Sharp minimax rates for Gaussian location mixtures with unknown covariance

Scott Schubr
Luminary Oddities

November 2026

Abstract

We determine the minimax rates for finite Gaussian location mixtures whose common noise covariance is unknown, the case that Doss, Wu, Yang and Zhou note is open beyond the isotropic case. For at most $k \geq 2$ components with means in a fixed ball and a bounded common covariance, write $r = 1 \wedge \sqrt{d/n}$. The sharp expected W_p error for the mixing law is $r^{1/\max(2k,p)}$ for every fixed finite $p \geq 1$, the sharp covariance error is $r^{1/k}$ in operator norm, and the sharp Frobenius error is $(d/\sqrt{n} \wedge \sqrt{d}) + r^{1/k}$; one proper fit attains the mixing-law and operator rates together. When the unknown covariance is isotropic and $d \geq k$, the noise variance is estimable at the parametric rate $(nd)^{-1/2}$ and the mixing law at the same rate as with known noise, $r^{1/\max(2,p)} + n^{-1/[2\max(2k-1,p)]}$; below the rank threshold a polynomial bound for the variance is sharp whenever a Gaussian cubature exists, and in the plane the variance rate is $n^{-1/(2m)}$ up to constants for every $k \leq 35$, where m is the least integer with $\binom{m+2}{2} > k$, so that the rate drops at $k = 3, 6, 10, 15, 21, 28$. We then add sparsity and Huber contamination: testing one Gaussian against a mixture whose means differ on s unknown coordinates has sharp W_p separation $\big(1 \wedge [\sqrt{s\log(ed/s)/n} + \epsilon]\big)^{1/\max(2k,p)}$, and the same comparison gives sharp robust estimation. For the isotropic variance under Huber contamination and $d \geq k$, the minimax absolute risk is $(nd)^{-1/2} + \min\{1, \psi(\epsilon)/\sqrt{d}\} + \beta_n(\epsilon)$, where $\psi(\epsilon) = \epsilon\sqrt{1 + \log(1/(1-2\epsilon))}$ and $\beta_n(\epsilon)$ is the probability of a contaminated majority among n draws; for fixed n the risk tends to exactly $(b-a)\beta_n(\epsilon)$ as the dimension grows, and the exact limit is identified for every power loss. Finally, when the covariance lies in a Loewner interval of width ω below a known matrix, the sharp separation interpolates between the known and unknown covariance rates, with additional regimes between the moment knots $2, 4, \ldots, 2k-2, 2k-1$. The upper bounds rest on directional moment fitting and a uniform differential-equation comparison between densities and distribution functions; the lower bounds rotate a Gaussian quadrature and absorb its second moment into the noise covariance.

Contents

1 Introduction **2**

2 Sharp minimax rates for Gaussian location mixtures with unknown common covariance **4**

2.1 Model and results 4
2.2 Directional moment fitting 5
2.3 Transport from scalar observed moments 6
2.4 Rotated quadrature lower bounds 7
2.5 Frobenius covariance loss 7

3 Sharp calibration of unknown isotropic variance in Gaussian location mixtures 8
3.1 Results . . . 8
3.2 Independent residual energy . . . 9
3.3 Fitting the mixing law . . . 10
3.4 Mixing-law lower bounds . . . 11
3.5 Calibrating below the rank threshold . . . 12

4 Sharp robust detection of sparse Gaussian location mixtures 13
4.1 Setting and results . . . 13
4.2 A linear moment bound from distribution functions . . . 14
4.3 Upper bounds . . . 15
4.4 Sparse testing lower bound . . . 16
4.5 The contamination lower bounds . . . 17

5 Testing Gaussian mixtures with a partially known covariance 17
5.1 Model and separation . . . 18
5.2 An odd-moment bound for the far tail . . . 18
5.3 Testing upper bound . . . 19
5.4 A testing bound for planted directions . . . 20
5.5 Separated scalar alternatives . . . 20

6 Sharp isotropic variance rates for planar Gaussian mixtures 22
6.1 Variance estimation and moment matching . . . 22
6.2 Six nodes and fifth moments . . . 24
6.3 Ten nodes and seventh moments . . . 24
6.4 Fifteen nodes and ninth moments . . . 26
6.5 Twenty-one and twenty-eight nodes . . . 27
6.6 Comparison with ordinary Gaussian cubature . . . 29

7 Robust isotropic noise calibration and the finite-sample majority limit 29
7.1 Model and result . . . 30
7.2 Lower bounds . . . 30
7.3 Norm medians with an exact majority term . . . 31
7.4 Removing latent energy in moderate dimension . . . 33
7.5 Exact large-dimension limits for power loss . . . 34
7.6 The majority transition . . . 34

1 Introduction

Doss, Wu, Yang and Zhou [6] determined the minimax rate $\big((d/n)^{1/4}\wedge 1\big)+n^{-1/(4k-2)}$ for estimating the mixing law of a k-component Gaussian location mixture in dimension d with known noise. In their Theorem 6.1 they also treat an unknown isotropic variance and state the rates $n^{-1/(4k)}$ for the mixing law and $n^{-1/(2k)}$ for the variance as minimax optimal. In their Section 6 they note that, beyond the isotropic case, the optimal rate with unknown covariance is open. Section 2 settles this case for covariances with bounded spectrum. The unknown covariance enters the mixing-law rate through the power $1/(4k)$ of d/n, giving the W_1 rate $(d/n)^{1/(4k)}\wedge 1$, so the dimension and the number of components now interact; the operator-norm covariance error is $(d/n)^{1/(2k)}\wedge 1$, and the

Frobenius error picks up the additional Gaussian term $d/\sqrt{n} \wedge \sqrt{d}$. The rates hold for every fixed W_p and uniformly in d and n, with no separation or minimum-weight conditions.

Section 3 returns to the isotropic model. The rates $n^{-1/(2k)}$ for the variance and $n^{-1/(4k)}$ for the mixing law in [6, Theorem 6.1] are the optimal rates in dimension one, established by Wu and Yang [5]; the lower bounds of [6, Theorem 6.1] embed this one-dimensional construction. In dimension $d \geq 2$ the orthogonal coordinates help to calibrate the noise, and the results of Section 3 show that these rates are not optimal for $d \geq k$, and, for the variance, for every fixed $d \geq 2$. Once $d \geq k$, the centered latent covariance has rank below the dimension, and the smallest eigenvalue of the observation covariance is the noise variance, an identity due to Hsu and Kakade [10, Theorem 1]; averaging trailing eigenvalues is the classical noise estimate of probabilistic principal component analysis [11, 12]. The contribution here is the uniform minimax variance rate $(nd)^{-1/2}$, with no eigenvalue gap, together with the mixing-law rate, which is the same, up to constants, as with known noise. Below the rank threshold the variance rate is controlled by a polynomial-vanishing bound under a heat flow, sharp whenever a Gaussian cubature exists and whenever $\lceil 3d/2 \rceil \leq k < \binom{d+2}{2}$; these statements concern the variance. The planar three-component case gives the intermediate rate $n^{-1/4}$ through an explicit pair of triangles with equal third moment tensors. Section 6 completes the planar picture through 35 components. At each threshold it builds two rotationally symmetric atomic laws whose moments agree through degree $2m - 1$ once one of them is smoothed by an isotropic Gaussian. These pairs use fewer nodes than any positive Gaussian cubature of the same degree, so the statistical lower bound is reached at component counts where a point-mass reference cannot reach it.

Section 4 treats detection. Verzelen and Arias-Castro [4] found the sparse detection boundary for two components with unknown covariance. We allow any number of components, every W_p separation and Huber contamination, with the threshold taken uniformly over the mixing weights; for two balanced components it recovers the symmetric boundary of [4, Section 3.1], while for weights bounded away from one half their boundary in [4, Section 3.2] is faster. The threshold has the same form as the estimation rate with d replaced by $s \log(ed/s)$ and an additive contamination level. Section 7 calibrates the isotropic variance under Huber contamination. For $d \geq k$, besides the sampling term and the contamination modulus of Chen, Gao and Ren [1], which in this model grows logarithmically as the contamination fraction approaches one half, the expected loss carries a binomial majority term: with a contaminated majority among n draws no estimator can tell the endpoints apart. A norm median, which does not need the contamination fraction, attains the resulting finite binary Bayes risk exactly in high dimension, and for power losses an order-statistic correction that uses the contamination fraction attains the exact limit. Section 5 interpolates between known and unknown covariance: an odd-moment polynomial controls far atoms, and a quadrature with prescribed remote nodes produces the intermediate regimes. The covariance-confounding lower bounds in Sections 2, 4 and 5 use the directional quadrature construction of Diakonikolas, Kane and Stewart [3] at vanishing amplitude; the remaining lower bounds use sign-matrix packings, rare-mass and finite-difference pairs, the planar paired laws of Section 6, and two-point Gaussian experiments. Concurrent work of He and Cheng [7] treats proper estimation in polynomial time with known covariance under the Hellinger distance. Earlier Luminary work on Gaussian deconvolution of log-concave laws [2] uses windowed cosine perturbations that fix mass, mean and covariance; here quadrature matches moments through degree $2k - 1$ instead.

This is the third edition of this article. The third edition follows an independent referee review: it corrects the description of the results of Doss, Wu, Yang and Zhou, removes an incorrect remark that the isotropic rates reduce to theirs in dimension one, states the scope of several statements precisely, and adds references. The second edition added Section 7.

2 Sharp minimax rates for Gaussian location mixtures with unknown common covariance

We determine the minimax rates for finite Gaussian location mixtures with a shared unknown anisotropic covariance, the case that Doss, Wu, Yang and Zhou note is open in their discussion of location-scale mixtures (Section 6). For at most $k \geq 2$ components with means in a fixed ball and bounded covariance, the minimax expected W_1 error for the mixing law is $1 \wedge (d/n)^{1/(4k)}$. The minimax covariance error in operator norm is $1 \wedge (d/n)^{1/(2k)}$. We also give sharp rates under W_p for every fixed finite $p \geq 1$ and for Frobenius covariance loss. The upper bounds use robust directional moments on a fixed sphere net. For the lower bounds, we rotate a Gaussian quadrature rule and offset its second moment in the noise covariance. Every alternative then agrees with the same Gaussian reference through degree $2k-1$.

2.1 Model and results

We observe n independent copies of $X = U + Z$, where $U \sim \mu$ has at most k atoms in the Euclidean ball of radius $R > 0$ in $\mathbb{R}^d$, and $Z \sim N(0, \Sigma)$ is independent of U. Fix $k \geq 2$ and $0 \leq \lambda_- < \lambda_+ < \infty$, and require $\lambda_- I \preceq \Sigma \preceq \lambda_+ I$. Here W_p uses Euclidean distance, $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm, and $\|\cdot\|_F$ denotes the Frobenius norm. Write $r = 1 \wedge \sqrt{d/n}$.

Theorem 2.1. *Uniformly over $d, n \geq 1$, for each fixed $1 \leq p < \infty$, the minimax expected losses are*

$$\inf_{\widehat{\mu}} \sup_{\mu,\Sigma} \mathbb{E} W_p(\widehat{\mu}, \mu) \asymp r^{1/\max(2k,p)},$$
$$\inf_{\widehat{\Sigma}} \sup_{\mu,\Sigma} \mathbb{E}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}} \asymp r^{1/k},$$
$$\inf_{\widehat{\Sigma}} \sup_{\mu,\Sigma} \mathbb{E}\|\widehat{\Sigma} - \Sigma\|_F \asymp (d/\sqrt{n} \wedge \sqrt{d}) + r^{1/k}.$$

The comparison constants depend only on $p, k, R, \lambda_-, \lambda_+$. One proper joint fit attains the mixing-law and operator-norm rates.

Section 6 of Doss, Wu, Yang and Zhou [6] notes that the optimal rate with unknown common covariance beyond the isotropic case is open ("Location-scale mixtures"); Theorem 2.1 settles this case for bounded spectrum. Their Theorem 6.1 gives isotropic upper bounds $((d/n)^{1/4} \wedge 1) + n^{-1/(4k)}$ for the mixing law and $n^{-1/(2k)}$ for the variance, and states them as minimax optimal, using the one-dimensional lower bounds of Wu and Yang. In dimension $d \geq k$, the isotropic variance rate is $(nd)^{-1/2}$ and the latent rate is the same as under known noise (Section 3), so those rates are optimal in dimension one but not for $d \geq k$. When the covariance is arbitrary, the dimension enters the $1/(4k)$ power. Our estimator uses an exponentially large net. Recent work of Bariletto, Le, Rinaldo and Ho [8] gives logarithmic Bayesian recovery bounds with arbitrary compactly supported mixing laws. Agostini, Améndola and Ranestad [9] study algebraic identifiability, including generic two-component recovery through moments of order four. We need a uniform comparison that remains valid through collisions and vanishing weights.

The upper bound uses Wu and Yang's scalar moment comparison with unknown variance [5, Proposition 3] and the single-direction lift in the proof of Doss et al. [6, Lemma 3.1], namely property (B.1) in their Appendix B. The lower bound uses a vanishing-amplitude version of the directional planting of Diakonikolas, Kane and Stewart [3, Section 4]. The Luminary Research Archive's deconvolution paper [2] uses Gaussian-windowed cosine perturbations that fix the mass, mean and covariance of a log-concave reference before smoothing with a known kernel. Here

quadrature matches moments through degree $2k-1$ and a direction-dependent covariance absorbs the signal's second moment. The scalar upper bound for general p uses both a polynomial vanishing to second order at the atoms and a transport bound across the successive gaps.

2.2 Directional moment fitting

For unit v write $m_j(v) = \mathbb{E}(v^\top X)^j$. The difference of two candidate jth moments is a homogeneous polynomial p of degree j. Its associated symmetric multilinear form T satisfies

$$\sup_{\|v_i\|\le 1} |T(v_1,\dots,v_j)| \le (j^j/j!) \sup_{\|v\|=1} |p(v)|.$$

Indeed polarization averages $p(\sum_i \epsilon_i v_i)\prod_i \epsilon_i/j!$ over independent signs. Telescoping its j inputs shows that the Lipschitz constant of p on the unit sphere is at most $j(j^j/j!)\sup|p|$. Choose a fixed $\varepsilon_k > 0$ small enough that every ε_k-net $\mathcal{N}$ of the sphere satisfies

$$\sup_{\|v\|=1} |p(v)| \le 2 \max_{v\in\mathcal{N}} |p(v)|, \qquad 1 \le j \le 2k. \tag{1}$$

A volumetric net has $|\mathcal{N}| \le (1+2/\varepsilon_k)^d$.

Take an odd block count $B \asymp d$, with $B \ge 3$, and first suppose $n \ge 2B$. Set $b = \lfloor n/B \rfloor$ and use B disjoint blocks of size b. For every net direction v and $1 \le j \le 2k$ take the median $\widetilde{m}_j(v)$ of the block means of $(v^\top X)^j$. The variables averaged in each block have variance at most $V = V(k, R, \lambda_+)$, since $v^\top U$ is bounded by R and the added scalar Gaussian has variance at most λ_+. If $M = \max_{v\in\mathcal{N}, j\le 2k} |\widetilde{m}_j(v) - m_j(v)|$, Chebyshev's inequality and a union bound over subsets of at least half the independent blocks give

$$\mathbb{P}\{M > t\sqrt{B/n}\} \le 2k|\mathcal{N}|2^B(2V/t^2)^{B/2} \le (C/t)^B \qquad (t \ge \sqrt{2V}).$$

Integrating the minimum of this bound and one yields $\mathbb{E}M \lesssim \sqrt{d/n}$.

Fit one admissible pair by minimizing the maximum absolute discrepancy from these net estimates, to additive accuracy $n^{-1/2}$. The parameter space is compact when parameterized by weights, atom positions and covariance entries. Using the first qualifying element of an enumerated dense sequence gives a measurable approximate minimizer. Since the true pair is feasible, (1) gives

$$\delta := \max_{j\le 2k} \sup_{\|v\|=1} |\widehat{m}_j(v) - m_j(v)| \le 4M + 2n^{-1/2}. \tag{2}$$

For the projections $\mu_v = (v^\top)_\# \mu$ and $\widehat{\mu}_v$, the scalar comparison in Lemma 2.2 below now gives

$$W_p(\widehat{\mu}_v, \mu_v) \lesssim \delta^{1/\max(2k,p)}, \qquad |v^\top(\widehat{\Sigma} - \Sigma)v| \lesssim \delta^{1/k}. \tag{3}$$

The hypotheses are bounded scalar means and variances, at most k atoms and raw moment control through order $2k$. Cases of zero variance or fewer distinct atoms follow by continuity. The inequalities for $\delta > 1$ follow from bounded losses.

To lift the first bound, take the span of the nonzero differences between points in the two supports. There are at most $\binom{2k}{2}$ differences to consider, up to sign, and the span has dimension at most $2k-1$. If its dimension is at least two, a uniform random unit vector in this span lies in each band $\{v : |v^\top z| < c\|z\|\}$ with probability at most $C_k c$ for $c \le 1/2$, by the first-coordinate density of a uniform sphere point. A union bound gives one direction retaining at least a fraction $c_k > 0$ of every pairwise distance. The same holds along the span in dimension one, and in dimension

zero the laws agree. The projection is injective on the union of the supports. Lifting its optimal coupling hence gives

$$W_p(\widehat{\mu}, \mu) \le c_k^{-1} \sup_{\|v\|=1} W_p(\widehat{\mu}_v, \mu_v).$$

The supremum of the second bound in (3) is the operator norm. Jensen's inequality with (2) gives both rates. If $n < 2B$, a fixed admissible pair suffices since r is bounded below by a constant.

2.3 Transport from scalar observed moments

Lemma 2.2. *Let μ, ν be laws on $[-R, R]$ with at most k atoms each. Suppose the first $2k$ moments of $\mu * N(0, \sigma^2)$ and $\nu * N(0, s^2)$ differ by at most δ, with both variances bounded. Then for $1 \le p < \infty$,*

$$|\sigma^2 - s^2| \le C\delta^{1/k}, \qquad W_p(\mu, \nu) \le C\delta^{1/\max(2k,p)}.$$

Proof. Assume $0 < \delta \le 1$ by boundedness and continuity, and relabel so $\sigma^2 \ge s^2$. Undo the smaller Gaussian variance in the moment equations using the coefficients of $\exp(tx - s^2t^2/2)$. Then the first $2k$ moments of $\mu' = \mu * N(0, \tau^2)$ and ν differ by at most $C\delta$, where $\tau^2 = \sigma^2 - s^2$. Denote the distinct atoms of ν by $a_1 < \cdots < a_m$ and put $\Pi(x) = \prod_{i=1}^m (x - a_i)^2$. Its coefficients are uniformly bounded, so $\int \Pi \, d\mu' \le C\delta$. The top Hermite coefficient of the monic degree m polynomial $\prod_i (y + \tau G - a_i)$, as a polynomial in the standard normal G, is τ^m. Thus its expected square is at least $m!\tau^{2m}$ for any y, and $\tau^2 \le C\delta^{1/m}$.

Let T map to the nearest atom, with either deterministic choice at ties, and put $\nu' = T_\# \mu'$. As $\Pi(x) \ge |x - Tx|^{2m}$, this quantization has p-cost at most $C\delta^{p/(2m)}$ when $p \le 2m$. For $p > 2m$, on $[-R-1, R+1]$ the p-cost is at most a multiple of the $2m$-cost. Outside that interval the Gaussian displacement is at least one, so the tail contribution is bounded by $C\exp(-c/\tau^2) \le C\delta$ using the bound on τ; if $\tau = 0$ the tail is empty. Hence the quantization cost is at most $C\delta^{\min(p/(2m),1)}$.

For a gap $g = a_{i+1} - a_i$, let F be a smooth left step whose transition is confined to the middle half of the gap. Use a fixed scaled profile, so $\|F^{(j)}\|_\infty \le C_j g^{-j}$. Its polynomial interpolant H of degree at most $2m - 1$, matching the values and first derivatives at all the atoms, has ordinary coefficients bounded by $Cg^{-(2m-1)}$. Indeed Newton coefficients are repeated-node divided differences, bounded by the corresponding derivative suprema, and all nodes are bounded. The global Hermite interpolation remainder satisfies

$$|F(x) - H(x)| \le Cg^{-2m}\Pi(x)$$

by the derivative bound of order $2m$. Since $H = F$ on the support of ν, moment comparison yields $|\int F(d\mu' - d\nu)| \le C\delta g^{-2m}$; we used $g \le 2R$ to absorb the coefficient bound of order $2m - 1$. The quantity $F(x)$ can differ from $\mathbf{1}_{\{T(x) \le a_i\}}$ only in the middle half of the gap, where $\Pi(x) \ge (g/4)^{2m}$. Therefore

$$|\nu'((-\infty, a_i]) - \nu((-\infty, a_i])| \le C\min(1, \delta g^{-2m}).$$

In the monotone coupling of ν' and ν, the mass crossing each gap is exactly this cumulative discrepancy. The pth power of the total distance along any path of at most $m - 1$ gaps is bounded by m^{p-1} times the sum of their pth powers. Thus

$$W_p(\nu', \nu)^p \le C\sum_i g_i^p \min(1, \delta g_i^{-2m}) \le C\delta^{\min(p/(2m),1)}.$$

The last inequality splits at $g_i = \delta^{1/(2m)}$ and uses at most $m - 1$ bounded gaps. Finally $W_p(\mu, \mu') \le C_p\tau$. The triangle inequality and $m \le k$ prove the lemma. □

2.4 Rotated quadrature lower bounds

Let Q have the k-point Gauss–Hermite quadrature law for the standard normal measure. It matches Gaussian moments through order $2k-1$, has bounded support depending only on k, and $\mathbb{E}|Q| > 0$ since $k \geq 2$. This scalar construction appears in the univariate lower bound of Wu and Yang [5, Proposition 9]. Diakonikolas, Kane and Stewart [3, Proposition 4.2 and Corollary 4.4] use the same quadrature and the same common-covariance adjustment in an unknown direction for their statistical query result. Set $s = (\lambda_- + \lambda_+)/2 > 0$. For a unit vector u put

$$\mu_u = \mathcal{L}(hQu), \qquad \Sigma_u = sI - h^2 uu^\top, \qquad P_u = \mu_u * N(0, \Sigma_u).$$

For $h > 0$ sufficiently small these are admissible. The reference $P_* = N(0, sI)$ is itself in the model with $\mu_* = \delta_0$. In the direction u, after scaling by $s^{-1/2}$, the alternative is the law of $\rho Q + \sqrt{1-\rho^2}\, G$ with G standard normal and independent of Q, and $\rho = h/\sqrt{s}$. The orthogonal component is independent and has the same law as under P_*.

For completeness let f_ρ denote the scalar density relative to standard normal γ. For fixed $0 < \rho_0 < 1$, $f_{\rho_0} \in L^2(\gamma)$, because the component means are bounded and the component variances equal $1 - \rho_0^2 < 1$. The Hermite coefficient in degree j for f_ρ in the orthonormal basis $\mathrm{He}_j/\sqrt{j!}$ is $\rho^j \mathbb{E}\,\mathrm{He}_j(Q)/\sqrt{j!}$. This follows by taking expectations of the generating function $\exp(tx - t^2/2)$. The coefficients in degrees $1, \ldots, 2k-1$ are zero. Comparing the remaining squared coefficients to those at ρ_0 gives, by Parseval,

$$D(P_u \| P_*) \leq \chi^2(P_u, P_*) \leq (\rho/\rho_0)^{4k} \| f_{\rho_0} - 1 \|^2_{L^2(\gamma)} \lesssim h^{4k} \qquad (0 < \rho \leq \rho_0). \tag{4}$$

In sufficiently large dimension there is a collection of at least $\exp(cd)$ unit vectors with all pairwise inner products at most $1/2$ in absolute value. To see this, take vectors of independent uniform signs divided by $\sqrt{d}$. The probability that two independent vectors violate the bound is at most $2e^{-d/8}$ by Hoeffding. A union bound over pairs gives the collection. For distinct u, v in it, testing with the 1-Lipschitz function $x \mapsto \mathrm{dist}(x, \mathbb{R}v)$ and computing on $\mathrm{span}(u, v)$ give

$$W_1(\mu_u, \mu_v) \geq h\, \mathbb{E}|Q| \sqrt{1 - (u^\top v)^2}, \qquad \|\Sigma_u - \Sigma_v\|_{\mathrm{op}} = h^2 \sqrt{1 - (u^\top v)^2}.$$

With $h = c_0 r^{1/(2k)}$, the n-sample KL divergence to the common reference is at most $C c_0^{4k} d$ by (4). Choose c_0 small enough for admissibility and for Fano's inequality. The two separations give the lower bounds of orders h and h^2. In bounded dimensions compare one alternative with the reference, use $h = c_0 n^{-1/(4k)}$, and apply the two-point bound from Pinsker's inequality. Since d is bounded, this gives the same orders.

As $W_p \geq W_1$, the quadrature argument gives the desired lower bound for $p \leq 2k$. For $p > 2k$, fix $0 < t < R$ and use $\mu_u = (1-w)\delta_0 + w\delta_{tu}$ with $\Sigma = sI$. The density relative to P_* is $1 + w(L_u - 1)$, where L_u is the Gaussian shift likelihood ratio. Thus $\chi^2(P_u, P_*) = w^2(\exp(t^2/s) - 1)$. Every transport plan to μ_v must move the mass w at tu to the line $\mathbb{R}v$, giving $W_p(\mu_u, \mu_v) \geq t w^{1/p} \sqrt{1 - (u^\top v)^2}$. With $w = c_0 r$, the same projective packing and Fano argument apply. In bounded dimension use the two-point comparison with $\mu_* = \delta_0$.

2.5 Frobenius covariance loss

Write $C_\mu = \mathrm{Cov}(U)$. The fit from the upper bound satisfies

$$\|C_{\widehat{\mu}} - C_\mu\|_{\mathrm{op}} \lesssim \delta + \delta^{1/k}.$$

In fact, for each direction, $v^\top C_\mu v = m_2(v) - m_1(v)^2 - v^\top \Sigma v$; use (2), bounded means and (3). The difference has rank at most $2(k-1)$, so its Frobenius norm satisfies the same bound up to a k-dependent factor. Use independent paired differences $(X_{2i} - X_{2i-1})/\sqrt{2}$ to estimate $\mathrm{Cov}(X)$ by the average of their outer products. Each vector has fourth norm moment $O(d^2)$, since the means are bounded and the Gaussian covariance has operator norm at most λ_+. Independence across pairs gives expected Frobenius error $O(d/\sqrt{n})$. No independence between this estimate and the previous fit is needed. Subtract $C_{\widehat{\mu}}$ and project onto the admissible spectral interval in Frobenius norm. Projection onto this closed convex set cannot increase the loss. For $n \geq 2B$ this proves the desired upper bound. For $n < 2B$ the fixed midpoint covariance has error $O(\sqrt{d})$, which is within a constant of $d/\sqrt{n} \wedge \sqrt{d}$.

The rotated lower bound already gives $r^{1/k}$ since Frobenius norm dominates operator norm. The term $d/\sqrt{n} \wedge \sqrt{d}$ comes from the Gaussian submodel $\mu = \delta_0$. For all sufficiently large d, consider symmetric matrices A with zero diagonal and independent signs above it. On a $1/4$-net of at most 9^d points, each $v^\top A v = 2\sum_{i<j} v_i v_j A_{ij}$ has centered sub-Gaussian tails with uniformly bounded variance proxy. The operator norm is at most twice its maximum absolute quadratic form on the net. Consequently at least half the sign matrices have $\|A\|_{\mathrm{op}} \leq C\sqrt{d}$, for a fixed C. Greedy deletion of Hamming balls of radius cd^2 from the good sign matrices supplies $\exp(c'd^2)$ of them with pairwise $\|A - A'\|_F \geq c''d$. Indeed each such ball has at most $\exp[H(\alpha)\binom{d}{2}]$ elements for a fixed small fraction α, where $H(\alpha) = -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$.

With $a = c_0 \min(n^{-1/2}, d^{-1/2})$ use covariances $sI + aA$. They are admissible for small enough c_0. Their n-sample Gaussian divergences to the midpoint reference are bounded by $Cna^2\|A\|_F^2 \leq Cc_0^2 d^2$, by the quadratic bound on $t - \log(1+t)$ for $|t| \leq 1/2$. Fano's inequality gives expected loss of order at least ad. For bounded d, perturb one diagonal entry by $c_0 n^{-1/2}$ and use the two-point bound, which gives the same order up to constants. The maximum of the two lower bounds is equivalent to their sum. This proves Theorem 2.1.

3 Sharp calibration of unknown isotropic variance in Gaussian location mixtures

Consider a Gaussian location mixture with at most k bounded means in dimension $d \geq k$, and an unknown shared isotropic noise variance. The minimax absolute error for that variance is $(nd)^{-1/2}$. For every fixed finite $p \geq 1$, the minimax latent W_p rate is the same as with known variance. The latent rate coincides with the known-noise rate of Doss, Wu, Yang and Zhou; in dimension one the rates are instead those of Wu and Yang. For covariance matrices within operator distance ω of an unknown scalar matrix, the sharp operator loss is $(nd)^{-1/2} + \min\{\omega, (1 \wedge \sqrt{d/n})^{1/k}\}$. The variance estimator removes a low-rank signal subspace and averages independent residual energy. For three components in the plane the sharp variance rate is $n^{-1/4}$, by a degree-two polynomial bound and a pair of triangles with matching third moments.

3.1 Results

We observe n independent copies of $X = U + Z$, where $U \sim \mu$ is supported on at most k points in the Euclidean ball of radius $R > 0$ in $\mathbb{R}^d$, and $Z \sim N(0, \Sigma)$ is independent. Fix $k \geq 2$, $d \geq k$, and $0 \leq a < b < \infty$. Write $r = 1 \wedge \sqrt{d/n}$.

Theorem 3.1. *When $\Sigma = tI$ with unknown $t \in [a, b]$,*

$$\inf_{\widehat{t}} \sup_{\mu, t} \mathbb{E}|\widehat{t} - t| \asymp (nd)^{-1/2}.$$

For every fixed $1 \le p < \infty$,

$$\inf_{\widehat{\mu}} \sup_{\mu,t} \mathbb{E} W_p(\widehat{\mu}, \mu) \asymp r^{1/\max(2,p)} + n^{-1/[2\max(2k-1,p)]}.$$

The latent rate also holds for known t *at each fixed positive variance, with constants allowed to depend on* t. *The comparison constants are uniform in* d, n *and otherwise depend only on the fixed class parameters and* p.

Theorem 3.2. *For the larger class* $aI \preceq \Sigma \preceq bI$ *with* $\|\Sigma - tI\|_{\mathrm{op}} \le \omega$ *for some* $t \in [a, b]$,

$$\inf_{\widehat{\Sigma}} \sup_{\mu,\Sigma} \mathbb{E}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}} \asymp (nd)^{-1/2} + \min\{\omega, r^{1/k}\},$$

uniformly also in $\omega \ge 0$.

Doss, Wu, Yang and Zhou [6, Theorem 6.1] give isotropic upper bounds $((d/n)^{1/4} \wedge 1) + n^{-1/(4k)}$ for latent W_1 and $n^{-1/(2k)}$ for the variance, and state that both rates are minimax optimal. Their lower bounds embed the one-dimensional construction of Wu and Yang [5, Proposition 9], and these are the sharp rates in dimension one. In dimension $d \ge 2$ the orthogonal coordinates help to calibrate the noise, so that construction is not least favourable: Theorems 3.1 and 3.3 show that both rates improve for $d \ge k$, and that the variance rate improves for every fixed $d \ge 2$, since the integer m of Theorem 3.3 is at most $k - 1$ once $d \ge 2$. For $d \ge k$ the centered latent covariance has rank at most $k - 1$, so even a fixed k-dimensional coordinate projection gives a uniform root-n variance estimator. The population eigenvalue identity is due to Hsu and Kakade [10, Theorem 1]. Averaging trailing eigenvalues is classical in probabilistic principal component analysis [11, 12]. Our calibration uses an independent residual block and applies uniformly even when the signal covariance has vanishing eigenvalues. The rate in Theorem 3.1 also captures the gain from averaging across the d coordinates and recovers the known-noise W_1 rate of Doss et al. [6, Theorem 1.1]. We place no separation or minimum-weight assumptions on the component means. Theorem 3.2 quantifies the interpolation between this isotropic variance rate and the arbitrary shared covariance rate of Section 2. Section 3.5 gives a polynomial upper bound below the rank threshold and shows it is sharp for three planar components.

3.2 Independent residual energy

For the class in Theorem 3.2, fix any admissible representative t and put $A = \Sigma - tI$. Let $C = \mathrm{Cov}(U)$. Then $\operatorname{rank} C \le k - 1$, because the centered support of U spans a space of at most that dimension. We construct a scalar estimate $\widetilde{t}$ such that

$$\mathbb{E}|\widetilde{t} - t| \lesssim \omega + (nd)^{-1/2}. \tag{5}$$

For unbiased estimation of the centered observation covariance $C + \Sigma$ use disjoint paired differences $Y = (X - X')/\sqrt{2}$. Their latent parts have norm at most $\sqrt{2}R$, and their Gaussian part has covariance Σ. The same identity holds when Σ is singular. With m such pairs, the empirical second moment M obeys

$$\mathbb{E}\|M - (C + \Sigma)\|_{\mathrm{op}} \lesssim \sqrt{d/m} + d/m. \tag{6}$$

Indeed a fixed-direction quadratic observation is uniformly subexponential after centering. Apply Bernstein's inequality to a volumetric 1/4-net with at most 9^d elements. The operator norm is bounded by twice the largest absolute net quadratic form. Integration of the resulting tails proves (6).

If $k \le d < 2k$, the smallest eigenvalue of $C + tI$ is t. Use the smallest eigenvalue of M, clipped to $[a, b]$, and apply Weyl's eigenvalue inequality and (6). The dimension is bounded by $2k$, so this gives (5). If only one draw is available in this bounded dimension, a constant estimate suffices.

If $d \ge 2k$ and $d \ge n$, use the raw average of $\|X_i\|^2/d$, clipped to $[a, b]$. Its mean differs from t by at most $\omega + R^2/d$; its variance is at most $C/(nd)$, by conditioning on the bounded latent part and using the first two Gaussian quadratic moments. Since $1/d \le 1/\sqrt{nd}$, this gives (5).

For $2k \le d < n$, use independent blocks with numbers of pairs comparable to n. Let $\widehat{\Pi}$ project onto the top $k-1$ directions of M from the first block, and set $Q = I - \widehat{\Pi}$, of rank $\ell = d - k + 1$. Writing $\zeta = \|M - (C + tI)\|_{\mathrm{op}}$, Weyl's inequality gives $\lambda_k(M) \le t + \zeta$, since $\lambda_k(C) = 0$. Thus $v^\top C v \le 2\zeta$ for each unit v in $\operatorname{im} Q$, and $\operatorname{tr}(QCQ) \le 2(k-1)\zeta$ by its rank. On the held-out block average $\|QY_i\|^2/\ell$ and clip to $[a, b]$. Given the first block, the mean differs from t by at most $\omega + 2(k-1)\zeta/\ell$. The conditional variance of $\|QY_i\|^2$ is $O(\ell)$: given the latent difference B, the trace term from the Gaussian quadratic part is $2\operatorname{tr}[(Q\Sigma Q)^2] \le 2b^2\ell$, the mixed term is bounded by $4b\|B\|^2$, and the variance of $\|QB\|^2$ is bounded by a constant. Independence on this block gives fluctuation $O((n\ell)^{-1/2})$ for the estimate. Finally $\mathbb{E}\zeta \lesssim \omega + \sqrt{d/n}$ by (6) and $d < n$. This proves (5), since $\ell \asymp d$.

The isotropic lower bound follows by restricting to $\mu = \delta_0$ and comparing two positive scalar variances separated by $c/\sqrt{nd}$. The n-sample divergence of the Gaussian laws is $O(nd)$ times the squared variance gap. Taking c small and applying the two-point bound proves the variance part of Theorem 3.1.

For Theorem 3.2, the operator risk of $\widetilde{t}I$ is $O(\omega + (nd)^{-1/2})$. Alternatively apply the unrestricted fit of Section 2, with risk $O(r^{1/k})$. Use the first estimate when $\omega \le r^{1/k}$ and the second otherwise. The scalar submodel gives the first lower term. For the second, take the reference and quadrature family of Section 2.4 with covariance $t_0 I - h^2 uu^\top$, where $t_0 = (a+b)/2$. Set $h = c\min(\sqrt{\omega}, r^{1/(2k)})$. The candidates lie in the width class, and the original projective packing (or the two-point comparison in bounded dimension) gives operator risk at least $c'h^2$. The maximum of the two lower bounds is equivalent to their sum.

3.3 Fitting the mixing law

We first record the comparison for two scalar discrete laws ν_1, ν_2 with at most k atoms in $[-R, R]$. If their raw moments through degree $2k-1$ differ by at most δ, then

$$W_p(\nu_1, \nu_2) \lesssim \delta^{1/\max(2k-1,p)}. \tag{7}$$

Order the $m \le 2k$ nodes in their combined support. At each gap g, take a smooth left step of transition width comparable to g, lying inside the gap. Its interpolating polynomial on the m nodes has degree at most $m-1$ and coefficients $O(g^{-(m-1)})$. To see the latter, use the Newton form, bound each divided difference by the corresponding derivative of the scaled step, and use bounded nodes. The two measures agree with their integrals against the step on those nodes, so the cumulative discrepancy across the gap is bounded by $C\min(1, \delta g^{-(m-1)})$. In the monotone coupling, summing the pth powers of the gap lengths weighted by their crossing masses, times m^{p-1}, bounds the p-cost. Splitting at $g = \delta^{1/(m-1)}$ proves (7). If $m = 1$, the laws coincide. This is the gap form of the atomic moment comparison used for W_1 by Wu and Yang [5].

We now work in the isotropic model. If $d \ge n$, a fixed mixing estimate suffices since $r = 1$. Otherwise split the sample into three groups of sizes comparable to n; a fixed estimate handles bounded n. In the first group estimate the raw second moment $\mathbb{E}[XX^\top] = tI + \mathbb{E}[UU^\top]$ by its empirical second moment. Its expected operator error is $O(\sqrt{d/n})$ by the same net proof as (6). Let $\widehat{P}$ project onto its top k eigenvectors. The latent raw second moment $S = \mathbb{E}(UU^\top)$ has rank at most

k. Put $\zeta' = \|\widehat{M} - (tI + S)\|_{\mathrm{op}}$, where $\widehat{M}$ is the empirical second moment. Weyl's inequality gives $\lambda_{k+1}(\widehat{M}) \le t+\zeta'$. For a unit vector v in $\mathrm{im}(I-\widehat{P})$, $v^\top S v \le v^\top \widehat{M} v - t + \zeta' \le \lambda_{k+1}(\widehat{M}) - t + \zeta' \le 2\zeta'$. Since $(I-\widehat{P})S(I-\widehat{P})$ is positive semidefinite of rank at most k, $\mathrm{tr}[(I-\widehat{P})S] \le 2k\zeta'$, and therefore $\mathbb{E}\,\mathrm{tr}[(I-\widehat{P})S] \lesssim \sqrt{d/n}$. When $k = d$ the trace is zero. The coupling $U \mapsto \widehat{P}U$ therefore costs

$$\mathbb{E}W_p(\mu, \widehat{P}_{\#}\mu) \lesssim r^{1/\max(2,p)}. \tag{8}$$

For $p \le 2$, use the quadratic cost and Jensen's inequality. For $p > 2$, use the bound R^{p-2} on the ratio of the pth power to the square of the discarded distance.

Use the second group to estimate t by (5) at $\omega = 0$, with expected error $O(n^{-1/2})$ or better, and keep the estimate in $[a, b]$. Project the third group into $\mathrm{im}\,\widehat{P}$. Conditional on the first two groups its dimension is at most k, and the observation law is Gaussian convolution of $\widehat{P}_{\#}\mu$ with scalar variance t. Estimate all raw directional moments through degree $2k-1$ by sample averages. The expected supremum moment error over unit vectors is $O(n^{-1/2})$, conditionally and uniformly: expand the finitely many monomials in an orthonormal frame and use their uniformly bounded variances. Undo the estimated variance in the moment equations, using the coefficients of $\exp(zx - \widetilde{t}z^2/2)$. For each fixed degree those coefficients are polynomials in $\widetilde{t}$, with uniformly bounded derivatives on $[a, b]$ after taking expectations against the clean bounded moments. The empirical error remains bounded by a constant times the original moment error. Consequently the estimated latent directional moments have expected maximum error $O(n^{-1/2})$.

Fit a law with at most k atoms in the projected ball by minimizing this maximum directional discrepancy to accuracy $n^{-1/2}$. A countable dense approximation in an orthonormal frame gives a measurable fit. Feasibility of $\widehat{P}_{\#}\mu$ doubles the moment discrepancy up to the tolerance. Apply (7) in every direction and lift a projected coupling along a direction preserving all distances in the combined support up to a fixed multiplicative constant. Such a direction exists by excluding finitely many angular bands in the span of the support differences, as in the proof of [6, Lemma 3.1], property (B.1) in their Appendix B. Jensen's inequality gives expected fit error $O(n^{-1/[2\max(2k-1,p)]})$. Together with (8) this proves the upper rate. The same procedure can use the actual t when it is known.

3.4 Mixing-law lower bounds

All lower bounds in this section take a fixed positive variance t_0 in the allowed interval. For $p \le 2$, the high-dimensional bound already follows from the W_1 lower bound of Doss et al. [6, Theorem 1.1]. Equivalently, use $\mu_u = (\delta_{hu} + \delta_{-hu})/2$ over unit directions with pairwise inner products of modulus at most $1/2$. There are $\exp(cd)$ such directions for large d, by a random sign packing. The separation in W_1 is at least ch, by testing with distance to the other supporting line. The chi-square divergence from $N(0, t_0 I)$ is $\cosh(h^2/t_0) - 1 = O(h^4)$. With $h = cr^{1/2}$, Fano's inequality gives the result. In bounded dimension a two-point comparison with a single centered Gaussian gives the same order. For $p > 2$, use $\mu_u = (1-w)\delta_0 + w\delta_{\tau u}$, with fixed $0 < \tau < R$. The same packing has W_p separation at least $cw^{1/p}$, while the chi-square divergence to the reference Gaussian is $w^2(e^{\tau^2/t_0} - 1)$. Take $w = cr$, and use Fano or the bounded-dimension two-point comparison.

For the sample-size term put $m = 2k-1$. On the nodes jhe_1, $0 \le j \le m$, assign the even and odd indices to two respective measures, with masses $2^{1-m}\binom{m}{j}$. Each law has k atoms. Their moments in the first coordinate agree through degree $m-1$ by the finite-difference identity. The W_p separation is at least h, since the two supports are separated by this distance. The even law's convolution dominates a fixed positive multiple of $N(0, t_0 I)$ because it has positive mass at zero. Expanding the two convolved density ratios in the scalar Gaussian Hermite basis, their difference

has no coefficients below degree m. The squared L^2 norm of the difference is $O(h^{2m})$: its jth orthonormal coefficient is $(h/\sqrt{t_0})^j$ times the difference of the jth moments of the fixed integer-node laws divided by $\sqrt{j!}$, and the squared series is finite at any fixed bounded h. By domination this bounds the chi-square divergence between the two convolved laws. Choose $h = cn^{-1/(2m)}$ and apply the two-point inequality. For $p \le m$ this is the sample-size term in Theorem 3.1. For $p > m$ the preceding high-dimensional rare-mass lower bound already dominates $n^{-1/(2p)}$.

3.5 Calibrating below the rank threshold

Theorem 3.3. *For fixed $d, k \ge 1$, let m be the smallest integer with $\binom{d+m}{m} > k$. The isotropic variance risk is at most $Cn^{-1/(2m)}$. The bound is sharp if there exists a standard Gaussian cubature law with at most k atoms whose moments match through degree $2m-1$. It is also sharp with $m = 2$ whenever $d \ge 2$ and $\lceil 3d/2 \rceil \le k < \binom{d+2}{2}$.*

Estimate all raw multivariate moments through order $2m$ and jointly fit the isotropic variance and atomic law over the compact parameter space. As in the projected fit above, the maximum population moment discrepancy δ has expectation $O(n^{-1/2})$. Compare any two admissible pairs with that discrepancy. Relabel so that s is the smaller variance and put $\tau^2 = t - s$. Removing the common Gaussian variance s from their moment equations leaves discrepancy at most $C\delta$ through degree $2m$ between the smaller-variance latent law ν and $\mu * N(0, \tau^2 I)$. There is a nonzero polynomial P of degree at most m vanishing on the at most k atoms of ν, by the dimension of the polynomial space. Normalize its coefficient vector to unit Euclidean length. Since $\int P^2 \, d\nu = 0$, $\mathbb{E}P(U + \tau G)^2 \le C\delta$. For each $\|y\| \le R$, the coefficients of $P(y + \cdot)$ have norm bounded below by a constant, by invertibility of translation with uniformly bounded inverse. Scaling the argument by τ multiplies each coefficient of degree j by τ^j. Equivalence of the coefficient norm and the standard Gaussian L^2 norm on this finite-dimensional polynomial space shows

$$\mathbb{E}P(y + \tau G)^2 \ge c \min(1, \tau)^{2m}.$$

Since τ is bounded, this gives $t - s \le C\delta^{1/m}$. Jensen's inequality proves the first claim.

For the cubature lower bound, compare a point mass with variance $t_0 I$ to the cubature law scaled by h with variance $(t_0 - h^2)I$, where $t_0 = (a + b)/2$. Relative to the reference Gaussian, the alternative has no nonconstant Hermite coefficients of total degree below $2m$. Indeed the multivariate Gaussian generating function gives the coefficient for multi-index α as the cubature Hermite moment multiplied by $(h/\sqrt{t_0})^{|\alpha|}$ and divided by $\sqrt{\alpha!}$. For some fixed $0 < h_0 < \sqrt{t_0}$ the density ratio is in Gaussian L^2, so comparison of the squared series gives chi-square divergence $O(h^{4m})$ when $h \le h_0$. Take $h = cn^{-1/(4m)}$ to separate the variances by order $n^{-1/(2m)}$. For example, the uniform law on $\{\pm\sqrt{d}\, e_i : 1 \le i \le d\}$ matches the Gaussian through degree three. Thus for $2d \le k < \binom{d+2}{2}$ the bound is $n^{-1/4}$.

For the planar lower bound construct a centered three-point law at $(u, v), (u, -v), (-wu/(1 - w), 0)$, with weights $w/2, w/2, 1 - w$. For any positive x, y, T, set

$$u = T/y, \qquad w = x/(x + u^2), \qquad v^2 = y/w.$$

Then its covariance is $\mathrm{diag}(x, y)$. Its third tensor has $T_{122} = T$ and $T_{111} = T(x/y - x^2 y/T^2)$, along with the permutations of T_{122}, and the other entries vanish by reflection in the second coordinate. Indeed the first-coordinate marginal has variance x and third moment $x(u - x/u)$, while the mixed moment is $wuv^2 = uy$. Choose $T = \sqrt{60}$ and $(x, y) = (1, 2)$ and $(2, 3)$. The third tensors agree, because $1/2 - 2/60 = 2/3 - 12/60$, while the covariances differ by I.

Scale both latent laws by h. Give the first law Gaussian variance $t_0 = (a+b)/2$ and the second variance $t_0 - h^2$. For small h these are admissible. Their observed moments agree through degree three. Their density ratios relative to $N(0, t_0 I_2)$ have difference $O(h^4)$ in Gaussian L^2. To see the cancellation, divide their moment generating functions by that of $N(0, t_0 I_2)$. The resulting functions are $\mathbb{E}e^{hz^\top U_1}$ and $e^{-h^2\|z\|^2/2}\mathbb{E}e^{hz^\top U_2}$ for the unscaled latent variables. Their Taylor coefficients through h^3 agree by the mean, covariance and third tensor identities. For each density ratio the h-derivatives through order four exist continuously in Gaussian L^2 near zero: derivatives of the Gaussian kernels have polynomial factors and their variances stay strictly below twice t_0. Differentiating the generating functions identifies the density derivatives through their integrals against exponentials, so the first four Taylor coefficients of the density difference vanish. Taylor's bound gives $O(h^4)$. The first observed density dominates $e^{-h^2\mathbb{E}\|U_1\|^2/(2t_0)}$ times the reference Gaussian, by Jensen's inequality and $\mathbb{E}U_1 = 0$. Thus the chi-square divergence is $O(h^8)$. Taking $h = cn^{-1/8}$ gives variance separation of order $n^{-1/4}$ by the two-point inequality.

For the general $m = 2$ range, split the coordinates into orthogonal planes and, if needed, a final line. On each plane use the two planar laws just constructed. On the line use δ_0 and the symmetric law on $\{-1, 1\}$. Each pair has mean zero, identical cubic moments and covariance difference equal to the identity on its block. Mix the respective block laws with the same positive block weights. Before mixing, scale the points on each block by the inverse square root of its weight so that the overall covariance difference is I. Cubic moment matching is preserved, since each point lies in a single block. The number of points in either resulting law is at most $\lceil 3d/2 \rceil$. For fixed d these are again bounded configurations, so the same density Taylor argument gives the claimed lower bound.

4 Sharp robust detection of sparse Gaussian location mixtures

We test one Gaussian against a finite location mixture with unknown shared covariance and arbitrary Huber contamination. The component means may differ on a common unknown set of at most s coordinates. For at most $k \geq 2$ components and latent separation from one point in W_p, the minimax threshold is of order $\left(1 \wedge [\sqrt{s\log(ed/s)/n} + \epsilon]\right)^{1/\max(2k,p)}$. Gaussian smoothing and a factored differential equation turn projected distribution-function control into linear moment control, including when the original covariance is singular. A sparse direction prior gives the matching testing lower bound. The same moment comparison gives sharp robust estimation rates for dense finite mixtures.

4.1 Setting and results

Fix $k \geq 2$, $R > 0$, and $0 \leq \lambda_- < \lambda_+ < \infty$. For a mixing law μ supported on at most k points in the Euclidean ball of radius R, let $P_{\mu,\Sigma} = \mu * N(0, \Sigma)$ with $\lambda_- I \preceq \Sigma \preceq \lambda_+ I$. For $1 \leq s \leq d$ we restrict the support differences of μ to a common coordinate set of size at most s; the common offset is arbitrary subject to the ball constraint. The observations have law $(1-\epsilon)P_{\mu,\Sigma} + \epsilon B$ independently, with arbitrary probability law B and $0 \leq \epsilon < 1/2$. The null consists of all point-mass mixing laws in the ball. For fixed $1 \leq p < \infty$, the alternatives have

$$\Delta_p(\mu) := \inf_{a \in \mathbb{R}^d} W_p(\mu, \delta_a) \geq \eta.$$

Write $A = s\log(ed/s)$, $q = \max(2k, p)$, and $r_s = 1 \wedge (\sqrt{A/n} + \epsilon)$.

Theorem 4.1. *There are constants $0 < c < C$ depending only on the fixed parameters such that the sum of the worst type I and type II errors can be made at most $1/3$ at separation $\eta \geq C r_s^{1/q}$. At*

separation $\eta \le c r_s^{1/q}$ every test has sum of errors at least $2/3$. The upper statement is interpreted as immediate when the alternative class is empty.

Theorem 4.2. *For the dense model $s = d$, the robust minimax expected errors for latent W_p and covariance operator norm are, respectively, $r_d^{1/q}$ and $r_d^{1/k}$, up to the same types of constants. A single proper fit attains both. It need not know ϵ.*

For two equally weighted components without contamination, Δ_p is half the separation of the means, and the testing result recovers the sparse unknown-covariance order of Verzelen and Arias-Castro [4, Section 3.1] in the bounded spectral model. We allow more than two components, arbitrary finite p, and contaminated data, with the threshold taken uniformly over the weights. The least favourable law is balanced, so for weights bounded away from $1/2$ the boundary is faster [4, Section 3.2]. General cumulant testing studies related directional moment signals [13]. For the latent loss the maximum of the collision exponent $2k$ and the loss exponent p determines the least favorable family. The collision family is the directional quadrature construction of Diakonikolas, Kane and Stewart [3, Section 4], used at vanishing amplitude as in the estimation theorem of Section 2.

4.2 A linear moment bound from distribution functions

Bonaccorsi, Hanzon and Lombardi [14] use differential equations to study zeros of signed Gaussian mixtures. We use a monic product of first-order factors with coefficients that stay in a compact family even when Gaussian parameters coalesce. This gives a uniform weighted comparison between densities and distribution functions.

Lemma 4.3. *Let F_1, F_2 be scalar Gaussian mixture distribution functions, each with at most k components, all means in $[-R, R]$ and all variances in the same interval $[a, b]$ with $0 < a \le b$. For any fixed integer $j \ge 0$,*

$$\int_{\mathbb{R}} |x|^j \, |F_1'(x) - F_2'(x)| \, dx \le C_{k,R,a,b,j} \|F_1 - F_2\|_\infty.$$

Proof. Put $g = F_1' - F_2'$ and $F = F_1 - F_2$. Write $g = \sum_{i=1}^{m} c_i e^{q_i(x)}$ with $m = 2k$, allowing zero terms. The q_i are quadratics $q_i(x) = -\alpha_i x^2 + \beta_i x$ with α_i bounded above and away from zero and β_i bounded. Any constant term is absorbed in c_i. For $i = 1, \dots, m$ apply $D - q_i'$ exactly 2^{i-1} times, in increasing order of i, where D is differentiation. When the ith block starts, the polynomial multiplier of e^{q_i} has degree at most $2^{i-1} - 1$, since each preceding factor increases polynomial degree by at most one. The ith block annihilates that term, and zero terms stay zero. Consequently g solves a monic differential equation of order $N = 2^m - 1$ with coefficients depending continuously on the quadratics. This remains true if several parameters coincide. The coefficients of the monic equation are polynomials in x whose coefficients are polynomials in the (α_i, β_i), hence depend continuously on them. The order $N = 2^{2k} - 1$ is large, but only its finiteness is used. A zero term keeps given parameters (α_i, β_i) in the same compact box, so the compactness argument below runs over a single compact parameter set.

Over this compact coefficient family, the sup norm of a primitive on $[-1, 1]$ bounds its initial value and the initial jet of its derivative:

$$|F(0)| + \sum_{i=0}^{N-1} |g^{(i)}(0)| \le C \sup_{|x| \le 1} |F(x)|.$$

Indeed, solutions of the monic equation depend continuously on its coefficients and on initial data on this interval. If the bound failed, a convergent subsequence of coefficient parameters and normalized initial vectors would give an identically zero primitive with nonzero initial vector, a contradiction.

List the applied factors as $D - q'_{\iota_l}$, $1 \le l \le N$, and let g_l be the result of the first l factors applied to $g = g_0$. We have $g_N = 0$, and each $g_l(0)$ for $l < N$ is bounded by the initial jet. Reversing the factors uses

$$g_{l-1}(x) = e^{q_{\iota_l}(x)} \left[g_{l-1}(0) + \int_0^x e^{-q_{\iota_l}(y)} g_l(y)\, dy \right].$$

On either ray, for y between 0 and x, $q_i(x) - q_i(y) \le -\alpha_0(x^2 - y^2) + \beta_0(|x| - |y|)$ for common constants $\alpha_0 > 0, \beta_0 \ge 0$. Backward induction therefore bounds $|g_0(x)|$ by the initial jet norm times a fixed polynomial in $1 + |x|$ times $\exp(-\alpha_0 x^2 + \beta_0 |x|)$. Taking the weighted integral gives the result. □

4.3 Upper bounds

For any law P, let $F^+_{P,v}$ be the distribution function of its scalar projection along v after convolution with a standard scalar Gaussian. Write

$$d_s^+(P, Q) = \sup_{\|v\|=1,\ \|v\|_0 \le s} \sup_t |F^+_{P,v}(t) - F^+_{Q,v}(t)|.$$

This is at most the corresponding unsmoothed discrepancy since convolution averages translates of the scalar distribution functions. For an empirical law P_n from a population law $\widetilde{P}$, for each fixed coordinate set S of size s the unsmoothed discrepancy in directions on S has mean at most $C\sqrt{s/n}$. This is the distribution-free VC empirical-process bound for half-spaces in $\mathbb{R}^s$. Changing one observation changes this discrepancy by at most $1/n$. Bounded differences and a union bound over $\binom{d}{s} \le (ed/s)^s$ sets yield, for $t \ge 0$,

$$\mathbb{P}_{\widetilde{P}} \left\{ d_s^+(P_n, \widetilde{P}) > C\sqrt{A/n} + t/\sqrt{n} \right\} \le e^{-ct^2}.$$

In particular the mean is at most $C\sqrt{A/n}$, and for a clean law P underlying the contamination $d_s^+(P, \widetilde{P}) \le \epsilon$. All these empirical bounds require no moments of B.

Now take any two candidate laws $P_{\mu,\Sigma}, P_{\nu,\Theta}$. On each probed direction both smoothed scalar mixtures have component variances between $1 + \lambda_-$ and $1 + \lambda_+$. Lemma 4.3 bounds their first $2k$ moment gaps by $C d_s^+(P_{\mu,\Sigma}, P_{\nu,\Theta})$. The scalar transport and variance comparison in Lemma 2.2 therefore implies

$$W_p(\mu_v, \nu_v) \le C d_s^+(P_{\mu,\Sigma}, P_{\nu,\Theta})^{1/q}, \qquad |v^\top(\Sigma - \Theta)v| \le C d_s^+(P_{\mu,\Sigma}, P_{\nu,\Theta})^{1/k}. \tag{9}$$

The common added variance cancels in the second inequality.

For the testing upper bound, fix an alternative μ. Among directions on its common coordinate set there is one retaining a fixed fraction of every pairwise distance of its at most k atoms. To see this, choose a uniform direction in the span of their differences, whose dimension is at most $k - 1$, and exclude a thin band for each difference. For dimension at least two each excluded band has probability $O_k(c)$ at relative angular width c, and for dimension one the unique line suffices. If U, U' are independent with law μ, this direction satisfies

$$\Delta_p(\mu) \le (\mathbb{E}\|U - U'\|^p)^{1/p} \le C(\mathbb{E}|v^\top(U - U')|^p)^{1/p} \le 2C\Delta_p(\mu_v).$$

The first inequality averages the p-cost of choosing the center U'; the last holds by the triangle inequality in L^p for any scalar center. Hence (9), applied to any null law Q, shows $d_s^+(P_{\mu,\Sigma}, Q) \geq c\Delta_p(\mu)^q$. Test with $\inf_{Q\in H_0} d_s^+(P_n, Q)$, where H_0 is the class of clean null laws. Under either hypothesis the empirical discrepancy from its clean law is $O(\sqrt{A/n} + \epsilon)$ with probability at least $5/6$. This proves the stated errors using a threshold at a sufficiently large constant times that quantity. When the quantity is large, the upper claim follows by choosing C larger than the maximal latent separation divided by the corresponding fixed lower bound on $r_s^{1/q}$. The infimum is measurable, since after smoothing one can use dense countable sets of parameters, directions and thresholds. Continuity in parameters holds uniformly in directions and scalar thresholds by the fixed positive smoothed variance floor and bounded means.

For estimation, use $s = d$ and minimize the same distance over all clean admissible k-component laws, to accuracy $n^{-1/2}$, again by a countable dense parameter set. The true parameter is feasible. The fitted population distance from the true law is at most $2d_d^+(P_n, P_{\mu,\Sigma}) + n^{-1/2}$. The quadratic-form supremum of (9) bounds the operator loss. For the transport loss, a single direction retains a fixed fraction of all pairwise distances in the combined support of the fitted and true laws by the same band argument with dimension at most $2k - 1$. Its injective projection allows a projected coupling to be lifted. Use (9), the empirical mean bound, concavity of the two powers and bounded parameter losses to conclude the upper bounds of Theorem 4.2.

4.4 Sparse testing lower bound

Let u have a uniform s-subset as support and independent uniform signs divided by $\sqrt{s}$ on that support. Let v be an independent copy and put $T = u^\top v$. For $L = \log(ed/s)$, the overlap size J gives

$$\mathbb{E}e^{zsT} = \mathbb{E}(\cosh z)^J \leq [1 + (s/d)(\cosh z - 1)]^s.$$

For example, expand in powers of $\cosh z - 1$ and use $\mathbb{E}\binom{J}{j} = \binom{s}{j}(s)_j/(d)_j \leq \binom{s}{j}(s/d)^j$ with falling factorials. If $|z| \leq L/2$, then $(s/d)(\cosh z - 1) \leq Cz^2/L$, since $\cosh z - 1 \leq z^2 e^{|z|}/2$ and $Le^{-L/2}$ is bounded. Taking $z = cLx$ for small universal c gives

$$\mathbb{P}(|T| \geq x) \leq 2e^{-c'Ax^2}, \qquad 0 \leq x \leq 1.$$

Tail integration implies

$$\mathbb{E}e^{\alpha AT^2} - 1 \leq 2\alpha/(c' - \alpha) \qquad (0 < \alpha < c'). \tag{10}$$

Fix $b_0 = (\lambda_- + \lambda_+)/2$ and the clean null $P_* = N(0, b_0 I)$. Consider alternatives differing from P_* only along u, with scalar density ratio relative to the standard Gaussian after division by $\sqrt{b_0}$. Write this ratio as $1 + \sum_{j\geq 2} a_j \operatorname{He}_j/\sqrt{j!}$ in Gaussian L^2. The coefficients will have no degree-one term. If $L_u = dP_u/dP_*$, Gaussian Hermite orthogonality in correlated directions gives

$$\mathbb{E}_{P_*}(L_u L_v) = 1 + \sum_{j\geq 2} a_j^2 (u^\top v)^j.$$

This identity follows by expanding the joint Gaussian exponential generating function and taking L^2 limits. For the prior which uses one common u for all n observations, product independence and $1 + x \leq e^x$ give

$$1 + \chi^2(\mathbb{E}_u P_u^{\otimes n}, P_*^{\otimes n}) \leq \mathbb{E}\exp\left(nT^2 \sum_{j\geq 2} a_j^2\right). \tag{11}$$

Thus the testing error sum is at least 2/3 when $n\sum_j a_j^2 \le c_0 A$ for a sufficiently small universal c_0, by (10) and the total-variation bound from chi-square divergence.

For $p \le 2k$ take $\mu_u = \mathcal{L}(hQu)$ with Q a k-point standard normal quadrature through degree $2k-1$, and $\Sigma_u = b_0 I - h^2 uu^\top$. These candidates are admissible for small fixed h, and $\Delta_p(\mu_u) \ge ch$ since Q is nondegenerate. The standardized scalar law along u is $\rho Q + \sqrt{1-\rho^2}G$, where $\rho = h/\sqrt{b_0}$ and G is independent standard Gaussian. Its Hermite coefficients $a_j = \rho^j \mathbb{E}\,\mathrm{He}_j(Q)/\sqrt{j!}$ vanish through degree $2k-1$. At any fixed $0 < \rho_0 < 1$ its relative density is in Gaussian L^2; comparing the coefficients yields $\sum_j a_j^2 \le Ch^{4k}$ for $\rho \le \rho_0$. Take $h = c(1 \wedge \sqrt{A/n})^{1/(2k)}$.

For $p > 2k$, fix $0 < t < R$ and let $\mu_u = \mathcal{L}(t(B_w - w)u)$, where B_w is Bernoulli with success probability $w \le 1/2$. Use $\Sigma_u = b_0 I - t^2 w(1-w)uu^\top$. The projected observed first and second moments agree with the null, so degrees one and two vanish. For small w, the scalar density ratio differs from one by $O(w)$ in Gaussian L^2. Indeed the dominant Gaussian component changes its mean and variance by $O(w)$ from those of the reference. In a small neighborhood of those parameters the density ratio and its first parameter derivatives have uniformly bounded Gaussian L^2 norms. This follows by integrating a polynomial times a Gaussian exponential, with the component variance bounded strictly below twice the reference variance. The rare component has bounded mean and the same variance and contributes $O(w)$ by its weight. Thus $\sum_j a_j^2 \le Cw^2$. For any center, at least one of the two atom distances is at least $t/2$, and the weight of either is at least w. So $\Delta_p(\mu_u) \ge (t/2)w^{1/p}$. Take $w = c(1 \wedge \sqrt{A/n})$. These two constructions with (11) prove the finite-sample clean lower bound.

4.5 The contamination lower bounds

Each clean law is in its own Huber neighborhood by taking the contaminant equal to the clean law. Thus clean minimax lower bounds also hold under contamination. For the additional term we use the following exact overlap property of Huber neighborhoods [1]. If $d_{\mathrm{TV}}(P,Q) \le \epsilon/(1-\epsilon)$, there is one law in both neighborhoods: take a measure dominating $(1-\epsilon)P$ and $(1-\epsilon)Q$ by using their pointwise maximum with respect to $P+Q$. Its mass is $(1-\epsilon)(1+d_{\mathrm{TV}}(P,Q)) \le 1$. Add any remaining mass. Its differences from the two dominated measures, divided by ϵ, are probability laws when $\epsilon > 0$.

For testing at a fixed one-sparse direction, quadrature gives $d_{\mathrm{TV}}(P_u, P_*) \le Ch^{2k}$ and separation at least ch. Choose $h = c_0\epsilon^{1/(2k)}$ when $p \le 2k$. For $p > 2k$ the compensated Bernoulli gives $d_{\mathrm{TV}}(P_u, P_*) \le Cw$ and separation at least $cw^{1/p}$; choose $w = c_0\epsilon$. Both choices yield identical contaminated observation laws under one null and one separated alternative. Together with the clean testing bound, this gives the claimed lower order, since the maximum of the clean and contamination rates is equivalent to $r_s^{1/q}$.

For dense estimation the clean lower bounds are those of Section 2. The same contaminated pairs give the latent lower bound $\epsilon^{1/q}$, since the alternative is separated in W_p from the point-mass reference. For covariance loss use the quadrature pair, whose matrices are separated in operator norm by h^2 with $h = c_0\epsilon^{1/(2k)}$. Identical observation laws and the triangle inequality give the lower order $\epsilon^{1/k}$. Combining the two contributions completes Theorems 4.1 and 4.2.

5 Testing Gaussian mixtures with a partially known covariance

We determine the minimax separation for testing one Gaussian against a sparse finite location mixture when the common covariance lies in a Loewner interval of width ω below a known positive-definite matrix. The result is uniform in the width, dimension, sample size and Huber contamination. For transport powers between certain moment orders there are intermediate

rates between those for known and unrestricted covariance. They arise from small masses at intermediate distances whose moments are compensated by an inner quadrature. A polynomial constrained by the number of mixture components gives the last transition at transport power $2k-1$.

5.1 Model and separation

Fix an integer $k \geq 2$, radius $R > 0$, and constants $0 < b_- < b_+ < \infty$. Let B be a known d-dimensional covariance with $b_- I \preceq B \preceq b_+ I$. The unknown covariance satisfies

$$B - \omega I \preceq \Sigma \preceq B, \qquad 0 \leq \omega \leq \omega_0 < \min(1, b_-/2).$$

The uncontaminated law is $P = \mu * N(0, \Sigma)$, with μ supported on at most k points in the radius-R ball. Their differences are supported on a common unknown coordinate set of size at most s, with $1 \leq s \leq d$. The observed independent draws have law $(1-\epsilon)P + \epsilon H$, where H is arbitrary and $0 \leq \epsilon < 1/2$. Under the null μ is a point mass. Separate the alternative by $\Delta_p(\mu) = \inf_a W_p(\mu, \delta_a)$, for fixed $p \geq 1$. Set

$$A = s \log(ed/s), \qquad e_n = 1 \wedge (\sqrt{A/n} + \epsilon), \qquad q = \max(2k, p), \quad q_0 = \max(2, p).$$

Let the knots be $2, 4, \ldots, 2k-2, 2k-1$. When p lies strictly between consecutive knots $j < J$, put

$$M_p(\omega, e) = \left(\omega^{j(J-p)/(2(J-j))} e^{(p-j)/(J-j)}\right)^{1/p}.$$

Put $M_p = 0$ if p is a knot or outside the knot range.

Theorem 5.1. *The minimax separation, up to constants depending only on the fixed class parameters and p, is*

$$\mathcal{R}_p(\omega, e_n) = \min\{e_n^{1/q},\ \sqrt{\omega} + e_n^{1/q_0} + M_p(\omega, e_n)\}.$$

More precisely the sum of worst type I and type II errors is at most $1/3$ above a sufficiently large multiple of this expression, and at least $2/3$ below a sufficiently small multiple. The upper statement is immediate for an empty alternative class.

At zero width the formula gives the known-covariance rate. At fixed positive width it gives the unrestricted-covariance rate as $e_n \to 0$. In the clean balanced two-component case these sparse rates agree with Verzelen and Arias-Castro [4]. Chen and Li [15] study univariate homogeneity with the EM test. The unrestricted robust finite-mixture rate appears in Section 4; the new assertion is uniform as the covariance interval shrinks. For example, when $k \geq 3$, $p = 3$, and e_n is of order $\omega^{3/2}$, the intermediate term is of order $\omega^{5/12}$, while $\sqrt{\omega} + e_n^{1/3}$ is of order $\omega^{1/2}$.

5.2 An odd-moment bound for the far tail

Lemma 5.2. *Let a variable V have at most k atoms in $[-2R, 2R]$. Suppose its moments through order $2k-1$ are within $C_0 e$ of those of $N(0,t)$, with $e \geq 0$ and $0 \leq t \leq \omega$. Then, for a constant D_k depending only on k,*

$$\mathbb{E}\left[|V|^{2k-1} \mathbf{1}_{\{|V| > D_k\sqrt{\omega}\}}\right] \leq Ce,$$

with C depending only on C_0, k and R.

Proof. Choose $L_k > 2\sqrt{k}$ and set $b_0 = L_k\sqrt{\omega}$. If there is a support point above $2b_0$, let Q be the monic polynomial whose roots are all nodes at or below b_0. Its degree m is at most $k-1$, and the empty product is allowed. The polynomial $G(x) = (x-b_0)Q(x)^2$ is nonnegative on the support of V, vanishing at the nodes at or below b_0. At $x > 2b_0$ it is at least $c|x|^{2m+1}$, and therefore at least $c_R|x|^{2k-1}$ since $|x| \le 2R$.

For a Gaussian variable of variance $t > 0$, multiplication by x compressed to polynomials of degree at most $k-1$ has operator norm at most $2\sqrt{kt}$. Indeed in the orthonormal Hermite basis its tridiagonal entries adjacent to the diagonal are $\sqrt{jt}$, $1 \le j < k$. Its largest absolute row sum bounds the norm. Hence the Gaussian expectation of G is nonpositive. For $t = 0$ the same follows by evaluating at zero. All the coefficients of G are bounded uniformly, since its roots and the cutoff are bounded, and its degree is at most $2k-1$. Thus $\mathbb{E}G(V) \le Ce$. If no support point lies above $2b_0$ that side contributes zero. Apply the same argument to $-V$. □

5.3 Testing upper bound

Use the maximum, over unit s-sparse directions, of the scalar Kolmogorov distances after Gaussian smoothing of variance one. Denote it by D_s^+. The distribution-free half-space bound of Section 4 gives

$$D_s^+(P_n, P) \le C(\sqrt{A/n} + \epsilon)$$

with probability at least $5/6$ under each allowed contaminated law. Test by the distance of P_n from the clean null class in D_s^+. As in that paper, compactness and smoothing allow the infimum and supremum to be taken over countable dense sets. It suffices to show $\Delta_p(\mu) \le C\mathcal{R}_p(\omega, e)$ whenever the clean mixture is within $C_0 e$ of any clean null in D_s^+. The case of e bounded below by a positive constant is automatic after increasing C, by bounded latent separation.

Choose a direction v supported on the alternative's coordinate set retaining a fixed fraction of all its latent pairwise distances, as in Section 4. Then $\Delta_p(\mu) \lesssim \Delta_p(\mu_v)$. Lemma 4.3 gives $O(e)$ discrepancy between the observed scalar moments after smoothing through order $2k$. Its hypotheses hold uniformly, since both projected smoothed variances have fixed upper and positive lower bounds. The arbitrary-covariance scalar inverse of Lemma 2.2 then bounds $\Delta_p(\mu_v)$ by $Ce^{1/q}$.

For the width bound, center at the null's projected mean and undo the alternative's scalar Gaussian variance including the smoothing. The moment equations compare the centered latent variable V to formal centered Gaussian moments with variance t, the null projected variance minus the alternative projected variance. The discrepancy is $O(e)$: undoing that variance and centering are linear triangular maps on the raw moments, with uniformly bounded coefficients. We have $|t| \le \omega$. If $t < 0$, the second moment gives $\mathbb{E}V^2 \le Ce$; bounded support and Jensen give $\|V\|_p \le Ce^{1/q_0}$. For $t \ge 0$ we have

$$\mathbb{E}|V|^j \le C(\omega^{j/2} + e) \quad \text{for } j = 2, 4, \dots, 2k-2. \tag{12}$$

The contribution of $|V| \le D_k\sqrt{\omega}$ to $\|V\|_p$ is $O(\sqrt{\omega})$. On the complement, the odd endpoint $J = 2k-1$ has Jth absolute moment at most Ce by Lemma 5.2. Thus $\|V\|_p \le C(\sqrt{\omega} + e^{1/q_0})$ at every knot, by the second-moment bound for $p \le 2$, and by bounded support for $p \ge 2k-1$.

For an interior p between consecutive knots $j < J$, put $\theta = (p-j)/(J-j)$. Hölder interpolation of the two exterior moments gives an upper bound on the pth moment of $C(\omega^{j/2}+e)^{1-\theta}(\omega^{J/2}+e)^{\theta}$. The bound also applies when J is odd, by replacing its actual Ce bound by $C(\omega^{J/2}+e)$. If $e \le \omega^{J/2}$, this product is $O(\omega^{p/2})$. If $e \ge \omega^{j/2}$, it is $O(e)$. Between the two thresholds it is $O(M_p(\omega, e)^p)$. These arguments prove the width bound on $\Delta_p(\mu_v) \le \|V\|_p$. Combining with the arbitrary covariance bound proves the desired upper comparison and hence the testing upper bound.

5.4 A testing bound for planted directions

Use the reference $P_* = N(0, B)$. For a unit vector u let P_u be the convolution of a centered scalar latent variable V placed along u with $N(0, B - tuu^\top)$, where $t \in [0, \omega]$. After whitening by $B^{-1/2}$, this model differs from a standard Gaussian only along the unit vector $a_u = B^{-1/2}u/\|B^{-1/2}u\|$. The scalar latent variable in that direction is $\|B^{-1/2}u\|V$, and the retained scalar Gaussian variance is $1 - t\|B^{-1/2}u\|^2$. Consider any such family for which its scalar density ratio relative to a standard Gaussian is within $C\delta$ of one in Gaussian L^2, uniformly in u. Its first Hermite coefficient vanishes by centering. The Hermite correlation identity of Section 4, now allowing different coefficients in different directions, and Cauchy–Schwarz give

$$|\mathbb{E}_{P_*}(L_u L_v) - 1| \le C\delta^2 |\langle a_u, a_v\rangle|^2, \qquad L_u = dP_u/dP_*.$$

Choose u, v independently with uniformly random s coordinates and independent uniform signs of magnitude $s^{-1/2}$. For any fixed vector z and real x, conditioning on the chosen coordinates and expanding a product gives

$$\mathbb{E}_v e^{x\langle z,v\rangle} \le \prod_{i=1}^{d} [1 + (s/d)(\cosh(xz_i/\sqrt{s}) - 1)].$$

Indeed the probability that a given set of m indices is included is $(s)_m/(d)_m \le (s/d)^m$, bounding each nonnegative term in the expansion. With $L = \log(ed/s)$, for every real y,

$$\log[1 + (s/d)(\cosh y - 1)] \le Cy^2/L.$$

For $|y| \le L/2$ use $\cosh y - 1 \le y^2 e^{|y|}/2$ and $Le^{-L/2} \le C$; otherwise use the bound by $|y|$. Thus $\langle z, v\rangle$ has sub-Gaussian tails at scale $\|z\|/\sqrt{A}$. With $z = B^{-1}u$ and the two normalizing denominators in $\langle a_u, a_v\rangle$ bounded above and away from zero, this yields a uniform tail $\mathbb{P}(|\langle a_u, a_v\rangle| > x) \le 2e^{-cAx^2}$. As in Section 4, integration gives $\mathbb{E}\exp(\alpha A\langle a_u, a_v\rangle^2) \le 1 + O(\alpha)$ for small α. The likelihood second moment for the prior using the same random u for all n observations is then at most

$$\mathbb{E}\exp(Cn\delta^2 |\langle a_u, a_v\rangle|^2).$$

For $n\delta^2 \le cA$ and small c, the prior and the null have total variation at most $1/3$. Alternatively, at a fixed direction if the clean total variation is at most $\epsilon/(1-\epsilon)$, their Huber neighborhoods intersect. Thus, writing $\delta_* = \max(1 \wedge \sqrt{A/n}, \epsilon)$, it suffices to construct separated families whose scalar Gaussian L^2 gap is at most a sufficiently small multiple of δ_*. This suffices both for the clean lower bound when the first term is larger and for indistinguishability under contamination when the second is larger. The scales δ_* and e_n are comparable by constants, uniformly in the parameters.

5.5 Separated scalar alternatives

Write $e = \delta_*$. First take $t = 0$. For $p \le 2$ use two equally weighted atoms at $\pm c\sqrt{e}$. The Gaussian L^2 gap is $O(c^2 e)$, since the squared gap in a direction with whitening factor b is $\cosh(b^2c^2e) - 1$. The separation is $c\sqrt{e}$. For $p > 2$, use $V = a(B_w - w)$ with $a > 0$ fixed below R, B_w Bernoulli of probability w, and $w = ce$. The scalar Gaussian L^2 gap is $O(w)$: the dominant component changes its mean by $O(w)$, and the other component has weight w and bounded mean. The separation is at least $(a/2)w^{1/p}$. This gives the term e^{1/q_0}.

Next, scale a k-point Gaussian quadrature by

$$h = c\min\{\sqrt{\omega}, e^{1/(2k)}\}, \qquad t = h^2.$$

Its separation is of order h. The Hermite expansion of Section 2.4 gives a scalar L^2 gap of $O(h^{2k})$ uniformly in the whitening factor. Together these two constructions cover the formula if $M_p = 0$ or if M_p is dominated by the first two width terms. The first construction suffices at $\omega = 0$ and for $p \geq 2k$.

Between knots, $M_p \leq \sqrt{\omega}$ for $e \leq \omega^{J/2}$ and $M_p \leq e^{1/p}$ for $e \geq \omega^{j/2}$. It remains to treat $\omega^{J/2} \leq e \leq \omega^{j/2}$ for $\omega > 0$.

Quadrature with prescribed nodes is classical [16]. Here the remote distance and mass must also have the scales specified by the statistical comparison.

Lemma 5.3. *For any consecutive knots $j < J$, let $t > 0$ and let $a > 0$ with $a/\sqrt{t}$ sufficiently large. There exists a scalar law of at most k atoms matching $N(0,t)$ through degree $J-1$. Its core is supported in $[-C\sqrt{t}, C\sqrt{t}]$, and it has total mass of order $t^{j/2}a^{-j}$ on one or two remote atoms at distance a from zero. In particular its Jth absolute moment is $O(t^{J/2} + t^{j/2}a^{J-j})$, and*

$$\inf_b (\mathbb{E}|V-b|^p)^{1/p} \gtrsim (t^{j/2}a^{p-j})^{1/p}.$$

Proof. First suppose $J = j+2$ is even and write $j = 2l$. Work with the squared and normalized variable $Y = V^2/t$, and let m_r be the moments of a standard normal squared. Write $U = a^2/t$, and place a mass λU^{-l} at U. Choose the inner rule to match the residual moments $m_r - \lambda U^{r-l}$, $0 \leq r < l$. When $l = 2q$ is even, start from the q-node Gaussian quadrature of the squared normal measure. When $l = 2q+1$ is odd, start from its Gauss–Radau rule with node zero and q positive nodes. The latter can be obtained by Gaussian quadrature for the measure weighted by y, dividing positive-node weights by their nodes and adding the residual mass at zero. That residual mass is positive, since it equals the squared nodal polynomial of the positive nodes integrated against the original measure and normalized by its value at zero squared.

For either rule the moment map through degree $l-1$ has an invertible Jacobian in the free weights and nodes. Indeed a polynomial of degree at most $l-1$ orthogonal to the Jacobian rows would have zero values and zero derivatives at every moving node, and also zero value at the fixed node in the Radau case. Counting these zeros gives at least l zeros with multiplicity. The inner rule continues smoothly under perturbations of size $O(U^{-1})$, uniformly for bounded λ. Its weights and moving nodes remain positive. This includes $l = 1$, when the inner rule has only the node zero with a variable weight.

At $\lambda = 0$, the inner lth moment is smaller than m_l by a fixed positive amount D: integrate the monic squared product over the inner nodes, times y in the Radau case. This polynomial vanishes on the inner rule and has degree l, with all lower moments matched. For large U the inner lth moment changes by only $O(U^{-1})$ for $0 \leq \lambda \leq 2D$, while the remote atom adds exactly λ. By continuity some $\lambda \asymp 1$ matches degree l. Lift the positive inner nodes and the remote atom to equal positive and negative square roots times $\sqrt{t}$, keeping a zero node if present. There are l inner nodes and two remote ones after lifting, so there are at most k nodes since $l \leq k-2$ in this even case. Symmetry gives matching odd moments through $J-1$, and the construction gives the even ones.

In the remaining case $j = 2k-2$, $J = 2k-1$, use the normalized variable $V/\sqrt{t}$ directly. Put a single remote atom at $U = a/\sqrt{t}$ of mass λU^{-j}. Use the $(k-1)$-node standard Gaussian quadrature as the inner rule and adjust its moments through degree $j-1$ by the perturbations $-\lambda U^{r-j}$. The same Jacobian and positive deficit argument uses the squared nodal polynomial of degree j. An

inner perturbation of size $O(U^{-1})$ together with $\lambda \asymp 1$ matches degree j. There are k nodes in this case. This proves the required support, moment and mass statements. For the transport bound, either the center is far from a remote atom with mass of the stated order, or all the core mass is a distance of order a away. The core has mass bounded away from zero for sufficiently large $a/\sqrt{t}$. □

We now apply the lemma. Choose fixed small constants $\gamma, \kappa > 0$ and set

$$t = \kappa\omega, \qquad a = \gamma(e/\omega^{j/2})^{1/(J-j)}.$$

Then $a \le \gamma$ and $a/\sqrt{t} \ge \gamma/\sqrt{\kappa}$. Choose $\gamma < R/2$ and κ small enough relative to γ to apply the lemma and keep the core inside the radius-R ball. The separation is a constant times $M_p(\omega, e)$. Indeed $(t^{j/2}a^{p-j})^{1/p} \asymp M_p(\omega, e)$.

For the Gaussian L^2 gap, compare the alternative to a latent $N(0, t)$ in the same direction with the same retained noise, which gives the clean null. In whitened scalar coordinates expand the retained-noise Gaussian location kernel as a function of its shift to order $J-1$ in L^2 of the standard Gaussian reference. The kernel here is the retained Gaussian density divided by the reference Gaussian density. For every intermediate shift z, the L^2 norm of its Jth shift derivative is at most $C\exp(Cz^2)$, by completing the square. The estimate is uniform for the retained variances in a fixed neighborhood of one. The location moments agree through degree $J-1$, so the gap is bounded by

$$C(t^{J/2} + t^{j/2}a^{J-j}).$$

Here the compact alternative remainder is controlled by Lemma 5.3, and the reference latent Gaussian remainder uses $\mathbb{E}[|\sqrt{t}G|^J e^{CtG^2}] \lesssim t^{J/2}$ for small t. The whitening factors remain bounded, so these estimates hold in every direction. By $e \ge \omega^{J/2}$, the two terms are at most $Ce(\kappa^{J/2} + \kappa^{j/2}\gamma^{J-j})$. Take κ small enough for the lower-bound argument in the preceding section. In this band the lower bounds for the individual terms combine by their maximum, which is equivalent to their sum. With the global cap, the three lower constructions prove Theorem 5.1.

6 Sharp isotropic variance rates for planar Gaussian mixtures

We determine the minimax rate for the common isotropic variance of a planar Gaussian location mixture with at most $k \le 35$ bounded means. The successive rates are $n^{-1/2}$, $n^{-1/4}$, $n^{-1/6}$, $n^{-1/8}$, $n^{-1/10}$, $n^{-1/12}$, and $n^{-1/14}$, with transitions at $k = 3, 6, 10, 15, 21, 28$. At the last five transitions we construct pairs of positive measures with moments matching through degrees five, seven, nine, eleven, and thirteen after one law is convolved with an isotropic Gaussian. Each pair uses fewer nodes than ordinary planar Gaussian cubature at the same degree.

6.1 Variance estimation and moment matching

Observe n independent draws from $\mu * N(0, tI_2)$, where μ has at most k atoms in a fixed radius-R disk, $R > 0$, and t belongs to a fixed interval $[t_-, t_+]$ with $0 \le t_- < t_+$. Put $m(k) = \min\{m : \binom{m+2}{2} > k\}$.

Theorem 6.1. *For each $1 \le k \le 35$, the minimax expected absolute variance error is of order $n^{-1/(2m(k))}$, with constants depending only on the class parameters.*

Atom cap k	1–2	3–5	6–9	10–14	15–20	21–27	28–35
Variance error	$n^{-1/2}$	$n^{-1/4}$	$n^{-1/6}$	$n^{-1/8}$	$n^{-1/10}$	$n^{-1/12}$	$n^{-1/14}$

For $k \le 2$ this is the rank case of the isotropic calibration theorem of Section 3, using the population identity of Hsu and Kakade [10]. The polynomial upper bound and the three-atom lower construction are also in Section 3. The constructions below handle the next five planar thresholds. The distinction from ordinary cubature is useful because the lower model can use two non-Gaussian latent laws. Both laws have positive atomic weights and use a shared scalar noise variance within each model. Agostini, Améndola and Ranestad [9] study generic moment identifiability for homoscedastic mixtures. Curto and di Dio [17] develop the heat evolution of truncated moment sequences. Here the finite supports on two points of a common heat curve give a worst-case statistical lower bound for the scalar noise.

Lemma 6.2. *If two centered finite atomic laws μ, ν have the same moments through degree $2m-1$ after convolving μ with $N(0, \tau I_2)$ for a fixed $\tau > 0$, the model above has minimax error at least $cn^{-1/(2m)}$ whenever each law has at most k atoms. If both atomic laws have at most k nodes and $\binom{m+2}{2} > k$, matching through degree $2m$ is impossible.*

Proof. Scale both laws by $h > 0$, use a fixed $t_0 \in (t_-, t_+)$ for the first law, and variance $t_0 - h^2\tau$ for the second. Dividing their observed moment generating functions by that of $N(0, t_0 I_2)$ gives $\mathbb{E}e^{hz^\top U}$ and $e^{-h^2\tau\|z\|^2/2}\mathbb{E}e^{hz^\top V}$, respectively. The first $2m$ coefficients, counting the constant term, agree by the assumed moment matching. The corresponding density ratios are $2m$ times continuously differentiable in h in Gaussian L^2 near zero. This follows by differentiating the Gaussian kernels, whose h-derivatives have polynomial factors, bounded means and covariance strictly below twice the reference covariance. The Taylor cancellations for the moment generating functions imply the same cancellations for the density ratios, by uniqueness of their exponential transforms. The difference of the ratios is therefore $O(h^{2m})$ in Gaussian L^2. The first density dominates a positive constant times the reference density by Jensen's inequality and $\mathbb{E}U = 0$. The one-sample chi-square divergence is $O(h^{4m})$. With $h = cn^{-1/(4m)}$, the two-point testing inequality yields an absolute variance risk of order at least h^2.

For the impossibility, choose a nonzero polynomial of degree at most m vanishing on every node of ν, by the dimension of the planar polynomial space. Its square has degree at most $2m$ and integrates to zero under ν but positively under a convolution with a nondegenerate Gaussian. □

For the upper bound of Theorem 6.1, fit the observed moments through degree $2m(k)$ by an admissible pair over the compact parameter class. The empirical moments have expected error $O(n^{-1/2})$ since the support and Gaussian variance are bounded. Use a minimizer to tolerance $n^{-1/2}$ over a fixed countable dense subset to ensure measurability. For fitted and true pairs with moment discrepancy δ, remove the smaller Gaussian variance in their moment equations by the Gaussian exponential generating function. Let τ be the difference between the two variances. This compares the smaller-noise latent law to the larger-noise latent law convolved with $N(0, \tau I_2)$, with discrepancy at most $C\delta$ in the same moments. Normalize a polynomial P vanishing on the nodes of the smaller-noise law to have unit coefficient norm. On each latent location x of norm at most R, the coefficients of $P(x + \cdot)$ have norm bounded below by a fixed constant, since translation has uniformly bounded inverse there. Scaling by $\sqrt{\tau}$ and using equivalence of polynomial coefficient norm and Gaussian L^2 norm gives $\mathbb{E}P(x + \sqrt{\tau}\, G)^2 \ge c \min(1, \tau)^m$. Its expectation is at most $C\delta$ by the moment comparison and vanishing on the other law. Since τ is bounded, $\tau \le C\delta^{1/m}$. Jensen's inequality proves the upper rate.

For $k \le 2$, use the simple Gaussian variance lower bound. For $k = 3, 4, 5$, the three-node construction of Section 3.5 matches moments through degree three with a covariance offset. It takes a centered law at $(u, v), (u, -v), (-wu/(1-w), 0)$ with weights $w/2, w/2, 1-w$. For diagonal covariance (x, y), set $T = \sqrt{60}$, $u = T/y$, $w = x/(x + u^2)$, $v^2 = y/w$. The choices $(x, y) =$

$(1,2),(2,3)$ give identical third tensors, whose nonzero entries have values T and $T(x/y - x^2y/T^2)$; their covariances differ by I_2. This gives the necessary moment matching after convolution.

6.2 Six nodes and fifth moments

Write the plane as $\mathbb{C}$ and $\zeta_3 = e^{2\pi i/3}$. For $a > 1$, place total mass $1/(1+a^3)$ uniformly on $\{a, a\zeta_3, a\zeta_3^2\}$ and the remaining mass uniformly on $\{-1, -\zeta_3, -\zeta_3^2\}$. The third harmonic moment $\mathbb{E}z^3$ is zero. The radial moments and the nonzero fifth-degree harmonic are

$$S_2 = \frac{a^2 + a^3}{1 + a^3}, \qquad S_4 = \frac{a^4 + a^3}{1 + a^3}, \qquad H_5 = \mathbb{E}z^4\bar{z} = \frac{a^5 - a^3}{1 + a^3}.$$

Scale the law by $H_5^{-1/5}$ so that its fifth-degree harmonic is one. Its scaled radial second moment is $A(a) = S_2/H_5^{2/5}$. The scaled fourth radial invariant is $B(a) = (S_4 - 2S_2^2)/H_5^{4/5}$. This is negative near $a = 1$ and is positive for all sufficiently large a, with $B(a) \sim a^{-3/5}$. Meanwhile $A(a) \to 0$ at infinity. Choose a finite sign-changing zero of B with positive values to its right. By continuity, arbitrarily small positive levels of B occur in a compact interval to the right of that zero and also at arbitrarily large arguments. To see the latter, fix an arbitrarily large point of positivity and use the convergence to zero. The values of A on the compact interval are bounded away from zero. Thus there is a pair with the same positive value of B and distinct A.

Under convolution with $N(0, \tau I_2)$, the second radial moment increases by 2τ. The fourth increases by $8\tau\,\mathbb{E}|z|^2 + 8\tau^2$, so $\mathbb{E}|z|^4 - 2(\mathbb{E}|z|^2)^2$ is unchanged. The third harmonic is unchanged, and the fifth harmonic increases by $8\tau\,\mathbb{E}z^3 = 0$. These identities follow from the finite expansion of the heat operator $\exp[(\tau/2)\Delta]$ on polynomials, with $\Delta = 4\partial_z\partial_{\bar{z}}$. Threefold rotational symmetry leaves, through degree five, only these radial and harmonic moments, their conjugates and the mass. Add to the smaller-second-moment law the positive Gaussian variance that equalizes the second moments. All the moments through degree five now agree. This gives the lower bound for $6 \le k \le 9$ by Lemma 6.2.

6.3 Ten nodes and seventh moments

Two regular pentagons with variable relative angle suffice. Fix $a = 1/20$, and let $b \in [59/10, 6]$. The squared radii are

$$x = 1 - a, \qquad y = 1 + b, \qquad w = \frac{b}{a + b},$$

with total masses $w, 1 - w$ on the two rings. Write θ for the fifth harmonic phase of the second pentagon, relative to that of the first, and put

$$A = wx^{5/2}, \qquad B = (1 - w)y^{5/2}, \qquad C = A, \qquad D = Be^{i\theta}.$$

The fifth harmonic $H_5 = \mathbb{E}z^5$ is $C + D$; the seventh $H_7 = \mathbb{E}z^6\bar{z}$ is $xC + yD$. For fivefold rotational symmetry the other moments through degree seven are the three radial moments, the conjugates of H_5, H_7, and the mass.

After adding $N(0, \tau I_2)$, the moments of the squared radius of orders one, two and three become

$$\begin{aligned} S_2(\tau) &= S_2 + 2\tau, \\ S_4(\tau) &= S_4 + 8\tau S_2 + 8\tau^2, \\ S_6(\tau) &= S_6 + 18\tau S_4 + 72\tau^2 S_2 + 48\tau^3. \end{aligned} \tag{13}$$

The fifth harmonic stays fixed and the seventh becomes $H_7 + 12\tau H_5$, by the same heat operator calculation. Match the radial moments with two positive squared radii x_τ, y_τ and mass w_τ at x_τ. Explicitly, if μ_τ, v_τ are the mean and variance determined by (13), and u_τ is the third central moment divided by v_τ, take unique positive a_τ, b_τ with

$$a_\tau b_\tau = v_\tau, \qquad b_\tau - a_\tau = u_\tau.$$

Then $x_\tau = \mu_\tau - a_\tau$, $y_\tau = \mu_\tau + b_\tau$, and $w_\tau = b_\tau/(a_\tau + b_\tau)$. These are smooth for sufficiently small τ uniformly in the displayed b interval, with positive radii and weights. At $\tau = 0$ they reduce to the original radial parameters.

The target complex harmonic coefficients C_τ, D_τ must satisfy $C_\tau + D_\tau = C + D$ and $x_\tau C_\tau + y_\tau D_\tau = (x + 12\tau)C + (y + 12\tau)D$. Thus

$$\begin{aligned} C_\tau &= \alpha C + \beta D, & D_\tau &= (1-\alpha)C + (1-\beta)D, \\ \alpha &= \frac{y_\tau - x - 12\tau}{y_\tau - x_\tau}, & \beta &= \frac{y_\tau - y - 12\tau}{y_\tau - x_\tau}. \end{aligned}$$

A target pair of pentagons exists precisely when these coefficients have respective magnitudes $A_\tau = w_\tau x_\tau^{5/2}$ and $B_\tau = (1 - w_\tau) y_\tau^{5/2}$. Each requirement specifies a cosine value:

$$\begin{aligned} c_1(\tau, b) &= \frac{A_\tau^2 - \alpha^2 A^2 - \beta^2 B^2}{2\alpha\beta AB}, \\ c_2(\tau, b) &= \frac{B_\tau^2 - (1-\alpha)^2 A^2 - (1-\beta)^2 B^2}{2(1-\alpha)(1-\beta)AB}. \end{aligned} \tag{14}$$

We show these agree at an interior value of b and lie in $(-1, 1)$.

Primes below denote derivatives in τ at zero. From (13), using $\mu_0 = 1$, $v_0 = ab$ and $u_0 = b - a$, we have

$$\begin{aligned} \mu_0' &= 2, & v_0' &= 4, & u_0' &= 12 - \frac{4(b-a)}{ab}, \\ a_0' &= \frac{4 - a u_0'}{a+b}, & b_0' &= a_0' + u_0', & w_0' &= \frac{a b_0' - b a_0'}{(a+b)^2}. \end{aligned}$$

In particular $x_0' = 2 - a_0'$, $y_0' = 2 + b_0'$. Write $D_1 = \alpha_0' = (x_0' - 12)/(a+b)$ and $E_1 = \beta_0' = (y_0' - 12)/(a+b)$. These are strictly negative throughout the bracket:

$$D_1 = -\frac{40(100b^2 + 79b - 2)}{b(20b+1)^2}, \qquad E_1 = -\frac{600(52b-5)}{(20b+1)^2}.$$

Since $\alpha_0 = 1$, $\beta_0 = 0$, dividing the numerators and denominators of (14) by τ extends them continuously to

$$c_1(0, b) = \frac{A_0' - D_1 A}{E_1 B}, \qquad c_2(0, b) = -\frac{B_0' + E_1 B}{D_1 A}.$$

The extension is uniform across the compact bracket. The sign of $c_1(0,b) - c_2(0,b)$ is the sign of $(A_0' - D_1 A) D_1 A + (B_0' + E_1 B) E_1 B$. Substitute the derivatives above and use

$$\frac{A_0'}{A} = \frac{w_0'}{w} + \frac{5x_0'}{2x}, \qquad \frac{B_0'}{B} = -\frac{w_0'}{1-w} + \frac{5y_0'}{2y}.$$

The sign expression becomes

$$\frac{421200b^4 - 688680b^3 - 9599227b^2 - 6681812b + 374004}{10(20b+1)^3}. \tag{15}$$

Its numerator equals $-425494307/100$ at $59/10$ and 11831280 at 6. The same substitution gives

$$c_1(0,b)^2 = \frac{6859(2000b^3 - 3400b^2 - 1267b + 162)^2}{7200000(b+1)^5(52b-5)^2} \le \frac{34295}{69984} < \frac{1}{2}.$$

For the inequality bound the absolute cubic by 600000 on the bracket, and use $b+1 \ge 6$, $52b-5 \ge 300$.

For any sufficiently small positive τ, the endpoint signs from (15) persist and the first cosine remains in $(-1,1)$ throughout. The intermediate value theorem furnishes an interior b for which the two values agree. Choose θ to have their common cosine. The target coefficients then have the prescribed magnitudes, so their arguments determine valid rotations of the two target pentagons. The resulting ten-node laws have equal moments through degree seven after Gaussian convolution of the first law by this fixed $\tau > 0$. The lower bound for every $10 \le k \le 14$ now follows from Lemma 6.2.

6.4 Fifteen nodes and ninth moments

Use two regular heptagons and an origin atom. The initial squared radii are $x = 1$ and $y \in [17/5, 7/2]$, with ring masses $p = 1/4$ and $q = 1/2000$; the remaining mass is at the origin. Write $S_j = \mathbb{E}|z|^{2j}$ for the radial moments, with $S_0 = 1$. Under the heat operator their derivatives satisfy $S_j' = 2j^2 S_{j-1}$. To match the four heated radial moments, continue the four parameters p, q, x, y as smooth functions of τ. The Jacobian of the map $(p,q,x,y) \mapsto (px^j + qy^j)_{1 \le j \le 4}$ is invertible. A dual polynomial of degree at most four with no constant term annihilating its columns would vanish doubly at the two positive radii and also at zero, which is impossible unless it is zero. Equivalently, if $S_j(\tau)$ are the heated values, the two target radii are roots of a quadratic with sum and product

$$\frac{S_1S_4 - S_2S_3}{S_1S_3 - S_2^2}, \qquad \frac{S_2S_4 - S_3^2}{S_1S_3 - S_2^2},$$

respectively, with each S_j evaluated at τ. These formulas follow from the two-node recurrence for the size-biased radial moments. They give a uniform smooth continuation on the compact bracket for small τ. The two radii, both ring masses, and the remaining origin mass remain positive by continuity.

The only nonradial moments through degree nine are $H_7 = \mathbb{E}z^7$, $H_9 = \mathbb{E}z^8\bar{z}$, and their conjugates. The heat step keeps H_7 and adds $16\tau H_7$ to H_9. Use the same two equations for the complex harmonic coefficients as for the pentagons, now with 16τ replacing 12τ and $A = px^{7/2}$, $B = qy^{7/2}$. In particular (14) remains valid with these A, B and the corresponding α, β. In the limit $\tau = 0$, the quantities D_1, E_1 are $(x_0' - 16)/(y - x)$, $(y_0' - 16)/(y - x)$. Differentiate the four equations for S_j to get

$$x^j p_0' + y^j q_0' + jpx^{j-1}x_0' + jqy^{j-1}y_0' = 2j^2 S_{j-1}, \qquad 1 \le j \le 4.$$

Solving gives

$$D_1 = \frac{(2y-1)(y^2 - 2000y + 500)}{250(y-1)^3}, \qquad E_1 = -\frac{2P(y)}{y(y-1)^3}, \qquad P(y) = y^3 - 6y^2 + 2504y - 1000.$$

Both are negative throughout the bracket. Use $A_0'/A = p_0'/p + 7x_0'/(2x)$ and $B_0'/B = q_0'/q + 7y_0'/(2y)$. The sign expression $(A_0' - D_1 A)D_1 A + (B_0' + E_1 B)E_1 B$ becomes $N(y)/[2000000(y-1)^3]$, where

$$N(y) = -(23y^8 - 136y^7 + 55079y^6 + 8y^5 - 6271001y^4 \\ + 9984000y^3 + 41761000y^2 - 34000000y + 5750000).$$

At 17/5 and 7/2 its values are respectively $-4721948853653/390625$ and $3522274957/256$. The first limiting cosine has square

$$c_1(0,y)^2 = \frac{Q(y)^2}{4y^5 P(y)^2}, \qquad Q(y) = 10y^4 + 3991y^3 - 28999y^2 + 33500y - 11500.$$

Bounding the monomials by their respective endpoint values with their indicated signs gives $-100000 < Q(y) < 0$ and $P(y) > 7400$. Since $y^{5/2} > 15$, the first cosine has absolute value less than $1/2$ throughout. The same continuous extensions and intermediate value argument as for the pentagons yield positive τ and two fifteen-node laws with the required matching through degree nine. By Lemma 6.2, the variance rate is $n^{-1/10}$ for $15 \le k \le 20$.

6.5 Twenty-one and twenty-eight nodes

Place positive masses w_1, w_2, w_3 uniformly on three regular s-gons with respective squared radii x_1, x_2, x_3, and place mass η at the origin when indicated. Take $x_1 = 1$, $w_1 = 1 - \eta - w_2 - w_3$, and let $x_3 = b$ vary in the following brackets.

Degree	s	η	x_2	w_2	w_3	$1000b$
11	7	0	17/5	1/800	10^{-11}	$[4791, 4792]$
13	9	3/5	2	1/4000	1/400000	$[3707, 3708]$

The radial matching orders are $0, \dots, 5$ and $0, \dots, 6$, respectively. Match the heated radial moments by continuing the three masses and three squared radii, together with a free origin mass in the second case. The Jacobian is invertible. In the first case a dual degree-five polynomial in its kernel would have double zeros at three distinct nodes. In the second case it would have the three double zeros and a zero at the origin despite having degree at most six. The implicit function theorem and compactness give continuations of positive weights and distinct positive squared radii uniformly for small τ. Local uniqueness patches the continuations along each bracket.

Write $u = |z|^2$. For the radial slopes use the heat derivative functional

$$\mathcal{R}(f) = 2\eta f'(0) + 2\sum_{j=1}^{3} w_j [x_j f''(x_j) + f'(x_j)].$$

Let L_i be the degree-two Lagrange cardinal at x_i on the three squared radii. With $e = 0$ for the first row and $e = 1$ for the second, put

$$h_i(u) = (u/x_i)^e L_i(u)^2, \qquad q_i(u) = (u - x_i)h_i(u), \qquad p_i(u) = h_i(u) - h_i'(x_i)q_i(u).$$

At the ring nodes p_i, p_i' have values $\delta_{ij}, 0$, and q_i, q_i' have values $0, \delta_{ij}$. They vanish at the origin when $e = 1$. Differentiating the radial match at zero therefore yields $w_i' = \mathcal{R}(p_i)$ and $x_i' = \mathcal{R}(q_i)/w_i$.

At the degrees in the table, only the frequency s and its conjugate survive rotational symmetry besides the radial moments, since $2s$ exceeds the total degree. The frequency s moments have the

form $\mathbb{E}[z^s f(u)]$ for f of degree at most two. Write $C_j = w_j x_j^{s/2} e^{is\theta_j}$ for the complex harmonic contribution of ring j. On these moments the heat operator acts on f by $e^{\tau\mathcal{A}}$, where $\mathcal{A}f = 2uf'' + 2(s+1)f'$. Let L_i^τ be the Lagrange cardinal on the target squared radii and set

$$M_{ij}(\tau) = (e^{\tau\mathcal{A}} L_i^\tau)(x_j), \qquad B_{ij} = M_{ij}'(0) = 2x_j L_i''(x_j) + [2(s+1) - x_j'] L_i'(x_j).$$

The derivative formula uses $L_i^\tau(x_j(\tau)) = \delta_{ij}$. Interpolation shows that the heated harmonic matches exactly when the target contribution vector is MC.

Let G be the real Gram matrix of the three C_j viewed as planar vectors. Put $a_i(\tau) = w_i(\tau)^2 x_i(\tau)^s$, so that $G_{ii} = a_i(0)$. The target magnitudes match if

$$2\sum_{j<l} M_{ij} M_{il} G_{jl} = a_i(\tau) - \sum_j M_{ij}^2 a_j(0), \qquad 1 \le i \le 3.$$

Since $M(0) = I$, divide by 2τ and extend to $\tau = 0$ by differentiation. The limiting system for G_{12}, G_{13}, G_{23} is

$$T = \begin{pmatrix} B_{12} & B_{13} & 0 \\ B_{21} & 0 & B_{23} \\ 0 & B_{31} & B_{32} \end{pmatrix}, \qquad d_i = a_i(0)\left(\frac{w_i'}{w_i} + \frac{s x_i'}{2x_i} - B_{ii}\right).$$

All entries here are rational functions of b. The following table gives exact interval certificates for $\det T$ and the three principal minors of size two in the limiting G, obtained by Cramer's rule. Each cell gives the numerator and denominator degrees and the common sign of the numerator Bernstein coefficients. All denominator Bernstein coefficients are strictly positive.

Degree	$\det T$	minor 12	minor 13	minor 23
11	$(9, 8; -)$	$(28, 18; +)$	$(32, 18; +)$	$(32, 18; +)$
13	$(12, 9; -)$	$(34, 24; +)$	$(40, 24; +)$	$(40, 24; +)$

For reproduction, reduce each rational function and orient its denominator to be positive at the left endpoint l. For each polynomial substitute $b = l + (r-l)v$, where r is the right endpoint, and expand as $\sum_{i=0}^D c_i v^i$. Its Bernstein coefficient at j is $\sum_{i=0}^j c_i \binom{j}{i} / \binom{D}{i}$. Strictly same-signed coefficients certify the sign on the whole bracket by nonnegativity and partition of unity of the Bernstein basis. Direct rational evaluation of $\det G$ at $b = l, r$ gives signs $+, -$ in both rows. The source package contains three verification scripts. The script `verification/certificate.py` checks these coefficient and endpoint signs using exact arithmetic. The script `verification/seeded_pairs.py` reproduces the paired laws of degrees seven and nine numerically. The script `verification/indep_pairs.py` solves the exact equations at fixed positive τ in high precision at each of the degrees seven, nine, eleven and thirteen, builds both laws node by node and compares all moments through the stated degree.

Invertibility, strict principal minor inequalities, and the two endpoint signs persist uniformly for all sufficiently small $\tau > 0$. Because the third ring carries mass 10^{-11} at degree eleven, the admissible range of τ there is very small, of order 10^{-16}; at degree thirteen it is of order 10^{-9}. For any such fixed τ, continuity gives an interior b with $\det G = 0$. At this point G is positive semidefinite of rank two by the principal minor criterion. It is the Gram matrix of three planar vectors of the prescribed source lengths. Choose the source rotations from these vectors and the target rotations from MC, whose lengths match by the linear equations. The radial and harmonic moments now match through the specified degree. Lemma 6.2 gives $n^{-1/12}$ for $21 \le k \le 27$ and $n^{-1/14}$ for $28 \le k \le 35$, completing Theorem 6.1.

6.6 Comparison with ordinary Gaussian cubature

The five new paired constructions use fewer nodes than an ordinary positive cubature for the planar Gaussian through the same degrees. For centrally symmetric planar weights, Möller's lower bound [18] states that a cubature rule of degree $2s-1$ has at least $\dim \mathcal{P}_{s-1} + \lfloor s/2 \rfloor$ nodes, where $\mathcal{P}_{s-1}$ is the space of planar polynomials of degree at most $s-1$. For $s = 3, \dots, 7$ this gives $7, 12, 17, 24, 31$. Here is a direct verification of these counts by Möller's argument.

For degree five, evaluations of even polynomials of degree at most two have rank four, since exactness for their squares makes evaluation injective. Delete the origin coordinate if it is a node, reducing rank by at most one. Multiply these polynomials by a linear form nonzero at every nonorigin node. This produces an odd evaluation subspace of rank at least three. The two spaces are orthogonal in the weighted evaluation inner product because their products are odd and have degree at most five. Hence at least seven nodes are necessary. One positive seven-node rule uses half the mass at the origin and half uniformly on the hexagon of radius two. The first two nonconstant radial moments are then two and eight, and the remaining moments through degree five follow from sixfold symmetry.

For degree seven, odd polynomials through degree three have evaluation rank six, by exactness of their squares. They already vanish at the origin. Multiplying by a generic linear form as above gives an even evaluation subspace of rank six, orthogonal to the odd one. At least twelve nodes are necessary. Stroud's rule [21] attains the bound. In standard Gaussian normalization, it uses the four axis points at radius $\sqrt{6}$ with weight $1/36$ each, and the eight diagonal points in two groups, $(\pm\sqrt{u}, \pm\sqrt{u})$ and $(\pm\sqrt{v}, \pm\sqrt{v})$, where

$$u = (9 - 3\sqrt{5})/4, \qquad v = (9 + 3\sqrt{5})/4.$$

Their per-point weights are $p = (5 + 2\sqrt{5})/45$ and $q = (5 - 2\sqrt{5})/45$, respectively. For $A_j = 4(pu^j + qv^j)$ we have $A_0 = 8/9,\ A_1 = 2/3$, and the two-root recurrence gives $A_2 = 1,\ A_3 = 3$. The axis points supply the remaining pure second, fourth and sixth moments, and reflection gives the vanishing odd-coordinate moments. At degree nine the even polynomials through degree four have evaluation rank nine. Multiplication by a linear form produces odd polynomials through degree five with rank at least eight after allowing for an origin node. Orthogonality again shows that at least seventeen nodes are necessary for ordinary cubature. At degree eleven the odd polynomials through degree five give rank twelve, and multiplication by a generic linear form gives orthogonal even rank twelve. At degree thirteen the even polynomials through degree six give rank sixteen, and the odd products with a generic linear form give rank at least fifteen after accounting for the origin. Thus ordinary cubature requires at least twenty-four and thirty-one nodes at these degrees, respectively. Thus the paired constructions attain the maximum moment degree permitted by the polynomial dimension bound even at component counts where a point-mass reference cannot do so. The comparison uses lower bounds only. Cools and Verlinden [19] study rules of degree $4k+1$ attaining Möller's bound for integrals with circular symmetry, and Cools and Schmid [20] sharpen the bound at degree $4n+1$ for some circularly symmetric integrals. The counts 17 and 31 at degrees nine and thirteen are therefore lower bounds for the planar Gaussian; any larger minimal count only widens the advantage of the fifteen- and twenty-eight-node pairs.

7 Robust isotropic noise calibration and the finite-sample majority limit

We determine the absolute-error minimax risk for the common isotropic variance in a bounded k-component Gaussian location mixture under independent Huber contamination, uniformly for

$d \geq k$ and contamination below one half. Alongside sampling error and the usual contamination modulus, a binomial majority term is necessary for expected loss, even for a single centered Gaussian. We give its exact coefficient in the large-dimension limit for every fixed power loss and exhibit a logarithmic increase of the contamination modulus as the contamination fraction approaches one half. The absolute-loss rate is attainable without knowing the contamination fraction.

7.1 Model and result

Fix $k \geq 1$, $R > 0$, and a variance interval $[a, b]$ with $0 \leq a < b < \infty$. The clean law is $P_{\mu,t} = \mu * N(0, tI_d)$, where μ is a probability on at most k points of the radius-R ball in $\mathbb{R}^d$. We observe $n \geq 1$ independent draws from $(1-\epsilon)P_{\mu,t} + \epsilon H$, with H arbitrary. Let $\mathcal{R}_{n,d}(\epsilon)$ be the minimax expected absolute error for t. Write $D = b - a$, and for $J \sim \mathrm{Bin}(n, \epsilon)$ put

$$\beta_n(\epsilon) = \mathbb{P}(J > n/2) + \tfrac{1}{2}\mathbb{P}(J = n/2), \qquad \psi(\epsilon) = \epsilon\sqrt{1 + \log(1/(1-2\epsilon))}.$$

Theorem 7.1. *For every $d \geq k$ and $0 \leq \epsilon < 1/2$, uniformly in n, d, ϵ,*

$$\mathcal{R}_{n,d}(\epsilon) \asymp (nd)^{-1/2} + \min\{1, \psi(\epsilon)/\sqrt{d}\} + \beta_n(\epsilon),$$

with constants depending only on k, R, a, b. Moreover, for $d \geq \max\{n, k\}$ or $\epsilon \geq 1/4$,

$$D\beta_n(\epsilon) \leq \mathcal{R}_{n,d}(\epsilon) \leq D\beta_n(\epsilon) + C\left[(nd)^{-1/2} + \min\{1, \psi(\epsilon)/\sqrt{d}\}\right].$$

In particular, for fixed n, ϵ the limit as $d \to \infty$ is $D\beta_n(\epsilon)$. If $\epsilon = 1/2$ is admitted, the minimax risk is $D/2$ in every dimension. The same estimator, independent of ϵ, can attain these bounds at every $\epsilon < 1/2$.

The clean rate follows from dimension-averaged residual calibration in Section 3, refining the isotropic variance bound of Doss, Wu, Yang and Zhou [6, Theorem 6.1] in the range $d \geq k$. The centered latent covariance has rank at most $k - 1$, so the scalar variance is the smallest eigenvalue of the clean observation covariance, as in Hsu and Kakade [10, Theorem 1]. Matrix-depth methods of Chen, Gao and Ren [22] treat operator loss for general and structured covariance families. Their general Huber risk theory [1] gives a lower bound by the clean risk and the contamination modulus defined by total variation. We evaluate this modulus in the scalar model all the way to one half. Peng and Neykov [23, Theorem 2.1] find the analogous logarithmic dependence for Gaussian means with a deterministic bound on the number of corruptions. Minasyan and Zhivotovskiy [24, Section 3] analyze chi-square order statistics for robust covariance in operator norm. Ruckdeschel [25, Section 1.2] discusses removing the contaminated-majority event in higher-order risk expansions for the sample median on an unbounded location model. The additional term in Theorem 7.1 is an expected-loss contribution from the random number of contaminants. Its coefficient comes from a finite binary testing problem.

7.2 Lower bounds

Restrict first to $\mu = \delta_0$, and compare two centered Gaussians at positive variances in the interior of the bracket, separated by $c/\sqrt{nd}$. Their n-sample Kullback–Leibler divergence is bounded by a constant times nd times the square of their variance separation. Taking a sufficiently small $c > 0$ and using the two-point bound gives the sampling lower bound. Each clean law is attainable for every contamination level by taking H equal to that law.

For the contamination modulus, two Huber neighborhoods intersect when $\mathrm{TV}(P_0, P_1) \le \epsilon/(1-\epsilon)$. Indeed, the density $(1-\epsilon)\max(p_0, p_1)$ has integral at most one, and filling out its mass gives a law in both neighborhoods. We again use positive Gaussian variances from a fixed interior subinterval. Pinsker's inequality permits separation $c\epsilon/\sqrt{d}$. Write $\delta = 1 - 2\epsilon$. For small δ use the Gaussian affinity

$$A(P_0, P_1) = \left(\frac{2\sqrt{t_0 t_1}}{t_0 + t_1}\right)^{d/2}.$$

Cauchy–Schwarz gives $1 - \mathrm{TV} \ge A^2/2$. On the positive subinterval this lower bound is at least $\frac{1}{2}\exp[-Cd(t_1 - t_0)^2]$. For sufficiently small δ, the overlap therefore exceeds $\delta/(1-\epsilon)$ at separation $c\min\{1, \sqrt{\log(1/\delta)/d}\}$ with a fixed sufficiently small $c > 0$. Intermediate values of δ use the Pinsker bound, since the logarithmic factor is bounded there. Identical contaminated laws yield the claimed modulus lower bound.

Finally use the endpoints $P_0 = N(0, aI_d)$, $P_1 = N(0, bI_d)$, and choose the opposite Gaussian as the contaminant under each endpoint. More informative than the observation is the binary label specifying which Gaussian generated each draw. Under the two hypotheses the labels are iid Bernoulli with parameters ϵ and $1-\epsilon$. With equal prior weights the minimum testing error is $\beta_n(\epsilon)$, by the likelihood ratio test on the label count. For absolute loss between the two endpoints the optimum conditional action selects the more probable endpoint, so the label Bayes risk is exactly $D\beta_n(\epsilon)$. The data are generated from the labels by a fixed channel, proving the risk lower bound. At $\epsilon = 1/2$ the midpoint estimator attains the bound $D/2$. Combining the three lower bounds proves the lower side of the rate.

7.3 Norm medians with an exact majority term

For the upper bound we may add independent $N(0, I_d)$ noise to every draw and estimate the new variance $t + 1$, then subtract one. This preserves the model and the length of the bracket. Conditional averaging of the clipped estimate over this auxiliary noise can only reduce absolute risk by Jensen's inequality. We thus prove the upper bounds under the convention $a > 0$.

The following norm lemma also applies after projecting an independent estimation block. The random mean U need only satisfy $\|U\| \le R$ almost surely.

Lemma 7.2. *Let N independent observations in $\mathbb{R}^\ell$ come from Huber contamination of $U + \sqrt{t}\,G$, with G standard Gaussian independent of U. Put $\eta = \mathbb{E}\|U\|^2$, and let m_ℓ be the median of χ^2_ℓ. Form each energy $\|X_i\|^2/m_\ell$ and clip to $[a, b]$. Let T be the middle order statistic or the average of the two middle order statistics. For $0 \le \epsilon \le 1/4$,*

$$\mathbb{E}|T - t| \le D\beta_N(\epsilon) + C\left[(N\ell)^{-1/2} + \epsilon/\sqrt{\ell} + \eta/\ell\right].$$

For $1/4 \le \epsilon < 1/2$, the same estimator satisfies

$$\mathbb{E}|T - t| \le D\beta_N(\epsilon) + C\min\{1, \psi(\epsilon)/\sqrt{\ell}\}.$$

The constants depend only on R, a, b.

Proof. Use the latent contamination labels. When $J < N/2$, every middle order statistic lies between the smallest and largest clean observation and between the clean order statistics with ranks $r - J$ and r for its own rank r. At a tie $J = N/2$, the interval between the two middle order statistics intersects the range of the clean observations and at least one of the two endpoints belongs to that range. To see the endpoint assertion, if one endpoint is strictly below and one strictly

above the range, the even sample would have no clean observations between the two adjacent middle ranks. Thus at a tie the loss is at most $D/2$ plus the maximum clean deviation from t. With a strict contaminated majority the loss is at most D. These assertions hold after clipping as well.

For moderate contamination, a uniform interior quantile comparison for the unnormalized clean energies is

$$\left|F^{-1}_{\|U+\sqrt{t}G\|^2}(p) - tm_\ell\right| \le C\left(\eta + \sqrt{\ell}\,|p-1/2|\right) \qquad (c_0 \le p \le 1-c_0) \tag{16}$$

for any fixed $c_0 > 0$, with its own constant. Here are details for large ℓ. The noncentral chi-square law with noncentrality $\|U\|^2/t$ is a Poisson mixture of $\chi^2_{\ell+2j}$. The increment of the CDF from ℓ to $\ell+2$ in absolute value is $(x/2)^{\ell/2}e^{-x/2}/\Gamma(1+\ell/2)$ at x. Its supremum is at most $C/\sqrt{\ell}$ by Stirling's bounds. Telescoping and averaging over the Poisson variable bounds the uniform CDF difference from central $t\chi^2_\ell$ by $C\eta/\sqrt{\ell}$. The central density is bounded below by $c/\sqrt{\ell}$ on any fixed interior quantile interval: Chebyshev's inequality locates that interval within $O(\sqrt{\ell})$ of its mean, and the gamma density formula and Stirling's bounds give the density bound there. The inverse CDF comparison yields (16) for all sufficiently large ℓ, since $\eta \le R^2$. For the finitely many smaller dimensions, the central density has a positive minimum on each fixed interior quantile interval. The same CDF argument applies when η is small. When η exceeds the corresponding fixed positive threshold, bounded clean expectations and Markov's inequality bound the clean interior quantiles by $C\eta$ in absolute difference from tm_ℓ. Also $m_\ell \asymp \ell$.

Choose $\alpha \in (1/4, 1/2)$. Conditional on $J = j \le \alpha N$, each bounding clean order statistic is the clean quantile at a beta order statistic of $N-j$ independent uniforms. Its mean differs from $1/2$ by at most $C(j+1)/N$, and its variance is at most C/N. For all such ranks the means are in a fixed central interval for all sufficiently large N. The probability of leaving a wider fixed central interval is exponentially small in N, by the binomial characterization of uniform order statistics and Hoeffding's inequality. Thus on the central event (16) gives expected loss at most $C[(j/N + N^{-1/2})/\sqrt{\ell} + \eta/\ell]$.

To control exceptional events while retaining the dimension factor, let V be the maximum clipped clean deviation from t (zero when no clean draws occur). For $M \le N$ clean observations, uniformly in the labels, it has L^2 norm at most

$$C\min\{1, \sqrt{\log(2N)/\ell}\}. \tag{17}$$

Indeed for $Z = \|U + \sqrt{t}\,G\|/\sqrt{m_\ell}$, the reverse triangle inequality controls its difference from $\sqrt{t}\|G\|/\sqrt{m_\ell}$ by $R/\sqrt{m_\ell}$. Gaussian norm concentration about its median yields $\mathbb{P}(|Z - \sqrt{t}| > C(1+u)/\sqrt{\ell}) \le 2e^{-cu^2}$. Clipped energy is the square of Z clipped to $[\sqrt{a}, \sqrt{b}]$, and hence its deviation is bounded by $2\sqrt{b}\,|Z - \sqrt{t}|$ as well as by D. A union bound and tail integration prove (17). For $\epsilon \le 1/4$, both $J > \alpha N$ and the conditional beta exceptional events have exponentially small probability. On strict clean majority and on ties the remaining loss beyond the bound already allocated to β_N is at most V. Cauchy–Schwarz and (17) bound this expected residual by $C/\sqrt{N\ell}$. Averaging j/N and treating bounded N by (17) proves the first assertion.

For $\epsilon \ge 1/4$ put $\delta = 1 - 2\epsilon$ and $h = C\sqrt{1 + \log(1/\delta)}/\sqrt{\ell}$, with C large. The clean norm concentration just used implies that the probability of a clean label with clipped deviation at most h is at least $1/2 + c\delta$. For example, bound the conditional tail by $\delta/4$ and use $1 - \epsilon = (1+\delta)/2$. Hoeffding's inequality shows that a strict majority of such good observations fails with probability at most $e^{-c'N\delta^2}$. On the good event both middle statistics are within h of t. On failure, the extra loss beyond the bound already allocated to β_N is at most V whenever there is no strict contaminated majority. If $\delta \le N^{-1/4}$, then (17) alone bounds that contribution by $C\min(1, h)$. If $\delta > N^{-1/4}$, Cauchy–Schwarz and $N\delta^2 > \sqrt{N}$ give the same bound. Clipping the bound on the good event and recalling $\epsilon \ge 1/4$ proves the second assertion. □

7.4 Removing latent energy in moderate dimension

If $d \geq n$, apply Lemma 7.2 to all raw observations with $\ell = d$ and $\eta \leq R^2$. Then $R^2/d \leq R^2/\sqrt{nd}$. The same full-sample estimate gives the high-contamination bound in every dimension. It remains to prove the rate for $k \leq d < n$ and $\epsilon \leq 1/4$.

On a pilot block of size comparable to n, fit a clean bounded k-component isotropic mixture in smoothed halfspace distance, the supremum over projected scalar CDF gaps after adding a scalar standard Gaussian. The empirical distance to the true clean law has expectation $O(\sqrt{d/n} + \epsilon)$, by the halfspace VC bound and the Huber mixture form. A fit to accuracy $n^{-1/2}$ over a countable dense parameter set is measurable and is within twice the empirical error plus its tolerance from the true clean law. The scalar Gaussian ODE bound of Lemma 4.3 controls the gap in the first two moments of every smoothed projection by a constant times this distance. The second-moment identity after removing the known smoothing variance gives the same bound for clean raw second moments, uniformly in every unit direction. In particular, with $M = \mathbb{E}UU^\top$, the fitted raw second moment $\widehat{S}$ satisfies

$$\mathbb{E}\|\widehat{S} - (M + tI_d)\|_{\mathrm{op}} \leq C(\sqrt{d/n} + \epsilon).$$

The centered covariance bound follows as well from the first moment comparison and the bounded mean norm.

If $d \leq 2k$, the smallest eigenvalue of the fitted centered covariance, clipped to $[a, b]$, gives $O(n^{-1/2} + \epsilon)$ risk by Weyl's inequality and the rank bound for the centered latent covariance. This matches the desired rate with constants depending on k. If $d > 2k$, let Q project onto the orthogonal complement of the top k eigenvectors of $\widehat{S}$, so that $\ell = d - k \geq d/2$. Write $\zeta = \|\widehat{S} - (M + tI_d)\|_{\mathrm{op}}$. Since $\operatorname{rank} M \leq k$, Weyl's inequality gives $\lambda_{k+1}(\widehat{S}) \leq t + \zeta$. Every unit v in the range of Q then satisfies $v^\top M v \leq 2\zeta$. Since QMQ has rank at most k, the residual latent energy satisfies $\eta = \mathbb{E}[\|QU\|^2 \mid Q] = \operatorname{tr}(QMQ) \leq 2k\zeta$.

On the held-out block, project onto this range and apply the first part of Lemma 7.2 conditionally, with N comparable to n. The held-out majority term is $O(e^{-cn})$, which is bounded by $C/\sqrt{nd}$ when $d < n$. Furthermore

$$\mathbb{E}\eta/\ell \leq C\left[(nd)^{-1/2} + \epsilon/d\right].$$

This proves the remaining upper bound. Bounded sample sizes are covered by a constant estimate when $d < n$.

For adaptation to unknown ϵ when $d < n$, let T_{fit} be the preceding pilot-based estimator and T_{raw} the full-sample clipped norm median. Project T_{fit} onto the closed interval

$$[T_{\mathrm{raw}} - C_0R^2/d, T_{\mathrm{raw}} + C_0R^2/d] \cap [a, b],$$

where C_0 is a constant such that the term $C\eta/m_d$ of the pointwise central quantile bound in (16), after normalization by m_d, satisfies $C\eta/m_d \leq C_0R^2/d$ for all $d \geq 1$. Projection increases the error by at most $(|T_{\mathrm{raw}} - t| - C_0R^2/d)_+$. For $\epsilon \leq 1/4$, the proof of Lemma 7.2 bounds the expectation of this positive part by $D\beta_n + C[(nd)^{-1/2} + \epsilon/\sqrt{d}]$: the subtraction removes the η/d term on the central event, and the other events have the same bound as before. For $\epsilon \geq 1/4$ the projected estimate is within C_0R^2/d of the raw one. That difference is absorbed by the logarithmic term and preserves the coefficient of β_n. When $d \geq n$ use the raw estimate directly.

7.5 Exact large-dimension limits for power loss

Let $\mathcal{R}_{n,d,p}$ denote the minimax expected loss $|\widehat{t} - t|^p$, for fixed $1 \le p < \infty$. For $0 < \epsilon < 1/2$ write

$$\pi_j = \frac{1}{1 + (\epsilon/(1-\epsilon))^{2j-n}}, \qquad g_p(j) = \frac{\pi_j^{1/(p-1)}}{\pi_j^{1/(p-1)} + (1-\pi_j)^{1/(p-1)}} \quad (p > 1).$$

For $p = 1$, take $g_1(j) = 0, \frac{1}{2}, 1$ according as $j < n/2, j = n/2, j > n/2$. At $\epsilon = 0$ use the limits. These functions are nondecreasing and satisfy $g_p(n-j) = 1 - g_p(j)$.

Theorem 7.3. *For $d \ge k$, $n \ge 1$, and $0 \le \epsilon < 1/2$, the p-th root minimax risk lies between*

$$D(\mathbb{E} g_p(J)^p)^{1/p} \quad \textit{and} \quad D(\mathbb{E} g_p(J)^p)^{1/p} + C_p \min\{1, \sqrt{\log(2n)/d}\}.$$

The upper value is attained by an estimator that uses ϵ (with ϵ known); for $p = 1$ it is the norm median, which does not. Thus the lower expression is the exact fixed-n, ϵ large-dimension limit.

Proof. The swapped endpoint experiment gives the lower bound. For equal prior weights, π_j is the posterior on the upper endpoint given j upper labels. The Bayes action on the normalized unit interval is $g_p(j)$, by minimizing posterior expected power loss. Under the lower endpoint j has law J, and symmetry makes the two endpoint risks equal.

For the upper bound let $S_{(1)} \ge \cdots \ge S_{(n)}$ be descending clipped raw norm energies normalized to $[0, 1]$, with true normalized variance $s = (t - a)/D$. Use the normalized estimate

$$g_p(0) + \sum_{i=1}^{n} (g_p(i) - g_p(i-1)) S_{(i)}.$$

First replace all clean values by s, with the bad values unchanged. Given $J = j$, monotonicity in each coordinate bounds the resulting estimate between the values from putting all bad coordinates at zero or all at one. The latter value equals $g_p(j) + s(g_p(n) - g_p(j)) = s + (1-s)g_p(j) - s g_p(0)$, so it exceeds s by at most $g_p(j)$. Reflection, using $g_p(n-i) = 1 - g_p(i)$, gives the same bound on the deviation below s in the former case. Thus the absolute error is at most $g_p(j)$.

Sorting is 1-Lipschitz for maximum coordinate distance, and the total weight is at most one. Undoing the clean replacement changes the unnormalized estimate by at most the maximum clipped clean deviation. The union bound used in (17) also bounds its L^p norm by $C_p \min\{1, \sqrt{\log(2n)/d}\}$. Minkowski's inequality yields the claim. □

7.6 The majority transition

For any fixed $0 < \epsilon < 1/2$, let $I_\epsilon = -\frac{1}{2}\log(4\epsilon(1-\epsilon))$. The ratio of successive binomial terms past $n/2$ is at most $\epsilon/(1-\epsilon) < 1$; Stirling's formula for the first term, with the half tie when n is even, gives $\beta_n(\epsilon) \asymp_\epsilon n^{-1/2} e^{-nI_\epsilon}$. Consequently for fixed ϵ the two leading effects are comparable when d has order ne^{2nI_ϵ}. When $d/(ne^{2nI_\epsilon}) \to \infty$, Theorem 7.1 gives the asymptotically exact risk $(1 + o(1))D\beta_n(\epsilon)$. The rate statement also applies as ϵ tends to one half, where the contamination modulus has size $\min\{1, \sqrt{\log(1/(1-2\epsilon))/d}\}$. This distinction between the population modulus and the finite-sample label experiment persists in the submodel with a single centered Gaussian.

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] M. Chen, C. Gao, and Z. Ren. A general decision theory for Huber's ϵ-contamination model. *Electronic Journal of Statistics*, 10:3752–3774, 2016. https://doi.org/10.1214/16-EJS1216.

[2] 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/v/4.

[3] I. Diakonikolas, D. M. Kane, and A. Stewart. Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures. *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science*, 73–84, 2017. https://doi.org/10.1109/FOCS.2017.16. Full version https://arxiv.org/abs/1611.03473.

[4] N. Verzelen and E. Arias-Castro. Detection and feature selection in sparse mixture models. *Annals of Statistics*, 45(5):1920–1950, 2017. https://doi.org/10.1214/16-AOS1513.

[5] Y. Wu and P. Yang. Optimal estimation of Gaussian mixtures via denoised method of moments. *Annals of Statistics*, 48(4):1981–2007, 2020. https://doi.org/10.1214/19-AOS1873.

[6] N. Doss, Y. Wu, P. Yang, and H. H. Zhou. Optimal estimation of high-dimensional Gaussian location mixtures. *Annals of Statistics*, 51(1):62–95, 2023. https://doi.org/10.1214/22-AOS2207. Appendices B and D of https://arxiv.org/abs/2002.05818v2.

[7] H. He and G. Cheng. Sharp proper estimation of fixed-component Gaussian location mixtures in polynomial time. arXiv:2608.12701, 2026. https://arxiv.org/abs/2608.12701.

[8] N. Bariletto, D. Le, A. Rinaldo, and N. Ho. Convergence rates for latent mixing measures in infinite homoscedastic location-scale mixture models. arXiv:2605.06845v2, 2026. https://arxiv.org/abs/2605.06845v2.

[9] D. Agostini, C. Améndola, and K. Ranestad. Moment identifiability of homoscedastic Gaussian mixtures. *Foundations of Computational Mathematics*, 21:695–724, 2021. https://doi.org/10.1007/s10208-020-09469-6.

[10] D. Hsu and S. M. Kakade. Learning mixtures of spherical Gaussians: moment methods and spectral decompositions. *ITCS*, 11–20, 2013. https://arxiv.org/abs/1206.5766.

[11] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society B*, 61(3):611–622, 1999. https://doi.org/10.1111/1467-9868.00196.

[12] D. Passemier, Z. Li, and J.-F. Yao. On estimation of the noise variance in high-dimensional probabilistic principal component analysis. *Journal of the Royal Statistical Society B*, 79(1):51–67, 2017. https://doi.org/10.1111/rssb.12153.

[13] R. Tang, Y. Han, and A. R. Zhang. Statistical and computational limits for cumulant tensor inference: A regime in which efficient detection is harder than efficient estimation. arXiv:2603.26029v4, 2026. https://arxiv.org/abs/2603.26029v4.

[14] S. Bonaccorsi, B. Hanzon, and G. Lombardi. A generalized Budan–Fourier approach to generalized Gaussian and exponential mixtures. *AIMS Mathematics*, 9(10):26499–26537, 2024. https://doi.org/10.3934/math.20241290.

[15] J. Chen and P. Li. Hypothesis test for normal mixture models: the EM approach. *Annals of Statistics*, 37(5A):2523–2542, 2009. https://arxiv.org/abs/0908.3428.

[16] A. Bultheel, R. Cruz-Barroso, and M. Van Barel. On Gauss-type quadrature formulas with prescribed nodes anywhere on the real line. *Calcolo*, 47:21–48, 2010. Article and erratum.

[17] R. E. Curto and P. J. di Dio. Time-dependent moments from the heat equation and a transport equation. *International Mathematics Research Notices*, 2023(17):14955–14990. https://doi.org/10.1093/imrn/rnac244.

[18] H. M. Möller. Kubaturformeln mit minimaler Knotenzahl. *Numerische Mathematik*, 25:185–200, 1976. https://doi.org/10.1007/BF01462272.

[19] R. Cools and P. Verlinden. On cubature formulae of degree $4k + 1$ attaining Möller's lower bound for integrals with circular symmetry. *Numerische Mathematik*, 61(3):395–408, 1992.

[20] R. Cools and H. J. Schmid. A new lower bound for the number of nodes in cubature formulae of degree $4n + 1$ for some circularly symmetric integrals. In H. Brass and G. Hämmerlin, editors, *Numerical Integration IV*, International Series of Numerical Mathematics 112, 57–66. Birkhäuser, Basel, 1993.

[21] A. H. Stroud. *Approximate Calculation of Multiple Integrals*. Prentice-Hall, 1971, p. 324, formula $E_2^{r^2}$ 7-1.

[22] M. Chen, C. Gao, and Z. Ren. Robust covariance and scatter matrix estimation under Huber's contamination model. *Annals of Statistics* 46:1932–1960, 2018. https://arxiv.org/abs/1506.00691.

[23] T. Peng and M. Neykov. Minimax rates for robust Gaussian mean estimation near the breakdown point under constraints. 2026. https://arxiv.org/abs/2610.36283.

[24] A. Minasyan and N. Zhivotovskiy. Statistically optimal robust mean and covariance estimation for anisotropic Gaussians. *Mathematical Statistics and Learning* 8:33–69, 2025. https://doi.org/10.4171/MSL/48.

[25] P. Ruckdeschel. Higher order asymptotics for the MSE of the sample median on shrinking neighborhoods. https://arxiv.org/abs/1006.0045.