} M_j(n) + O(h^{\alpha-j+1}).$$

For the second formula, the highest singular power occurs when each spatial derivative hits a separate factor of the linear part of $1 - |y - bz|^2$ in its Taylor expansion. Every term using the quadratic factor has at least one additional power of h. Equivalently, expand the scalar function of this quadratic through order j. In the interior-dominant regime the binomial coefficient is nonzero, and a nonzero moment tensor gives a nonzero polynomial on the sphere. This proves positivity.

At $2j = \alpha + 1$, the same expansions show that the leading integrand after division by b^{2j} is $2^{\alpha-2}c_\alpha\binom{\alpha}{j}^2 h^{-1} M_j(n)^2$. Corrections from the spherical Jacobian and the $O(h^{\alpha-j+1})$ term are integrable at zero. Corrections from Taylor in b and the denominator are bounded by Cbh^{-2} on $h \geq Tb$, of integral $O_T(1)$. The fixed shell also costs $O_T(b^{2j})$. Integrating h^{-1} yields the claimed logarithmic coefficient, which is positive since $\alpha = 2j - 1$.

Now assume $2j > \alpha + 1$. In $y = (1 + bt)n$ with $|t| \leq T$, the normalized densities $b^{-\alpha}T_b P$ and $b^{-\alpha}T_b Q$ converge to $S_P(n,t)$ and $S_Q(n,t)$. For $\alpha > 0$ this follows by bounded convergence in z; for $\alpha = 0$ use absolute continuity to exclude the threshold hyperplane. For $-1 < \alpha < 0$, first restrict to arguments at least a small positive η and use bounded convergence away from the singularity. The omitted integrals are $O(\eta^{\alpha+1})$, uniformly, by integration in $n \cdot z$ and the derivative bound above. The scaled density has a uniform bound for bounded t in all cases. The volume element is $b(1 + bt)^{d-1} dn\, dt$, so dominated convergence gives the coefficient on the bounded shell. Outside it the interior discrepancy is at most

$$Cb^{2j} \int_{Tb}^{1} h^{\alpha-2j}\, dh \leq Cb^{\alpha+1} T^{\alpha-2j+1}.$$

The exterior is empty for $T > 1$. Sending $T \to \infty$ proves the full boundary formula. Positivity follows from injectivity of the one-dimensional cumulative power transform, applied in each normal direction. Indeed a translate and reversal of the threshold coordinate writes the transform as convolution of $t^\alpha \mathbf{1}_{t>0}$ with a compact signed measure supported in the nonnegative half-line. Its Laplace transform is the product of that signed measure's transform and $\Gamma(\alpha+1)s^{-\alpha-1}$, so a zero profile implies a zero projected measure. Almost-everywhere agreement of normal projections then gives agreement in every direction by continuity of characteristic functions, hence $P = Q$.

Finally consider the finite family. For fixed T, the count of retained observations is binomial with probability $O_T(b^{\alpha+1})$. The binomial Poisson approximation has vanishing error at $Nb^{\alpha+1} \to \lambda$. Given the count the scaled marks are independent with the normalized retained density. The resulting intensity measures converge in L^1 on the fixed window to $\lambda S_P(n,t)\,dn\,dt$. The cropped laws therefore converge in total variation as laws of point processes. To reconstruct the discarded observations, sample the normalized exterior density of a fixed family reference and restore the sample order uniformly. Conditional on retained count and marks, the actual remaining observations are independent with their normalized exterior density. The normalization masses tend to one, and their squared Hellinger comparability to the reference is bounded by a constant times the unnormalized squared square-root discrepancy. The preceding tail bound and the product Hellinger inequality give reconstruction error tending to zero as $b \to 0$ and then $T \to \infty$. The Poisson experiment can likewise be uncropped by adding reference-intensity points outside the window. The integrated squared square-root intensity discrepancy there tends to zero uniformly over the finite family by the boundary formula and tail bound. Poisson affinity on finite regions followed by exhaustion bounds the error, including when the full intensities have infinite mass. This proves convergence in Le Cam distance.

6.6 Estimating a fixed noise moment

For each integer $j \geq 1$, let $\mathfrak{m}_j(f)$ be the vector of all degree-j moments of f, and write

$$\sigma_j(b) = \begin{cases} b^j, & 2j < \alpha + 1, \\ b^j\sqrt{\log(1/b)}, & 2j = \alpha + 1, \\ b^{(\alpha+1)/2}, & 2j > \alpha + 1. \end{cases}$$

Theorem 6.5. *The minimax expected Euclidean loss for estimating $\mathfrak{m}_j(f)$ on $\mathcal{F}_s$, uniformly for small b, has order*

$$\min\{1, (\sqrt{N}\sigma_j(b))^{-1}\}.$$

When $N\sigma_j(b)^2 \to \infty$, this is also the separation rate for testing $\mathfrak{m}_j(f) = \mathfrak{m}_j(f_0)$ with both errors at most $1/4$. The other moments are unrestricted.

Proof. Below the critical degree, expand the observable monomials by independence:

$$\mathbb{E}Y^\beta = \sum_{\gamma \leq \beta} \binom{\beta}{\gamma} \mathbb{E}X^{\beta-\gamma} b^{|\gamma|} \mathbb{E}Z^\gamma.$$

Triangular inversion expresses each degree-j moment as the expectation of a bounded-data polynomial probe of size $O(b^{-j})$. Above the critical degree, the same Riesz probes as in the proof of Theorem 6.1, with the Fourier tests replaced by fixed cutoff monomials, give variance $O(b^{-(\alpha+1)})$. It remains to get the logarithmic improvement when $\alpha = 2J - 1$ and $j = J$.

Write $q_0 = T_b f_0$ and use the squared output norm $\|T_b u\|^2_{1/q_0} = \int (T_b u)^2/q_0$. It is finite for L^2 signals supported on B_a by the shell and interior lower bounds on q_0 already proved and by Young's inequality. Lemma 6.4 and $q_0 \leq Cb^{2J-1}$ on the shell give

$$b^J \|u\|_{H^{-r}} \leq C\|T_b u\|_{1/q_0}.$$

For mass-zero u, on the interior $h = 1 - |y| \geq Tb$, Taylor expansion gives $T_b u = \sum_{\ell=1}^J b^\ell A_\ell(u) + R_b$; here $A_\ell(u)$ is defined as in Theorem 6.2 using the signed measure $u(z)\,dz$ at order ℓ, regardless of

its other moments. The weighted norm of R_b on this region is at most $Cb^J\|u\|_{H^{-r}}$: its pointwise bound is $Cb^{J+1}h^{\alpha-J-1}\|u\|_{H^{-r}}$ by the Taylor remainder estimate, and the weighted square integral contains $b^{2J+2}\int_{Tb}^1 h^{-3}\,dh$.

The weighted Gram matrix of the maps $\mathfrak{m}_\ell \mapsto A_\ell$, $\ell < J$, converges to their Gram matrix in $L^2(1/g)$ by dominated convergence, since $2\ell < 2J$. This form is positive definite on the direct sum of tensor spaces. Indeed, for a nonzero tuple, take the largest $\ell < J$ with nonzero tensor. Its leading boundary power $h^{\alpha-\ell}$ has a nonzero angular polynomial, so lower degrees cannot cancel it. The degree-J Gram matrix after division by $\log(1/b)$ converges to the positive definite angular form

$$2^\alpha c_\alpha \binom{\alpha}{J}^2 \int_{\mathbb{S}^{d-1}} M_J(n)^2\,dn$$

by the logarithmic calculation of Theorem 6.2. Its cross terms with each lower degree, after division by $\sqrt{\log(1/b)}$, tend to zero: their unscaled integrands are integrable near $h = 0$, since $\alpha-\ell-J > -1$. These finite-dimensional limits give, uniformly in u,

$$\sum_{\ell<J} b^{2\ell}|\mathfrak{m}_\ell(u)|^2 + b^{2J}\log(1/b)|\mathfrak{m}_J(u)|^2 \le C\|T_b u\|^2_{1/q_0}.$$

In fact the left side is first controlled by the weighted norm of the Taylor polynomial, and the remainder has already been controlled by the full weighted norm.

Use Riesz representation on the closure of the weighted output space to obtain a centered probe for each critical moment. If its representative is v in $L^2(1/q_0)$, the observable is $v(Y)/q_0(Y)$, centered at the baseline. Its second moment under f is at most a constant times $\|v\|^2_{1/q_0}$, since the noise density bounds give $T_b f \le C' q_0$ everywhere. Its expectation is exactly the target moment difference and its variance is $O(\sigma_J^{-2})$. The sample means in all regimes give the upper estimation and testing bounds; when the effective information is bounded a constant estimate gives bounded error.

For the matching lower bounds, choose a single smooth compact interior mass-zero u_j with moments below degree j zero and nonzero degree-j tensor. A derivative of order j of a smooth bump with nonzero integral suffices. For small fixed $|\epsilon|$, $f_0+\epsilon u_j$ is admissible. The fixed-direction weighted cost $\|T_b u_j\|^2_{1/q_0} \le C\sigma_j(b)^2$ follows from the same interior and shell bounds as the three Hellinger scales, using the cap lower bound where u_j contributes. Thus for a sufficiently small constant times $\min\{1, (\sqrt{N}\sigma_j(b))^{-1}\}$ as amplitude, the two experiments have total variation less than $1/2$. Their target moments differ by a constant times that amplitude. The two-point estimation and testing lower bounds follow. □

6.7 The mixed experiment at the logarithmic threshold

Suppose $\alpha = 2J - 1$ and $Nb^{2J} \to \lambda \in (0, \infty)$ as $N \to \infty$ and $b \downarrow 0$. Let f_θ range over a fixed finite family of noise densities bounded above and below on B and sharing moments through J. For $1 \le j \le J$ fix finitely many bounded signed directions $u_{j,\ell}$ on B with moments through $j-1$ zero. From a finite parameter set (θ, a) use the noise densities

$$f_{\theta,a,b} = f_\theta + \sum_{j<J,\ell} \frac{a_{j,\ell}u_{j,\ell}}{\sqrt{N}b^j} + \sum_\ell \frac{a_{J,\ell}u_{J,\ell}}{\sqrt{N}b^J\sqrt{\log(1/b)}}.$$

All additive terms tend uniformly to zero, so the densities are admissible for small b.

Theorem 6.6. *The experiments converge in Le Cam distance to the independent product of the boundary Poisson experiment of Theorem 6.2 at f_θ and two Gaussian shifts for the lower and critical direction coefficients. The lower information matrix has entries $\int_B A_j(u_{j,\ell})A_{j'}(u_{j',\ell'})/g$ for $j, j' < J$. The critical information entries are*

$$2^\alpha c_\alpha \binom{\alpha}{J}^2 \int_{\mathbb{S}^{d-1}} M_J(u_{J,\ell}; n) M_J(u_{J,\ell'}; n)\, dn.$$

A Gaussian block with information I is represented by an observation of mean Ia and covariance I, including singular information matrices.

Proof. For fixed T, split the observations into the boundary window and its complement. The retained count and marks have the same Poisson limit as at f_θ by the bounded perturbations and the fixed-shell density bound. On the complement replace f_θ by a fixed f_{θ_0}, keeping the same local perturbations. The signed difference has moments through J zero. The weighted interior bound from the phase proof gives squared Hellinger cost $O(b^{2J}T^{-2})$ per point before normalization, and the same bound up to a constant after normalization. The total variation error is $O(T^{-1})$ uniformly for small b.

At the exterior reference, the score in direction $u_{j,\ell}$ is $T_b u_{j,\ell}/T_b f_{\theta_0}$ minus its exterior conditional mean. The ratio is uniformly bounded by the density floor. Deleting the fixed shell removes $O_T(b^{2J})$ score energy, and changes the mean by $O_T(b^{2J})$, by the same floor and the bound on the shell mass. The Taylor calculation of Theorem 6.5 yields the two information limits at the normalizations σ_j. Higher terms of the individual Taylor series have negligible scaled variance: for $j < J$ use the dominant integrable term, and for $j = J$ use the logarithmic leading term. The lower-critical cross terms vanish by integrability of $h^{\alpha-j-J}$ and the critical logarithmic division. The group of lower degrees need not be mutually orthogonal.

Writing r_b for the locally scaled linear perturbation divided by the reference density, the conditional likelihood ratio is $1 + (r_b - \mathbb{E}_0 r_b)/(1 + \mathbb{E}_0 r_b)$, with expectation under the exterior reference. Each centered increment is bounded by a deterministic $o(1)$, since all local perturbations tend to zero. The increments of the normalized linear likelihood are centered. Their joint summed covariances converge to the stated block matrix. Boundedness of the maximum increment gives the triangular-array central limit theorem and concentration of the sum of squares; the cubic log remainder is bounded by the maximum increment times that sum. Consequently the finite likelihood vector converges to the joint Gaussian shift likelihood vector. Its coordinates have mean one in both experiments, hence are uniformly integrable. Truncating and quantizing the likelihood vector gives convergence of the finite experiments in Le Cam distance, by regenerating from the reference conditionals within each bin.

Given the boundary count and marks, the exterior sample has length N minus the count and independent conditional draws. Pad it to length N with the zero-perturbation exterior reference. The squared Hellinger cost per padded point relative to a perturbed exterior draw is $O(1/N)$, from the same scaled score norm bounds and the reference denominator. The retained count is tight, so the padding cost tends to zero. In the reverse direction discard the corresponding number of exterior points and restore uniform order. These kernels give independence of the two parts at fixed window width up to the replacement error. Finally let $T \to \infty$ and uncrop the boundary process as in Theorem 6.2. □

6.8 Smooth positively curved reference supports

Let D be a finite union of bounded convex bodies with pairwise disjoint closures, each with C^∞ boundary and strictly positive Gauss curvature. Let g be a known density supported on D, smooth

and positive in its interior, with $g(y) = w(y)h(y)^\alpha$ near the boundary on the inside. Here w is smooth and strictly positive up to the boundary, and h is the inward distance to the boundary.

Theorem 6.7. *Theorems 6.1 and 6.5 hold with unchanged exponents for this observation model. Theorems 6.2 and 6.6 also hold. In their boundary integrals, replace unit normals n on the sphere by outward normals n_x at $x \in \partial D$, integrate against surface measure $dS(x)$, and replace $2^\alpha c_\alpha$ by $w(x)$. The bulk integrals use the new g and D.*

Proof. We give the Fourier step in detail. For one component and a unit direction n, each extreme point $x_\pm(n)$ of $n \cdot x$ is unique and depends smoothly on n. In tangent coordinates near $x_+(n)$, write $y = x_+(n) + v - \eta n$. The boundary is a graph $\eta = F_n(v)$, with $F_n(0) = 0$, zero gradient and positive definite Hessian. The inward distance is $(\eta - F_n(v))$ times a smooth positive factor equal to one at $(v, \eta) = (0, 0)$. The parameter-dependent Morse change of coordinates writes $F_n(v) = |z|^2/2$ and has inverse Jacobian $\kappa(x_+)^{-1/2}$ at zero, where κ is Gauss curvature. For the slice at depth η from the supporting plane, use $z = \sqrt{\eta}\,\zeta$. Its projected density is $\eta^{r-1}A(n, \eta)$ where A is smooth down to zero. Indeed the integral is over $|\zeta| \le \sqrt{2}$ with weight $(1 - |\zeta|^2/2)^\alpha$, times a smooth amplitude in $(\sqrt{\eta}\,\zeta, \eta)$. The odd Taylor terms in $\sqrt{\eta}$ integrate to zero, and the remainders and their derivatives are bounded against the integrable weight. At zero this coefficient is

$$A(n, 0) = \frac{(2\pi)^{(d-1)/2}\Gamma(\alpha + 1)}{\Gamma(r)} \frac{w(x_+(n))}{\sqrt{\kappa(x_+(n))}}.$$

The lower endpoint has the analogous expression. On compact interior ranges of projected levels away from the endpoints the slice density is smooth uniformly in n. In boundary charts there, the projected level is a submersion on the boundary, so one may integrate the same inward weight h^α against smooth parameterized amplitudes. Interior charts have smooth density.

Applying the one-dimensional Fourier endpoint expansion to these slices gives uniform asymptotics

$$\widehat{g}(\rho n) = \rho^{-r} \sum_a C_a(n) e^{-i\rho n \cdot x_a(n)} + O(\rho^{-r-1}),$$

where a runs through the two support points of every component and $|C_a(n)| = \Gamma(r)A_a(n, 0)$. The endpoint multipliers are $\Gamma(r)e^{\pm i\pi r/2}$. This is the Fourier endpoint formula for a smooth amplitude times η_+^{r-1}, obtained by applying the oscillatory integral formula to its first two Taylor terms and integrating by parts on the smooth remainder. The analogous curved power-edge expansion for radial gauges appears in [23]. In the squared Fourier form, the diagonal terms give

$$(2\pi)^{-d} \int |\widehat{u}(\xi)|^2 |\xi|^{-2r} \sum_a |C_a(\xi/|\xi|)|^2 \, d\xi.$$

This is bounded above and below by positive multiples of $D_r(u)^2$. For the cross terms, write $\Phi_a(\xi) = \xi \cdot x_a(\xi/|\xi|)$. The gradient of $\Phi_a - \Phi_{a'}$ is the difference of the corresponding support points, since differentiating the support point itself is tangential. It is bounded away from zero for distinct phases: points on different components are separated, and opposite support points of a single component differ by at least its minimum width. On $|\xi| \ge \eta$ integrate by parts along $\nabla(\Phi_a - \Phi_{a'})/|\nabla(\Phi_a - \Phi_{a'})|^2$ in the cross oscillation with frequency $1/b$. All first derivatives of this vector field and the angular amplitudes are bounded there. The derivatives of $\widehat{u}$ in this integral are controlled in weighted L^2 by compactly supported coordinate multiplication, as in Lemma 6.3. A smooth high-frequency cutoff and the same low-frequency moment estimate therefore give the weighted leading form with uniform relative error $o(1)$ after rescaling by $b^{-(2\alpha+1)}$. The uniformly

bounded low-frequency contribution uses $|\widehat{g}(\rho n)| \le C(1+\rho)^{-r}$, from the displayed expansion and integrability of g.

For the shell proof use $y = x + btn_x$, with volume element $b(1 + O_T(b))\, dS(x)\, dt$. On bounded collars the signed depth of $y - bz$ has expansion

$$b(n_x \cdot z - t) + O_T(b^2).$$

The derivative in the n_x direction of z, divided by b, is bounded below by a positive constant. Apply the negative-power bound from Lemma 6.4. On a fixed far-inner collar, the normalized kernel is

$$w(x)(R\sigma + n_x \cdot z)^{\alpha} + O_R(b)$$

in the cutoff H^r test norm. Norms of the angular moment polynomials on one component's boundary control their coefficients by its Gauss map and positive area Jacobian. Thus the moment lift and absorption work unchanged. The interior Taylor estimate (1) uses the new h and follows by the same derivative bounds. For the lower reference density bounds, a noise cap puts the arguments at depth comparable to b on the relevant shell, and the entire noise support puts them at depth comparable to h in the deeper interior. These estimates prove (2)–(4) for the new g, and the statistical bounds use only those estimates.

The phase proof likewise uses the same derivatives and collar. The leading depth coefficient of A_j is $w(x)\binom{\alpha}{j} h^{\alpha-j} M_j(n_x)$. The same nonvanishing angular argument applies on one component. The Poisson proof, moment Gram estimates and exterior Gaussian likelihood argument use precisely these directional coefficients, the fixed-shell mass order and the bulk Taylor estimates. They give the replacements stated in the theorem. □

7 Critical Wasserstein rates for one-dimensional endpoint noise

A known interval density plus small unknown additive noise defines an inverse problem whose difficulty is set by the smaller boundary exponent. We give the sharp Wasserstein estimation and testing rates on bounded positive densities with interior Sobolev perturbations. At inverse degree $1/2$, estimation incurs the factor $(\log M)^{1/2}$ beyond $M^{-1/2}$. At inverse degree $3/4$, testing incurs the factor $(\log M)^{1/4}$, where M is the effective endpoint sample size. Multiscale wavelet alternatives establish both factors for the physical observations and W_1 loss. A weighted Fourier comparison transfers the critical estimation lower bound to classical fixed-error deconvolution under the corresponding Fourier and moment assumptions.

7.1 Model and rates

A known X has density g on a fixed interval $[L, R]$, smooth and positive on the interior. Near its two endpoints write the density as $w_{\pm}(h)h^{\alpha_{\pm}}$, where h is depth measured from that endpoint, $w_{\pm}$ are smooth strictly positive up to zero and $\alpha_{\pm} > -1$. We observe N independent copies of $Y = X + bZ$, with Z independent of X. Set

$$r = \min(\alpha_-, \alpha_+) + 1, \qquad M = Nb^r.$$

Fix a known density f_0 on $[-1, 1]$ bounded between two strictly positive constants. The unknown density $f = f_0 + u$ satisfies $c \le f \le C$ on $[-1, 1]$, where the bounds have strict slack around those of f_0. The mass-zero perturbation u is supported in $[-a, a]$ for a fixed $0 < a < 1$, with H^s norm at most a fixed $A > 0$, $s \ge 0$, on the torus of length eight containing $[-1, 1]$. All transport distances are Euclidean. Let $T_b u(y) = \int g(y - bz)u(z)\, dz$.

Theorem 7.1. *For $1 \le p \le 2$, uniformly at sufficiently small $b > 0$ and as $M \to \infty$, the minimax expected W_p loss is of order*

$$\begin{cases} M^{-1/2}, & 0 < r < 1/2, \\ M^{-1/2}(\log M)^{1/2}, & r = 1/2, \\ M^{-(s+1)/(2s+1+2r)}, & r > 1/2. \end{cases}$$

For testing $f = f_0$ against $W_p(f, f_0) \ge \rho$ with each error at most $1/4$, the minimax separation radius is of order

$$\begin{cases} M^{-1/2}, & 0 < r < 3/4, \\ M^{-1/2}(\log M)^{1/4}, & r = 3/4, \\ M^{-(s+1)/(2s+1/2+2r)}, & r > 3/4. \end{cases}$$

Both lower bounds also hold if any fixed finite number of noise moments are prescribed to equal those of f_0.

In fixed-scale ordinary-smooth deconvolution, Dedecker, Fischer and Michel [29, Theorems 3.1 and 4.1] give a W_1 upper bound with a square-root logarithm at degree $1/2$ and a root-sample-size lower bound at equality. The Sobolev W_1 lower bound of Rousseau and Scricciolo [32, Theorem 5.1] also gives the root rate at this endpoint. For testing in the independent Gaussian sequence model, the heterogeneous-variance bounds of Laurent, Loubès and Marteau [30] exhibit the fourth-power variance sum. At degree $3/4$ its harmonic divergence produces the testing logarithm. Here the data come from finite-scale shifts of a compact density, and the unknown noise itself is compressed by b. The proof derives exact endpoint probes and matches both logarithms with alternatives separated in the CDF loss W_1. The critical lower bounds also apply to fixed ordinary-smooth errors under a weighted Fourier condition. In the published Archive Gaussian model [14], the inverse instead has a logarithmic modulus because the known Gaussian multiplier decays exponentially. The higher-dimensional endpoint analogue and the moment visibility hierarchy are treated in Section 6.

7.2 One-sided probes and a weighted information bound

Let I^r denote the left Riemann–Liouville integral with kernel $t^{r-1}\mathbf{1}_{t>0}/\Gamma(r)$ on a fixed interval $[0, 3]$. Reflect the variables if necessary so that the right endpoint has order $\alpha = r - 1$. Near this endpoint introduce $x = (R + b - y)/b$, and write the noise coordinate as $z = 1 - v$. For $0 \le x \le 3$ and $0 \le v \le 2$, the depth of $y - bz$ from the right endpoint is $b(x - v)$. Hence, after division by b^α, the output on the window is exactly the left convolution of $u(1 - v)$ with

$$t^{r-1} w_+(bt)\mathbf{1}_{t>0}, \qquad t = x - v.$$

It suffices to take b small enough for this interval to lie within the power-law region whenever the depth is positive.

On $[0, 3]$, write $w_+(bt) = w_+(0) + btA_b(t)$ with uniform smooth bounds. The causal fractional derivative of $k_b(t) = t^r A_b(t)$ is bounded uniformly there. Indeed choose integer $m > r$. Applying I^{m-r} to k_b gives t^m times a smooth bounded integral over a fixed unit interval by the change of variable $s = tv$. Its mth derivative is bounded. Denote it by ℓ_b; the first $m - 1$ initial derivatives vanish, so $I^r \ell_b = k_b$, by the fractional convolution semigroup. Thus the output operator in scaled coordinates is

$$b^\alpha \Gamma(r) w_+(0)\, I^r (I + bV_b),$$

where V_b is a causal convolution of uniformly bounded L^2 norm, by Young's inequality. It commutes with I^r. In particular $I + bV_b$ has a uniformly bounded inverse on $L^2([0,3])$ for small b.

For a smooth trigonometric test $\phi(z)$ choose a fixed smooth cutoff extension $\psi(v)$ of $\phi(1-v)$ on $[0,2]$, vanishing past $v = 3$. Its $H^r(\mathbb{R})$ norm is at most a constant times the torus H^r norm of ϕ. Under the adjoint, use the right fractional derivative of ψ, followed by $(I + bV_b^*)^{-1}$. The right fractional integral inverts this derivative on the interval because the test vanishes near the upper end. Its L^2 derivative norm is at most $C\|\psi\|_{H^r(\mathbb{R})}$ from its Fourier multiplier of modulus $|\xi|^r$. These statements follow from the one-sided Gamma convolution and its semigroup, or from the exponentially damped formula for the multiplier followed by a limit. Commutation of the right convolution operators gives an exact dual probe. Changing back to the physical observation scale and dividing by $b^\alpha \Gamma(r) w_+(0)$ gives a function a_ϕ supported on the observed window with

$$\int a_\phi T_b u = \int \phi u, \qquad \|a_\phi\|_{L^2(dy)} \le C b^{-\alpha - 1/2} \|\phi\|_{H^r}.$$

The probe is linear in ϕ. The observation density on this window is at most Cb^α because f is bounded and $\int_0^3 t^\alpha dt$ is finite. Therefore the centered observation $a_\phi(Y) - \int a_\phi T_b f_0$ is unbiased and has variance at most $Cb^{-r}\|\phi\|_{H^r}^2$.

For the lower bounds we also need an estimate using the full data. Fix an integer K large enough that $2K + 1 > \max(\alpha_-, \alpha_+)$. For signals u supported on $[-a, a]$ with moments through K zero,

$$\int \frac{|T_b u|^2}{T_b f_0} \le C b^r \|u\|_{H^{-r}(\mathbb{R})}^2. \tag{5}$$

To see this on the two fixed scaled endpoint windows, the convolution kernel after rescaling has the form $t_+^{\alpha_\pm} w_\pm(bt)$ times a smooth cutoff in a fixed bounded interval. Its Fourier transform is bounded by $C\langle\xi\rangle^{-1-\alpha_\pm}$ uniformly in b, with $\langle\xi\rangle = (1 + |\xi|^2)^{1/2}$. For example split into dyadic intervals near zero, integrate by parts more than $1 + \alpha_\pm$ times on each interval, and sum $Ch^{1+\alpha_\pm} \min(1, (h|\xi|)^{-m})$ over dyadic lengths h. The remaining smooth part is estimated by integration by parts. Thus the scaled output is bounded in L^2 by $C\|u\|_{H^{-1-\alpha_\pm}}$. On the part where $T_b u$ can be nonzero, the baseline is bounded below by $c' b^{\alpha_\pm}$: a fixed interval of noise values deeper into the permitted endpoint direction puts the latent argument at depth comparable to b. Change variables back from the endpoint window. Its cost in (5) is at most $Cb^{1+\alpha_\pm}\|u\|_{H^{-1-\alpha_\pm}}^2$. At physical interior depths $h \ge Tb$ from either endpoint, Taylor subtraction through K and compact-support Sobolev duality bound $|T_b u| \le C b^{K+1} h^{\alpha_\pm - K - 1} \|u\|_{H^{-r}}$ where the endpoint representation holds. The cutoff Taylor remainder has this bound in H^r because derivatives in the noise variable contribute b/h and $h \ge Tb$. There $T_b f_0 \ge c' h^{\alpha_\pm}$. Integration gives $Cb^{1+\alpha_\pm}\|u\|_{H^{-r}}^2$. The deep interior cost is $O(b^{2K+2})\|u\|_{H^{-r}}^2$; outside the windows and the interior the signal is zero. Since $1 + \alpha_\pm \ge r$, this proves (5). Baseline observation densities remain uniformly comparable across the whole admissible class because f/f_0 is bounded above and below.

7.3 Upper bounds and the noncritical lower bounds

With the linear probes of the previous section, reconstruct Fourier coefficients through m from sample means. The squared H^{-1} risk is bounded by

$$C\left[m^{-2(s+1)} + M^{-1} \sum_{|k| \le m} (1 + |k|)^{2r-2}\right].$$

On this class the H^{-1} norm controls W_2. In one dimension, for a mass-zero supported difference u, its cumulative integral P_u is supported on the same interval and $\|P_u\|_2 \le C\|u\|_{H^{-1}(\mathbb{T})}$ by the

periodic zero-mode comparison and the common support. Indeed its mean can be controlled by its centered part because it vanishes on a fixed positive fraction of the torus. The current P_u, together with the path of linearly interpolated densities bounded below on the interval, gives the transport bound by the dynamic formula. Project the preliminary estimate to a countable dense subset of the class with H^{-1} tolerance of order m^{-s-1} to get a measurable density estimate. For $r < 1/2$ the sum is bounded, and at $r = 1/2$ it is $O(\log m)$. Taking m of order $M^{1/(2(s+1))}$ gives the first two rates. For $r > 1/2$ the sum is $O(m^{2r-1})$; balancing the terms gives the third.

For testing, take the H^{-1}-weighted coefficient vector of the individual probes through frequency m. Its mean is the truncated coefficient vector θ of u, and its covariance satisfies $\Sigma \preceq Cb^{-r}\,\mathrm{diag}((1+|k|)^{2r-2})$. The unbiased statistic averaging inner products on distinct sample pairs has mean $|\theta|^2$. The variance equals $4\theta^\top \Sigma\theta/N + 2\,\mathrm{tr}(\Sigma^2)/(N(N-1))$. The standard deviation is bounded by

$$C\left[M^{-1}\left(\sum_{|k|\le m}(1+|k|)^{4r-4}\right)^{1/2} + M^{-1/2}m^{(r-1)_+}|\theta|\right].$$

If $r < 3/4$ the first sum is bounded and if $r = 3/4$ it is $O(\log m)$. With m as in the root estimation case, the squared Sobolev tail is $O(M^{-1})$. For $r > 3/4$ take $m \asymp M^{1/(2s+1/2+2r)}$; the squared tail and first standard deviation then have the same order. In all cases the W_2 bound implies that a sufficiently separated W_p alternative has truncated mean norm exceeding a large constant times the claimed separation. The linear standard deviation is smaller than the corresponding mean squared by a constant that can be made arbitrarily small by increasing this separation constant. For $r > 3/4$ it even tends to zero relatively, by the factor $m^{-1/4}$ when $r \ge 1$ and $m^{3/4-r}$ otherwise. A threshold and Chebyshev's inequality give both error bounds.

Below each respective critical degree, a fixed nonzero smooth compact moment-cancelled direction at amplitude $cM^{-1/2}$ is a two-point lower bound by (5). Above each critical degree, the single-resolution construction can be used in dimension one. For clarity, take $J_h \asymp h^{-1}$ disjoint translated L^2-normalized smooth bumps p_ℓ of width h with vanishing moments through a sufficiently large order. Local polynomial Poincaré followed by interpolation gives $\|\sum_\ell c_\ell p_\ell\|_{H^{-r}} \le Ch^r|c|$. At amplitude $\gamma = \varepsilon h^{s+1/2}$, the signs yield admissible perturbations, and the CDFs or signed Lipschitz tests give W_1 separation $c\gamma h^{3/2}$ times Hamming distance. Neighboring product Hellinger distances squared are at most $CM\gamma^2h^{2r}$. At $h \asymp M^{-1/(2s+1+2r)}$, the coordinate error lower bound of Assouad gives risk of order h^{s+1}. For testing the Gram matrix of the normalized signals in the output norm (5) satisfies $G \preceq Cb^r h^{2r} I$. Thus the independent sign mixture at $h \asymp M^{-1/(2s+1/2+2r)}$ has chi-square divergence at most $\exp(CM^2h^{4s+1+4r}\varepsilon^4) - 1$. Indeed conditioning on either sign vector and then applying the Gaussian integral for its squared norm bounds the likelihood second moment by $\det(I - (N\gamma^2 G)^2)^{-1/2}$; the same argument is spelled out below for the critical case. Small ε makes the divergence small, while every sign alternative has W_1 separation of order h^{s+1}.

7.4 Critical lower bounds across resolutions

For both critical arguments use compactly supported real orthonormal wavelets $\psi_{j,k}$ on $\mathbb{R}$ of sufficiently high regularity and with sufficiently many vanishing moments. Use only those with support inside a fixed interval in $(-a, a)$, at levels $j_0 \le j \le J$ with j_0 fixed. There are $\asymp 2^j$ at each level and bounded overlap at a given level. Use levels $J = \lfloor c_0 \log_2 M \rfloor$ with a fixed small $c_0 > 0$. For independent signs $\omega_{j,k}$ let the perturbation and its primitive be

$$f_\omega - f_0 = P'_\omega, \qquad P_\omega = \sum_{j,k} \delta_j \omega_{j,k}\psi_{j,k}.$$

We use the standard orthonormal wavelet characterization of H^t for $0 \le t \le s+1$, choosing the wavelet regularity accordingly [28, 31]. In particular the H^{-r} norm of P'_ω is at most a constant times the weighted coefficient norm in H^{1-r} for the two critical values $r < 1$. Integration by parts shows that P'_ω has all the required vanishing moments once the wavelets have sufficiently many. In each construction below $\delta_j \le \varepsilon M^{-1/2} 2^{-j/2}$. The H^s squared bound for P'_ω follows from $\sum_j 2^j \delta_j^2 2^{2j(s+1)}$. The uniform bound follows from bounded overlaps and $\|\psi'_{j,k}\|_\infty = O(2^{3j/2})$. With $c_0 < 1/(2(s+1))$, both are uniformly small as $M \to \infty$, so all densities are admissible.

7.4.1 Estimation at degree $1/2$

Take $\delta_j = \varepsilon M^{-1/2} 2^{-j/2}$. For neighboring sign vectors, the product squared Hellinger distance is at most $CM\delta_j^2 2^{2j(1-r)} \le C\varepsilon^2$. Under the uniform sign prior, draw ω, ω' independently from the posterior conditional on the data. For each bit, the expected posterior disagreement probability is bounded below by a positive constant. Indeed conditioning additionally on all the other bits gives a two-point experiment with total variation bounded away from one. Its average bit error is bounded below, hence so is its posterior disagreement ($2p(1-p) \ge \min(p, 1-p)$). Removing this extra conditioning can only increase the expected conditional variance. Orthonormality of the primitives now gives

$$\mathbb{E} \int |P_\omega(x) - P_{\omega'}(x)|^2 dx \ge c \sum_j 2^j \delta_j^2 \ge c' \varepsilon^2 J/M.$$

For the fourth moment, each of the two posterior draws has the original independent sign distribution marginally after averaging over the data. At each x, the sum of squared pointwise coefficients is at most $C \sum_j \delta_j^2 2^j \le C' \varepsilon^2 J/M$ by overlap and the wavelet supremum bound. The fourth-moment bound for a sum of independent signs and the triangle inequality give $\mathbb{E} \int |P_\omega - P_{\omega'}|^4 \le C(\varepsilon^2 J/M)^2$. Interpolate L^1, L^2, L^4 on the joint space of x, data and posterior draws. This yields $\mathbb{E} \int |P_\omega - P_{\omega'}| \ge c'' \varepsilon \sqrt{J/M}$. The integral is $W_1(f_\omega, f_{\omega'})$. The posterior triangle inequality says that its Bayes expectation is at most twice the risk of any estimator. This proves the critical estimation lower bound.

7.4.2 Testing at degree $3/4$

Take $\delta_j = \varepsilon M^{-1/2} J^{-1/4} 2^{-j/2}$. In the sign-mixture likelihood second moment, let B_N be N times the Gram matrix of the scaled directions $\delta_j \psi'_{j,k}$ in the output norm (5). The joint wavelet bound gives

$$0 \preceq B_N \preceq CM \operatorname{diag}(\delta_j^2 2^{2j(1-r)}).$$

At $r = 3/4$, this gives $\|B_N\|_{\mathrm{op}} \le C\varepsilon^2$ and

$$\operatorname{tr}(B_N^2) \le CM^2 \sum_j 2^{2j} \delta_j^4 \le C' \varepsilon^4.$$

For each pair of sign vectors the likelihood inner product is bounded by $e^{\omega^\top B_N \omega'}$ using $1 + x \le e^x$. Conditioning on ω and using $\cosh x \le e^{x^2/2}$ bounds the sign average by $\mathbb{E} e^{\|B_N \omega\|^2/2}$. Write this exponential as an expectation of $e^{G^\top B_N \omega}$ with G standard Gaussian, apply the same hyperbolic-cosine bound to the remaining signs, and integrate $e^{\|B_N G\|^2/2}$. This gives the stated determinant bound, at most $\exp(C \operatorname{tr} B_N^2)$ for $\|B_N\|_{\mathrm{op}} \le 1/2$. With small ε the mixture is close in total variation to the null.

It remains to obtain deterministic separation within the mixture. By the lower Khintchine inequality,

$$\mathbb{E}\|P_\omega\|_1 \geq c \int \left(\sum_{j,k} \delta_j^2 |\psi_{j,k}(x)|^2 \right)^{1/2} dx \geq c' \varepsilon M^{-1/2} J^{1/4}.$$

For the last bound the integral of the sum of squares is $\asymp \varepsilon^2 \sqrt{J}/M$ by orthonormality, and its pointwise bound is at most a constant times this value by overlap. Each sign flip changes $\|P_\omega\|_1$ by at most $C\delta_j 2^{-j/2}$. The sum of squared changes is $O(\varepsilon^2/(M\sqrt{J}))$, by summing over the $O(2^j)$ signs at each level. The bounded differences variance inequality therefore implies that the proportion of signs with less than half the displayed order of separation tends to zero. Condition the prior on the remaining signs. Its observation mixture changes by $o(1)$ in total variation, so the testing lower bound still follows. Since $J \asymp \log M$ and $W_p \geq W_1$, both critical bounds prove Theorem 7.1.

7.5 Fixed-scale gamma errors

Consider the ordinary-smooth model $Y = Z + X$ with the same class of densities for Z, and independent known gamma error of density $g(x) = x_+^{r-1} e^{-x}/\Gamma(r)$ for $x > 0$ and zero otherwise, with $r > 0$. No noise scale shrinks in this model.

Theorem 7.2. *All the estimation and testing rates of Theorem 7.1 hold for the gamma error experiment with $M = N$. In particular the critical W_1 estimation loss for shape $r = 1/2$ has order $\sqrt{\log N/N}$ on the stated positive compact density class, and the critical W_1 testing separation at $r = 3/4$ has order $N^{-1/2}(\log N)^{1/4}$.*

Proof. Use the observation window $-1 \leq y \leq 2$, writing $x = y + 1$ and $v = z + 1$. The window operator is exactly $e^{-x} I^r$ applied to $e^v u(v - 1)$. The positive multipliers and their inverses are smooth and bounded on the window. The same right fractional derivative gives exact probes of L^2 norm $O(\|\phi\|_{H^r})$, with uniformly bounded observation density there by $\|f\|_\infty$. This proves the upper bounds with $M = N$.

To prove (5) with its factor b^r replaced by one, write $q_0 = g * f_0$. On a fixed bounded observation interval, the signal can be nonzero only for $y \geq -a$. For $y \in [-a, 3]$, integrating over a fixed interval of signal values near -1 places $y - z$ in a compact subset of the positive half-line, uniformly in y. Thus q_0 is bounded below on the relevant region. Plancherel and the gamma characteristic multiplier $(1 + i\xi)^{-r}$ bound the squared output norm by $C\|u\|_{H^{-r}}^2$. For $y > 3$, $q_0(y) \asymp e^{-y} y^{r-1}$. The H^r norm in z of the convolution kernel, smoothly cut off on a fixed neighborhood of $[-a, a]$, is $O(e^{-y} y^{r-1})$ by differentiating a fixed number of times. By duality the weighted squared output on the tail is bounded by $C\|u\|_{H^{-r}}^2 \int_3^\infty e^{-y} y^{r-1}\, dy$. The left tail of the difference is zero. This gives the same weighted information bound, even without moment cancellation. All the lower bound constructions therefore apply. □

This fixed-scale conclusion concerns the one-sided gamma subfamily of ordinary-smooth errors. The exact Laplace conjugation and the weighted tail estimate are what make the logarithmic lower bounds applicable to that family.

7.6 The critical lower bound for fixed ordinary-smooth errors

The same wavelet alternatives give a lower bound under the Fourier derivative condition of Dedecker, Fischer and Michel [29, Theorem 4.1], using their polynomial-tailed reference family. This fills the logarithmic gap between the printed estimation bounds at degree 1/2 on the moment classes specified below.

Theorem 7.3. *Fix $r > 0$ and an integer $m \geq 2$. Let Q be a known fixed error law with characteristic function φ. Suppose that*

$$|\varphi^{(\ell)}(\xi)| \leq C\langle\xi\rangle^{-r}, \qquad 0 \leq \ell \leq m.$$

In the model $Y = Z + X$, $X \sim Q$, consider the density $f_0(x) = c_m(1+x^2)^{-m}$ for Z and a class of sufficiently small, compactly supported, mass-zero Sobolev perturbations of f_0 as above. The estimation and testing lower rates of Theorem 7.1 hold with $M = N$. If also $|\varphi(\xi)| \geq c\langle\xi\rangle^{-r}$, these are sharp estimation and testing rates on this perturbation class.

Under the displayed characteristic-function derivative bounds, take $r = 1/2$, $2 < q < 2m-1$, and fix any positive q-moment bound for Z. If the known error law has finite qth moment and its reciprocal characteristic function obeys the upper-bound hypotheses of [29, Theorem 3.1], the minimax expected W_1 estimation rate over that moment-bounded class is $\sqrt{\log N/N}$. For $2 < q < 3$, take $m = 2$ so the lower bound uses only the two derivatives in their Theorem 4.1.

Proof. Let $q_0 = f_0 * Q$. Some bounded interval carries positive Q-mass, so $q_0(y) \geq c(1+y^2)^{-m}$ uniformly in y. For any signed u supported on a fixed compact interval, Plancherel and the Leibniz rule give

$$\begin{aligned}\int \frac{|u * Q|^2}{q_0} &\leq C' \sum_{\ell=0}^{m} \|\partial_\xi^\ell(\widehat{u}\,\varphi)\|_2^2 \\ &\leq C''\|u\|_{H^{-r}(\mathbb{R})}^2.\end{aligned}$$

For the second bound, a derivative of $\widehat{u}$ is the transform of u multiplied by a power of the spatial coordinate. This is a bounded multiplier on H^{-r} after cutting it off smoothly on the common support. Every remaining factor $\varphi^{(\ell)}$ has the assumed decay. This is exactly (5) at effective sample size N, without moment cancellation. All the lower constructions use only compact interior perturbations, so their densities are admissible around the positive polynomial reference. The W_1 discrepancies still equal the L^1 norm of the primitive.

For sharp upper rates on the perturbation class, use the Fourier inverse of the adjoint convolution on each smooth compactly localized coefficient test. Its L^2 norm is bounded by $C\|\phi\|_{H^r}$ by the lower bound on $|\varphi|$. The observation densities are uniformly bounded since the noise law is a probability and the target densities are uniformly bounded. Thus the same centered-observation and joint covariance bounds hold as for the endpoint probes at $M = N$. Outside the common perturbation interval the densities agree, so the cumulative current is zero there. On the interval the positive density floor gives the same W_2 transport bound. This proves the upper rates.

For the moment-bounded conclusion, dilate the reference toward zero if necessary and choose the compact perturbation interval inside its positive region. Since $q < 2m-1$, its qth moment can be made strictly smaller than the moment bound with fixed slack for all perturbations in the lower proof. The upper bound follows from [29, Theorem 3.1] at $p = 1$: the $q > 2$ moment bound for the target and the fixed error gives a uniform bound on $\int_0^\infty \sqrt{\mathbb{P}(|Y| > x)}\,dx$, and the reciprocal derivative hypotheses supply their other bounds. Thus the square-root logarithm is necessary on the stated moment-bounded classes. □

8 Recovering joint noise from corners of product densities

We observe a known product density on a box plus a small joint perturbation that is independent of the latent variable. A single corner gives exact linear probes for the unknown noise density, including its dependence. The coordinate boundary powers determine both the mixed Sobolev inverse and the effective sample size. We give sharp Wasserstein estimation and testing rates

for isotropic Sobolev classes. At the testing threshold a factor $(\log M)^{1/4}$ is necessary, including for W_1 loss in dimensions two and three. The same geometry determines the exact Poisson experiment at the critical corner sampling scale.

8.1 Model and results

Let $d \geq 2$. The known latent variable X has independent coordinates with densities g_i, each supported on an interval $[L_i, U_i]$, smooth and positive in its interior. Near the two endpoints $g_i = w_{i,\pm}(h)h^{\alpha_{i,\pm}}$, where h is inward distance, $w_{i,\pm}$ is smooth and strictly positive up to zero, and $\alpha_{i,\pm} > -1$. Observe independent samples of $Y = X + (b_1 Z_1, \dots, b_d Z_d)$ with X independent of Z and each $b_i > 0$ known. Let

$$r_i = 1 + \min(\alpha_{i,-}, \alpha_{i,+}), \qquad R_* = \sum_i r_i, \qquad \eta = \prod_i b_i^{r_i}, \qquad M = N\eta.$$

The coordinates of Z need not be independent. Write $Q = [-1, 1]^d$ and $T_b u$ for the signed output density of a signed integrable u in the Z coordinates.

Fix $0 < a < 1$ and a known reference density f_0 on Q, bounded above and below by positive constants. The candidate density $f = f_0 + u$ satisfies $c \leq f \leq C$ on Q with strict slack around the baseline bounds, and u is mass-zero, supported in aQ, with H^s norm at most L for fixed $s \geq 0, L > 0$ on the torus of side length eight. Distances W_p use the Euclidean metric.

Theorem 8.1. *Uniformly at sufficiently small $b_i > 0$, the minimax expected W_p risk for $1 \leq p \leq 2$ has order*

$$\min\{1, M^{-(s+1)/(2s+d+2R_*)}\}.$$

As $M \to \infty$, the separation for testing $f = f_0$ with both errors at most $1/4$ has order

$$\begin{cases} M^{-1/2}, & d + 4R_* < 4, \\ M^{-1/2}(\log M)^{1/4}, & d + 4R_* = 4, \\ M^{-(s+1)/(2s+d/2+2R_*)}, & d + 4R_* > 4. \end{cases}$$

The deterministic W_p modulus over pairs in the class at output Hellinger distance at most $\sqrt{\eta}\,\delta$ has order $\delta^{\min\{1,(s+1)/(s+R_*)\}}$ as $\delta \downarrow 0$, uniformly for small b_i. The lower bounds still hold if all proper coordinate marginals are prescribed to equal those of f_0, and even if the perturbation must have zero coordinate moments through any fixed order on every coordinate slice.

The mixed-cumulative inversion formula for uniform bivariate errors is classical [11]. The uniform polytope experiment Section 4 shows how arbitrary joint noise can be recovered at vertices without known lower margins. Here individual endpoint tapers may have different exponents and each observation scale may differ; the exact product corner probes capture these choices. Comte and Lacour [33] study anisotropic density recovery with fixed errors. Rousseau and Scricciolo [32] treat W_1 in a fixed-error model with independent error coordinates. The anisotropic W_1 result of [35, Section 8] assumes product target laws; here the target's dependence is part of the object to recover. The critical testing lower bound uses the potential-to-transport bound of Park, Moon and Kim [34, Section 4], which they apply to direct multiscale estimation. We use it with a weighted likelihood mixture in the physical corner experiment. In the published Archive Gaussian model [14], fixed Gaussian smoothing instead leads to a logarithmic inverse modulus.

8.2 Tensorized observation bounds

For $\mathbf{a} = (a_1, \dots, a_d)$, use $H^{\mathbf{a}}_{\mathrm{mix}}$ for the Hilbert norm with Fourier square weight $\prod_i (1 + \xi_i^2)^{a_i}$ on $\mathbb{R}^d$, or the corresponding product weight over frequencies on the torus. This is the Hilbert tensor product of the corresponding one-dimensional Sobolev spaces.

For each coordinate take a fixed-width scaled observation window at the endpoint achieving r_i, as in Section 7. There is a linear map taking a scalar trigonometric test ϕ_i to a probe a_i supported on the window, with

$$\int a_i T_{b_i,i} u_i = \int \phi_i u_i, \qquad \|a_i\|_2 \le C b_i^{1/2-r_i} \|\phi_i\|_{H^{r_i}}.$$

To recall its origin, in scaled coordinates the operator is $b_i^{r_i-1}$ times a causal fractional convolution of order r_i. The endpoint coefficient factors into a positive constant times I^{r_i} followed by $I + b_i V_i$ for a uniformly bounded causal convolution V_i. Invert the perturbation by a Neumann series and use the right fractional derivative on a compact cutoff extension of the test. The right integral then exactly reproduces the test on the noise interval. This gives the displayed physical L^2 bound after change of variables. No moment restrictions are used in the scalar identity.

Tensoring the maps gives a joint linear probe a_ϕ for any trigonometric polynomial ϕ, supported on the product window, with

$$\int a_\phi T_b u = \int \phi u, \qquad \|a_\phi\|_2 \le C \left(\prod_i b_i^{1/2-r_i} \right) \|\phi\|_{H^{\mathbf{r}}_{\mathrm{mix}}}.$$

The operator norm bound follows from the tensor-product Hilbert norm. Exactness for test products follows by Fubini and the one-dimensional exact identity applied to each noise coordinate in the kernel pairing. Finite sums give the stated class of tests. On the product window the output density is bounded by $C \prod_i b_i^{r_i-1}$, by bounding the joint noise density by a constant and integrating the separate kernels. This is valid without any independence of the noise. Hence the centered evaluation at Y gives an unbiased probe of $\int \phi u$ with variance at most $C\eta^{-1} \|\phi\|^2_{H^{\mathbf{r}}_{\mathrm{mix}}}$. The density bound also gives by duality, on the torus,

$$\|f - g\|_{H^{-\mathbf{r}}_{\mathrm{mix}}} \le C\eta^{-1/2} H(T_b f, T_b g), \qquad H^2(p, q) = \int (\sqrt{p} - \sqrt{q})^2. \tag{6}$$

Indeed pair the output difference with the probe, use the square root identity, and use the product-window density bound.

For lower bounds fix an integer K so large that $2K + 1 > \max_{i,\pm} \alpha_{i,\pm}$. Consider tensor signals, finite sums of products of compactly supported functions of the separate coordinates, each scalar factor supported in $[-a, a]$ and with moments through K zero. For such signals,

$$\int \frac{(T_b u)^2}{T_b f_0} \le C\eta \, \|u\|^2_{H^{-\mathbf{r}}_{\mathrm{mix}}(\mathbb{R}^d)}, \qquad 0/0 := 0. \tag{7}$$

For each coordinate the scalar output divided by the square root of the observed uniform-noise reference density has norm at most $C b_i^{r_i/2} \|u_i\|_{H^{-r_i}(\mathbb{R})}$ on this scalar moment kernel, by the one-dimensional weighted bound Section 7. That bound uses both endpoint windows and a Taylor subtraction away from each endpoint, allowing asymmetric powers. The physical denominator for any f_0 here is bounded below by a constant times the product of the scalar reference densities, since f_0 is bounded below on Q. Tensor the bounded scalar operators on the closures of the moment-kernel subspaces in the negative Sobolev norms. Finite sums of the permitted products

belong to this tensor product, so the squared norm bound yields (7). Again the result does not require an independent joint baseline. All alternatives used below are finite sums of the permitted products. Their coordinate cancellations in particular preserve every proper marginal. Any fixed larger number of moments can be preserved by increasing the scalar cancellation orders.

8.3 Upper bounds

The mixed norm in (6) controls the isotropic H^{-R_*} norm since $\prod_i(1+\xi_i^2)^{r_i} \le (1+|\xi|^2)^{R_*}$. On densities bounded below on Q, the H^{-1} torus norm of the difference controls W_2 by solving the mean-zero Neumann variational problem on Q. Poincaré and bounded extension from $H^1(Q)$ to the torus control the gradient norm, and the linear interpolation of the densities with this current gives the transport bound. Interpolation with the H^s class bound yields the population upper when $R_* > 1$. When $R_* \le 1$, the direct comparison to H^{-1} gives the linear upper.

Write $v_j = \langle j\rangle^{-2}\prod_i \langle j_i\rangle^{2r_i}$ for Fourier indices $j \in \mathbb{Z}^d$. For a cutoff m, the joint covariance of the centered H^{-1}-weighted individual coefficient probes satisfies $\Sigma \preceq C\eta^{-1}\operatorname{diag}(v_j)$. Estimating the coefficients by sample averages and truncating gives squared H^{-1} risk at most

$$C\{m^{-2(s+1)} + M^{-1}m^{d+2R_*-2}\}.$$

Indeed $v_j \le C\langle j\rangle^{2R_*-2}$ and $d + 2R_* - 2 > 0$. Optimize m and project to a countable H^{-1}-dense subset of the candidate densities, choosing the first within the tail scale of the infimum. The triangle inequality and the W_2 bound give the estimation upper. A constant estimator is sufficient for $M < 1$.

For testing use the average inner product on distinct pairs of these weighted probe vectors, centered by the reference. If θ is the truncated weighted mean, its expectation is $|\theta|^2$, and its variance is $4\theta^\top\Sigma\theta/N + 2\operatorname{tr}(\Sigma^2)/(N(N-1))$. The two standard deviation bounds are

$$CM^{-1/2}m^{(R_*-1)_+}|\theta|, \qquad CM^{-1}\Big(\sum_{|j|\le m} v_j^2\Big)^{1/2}.$$

The latter sum is bounded for $d+4R_* < 4$, is $O(\log m)$ at equality, and is $O(m^{d+4R_*-4})$ otherwise, by the radial upper bound on v_j. In the first two cases $m \asymp M^{1/(2s+2)}$ makes the squared truncation tail $O(M^{-1})$. In the last case $m \asymp M^{1/(2s+d/2+2R_*)}$ balances the squared tail with the second standard deviation. For an alternative at a large constant times the claimed separation, the transport upper bound and Sobolev tail give the same lower order on $|\theta|$. The linear standard deviation is smaller than the mean squared by an arbitrarily small constant after increasing the separation constant. In the supercritical case its ratio even vanishes by $m^{-d/4}$ if $R_* \ge 1$, and by $m^{1-R_*-d/4}$ otherwise. A threshold and Chebyshev's inequality yield the stated testing upper.

8.4 Single-resolution lower bounds

Let p_i be smooth compact real nonzero scalar bumps on $[-1/4, 1/4]$ with moments through K zero. In an interior grid form L^2-normalized translated bumps $p_\ell(z) = h^{-d/2}\prod_i p_i((z_i - z_{\ell,i})/h)$. There are $J_h \asymp h^{-d}$ disjoint cells. The map from a scalar coefficient vector in one direction to a linear combination of its scalar translates has norm $O(h^{r_i})$ into H^{-r_i}: for an integer $k > r_i$ use local polynomial subtraction on each cell to bound the H^{-k} norm by Ch^k times the coefficient norm, and interpolate with the disjoint L^2 bound. Increase K when necessary. Tensor these maps to obtain

$$\Big\|\sum_\ell a_\ell p_\ell\Big\|_{H^{-\mathbf{r}}_{\mathrm{mix}}} \le Ch^{R_*}|a|.$$

Use alternatives $f_\omega = f_0 + \gamma \sum_\ell \omega_\ell p_\ell$ with $\gamma = \varepsilon h^{s+d/2}$ for a small fixed $\varepsilon > 0$. Disjoint supports and derivative scaling give positivity and the isotropic H^s bound by interpolation. A signed sum of tests $h^{1+d/2} p_\ell$ has bounded Lipschitz constant and gives W_1 separation at least $c\gamma h^{1+d/2}$ times Hamming distance.

The squared Hellinger discrepancy of neighbors in the physical experiment is at most $C\eta\gamma^2 h^{2R_*}$ by (7). Take h a small constant times $\min\{1, M^{-1/(2s+d+2R_*)}\}$. Neighboring product total variations are then bounded away from one. Mapping an arbitrary estimator to a nearest hypercube vertex and averaging the coordinate errors gives expected loss at least $c\gamma h^{1+d/2} J_h$, of the claimed order.

For later use, a uniformly signed family with scaled output Gram matrix B_N, defined as N times the Gram in (7), has mixture chi-square divergence at most $\exp(C \operatorname{tr} B_N^2) - 1$ if $\|B_N\|_{\mathrm{op}} \le 1/2$. In fact the product likelihood inner product is bounded by $\exp(\omega^\top B_N \omega')$. Condition on one sign vector, use $\cosh x \le e^{x^2/2}$, express the resulting squared-norm exponential as a Gaussian integral, and apply the same cosine bound to the other sign vector. The Gaussian integral is $\det(I - B_N^2)^{-1/2}$, giving the stated estimate. For the single-resolution family, $0 \preceq B_N \preceq CM\gamma^2 h^{2R_*} I$. At $h \asymp M^{-1/(2s+d/2+2R_*)}$ its squared Hilbert–Schmidt norm is bounded by $C\varepsilon^4$, and the alternatives are W_1-separated from f_0 by order h^{s+1}. This proves the supercritical testing lower. In the subcritical case a single fixed nonzero moment-cancelled tensor bump at amplitude $cM^{-1/2}$ is a matching two-point lower.

For the population lower when $R_* > 1$, any sign vector at resolution h has output Hellinger distance at most $C\sqrt{\eta}\, h^{s+R_*}$ from the reference by (7), and has W_1 distance at least ch^{s+1}. Take $h \asymp \delta^{1/(s+R_*)}$ with a small prefactor. For $R_* \le 1$ use a fixed bump of amplitude $c\delta$. All the lower bounds apply to $W_p \ge W_1$.

8.5 Critical testing across scales

Suppose $d + 4R_* = 4$, which can occur in dimensions two and three. Let $\psi_{j,k}$ be fully high-pass tensor products of one-dimensional compact orthonormal wavelets, at the same level j in every direction. Choose enough smoothness and vanishing moments to cover all Sobolev orders and slice moments here. Use only translates supported in a fixed interior cube, with $j_0 \le j \le J = \lfloor c_0 \log_2 M \rfloor$ and fixed j_0. At each level there are $\asymp 2^{dj}$ functions, with uniformly bounded overlap. Set

$$U_\omega = \sum_{j,k} \delta_j \omega_{j,k} \psi_{j,k}, \qquad u_\omega = -\Delta U_\omega, \qquad \delta_j = \varepsilon M^{-1/2} J^{-1/4} 2^{-j(d+2)/2}.$$

Each summand of the Laplacian is a product of scalar factors with the required moments, by integration by parts. Isotropic wavelet bounds give $\|u_\omega\|_{H^s}^2 \le C \sum_j \delta_j^2 2^{dj} 2^{2j(s+2)}$. The pointwise bound follows by summing $C\delta_j 2^{j(d/2+2)}$ using bounded overlap. Both are small for $c_0 < 1/(2s+2)$ and ε small, so the densities are admissible.

For the directional Laplacian term $\partial_i^2 U$, bound its negative mixed norm by the mixed norm of U of orders $2 - r_i$ in the ith coordinate and $-r_\ell$ in every other coordinate. Tensorizing the scalar wavelet Sobolev characterization gives the coefficient norm with weight $2^{2j(2-R_*)}$ for the chosen equal levels [28, 31]. Summing over i and using (7) gives

$$0 \preceq B_N \preceq CM \operatorname{diag}\{\delta_j^2 2^{2j(2-R_*)}\}.$$

Its squared Hilbert–Schmidt norm is at most $C\varepsilon^4 J^{-1} \sum_j 2^{j(4-d-4R_*)} \le C'\varepsilon^4$. Its operator norm is small with ε. The likelihood mixture thus has small total variation distance from the null.

It remains to separate in W_1. The standard isotropic wavelet H^1 norm comparison and Poincaré on the fixed support give for every sign vector

$$\|\nabla U_\omega\|_2^2 \asymp E, \qquad E = \varepsilon^2 \sqrt{J}/M.$$

Indeed the sum of squared coefficient contributions is $\sum_j 2^{dj}\delta_j^2 2^{2j}$. By the fourth moment inequality for independent signs and bounded overlap, $\mathbb{E}\|\nabla U_\omega\|_4^4 \le CE^2$; the pointwise sum of squared coefficients has upper bound $C\sum_j \delta_j^2 2^{j(d+2)}$. The potential bound of [34, Section 4] states

$$\|(-\Delta U)\,dz\|_{\mathrm{KR}} \ge c_d \frac{\|\nabla U\|_2^3}{\|\nabla U\|_4^2} \quad \text{for } U \in C_c^2.$$

Here $\|\cdot\|_{\mathrm{KR}}$ is the supremum against 1-Lipschitz tests. It follows by cutting off the gradient at level λ via the Hardy–Littlewood maximal function, Lipschitz extending the potential from the good set, and optimizing λ. The bounded-support wavelets can be chosen C^2, as required. The norm is $W_1(f_0 + u_\omega, f_0)$ for the density differences. Markov's inequality for the fourth moment and the deterministic lower energy show $\mathbb{E}W_1(f_0 + u_\omega, f_0) \ge c\sqrt{E}$.

Changing a single sign affects W_1 by at most $C\delta_j\|\nabla\psi_{j,k}\|_1 \le C\delta_j 2^{j(1-d/2)}$, by the Kantorovich dual and integration by parts. The sum of squared changes is $O(\varepsilon^2/(M\sqrt{J}))$, since $\sum_j 2^{2j}\delta_j^2 \le C\varepsilon^2/(M\sqrt{J})$. The bounded differences variance bound implies that all but $o(1)$ of the signs have W_1 separation at least $c'\sqrt{E}$. Conditioning the prior on these signs changes its observed mixture by $o(1)$ in total variation. The mixture still cannot be separated from the null with both errors at most $1/4$. Since $J \asymp \log M$, this proves the critical testing lower and finishes Theorem 8.1.

8.6 Recovering a marginal from the full experiment

For a nonempty subset $S \subseteq \{1, \dots, d\}$, define $\eta_S = \prod_{i\in S} b_i^{r_i}$, $M_S = N\eta_S$ and $R_S = \sum_{i\in S} r_i$. The sharp estimation and testing rates for the noise marginal on S, using the full observations, are precisely the rates with dimension $|S|$, degree R_S, and effective sample size M_S. For a singleton use the one-dimensional rates in Section 7.

For the upper bounds, ignore the other coordinates. Projection of f_0 retains the positive bounds, and integration over the other coordinates retains the isotropic H^s bound, since it selects the zero Fourier frequencies in those coordinates. Thus the corresponding upper risk bound applies. For the lower bounds take any of the alternatives for the subset S, lift their perturbations to the full cube by multiplying by fixed smooth compact interior probability densities in the remaining coordinates, and add them to f_0. The marginal differences on S are unchanged. The full Sobolev norms are bounded by a fixed multiple of the subset norms, and with small amplitudes these densities stay admissible. The physical reference denominator is bounded below by the product of the scalar uniform-reference outputs. For each inactive coordinate the fixed probability lift is bounded by a fixed multiple of the uniform density, so its output has squared weighted norm bounded independently of b_i. Tensoring shows that the full sample information bound for these alternatives uses only η_S, with fixed multiplicative constants. This leaves the two-point, hypercube, and mixture lower proofs unchanged. The case $|S| = 1$ uses the primitive wavelet alternatives for the corresponding critical threshold.

8.7 The corner Poisson experiment

Let $E_i \subset \{-1, 1\}$ list the endpoints attaining the smaller exponent in coordinate i, with sign -1 at L_i and sign 1 at U_i. For $e \in \prod_i E_i$, let x_e be that corner. Define the profile for a joint noise density f by

$$S_{e,f}(t) = \left(\prod_i w_{i,e_i}(0)\right) \int_Q f(z) \prod_i (e_i z_i - t_i)_+^{r_i - 1}\,dz, \qquad t \in \mathbb{R}^d.$$

As before, a positive-part power is zero on the nonpositive set, including when its exponent is nonpositive.

Theorem 8.2. *Let a finite family of noise densities on Q be bounded above and below by positive constants. In each coordinate i, suppose their pairwise differences have zero slice moments through degree $K_i = \lfloor r_i/2 \rfloor$ almost everywhere in the other coordinates. As all $b_i \downarrow 0$,*

$$\eta^{-1} H^2(T_b f, T_b g) \longrightarrow \sum_{e \in \prod E_i} \int_{\mathbb{R}^d} (\sqrt{S_{e,f}(t)} - \sqrt{S_{e,g}(t)})^2 \, dt.$$

The right side is finite and is positive if $f \neq g$. If $N\eta \to \lambda \in (0, \infty)$, the sample experiment converges in Le Cam distance to independent Poisson point processes with intensities $\lambda S_{e,f}(t)\, dt$ at these corners.

Proof. For each coordinate take the union $W_{i,T}$ of the width Tb_i windows on both sides of all minimal-exponent endpoints. Here $T > 2$ is fixed initially. In the observed location $y_i = x_{e,i} + b_i e_i t_i$, near the corresponding corner, each rescaled kernel tends to $w_{i,e_i}(0)(e_i z_i - t_i)_+^{r_i - 1}$. Its integral over the noise coordinate is uniformly bounded on bounded t_i after rescaling, even for negative powers since $r_i > 0$. For a negative power restrict first to arguments bounded away from zero, and then use that the omitted scalar integral is $O(\epsilon^{r_i})$ on an interval of width ϵ. Bounded noise densities and the separate integrals show convergence also for the product kernels. Thus on bounded mark windows the normalized joint observations converge in density after factoring $\prod_i b_i^{r_i - 1}$, with a uniform dominating constant. The physical volume factor is $\prod_i b_i$. This gives the asserted squared Hellinger formula on the bounded product windows.

For the tail use weighted scalar operators on the moment kernels in $L^2([-1, 1])$. This allows the signals to reach the noise endpoints. On a fixed-width window about a scalar endpoint of exponent α, Cauchy–Schwarz against the uniform noise law gives

$$\frac{|T_{b_i,i} u_i|^2}{q_{0,i}} \leq C T_{b_i,i}(|u_i|^2),$$

where $q_{0,i}$ is the scalar uniform-reference output. The integral of the right side over the window is at most $C_T b_i^{1+\alpha} \|u_i\|_2^2$, since every latent argument counted there lies within depth $O_T(b_i)$ of the endpoint. For physical inner depth $h \geq Tb_i$, the moment cancellations give the scalar Taylor remainder $C b_i^{K_i+1} h^{\alpha - K_i - 1} \|u_i\|_2$, and $q_{0,i} \geq c h^\alpha$. At a minimal endpoint, integrating its square against the denominator gives $C b_i^{r_i} T^{r_i - 2K_i - 2} \|u_i\|_2^2$. Here $r_i - 2K_i - 2 < 0$. At a nonminimal endpoint integrate from a fixed multiple of b_i inward and add its fixed-width window cost. This gives $O(b_i^{\min\{1+\alpha, 2K_i+2\}}) \|u_i\|_2^2$ with a possible logarithm when the powers tie. It is $o(b_i^{r_i}) \|u_i\|_2^2$ because both powers exceed r_i. The compact deep interior cost is $O(b_i^{2K_i+2}) \|u_i\|_2^2$.

It follows that the full scalar weighted map has squared norm $O(b_i^{r_i})$, and its restriction to $W_{i,T}^c$ has squared norm $C b_i^{r_i}(T^{-\gamma_i} + o_{b_i}(1))$, with $\gamma_i = 2K_i + 2 - r_i > 0$. Tensor the restricted map in one coordinate with the full maps in the others, then sum over coordinates in the product-window complement. The resulting weighted squared difference is bounded by

$$C\eta \left(\sum_i T^{-\gamma_i} + \sum_i o_{b_i}(1) \right) \|f - g\|_2^2.$$

To justify density in the tensor estimate, in $L^2(Q)$ apply the commuting coordinatewise orthogonal projections onto the scalar moment kernels. Their product fixes each slice-cancelled difference, and applied to approximating finite product sums it gives convergence by permitted tensors. The weighted limit agrees with the unweighted physical convolution by Young's inequality. The noise density floors bound the squared square-root difference by this same weighted difference. After

the bounded-window limit, Fatou's lemma on every enlarged bounded mark annulus followed by exhaustion gives the same tail estimate for the limiting profiles. This proves finiteness and the full Hellinger limit.

Each corner's profile determines the noise law. After reversing and translating each threshold coordinate, it is the convolution of the transformed compact law with the product of the causal kernels $t_i^{r_i-1}\mathbf{1}_{t_i>0}$, up to a fixed positive factor. The multivariate Laplace transform multiplies the signed law's transform by $\prod_i \Gamma(r_i)s_i^{-r_i}$ for $s_i > 0$. This factor is nonzero, and equality of Laplace transforms of compact measures implies equality of the measures. Thus even agreement of one full corner profile forces $f = g$.

Finally, at fixed T retain the points in the bounded corner windows. The retained count is binomial with probability $O_T(\eta)$, hence can be Poissonized with vanishing total variation error when $N\eta$ stays bounded. Given the count the marks are independent with the normalized retained density. The resulting intensity converges in L^1 on the disjoint bounded windows, giving the independent cropped Poisson processes. Conversely restore the discarded exterior points from the normalized exterior law of a fixed reference family member and restore uniform order. Normalizing the exterior densities costs only a fixed constant in the squared Hellinger estimate, since their masses tend to one. The product Hellinger bound and the weighted tail estimate give vanishing reconstruction error as $b_i \downarrow 0$ and then $T \to \infty$. The limit processes can likewise be uncropped by adding the reference-intensity complement, whose Hellinger error tends to zero by the just-proved profile tail control, using Poisson affinity first on finite windows and then exhaustion. This proves the Le Cam assertion. □

The slice threshold cannot in general be lowered while retaining this scale. With a product uniform noise reference, take a small signed tensor perturbation $\epsilon \prod_i u_i(z_i)$. Choose all the u_i smooth and compactly supported inside $[-1, 1]$ with zero mass. In one coordinate choose the first nonzero moment to have degree $j \geq 1$ with $2j \leq r_i$; in the other coordinates use the required slice cancellations. The physical chi-square cost factorizes into the scalar weighted squared costs, and is comparable to squared Hellinger by the bounded ratio of the perturbed and reference density. In the distinguished coordinate the leading derivative term has weighted integral of order b_i^{2j} for $2j < r_i$, or $b_i^{r_i}\log(1/b_i)$ at equality. This follows by the inner-depth integration of $b_i^{2j}h^{r_i-1-2j}$; the nonzero degree-j moment and nonzero leading derivative at a minimal endpoint give matching lower bounds there. The other coordinates have cost of order $b_\ell^{r_\ell}$, by a nonzero profile on some bounded window and the tail estimate. Thus the discrepancy is larger than order η when the specified slice degree is left free.

9 Sharp Wasserstein rates under mixed smoothness at product corners

We recover an unknown joint noise density through a known product forward density with power-law endpoints. Under coordinatewise mixed Sobolev orders, the sharp Wasserstein rates depend on subsets of coordinates, and tied subsets can produce logarithmic losses even away from a parametric threshold. We determine both the powers and the logarithms for estimation and testing in the physical sample. A hyperbolic frequency cutoff and multiresolution potential lower bounds give matching W_1 and W_2 results.

9.1 Model and rate formulas

Let X have independent coordinates with known densities g_i on fixed intervals. Each is smooth and positive in its interior and equals $w_{i,e}(h)h^{\alpha_{i,e}}$ near each endpoint e in inward depth h, with $\alpha_{i,e} > -1$ and $w_{i,e}$ smooth and strictly positive up to the endpoint. Observe N independent copies of $Y_i = X_i + b_i Z_i$, where the small $b_i > 0$ are known and Z is independent of X. The joint noise need not have independent coordinates. Put

$$r_i = 1 + \min_e \alpha_{i,e}, \qquad \eta = \prod_i b_i^{r_i}, \qquad M = N\eta.$$

Fix a positive bounded reference density f_0 on $Q = [-1,1]^d$. The unknown density is $f = f_0 + u$, with u of zero mass supported on aQ for a fixed $0 < a < 1$. Require $c \le f \le C$ on Q, with strict slack at the reference. The perturbation has norm at most a fixed radius in the torus space $H^{\mathbf{s}}_{\mathrm{mix}}$ with Fourier square weight $\prod_i (1 + k_i^2)^{s_i}$, where every $s_i \ge 0$. The torus has side length eight. Transport uses Euclidean distance.

For a dyadic index $\ell \in \mathbb{Z}^d_{\ge 0}$ set

$$n(\ell) = \max_i \ell_i, \qquad P(\ell) = n(\ell) + \mathbf{s}\cdot\ell, \qquad E_\nu(\ell) = \sum_i (1 + \nu r_i)\ell_i - \nu n(\ell).$$

Here $\nu = 2$ is used for estimation and $\nu = 4$ for testing. For $J \ge 0$ let $Z_\nu(J) = \sum_{P(\ell)\le J} 2^{E_\nu(\ell)}$. The following constants describe these sums explicitly. For $S \subseteq [d]$ write $r(S) = \sum_{i\in S} r_i$ and $s(S) = \sum_{i\in S} s_i$, and put

$$\tau_\nu = \max\left\{0,\ \max_{\varnothing \ne S\subseteq[d]} \frac{|S| + \nu r(S) - \nu}{1 + s(S)}\right\}.$$

In each cone $x_i = \max_j x_j$, $x_j \ge 0$, take the face on which $E_\nu = \tau_\nu$ in the polytope $P(x) \le 1$. Let κ_ν be the largest dimension of these faces over cones attaining τ_ν. Equivalently find the affine dimension of the winning vertices $\mathbf{1}_S/(1 + s(S))$ among subsets containing a common i, including the origin if $\tau_\nu = 0$, then take the largest dimension. The origin alone has dimension zero.

Theorem 9.1. *For $1 \le p \le 2$, uniformly for sufficiently small b_i and as $M \to \infty$, the minimax expected W_p estimation risk has order*

$$M^{-1/(2+\tau_2)}(\log M)^{\kappa_2/(2+\tau_2)}.$$

The minimax separation for testing $f = f_0$ against $W_p(f, f_0) \ge \rho$, with both errors at most $1/4$, has order

$$M^{-1/(2+\tau_4/2)}(\log M)^{\kappa_4/[2(2+\tau_4/2)]}.$$

The deterministic W_p modulus for output Hellinger radius $\sqrt{\eta}\,\delta$ has order δ^θ, where

$$\theta = \min\left\{1,\ \min_{\varnothing \ne S\subseteq[d]} \frac{1 + s(S)}{r(S) + s(S)}\right\}.$$

All lower bounds hold even with all proper coordinate marginals prescribed to equal those of the reference, and with any fixed finite number of matched coordinate-slice moments.

A tie in the subset maximum can affect the logarithm when the leading power is already nonparametric. For example, in two dimensions take $(r_1, r_2) = (1, 1/2)$ and $(s_1, s_2) = (1, 4)$. For estimation the sets $\{1\}$ and $\{1, 2\}$ tie at $\tau_2 = 1/2$ in one cone, with $\kappa_2 = 1$, so the rate is $M^{-2/5}(\log M)^{2/5}$.

Ding, Zou, Wang, Shahrampour and Tuo [38] use hyperbolic estimators for direct mixed-smooth density estimation under integral probability metrics. Cleanthous, Georgiadis and Porcu [37] give direct density results under mixed smoothness. Benhaddou, Pensky and Picard [36] study mixed-Besov regularity in noisy functional deconvolution under squared Lš loss. The model here is an indirect probability experiment with mixed inverse orders and Euclidean transport. Both the subset powers and the tied-face logarithms follow from this combination. We use exact corner probes as in Section 8; the multiscale transport inequality for the lower bounds is from Park, Moon and Kim [39, Section 4]. The published Gaussian Archive inverse [14] yields a different logarithmic modulus from a fixed super-smooth operator.

9.2 Observation tools and the dyadic sums

Write T_b for the physical convolution operator on signed noise densities. The exact dual probes on the strongest corner of Section 8 are linear in the smooth tests ϕ. A centered observation probes $\int \phi(f - f_0)$ unbiasedly with variance at most $C\eta^{-1}\|\phi\|^2_{H^{\mathbf{r}}_{\mathrm{mix}}}$. The same argument yields

$$\|f - g\|_{H^{-\mathbf{r}}_{\mathrm{mix}}} \le C\eta^{-1/2} H(T_b f, T_b g).$$

For a contrast supported on the smaller cube and with sufficiently many zero moments in each coordinate slice, the full-data bound is

$$\int \frac{(T_b u)^2}{T_b f_0} \le C\eta \|u\|^2_{H^{-\mathbf{r}}_{\mathrm{mix}}(\mathbb{R}^d)}.$$

These facts are tensor products of the scalar fractional endpoint probes and scalar weighted contrast bounds. The dependent reference denominator is comparable to that of a product uniform noise. Thus the bounds apply with no independence of Z. On the standardized cube,

$$W_2(f, g) \le C\|f - g\|_{H^{-1}(\mathbb{T}^d)}$$

by the Neumann transport current with the density floor. In detail, the mean-zero Neumann variational solution has L^2 gradient controlled by the torus dual norm using Poincaré and extension of $H^1(Q)$ to the torus. Use the gradient current and the linear path of positive densities in the dynamic transport formula.

Lemma 9.2. *For $\nu = 2, 4$, $Z_\nu(J) \asymp 2^{\tau_\nu J}(1 + J)^{\kappa_\nu}$.*

Proof. Extend P, E_ν to $\mathbb{R}^d_{\ge 0}$. On the cone with coordinate i maximal, both are linear. Its unit polytope $P \le 1$ has vertices the origin and $\mathbf{1}_S/(1 + s(S))$ for $i \in S$. Indeed the rays of the cone are the nonzero binary vectors with coordinate i equal to one, by considering the box section $x_i = 1$; intersect each ray with $P = 1$. This yields the displayed maximum formula.

For any one of these polytopes D, let F be the maximizing face of the linear function E_ν, of dimension k. If its maximum is τ, the deficit $\tau - E_\nu(x)$ is comparable to the distance to F on D, by the polyhedral error bound applied to the extra maximizing equality (or by linear triangulation of the polytope). For small positive ϵ, the volume within distance ϵ of F inside D has order ϵ^{d-k}. For the upper bound use a Euclidean tube. For the lower bound join a smaller relative ball in the face with a fixed interior ball of D at distances of order ϵ; this supplies the required transverse thickness. If $F = D$ the claim means constant volume. After rescaling JD to D, integration of 2^{E_ν} there has order $2^{J\tau} J^k$ by that volume estimate and exponential integration of the deficit. Cones with smaller maxima are negligible, and dimensions of tied maximal faces determine the largest

power. Finally unit cubes based at integer ℓ change E_ν and P only by bounded amounts. Since P is coordinatewise nondecreasing, the sum over $P(\ell) \le J$ is bounded above and below, up to constants, by the continuous integrals on $P(x) \le J + O(1)$ and $P(x) \le J$, respectively. This proves the lemma. □

9.3 Upper bounds

Use real Fourier probes and retain indices with dyadic blocks satisfying $P(\ell) \le J$, where $\langle k_i \rangle \asymp 2^{\ell_i}$. The squared H^{-1} tail for u is $O(2^{-2J})$ because its squared Sobolev coefficient weight relative to the mixed positive norm is at most $C2^{-2P(\ell)}$. Here $\langle k \rangle \asymp 2^{n(\ell)}$. In a dyadic block there are $\asymp 2^{\sum_i \ell_i}$ real modes, counting a fixed low-frequency convention. The joint covariance of the H^{-1}-weighted probes is bounded above in positive semidefinite order by $C\eta^{-1}$ times the diagonal with entries $v_k = \langle k \rangle^{-2} \prod_i \langle k_i \rangle^{2r_i}$. Hence the trace and squared Hilbert–Schmidt sums have orders at most $Z_2(J)$ and $Z_4(J)$, respectively, after removing the powers of η.

The squared reconstruction error is $O(2^{-2J} + M^{-1} Z_2(J))$. Project to a countable dense subset of admissible densities in H^{-1}, within tolerance 2^{-J}, then use the transport bound. Choose the largest integer J with $2^{2J} Z_2(J) \le M$. Lemma 9.2 shows that the two orders balance to constants and yields the estimation rate.

For testing use the unbiased average of inner products on distinct pairs of the weighted, baseline-centered probe vectors. If their truncated mean is ϑ, its expectation is $|\vartheta|^2$. The degenerate part has standard deviation $O(M^{-1}\sqrt{Z_4(J)})$, and the linear part has standard deviation bounded by this order's square root times $C|\vartheta|$, because $\max v_k \le (\sum v_k^2)^{1/2}$. Choose the largest integer J with $2^{2J}\sqrt{Z_4(J)} \le M$. By the transport bound and tail estimate, sufficiently separated alternatives have $|\vartheta|$ at least a large constant times 2^{-J}. Then a threshold and Chebyshev's inequality give both errors at most $1/4$. Lemma 9.2 gives the testing rate.

Finally the pointwise weight inequality on every dyadic block

$$2^{-n(\ell)} \le 2^{-\theta \mathbf{r} \cdot \ell} 2^{(1-\theta)\mathbf{s} \cdot \ell}$$

holds with the displayed θ. Indeed on each maximum-coordinate cone the minimum ratio $(n + \mathbf{s} \cdot \ell)/[(\mathbf{r} + \mathbf{s}) \cdot \ell]$ occurs on one of its binary rays. Apply Hölder's inequality to the coefficient sum and use the weighted Hellinger lower and the W_2 transport bound to get the modulus upper. The case $\theta = 1$ is immediate by the same weight comparison.

9.4 Multiresolution lower bounds

We describe a lower-bound set of levels for each $\nu = 2, 4$ at the cutoff J used in the corresponding upper bound. Keep levels with $P(\ell) \le J$ whose sum of $2^{E_\nu(\ell)}$ is comparable to $Z_\nu(J)$. They may be chosen so that either there is only one level, or all retained levels have $n(\ell) \le (1 - c_*)J$ for a fixed $c_* > 0$. This ensures positivity even when some smoothness orders are zero.

Here are the details of the choice, also giving a simpler logarithmic multiplicity formula. If $d + \nu r([d]) < \nu$, all the nonzero rays have negative objective and $\tau_\nu = \kappa_\nu = 0$. Keep the level zero. If equality holds, the only maximizing nonzero ray is the full binary vector, and the face includes the origin. Thus $\tau_\nu = 0, \kappa_\nu = 1$. Keep the integer points on a fixed inner portion of its dilated segment with $n(\ell) \le J/2$, of which there are order J. If $d + \nu r([d]) > \nu$, the positive maximum solves

$$\sum_i (1 + \nu r_i - \tau_\nu s_i)_+ = \nu + \tau_\nu.$$

In fact this is the equality for maximizing the subset numerator after subtracting the ratio times the denominator. Every winning subset contains the positive-weight indices, can include any zero-weight indices, and excludes all negative-weight ones. A cone with one positive-weight axis contains all winning rays, so the normalized vertices in that cone span a face of dimension equal to the number of zero weights. Other maximizing cones cannot give a larger dimension. Thus $\kappa_\nu = \#\{i : 1 + \nu r_i = \tau_\nu s_i\}$. If $\kappa_\nu = 0$, round down one scaled maximizing vertex and keep its level. If $\kappa_\nu > 0$, the optional axes have positive smoothness, so a relative-open patch of the maximizing face has $\mathbf{s} \cdot x$ bounded away from zero and $n(x) \le 1 - c_*$. Round down all points in the dilated patch. It meets order J^{κ_ν} distinct unit cubes, since the intersection with each cube has bounded relative volume. Each resulting index satisfies $P \le J$ and has objective $\tau_\nu J + O(1)$. This choice captures the required order in every case.

Take compactly supported scalar orthonormal wavelets with enough regularity and vanishing moments for the fixed orders in the problem. At each retained level ℓ take tensor products $\psi_{\ell,k}$ at coordinate levels $\ell_i + \ell_0$, with fixed ℓ_0 large enough to fit inside a fixed smaller cube. Retain only translations supported there. The number at such a level is $\asymp 2^{|\ell|_1}$ and overlap is uniformly bounded at that level. The fixed level offset changes only constants. We use the tensor wavelet characterization of mixed Sobolev norms [28, 31], either on the real line in each coordinate or periodized for the supported functions on the larger torus. In particular, for a potential

$$V_\omega = \sum_{\ell,k} d_\ell \omega_{\ell,k} \psi_{\ell,k}, \qquad f_\omega = f_0 - \Delta V_\omega,$$

the squared mixed positive norm of the perturbation is bounded by

$$C \sum_\ell d_\ell^2 2^{|\ell|_1 + 4n(\ell) + 2\mathbf{s}\cdot\ell}.$$

The same upper bound with $\mathbf{s}$ replaced by $-\mathbf{r}$ holds for the mixed negative norm, and for arbitrary real coefficients in place of $d_\ell \omega_{\ell,k}$ use the corresponding sum over k. These statements follow by bounding each $\partial_i^2 V$ with mixed orders increased by two in that coordinate and summing over i. The tensor wavelet norm of the gradient squared is equivalent to $\sum_{\ell,k} |d_\ell \omega_{\ell,k}|^2 2^{2n(\ell)}$: sum the single-coordinate H^1 norms and use Poincaré on the fixed support. The same equivalence holds for any difference of potentials. All Laplacian terms are products with the required zero slice moments, by integration by parts in the differentiated coordinate and vanishing moments in the others.

For all priors below the total gradient energy has order $\mathcal{E} \asymp \varepsilon^2 2^{-2J}$, with sufficiently small fixed $\varepsilon > 0$. The chosen level property gives a uniformly small bound on the pointwise density perturbation. At one level it is bounded by

$$C d_\ell 2^{2n(\ell) + |\ell|_1/2} \asymp 2^{n(\ell)} \big(d_\ell^2 2^{2n(\ell) + |\ell|_1}\big)^{1/2}.$$

If only one level is retained this is $O(\varepsilon)$. Otherwise sum over levels by Cauchy–Schwarz, using their polynomial number in J, the bound $n \le (1 - c_*)J$ and the total energy order. It is still $O(\varepsilon)$. Thus positivity and the density bounds follow for all signs, once the claimed energies and positive norm budgets are verified.

9.4.1 Estimation lower bound

On the estimation levels choose independent uniform signs and

$$d_\ell^2 = \varepsilon^2 M^{-1} 2^{2\mathbf{r}\cdot\ell - 4n(\ell)}.$$

The gradient energy $\mathcal{E}$ is comparable to $\varepsilon^2 M^{-1} Z_2(J) \asymp \varepsilon^2 2^{-2J}$. The positive mixed Sobolev budget is at most $C 2^{2J}\mathcal{E} \le C'\varepsilon^2$. This verifies admissibility. For a single sign flip the physical N-sample squared Hellinger cost is $O(\varepsilon^2)$ by the full-data information bound and the individual negative norm weight $2^{4n-2\mathbf{r}\cdot\ell}$.

Draw two independent sign vectors ω, ω' from the posterior given the data. For each bit, the probability of posterior disagreement averaged over the data is bounded below by a positive constant. To see this, condition additionally on all the other bits: the two-point data laws have total variation bounded away from one, so their posterior bit error is bounded below. Posterior disagreement is at least that bit error, and removing the extra conditioning increases the averaged conditional variance. Let $V = V_\omega - V_{\omega'}$. The energy comparison gives $\mathbb{E}\|\nabla V\|_2^2 \gtrsim \mathcal{E}$.

A sign vector drawn once from the posterior and averaged over the data has the original independent prior. By the fourth-moment inequality for independent signs and bounded overlap at each level, $\mathbb{E}\|\nabla V\|_4^4 \le C\mathcal{E}^2$. Indeed the pointwise sum of squared gradient coefficients is bounded by $C\sum_\ell d_\ell^2 2^{2n+|\ell|_1} \asymp \mathcal{E}$, and the functions are supported on a fixed cube. For the posterior difference use the fourth power bound by a constant times the two marginal fourth powers. The potential inequality of [39, Section 4] gives

$$W_1(f_\omega, f_{\omega'}) \ge c\frac{\|\nabla V\|_2^3}{\|\nabla V\|_4^2}.$$

It applies to these C^2 compactly supported potentials. Hölder's inequality for this ratio bounds its expectation below by $c(\mathbb{E}\|\nabla V\|_2^2)^{3/2}/(\mathbb{E}\|\nabla V\|_4^4)^{1/2} \gtrsim \sqrt{\mathcal{E}}$. The posterior triangle inequality implies that half this pairwise Bayes distance is a lower bound for the risk of any estimator. This establishes sharp estimation.

9.4.2 Testing lower bound

Use the testing levels, and write $Z = \sum_{\ell \text{ retained}} 2^{E_4(\ell)}$. Choose

$$d_\ell^2 = \varepsilon^2 M^{-1} Z^{-1/2} 2^{4\mathbf{r}\cdot\ell - 6n(\ell)}.$$

Now $\mathcal{E} \asymp \varepsilon^2 M^{-1}\sqrt{Z} \asymp \varepsilon^2 2^{-2J}$, and the mixed Sobolev budget is again at most $C 2^{2J}\mathcal{E}$. Let B_N be N times the weighted physical Gram matrix of the individual signed basis directions with their amplitudes. The joint mixed norm bound gives

$$0 \preceq B_N \preceq CM \operatorname{diag}\{d_\ell^2 2^{4n(\ell)-2\mathbf{r}\cdot\ell}\}.$$

Thus $\operatorname{tr}(B_N^2) \le C\varepsilon^4$ by squaring these entries and summing over k, ℓ; the operator norm is also small. For the uniform sign mixture, the N-sample likelihood second moment is at most $\mathbb{E}e^{\omega^\top B_N \omega'}$, by $1 + x \le e^x$. The hyperbolic-cosine bound conditional on one sign vector, followed by a Gaussian integral and the same bound on the other vector, gives $\det(I - B_N^2)^{-1/2} \le \exp(C\varepsilon^4)$. The resulting mixture is close to the null in total variation.

The energy and fourth-moment argument now against the reference, using $\mathbb{E}\|\nabla V_\omega\|_4^4 \le C\mathcal{E}^2$, gives $\mathbb{E}W_1(f_\omega, f_0) \gtrsim \sqrt{\mathcal{E}}$. Changing one sign changes this distance by at most $Cd_\ell\|\nabla\psi_{\ell,k}\|_1 \le Cd_\ell 2^{n(\ell)-|\ell|_1/2}$, by integration by parts in the transport dual. The bounded differences variance estimate therefore gives

$$\operatorname{Var}(W_1(f_\omega, f_0)) \le C\mathcal{E}\,\frac{\sum_{\ell \text{ retained}} 2^{E_4(\ell)-|\ell|_1}}{Z}.$$

If $Z \to \infty$, the ratio tends to zero: split off any fixed finite set of levels and use the factor $2^{-|\ell|_1}$ on the others. Thus all but $o(1)$ of the prior signs have distance at least a constant times $\sqrt{\mathcal{E}}$. Condition on these signs, changing the mixture by $o(1)$ in total variation. Small fixed ε gives the claimed testing lower bound. If Z is bounded use a fixed smooth interior moment-cancelled contrast of amplitude $cM^{-1/2}$ and the two-point information bound instead.

9.4.3 Modulus lower bound

If $\theta < 1$, take a subset S attaining it. At a fine common level j in those coordinates use tensor wavelets with disjoint supports in the varying coordinates, and use fixed compact mean-zero wavelets in the other coordinates. There are $\asymp 2^{j|S|}$ products. An equal sign-amplitude sum with L^2 norm $\varepsilon 2^{-js(S)}$ has mixed positive norm $O(\varepsilon)$ and mixed negative norm $O(2^{-j(r(S)+s(S))})$. Its pointwise size is bounded by a constant times $\varepsilon 2^{-js(S)}$ by disjointness. A signed sum of the same bumps scaled to have uniformly bounded gradient gives W_1 distance at least $c2^{-j(1+s(S))}$ from f_0: each normalized bump test has amplitude $2^{-j-j|S|/2}$, and summing the pairings gives the stated bound. The weighted physical information bound makes $H/\sqrt{\eta}$ at most a constant times $2^{-j(r(S)+s(S))}$. Choose j for the required radius. If $\theta = 1$, use a fixed nonzero smooth moment-cancelled contrast of amplitude $c\delta$. These lower bounds in W_1 also hold for $W_p \geq W_1$, completing Theorem 9.1.

10 Information scales for mixed noise moments

> For a known product density observed with small additive joint noise, the information about a mixed noise moment is the product of its coordinate information scales. Each scale comes from the interior, a logarithmic range of depths, or an endpoint window. We give a uniform decomposition of the full weighted observation norm, combining finitely many moment terms with a negative Sobolev term in each coordinate. The rescaled information forms converge in operator norm. Each limit block depends on the reference marginal on its residual axes. At the corner sampling scale this separates independent Gaussian blocks from a corner Poisson experiment, even when the unknown noise has dependent coordinates.

10.1 Model and scalar information scales

For each coordinate i, a known density g_i is supported on an interval $[L_i, U_i]$, smooth and positive in its interior. Near an endpoint it is $w_{i,e}(h)h^{\alpha_{i,e}}$, where $e \in \{-1, 1\}$ is the outward sign, h is inward depth, $\alpha_{i,e} > -1$, and $w_{i,e}$ is smooth and strictly positive up to zero. The latent variable X has independent coordinates with these densities. Independently of X, let Z have an unknown joint density f on $Q = [-1, 1]^d$. The observation is $Y_i = X_i + b_i Z_i$ with known $b_i > 0$. The scales b_i may go to zero at different speeds. Write $T_{\boldsymbol{b}}$ for the resulting convolution operator on noise-scale densities, and $q_{\boldsymbol{b},f} = T_{\boldsymbol{b}} f$.

Set

$$r_i = 1 + \min_e \alpha_{i,e}, \qquad K_i = \lfloor r_i/2 \rfloor, \qquad \sigma_{i,j}(b)^2 = \begin{cases} b^{2j}, & 2j < r_i, \\ b^{r_i} \log(1/b), & 2j = r_i, \\ b^{r_i}, & 2j > r_i. \end{cases}$$

Thus $\sigma_{i,0} = 1$. For $S \subset \{1, \dots, d\}$ and $j = (j_i)_{i \in S}$, let

$$m_{S,j}(u)(z_{S^c}) = \int u(z) \prod_{i \in S} z_i^{j_i} \, dz_S.$$

Use the mixed Sobolev norm with Fourier square weight $\prod_{i\in S^c}(1+\xi_i^2)^{-r_i}$ in the remaining coordinates, with ordinary absolute value if $S^c=\varnothing$. Define the squared norm

$$\mathcal{N}_{\boldsymbol{b}}(u)^2=\sum_{S\subset[d]}\sum_{0\le j_i\le K_i,\ i\in S}\left(\prod_{i\in S}\sigma_{i,j_i}(b_i)^2\right)\left(\prod_{i\notin S}b_i^{r_i}\right)\|m_{S,j}(u)\|^2_{H^{-\mathbf{r}_{S^c}}_{\mathrm{mix}}}.$$

Theorem 10.1. *Fix $0<a<1$ and a reference density f_0 bounded above and below by positive constants on Q. For every L^2 signal u supported on aQ, uniformly for sufficiently small b_i,*

$$\int\frac{|T_{\boldsymbol{b}}u|^2}{q_{\boldsymbol{b},f_0}}\asymp\mathcal{N}_{\boldsymbol{b}}(u)^2,\qquad 0/0:=0.$$

For any densities f,g bounded above and below on Q whose difference is supported on aQ, their squared output Hellinger distance is comparable to $\mathcal{N}_{\boldsymbol{b}}(f-g)^2$.

In classical fixed-error deconvolution, polynomial moments are root-sample-size estimable under moment assumptions; Hall and Lahiri [40] examine more delicate fractional moments. Pensky [41] studies more general linear functionals. Trabs [42] describes information bounds through the adjoint score operator in fixed regular inverse experiments. Here the operator itself changes with sample size. The logarithmic gain and the product information factors follow from this change. The one-dimensional endpoint probes, marginal recovery and the corner Poisson experiment are developed in Sections 7, 8. For positively curved reference supports, a single radial shell has the information scales in Section 6. The Gaussian Archive result [14] has a different, exponentially attenuated inverse operator.

10.2 Uniform scalar comparison and its leading form

Suppress the index i and use a scalar reference noise density f_* bounded above and below on $[-1,1]$. Write b,r,K,σ_j for the scalar parameters and $q_b=T_bf_*$. For a compactly supported signal u, use $m_j(u)=\int z^ju(z)\,dz$. Choose fixed smooth lifts p_j supported inside $[-a,a]$ with $m_\ell(p_j)=\mathbf{1}_{\ell=j}$ for $0\le\ell\le K$. They exist by a positive definite moment Gram matrix of monomials times a smooth interior bump. Write $v=u-\sum_{j\le K}m_j(u)p_j$.

Lemma 10.2. *For u supported on $[-a,a]$, uniformly for small b,*

$$\|T_bu\|^2_{L^2(1/q_b)}\asymp b^r\|u\|^2_{H^{-r}(\mathbb{R})}+\sum_{j=0}^{K}\sigma_j(b)^2|m_j(u)|^2.$$

Equivalently replace u by v in the Sobolev term.

Proof. The stronger scalar endpoint window furnishes the exact dual probes of Section 7, valid for arbitrary signed input. Its observation density is bounded by Cb^{r-1} and the probes have physical L^2 bound $Cb^{1/2-r}\|\phi\|_{H^r}$. Hence $b^{r/2}\|u\|_{H^{-r}}\le C\|T_bu\|_{L^2(1/q_b)}$, by duality after localizing a test to the fixed noise support.

For an upper bound on fixed-width endpoint windows, the compactly cut off scaled kernel at an endpoint of exponent α_e has Fourier decay of order $1+\alpha_e$, by dyadic integration by parts as in Section 7. Where the signed output can be nonzero, $q_b\ge cb^{\alpha_e}$ from a fixed cap of noise values a positive distance beyond the inner noise support. Thus the weighted cost there is $O(b^{1+\alpha_e}\|u\|^2_{H^{-r}})$. For inner depth $h\ge Tb$ subtract the Taylor polynomial through K in the noise variable. The remainder is bounded by $Cb^{K+1}h^{\alpha_e-K-1}\|u\|_{H^{-r}}$, using a compact cutoff H^r test bound. Since

$q_b \asymp h^{\alpha_e}$ there, its squared integral is $O(b^r\|u\|^2_{H^{-r}})$. In fact at a minimal endpoint it is at most $Cb^rT^{r-2K-2}\|u\|^2_{H^{-r}}$. At a flatter endpoint it is $O(b^{\min(1+\alpha_e,2K+2)})\|u\|^2_{H^{-r}}$, with a logarithm if the powers tie; this is $o(b^r)$. The deep interior has squared remainder $O(b^{2K+2})\|u\|^2_{H^{-r}}$. These statements use $2K+2 > r$. Each degree-j Taylor output is $b^jD_j(y)m_j(u)$, where $D_j = (-1)^j g^{(j)}/j!$. Its squared weighted integral is bounded by $C\sigma_j^2|m_j(u)|^2$, by integrating $b^{2j}h^{\alpha_e-2j}$ over the interior. This proves the upper estimate.

For the reverse moment estimates, work on the interior outside fixed endpoint windows and use the previously controlled Sobolev term to bound the Taylor remainder. For all $2j < r$, the scaled Gram matrix of the functions D_j in $L^2(1/q_b)$ on this interior converges to $\int D_jD_\ell/g$, by dominated convergence. It is positive definite: at a minimal endpoint a nonzero linear combination has the nonzero leading power h^{r-1-j} of its highest degree, since the binomial coefficient is nonzero there. When $2J = r$, the degree-J function divided by $\sqrt{\log(1/b)}$ has a positive limiting squared norm from the integral of h^{-1} at each minimal endpoint. Its cross terms with all lower degrees tend to zero, since their unnormalized products are integrable there and at all other endpoints. Thus the scaled Gram matrix is uniformly bounded below. The Taylor polynomial norm controls the weighted moment sum. Together with the exact endpoint probe bound this proves the reverse inequality. The equivalent version follows because the lifts are fixed and $b^r \lesssim \sigma_j^2$ for $j \le K$. □

We record the exact limiting form in these coordinates. Let E be the set of endpoints attaining $r-1$. At $e \in E$ put

$$\mathcal{S}_{e,u}(t) = w_e(0)\int (ez-t)_+^{r-1}u(z)\,dz.$$

The convention at nonpositive arguments is zero. Let $\mathcal{H}$ be the closure in $H^{-r}(\mathbb{R})$ of L^2 signals supported on $[-a,a]$ and with first $K+1$ moments zero, including mass. The coordinate map is $D_bu = (\sigma_j m_j(u) : j \le K,\ b^{r/2}v)$ with range dense in $\mathbb{R}^{K+1}\oplus\mathcal{H}$.

Theorem 10.3. *In the coordinates $D_bu = (c_j, v')$, the scalar output quadratic form converges in operator norm to the sum of three independent forms:*

$$\sum_{2j,2\ell<r} c_jc_\ell\int_L^U \frac{D_jD_\ell}{g}, \qquad \mathbf{1}_{2J=r}\,c_J^2\binom{r-1}{J}^2\sum_{e\in E}w_e(0), \qquad \sum_{e\in E}\int_{\mathbb{R}}\frac{|\mathcal{S}_{e,v'}(t)|^2}{\mathcal{S}_{e,f_*}(t)}\,dt.$$

In the middle term include it only if $r/2$ is a positive integer J. The last term is bounded above and below by positive multiples of the H^{-r} norm squared on $\mathcal{H}$.

Proof. The normalized lift outputs T_bp_j/σ_j have the stated bulk and critical Gram limits. On the interior their first nonzero Taylor coefficient is b^jD_j, by the choice of lifts. The other Taylor terms through K are zero, and the remainder squared is $O(b^r)$, negligible relative to σ_j^2. At a critical endpoint, $q_b/g = 1+O(b/h)$ for depth $h \ge Tb$ and $D_J^2/g \sim w_e(0)\binom{r-1}{J}^2h^{-1}$. The fixed endpoint windows themselves cost only $O(b^r)$ for a lift. This gives the stated coefficient and shows the critical-to-bulk cross terms vanish.

For the residual block, in a fixed scaled window around a minimal endpoint the normalized convolution kernel is $(ez-t)_+^{r-1}w_e(b(ez-t))$. After a fixed smooth cutoff on the bounded argument range, the difference from its $b=0$ kernel is $O_T(b)$ as an operator $H^{-r}\to L^2$. Indeed the difference is b times a smooth uniformly bounded factor times a one-sided power of degree r, whose cutoff Fourier transform is $O_T(\langle\xi\rangle^{-r-1})$ by dyadic integration by parts. The normalized reference densities differ from the limiting profiles by $O_T(b)$ relatively, since the taper weight differs relatively by $O_T(b)$ throughout the same window. Where the signed output can be nonzero, the denominator profiles are

bounded below by the interior buffer. Thus the fixed-window residual quadratic forms converge in operator norm. The residual tail estimate in the proof of Lemma 10.2, with all its Taylor coefficients zero, bounds the omitted output forms by $C(T^{r-2K-2} + o_b(1))\|v'\|^2_{H^{-r}}$. The same bound passes to the limiting profiles on bounded annuli and then by exhaustion. This proves norm convergence of the entire residual block.

Finally the normalized lift outputs have vanishing energy in each fixed strongest-endpoint window, since $b^r/\sigma_j^2 \to 0$ for all retained moment degrees. The normalized residual operator has small norm outside those windows as $T \to \infty$. Cauchy–Schwarz on the window and on its complement makes all lift–residual cross terms tend uniformly to zero. This proves the claimed operator-norm convergence, and Lemma 10.2 gives the residual coercivity. □

10.3 Tensor information and mixed moments

For the product uniform reference $f_* = \prod_i (1/2)$ on Q, the observed reference is the product of scalar references. The scalar weighted output maps are bounded above and below in the norms of Lemma 10.2. Taking Hilbert tensor products of their operator inequalities gives the tensor norm bound on finite sums of separated signals. Expanding the tensor product of the squared norms in Lemma 10.2 gives exactly $\mathcal{N}^2_{\boldsymbol{b}}$: each axis contributes its negative Sobolev term or one of its scalar moment evaluations. Finite separable signals supported on the smaller cube are dense in $L^2(aQ)$. The norm bound passes to this limit since all individual moment evaluations and Sobolev embeddings are bounded on supported L^2, and weighted convergence agrees with physical convolution by Young's inequality. For a positive dependent f_0, its observed density is pointwise comparable to that of the product uniform reference, by the original noise density bounds. The information comparison therefore holds for every such reference. The square-root identity and the same pointwise comparisons prove the Hellinger assertion of Theorem 10.1.

We now identify the exact leading form for a dependent reference. Use the tensor coordinate map $D_{\boldsymbol{b}} = \bigotimes_i D_{b_i,i}$. A tensor block selects a type in each axis: bulk, critical, or residual. Let S be its residual axes. The bulk axes have the finite-dimensional scalar Gram matrices from Theorem 10.3, and the critical axes have its scalar critical coefficients. For residual axes S replace the product residual form by the joint form

$$\mathcal{J}_S(v) = \sum_{e_S} \int_{\mathbb{R}^{|S|}} \frac{|\mathcal{S}_{e_S,v}(t)|^2}{\mathcal{S}_{e_S,(f_0)_S}(t)}\, dt. \tag{8}$$

Here e_S runs over the strongest endpoints in each residual axis, $(f_0)_S$ is the noise reference marginal, and $\mathcal{S}_{e_S,v}$ is the product causal power profile of Section 8 in those coordinates. For $S = \varnothing$ the residual form is just the Euclidean form on scalars.

Theorem 10.4. *In coordinates $D_{\boldsymbol{b}}$, the weighted output quadratic form with denominator $q_{\boldsymbol{b},f_0}$ converges in operator norm to a block-diagonal coercive form $\mathcal{Q}_{f_0}$. In each tensor block it is the tensor product of the bulk Gram matrices, critical scalars, and the joint residual form* (8)*. The limiting form depends on f_0 only through the marginal on the residual axes of the block.*

Proof. Fix a small $\delta > 0$ and large T. In each coordinate, use disjoint observation regions for the three types: physical interior depth at least δ for bulk; depth between Tb_i and δ for critical; and the full rescaled windows of radius Tb_i at the minimal endpoints for residual. The scalar proof of Theorem 10.3 shows that each normalized row of the operator has arbitrarily small squared norm outside its designated region as $\boldsymbol{b} \to 0$, then $T \to \infty$ and $\delta \downarrow 0$. For bulk this is dominated convergence with integrable boundary weights. For critical the discarded

parts are $o(1)$ after division by the logarithm. For residual this is the uniform Taylor tail bound. Tensoring these restriction inequalities with the unrestricted maps from all other coordinates gives the same localization for each tensor block. Comparison to the product uniform reference transfers those estimates to the dependent denominator. Two different blocks have disjoint regions in at least one coordinate, so Cauchy–Schwarz makes their cross form vanish uniformly.

On the product region assigned to one block with residual axes S, the reference output is

$$q_{\boldsymbol{b},f_0}(y) = \left[\prod_{i\notin S} g_i(y_i)\right] q_{\boldsymbol{b}_S,(f_0)_S}(y_S)\big(1 + O(1/T) + o_{\boldsymbol{b}}(1)\big)$$

with the relative error uniform at fixed T, δ. Indeed for a bulk coordinate the kernel ratio $g_i(y_i - b_i z_i)/g_i(y_i)$ tends to one uniformly over the noise interval. In a critical coordinate it is $1+O(b_i/h_i) = 1 + O(1/T)$. Multiply these bounds inside the nonnegative integral defining the reference output, and integrate out the nonresidual coordinates. In the residual coordinates the endpoint taper weights converge relatively to $w_{i,e_i}(0)$ on the fixed rescaled windows. The residual denominator is thus relatively approximated by the corresponding marginal profile there. It is bounded below where a signal on the smaller cube can contribute, by fixed caps in the residual coordinates and the noise density floor. The normalized residual kernel converges in mixed-norm operator norm on those windows by tensoring the scalar cutoff bounds. The nonresidual Gram integrals converge as in the scalar proof. The relative error in the denominator bounds the error of the whole restricted operator form by $O(1/T) + o_{\boldsymbol{b}}(1)$ times its bounded norm. Passing first to the small scales, then removing the cutoffs by their uniform localization, proves the asserted limit on finite sums of tensors, uniformly in their Hilbert norm. Completing the spaces gives the operator-norm result. Coercivity follows from Theorem 10.1 in the equivalent coordinates of Lemma 10.2. □

For a nonzero multi-index β of nonnegative integers, write $\theta_\beta(f) = \int z^\beta f(z)\,dz$ and

$$I_\beta = N \prod_i \sigma_{i,\beta_i}(b_i)^2.$$

Theorem 10.5. *Consider densities bounded above and below on Q with a common fixed compactly supported Sobolev perturbation class as in Section 8. Uniformly for all sufficiently small b_i, the minimax expected absolute loss for $\theta_\beta(f)$ is $\asymp \min\{1, I_\beta^{-1/2}\}$. When $I_\beta \to \infty$, the same rate is the testing separation for the composite null $\theta_\beta(f) = \theta_\beta(f_0)$ with both errors at most $1/4$. No other noise moments need be known. The upper bounds also hold without the fixed buffer between the perturbation support and the noise-box boundary.*

Proof. In a coordinate with $2j < r_i$, polynomial inversion of the independent observation moments gives an unbiased probe of variance $O(b_i^{-2j})$. Use the constant one at degree zero. For $2j > r_i$, the exact endpoint probe for a smooth cutoff monomial has variance $O(b_i^{-r_i})$. For the critical coordinate $2J = r_i$, use the original weighted output norm on arbitrary supported L^2. It is finite even without the buffer: Cauchy–Schwarz in the nonnegative convolution kernel gives $(T_b u)^2/q_b \le C T_b(|u|^2)$ under the uniform scalar reference. The exact endpoint probes give $b^{r_i/2}\|u\|_{H^{-r_i}}$ bounded by this norm without a denominator floor on the endpoint window, because the proof uses only the upper bound on q_b there. On the interior annulus the Taylor remainder and scaled Gram argument of Lemma 10.2 then gives the bound on $\sigma_{i,J}|m_J(u)|$. The Riesz representative v in the closure of the weighted output space gives the observable $v(Y_i)/q_b(Y_i)$, zero when $q_b = 0$. Its expectation is the target moment by duality, and its second moment is $O(\sigma_{i,J}^{-2})$ under every bounded positive noise density, by comparison to q_b. The polynomial and window probes also have the indicated second-moment bounds.

Multiply the uncentered scalar probes across coordinates. The product is unbiased for the mixed moment by the exact scalar identities and Fubini. For the second moment compare the joint output density to the product of the scalar uniform-reference outputs. The scalar second-moment bounds multiply, giving $C\prod_i \sigma_{i,\beta_i}^{-2}$. The sample mean, with clipping to the bounded moment range if needed, proves estimation and composite testing upper bounds.

For lower bounds choose a fixed smooth compact density bump in each coordinate, and differentiate it β_i times, using the undifferentiated bump when $\beta_i = 0$. The resulting product direction has total mass zero, and its mixed moment of index β is nonzero. A scalar factor has first nonzero moment β_i; its weighted squared output is $O(\sigma_{i,\beta_i}^2)$ by Lemma 10.2 and the moment cancellations. Tensoring bounds the full weighted cost by $C\prod_i \sigma_{i,\beta_i}^2$. Use a small constant times $\min\{1, I_\beta^{-1/2}\}$ as the amplitude of the direction around f_0. This gives admissible probabilities, two-point N-sample total variation less than $1/2$, and target separation of the same order as the amplitude. The two-point lower bounds give both conclusions. □

10.4 Variance constants for unbiased moment probes

Fix the interior buffer. Let $\mathcal{X}_0$ be the completed tensor coordinate space restricted to total mass zero. The restriction removes just the coefficient of the tensor of degree-zero moments. Let $Q_{\boldsymbol{b}}$ and Q_0 denote the bounded self-adjoint Gram operators of the exact and limiting information forms on this space at reference f_0. By Theorem 10.4, they are uniformly coercive and $Q_{\boldsymbol{b}} \to Q_0$ in operator norm.

For the mixed moment β put $k_{\boldsymbol{b}} = \prod_i \sigma_{i,\beta_i}(b_i)$. The standardized functional on $\mathcal{X}_0$ is

$$\ell_{\boldsymbol{b}}(x) = k_{\boldsymbol{b}} \int z^\beta (D_{\boldsymbol{b}}^{-1}x)(z)\, dz.$$

It converges in the Hilbert dual norm to a functional ℓ_β supported in one tensor block. In a coordinate $\beta_i \le K_i$ it is the corresponding moment-coordinate evaluation. In a coordinate $\beta_i > K_i$ it is evaluation of the degree-β_i moment on the residual block. The latter statement holds because the normalized contributions of all the moment lifts have factors $b_i^{r_i/2}/\sigma_{i,j}(b_i) \to 0$. These moment evaluations are continuous on supported negative Sobolev spaces by compact cutoff duality. Let ℓ_β^* denote the Hilbert representative of this limiting functional.

Theorem 10.6. *Among single-observation probes that are unbiased for the standardized mixed moment on the interior perturbation class and centered at f_0, the minimum reference variance converges to*

$$V_\beta = \langle \ell_\beta^*, Q_0^{-1}\ell_\beta^* \rangle.$$

For $\theta_\beta(f)$, the corresponding estimator is the reference value $\theta_\beta(f_0)$ plus the sample mean of the centered probe divided by $k_{\boldsymbol{b}}$. As all $b_i \downarrow 0$, its normalized error at f_0 is asymptotically Gaussian with scaling $\sqrt{N}\, k_{\boldsymbol{b}}$ and variance V_β if $N\eta \to \infty$. This Gaussian conclusion also holds if $N\eta \to \lambda \in (0,\infty)$ and $\beta_i \le K_i$ for at least one coordinate i.

Proof. The score operator on $\mathcal{X}_0$ is $A_{\boldsymbol{b}}x = (T_{\boldsymbol{b}}D_{\boldsymbol{b}}^{-1}x)/q_{\boldsymbol{b},f_0}$, considered in $L^2(q_{\boldsymbol{b},f_0})$. It has closed range by coercivity and its range consists of centered functions by the mass constraint and completion. A centered probe is unbiased for the standardized difference exactly when its inner products with these scores equal $\ell_{\boldsymbol{b}}$. Necessity follows by varying small smooth admissible directions, which are dense in the coordinate space. The minimum-norm probe is the projection onto the score range, hence equals

$$A_{\boldsymbol{b}}Q_{\boldsymbol{b}}^{-1}\ell_{\boldsymbol{b}}^*.$$

Its squared norm is $\langle \ell^*_{\boldsymbol{b}}, Q_{\boldsymbol{b}}^{-1}\ell^*_{\boldsymbol{b}}\rangle$. Norm convergence and coercivity give the variance limit.

For the central limit claim, approximate the fixed vector $Q_0^{-1}\ell^*_\beta$ by finite sums of elementary tensors with smooth compactly supported residual factors in the corresponding block. The exact representers converge in the coordinate norm, so the normalized sample averages of the differences from the approximating scores have uniformly small variance. A fixed approximating score is bounded in supremum by $C\eta^{-1/2}$ by the density floor and the elementary tensor scalings. Thus its supremum divided by $\sqrt{N}$ tends to zero if $N\eta \to \infty$. If $N\eta \to \lambda$ and a retained moment is present, take the approximations within the unique target block. This is possible by the block structure of Q_0, which is preserved by the mass constraint. Every elementary term in that block has a retained scalar degree and hence its supremum is $o(\eta^{-1/2})$. The triangular-array central limit theorem therefore applies to each fixed approximate score, with variance converging by Theorem 10.4. Letting the approximation error tend to zero proves the result for the canonical probes. □

The probe at the corner sampling scale. If $N\eta \to \lambda \in (0,\infty)$ and $\beta_i > K_i$ in every coordinate, the normalized minimum-variance estimator in Theorem 10.6 instead has the limit

$$\frac{1}{\sqrt{\lambda}} \sum_e \int \frac{\mathcal{S}_{e,h_\beta}(t)}{\mathcal{S}_{e,f_0}(t)} \big[d\Pi_e(t) - \lambda \mathcal{S}_{e,f_0}(t)\, dt \big], \qquad h_\beta = Q_0^{-1}\ell^*_\beta. \tag{9}$$

Here Π_e are the independent processes with intensities $\lambda \mathcal{S}_{e,f_0}(t)\, dt$ from Section 8. The profile of h_β is defined by the weighted L^2 extension in Theorem 10.4. The compensated integral is defined by completion over bounded windows. It has variance V_β and is non-Gaussian.

Indeed h_β lies in the block with every axis residual. First approximate it there by a finite smooth tensor sum h with the prescribed moment cancellations. For its score, the exact $D_{\boldsymbol{b}}^{-1}$ scaling is $\eta^{-1/2}$. On a fixed bounded corner window the rescaled score $\eta^{1/2} A_{\boldsymbol{b}} h$, in the mark coordinates, converges to $\mathcal{S}_{e,h}/\mathcal{S}_{e,f_0}$, with bounded values there. This follows from the cutoff kernel estimates and the positive denominator wherever the smaller-cube signal contributes. The marked rare point convergence gives convergence of the centered score sum restricted to this window to the corresponding bounded-window compensated Poisson integral divided by $\sqrt{\lambda}$. The deterministic centering also converges by the same bounded-window density and score convergence. The variance of the centered empirical sum outside these windows is bounded by the omitted score energy. It tends uniformly to zero as the windows grow by the residual operator tail bounds in Theorem 10.4. The analogous statement for the limiting compensated integrals is their variance identity. Finally approximate h_β in the residual Hilbert norm and use the uniform exact operator bound to pass to the actual representers. This proves (9) and its variance formula. Its characteristic exponent is the compensated Poisson exponent with jump size $\lambda^{-1/2}\mathcal{S}_{e,h_\beta}/\mathcal{S}_{e,f_0}$. The jump measure is nonzero since $V_\beta > 0$, and the Lévy–Khintchine uniqueness theorem rules out a Gaussian law.

10.5 Gaussian blocks together with corner Poisson data

Let $\eta = \prod_i b_i^{r_i}$ and suppose all $b_i \downarrow 0$ with $N\eta \to \lambda \in (0,\infty)$. Let f_θ be a fixed finite family of positive bounded noise densities on Q. Choose a fixed member f_{θ_*} as reference. Assume their differences have zero coordinate-slice moments through the K_i in each corresponding coordinate. The family may have dependent coordinates; no interior buffer is needed for differences between the base densities.

The scalar limiting spaces in Theorem 10.3 have three blocks: the bulk moment vector, the critical coordinate when it exists, and the residual moment kernel. Use a fixed finite list of directions v_k in their tensor product. Each direction is a finite sum of elementary tensors with smooth interior-supported residual factors. Assume it has zero component in the block with every axis residual,

and zero coefficient on the tensor of degree-zero moments. For a finite parameter set (θ, a), define the noise density

$$f_{\theta,a,\boldsymbol{b}} = f_\theta + N^{-1/2} D_{\boldsymbol{b}}^{-1} \sum_k a_k v_k.$$

Each perturbation has zero total mass. Its supremum tends to zero: every elementary tensor uses at least one retained moment degree, whose scale divided by $b_i^{r_i/2}$ tends to infinity. Thus the indicated functions are densities for sufficiently small scales.

Theorem 10.7. *The experiments converge in Le Cam distance to the independent product of the corner Poisson process experiment at f_θ from Section 8 and a Gaussian shift with information $I_{k\ell} = \mathcal{Q}_{f_{\theta_*}}(v_k, v_\ell)$ for any fixed member f_{θ_*}. This value is common to the family. If the directions are chosen within individual tensor blocks, the corresponding Gaussian blocks are independent. A possibly singular information matrix I denotes the observation with mean Ia and covariance I.*

Proof. Fix T and split off the product of the bounded windows at all strongest endpoints in each coordinate. It has probability $O_T(\eta)$ uniformly in the parameters. The local perturbation is uniformly $o(1)$ in the noise density, so the limit of the rare counts and marks is the same as for the base f_θ. The binomial count approximation and the product profile calculation are those of Section 8.

Outside the windows replace θ by the fixed reference θ_* at the same local parameter a. The signed difference is unchanged by the common perturbation and is slice-cancelled. The scalar Taylor and tensor estimates of Section 8 bound its squared Hellinger cost outside the windows by $\eta\, o_T(1) + o_{\boldsymbol{b}}(\eta)$. The positivity bounds still hold under every local perturbation. Normalize the exterior densities, whose masses tend to one. The resulting joint error has total variation $o_T(1) + o_{\boldsymbol{b}}(1)$ by the product Hellinger bound.

For a normalized direction put $w_{k,\boldsymbol{b}} = D_{\boldsymbol{b}}^{-1} v_k$. Its score at the fixed reference is $s_{k,\boldsymbol{b}} = T_{\boldsymbol{b}} w_{k,\boldsymbol{b}} / q_{\boldsymbol{b},f_{\theta_*}}$. The full score second moments converge to $I_{k\ell}$ by the operator-norm limit. Removing the fixed corner windows makes a vanishing change in these normalized second moments. Indeed on those windows the squared energy of every elementary tensor is bounded by $C_T \eta$ divided by its squared product scale, using uniform boundedness of the fixed unnormalized factors relative to the positive reference and the $O_T(\eta)$ window probability. That ratio tends to zero by the presence of a retained moment degree. The means of the conditional scores also tend to zero because the signed directions have total mass zero.

Under the reference exterior law, write $r_{\boldsymbol{b},a} = N^{-1/2} \sum_k a_k s_{k,\boldsymbol{b}}$ and let $\mathbb{E}_0$ denote exterior expectation. The local exterior likelihood ratio for one observation is

$$1 + \frac{r_{\boldsymbol{b},a} - \mathbb{E}_0 r_{\boldsymbol{b},a}}{1 + \mathbb{E}_0 r_{\boldsymbol{b},a}}.$$

The displayed centered increment is uniformly $o(1)$, since the actual local density perturbation is uniformly $o(1)$ and the base density has a positive floor. Its summed covariance tends to $a^\top I a$; the joint claim for any finite list of local parameters follows by polarization. The triangular central limit theorem gives the linear log likelihood term, the squared term concentrates by the uniform smallness of the increments, and the cubic remainder tends to zero. Thus the finite likelihood vectors converge to those of the stated Gaussian shift. Their coordinates have mean one, also in the limit, giving uniform integrability. Quantizing the likelihood vectors and regenerating conditional on the bins gives convergence of these finite experiments in Le Cam distance.

Conditionally on the retained count and marks, the number of independent exterior observations is N minus the retained count. Pad the exterior sample to length N by adding the retained count of

draws from the zero-perturbation exterior reference. Each padded draw has squared Hellinger cost $O(1/N)$, by the bounded normalized score norms and the exact ratio formula. The count is tight at fixed T, so the padding cost tends to zero. The reverse kernel discards the same number and restores uniform order. This factors the finite experiments into independent cropped Poisson and Gaussian parts, up to the exterior replacement error. Let $T \to \infty$ and uncrop by the reference-tail Poisson kernel in Section 8. Finally the limiting information is block diagonal in the tensor blocks, by the scalar orthogonal forms, giving the last statement. In each Gaussian block the residual axes are a proper subset. The base family shares every proper noise marginal by its coordinate-slice conditions, so Theorem 10.4 gives the same Gaussian information at every base member. □

10.6 Why the weighted norm uses an interior buffer

The weighted lower bound in Lemma 10.2 holds on the full noise interval. The uniform upper comparison can fail there. For example, choose a minimal right endpoint of exponent $r-1$ and a smooth nonzero p supported in a fixed positive interval with sufficiently many zero moments. Take $u_\epsilon(z) = p((1-z)/\epsilon)$ near the right edge of the noise interval. It has every required zero moment and $\|u_\epsilon\|^2_{H^{-r}} \le C\epsilon^{2r+1}$ by Fourier scaling and those cancellations. In the scaled right observation coordinate $x = (U+b-y)/b$, on $x = \epsilon t$ with t in a suitable fixed bounded positive interval, the reference is comparable to $b^{r-1}\epsilon^r$. The signed output is of size $b^{r-1}\epsilon^r$ on a subinterval of positive relative width, since the fractional primitive of p is nonzero there and the endpoint weight tends to its positive boundary value. The resulting weighted cost on this part alone is $\gtrsim b^r\epsilon^{r+1}$, above the proposed $O(b^r\epsilon^{2r+1})$ by a diverging factor. The exact fixed-moment probes use only the lower weighted comparison, and thus do not need a buffer.

11 Sampling transitions for minimum-variance mixed moment probes

> For small-noise observations from a known product latent density, the minimum-variance unbiased probe of a mixed noise moment need not have a Gaussian limit at its usual normalization. We classify the limit by the endpoint information scale in the moment's coordinate block. Without critical coordinates there is a transition from Gaussian to compensated Poisson to zero. With critical coordinates the limiting law is Gaussian with possibly reduced variance. Its variance fraction is the volume of a halfspace cut of the cube of logarithmic depths. The full variance is recovered whenever the target moment's effective information grows.

11.1 The observation and the canonical probes

Let X have independent coordinates with known densities g_i on bounded intervals $[L_i, U_i]$. Each is smooth and positive in the interior, and near endpoint e has form $w_{i,e}(h)h^{\alpha_{i,e}}$ in inward depth h, with smooth positive $w_{i,e}$ and $\alpha_{i,e} > -1$. Observe N independent copies of $Y_i = X_i + b_i Z_i$, where Z is independent of X, the $b_i \downarrow 0$ can differ, and the unknown joint density f on $Q = [-1,1]^d$ is bounded above and below by positive constants. Use a fixed reference f_0 in this class, with perturbations supported on a fixed strictly smaller cube and of a fixed Sobolev order and radius. No independence of f_0's coordinates is assumed. Throughout, $N \to \infty$ and every $b_i \downarrow 0$.

Set $r_i = 1 + \min_e \alpha_{i,e}$, $K_i = \lfloor r_i/2 \rfloor$. For a nonzero multi-index β, define $m_\beta(f) = \int z^\beta f(z)\,dz$ and

$$k_{\boldsymbol{b}} = \prod_i \sigma_{i,\beta_i}(b_i), \qquad \sigma_{i,j}(b)^2 = \begin{cases} b^{2j}, & 2j < r_i, \\ b^{r_i}\log(1/b), & 2j = r_i, \\ b^{r_i}, & 2j > r_i. \end{cases}$$

We use the minimum-reference-variance centered single-observation probe $a_{\boldsymbol{b}}$ that is exactly unbiased for $k_{\boldsymbol{b}}(m_\beta(f) - m_\beta(f_0))$, as constructed in Section 10. Thus the normalized error at f_0 of its sample-average estimator is

$$S_N = N^{-1/2} \sum_{j=1}^{N} a_{\boldsymbol{b}}(Y_j).$$

Its variance converges to a positive constant V_β. Trabs [42] gives the general score-adjoint interpretation of inverse variance bounds in regular indirect experiments. Here the changing kernel and sample scale make the limit law of the exactly unbiased probe depend on how the data sample the physical observation regions. Chakrabarty and Samorodnitsky [43] study related transitions for deliberately truncated heavy tails. Here the logarithmic truncation comes from the observation kernel and can retain a nontrivial fraction of the reference variance.

We recall the pieces of the coordinate representation that we use. In each axis, split a compactly supported input into smooth lifts $p_{i,j}$ of each moment $j \le K_i$, and a residual whose moments through K_i vanish. The normalized coordinates multiply those moment coefficients by $\sigma_{i,j}$, and multiply the residual by $b_i^{r_i/2}$. After taking tensor products, restrict the Hilbert coordinate space to total mass zero. The exact Gram operators converge in operator norm to the positive coercive block form Q_0 of Section 10. Its blocks assign each coordinate one of the three types bulk, critical or residual. The bulk block in an axis has matrix $\int D_{i,j} D_{i,\ell}/g_i$ for $2j, 2\ell < r_i$, where $D_{i,j} = (-1)^j g_i^{(j)}/j!$. A critical axis carries the sum of the endpoint logarithmic constants. For a set S of residual axes the block is coupled through the joint strongest-endpoint profiles of the marginal $(f_0)_S$. The tensor coordinate map is denoted $D_{\boldsymbol{b}}$; the normalized score map is

$$A_{\boldsymbol{b}} v = \frac{T_{\boldsymbol{b}} D_{\boldsymbol{b}}^{-1} v}{q_{\boldsymbol{b},f_0}}, \qquad q_{\boldsymbol{b},f_0} = T_{\boldsymbol{b}} f_0.$$

In particular $a_{\boldsymbol{b}} = A_{\boldsymbol{b}} h_{\boldsymbol{b}}$, where the representers $h_{\boldsymbol{b}}$ converge in coordinate norm to a fixed h_β. The limit is in the tensor block of β, and $\langle Q_0 h_\beta, h_\beta \rangle = V_\beta$.

11.2 Limit laws by depth and sample size

Let B, C, S be the bulk, critical and residual axes selected by β, respectively. A coordinate is bulk if $2\beta_i < r_i$, critical if $2\beta_i = r_i$, and residual otherwise. Write $\eta_S = \prod_{i \in S} b_i^{r_i}$, with empty product one.

First suppose that $C = \varnothing$. Let e range over the product of the strongest endpoints in the residual axes S. Use mark variables x_B in the physical bulk intervals and $t_S \in \mathbb{R}^{|S|}$ in the residual endpoint windows. Put

$$d\mu_e(x_B, t_S) = \left(\prod_{i \in B} g_i(x_i)\, dx_i \right) \mathcal{S}_{e,(f_0)_S}(t_S)\, dt_S.$$

Here $\mathcal{S}$ is the coordinatewise fractional endpoint profile of Section 8, including the positive taper weights. For a finite smooth v in this tensor block, write $v = \sum_{\mathbf{j}} e_{\mathbf{j}} \otimes v_{\mathbf{j}}$ with $e_{\mathbf{j}}$ the bulk moment basis. Set

$$F_{e,v}(x_B, t_S) = \sum_{\mathbf{j}} \left(\prod_{i \in B} \frac{D_{i,j_i}(x_i)}{g_i(x_i)} \right) \frac{\mathcal{S}_{e,v_{\mathbf{j}}}(t_S)}{\mathcal{S}_{e,(f_0)_S}(t_S)}.$$

Interpret $0/0 = 0$. By the exact block Gram formula this map extends to the entire block with the norm identity $\sum_e \|F_{e,v}\|^2_{L^2(\mu_e)} = \langle Q_0 v, v \rangle$.

Theorem 11.1. *Suppose $C = \varnothing$. If $N\eta_S \to \infty$, then S_N converges to $N(0, V_\beta)$. If $N\eta_S \to \lambda \in (0, \infty)$, then its limit is*

$$\lambda^{-1/2} \sum_e \int F_{e,h_\beta} \, (d\Pi_e - \lambda \, d\mu_e)$$

with independent Poisson measures Π_e of intensity $\lambda\mu_e$. The compensated integral has variance V_β and is non-Gaussian. If $N\eta_S \to 0$, then $S_N \to 0$ in probability.

For critical axes $C \neq \varnothing$, define

$$\ell_i = \log(1/b_i) \ (i \in C), \qquad W = \sum_{i \in C} r_i \ell_i, \qquad A = \log\left(N\eta_S \prod_{i \in C} \ell_i \right).$$

Along a sequence on which $A/W \to a \in [-\infty, \infty]$ and $r_i\ell_i/W \to \rho_i$, set

$$p(a, \rho) = \mathbb{P}\left(\sum_{i \in C} \rho_i U_i \leq a \right),$$

where the U_i are independent uniform variables on $[0, 1]$.

Theorem 11.2. *Suppose $C \neq \varnothing$. Under the indicated limits, S_N converges to a centered Gaussian variable of variance $p(a, \rho)V_\beta$, allowing variance zero. Its actual variance still converges to V_β. In particular, whenever $Nk_{\boldsymbol{b}}^2 \to \infty$, the conclusion for every block is the full-variance Gaussian limit $N(0, V_\beta)$, with no condition on the full product corner sample size.*

For example, with two equal scales b, two critical degrees of the same r, and $N \asymp b^{-r}$, the two log weights are equal and $a = 1/2$. Exactly half of the limiting reference variance appears in the Gaussian law. In this example $Nk_{\boldsymbol{b}}^2 \to 0$, so it does not describe a consistent estimator of the original unscaled moment. For unequal rates the halfspace probability accounts for the different lengths of the logarithmic depth ranges.

11.3 Localization and the proof of Theorem 11.1

It suffices to prove the assertions for $A_{\boldsymbol{b}}v$ with a fixed finite sum v of smooth tensors in the target block, in place of $a_{\boldsymbol{b}}$. Indeed $h_{\boldsymbol{b}} \to h_\beta$ and h_β is approximable by these v within its block. Each score is centered. The difference of the normalized sample sums has variance bounded by $C\|h_{\boldsymbol{b}} - v\|^2$ by the uniform Gram bound. Thus the approximation can be removed at the end, in probability for real limits and in mean square between fixed approximants.

For bulk coordinates retain only depths at least $\delta > 0$. In residual coordinates retain the strongest endpoint windows of rescaled radius T. In later proofs, retain in critical coordinates only the inner depths in $[Tb_i, \delta]$ at the strongest endpoints. For each of these block rectangles consider the statistic with its restricted score centered separately. By the exact block localization in Section 10, the difference from the full centered score sum has squared mean bounded by a norm error tending to zero as all $b_i \downarrow 0$, then $T \to \infty$ and $\delta \downarrow 0$. This also follows directly by combining the scalar bulk dominated bound, critical logarithmic energy concentration, and residual Taylor tail estimate, then comparing the dependent denominator to its product envelope. All references to cutoffs below are at fixed T, δ before these last limits.

If $C = \varnothing$, the probability of a retained rectangle is $O_{T,\delta}(\eta_S)$. In its mark coordinates its reference density divided by η_S converges in L^1 to the restriction of μ_e. The signal, after multiplying its score by $\sqrt{\eta_S}$, converges to $F_{e,v}$ there and is uniformly bounded. For bulk axes use the normalized Taylor

lift on the retained physical interval. For residual axes use the endpoint kernel and its fixed-window normal profile. On the smaller-cube signal the mark denominator is bounded below on the fixed window, and it has the indicated limit by positivity and factoring out the bulk g_i. These facts also show convergence of the restricted score second moments to the corresponding restricted Gram form.

If $N\eta_S \to \infty$, the largest possible normalized individual increment is $O_{T,\delta}((N\eta_S)^{-1/2})$. The centered triangular central limit theorem gives the restricted Gaussian limit. For nonempty S the centered subtraction in its variance tends to zero since the window probability tends to zero. If $S = \varnothing$, the full score has mean zero and the cutoff localization makes any missing mean negligible as $\delta \downarrow 0$. The cutoff and Hilbert approximations give the full Gaussian claim.

If $N\eta_S \to \lambda \in (0, \infty)$, S is nonempty. The counts in the retained windows and their marks converge to the indicated Poisson measures on their bounded regions, by the binomial Poisson approximation and the normalized mark convergence. Centering the restricted score sum converges to compensating the intensity integral, by bounded convergence. This gives the compensated Poisson limit on the retained marks. Removing the cutoffs passes to the full integrals in L^2 by the variance identity and the block norm bound. Their variance is V_β. The jump measure of the compensated Poisson integral is nonzero because $V_\beta > 0$, so its law is non-Gaussian by Lévy–Khintchine uniqueness. If $N\eta_S \to 0$, the probability of any retained mark tends to zero. The absolute deterministic centering of the normalized retained sum is at most $C_{T,\delta}\sqrt{N\eta_S}$ by the same probability and score bounds. It also tends to zero. The cutoff and Hilbert approximations prove the last case.

11.4 The logarithmic depth cutoff

Assume $C \neq \varnothing$, and retain the product rectangles defined above. For critical axes introduce $\lambda_i = r_i \log(1/h_i)$, where h_i is the physical depth from the chosen strongest endpoint. The allowed range of λ_i is

$$r_i \log(1/\delta) \le \lambda_i \le r_i \ell_i - r_i \log T.$$

In these variables the reference density on the retained region, integrated over the bounded bulk and residual marks, is bounded by

$$C_{T,\delta}\eta_S \exp\left(-\sum_{i\in C}\lambda_i\right)\prod_{i\in C} d\lambda_i.$$

Indeed $g_i(y_i)\,dy_i \asymp e^{-\lambda_i} d\lambda_i$ on the critical part, and the dependent reference is comparable to a product reference. For a fixed finite smooth v, the magnitude of a single score increment divided by $\sqrt{N}$ there is at most

$$C_{T,\delta,v} \exp\left(\tfrac{1}{2}\left[\sum_{i\in C}\lambda_i - A\right]\right). \tag{10}$$

Bulk factors are bounded. Residual factors cost at most $C_T \eta_S^{-1/2}$, and in a critical coordinate the normalized score is at most $C h_i^{-r_i/2}/\sqrt{\ell_i}$ by the Taylor derivative bound on the interval $h_i \ge T b_i$. Together these prove (10).

For $w = \sqrt{W}$, separate the rectangles at $\sum_i \lambda_i = A - w$ and $A + w$. On the upper part the expected number of observations is at most $C_{T,\delta} e^{-w}$: integrate the reference bound using $e^{-\sum \lambda_i} \le e^{-A-w}$ and the enclosing volume $C\prod_i \ell_i$. The deterministic centering of the normalized sum on this part is at most $C_{T,\delta,v} e^{-w/2}$, by combining the reference bound with (10) and using

$e^{-\sum \lambda_i/2} \le e^{-(A+w)/2}$. Thus this part vanishes in probability even when it carries nonvanishing second moment.

The middle part has vanishing variance. Multiply the reference bound by the squared score envelope before the sample summation. This bounds its variance by a constant times its logarithmic volume divided by $\prod_i \ell_i$. At least one coordinate has log width of order W; holding the others fixed, the slab has width at most $2w$ in that coordinate. Its normalized volume is $O(w/W) = o(1)$. The centering can only reduce variance.

On the lower part individual normalized increments are bounded by $C_{T,\delta,v} e^{-w/2}$. Its limiting variance is the fraction $p(a,\rho)$ of the block energy, up to the cutoff errors. Here is the calculation. Write the reference denominator on the rectangle as the product of the nonresidual g_i's and the residual marginal output, with a uniform relative error $O(T^{-1}) + o_{\boldsymbol{b}}(1)$, as in Section 10. In a critical coordinate $2J_i = r_i$ the leading lift score on $Tb_i \le h_i \le \delta$ is

$$\frac{D_{i,J_i}(y_i)}{\sqrt{\ell_i}\, g_i(y_i)} = \frac{c_{i,e} h_i^{-r_i/2}}{\sqrt{\ell_i}} \, (1 + O(\delta)),$$

where $c_{i,e}$ is the signed nonzero binomial coefficient at the chosen endpoint. The remaining Taylor terms have vanishing L^2 information cost after normalization, as in the scalar Gram proof of Section 10. Squaring and integrating against $g_i(y_i)\, dy_i$ gives constant energy density in $d\lambda_i/(r_i \ell_i)$ up to $O(\delta)$. The constants multiply, and the bulk and residual coordinates give the cutoff versions of their block Gram forms. The normalized logarithmic volume of the lower region tends to $p(a,\rho)$: rescale each λ_i by $r_i \ell_i$, note that the removed end ranges have vanishing relative length, and use continuity of the distribution of $\sum_i \rho_i U_i$, since at least one weight is positive. This argument includes a infinite.

For the centered variance the first-moment subtraction is negligible. Indeed integrating the absolute score envelope against the reference bound, before the sample normalization, gives at most $C_{T,\delta,v} \sqrt{\eta_S} / \sqrt{\prod_i \ell_i}$ on the whole rectangle, using the finite integral of $\exp(-\sum_i \lambda_i/2)$. The triangular central limit theorem thus applies on the lower region, with the variance just calculated. One may take the cutoff limits using characteristic functions: the relative denominator and endpoint errors are $O(T^{-1} + \delta)$, and the L^2 score localization error tends to zero when these cutoffs are removed. Hence the full law for a fixed smooth v is the centered Gaussian with variance $p(a,\rho)\langle Q_0 v, v\rangle$. The Hilbert approximation extends the conclusion to the canonical score and proves Theorem 11.2.

Finally, if $N k_{\boldsymbol{b}}^2 \to \infty$ with no critical axes, $N\eta_S \to \infty$ since the bulk factors in $k_{\boldsymbol{b}}^2$ are at most one. If there are critical axes, the same comparison gives

$$N k_{\boldsymbol{b}}^2 \le N \eta_S \prod_{i \in C} (b_i^{r_i} \ell_i) = e^{A-W}.$$

Thus $A - W \to \infty$, and every convergent subsequence of the log-width proportions gives $p = 1$. This proves the full variance assertion without requiring the log-width proportions to converge in this case.

11.5 Empirical variance and studentization

Let $\widehat{V}_N$ be the empirical centered second moment of the probe values $a_{\boldsymbol{b}}(Y_j)$. In the critical case of Theorem 11.2, $\widehat{V}_N \to p(a,\rho) V_\beta$ in probability, although the mean of the uncentered empirical second moment tends to V_β. If $p(a,\rho) > 0$, studentization by $\sqrt{\widehat{V}_N}$ therefore gives a standard normal limit. In the noncritical Gaussian regime the corresponding variance limit is V_β. In the finite Poisson regime, the statistic and empirical variance converge jointly to the compensated

integral of Theorem 11.1 and to $\lambda^{-1}\sum_e \int F_{e,h_\beta}^2\, d\Pi_e$, respectively. In the vanishing regime the empirical variance tends to zero.

For a fixed approximate score and fixed cutoffs, the upper log region contains no observations with high probability. The middle region contributes $o(1)$ expected empirical second moment. In the lower region the second moment concentrates: the variance of its empirical second moment is bounded by the largest squared score divided by N, times the bounded expected second moment. The first factor tends to zero by the lower-region envelope. The expected value has the same limit calculated for the retained Gaussian variance. The noncritical Gaussian regime has the same concentration by its uniform increment bound; the finite Poisson regime uses the same marked-process convergence applied also to the squared score, and in the vanishing regime there are no retained observations with high probability. The discarded second moments are bounded in expectation by the corresponding operator localization error. Removing a Hilbert approximation changes the empirical second moment by at most the empirical L^2 distance times the sum of the two empirical L^2 norms. The first is small in mean square, and the second is bounded in mean square, by the exact uniform Gram bound. These facts remove both cutoffs and approximations. Finally subtracting the square of the empirical mean changes the second moment by $S_N^2/N = o_P(1)$. The studentized claim follows.

11.6 Joint limits of finitely many probes

For any fixed finite list of moments, the same limits hold jointly whenever the sampling conditions converge for their occupied blocks. Within a block, the Gaussian covariance is its limiting Gram covariance, multiplied by the common logarithmic fraction in a critical block. A Poisson block uses the same processes Π_e for all its probe integrands. Distinct blocks are independent in the limit.

To prove this, approximate each representer in its own block as before. For a linear combination of probes in one block repeat the preceding proof with the resulting finite smooth sum v; its block type, sample size and depth fraction are unchanged. This gives the within-block vector limit. For independence fix the localization cutoffs simultaneously. The observation rectangles assigned to distinct blocks are disjoint in at least one coordinate, up to their null boundaries. Let $X_{i,\boldsymbol{b}}$ be one restricted, not yet centered, one-observation real linear combination in block i. The characteristic function for one observation has the form $1+\sum_i d_{i,\boldsymbol{b}}$, where $d_{i,\boldsymbol{b}} = \mathbb{E}(e^{iX_{i,\boldsymbol{b}}/\sqrt{N}}-1)$. For the product of separate block experiments it would instead be $\prod_i(1+d_{i,\boldsymbol{b}})$. Each $d_{i,\boldsymbol{b}}$ equals $i\mathbb{E}X_{i,\boldsymbol{b}}/\sqrt{N}+O(1/N)$, by the uniform score second moments. Except for a possible all-bulk block the restricted means tend to zero. For residual axes this follows by Cauchy–Schwarz and the vanishing rectangle probability, and for a critical-only and bulk rectangle it follows from the logarithmic first-moment bound. There is only one all-bulk block. Hence $Nd_{i,\boldsymbol{b}}d_{j,\boldsymbol{b}} \to 0$ for $i \neq j$. The N-fold characteristic functions, after the common deterministic centering, differ by $o(1)$. If empirical second moments are included, the per-observation exponent gains a term of the form $Z_{i,\boldsymbol{b}}/N$ with uniformly bounded $\mathbb{E}|Z_{i,\boldsymbol{b}}|$. This follows from the uniform score second moments and Cauchy–Schwarz for cross products. The change in $d_{i,\boldsymbol{b}}$ is then only $O(1/N)$, leaving the independence argument unchanged. Removing the observation and approximation cutoffs proves the claimed independent joint limits.

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] L. Aarts, P. Groeneboom and G. Jongbloed, Estimating the upper support point in deconvolution, *Scand. J. Stat.* **34** (2007), 552–568. `https://doi.org/10.1111/j.1467-9469.2006.00545.x`.

[3] A. Meister, Deconvolving compactly supported densities, *Math. Methods Statist.* **16** (2007), 63–76. `https://doi.org/10.3103/S106653070701005X`.

[4] S. Schubr, *Testing and inverse stability at piecewise log-affine and uniform laws under Gaussian noise*, the part titled *Gaussian covariance information at faces and edges*, Luminary Research Archive, 2026.

[5] N. Brosse and A. Dalalyan, *Boundary-layer asymptotics for Gaussian-smoothed singular measures.* arXiv:2607.04514v1.

[6] A. Meister and M. Reiss, *Asymptotic equivalence for nonparametric regression with non-regular errors.* arXiv:1101.5248.

[7] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press (1998), Chapters 7 and 9. DOI.

[8] A. J. Hoffman, *On approximate solutions of systems of linear inequalities*, J. Res. Natl. Bur. Standards 49 (1952), 263–265. Original paper.

[9] S. Robins, *Fourier Analysis on Polytopes and the Geometry of Numbers*, Student Mathematical Library 107, AMS (2024). DOI.

[10] J. A. Cuesta-Albertos, R. Fraiman and T. Ransford, *A sharp form of the Cramér–Wold theorem*, J. Theoret. Probab. 20 (2007), 201–209. DOI.

[11] M. Benešová, B. van Es and P. Tegelaar, Bivariate uniform deconvolution, Statistics 50 (2016), 812–840. Author manuscript.

[12] M. Brion and M. Vergne, Residue formulae, vector partition functions and lattice points in rational polytopes, J. Amer. Math. Soc. 10 (1997), 797–833. Author manuscript.

[13] Y. I. Ingster, T. Sapatinas and I. A. Suslina, Minimax signal detection in ill-posed inverse problems, Ann. Statist. 40 (2012), 1524–1549. Author manuscript.

[14] S. Schubr, *Sharp Wasserstein Stability for Gaussian Deconvolution of Log-Concave Laws*, Luminary Research Archive, edition 4. Archive edition.

[15] G. Bianchi, F. Segala and A. Volčič, The solution of the covariogram problem for plane C^2_+ bodies, J. Differential Geom. 60 (2002), 177–198. DOI.

[16] J. Dick, M. Ehler, M. Gräf and C. Krattenthaler, Spectral decomposition of discrepancy kernels on the Euclidean ball, the special orthogonal group, and the Grassmannian manifold, Constr. Approx. 57 (2023), 983–1026. DOI.

[17] B. Galerne, Computation of the perimeter of measurable sets via their covariogram, Image Anal. Stereol. 30 (2011), 39–51. DOI.

[18] 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.

[19] D. Belomestny and A. Goldenshluger, Density deconvolution under general assumptions on the distribution of measurement errors, Ann. Statist. 49 (2021), 615–649. Author manuscript.

[20] L. Cavalier, Efficient estimation of a density in a problem of tomography, Ann. Statist. 28 (2000), 630–647. DOI.

[21] A. Feuerverger, P. T. Kim and J. Sun, On optimal uniform deconvolution, J. Statist. Theory Practice 2 (2008), 433–451. Author manuscript.

[22] S. Park and D. Slepčev, Geometry and analytic properties of the sliced Wasserstein space, J. Funct. Anal. 289 (2025), 110975. Author manuscript.

[23] D. Beltran, J. Roos and A. Seeger, A note on endpoint Bochner–Riesz estimates, Section 6. Author manuscript.

[24] C. Butucea, Goodness-of-fit testing and quadratic functional estimation from indirect observations, Ann. Statist. 35 (2007), 1907–1930. Author manuscript.

[25] R. L. Smith, Maximum likelihood estimation in a class of nonregular cases, Biometrika 72 (1985), 67–90. DOI.

[26] M. Carrasco and J.-P. Florens, A spectral method for deconvolving a density, Econometric Theory 27 (2011), 546–581. Author manuscript.

[27] NIST Digital Library of Mathematical Functions, Chapter 10, integral 10.9.4 and asymptotic 10.17.3.

[28] I. Daubechies, *Ten Lectures on Wavelets*, SIAM (1992). DOI.

[29] J. Dedecker, A. Fischer and B. Michel, Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one, Electron. J. Statist. 9 (2015), 234–265. Author manuscript.

[30] B. Laurent, J.-M. Loubès and C. Marteau, Non asymptotic minimax rates of testing in signal detection with heterogeneous variances, Electron. J. Statist. 6 (2012), 91–122. Author manuscript.

[31] Y. Meyer, *Wavelets and Operators*, Cambridge University Press (1992).

[32] J. Rousseau and C. Scricciolo, Wasserstein convergence in Bayesian and frequentist deconvolution models, Ann. Statist. 52 (2024), 1691–1715. Author manuscript.

[33] F. Comte and C. Lacour, Anisotropic adaptive kernel deconvolution, Ann. Inst. Henri Poincaré Probab. Stat. 49 (2013), 569–609. DOI.

[34] S. Park, K.-S. Moon and H. Kim, Minimax rates for conditional generative learning under Wasserstein loss. arXiv:2610.35490v1.

[35] C. Scricciolo, Adaptive minimax multivariate L^1-deconvolution: harmonic-mean rates under noise filtering, and a 1-Wasserstein equivalence. arXiv:2610.11760v1.

[36] R. Benhaddou, M. Pensky and D. Picard, Anisotropic de-noising in functional deconvolution model with dimension-free convergence rates, Electron. J. Statist. 7 (2013), 1686–1715. DOI.

[37] G. Cleanthous, A. G. Georgiadis and E. Porcu, Minimax density estimation on Sobolev spaces with dominating mixed smoothness. Author manuscript.

[38] L. Ding, L. Zou, W. Wang, S. Shahrampour and R. Tuo, High-dimensional non-parametric density estimation in mixed smooth Sobolev spaces. Author manuscript.

[39] S. Park, K.-S. Moon and H. Kim, Minimax rates for conditional generative learning under Wasserstein loss. Author manuscript.

[40] P. Hall and S. N. Lahiri, Estimation of distributions, moments and quantiles in deconvolution problems, Ann. Statist. 36 (2008), 2110–2134. Author manuscript.

[41] M. Pensky, Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity, Ann. Statist. 45 (2017). DOI.

[42] M. Trabs, Information bounds for inverse problems with application to deconvolution and Lévy models, Ann. Inst. Henri Poincaré Probab. Stat. 51 (2015), 1620–1650. DOI.

[43] A. Chakrabarty and G. Samorodnitsky, Understanding heavy tails in a bounded world or, is a truncated heavy tail heavy or not?, Stochastic Models 28 (2012), 109–143. DOI.