Sharp transport identification with unknown Gaussian covariance

Scott Schubr
Luminary Oddities
scottschubr@luminaryoddities.com

October 31, 2026

Abstract

Unknown Gaussian covariance creates an ambiguity between latent variation and noise. We quantify how covariance information controls this ambiguity in quadratic Wasserstein distance. On a fixed two-sided potential-curvature class, the sharp identification modulus is the latent-covariance uncertainty plus an inherited inverse logarithm of order $3/2$. With n primary observations and m independent pure-noise observations, the sharp minimax rate replaces that uncertainty by its minimum with $m^{-1/2}$. Retained independent-error pairs remove the covariance prior and recover the known-noise order, even with unrestricted unknown latent location. For rough log-concave inputs, a Gaussian-divisible reference gives linear covariance dependence and an exact compatible-law set. A bounded-variation heat estimate gives covariance dependence of order $3/4$ for arbitrary smooth references, with sharp logarithmic testing under exact normalization. Finally, we allow unknown error shapes in a Wasserstein neighborhood of a Gaussian while preserving their covariance exactly. When the permitted noise covariance exceeds the unit latent scale, we prove a matching logarithmic shape-uncertainty floor and sharp primary-sample minimax risk. Positive Gaussian completion and covariance alignment transfer an established scalar-noise inverse theorem; exact-moment oscillations and convolution allocation supply the lower bounds.

Keywords: Gaussian deconvolution; partial identification; covariance calibration; Wasserstein distance; minimax estimation.
2020 Mathematics Subject Classification: 62G07, 62G20; 62C20, 60E15.

1 Introduction and information regimes

An observed Gaussian convolution does not by itself distinguish variation in the latent distribution from variation in the noise. For example, increasing a Gaussian latent covariance by tI and decreasing the noise covariance by the same amount preserves the observation law. We quantify how three sources of covariance information resolve this ambiguity: an approximate latent-covariance prior, independent pure-noise calibration observations, and two retained noisy repetitions of each latent sample. The loss throughout is quadratic Wasserstein distance between latent laws.

The resulting rates separate the cost of covariance uncertainty from the inverse-logarithmic cost of Gaussian smoothing. The main analytic inverse estimate is inherited from the known-noise theory; the present proofs transfer it to these three information regimes and give matching lower bounds. The estimators establish minimax rates through measurable minimum-distance constructions.

For a symmetric nonnegative matrix A, let γ_A be the centered Gaussian probability measure with covariance A, including the point mass when $A = 0$. Thus the parameter is a covariance, not a standard deviation. We write W_p for the usual Wasserstein distance with Euclidean cost, and $\|\cdot\|$ and $\|\cdot\|_{\mathrm{F}}$ for operator and Frobenius norms. All dimensions and constants below are fixed.

Fix $d \geq 1$, $S > 0$ and $0 < c < 1 < C < \infty$. Define

$$\mathcal{F}_\zeta = \Big\{ f = e^{-V} : \int f = 1, \quad \mathbb{E}_f X = 0, \quad V \in C^2(\mathbb{R}^d), \\ cI \preceq D^2 V \preceq CI, \quad \|\mathrm{Cov}(f) - I\|_{\mathrm{F}} \leq \zeta \Big\},$$

$$\mathcal{K}_S = \{0 \preceq A \preceq SI\}.$$

Normalization is incorporated in V. The observation law is $f*\gamma_\Sigma$, with $\Sigma \in \mathcal{K}_S$ unknown. For $0 < \delta < 1$, put $\ell(\delta) = [\log(1/\delta)]^{-3/2}$ and set $\ell(0) = 0$.

Theorem 1.1 (Pairwise partial identification)**.** *There are positive k, K, ζ_0, δ_0, depending only on d, c, C, S, such that for $0 \le \zeta \le \zeta_0$ and $0 \le \delta \le \delta_0$,*

$$k[\zeta + \ell(\delta)] \le \sup\left\{W_2(f,g) : \ f,g \in \mathcal{F}_\zeta,\ \Sigma,\Lambda \in \mathcal{K}_S,\ W_1(f*\gamma_\Sigma, g*\gamma_\Lambda) \le \delta\right\} \le K[\zeta + \ell(\delta)].$$

In particular, exact covariance normalization gives identifiability even when the Gaussian noise covariance is unknown. At identical outputs the sharp ambiguity order under covariance uncertainty ζ is linear in ζ. More precisely, every pair in the supremum satisfies

$$W_2(f,g) \le K\left[\min\{\zeta, \|\Sigma - \Lambda\|_{\mathrm{F}}\} + \ell(\delta)\right]. \tag{1}$$

There are two independent obstructions in this formula: exact Gaussian allocation gives the linear covariance term, while Gaussian smoothing gives the inverse-logarithmic term. The latter exponent and its hard analytic proof come from the known-noise theorem of [15]. Our central question is which additional covariance information removes the former obstruction.

Theorem 1.2 (Sharp calibration-sample tradeoff)**.** *In addition to n independent observations from $f*\gamma_\Sigma$, observe $m \ge 1$ independent samples from γ_Σ, independent of the primary observations. For $r = 1, 2$, all sufficiently large n, every $m \ge 1$, and $0 \le \zeta \le \zeta_0$,*

$$\inf_{\widehat{f}} \sup_{f\in\mathcal{F}_\zeta,\ \Sigma\in\mathcal{K}_S} \mathbb{E}_{f,\Sigma} W_2(\widehat{f}, f)^r \asymp \left[\min\{\zeta, m^{-1/2}\} + (\log n)^{-3/2}\right]^r. \tag{2}$$

The constants are uniform in m and ζ. Without calibration observations the same statement has ζ in place of the minimum.

Theorem 1.2 is proved in Section 5. It gives the precise criterion $\min\{\zeta, m^{-1/2}\} = O((\log n)^{-3/2})$ for preserving the known-noise minimax order. In the regime $\zeta/(\log n)^{-3/2} \to \infty$, this is equivalent to $m = \Omega((\log n)^3)$. If the covariance prior is already at that resolution, calibration samples are unnecessary for this rate.

The known latent covariance is information about the model. It cannot be obtained by whitening noisy data without also estimating the noise. If it is removed, the Gaussian pairs

$$(X,Z) \sim N(0,I) \otimes N(0,sI), \qquad (X',Z') \sim N(0,(1+t)I) \otimes N(0,(s-t)I)$$

produce the same output whenever $0 < t < s$.

Our third information regime retains two noisy repetitions of each latent sample. Let $0 < a < 1 < b < \infty$ be fixed and define

$$\mathcal{H}_{a,b} = \{f = e^{-V} : \int f = 1,\ \mathbb{E}_f X = 0,\ V \in C^2,\ cI \le D^2 V \le CI,\ aI \le \mathrm{Cov}(f) \le bI\}.$$

Write $Q_{f,\Sigma_1,\Sigma_2} = \mathcal{L}(X + Z_1, X + Z_2)$, where $X \sim f$ and the two centered Gaussian errors are mutually independent and independent of X, with unknown $\Sigma_j \in \mathcal{K}_S$. The paired observation space has its Euclidean norm on $\mathbb{R}^{2d}$.

Theorem 1.3 (Paired full identification and sharp estimation)**.** *Uniformly over $f, g \in \mathcal{H}_{a,b}$ and all permitted noise matrices, for sufficiently small $\delta \ge 0$,*

$$W_1(Q_{f,\Sigma_1,\Sigma_2}, Q_{g,\Lambda_1,\Lambda_2}) \le \delta \quad \Longrightarrow \quad W_2(f,g) \le K\ell(\delta).$$

From n independent paired observations, for $r = 1, 2$ and all sufficiently large n,

$$\inf_{\widehat{f}} \sup_{\substack{f\in\mathcal{H}_{a,b}\\ \Sigma_1,\Sigma_2\in\mathcal{K}_S}} \mathbb{E}_{f,\Sigma_1,\Sigma_2} W_2(\widehat{f}, f)^r \asymp (\log n)^{-3r/2}. \tag{3}$$

All constants depend only on the displayed fixed class parameters. The upper estimator may take values in $\mathcal{H}_{a,b}$.

Corollary 1.4 (Paired recovery with unrestricted latent location)**.** *Let $\mathcal{H}_{\mathrm{loc}}$ be all probability densities $f = e^{-V}$ with $V \in C^2$ and $cI \preceq D^2V \preceq CI$, without a prescribed mean or covariance. The pairwise inverse and minimax rate in Theorem 1.3 hold over $\mathcal{H}_{\mathrm{loc}}$, with constants depending only on d, c, C, S. An upper estimator takes values in this class and has mean*

$$\overline{M}_n = \frac{1}{2n}\sum_{i=1}^{n}(Y_{i1} + Y_{i2}).$$

The two errors remain centered and mutually independent.

Thus neither the latent mean nor a separate covariance interval is required for paired recovery. Curvature supplies the covariance bounds; mean anchoring pays the unbounded location class, including exceptional samples.

Section 6 proves the paired results by recovering latent covariance from the cross-covariance of the retained observations. The noise matrices may be unequal, noncommuting or singular.

A complementary result concerns a fixed reference and a larger input class. Suppose

$$\begin{gathered}\rho = \nu * \gamma_{\tau I}, \quad \tau > 0, \quad \nu = Z^{-1}e^{-U}, \\ \mathbb{E}_\nu X = 0, \quad \mathrm{Cov}(\rho) = I, \quad U \in C^2(\mathbb{R}^d), \\ \alpha I \preceq D^2U \preceq \beta I.\end{gathered} \tag{4}$$

Here $0 < \alpha \le \beta < \infty$. Necessarily $0 < \tau < 1$. Let

$$\mathcal{L}_\zeta = \{f : \ f \text{ is a centered log-concave density on } \mathbb{R}^d, \ \|\mathrm{Cov}(f) - I\|_{\mathrm{F}} \le \zeta\}.$$

No differentiability or positive curvature is required of f.

Theorem 1.5 (A Gaussian-divisible fixed reference)**.** *Under (4), for all sufficiently small $\zeta, \delta \ge 0$,*

$$\sup_{\substack{f \in \mathcal{L}_\zeta,\ \Sigma,\Lambda \in \mathcal{K}_S \\ W_1(f*\gamma_\Sigma, \rho*\gamma_\Lambda) \le \delta}} W_2(f,\rho) \asymp \zeta + \ell(\delta),$$

with constants depending only on the fixed reference, d and S. Moreover, if $q = \rho * \gamma_{\Lambda_0}$, $\Lambda_0 \in \mathcal{K}_S$, and ζ is sufficiently small, its compatible inputs are exactly

$$\left\{\nu * \gamma_{\tau I - D} : \ D = D^{\mathsf{T}}, \ \|D\|_{\mathrm{F}} \le \zeta, \ 0 \le \Lambda_0 + D \le SI\right\}. \tag{5}$$

Theorem 1.1 compares two members of the curvature class. The fixed-reference theorem uses the positive Gaussian component in (4) to allow rough log-concave inputs. The table summarizes the three full-density estimation regimes; fixed-reference testing retains its separate displayed assumptions.

Information	Latent class	Minimax W_2 estimation order
Single noisy sample of size n	$\mathcal{F}_\zeta$: two-sided curvature and covariance prior	$\zeta + (\log n)^{-3/2}$
The same sample and m pure-noise observations	The same $\mathcal{F}_\zeta$	$\min\{\zeta, m^{-1/2}\} + (\log n)^{-3/2}$
n retained independent-error pairs	$\mathcal{H}_{\mathrm{loc}}$: two-sided curvature, unrestricted latent location	$(\log n)^{-3/2}$

The minimum-distance procedures establish measurable estimators in fixed dimension.

1.1 Relation to earlier work

Unknown-variance deconvolution and Gaussian-allocation ambiguity are classical. Matias [13] identifies scalar Gaussian variance by excluding a Gaussian component from the latent density, with losses for variance and pointwise density estimation. Our approximately normalized class permits that allocation and quantifies its sharp transport floor. Gassiat, Le Corff and Lehéricy [11] obtain identification from dependent signal blocks and independent noise blocks under a Laplace-growth condition below order two. This is a different information contract from the covariance calibration used here.

Pure-noise auxiliary samples have a substantial deconvolution literature. Johannes [12] treats the primary-plus-auxiliary experiment with nonparametric ordinary or supersmooth error distributions, Sobolev losses and Fourier source conditions. His results already separate primary-sample and noise-sample errors. Here the noise belongs to a finite-dimensional Gaussian covariance family; the shape-constrained W_2 loss and prior covariance information give the sharp minimum in (2). Scricciolo [16] studies auxiliary-sample deconvolution under ordinary-smooth errors in univariate W_1 loss. These comparisons identify the model-specific calibration contribution.

Repeated measurements likewise provide classical identification. Delaigle, Hall and Meister [10] show that replicated observations can preserve known-error first-order performance in substantial deconvolution classes. Their density-loss results establish the broader statistical principle; here the curvature-constrained transport loss and Gaussian covariance family determine the precise rate. Capitao-Miniconi, Gassiat and Lehéricy [7] treat unknown errors through repeated observations and finite-order Laplace-transform conditions. Their repeated-measurement condition allows arbitrary finite positive growth order. Section 6 uses the Gaussian cross-covariance identity to prove sharp full smooth-class W_2 estimation and a testing law against rough log-concave inputs with bounded covariance.

The difficult inverse exponent $3/2$, reference-uniform constants and exact-moment alternatives are inherited from [15]. The present contribution is the sharp calibrated partial-identification law, its two-sample resource tradeoff, paired smooth-class estimation, and the rough-input fixed-reference extensions. Bobkov, Chistyakov and Götze [3] and Courtade, Fathi and Pananjady [9] provide related information-distance and entropy-power stability results with different observation functionals. Scricciolo [17, Theorem 2] obtains supersmooth logarithmic rates in univariate W_1 deconvolution under known errors and dependent signals. That Sobolev-type distribution-estimation result has different loss, regularity and information assumptions from the paired unknown-covariance W_2 experiment here.

2 Analytic tools and the scalar-noise input

We first state precisely the earlier result used below. The isotropic log-concave class consists of all centered log-concave densities with covariance I, including densities with nonsmooth convex potential or bounded support.

Proposition 2.1 (Scalar-noise inverse and alternatives, [15])**.** *Fix a dimension and scalar Gaussian variance $B > 0$. For isotropic references g whose C^2 potentials have common positive lower and finite upper Hessian bounds, there are common constants $K, \varepsilon_0 > 0$ such that every isotropic log-concave f satisfies*

$$W_2(f,g) \le K\ell\big(W_1(f * \gamma_{BI}, g * \gamma_{BI})\big)$$

when the output discrepancy is positive and at most ε_0. At zero discrepancy $f = g$. At every fixed permitted reference g there are centered isotropic log-concave f_T, for every sufficiently large real T, with

$$\begin{aligned} W_2(f_T, g) &\ge kT^{-3}, \\ W_1(f_T * \gamma_{BI}, g * \gamma_{BI}) &\le KT^{-2}e^{-aT^2}, \\ \chi^2(f_T * \gamma_{BI} \mid g * \gamma_{BI}) &\le KT^{-4}e^{-2aT^2}. \end{aligned} \tag{6}$$

At a standard Gaussian reference the alternatives can be chosen in any fixed curvature interval $0 < c < 1 < C$ by reducing their fixed amplitude.

For clarity, the alternatives in [15] are actual positive densities, not signed formal perturbations. They add a narrow Gaussian times $T^{-2}[\cos(Tx_1) - b_T\cos(x_1) + c_T]$, with b_T, c_T chosen to cancel mass and covariance exactly. Uniform curvature bounds control the perturbation, and the Gaussian envelope gives the chi-square bound in (6). We use the theorem and its statistical alternatives as established inputs; all covariance transfers needed in the present paper are proved below.

Lemma 2.2 (Moments and covariance interpolation)**.** *Centered log-concave laws with covariance bounded above have uniformly bounded fourth moments in fixed dimension. More generally, for any two centered probability laws p, q with a common fourth-moment bound and $\delta = W_1(p, q) \in (0, 1)$,*

$$\|\mathrm{Cov}(p) - \mathrm{Cov}(q)\|_{\mathrm{F}} \le K\delta^{2/3}. \tag{7}$$

This interpolation estimate uses only the fourth-moment bound and applies to singular laws as well. The uniform moment bound is preserved after adding Gaussian noise with uniformly bounded covariance.

Proof. For a centered one-dimensional marginal with variance at most v^2, Chebyshev gives upper-tail probabilities at least $3/4$ at $-2v$ and at most $1/4$ at $2v$. The survival function is log-concave, by the marginal/integration property of log-concave laws. Its logarithm therefore has a uniform negative secant slope, which extends to the right by concavity. If that tail is already zero, the conclusion is immediate. Applying the same argument to the negative marginal gives exponential tails on both sides. Integration and the coordinate inequality $|x|^4 \le d\sum_i |x_i|^4$ prove the fourth-moment bound. Gaussian addition preserves it by $|x+y|^4 \le 8(|x|^4 + |y|^4)$. For (7), multiply $x_i x_j$ by a smooth cutoff equal to one on the radius-R ball and zero outside the radius-$2R$ ball. Its Lipschitz constant is $O(R)$; the omitted second moment is $O(R^{-2})$. Kantorovich duality bounds each entry difference by $K(R\delta + R^{-2})$. Set $R = \delta^{-1/3}$ and sum the finitely many entries. □

Lemma 2.3 (Curvature and forward Gaussian displacement)**.** *Suppose $f = e^{-V}$ is centered and $cI \le D^2V \le CI$. Then*

$$C^{-1}I \le \mathrm{Cov}(f) \le c^{-1}I, \qquad I(f) := \int |\nabla \log f|^2 f \le dC.$$

If $A \succeq 0$, $f_A = f * \gamma_A$ has potential Hessian between $c(1 + c\|A\|)^{-1}I$ and CI, and

$$W_2(f, f_A) \le \tfrac{1}{2}\sqrt{dC}\,\|A\|. \tag{8}$$

Proof. The Gaussian tails and linear gradient growth justify integration by parts: $\mathbb{E}\nabla V = 0$, $\mathbb{E}[X\nabla V^{\mathsf{T}}] = I$, and $\mathbb{E}[\nabla V \nabla V^{\mathsf{T}}] = \mathbb{E}D^2V \le CI$. Directional Cauchy–Schwarz gives the covariance lower bound. The upper is Brascamp–Lieb [6]; the Fisher bound follows by taking the trace. For $A \succ 0$, the posterior curvature is at least $cI + A^{-1}$. Brascamp–Lieb and

$$D^2(-\log f_A) = A^{-1} - A^{-1}\,\mathrm{Cov}(X \mid X + Z)A^{-1}$$

give the stronger lower bound $(c^{-1}I + A)^{-1}$. The alternative identity $D^2(-\log f_A) = \mathbb{E}[D^2V \mid X + Z] - \mathrm{Cov}(\nabla V \mid X + Z)$ gives the upper. Singular A follows by replacing it with $A + \epsilon I$ and taking the C^2 limit; the first two density derivatives are integrable by the stated envelopes. Finally $f_t = f * \gamma_{tA}$ solves the continuity equation with velocity $-A\nabla \log f_t/2$. Conditional Jensen gives $I(f_t) \le I(f)$. Its transport action on $0 \le t \le 1$ is at most $dC\|A\|^2/4$, so the dynamic Wasserstein bound [2] proves (8). Positive regularization treats singular A. □

Lemma 2.4 (Rough log-concave heat displacement)**.** *For centered log-concave densities whose covariance eigenvalues lie in a fixed positive compact interval, in fixed dimension,*

$$W_2(f, f * \gamma_A) \le K\|A\|^{3/4}, \qquad A \succeq 0.$$

The exponent $3/4$ is optimal for a uniform forward-displacement bound on this class.

Bounds relating bounded-variation convolutions to Fisher information have established precedents [4, 5]. The proof below combines Fisher convexity with an interval-level-set decomposition and a Gaussian-specific interval estimate.

Proof. For a unit vector v, let $V_v(f)$ be the integral of the one-dimensional total variations of the slices $t \mapsto f(z + tv)$ over $z \in v^\perp$. A fixed-dimensional isotropic log-concave density satisfies $f(x) \le C_d e^{-c_d|x|}$ [1, Lemma 2(b)]. Covariance normalization on the stated spectral interval preserves such a uniform envelope. Since every slice is unimodal,

$$V_v(f) = 2\int_{v^\perp} \operatorname*{ess\,sup}_t f(z+tv)\,dz \le K,$$

uniformly in v. Boundary jumps are included in the total variation.

Write ϕ_s for the one-dimensional $N(0,s)$ density and $p_s = \mathbf{1}_{[a,b]} * \phi_s$. Its score is $s^{-1}(\mathbb{E}[Y \mid a \le Y \le b] - x)$ for $Y \sim N(x,s)$. With $c_0 = \sqrt{2/\pi}$,

$$|p_s'(x)|/p_s(x) \le \frac{\operatorname{dist}(x,[a,b])}{s} + \frac{c_0}{\sqrt{s}}.$$

Indeed, for $x \in [a,b]$, monotonicity of a truncated mean in each endpoint compares it with the positive and negative half-normal means. For $x < a$, the conditional overshoot beyond a is bounded by its mean conditional on $Y \ge a$. Its density on $u \ge 0$ is proportional to $e^{-u^2/(2s)}e^{-(a-x)u/s}$, a decreasing tilt of the half-normal density, so that mean is at most $c_0\sqrt{s}$. Reflection treats $x > b$. This argument is uniform in the interval length.

Put $q_s(x) = \phi_s(x-a) + \phi_s(x-b)$. Then

$$\int q_s = 2, \qquad \int \operatorname{dist}(x,[a,b])q_s(x)\,dx \le 2c_0\sqrt{s},$$

since the distance to the interval is at most the distance to either endpoint. Combining $|p_s'| \le q_s$ with the score bound gives the unnormalized Fisher bound

$$\int \frac{|p_s'|^2}{p_s} \le 4c_0 s^{-1/2}.$$

For any integrable nonnegative unimodal slice g, use $g = \int_0^\infty \mathbf{1}_{\{g>u\}}\,du$. Its nonempty positive level sets are bounded intervals almost everywhere. Cauchy–Schwarz in the level parameter and Tonelli give

$$I(g*\phi_s) \le Cs^{-1/2}\int_0^\infty \operatorname{TV}(\mathbf{1}_{\{g>u\}})\,du = Cs^{-1/2}\operatorname{TV}(g).$$

Both Fisher information and variation here are homogeneous in the unnormalized slice mass. Integrating transversely proves

$$I_v(f * \gamma_{svv^\mathsf{T}}) \le CV_v(f)s^{-1/2}.$$

The directional heat equation has continuity velocity $-v\partial_v \log f_s/2$. Its dynamic transport length from ϵ to t is therefore at most

$$\frac{\sqrt{CV_v(f)}}{2}\int_\epsilon^t s^{-1/4}\,ds \le \frac{2}{3}\sqrt{CV_v(f)}\,t^{3/4}.$$

The same-Gaussian coupling lets $\epsilon \downarrow 0$. For transverse regularity one may convolve this curve with a full Gaussian of variance ε: directional Fisher information contracts under convolution by conditional Jensen, so the same bound holds for the smooth curve. Finite-second-moment coupling then lets $\varepsilon \downarrow 0$.

For every positive semidefinite B, distributional differentiation gives $D_v(f * \gamma_B) = (D_v f) * \gamma_B$, hence $V_v(f * \gamma_B) \le V_v(f)$ by contraction of the total variation of a signed measure under probability convolution. Diagonalize $A = \sum_j a_j v_j v_j^\mathsf{T}$ and add these covariances successively. Log-concavity, hence

slice unimodality, is preserved; directional variation contracts under each convolution. The triangle inequality gives $K\sum_j a_j^{3/4} \le Kd\|A\|^{3/4}$. Zero eigenvalues cause no loss.

For optimality take the isotropic uniform density on $[-\sqrt{3},\sqrt{3}]$. Its convolution assigns a density bounded below on the interval $[\sqrt{3}+\sqrt{t},\sqrt{3}+2\sqrt{t}]$ for all small t. Any coupling from the original supported law pays at least the squared distance of the convolved endpoint from the support. Integration over this interval gives $W_2^2 \ge kt^{3/2}$, proving the matching forward lower bound. □

Lemma 2.5 (Covariance matching and Gaussian coupling)**.** *On any fixed positive spectral interval, matrix square roots and inverse square roots are Frobenius-Lipschitz. In particular, if C_f, C_g lie in that interval and $L = C_g^{1/2}C_f^{-1/2}$, then*

$$LC_fL^{\mathsf{T}} = C_g, \quad \|L-I\|_{\mathrm{F}} \le K\|C_f - C_g\|_{\mathrm{F}}, \quad W_2(f, L_{\#}f) \le K\|C_f - C_g\|_{\mathrm{F}}$$

for a centered law f with covariance C_f. If $A, D \succeq aI$, $a > 0$, then

$$W_1(\gamma_A, \gamma_D) \le \|A^{1/2} - D^{1/2}\|_{\mathrm{F}} \le \frac{\|A-D\|_{\mathrm{F}}}{2\sqrt{a}}. \tag{9}$$

Proof. For $X = A^{1/2} - D^{1/2}$ the Sylvester identity $A^{1/2}X + XD^{1/2} = A - D$ holds without commutation. In left and right orthonormal eigenbases every entry is divided by a sum of square roots at least $2\sqrt{a}$. This proves the square-root estimate. The identity for inverses gives the analogous inverse-square-root bound on a compact positive interval. Also $L - I = (C_g^{1/2} - C_f^{1/2})C_f^{-1/2}$. The same-input coupling proves the transport bound for L, and coupling the same standard Gaussian proves (9). □

Lemma 2.6 (Known scalar noise without isotropy)**.** *For centered densities with common two-sided positive potential-curvature bounds and fixed scalar noise BI, the inverse upper in Proposition 2.1 holds uniformly without requiring isotropy.*

Proof. Put $\delta = W_1(f * \gamma_{BI}, g * \gamma_{BI})$. The moments are uniform by Lemmas 2.2–2.3, so $\|C_f - C_g\|_{\mathrm{F}} = O(\delta^{2/3})$. Replace f by $L_{\#}f$, $L = C_g^{1/2}C_f^{-1/2}$, paying $O(\delta^{2/3})$ in W_2 and therefore in output W_1. Apply the common map $C_g^{-1/2}$. Both inputs become isotropic and their potential curvatures remain in one fixed positive interval. The common noise covariance becomes $BC_g^{-1} \preceq BCI$. Convolve both outputs with the same Gaussian of covariance $(BC+1)I - BC_g^{-1}$. The discrepancy only decreases, and the noise is now the exact fixed scalar noise in Proposition 2.1. Undo the uniformly bounded affine map and pay the matching cost. Since $\ell(K\delta^{2/3}) \le K'\ell(\delta)$ for sufficiently small δ, the result follows. At zero discrepancy the covariances agree and Gaussian injectivity gives $f = g$. □

3 Pairwise identification under covariance uncertainty

Proof of Theorem 1.1: upper bound. Write $p = f * \gamma_\Sigma$, $q = g * \gamma_\Lambda$, and $D = \Sigma - \Lambda$. By Lemma 2.2,

$$\|D\|_{\mathrm{F}} \le 2\zeta + K\delta^{2/3}, \qquad \delta = W_1(p, q). \tag{10}$$

Let D_+, D_- be the spectral positive and negative parts, so $D = D_+ - D_-$. Put $B = S + 1$. The matrix

$$K_0 = BI - \Sigma + D_+ = BI - \Lambda + D_- \succeq I \tag{11}$$

is the same positive covariance in both formulas, even for noncommuting Σ, Λ. Thus

$$p * \gamma_{K_0} = (f * \gamma_{D_+}) * \gamma_{BI}, \qquad q * \gamma_{K_0} = (g * \gamma_{D_-}) * \gamma_{BI}.$$

The new output discrepancy is at most δ. Since $\|D_\pm\| \le S$, Lemma 2.3 puts both smoothed inputs in common curvature bounds. Lemma 2.6 bounds their W_2 distance by $K\ell(\delta)$. Returning to f, g by (8) costs at most $K\|D\|_{\mathrm{F}}$. This directly gives $K[\|D\|_{\mathrm{F}} + \ell(\delta)]$; alternatively use (10) and absorb $\delta^{2/3}$ into $\ell(\delta)$. Taking the better bound proves both the stated upper and (1). If $\delta = 0$, the two smoothed laws agree by Gaussian injectivity and the same displacement argument gives (1) directly. □

Matching lower bounds. Choose $0 < s < S$ and $t = \zeta/\sqrt{d}$. For small ζ, the centered Gaussian priors of covariances I and $(1+t)I$ belong to $\mathcal{F}_\zeta$. Noise covariances sI and $(s-t)I$ give identical outputs. Their W_2 distance is $\sqrt{d}(\sqrt{1+t}-1) \geq k\zeta$. For the logarithmic term, use (6) at the standard Gaussian reference with a common fixed noise sI. Both priors have covariance exactly I and lie in $\mathcal{F}_\zeta$ after reducing the amplitude. For every small δ, choosing real T of order $\sqrt{\log(1/\delta)}$ gives the logarithmic lower. The maximum of these lower bounds pays half their sum. □

4 Gaussian-divisible anchors and exact compatible laws

Theorem 1.5: upper bound. Write $p = f * \gamma_\Sigma$, $q = \rho * \gamma_\Lambda$ and $D = \Sigma - \Lambda$. For $\zeta \leq 1/2$, both priors have bounded covariance, so Lemma 2.2 gives

$$\|D\|_{\mathrm{F}} \leq \zeta + K\delta^{2/3}, \qquad \delta = W_1(p,q).$$

Choose the thresholds so this is at most $\min(\tau/2, 1/4)$, and define the actual probability density

$$\rho_D = \nu * \gamma_{\tau I - D}.$$

Its Gaussian covariance lies between $\tau I/2$ and $3\tau I/2$. Lemma 2.3 gives common lower and upper curvature bounds for ρ_D. Moreover

$$\rho_D * \gamma_\Sigma = \rho * \gamma_\Lambda, \qquad \mathrm{Cov}(f) - \mathrm{Cov}(\rho_D) = \mathrm{Cov}(p) - \mathrm{Cov}(q). \tag{12}$$

Convolve both observed laws with $\gamma_{BI-\Sigma}$, $B = S + 1$. They now have the common scalar noise BI and latent laws f, ρ_D, with output discrepancy at most δ. The covariance difference in (12) is $O(\delta^{2/3})$, without a ζ loss.

Let $C_D = I - D$, $C_f = \mathrm{Cov}(f)$, and replace f by $L_\# f$, $L = C_D^{1/2} C_f^{-1/2}$. This costs $O(\delta^{2/3})$ by Lemma 2.5. Apply $C_D^{-1/2}$ to both laws. Both inputs are isotropic; $L_\# f$ remains log-concave and the reference keeps uniform two-sided curvature. The common noise is $BC_D^{-1} \preceq 2BI$. Adding the same Gaussian covariance $(2B+1)I - BC_D^{-1}$ now gives the exact fixed scalar-noise model of Proposition 2.1. Its full log-concave input scope is used here, rather than Lemma 2.6. Undoing the maps yields $W_2(f, \rho_D) \leq K\ell(\delta)$. Finally couple the common ν variable and a standard Gaussian to get

$$W_2(\rho_D, \rho) \leq \|(\tau I - D)^{1/2} - \sqrt{\tau} I\|_{\mathrm{F}} \leq K_\tau \|D\|_{\mathrm{F}}.$$

This proves the upper, without a heat-displacement estimate for the rough input f. □

Exact set and lower bounds. If $f * \gamma_\Sigma = \rho * \gamma_{\Lambda_0}$, covariance equality gives $D = \Sigma - \Lambda_0 = I - \mathrm{Cov}(f)$, so $\|D\|_{\mathrm{F}} \leq \zeta$ and $\tau I - D > 0$ for small ζ. Dividing the nonvanishing Gaussian characteristic multiplier identifies f with $\nu * \gamma_{\tau I - D}$. Conversely every matrix in (5) gives an actual centered log-concave density with covariance $I - D$ and precisely the required output. This proves (5); no arbitrary backward Gaussian transform is declared positive. For the ambiguity lower take $f = \rho * \gamma_{tI}$, $t = \zeta/\sqrt{d}$, and noise covariances $(s-t)I, sI$, $0 < s < S$. Outputs agree exactly. The reverse triangle inequality in L^2 for any coupling gives

$$W_2(f,\rho) \geq \left|\sqrt{\mathbb{E}_f |X|^2} - \sqrt{\mathbb{E}_\rho |X|^2}\right| = \sqrt{d}(\sqrt{1+t} - 1) \geq k\zeta.$$

The alternatives (6) at the fixed reference ρ remain isotropic log-concave, hence lie in every $\mathcal{L}_\zeta$. They supply the logarithmic lower at every small output discrepancy. The maximum pays half the sum. At zero discrepancy the exact set and the Gaussian coupling above give the upper $K\zeta$ directly. □

4.1 Arbitrary smooth references under exact covariance normalization

The positive-completion argument also gives a rough-input bound for every smooth reference in the stated curvature class. It uses forward smoothing of the rough law, rather than a backward Gaussian transform of the reference.

Proposition 4.1 (General-reference robustness)**.** *Let ρ be centered and isotropic with C^2 potential satisfying $cI \le D^2(-\log\rho) \le CI$. Uniformly over $f \in \mathcal{L}_\zeta$ and $\Sigma, \Lambda \in \mathcal{K}_S$, for sufficiently small $\zeta, \delta \ge 0$,*

$$W_1(f * \gamma_\Sigma, \rho * \gamma_\Lambda) \le \delta \quad \Longrightarrow \quad W_2(f, \rho) \le K\big[\zeta^{3/4} + \ell(\delta)\big].$$

In particular, when $\zeta = 0$, the sharp inverse-logarithmic order holds without a Gaussian-divisibility assumption on ρ.

Proof. Put $D = \Sigma - \Lambda$ and $E = \mathrm{Cov}(f * \gamma_\Sigma) - \mathrm{Cov}(\rho * \gamma_\Lambda)$. Lemma 2.2 gives $\|E\|_{\mathrm{F}} \le K\delta^{2/3}$ and $\|D\|_{\mathrm{F}} \le \zeta + K\delta^{2/3}$. With the positive and negative spectral parts of D, set

$$a = f * \gamma_{D_+}, \qquad b = \rho * \gamma_{D_-}, \qquad B = S + 1.$$

The common positive completion in (11) gives $W_1(a * \gamma_{BI}, b * \gamma_{BI}) \le \delta$ and $\mathrm{Cov}(a) - \mathrm{Cov}(b) = E$. The law a is log-concave, while b has common two-sided curvature bounds by Lemma 2.3. Their covariances lie in a fixed positive spectral interval.

Write $C_a = \mathrm{Cov}(a)$, $C_b = \mathrm{Cov}(b)$ and $L = C_b^{1/2} C_a^{-1/2}$. Lemma 2.5 bounds $W_2(a, L_\# a)$ by $K\delta^{2/3}$. After applying $C_b^{-1/2}$ to $L_\# a$ and b, both laws are isotropic. Their common noise covariance BC_b^{-1} has a fixed upper bound; adding the same positive Gaussian complement makes that noise one fixed scalar covariance. The first law remains log-concave and the reference retains uniform curvature bounds. Proposition 2.1 therefore gives

$$W_2(a, b) \le K\ell(\delta + K\delta^{2/3}) \le K'\ell(\delta)$$

for positive small δ, including the covariance-matching cost. At $\delta = 0$, the completed laws agree by Gaussian injectivity.

Lemma 2.4 and smooth-reference displacement give

$$W_2(f, a) \le K\|D\|_{\mathrm{F}}^{3/4}, \qquad W_2(\rho, b) \le K\|D\|_{\mathrm{F}}.$$

The triangle inequality proves the claim, since $\delta^{1/2}$ is absorbed into $\ell(\delta)$. These couplings also prove the zero-discrepancy case directly. □

For positive covariance uncertainty the general-reference bound is sufficient; its covariance exponent $3/4$ is not established as optimal. Exact Gaussian divisibility supplies the stronger linear dependence and explicit compatible-law set in Theorem 1.5.

5 Statistical consequences

Throughout this section $Y_1, \dots, Y_n$ are independent observations from the stated noisy law, and $Q_n = n^{-1}\sum_i \delta_{Y_i}$.

Lemma 5.1 (A sufficient empirical estimate)**.** *For laws Q with a common second-moment bound and $b = 1/[4(d+1)]$,*

$$\mathbb{E}W_1(Q_n, Q) \le Kn^{-b}, \qquad \Pr\{W_1(Q_n, Q) > n^{-b/2}\} \le Kn^{-b/2}.$$

Proof. Move the part outside $[-R,R]^d$ to zero and quantize the rest at mesh η, including zero as a site. The expected cost is at most $K\eta + K/R$. The grid has $O((1+R/\eta)^d)$ sites and diameter $O(R)$. Cauchy–Schwarz and the empirical Bernoulli variances bound its expected total variation error by half the square root of its size divided by n. Thus

$$\mathbb{E}W_1(Q_n, Q) \le K[\eta + R^{-1} + R\sqrt{(1+R/\eta)^d/n}].$$

Take $R = n^b$, $\eta = n^{-b}$. The last term is $O(n^{-1/4})$, and Markov proves the probability bound. This deliberately elementary estimate suffices for a logarithmic inverse rate. □

Theorem 5.2 (Estimation with approximately known latent covariance)**.** *For $r = 1, 2$, every sequence $0 \le \zeta_n \le \zeta_0$, and all sufficiently large n,*

$$\inf_{\widehat{f}_n} \sup_{f \in \mathcal{F}_{\zeta_n},\, \Sigma \in \mathcal{K}_S} \mathbb{E}_{f,\Sigma} W_2(\widehat{f}_n, f)^r \asymp [\zeta_n + (\log n)^{-3/2}]^r.$$

The constants are uniform in the sequence. The infimum is over measurable probability-valued estimators, and an upper estimator may take values in $\mathcal{F}_{\zeta_n}$.

Proof. For each n choose a countable dense subset of $\mathcal{F}_{\zeta_n} \times \mathcal{K}_S$ in W_2 times Frobenius distance. Such a subset exists because the ambient spaces are separable. Gaussian convolution is continuous into W_1, including singular covariances, by same-Gaussian square-root coupling. Select the first joint candidate whose output W_1 distance from Q_n is within $\eta_n = n^{-b/2}$ of the countable infimum. The distances are continuous in the observations, so selection is measurable. Density gives

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

On the good event in Lemma 5.1, Theorem 1.1 bounds the input error by $K[\zeta_n + (\log n)^{-3/2}]$. The complement has polynomially small probability. The common centered covariance bound gives a uniform deterministic W_2 diameter by independent coupling, proving both upper risks. For a lower of order ζ_n^r, use the identical-output Gaussian pair in the proof of Theorem 1.1. For the logarithmic lower use (6) with common fixed noise and $T^2 = A\log n$, A sufficiently large. Then $n\chi^2 \to 0$ and product total variation tends to zero. The two-point estimation inequality gives a lower $k(\log n)^{-3r/2}$. Their maximum is comparable to the stated sum to power r. □

Proof of Theorem 1.2. For $m \ge 1$ let $\widehat{\Sigma}$ be the Frobenius projection of $m^{-1}\sum_{j=1}^m Z_j Z_j^{\mathsf{T}}$ onto $\mathcal{K}_S$. The Gaussian fourth-moment identity and nonexpansiveness of the projection give, uniformly for all permitted covariances,

$$\mathbb{E}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}^2 \le K/m. \tag{13}$$

Indeed $\sum_{i,j}\mathrm{Var}(Z_iZ_j) = (\mathrm{tr}\,\Sigma)^2 + \mathrm{tr}(\Sigma^2) \le K$ for one observation. This formula covers singular covariances.

Set $b = 1/[4(d+1)]$, $\eta_n = n^{-b/2}$ and $A_n = \{W_1(Q_n, f * \gamma_\Sigma) \le \eta_n\}$. By Lemma 5.1, $\Pr(A_n^c) \le Kn^{-b/2}$. Choose a deterministic countable dense list $(f_k, \Sigma_k) \in \mathcal{F}_\zeta \times \mathcal{K}_S$. Among indices satisfying

$$W_1(f_k * \gamma_{\Sigma_k}, Q_n) < 2\eta_n, \tag{14}$$

let d_n be the infimum of $\|\Sigma_k - \widehat{\Sigma}\|_{\mathrm{F}}$ and select the first eligible index for which this distance is less than $d_n + \eta_n$. If the eligible set is empty, return the standard Gaussian. All choices are measurable because the list is countable and each displayed distance is measurable. On A_n, density provides eligible approximants to the true parameter, so $d_n \le \|\Sigma - \widehat{\Sigma}\|_{\mathrm{F}}$. The selected pair then has

$$\begin{aligned} W_1(f_k * \gamma_{\Sigma_k}, f * \gamma_\Sigma) &< 3\eta_n, \\ \|\Sigma_k - \Sigma\|_{\mathrm{F}} &\le 2\|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}} + \eta_n. \end{aligned} \tag{15}$$

Apply (1) and $\min\{\zeta, 2x + \eta_n\} \le 2\min\{\zeta, x\} + \eta_n$. On A_n the resulting error is at most

$$K\left[\min\{\zeta, \|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}\} + (\log n)^{-3/2}\right].$$

For $\epsilon = \|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}$ and $r = 1, 2$, $\mathbb{E}\min\{\zeta, \epsilon\}^r \le \min\{\zeta^r, \mathbb{E}\epsilon^r\}$. Equation (13) and Cauchy–Schwarz therefore give $\mathbb{E}\min\{\zeta, \epsilon\}^r \le K\min\{\zeta, m^{-1/2}\}^r$. On A_n^c all returned laws remain in $\mathcal{F}_\zeta$ and have uniformly bounded centered covariance, hence uniformly bounded W_2 diameter. The polynomial exceptional probability is smaller than both logarithmic risk rates. This proves the upper in (2) uniformly in m, ζ.

For the calibration lower, fix $0 < s < S$ and put $t = a\min\{\zeta, m^{-1/2}\}$, for a sufficiently small fixed $a > 0$. Use latent Gaussians with covariances $I, (1+t)I$ and noise covariances $sI, (s-t)I$. The priors obey all curvature and covariance conditions and $0 \le t \le s/2$ after reducing a and ζ_0. Their primary output laws are identical. For one pure-noise observation the Kullback–Leibler divergence is

$$\frac{d}{2}\left[\frac{s}{s-t} - 1 - \log\frac{s}{s-t}\right] \le Kt^2.$$

Thus the divergence for m calibration observations is at most $Kmt^2 \le Ka^2$. Choose a so Pinsker's inequality bounds their total variation by $1/2$. Their latent W_2 separation is at least kt, so the two-point estimation inequality gives the lower $k\min\{\zeta, m^{-1/2}\}^r$.

For the logarithmic lower use (6) at the standard Gaussian, with the same fixed noise sI in both hypotheses. All calibration laws then coincide. Taking $T^2 = A\log n$, A sufficiently large, makes the product primary total variation vanish and gives the lower $k(\log n)^{-3r/2}$. The maximum of the two lower bounds is comparable to the sum in (2) to power r. When $\zeta = 0$ the calibration lower is zero, and the logarithmic lower still applies. Without calibration observations Theorem 5.2 gives the asserted result. □

Theorem 5.3 (Blind-noise fixed-reference testing)**.** *Under* (4)*, define*

$$\mathcal{T}_n(r;\zeta) = \inf_{\psi}\left[\sup_{\Lambda\in\mathcal{K}_S}\Pr_{\rho,\Lambda}(\psi = 1) + \sup_{\substack{f\in\mathcal{L}_\zeta,\ \Sigma\in\mathcal{K}_S\\ W_2(f,\rho)\ge r}}\Pr_{f,\Sigma}(\psi = 0)\right].$$

There are fixed $0 < k < K$ such that, for every sufficiently small known sequence $\zeta_n \ge 0$ and $r_n = \zeta_n + (\log n)^{-3/2}$,

$$\mathcal{T}_n(Kr_n;\zeta_n) \longrightarrow 0, \qquad \mathcal{T}_n(kr_n;\zeta_n) \longrightarrow 1.$$

Proof. Reject when $\inf_{\Lambda\in\mathcal{K}_S} W_1(Q_n, \rho * \gamma_\Lambda) > 2\eta_n$, $\eta_n = n^{-b/2}$. Compactness of $\mathcal{K}_S$ and continuity of convolution make the statistic measurable. Under the null the empirical estimate controls rejection. On the good event for the true output, acceptance provides a permitted reference output within $3\eta_n$ of it. Theorem 1.5 then forces $W_2(f,\rho) \le K_0 r_n$. A strictly larger separation constant makes both errors tend uniformly to zero. For the lower, when $\zeta_n \ge (\log n)^{-3/2}$ the identical-output alternative from Theorem 1.5 has separation at least kr_n for a sufficiently small fixed k, and the two errors sum to one. Otherwise use the common-noise oscillatory alternative with $T^2 = A\log n$. Its product total variation tends to zero and its separation is at least kr_n. A common small constant treats alternating cases. Constant tests give total error at most one, proving the limit. □

Corollary 5.4 (Exactly normalized rough-input testing)**.** *For the reference in Proposition 4.1, define $\mathcal{T}_n(r;0)$ as in Theorem 5.3, with alternatives in $\mathcal{L}_0$. There are constants $0 < k < K$ such that*

$$\mathcal{T}_n\big(K(\log n)^{-3/2};0\big) \longrightarrow 0, \qquad \mathcal{T}_n\big(k(\log n)^{-3/2};0\big) \longrightarrow 1.$$

Proof. For the upper use the same empirical-distance test and good event as in Theorem 5.3, replacing its anchor bound by Proposition 4.1 at $\zeta = 0$. For the lower, Proposition 2.1 supplies exact-isotropic log-concave alternatives at this reference, with common scalar noise sI, $0 < s < S$. Taking $T^2 = A\log n$ with A sufficiently large makes the product total variation tend to zero, while the latent separation is at least $k(\log n)^{-3/2}$. A constant test bounds the risk above by one. □

Corollary 5.5 (Operational covariance completion under exact isotropy)**.** *If $\zeta = 0$, let $\widehat{\Sigma}$ be the Frobenius projection of $n^{-1}\sum_i Y_i Y_i^\top - I$ onto $\mathcal{K}_S$, and $B = S + 1$. Then*

$$\mathbb{E}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}^2 \le K/n.$$

The completed empirical probability measure $\widetilde{Q}_n = Q_n * \gamma_{BI-\widehat{\Sigma}}$ obeys, pathwise,

$$W_1(\widetilde{Q}_n, f * \gamma_{BI}) \le W_1(Q_n, f * \gamma_\Sigma) + \tfrac{1}{2}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}.$$

Proof. Projection onto the closed convex covariance set is nonexpansive; summing entry variances and using the fourth-moment bound gives the first estimate. For the second, first convolve both empirical and true outputs with the same estimated complementary Gaussian. This contracts W_1 even though the kernel depends on the data. Compare its true-output convolution with the ideal completion using (9); both complementary covariances are at least I. □

6 Paired measurements as covariance calibration

For every permitted paired law,

$$\mathrm{Cov}(Q_{f,\Sigma_1,\Sigma_2}) = \begin{pmatrix} C_f + \Sigma_1 & C_f \\ C_f & C_f + \Sigma_2 \end{pmatrix}, \qquad C_f = \mathrm{Cov}(f). \tag{16}$$

Independence makes the off-diagonal block equal to the latent covariance. This identity supplies the calibration step in both proofs below.

6.1 Full smooth-class estimation

Proof of Theorem 1.3. Put $C_f = \mathrm{Cov}(f)$ and $C_g = \mathrm{Cov}(g)$. Uniform fourth moments and Lemma 2.2, applied to the joint laws in dimension $2d$, bound their covariance difference by $K\delta^{2/3}$. By (16), their off-diagonal blocks are respectively C_f and C_g, hence $\|C_f - C_g\|_{\mathrm{F}} \le K\delta^{2/3}$. Use the same analysis-only linear map $L = C_g^{-1/2}$ on both inputs and first-channel outputs. Then $\|L\| \le a^{-1/2}$, $\|L^{-1}\| \le \sqrt{b}$, $\mathrm{Cov}(L_\# g) = I$, and

$$\|\mathrm{Cov}(L_\# f) - I\|_{\mathrm{F}} \le K a^{-1}\delta^{2/3}.$$

The transformed potentials have Hessians between caI and CbI; their noise covariances lie between zero and $(S/a)I$. First-marginal output W_1 is at most $a^{-1/2}\delta$. Thus Theorem 1.1, with these fixed enlarged class constants, gives

$$W_2(L_\# f, L_\# g) \le K[\delta^{2/3} + \ell(a^{-1/2}\delta)] \le K'\ell(\delta).$$

Mapping back proves the pairwise assertion. At $\delta = 0$, equality of joint covariances gives $C_f = C_g$ and $\Sigma_1 = \Lambda_1$. Equality of first-marginal convolutions and the nonvanishing Gaussian characteristic function give $f = g$ directly.

For estimation choose a deterministic countable dense list in $\mathcal{H}_{a,b} \times \mathcal{K}_S^2$ under W_2 and Frobenius metrics. This space is a subspace of a separable metric space and contains the standard Gaussian with zero noise. The map to joint outputs is W_1-continuous: couple latent variables in W_2 and use common independent standard Gaussians for both errors. Continuity of matrix square roots includes singular covariances. Let Q_n be the empirical paired law and $h = 1/[4(2d+1)]$, $\eta_n = n^{-h/2}$. Select the first candidate whose W_1 distance to Q_n is within η_n of the countable infimum. The index is measurable, and density gives

$$W_1(Q_{\widehat{\theta}}, Q_\theta) \le 2W_1(Q_n, Q_\theta) + \eta_n.$$

Lemma 5.1 in dimension $2d$ bounds the probability of $W_1(Q_n, Q_\theta) > \eta_n$ by $Kn^{-h/2}$. On its complement, the pairwise bound gives error at most $K(\log n)^{-3/2}$. On the exceptional event, independent coupling

gives $W_2(\widehat{f}, f)^2 \le 2db$. Its polynomially small probability is negligible for either asserted risk, proving the upper bound.

For the lower fix $0 < s < S$ and restrict both noise matrices to sI. The invertible transformation

$$(Y_1, Y_2) \longmapsto ((Y_1 + Y_2)/2, Y_1 - Y_2)$$

gives independent laws $f * \gamma_{(s/2)I}$ and γ_{2sI}. The latter is common to all parameters, so the chi-square divergence equals that of the first coordinate. Use the exact-isotropic alternatives of Proposition 2.1 at the standard Gaussian reference and variance $s/2$. They belong to the fixed curvature interval and have covariance $I \in [aI, bI]$. Their separation is at least kT^{-3} and their one-observation chi-square divergence is at most $KT^{-4}e^{-2AT^2}$ for a fixed $A > 0$. Taking $T^2 = B \log n$ with sufficiently large B makes product total variation tend to zero. The two-point estimation bound proves the lower bound for each $r = 1, 2$. □

Proof of Corollary 1.4. Write $m_f = \mathbb{E}_f X$. Curvature gives $C^{-1}I \le \mathrm{Cov}(f) \le c^{-1}I$ independently of location. The centered laws therefore lie in $\mathcal{H}_{a,b}$ for the fixed choices $a = 1/(2C)$ and $b = 2/c$. If paired output laws have W_1 distance δ, their means differ by at most δ, so $\sqrt{2}|m_f - m_g| \le \delta$. Center each law by its own mean. The centered paired distance is at most 2δ. Applying Theorem 1.3 to these centered laws and then translating back gives

$$W_2(f, g) \le |m_f - m_g| + K\ell(2\delta) \le K'\ell(\delta)$$

for positive small δ. At zero discrepancy the means and centered laws agree, giving equality directly.

For estimation let $e_n = W_1(Q_n, Q)$ in the paired space and $b_* = 1/[4(2d + 1)]$. Translating theoretically by (m_f, m_f) preserves this distance and gives a common centered second-moment bound. Lemma 5.1 therefore yields $\sup \mathbb{E}e_n \le Kn^{-b_*}$ uniformly over every latent mean. Set $\eta_n = n^{-b_*/2}$. Choose a countable dense list in $\mathcal{H}_{\mathrm{loc}} \times \mathcal{K}_S^2$ under W_2 and Frobenius metrics. As in the theorem, the joint output map is continuous, including singular error covariances. Select the first parameter whose output distance to Q_n is strictly less than the countable infimum plus η_n; denote its latent law by f_* and joint output by Q_*. This is measurable and

$$W_1(Q_*, Q_n) \le e_n + \eta_n, \qquad W_1(Q_*, Q) \le 2e_n + \eta_n.$$

The linear map $(y_1, y_2) \mapsto (y_1 + y_2)/2$ has Lipschitz constant $1/\sqrt{2}$. Its expectation under Q_* is m_{f_*}, and under Q_n is $\overline{M}_n$. Hence

$$\sqrt{2}|m_{f_*} - \overline{M}_n| \le e_n + \eta_n.$$

Translate f_* by $\overline{M}_n - m_{f_*}$, preserving curvature and its selected noise matrices, and call the resulting law $\widehat{f}_n$. On each measurable selection set this translation is continuous in the observed mean; it is a Borel W_2-valued estimator. The paired output translates in both coordinates, so its discrepancy from the true output is at most $3e_n + 2\eta_n$. On $\{e_n \le \eta_n\}$ the pairwise inverse gives $W_2(\widehat{f}_n, f) \le K(\log n)^{-3/2}$.

The full location class has no bounded W_2 diameter. Instead, independent coupling of the centered versions gives, pathwise,

$$W_2(\widehat{f}_n, f)^2 \le |\overline{M}_n - m_f|^2 + 2d/c.$$

By error independence,

$$\mathbb{E}|\overline{M}_n - m_f|^2 = \frac{1}{n}\,\mathrm{tr}\left(\mathrm{Cov}(f) + \frac{\Sigma_1 + \Sigma_2}{4}\right) \le \frac{d/c + dS/2}{n}.$$

The exceptional probability is at most $Kn^{-b_*/2}$. Thus its squared-loss contribution is bounded by $K/n + (2d/c)Kn^{-b_*/2}$, and its unsquared contribution is also polynomially small by $W_2 \le |\overline{M}_n - m_f| + \sqrt{2d/c}$. This proves both upper risks uniformly over every location. The centered subclass already used in Theorem 1.3 supplies both lower bounds. Mean anchoring is a minimum-distance construction, not an efficient computational prescription. □

6.2 Rough-input fixed-reference testing

Let ρ be any centered isotropic density whose C^2 potential has fixed two-sided positive Hessian bounds. Define

$$\mathcal{G} = \{f : \ f \text{ is a centered log-concave density}, \quad \tfrac{1}{2}I \preceq \mathrm{Cov}(f) \preceq \tfrac{3}{2}I\}.$$

The endpoints can be replaced by any fixed positive bounds containing I. For $\Sigma_1, \Sigma_2 \in \mathcal{K}_S$, write

$$Q_{f,\Sigma_1,\Sigma_2} = \mathcal{L}(X + Z_1, X + Z_2), \tag{17}$$

where $X \sim f$ and the two Gaussian errors are mutually independent and independent of X. Their coordinates may be correlated and their covariances may be unequal or singular. The joint pairing is retained.

Theorem 6.1 (Paired fixed-reference inverse and sharp testing)**.** *For all sufficiently small $\delta \geq 0$,*

$$W_1(Q_{f,\Sigma_1,\Sigma_2}, Q_{\rho,\Lambda_1,\Lambda_2}) \leq \delta \quad \Longrightarrow \quad W_2(f,\rho) \leq K\ell(\delta) \tag{18}$$

uniformly over $f \in \mathcal{G}$ and permitted noise covariances. Here W_1 on paired laws uses the Euclidean norm on $\mathbb{R}^{2d}$. For n independent paired observations define the sum-of-errors testing risk

$$\mathcal{P}_n(r) = \inf_{\chi} \left[\sup_{\Lambda_1,\Lambda_2 \in \mathcal{K}_S} \Pr_{\rho,\Lambda_1,\Lambda_2}(\chi = 1) + \sup_{\substack{f \in \mathcal{G},\ \Sigma_1,\Sigma_2 \in \mathcal{K}_S \\ W_2(f,\rho) \geq r}} \Pr_{f,\Sigma_1,\Sigma_2}(\chi = 0) \right].$$

There are constants $0 < k < K$ such that

$$\mathcal{P}_n\big(K(\log n)^{-3/2}\big) \longrightarrow 0, \qquad \mathcal{P}_n\big(k(\log n)^{-3/2}\big) \longrightarrow 1. \tag{19}$$

Proof. All joint laws have uniform fourth moments, by Lemma 2.2. Truncated quadratic tests on $\mathbb{R}^{2d}$ therefore bound their full covariance difference by $K\delta^{2/3}$. Comparing the cross block and then the first diagonal block in (16) with the reference law gives

$$\|C_f - I\|_{\mathrm{F}} \leq K\delta^{2/3}, \qquad \|D\|_{\mathrm{F}} \leq K\delta^{2/3}, \quad D = \Sigma_1 - \Lambda_1. \tag{20}$$

The first marginals have W_1 discrepancy at most δ. Put $B = S + 1$ and convolve both with the same positive Gaussian of covariance $BI - \Sigma_1 \succeq I$. They become $f * \gamma_{BI}$ and $\rho * \gamma_{BI-D}$; the discrepancy contracts. Both Gaussian covariances are at least I, so Lemma 2.5 and (20) imply

$$W_1(f * \gamma_{BI}, \rho * \gamma_{BI}) \leq K\delta^{2/3}.$$

Let $L = C_f^{-1/2}$. Its coupling cost satisfies $W_2(f, L_\# f) \leq K\delta^{2/3}$. The law $L_\# f$ is exactly isotropic and log-concave. Replacing f by this law in the common scalar-noise comparison costs the same amount in output W_1. Proposition 2.1 therefore gives $W_2(L_\# f, \rho) \leq K\ell(K\delta^{2/3}) \leq K'\ell(\delta)$. The triangle inequality proves (18). At $\delta = 0$, (20) gives $C_f = I$ and $\Sigma_1 = \Lambda_1$; first-marginal Gaussian injectivity gives $f = \rho$ directly.

For the testing upper, apply Lemma 5.1 in dimension $2d$ to the empirical paired law. With the corresponding polynomial tolerance η_n, reject when its W_1 distance from $\{Q_{\rho,\Lambda_1,\Lambda_2} : \Lambda_1, \Lambda_2 \in \mathcal{K}_S\}$ exceeds $2\eta_n$. This compactly parametrized family is continuous under same-Gaussian coupling, including singular covariances, so the statistic is measurable. On the good empirical event, acceptance places the true paired law within $3\eta_n$ of some null law. Equation (18) then gives $W_2(f,\rho) \leq K_0(\log n)^{-3/2}$. A larger constant proves the upper, and the same event controls the null error.

For the lower, fix $0 < s < S$ and use $\Sigma_1 = \Sigma_2 = sI$ under both hypotheses. The invertible map

$$(Y_1, Y_2) \mapsto \big((Y_1 + Y_2)/2, Y_1 - Y_2\big)$$

gives independent coordinates with laws $f * \gamma_{(s/2)I}$ and γ_{2sI}. The second law is the same under both hypotheses. Thus the paired chi-square divergence equals that of the first coordinate. Use the every-reference alternatives in (6), now with variance $s/2$, and take $T^2 = A \log n$ for large fixed A. Their product total variation tends to zero and their latent separation is at least $k(\log n)^{-3/2}$. The testing lower bound and a constant test prove the second limit in (19). □

Only observing the separate marginal laws is a different experiment. If all channel covariances have a common positive floor tI, then replacing f by $f * \gamma_{uI}$ and every covariance Σ_j by $\Sigma_j - uI$, $0 < u < t$, preserves every marginal output. The paired cross-covariance in (20) removes precisely this allocation ambiguity.

7 Covariance-preserving departures from Gaussian noise

Covariance information and Gaussian shape are separate assumptions. We now quantify the latter through a neighborhood that preserves the error covariance exactly. This section concerns independent latent and error variables and primary observations only.

Misspecification of the error law has long been studied in deconvolution. Meister [14] analyzes its effect on density-estimator mean integrated squared error, while Caillerie et al. [8, Section 3] give a Wasserstein residual-smoothing bound for one-sided Gaussian variance misspecification. The error-shape result below instead imposes exact covariance matching and a W_1 neighborhood of a Gaussian error, and identifies the sharp latent W_2 ambiguity and primary-sample minimax risk.

For $\Sigma \in \mathcal{K}_S$, define

$$\mathcal{E}_\eta(\Sigma) = \{E : E \text{ is a centered probability law}, \quad \mathrm{Cov}(E) = \Sigma, \quad W_1(E, \gamma_\Sigma) \le \eta\}.$$

Let $\mathfrak{M}(\zeta, \delta, \eta)$ be the supremum of $W_2(f, g)$ over $f, g \in \mathcal{F}_\zeta$, $\Sigma, \Lambda \in \mathcal{K}_S$, $E \in \mathcal{E}_\eta(\Sigma)$ and $F \in \mathcal{E}_\eta(\Lambda)$ such that $W_1(f * E, g * F) \le \delta$.

Theorem 7.1 (Sharp shape-misspecification modulus)**.** *Assume $S > 1$. For all sufficiently small $\zeta, \delta, \eta \ge 0$,*

$$k[\zeta + \ell(\delta + \eta)] \le \mathfrak{M}(\zeta, \delta, \eta) \le K[\zeta + \ell(\delta + \eta)].$$

The logarithmic η lower bound already holds at $\zeta = \delta = 0$, with the same exact error covariance in both experiments and with smooth log-concave errors having common two-sided curvature bounds.

Proof. Same-input coupling gives

$$W_1(f * \gamma_\Sigma, g * \gamma_\Lambda) \le W_1(f * \gamma_\Sigma, f * E) + W_1(f * E, g * F) + W_1(g * F, g * \gamma_\Lambda) \le \delta + 2\eta.$$

Theorem 1.1 proves the upper, since $\ell(\delta + 2\eta) \le K\ell(\delta + \eta)$ at small positive arguments. Its zero-discrepancy clause treats $\delta = \eta = 0$.

For the shape lower, fix $0 < s < S - 1$. Proposition 2.1 at the standard Gaussian reference supplies $f_T \in \mathcal{F}_0$ such that

$$W_2(f_T, \gamma_I) \ge kT^{-3}, \qquad W_1(f_T * \gamma_{sI}, \gamma_{(1+s)I}) \le KT^{-2}e^{-aT^2}.$$

Compare the two latent/error pairs

$$(\gamma_I,\ f_T * \gamma_{sI}), \qquad (f_T,\ \gamma_{(1+s)I}).$$

Both errors are centered with covariance $(1 + s)I \in \mathcal{K}_S$. Their common primary output is explicitly

$$\gamma_I * (f_T * \gamma_{sI}) = f_T * \gamma_{(1+s)I}.$$

The first error is within $KT^{-2}e^{-aT^2}$ of its Gaussian comparator, while the second has zero discrepancy. Both are smooth log-concave with common curvature bounds by Lemma 2.3. For each small $\eta > 0$, choose $T^2 = a^{-1}\log(K_1/\eta)$ with fixed sufficiently large K_1. The error condition is then satisfied and the latent separation is at least $k\ell(\eta)$.

The Gaussian-allocation pair in Theorem 1.1 supplies the lower $k\zeta$, and its common-noise oscillatory pair supplies $k\ell(\delta)$. Gaussian errors belong to every allowed shape neighborhood. Taking the maximum of these three lower bounds proves the result, because $\ell(\delta+\eta)$ is comparable to $\max\{\ell(\delta), \ell(\eta)\}$ at small arguments. Zero radii use $\ell(0) = 0$. □

The condition $S > 1$ provides room for the lower construction's error covariance above the normalized latent covariance. The upper comparison itself holds for every $S > 0$. The sharp statement separates covariance calibration from noise-shape fidelity: exact error covariance does not remove the shape ambiguity.

Corollary 7.2 (Sharp primary-sample risk under shape uncertainty)**.** *Assume $S > 1$ and observe n independent variables with law $f * E$, where $f \in \mathcal{F}_\zeta$ and $E \in \mathcal{E}_\eta(\Sigma)$ for some unknown $\Sigma \in \mathcal{K}_S$. For known sufficiently small radii $\zeta, \eta \geq 0$, $r = 1, 2$, and all sufficiently large n,*

$$\inf_{\widehat{f}} \sup_{f,\Sigma,E} \mathbb{E}W_2(\widehat{f}, f)^r \asymp \left[\zeta + \ell(\eta) + (\log n)^{-3/2}\right]^r,$$

uniformly in the two radii. An upper estimator may take values in $\mathcal{F}_\zeta$.

Proof. The allowed pairs (f, E) form a subspace of the separable product of two W_2 spaces. Choose a countable dense list in that subspace. Covariance and convolution into W_1 are continuous under these metrics, so the true output is approximated by the list's outputs even if the parameter constraints are not closed. Every output has uniformly bounded second moment because $\mathrm{Cov}(f * E) = \mathrm{Cov}(f) + \Sigma$.

Set $b = 1/[4(d+1)]$ and $\epsilon_n = n^{-b/2}$. As in Theorem 5.2, choose the first candidate whose empirical output distance is within ϵ_n of the countable infimum. On the event $W_1(Q_n, f * E) \leq \epsilon_n$, its true output distance is at most $3\epsilon_n$. Theorem 7.1 gives

$$W_2(\widehat{f}, f) \leq K[\zeta + \ell(\eta + 3\epsilon_n)] \leq K'[\zeta + \ell(\eta) + (\log n)^{-3/2}].$$

Lemma 5.1 bounds the exceptional probability by $Kn^{-b/2}$. The returned latent laws have uniformly bounded W_2 diameter, so that exceptional risk is negligible for either power.

For the lower, the Gaussian-allocation and shape-swap pairs have identical primary outputs and separations $k\zeta$ and $k\ell(\eta)$, respectively. The ordinary exact-Gaussian donor pair with $T^2 = A\log n$ has vanishing product total variation and separation $k(\log n)^{-3/2}$. The two-point estimation inequality gives all three risks; their maximum is comparable to the displayed sum to power r. □

8 Gaussian divisibility and transport proximity

The Gaussian component in (4) has a specific mathematical role: it makes every small signed covariance adjustment an actual probability density. Approximate transport proximity alone does not provide that property. The following explicit example records the distinction.

Proposition 8.1 (Smooth nearby laws with zero Gaussian component)**.** *There are centered isotropic C^∞ densities q_j on $\mathbb{R}$, with $\frac{1}{2} \leq (-\log q_j)'' \leq \frac{3}{2}$, such that $W_2(q_j, \gamma_1) \to 0$ and no q_j has a positive scalar Gaussian convolution component.*

Proof. Let ϕ be the standard Gaussian density, $t_k = 2^k$, $b_k = e^{-t_k}$, and $G(x) = \sum_{k\geq 1} b_k \cos(t_k x)$. Every derivative series converges uniformly. Write $B_r = \sum_k b_k t_k^r < \infty$ and, for small $\varepsilon > 0$, put

$$Z_\varepsilon = 1 + \varepsilon \sum_k b_k e^{-t_k^2/2}, \qquad p_\varepsilon(x) = \phi(x)\frac{1 + \varepsilon G(x)}{Z_\varepsilon}.$$

These are positive even probability densities. Direct differentiation gives

$$|(-\log p_\varepsilon)'' - 1| \le \frac{\varepsilon B_2}{1-\varepsilon B_0} + \frac{\varepsilon^2 B_1^2}{(1-\varepsilon B_0)^2}.$$

Their variance is $v_\varepsilon = 1 - \varepsilon Z_\varepsilon^{-1} \sum_k b_k t_k^2 e^{-t_k^2/2}$. For uniformly small ε it belongs to $[9/10, 1]$, and $q_\varepsilon(x) = \sqrt{v_\varepsilon}\, p_\varepsilon(\sqrt{v_\varepsilon}\, x)$ is isotropic with the displayed common curvature interval.

Its unscaled characteristic function has only positive summands:

$$\widehat{p}_\varepsilon(t) = Z_\varepsilon^{-1}\left[e^{-t^2/2} + \frac{\varepsilon}{2}\sum_k b_k\{e^{-(t-t_k)^2/2} + e^{-(t+t_k)^2/2}\}\right].$$

In particular $\widehat{p}_\varepsilon(t_k) \ge \varepsilon e^{-t_k}/(2Z_\varepsilon)$. If $p_\varepsilon = \mu * \gamma_a$ for any probability measure μ and $a > 0$, its characteristic modulus would be at most $e^{-at^2/2}$, a contradiction along t_k. Scaling preserves the absence of every positive Gaussian factor.

Finally put $\beta = 2\varepsilon B_0 < 1$. Since $p_\varepsilon/\phi \ge (1-\varepsilon B_0)/(1+\varepsilon B_0) \ge 1-\beta$, write $p_\varepsilon = (1-\beta)\phi + \beta h_\varepsilon$. The residual density has second moment at most one. Couple the common component identically and the residual component independently with ϕ to obtain $W_2(p_\varepsilon, \phi)^2 \le 2\beta$. The scaling coupling costs $W_2(q_\varepsilon, p_\varepsilon) \le |1 - \sqrt{v_\varepsilon}| = O(\varepsilon)$. Taking any $\varepsilon_j \downarrow 0$ proves the claim. □

Acknowledgements and disclosure

OpenAI tools assisted with literature comparison, proof development, exposition and verification. The author assumes responsibility for all content.

Funding

The author received no external research funding for this work.

Competing interests

The author operates Luminary Oddities and its Research Archive, which hosts a preprint of this work.

References

[1] M.-F. Balcan and P. M. Long, Active and passive learning of linear separators under log-concave distributions, *Proceedings of the 26th Annual Conference on Learning Theory*, PMLR **30** (2013), 288–316. https://proceedings.mlr.press/v30/Balcan13.html.

[2] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, *Numer. Math.* **84** (2000), 375–393.

[3] S. G. Bobkov, G. P. Chistyakov and F. Götze, Stability of Cramér's characterization of normal laws in information distances, in *High Dimensional Probability VII*, Progress in Probability 71, Birkhäuser, 2016, pp. 3–35. doi:10.1007/978-3-319-40519-3_1.

[4] S. G. Bobkov, G. P. Chistyakov and F. Götze, Fisher information and convergence to stable laws, *Bernoulli* **20**(3) (2014), 1620–1646. https://doi.org/10.3150/13-BEJ535.

[5] S. G. Bobkov, Upper bounds for Fisher information, *Electron. J. Probab.* **27** (2022), article 115. https://doi.org/10.1214/22-EJP834.

[6] H. J. Brascamp and E. H. Lieb, On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation, *J. Funct. Anal.* **22** (1976), 366–389.

[7] J. Capitao-Miniconi, E. Gassiat and L. Lehéricy, Deconvolution of repeated measurements corrupted by unknown noise, *Electron. J. Statist.* **19**(2) (2025), 4141–4173. https://doi.org/10.1214/25-EJS2415.

[8] C. Caillerie, F. Chazal, J. Dedecker and B. Michel, Deconvolution for the Wasserstein metric and geometric inference, *Electron. J. Statist.* **5** (2011), 1394–1423. https://doi.org/10.1214/11-EJS646.

[9] T. A. Courtade, M. Fathi and A. Pananjady, Wasserstein stability of the entropy power inequality for log-concave random vectors, *2017 IEEE International Symposium on Information Theory*, 2017, pp. 659–663. https://people.eecs.berkeley.edu/~courtade/pdfs/W2Stability_EPI.pdf.

[10] A. Delaigle, P. Hall and A. Meister, On deconvolution with repeated measurements, *Ann. Statist.* **36** (2008), 665–685. https://doi.org/10.1214/009053607000000884.

[11] E. Gassiat, S. Le Corff and L. Lehéricy, Deconvolution with unknown noise distribution is possible for multivariate signals, *Ann. Statist.* **50** (2022), 303–323. https://arxiv.org/abs/2006.14226v2.

[12] J. Johannes, Deconvolution with unknown error distribution, *Ann. Statist.* **37** (2009), 2301–2323. https://doi.org/10.1214/08-AOS652.

[13] C. Matias, Semiparametric deconvolution with unknown noise variance, *ESAIM Probab. Stat.* **6** (2002), 271–292. doi:10.1051/ps:2002015.

[14] A. Meister, On the effect of misspecifying the error density in a deconvolution problem, *Canad. J. Statist.* **32**(4) (2004), 439–449. https://doi.org/10.2307/3316026.

[15] S. Schubr, *Sharp Wasserstein Stability for Gaussian Deconvolution of Log-Concave Laws*, edition 2, Luminary Oddities research archive, 2026. https://research.luminaryoddities.com/a-sharp-logarithmic-inverse-bound-for-gaussian-estimation/v/2.

[16] C. Scricciolo, Adaptive minimax-optimal Wasserstein deconvolution with unknown error distributions, *Statist. Probab. Lett.* **230** (2026), article110589. https://doi.org/10.1016/j.spl.2025.110589.

[17] C. Scricciolo, Minimax rates for Wasserstein–Kantorovich distribution deconvolution under known ordinary smooth errors and dependent signal processes, *Statist. Methods Appl.* (2026). https://doi.org/10.1007/s10260-026-00858-9.