DIFFUSION COMPOSITION AS AN INVERSE PROBLEM: NOISE RESOLUTION AND LOW-RANK RECOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

When can separately available conditional diffusion scores recover a joint condition without joint training? We study this question as an inverse problem under a known Gaussian background and two positive bounded likelihoods on known, disjoint variable blocks. The likelihoods may be non-Gaussian, nonperiodic, and nonsmooth. Three aspects of the posterior correlation spectrum have different roles. Its largest canonical correlation ρ_1 determines the optimal same-noise error exponent, $1-\rho_1^2$, even with access to complete single-condition score fields. Its rank r determines the optimal deterministic query complexity for recovering one full score vector: $\Theta((\log(1/\eta))^r)$ for accuracy η. Its first discarded singular value determines, up to constants, the best worst-case score bias of a rank-k Gaussian replacement. A finite residual-based construction attains the information and query bounds, retains all output coordinates while querying only the correlated directions, and requires no score derivatives. Fully recorded analytic-source studies track the predicted exponents and demonstrate accurate full-vector recovery, while showing that high-order exact-source precision can be lost under small fixed source errors. A discrete-chain identity separates score-substitution error from sampling quality. A hard-condition extension replaces inverse-evidence sensitivity by a sharp logarithmic power; bounded non-Gaussian perturbation guarantees delimit the scope. The sharp laws are Gaussian score-oracle results.

1 INTRODUCTION

A model may generate condition A and condition B separately, yet fail when their scores are combined. Score addition is a useful starting point for compositional generation (Liu et al., 2022; Geffner et al., 2023), but the score of a product at a noisy time is generally not the score obtained by noising the intended clean product. Existing work studies this mismatch through MCMC, Gaussian posterior approximations, projective composition, and logical guidance (Du et al., 2023; Linhart et al., 2026; Bradley et al., 2025; Alesiani et al., 2026; Blohm and Garg, 2026). Recent evidence also separates inference error from potentially catastrophic learned-score error in out-of-distribution composition (Soiffer et al., 2026). The remaining question is not only which correction to add. It is also *whether the separately supplied information resolves the joint target accurately enough*.

Bayes' rule gives

$$s_{AB} = s_A + s_B - s_0 + \nabla_x \log \frac{p(A, B \mid x_t)}{p(A \mid x_t)p(B \mid x_t)}. \tag{1}$$

This identity is standard. It isolates an unknown function that may be difficult to recover from noisy single-condition scores. Nor can arbitrary joint semantics be inferred from two marginals: throughout this work, the clean joint likelihood is specified to be the product of the two likelihoods. Even under that specification, marginal smoothing can hide features whose interaction remains visible in the joint target.

We characterize this obstruction and its computational cost in a structured setting. The background denoising posterior is known and Gaussian; each likelihood is an arbitrary measurable function in a fixed positive interval on its own variable block. Gaussianity of the background still permits non-Gaussian conditional distributions. Our main information interface is the ordinary, same-noise single-condition score. The algorithm may query this interface at different locations, but receives neither clean likelihood values nor accurate score derivatives.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

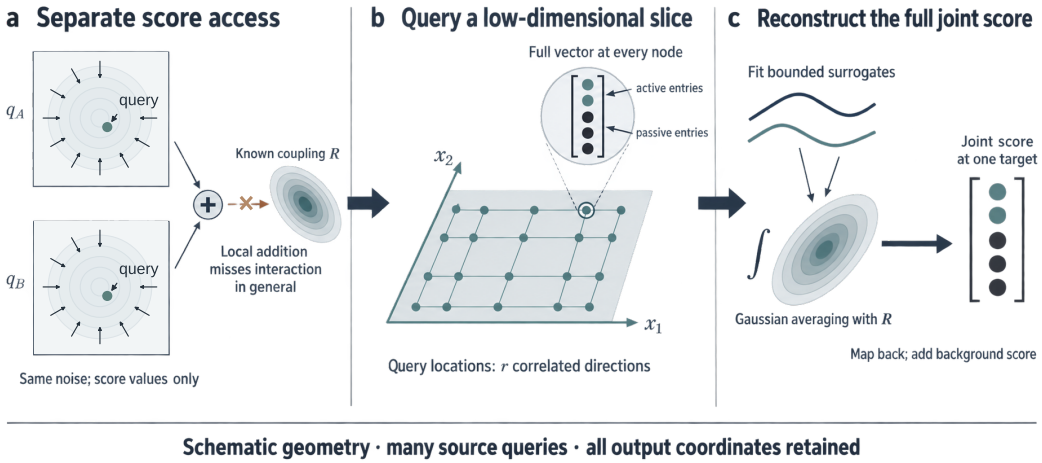


Figure 1: Recovering one full joint score from separate score-value access. (a) Local score addition generally misses the interaction induced by known R; $R = 0$ factorizes. (b) Multiple locations vary only along the r correlated directions, but every response retains its active and passive entries. (c) Bounded scalar/vector surrogates are fitted to vector residuals, Gaussian-averaged with the known coupling, and mapped back with the background score. All geometry and field glyphs are schematic, not measured scores. The algorithm never queries the unknown likelihoods or the joint target; a checked residual is not a certified target-error bound.

Why this model? A Gaussian prior with two nonlinear, separately specified sensor likelihoods already defines a non-Gaussian Bayesian inverse problem. Here the likelihoods can encode arbitrary bounded decision regions, and each block can be high dimensional; only their dependence channel is Gaussian. The model isolates whether inference can recover a prescribed composition *even before* learning and rare-event normalization create additional obstacles. Gaussianity is essential to the sharp Schur and Hermite laws, not a generic local approximation. Appendix H quantifies bounded non-Gaussian misspecification and gives a covariance-only failure example. Appendix I sharpens the hard-condition penalty to a logarithmic power with a matching witness, without extending all oracle theorems to rare events.

The answer separates three properties of the posterior cross-correlation matrix. The largest canonical correlation controls the precision lost in composition; the number of nonzero correlations controls how many locations must be queried; and the largest omitted correlation controls low-rank approximation bias. We prove matching information exponents, matching deterministic query orders for a *single output point*, and matching low-rank replacement bias. A common finite-query construction connects these guarantees. It searches only the correlated coordinates, yet recovers the score in every coordinate.

The distinction between recoverability and numerical robustness is consequential. Our fully recorded reference study shows that a value-only implementation can recover the full score far more accurately than local Gaussian approximations on exact sources. The same implementation can lose that advantage under small, fixed high-frequency source errors. The experiments therefore test recoverable information and its numerical use separately; they do not identify exact-source precision with a reliable default for learned scores.

Relation to prior work. JAC's Gaussian backward-posterior fusion and its omitted covariance-derivative term are due to Linhart et al. (2026); recovering that term is not our contribution. Touron et al. (2026) analyze a specified compositional algorithm. Soiffer et al. (2026) distinguish inference-time approximation from learned-score error, with theory-guided evidence and a broader infeasibility conjecture for OOD composition. Our lower bounds are proved for a specified oracle class; they do not prove that broader conjecture. Lift Scores provides a training-free resampling criterion for condition alignment (Yu and Gao, 2025); projective and logical approaches study compatible targets and operators (Bradley et al., 2025; Alesiani et al., 2026; Blohm and Garg, 2026). Feynman–Kac

methods correct evolving distributions (Skreta et al., 2025). The small-time support geometry of mixed-score dynamics studied by Liu and Zuazua (2026) is distinct from our fixed-positive-noise minimax recovery problem. Gaussian-kernel approximation, entire-function recovery, and maximal correlation supply classical tools (Karvonen et al., 2021; Batenkov et al., 2019; Nourdin et al., 2019); their exponential approximation rates are not themselves new here.

Figure 1 shows how low-dimensional queries retain the full output vector and recover the missing interaction.

Contributions and scope. In the Gaussian score-oracle model we prove three matched laws: the precision exponent $1-\rho_1^2$ from complete fields, the deterministic query order $\Theta((\log(1/\eta))^r)$ for one full score vector, and the rank-k replacement bias $\Theta(\rho_{k+1})$. A common value-only construction links the information and query guarantees. The numerical studies distinguish exact-source approximation, sensitivity to fixed source errors, and substitution for a joint-score discretization. The laws do not assume a learned-score error model, and deterministic query optimality does not settle randomized query complexity.

2 A STRUCTURED COMPOSITION PROBLEM

For a covariance $\Gamma \succ 0$, write $H_\Gamma f(x) = \mathbb{E} f(x + \Gamma^{1/2} G)$, with G standard Gaussian. At a fixed diffusion time, whiten each of two known posterior marginal blocks. The target covariance and its score correction become

$$\Sigma_R = \begin{pmatrix} I_p & R \\ R^\top & I_q \end{pmatrix}, \qquad w_R(\mu) = \mathbb{E}_{\Sigma_R} f(\mu_U + U) g(\mu_V + V), \qquad Z_R = \nabla_\mu \log w_R. \tag{2}$$

We assume $\|R\|_{\mathrm{op}} < 1$ and $m \le f, g \le M$, where $0 < m < 1 < M$ is fixed. There is no periodicity, smoothness, decay, or bandlimit assumption. Constants multiplying f and g do not change Z_R.

Let $\rho_1 \ge \cdots \ge \rho_r > 0$ be the nonzero singular values of R. Separate orthogonal transformations within the whitened blocks put $U = (u, a)$ and $V = (v, b)$, where $u, v \in \mathbb{R}^r$ have cross-covariance $D = \mathrm{diag}(\rho_1, \ldots, \rho_r)$. All passive background coordinates a, b are independent of the active pair and of one another. Define

$$\Sigma_* = \begin{pmatrix} I_r & D \\ D & I_r \end{pmatrix}, \qquad S = I_r - D^2. \tag{3}$$

The main source interface supplies $q_A = \nabla \log H_{I_p} f$ and $q_B = \nabla \log H_{I_q} g$. One query returns a *full vector*, with Euclidean error at most δ. Errors may be systematic; repeated queries need not reduce them. Algorithms know the background and block maps and may query arbitrary locations. Finite-query optimality is always over *deterministic* algorithms, including adaptive ones. The complete-field two-world lower bound also applies to randomized outputs in expected error; a randomized finite-query lower bound remains open.

For comparison, an extended interface supplies $\nabla \log H_{\mathrm{diag}(\Gamma_A, I)} f$ and the analogous B score, with $\Gamma_A, \Gamma_B \succ 0$ acting only on active coordinates. This anisotropic capability is *not* automatically available by choosing an earlier scalar diffusion time. It is used to state the information boundary, not assumed in the same-noise experiments.

Let $\mathcal{R}_{Q,\delta}$ be the infimum, over deterministic algorithms using at most Q queries, of the worst-case Euclidean error at a specified target mean, uniformly over admissible likelihoods and oracle errors. Let $\mathcal{R}_{\infty,\delta}$ allow entire source fields. Background parameters, r, source covariances, m, and M are fixed in all asymptotic statements. Constants may deteriorate near zero source noise or perfect correlation; they are not claimed to be uniform as r grows.

Connection to the diffusion score. For $X_0 \sim N(0, Q_0^{-1})$ and $X_t = X_0 + \sqrt{t} G$, the background posterior is $N(A_t x, C_t)$, where $C_t = (Q_0 + t^{-1} I)^{-1}$ and $A_t = C_t / t$. If W_t is the blockwise whitening and canonical-coordinate map, then, with correspondingly transformed likelihoods,

$$s_t^{AB}(x) = s_t^0(x) + (W_t A_t)^\top Z_{R_t}(W_t A_t x). \tag{4}$$

Thus source projections and output coordinate changes carry their operator-norm factors. A bound on each coordinate of an oracle error is not a dimension-independent vector bound. Appendix A makes these maps explicit.

3 NOISE RESOLUTION IS AN INFORMATION CONSTRAINT

Theorem 1 (Optimal precision exponent and stability threshold)**.** *In the model of Section 2, for $r \geq 1$ set*

$$\alpha_* = \min\left\{1, \lambda_{\min}(\Gamma_A^{-1/2} S \Gamma_A^{-1/2}), \lambda_{\min}(\Gamma_B^{-1/2} S \Gamma_B^{-1/2})\right\}. \tag{5}$$

Then $\lim_{\delta \downarrow 0} \log \mathcal{R}_{\infty,\delta} / \log \delta = \alpha_*$. Moreover, $\mathcal{R}_{\infty,\delta} = O(\delta)$ if and only if $\Gamma_A \preceq S$ and $\Gamma_B \preceq S$, including equality. For same-noise sources, $\alpha_* = 1 - \rho_1^2$.

We claim a critical exponent, not an unqualified $\Theta(\delta^{\alpha_*})$ law without subleading corrections. In particular, arbitrarily many same-noise spatial queries cannot produce uniform linear stability when $\rho_1 > 0$.

Mechanism. At active frequencies (k, l), the joint Gaussian multiplier has exponent

$$\|k\|^2 + 2k^\top D l + \|l\|^2 = k^\top S k + \|l + Dk\|^2. \tag{6}$$

A frequency hidden by source smoothing Γ_A can therefore remain relevant after only the smaller effective smoothing S. For a scalar correlation $\rho = 0.8$, the target marginal variance is 1 but the residual conditional variance is 0.36. A source variance γ has critical exponent $\min\{1, 0.36/\gamma\}$. The threshold measures required information resolution, rather than tuning a correction coefficient.

The lower bound uses positive sine/cosine likelihoods whose source scores are indistinguishable within δ, while their joint scores at zero differ. The upper bound uses Gaussian-weighted interpolation, not a small-interaction assumption. Both arguments are given in Appendix C. The statement applies to complete fields, so it also limits any local formula or finite set of queries.

4 FINITE QUERIES RECOVER THE FULL VECTOR THROUGH THE CORRELATION SUBSPACE

Theorem 2 (Optimal deterministic pointwise query order)**.** *For fixed $r \geq 1$, fixed nonzero correlations, and fixed positive source covariances,*

$$c e^{-C(Q+1)^{1/r}} \leq \mathcal{R}_{Q,0} \leq C' e^{-c' Q^{1/r}}. \tag{7}$$

Consequently the optimal deterministic *query count is $Q_*^{\det}(\eta) = \Theta((\log(1/\eta))^r)$. The upper bound recovers all coordinates; the lower bound holds even for one component at one target point and permits adaptive queries anywhere. In the Schur-stable interface,*

$$c \max\{\delta, e^{-C(Q+1)^{1/r}}\} \leq \mathcal{R}_{Q,\delta} \leq C'\{\delta + e^{-c' Q^{1/r}}\}. \tag{8}$$

For any $0 < \beta < \alpha_*$, a finite-query construction instead satisfies $\left\|\widehat{Z} - Z_R\right\| \leq C_\beta \delta^\beta + C_\beta e^{-c_\beta Q^{1/r}}$. The query bounds and error constants do not depend on passive output dimensions.

These are oracle-complexity statements, not dimension-free arithmetic or solver-time bounds. The risk exponents have different constants on their two sides. For $R = 0$, two same-noise queries suffice, with error at most $\sqrt{2}\delta$.

4.1 REDUCTION WITHOUT DISCARDING PASSIVE SCORE COMPONENTS

By translation take the target passive mean to be zero. Define

$$F_0(s) = \mathbb{E} f(s, G_a), \quad J_0(s) = \nabla_a \mathbb{E} f(s, a + G_a)|_{a=0}, \quad F = F_0/c_A, \quad J_A = J_0/c_A, \tag{9}$$

where $c_A = (H_{\Gamma_A} F_0)(0)$; define G, J_B similarly. Then $H_{\Gamma_A} F(0) = 1$, $m/M \leq F \leq M/m$, and $\|J_A(s)\| \leq (M/m)\sqrt{2/\pi}$ independently of the passive dimension. Importantly, J_A differentiates the *unnormalized* slice before division by c_A; differentiating a normalized slice would lose its mass derivative.

Writing $T(h, k) = H_{\Sigma_*}(h \otimes k)$ and $w = T(F, G)$, the desired components are

$$Z_u = \nabla_u w / w, \quad Z_v = \nabla_v w / w, \quad Z_a = T(J_A, G)/w, \quad Z_b = T(F, J_B)/w. \tag{10}$$

Algorithm 1 Finite-query recovery (FQR) at a target mean

Require: Known posterior marginals and cross-correlation R; score-value oracles; order n; retained rank k ($k = r$ without truncation).

1: Whiten the blocks, retain k canonical pairs, and keep the original background score.
2: Query full source vectors on each active cosine grid; cache repeated locations.
3: Fit $(h_A, j_A), (h_B, j_B)$ with boundedness, origin normalization, and residual (12).
4: Check actual constraints and maximum *vector* residuals E_A, E_B.
5: Evaluate (10) by finite Gaussian Fourier sums, replacing the denominator by $\max\{\widehat{w}, (m/M)^2/2\}$.
6: Map the full vector back using (4).

A single full-vector query on the active slice supplies both

$$q_{A,\parallel} = \nabla u/u, \qquad q_{A,\perp} = V/u, \quad u = H_{\Gamma_A} F, \quad V = H_{\Gamma_A} J_A. \tag{11}$$

The search dimension is r; the output is not restricted to that subspace. In particular, one p-vector source call has r active entries that identify F and $p - r$ passive entries that identify J_A on the *same* slice. Dropping those passive entries would change the task. Table 1 maps the upper and lower arguments.

4.2 A RESIDUAL-BASED CONSTRUCTION

At order n, query a tensor cosine grid in an active box of radius $L_n \asymp \sqrt{n}$. Fit bounded scalar and vector trigonometric polynomials h, j with an auxiliary period proportional to L_n, impose $(H_\Gamma h)(0) = 1$, and minimize the maximum node residual

$$E_n = \max_{x \text{ queried}} \left\| \begin{pmatrix} \nabla H_\Gamma h(x) - \widetilde{q}_\parallel(x) H_\Gamma h(x) \\ H_\Gamma j(x) - \widetilde{q}_\perp(x) H_\Gamma h(x) \end{pmatrix} \right\|. \tag{12}$$

This is a finite convex problem. The auxiliary functions may be signed. Only the representation is periodized, with growing period; the true likelihoods remain nonperiodic. A concrete construction uses at most $2(8n+1)^r$ logical query slots, or $2(4n+1)^r$ unique locations after removing symmetry duplicates. All derivatives in (12) act on known basis functions, not on the oracle.

The proof does not interpolate the score as though it had a Gaussian Fourier tail: taking a logarithmic derivative can destroy that property. Instead, it controls

$$W = u\nabla\widehat{u} - \widehat{u}\nabla u, \qquad B = u\widehat{V} - V\widehat{u}. \tag{13}$$

These cross-products inherit controlled entire-function growth. Stable reconstruction from the finite nodes bounds them over the expanding box. Integrating $\nabla(\widehat{u}/u) = W/u^2$ from the normalized origin, and then using Gaussian-weighted rather than box-uniform error, yields the claimed precision without a spurious $L_n\delta$ loss. Appendix B gives the construction, explicit intermediate tails, and the transfer to (10).

The upper proof is in Appendix B; the independent single-point lower proof is in Appendix D. The latter uses Hermite *functions* for the nonperiodic density kernel, whereas the replacement proof below uses Hermite *polynomials* in Gaussian L^2.

5 THE FIRST OMITTED CORRELATION CONTROLS LOW-RANK BIAS

Here $r = \operatorname{rank} R$ remains the true correlation rank; k denotes a chosen replacement rank, and n the representation order. Singular values beyond rank r are zero.

Theorem 3 (Task-specific optimal Gaussian replacement)**.** *Let $\rho_1 \le \bar{\rho} < 1$, fix $0 \le k \le \min(p, q)$, and let R_k be the rank-k singular-value truncation. There are constants depending only on $m, M, \bar{\rho}$, not on the block dimensions, such that*

$$\sup_{f,g,\mu} \|Z_R(\mu) - Z_{R_k}(\mu)\| \le K_{\mathrm{tr}} \rho_{k+1}, \tag{14}$$

Table 1: Four-step proof map for Theorem 2. Both sides concern one target point, not reconstruction of an entire ambient score field.

Finite-query upper bound	**Deterministic adaptive lower bound**
1. Preserve the vector. Integrate passive variables into scalar F, G and vector J_A, J_B on $\mathbb{R}^r$.	**1. Freeze the transcript.** Follow the algorithm's zero-answer path; Q queries impose at most rQ constraints.
2. Fit finite surrogates. Query $Q = O(n^r)$ nodes in a box of radius $O(\sqrt{n})$; use bounded basis coefficients.	**2. Leave invisible directions.** Use two N^r-dimensional Hermite spaces with $N^r > rQ + 1$.
3. Control cross-products. Entire-function tails transfer node residuals to $u\nabla\widehat{u} - \widehat{u}\nabla u$ and $u\widehat{V} - V\widehat{u}$.	**3. Preserve the target signal.** The derivative pairing matrix has $\sigma_{\min} \geq e^{-CN}$; parity equalizes denominators.
4. Transfer to the target. Normalize at the origin and use Gaussian-weighted Schur interpolation: $C_\beta(\delta^\beta + e^{-c_\beta n})$.	**4. Separate one component.** Positive bounded factors share the transcript but differ by $N^{-O(r)}e^{-CN}$; take $N \asymp Q^{1/r}$.

Steps 3–4 determine the matched rate; steps 1–2 set up each side.

$$c\rho_{k+1} \leq \inf_{\substack{\operatorname{rank}\widetilde{R}\leq k \\ \|\widetilde{R}\|_{\mathrm{op}}<1}} \sup_{f,g,\mu} \left\|Z_R(\mu) - Z_{\widetilde{R}}(\mu)\right\| \leq K_{\mathrm{tr}}\rho_{k+1}. \tag{15}$$

Furthermore, if both cross-correlation matrices have norm at most $\bar{\rho}$, then $\sup_{f,g,\mu} \left\|Z_R - Z_{\widetilde{R}}\right\| \leq K_{\mathrm{est}} \left\|R - \widetilde{R}\right\|_{\mathrm{op}}$*, without a spectral-gap assumption.*

The infimum is over Gaussian replacements with the *same marginals*, not over every possible compressed representation. The relevant tail is ρ_{k+1}, not $\sqrt{\sum_{j>k}\rho_j^2}$. This is possible because composition tests separable functions and their first moments, rather than approximating the whole joint density in total variation.

Specifically, the Gaussian conditional-expectation operator obeys $\|T_R - T_{R_k}\|_{L^2(\gamma_q)\to L^2(\gamma_p)} = \rho_{k+1}$. This controls $w_R = \mathbb{E}_R fg$ and $b_R = \mathbb{E}_R[(U,V)fg]$. Passing to the full gradient also requires the covariance-inverse term:

$$n_R - n_{R_k} = \Sigma_R^{-1}(b_R - b_{R_k}) - \Sigma_R^{-1}(\Sigma_R - \Sigma_{R_k})n_{R_k}, \qquad n_R = \nabla w_R. \tag{16}$$

The lower bound uses a vector in $\ker\widetilde{R}$ and positive sine/cosine likelihoods. Appendix E proves both bounds and supplies explicit constants.

Combining the results gives a single residual-sensitive budget. If $\varepsilon_R = \left\|R - \widetilde{R}\right\|_{\mathrm{op}}$, then for a feasible order-n candidate with $k \geq 1$ retained directions under same-noise sources,

$$\left\|\widehat{Z} - Z_R\right\| \leq K_{\mathrm{est}}\varepsilon_R + K_{\mathrm{tr}}\sigma_{k+1}(\widetilde{R}) + C_{k,\beta}\big(E_A + E_B + \delta + e^{-c_k n}\big)^\beta, \tag{17}$$

where $0 < \beta < 1 - \left\|\widetilde{R}_k\right\|_{\mathrm{op}}^2$. This separates background estimation, rank truncation, source error, and finite representation. It does not imply that every increase in n improves a numerical implementation. Nor is a checked node residual, by itself, a calibrated output-error certificate: the constants, tails, and source-error bound must also be supplied.

5.1 Conditioning and Gaussianity are separate restrictions

The lower bounds already hold for factors in $[1-\epsilon, 1+\epsilon]$ for each fixed $\epsilon > 0$: the information obstruction is not caused by rarity. For replacement bias with hard factors $0 \leq f, g \leq 1$, a sharper conditioning law is possible. If $w_R \geq \zeta > 0$, both correlations have norm at most $\bar{\rho} < 1$, and $\tau = \left\|R - \widetilde{R}\right\|_{\mathrm{op}}$, Theorem 6 proves

$$\left\|Z_R - Z_{\widetilde{R}}\right\| \leq C_{\bar{\rho}}\tau\sqrt{\log(1/\zeta)}\big(1 + \log(1/\zeta)\big). \tag{18}$$

Table 2: Executable reference study: mean full-vector canonical score error over 16 fixed target means, for the same likelihood pair and four source settings. Queries count unique full source-vector evaluations. Every cell is computed from the supplied case records; there are 512 evaluations in total. "High" multiplies both perturbation frequencies by ten, without changing the stated vector budget.

Method	**Queries**	**Exact**	**Low** 10^{-8}	**High** 10^{-8}	**High** 10^{-6}
Addition	2	1.74×10^{-2}	1.74×10^{-2}	1.74×10^{-2}	1.74×10^{-2}
JAC	10	3.27×10^{-3}	3.27×10^{-3}	3.27×10^{-3}	3.27×10^{-3}
Full overlap	26	1.06×10^{-3}	1.06×10^{-3}	1.06×10^{-3}	1.06×10^{-3}
FQR, $n = 1$	10	4.37×10^{-3}	4.37×10^{-3}	4.37×10^{-3}	4.37×10^{-3}
FQR, $n = 2$	18	1.80×10^{-4}	1.80×10^{-4}	1.80×10^{-4}	1.97×10^{-4}
FQR, $n = 4$	34	8.19×10^{-7}	8.22×10^{-7}	5.54×10^{-4}	2.19×10^{-3}
FQR, $n = 8$	66	4.69×10^{-7}	4.68×10^{-7}	8.52×10^{-5}	4.49×10^{-3}
FQR, $n = 16$	130	2.44×10^{-7}	2.43×10^{-7}	1.57×10^{-3}	4.82×10^{-3}

There is no dimension factor or requirement $\tau \leq \zeta/2$. The proof reserves independent Gaussian smoothing, controls log likelihoods by entropy, and interpolates covariance while retaining the product structure. The logarithmic power $3/2$ is necessary: for $f_L = \mathbf{1}_{\{|u| \geq L\}}$, $g_L = \mathbf{1}_{\{v \geq L\}}$, and $\rho = \theta/L^2$, the rank-zero score gap divided by L tends to $\tanh \theta$, while $\log(1/w_\rho) \sim L^2$. Appendix I supplies the bound, evidence transport, and witness. This is not a hard-event query theorem; source evidence across the query region still matters.

For non-Gaussian $p_\varepsilon \propto q_0 e^{\varepsilon v}$ with $\operatorname{osc} v \leq 1$, Proposition 4 transfers the reference algorithm with a dimension-independent mismatch term. Proposition 7 identifies the first-order change as $\varepsilon t^{-1} \operatorname{Cov}(X_0, v)$ under the reference conditional posterior, with a uniform quadratic remainder. A retained mismatch term is not a lower bound on recovery error. The joint posterior covariance is *additional information*, not an oracle-only correction available to FQR.

6 Numerical validation: recoverability, robustness, and sampler substitution

The experiments address three different questions: whether the predicted information loss is visible, how a finite value-only implementation uses the available precision, and what score substitution means for sampling. Our primary executable studies contain 200 high-precision two-world instances and 512 recovery evaluations on a fully specified nonperiodic likelihood pair. All use CPU computation and analytic sources; no network is trained. The broader historical benchmark and sampling summaries are retained separately in Appendix G; their complete execution archive is unavailable in the supplied artifacts. The reference studies below are independently specified executions, not reconstructions of those summaries.

Information resolution. We execute the sine/cosine two-world construction of Appendix C at 100-digit precision, with source budgets from 10^{-2} to 10^{-64}. Terminal local slopes at $\rho = 0.2, 0.6, 0.8, 0.95$ are $0.96195, 0.63877, 0.35781, 0.09441$, against $1 - \rho^2 = 0.96, 0.64, 0.36, 0.0975$. Each budget defines a different likelihood pair: these are checks of the lower-bound mechanism, not a precision curve for one dataset. The executable matrix and zero-transcript checks are described in Appendix J; the universal deterministic-adaptive statement comes from the proof, not from testing query schedules.

Full-vector recovery from exact scores. The reference study fixes $d = 4$, $r = 1$, $\rho = 0.8$, and the erf likelihood pair specified in Appendix J.1. Sixteen deterministic target means vary both active and passive coordinates. Table 2 reports means over this grid, with no statistical confidence interval. The practical FQR uses the supplied truncated-SVD implementation, selects only by checked feasible residual, and never receives the evaluator's joint target. All 512 method/source/target combinations complete without failure or denominator clipping.

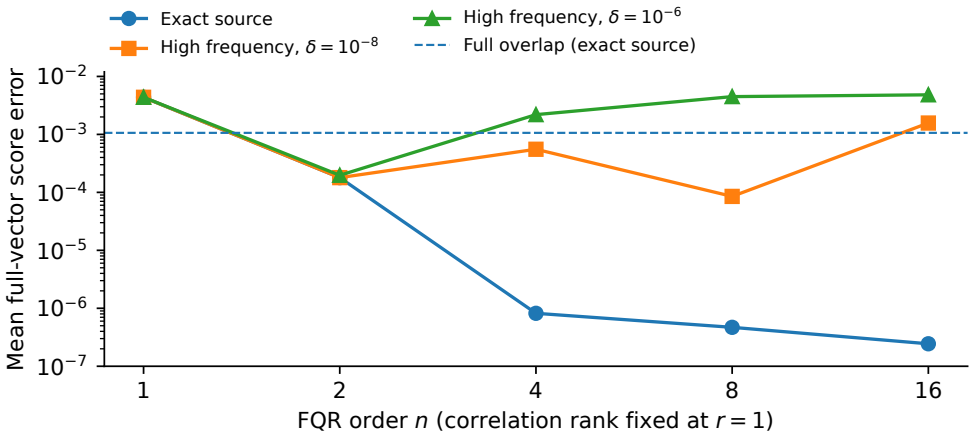


Figure 2: Exact-source accuracy and fixed-error sensitivity in the executable reference study. Each point averages the same 16 targets; rank remains $r = 1$ while representation order varies. The high-frequency fields have a global Euclidean source-error bound δ. The dashed line is exact-source full Gaussian overlap; its perturbed means are given in Table 2. More order improves the exact-source means but does not monotonically improve perturbed recovery. These are measured implementation curves, not fitted minimax rates.

With exact sources, the order-4 candidate has mean error 8.19×10^{-7}, compared with 1.06×10^{-3} for full Gaussian overlap. A more economical comparison is $n = 2$: its mean error is 1.80×10^{-4} with 18 queries, versus 26 for overlap. At the matched ten-query budget, JAC is more accurate than $n = 1$. Both $n = 2$ and $n = 4$ beat overlap at all 16 exact-source targets. These are fixed-model comparisons, not an estimate of performance over arbitrary attributes.

The baselines use the same score values and cached finite differences with step 10^{-4}. Independent 60-digit differentiation, including the covariance derivatives of full overlap, differs from the implemented baselines by at most 9.53×10^{-11} at three prescribed targets. The analytic joint evaluator agrees with independent Gaussian quadrature within 3.22×10^{-16} over all 16 targets. Removing the passive outputs from the order-4 reconstruction raises mean error to 0.185: querying a one-dimensional slice does not make the target a one-dimensional score.

Precision is not robustness. The same globally bounded error field is supplied to all methods, and repeated locations return identical errors. Low and high frequencies differ by a factor of ten; the complete formula and phases are fixed in Appendix J.1. At the same budget $\delta = 10^{-8}$, changing frequency raises the order-4 mean from 8.22×10^{-7} to 5.54×10^{-4}. At high-frequency budget 10^{-6}, its mean is 2.19×10^{-3}, worse than overlap, and its maximum is 1.22×10^{-2}. The order-8 maximum is 5.78×10^{-2} at that budget, so means alone would hide important variation.

The lower-order $n = 2$ candidate beats overlap at every target in all four source settings, but is not uniformly the best FQR order: $n = 8$ has lower mean error at high-frequency budget 10^{-8}. Thus the evidence supports an explicit precision–sensitivity tradeoff, not a universally preferred order. Theorem 1 limits worst-case recoverability from source information; it does not prove that these particular numerical amplifications are unavoidable. Likewise, the finite-query existence theorem does not make a residual-minimizing SVD candidate noise-adaptive. Source uncertainty must be accounted for when choosing representation and regularization; an error-adaptive rule for learned scores is not established here.

What a sampler comparison measures. For Euler chains with common initialization, grid, and Gaussian transition covariance, the Gaussian KL formula and chain rule give

$$\mathrm{KL}(\widehat{P} \| P_*) = \frac{1}{2} \sum_k h_k \mathbb{E}_{\widehat{P}} \left\| \widehat{s}_{t_k}(X_k) - s^{AB}_{t_k}(X_k) \right\|^2 . \tag{19}$$

Here P_* is the *exact-score Euler chain*, not continuous reverse diffusion or the clean target. This identity isolates score-substitution error; initialization, discretization, and early stopping are separate terms in Appendix F. Six independently integrated Gaussian-transition checks include a step larger than one, for which the discrete identity still applies.

The historical coupled-sampling summaries in Table 5 report very small substitution energies, but no consistent clean-distribution improvement across seeds. Close paired samples are evidence of agreement with the chosen discretization, not independent evidence of better generation. Those trajectory-level records are not available for re-execution here. We therefore use the sampler identity to delimit the interpretation, rather than promote the reported energy reduction as a general sampling-quality or speed advantage.

Scope diagnostics. Executable checks of Theorem 6 cover all 12 specified evidence settings, including three outside Proposition 5's domain. The two-branch rare-event witness approaches the sharp logarithmic power: at $L = 64$ and $\theta = 0.1$, the gap divided by L is 0.09972, versus $\tanh(0.1) = 0.09967$. The 108 signed perturbation-response comparisons isolate a quadratic *target* remainder using joint evaluator information, not an improved score-only FQR method. The historical non-Gaussian recovery summaries are kept distinct in Appendix H.5.

7 Discussion

The three spectral laws answer different questions: how much source precision survives composition, how many deterministic queries recover one full score, and how much Gaussian low-rank replacement costs. The common value-only construction preserves every output coordinate. The executable reference study demonstrates accurate full-vector recovery in the stated model, while its fixed-error tests show why exact-source precision is not a robustness guarantee. Known blocks, known Gaussian dependence, and controlled source errors are substantive assumptions. Hard-condition and bounded-potential results quantify separate extensions without transporting the principal oracle laws to arbitrary backgrounds. Randomized query optimality, unknown semantic views, and error-adaptive use of learned scores remain research directions. The contribution is a recoverability theory, with an executable implementation study that makes its numerical and modeling boundaries explicit.

AI use statement

Generative AI tools, including ChatGPT and Codex, were used substantially for conceptual modeling; proposing mathematical statements, proof strategies, and proof text; argument auditing; experiment and diagnostic design, implementation, and analysis; literature discovery; and drafting, translation, editing, typesetting, and figure preparation. The paper distinguishes executed analytic reference studies from historical aggregate reports. Numerical and high-precision checks are not formal proof verification. Human authors remain responsible for the claims, mathematical correctness, research integrity, artifacts, and disclosure.

Reproducibility statement

Appendices A–F provide the constructions and proofs. The source package includes the score-value-only implementation, a 200-case information study, the complete 512-row reference recovery study with fixed configurations and queried vectors, and executable independent checks. Appendix J specifies the inputs, commands, recorded outputs, and baseline validation. Historical benchmark and trajectory aggregates are retained separately in Appendices G and H.5; their original execution archive is not supplied, and the reference studies do not reconstruct it.

Ethics statement

The experiments use synthetic analytic distributions and contain no human-subject data. The paper does not establish reliability or semantic safety for deployed generative models. Its restricted

information model, source-error assumptions, negative robustness results, and non-certification status are stated explicitly to limit such misinterpretation.

REFERENCES

Francesco Alesiani, Jonathan Warrell, Tanja Bien, Henrik Christiansen, Matheus Ferraz, and Mathias Niepert. Logical Guidance for the Exact Composition of Diffusion Models. *arXiv preprint arXiv:2602.05549v2*, 2026.

Dmitry Batenkov, Laurent Demanet, and Hrushikesh N. Mhaskar. Stable soft extrapolation of entire functions. *Inverse Problems, 35:015011*, 2019.

Peter Blohm and Vikas K. Garg. Composition of Pretrained Diffusion Models: A Logic-Based Calculus. *International Conference on Learning Representations*, 2026.

Arwen Bradley, Preetum Nakkiran, David Berthelot, James Thornton, and Joshua M. Susskind. Mechanisms of Projective Composition of Diffusion Models. *Proceedings of the 42nd International Conference on Machine Learning, PMLR 267:5319–5354*, 2025.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. *International Conference on Learning Representations*, 2023.

Yilun Du, Conor Durkan, Robin Strudel, Joshua B. Tenenbaum, Sander Dieleman, Rob Fergus, Jascha Sohl-Dickstein, Arnaud Doucet, and Will Sussman Grathwohl. Reduce, Reuse, Recycle: Compositional Generation with Energy-Based Diffusion Models and MCMC. *Proceedings of the 40th International Conference on Machine Learning, PMLR 202:8489–8510*, 2023.

Tomas Geffner, George Papamakarios, and Andriy Mnih. Compositional Score Modeling for Simulation-Based Inference. *Proceedings of Machine Learning Research, 202:11098–11116*, 2023.

Nathael Gozlan, Cyril Roberto, and Paul-Marie Samson. A new characterization of Talagrand's transport-entropy inequalities and applications. *Annals of Probability, 39:857–880*, 2011.

Toni Karvonen, Chris J. Oates, and Mark Girolami. Integration in reproducing kernel Hilbert spaces of Gaussian kernels. *Mathematics of Computation, 90:2209–2233*, 2021.

Julia Linhart, Gabriel Victorino Cardoso, Alexandre Gramfort, Sylvain Le Corff, and Pedro L. C. Rodrigues. Diffusion posterior sampling for simulation-based inference in tall data settings. *arXiv preprint arXiv:2404.07593v3*, 2026.

Kang Liu and Enrique Zuazua. Geometric Asymptotics of Score Mixing and Guidance in Diffusion Models. *arXiv preprint arXiv:2605.12231*, 2026.

Nan Liu, Shuang Li, Yilun Du, Antonio Torralba, and Joshua B. Tenenbaum. Compositional Visual Generation with Composable Diffusion Models. *European Conference on Computer Vision*, 2022.

Ivan Nourdin, Giovanni Peccati, and Xiaochuan Yang. Berry–Esseen bounds in the Breuer–Major CLT and Gebelein's inequality. *Electronic Communications in Probability, 24(34):1–12*, 2019.

Manfred Opper, Ulrich Paquet, and Ole Winther. Perturbative Corrections for Approximate Inference in Gaussian Latent Variable Models. *Journal of Machine Learning Research, 14:2857–2898*, 2013.

Felix Otto and Cédric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis, 173:361–400*, 2000.

Marta Skreta, Tara Akhound-Sadegh, Viktor Ohanesian, Roberto Bondesan, Alán Aspuru-Guzik, Arnaud Doucet, Rob Brekelmans, Alexander Tong, and Kirill Neklyudov. Feynman–Kac Correctors in Diffusion: Annealing, Guidance, and Product of Experts. *Proceedings of the 42nd International Conference on Machine Learning, PMLR 267:55906–55949*, 2025.

Duncan Soiffer, Chandler Squires, Yuan Guan, Jason Hartford, and Pradeep Ravikumar. Catastrophic Compositional Generation: Why Vanilla Diffusion Models Fail to Extrapolate. *arXiv preprint arXiv:2606.23920*, 2026.

Camille Touron, Gabriel Victorino Cardoso, Julyan Arbel, and Pedro L. C. Rodrigues. Error analysis of a compositional score-based algorithm for simulation-based inference. *arXiv preprint arXiv:2510.15817v2*, 2026.

Felix Voigtlaender. A general version of Price's theorem. *Journal of Theoretical Probability, 34:1474–1485*, 2021.

Chenning Yu and Sicun Gao. Improving Compositional Generation with Diffusion Models Using Lift Scores. *Proceedings of the 42nd International Conference on Machine Learning, PMLR 267:72944–72971*, 2025.

A COORDINATES, INFORMATION INTERFACES, AND EXACT FULL-VECTOR REDUCTION

This appendix fixes the objects underlying all three main theorems. The Gaussian background, likelihood bounds, source covariances, and active rank are known, fixed parameters. The unknown objects are the two likelihood functions. An oracle supplies only a single-condition score vector evaluated at a requested position. No clean likelihood evaluations, joint-condition observations, or exact derivatives of the oracle are supplied to a recovery algorithm. The analytic reference functions used in the experiments are available to the evaluator, not additional inputs to the recovery rule.

A.1 MAPPING AN ORDINARY DIFFUSION MODEL TO THE CANONICAL PROBLEM

Let $X_0 \sim p_0 = N(0, Q_0^{-1})$ and $X_t = X_0 + \sqrt{t}G$. Completing the square gives

$$C_t = (Q_0 + t^{-1}I)^{-1}, \qquad A_t = C_t/t, \qquad X_0 \mid X_t = x \sim N(A_t x, C_t). \tag{20}$$

Partition C_t into the two known likelihood blocks. Choose blockwise invertible maps $W_{U,t}, W_{V,t}$ which whiten its marginal covariances and rotate to canonical-correlation coordinates. Set $W_t = \mathrm{diag}(W_{U,t}, W_{V,t})$. Then

$$W_t C_t W_t^\top = \Sigma_{R_t}, \qquad \mu = W_t A_t x. \tag{21}$$

In (21) and below the likelihoods in canonical coordinates are $f_t(u) = f_0(W_{U,t}^{-1}u)$ and $g_t(v) = g_0(W_{V,t}^{-1}v)$. An invertible change of variables preserves the bounds $[m, M]$ and measurability. It does not impose Gaussianity on a likelihood-tilted distribution.

Bayes' rule expresses the noisy joint density as $p_t(x)$ times a posterior expectation of $f_0 g_0$, divided by a constant. Differentiation therefore gives (4). For a single condition the corresponding posterior expectation uses only its marginal, so its canonical score is $\nabla \log H_I f_t$. The map $W_t A_t$ is invertible. Consequently any requested canonical mean can be implemented by an original-space score query, followed by the known inverse-transpose transformation and selection of the relevant block. Bounds for errors in original scores must be multiplied by the norm of this inverse map; output errors are multiplied by $\|W_t A_t\|_{\mathrm{op}}$. We never identify these condition numbers with one.

For an extended source covariance, one must actually have a source model with the specified covariance. A change to an earlier scalar variance time generally changes passive as well as active marginal smoothing, and does not realize $\mathrm{diag}(\Gamma, I)$ for free. The same-noise results avoid this additional assumption.

A useful structural fact is that low cross-block rank persists through Gaussian denoising. If the cross block of the clean precision Q_0 has rank r, block inversion gives

$$(C_t)_{UV} = -(Q_{UU} + t^{-1}I)^{-1} Q_{UV} \big[Q_{VV} + t^{-1}I - Q_{VU}(Q_{UU} + t^{-1}I)^{-1} Q_{UV} \big]^{-1}. \tag{22}$$

Both outside factors are invertible. Thus the posterior cross-covariance has rank r at every positive time. Its directions and canonical correlations can vary with time. This rank identity does not imply time-uniform recovery constants near $t = 0$.

A.2 LIKELIHOOD NORMALIZATION AND PASSIVE DIRECTIONS

Fix a target mean and translate it to zero. Write the canonical variables as $U = (u, a)$, $V = (v, b)$. The active covariance is (3), and the passive coordinates are mutually independent standard Gaussians, independent of the active pair. Define F_0, J_0, c_A, F, J_A by (9); G, J_B are defined analogously. Gaussian differentiation under a bounded integrand is legitimate even when f is not classically differentiable. For every unit passive vector e,

$$e^\top J_0(s) = \mathbb{E}[(e^\top G_a) f(s, G_a)], \qquad |e^\top J_0(s)| \le M\mathbb{E}|N(0,1)| = M\sqrt{2/\pi}. \tag{23}$$

Taking the supremum over directions, not a sum over coordinates, proves

$$m \le c_A \le M, \quad m/M \le F \le M/m, \quad \sup_s \|J_A(s)\| \le (M/m)\sqrt{2/\pi}, \quad H_{\Gamma_A} F(0) = 1. \tag{24}$$

The factor c_A in J_A is evaluated at the current slice and then held fixed. In particular, $J_A \neq \nabla_a(F_0/c_A)$ in general. Differentiating the normalized slice instead would subtract $F\nabla_a \log c_A$ and lose a genuine component of the target score.

Integrating passive Gaussian coordinates and differentiating the original expectation establishes (10). The constants $c_A c_B$ cancel from all numerator-to-denominator ratios. Likewise, a source full-vector query at passive mean zero yields (11). This proves exact reduction to two scalar and two vector functions defined on $\mathbb{R}^r$. Their output dimensions may be large, but all bounds above are Hilbert-norm bounds independent of those dimensions.

B A FINITE NONPERIODIC RECOVERY CONSTRUCTION

We prove the upper statements of Theorem 2, including the residual-sensitive form used in (17). Fix an order n and a positive constant ℓ_0. Let

$$L_n = \ell_0\sqrt{n}, \qquad P_n = 8L_n, \qquad x_{j_1,\dots,j_r} = L_n(\cos\theta_{j_1},\dots,\cos\theta_{j_r}), \quad \theta_j = \frac{2\pi j}{8n+1}. \tag{25}$$

There are $(8n+1)^r$ logical slots per oracle, and $(4n+1)^r$ distinct locations. Removing duplicates does not change any information in the exact or fixed-error-field model. Using all slots is also a valid conservative upper bound for arbitrary bounded per-call errors.

B.1 FINITE FEASIBLE SET AND SMALL-RESIDUAL CANDIDATES

Let $\mathcal{P}_{2n}(P_n)$ be the real trigonometric polynomials with period P_n in every coordinate and coordinatewise frequency index at most $2n$. Fit a scalar h and a passive-vector-valued polynomial j from this space. On a full equispaced auxiliary grid of $(16n+1)^r$ points per period impose

$$|h(z)| \le B_F, \qquad \|j(z)\| \le B_J, \qquad (H_\Gamma h)(0) = 1, \tag{26}$$

where one may fix $B_F > 3^r M/m + 1$ and $B_J > 3^r(M/m)\sqrt{2/\pi}$. Minimize the maximum Euclidean residual (12) on (25). Residuals are affine in the coefficients, and all inequalities in (26) are convex. The problem is a finite second-order cone program. The auxiliary evaluation map is injective on $\mathcal{P}_{2n}(P_n)$; its bounded grid values bound all coefficients. Hence the feasible set is compact. The constant candidate $h = 1, j = 0$ is feasible, so a minimizer exists for every $n \ge 1$. Among residual minimizers choose the one of smallest coefficient norm. This unique secondary choice is the pointwise limit of minimizers after adding $\xi\,\|c\|_2^2$ to the objective as $\xi \downarrow 0$. Compactness, continuity in the finite oracle values, and strict convexity of the perturbed problem make this selector Borel measurable. It therefore defines a measurable recovery rule for measurable source fields. A numerical procedure can instead be analyzed through any achieved feasible residual.

For completeness, take the Fejér kernel K_N with Fourier coefficients $(1-|k|/N)_+$ on one period. The de la Vallée–Poussin kernel $V_N = 2K_{2N} - K_N$ reproduces indices $|k| \le N$, has support in $|k| \le 2N$, and has L^1 norm at most three. Tensoring preserves a 3^r bound. Applied with index $2n$, its oversampled discrete reconstruction on the auxiliary grid yields

$$\|h\|_\infty \le 3^r B_F, \qquad \|j\|_{L^\infty(\ell_2)} \le 3^r B_J. \tag{27}$$

Here V_{2n} is used because the auxiliary functions have degree $2n$. Products in the reconstruction quadrature have degree below $16n+1$, so no zero-mode aliasing occurs.

To construct a feasible comparison, restrict the true bounded F to $[-4L_n, 4L_n]^r$ and periodically extend that restriction. Denote the result by F^{per}. For means in the query cube, a Gaussian displacement must cross a distance at least $3L_n$ before encountering a copied value. Gaussian tail and differentiated-kernel bounds imply

$$\|H_\Gamma(F^{\mathrm{per}} - F)\|_{C^1([-L_n,L_n]^r)} \le Ce^{-cL_n^2}. \tag{28}$$

Filter F^{per} with the tensor V_n. The filtered function h_0 belongs to $\mathcal{P}_{2n}(P_n)$ and is bounded by $3^r M/m$. Its omitted smoothed modes have a multiplier at most $\exp[-cn^2/P_n^2]$. Summing the bounded Fourier coefficients with the gradient factors gives at most a polynomial in n times this exponential, hence Ce^{-cn} because $P_n^2 \asymp n$. Add the constant $1-(H_\Gamma h_0)(0)$, which is exponentially

small, to get the origin-normalized h. Apply the same extension and filter to J_A without adding a normalization term. This proves feasibility for sufficiently large n, with

$$E_n \le C\delta + Ce^{-cn}. \tag{29}$$

The statement is uniform in the unknown measurable likelihoods. It does not assert that the true functions are periodic or that an empirical solver finds these comparison functions.

B.2 The cross-product residual has controlled entire growth

Write $u = H_\Gamma F$, $V = H_\Gamma J$, $\widehat{u} = H_\Gamma h$, and $\widehat{V} = H_\Gamma j$. Define W, B by (13). Gaussian convolution of any bounded, possibly vector-valued function extends to an entire function. Taking the modulus of the complex-shifted Gaussian kernel gives

$$\|H_\Gamma a(x+iy)\| \le \|a\|_\infty \, e^{y^\top \Gamma^{-1} y/2}, \tag{30}$$

$$\|\nabla H_\Gamma a(x+iy)\| \le \|a\|_\infty \left(\sqrt{\operatorname{tr}\Gamma^{-1}} + \left\|\Gamma^{-1} y\right\|\right) e^{y^\top \Gamma^{-1} y/2}. \tag{31}$$

For vector-valued a, the first estimate uses its Hilbert norm; no output-dimension factor is needed. The active gradient dimension r is fixed.

Let b_F, b_J, b_h, b_j bound the real-space norms of F, J, h, j, including the factors in (27). On the complex strip $|\operatorname{Im}\theta_i| \le a$ after the substitution $x_i = L_n \cos\theta_i$, a simultaneous upper bound for $\|W\| + \|B\|$ is

$$\begin{aligned} M_{n,a} = \Big[2b_F b_h \left(\sqrt{\operatorname{tr}\Gamma^{-1}} + \left\|\Gamma^{-1}\right\| L_n \sqrt{r}\sinh a\right) + b_F b_j + b_J b_h\Big] \\ \cdot \exp\left(\left\|\Gamma^{-1}\right\| r L_n^2 \sinh^2 a\right). \end{aligned} \tag{32}$$

Moving each Fourier integration contour yields coefficient bounds $M_{n,a} e^{-a|k|_1}$. The tail outside $\|k\|_\infty \le n$ is bounded by $M_{n,a}\Xi_{r,a,n}$, where

$$\Xi_{r,a,n} = r \frac{2e^{-a(n+1)}}{1-e^{-a}} \left(\frac{1+e^{-a}}{1-e^{-a}}\right)^{r-1}. \tag{33}$$

Fix $a > 0$ sufficiently small that $\left\|\Gamma^{-1}\right\| r\ell_0^2 \sinh^2 a < a/4$. It is chosen before varying n or δ. Equations (32)–(33) then give a tail Ce^{-cn}, with polynomial factors absorbed into a smaller exponent.

At a query node, the stacked exact residual $[W; B]$ is the noisy fitting residual multiplied by u, plus $u\widehat{u}(\widetilde{q} - q)$. Its norm is at most $b_F(E_n + b_h\delta)$. The tensor V_n has discrete absolute mass at most 3^r on the angular grid in (25), and reproduces the low modes without aliasing. Applying this to the low-frequency part and bounding both omitted tails proves

$$\|W\|_{\infty,\mathrm{box}} + \|B\|_{\infty,\mathrm{box}} \le \sqrt{2}\, 3^r b_F(E_n + b_h\delta) + (1+3^r) M_{n,a}\Xi_{r,a,n}. \tag{34}$$

The factor $\sqrt{2}$ changes stacked Euclidean norm to the sum of its two block norms. Every displayed constant here is independent of passive output dimension.

This argument cannot be replaced by a Gaussian-tail assumption on q. For example, $u(x) = 1 + 2a\cos x/(1+a^2)$, $0 < a < 1$, satisfies

$$\partial_x \log u(x) = 2\sum_{k\ge 1} (-1)^k a^k \sin(kx), \tag{35}$$

even though u has only one nonzero Fourier frequency. For small enough a, it is the Gaussian smoothing of a positive bounded likelihood. A score itself need not have Gaussian Fourier decay; the cross-products are the required information bridge.

B.3 Origin normalization and weighted transfer

Set $\epsilon = E_n + \delta + e^{-cn}$. The origin constraint and the positive true-source bound give

$$u(0) = \widehat{u}(0) = 1, \qquad u \ge m/M, \qquad \nabla(\widehat{u}/u) = W/u^2. \tag{36}$$

Integrate along the line segment from zero to a point x in the query cube, then differentiate the product $u(\widehat{u}/u-1)$. Equation (34) yields

$$|\widehat{u}(x)-u(x)| \le C\epsilon|x|, \tag{37}$$

$$\|\nabla\widehat{u}(x)-\nabla u(x)\| + \left\|\widehat{V}(x)-V(x)\right\| \le C\epsilon(1+|x|). \tag{38}$$

For the second vector term use $B = u(\widehat{V}-V)+V(u-\widehat{u})$. Approximate origin normalization with error ξ_0 adds $C\xi_0$ to these estimates; it is not omitted from a numerical residual check.

Outside the cube, all smoothed functions and their required gradients have fixed bounds by (27). Consequently for any fixed nondegenerate Gaussian measure ν,

$$\|\widehat{u}-u\|_{L^2(\nu)} + \left\|\widehat{V}-V\right\|_{L^2(\nu)} \le C_\nu\epsilon + C_\nu e^{-c_\nu L_n^2}. \tag{39}$$

Gaussian weights absorb the factor $1+|x|$ in (38). Replacing that factor by its box maximum before integration would unnecessarily introduce $L_n\delta$ and would obscure linear stability at the Schur boundary.

B.4 Target transfer in the Schur-stable region

Conditioning the target Gaussian on its second active block gives, for bounded h, G,

$$T(h,G)(u,v) = \mathbb{E}_Z\big[G(v+Z)(H_S h)(u+DZ)\big], \qquad Z\sim N(0,I_r). \tag{40}$$

If $\Gamma \preceq S$, write $H_S = H_{S-\Gamma}H_\Gamma$. At the zero target mean the source error is evaluated at

$$Y = DZ+\xi, \qquad \xi\sim N(0,S-\Gamma), \qquad \mathrm{Cov}(Y) = I-\Gamma. \tag{41}$$

This Gaussian is well-defined even if $S-\Gamma$ is singular. Equation (38) within the cube and boundedness outside control the target difference by $C\epsilon + Ce^{-cL_n^2}$. Active u gradients use the source-gradient estimate. A v derivative can be moved onto the Gaussian density after changing variables $y = v+z$:

$$\nabla_v T(h,G)(u,v) = \int G(y)\phi(y-v)\big[(y-v)(H_S h)(u+D(y-v)) - D^\top\nabla H_S h(u+D(y-v))\big]dy. \tag{42}$$

Thus differentiability of G is not required. The additional Gaussian polynomial weights are integrable. The same argument applies to vector-valued numerators.

Use a telescoping decomposition for the denominator and each numerator, for example

$$T(F,G) - T(h_A,h_B) = T(F-h_A,G) + T(h_A,G-h_B), \tag{43}$$

$$T(J_A,G) - T(j_A,h_B) = T(J_A-j_A,G) + T(j_A,G-h_B). \tag{44}$$

The other factors are globally bounded, including auxiliary candidates. The true denominator is at least $w_{\min} = (m/M)^2$. If the approximate denominator is replaced by $\max\{\widehat{w}, w_{\min}/2\}$, its distance to w is no larger than $|\widehat{w}-w|$. A ratio difference then bounds the complete output by

$$\left\|\widehat{Z}-Z_R\right\| \le C(E_A+E_B+\delta+e^{-cn}). \tag{45}$$

Combined with (29) and $Q_n \asymp n^r$, this proves the stable-region upper bound of Theorem 2. Any feasible candidate obeys (45) with its actual residual; a solver's success flag is not needed, and is not sufficient.

B.5 Arbitrary positive source noise

Fix $0<\beta<\alpha_*$ and choose β' strictly between β and α_*. For one source, $\Delta = S-\beta'\Gamma \succ 0$. Choose a scalar P sufficiently large that $PI \succ \Gamma$. In an eigenbasis of Γ, write its eigenvalues as γ_j, and define

$$a_j = (1-\gamma_j/P)^{1/2}, \qquad b_j = (1-\beta'\gamma_j/P)^{1/2}. \tag{46}$$

Because $\log b_j/\log a_j \to \beta'$ as $P\to\infty$, we can fix P such that $b_j \le a_j^\beta$ for every j.

For a bounded source difference h, the function $H_\Gamma h((PI-\Gamma)^{1/2}z)$ is an anisotropic Ornstein–Uhlenbeck average of $h(\sqrt{P}\,z)$. In the orthonormal Hermite polynomial basis of Gaussian L^2, its multiplier is a^ν. The multiplier for $H_{\beta'\Gamma}$ is b^ν. Hölder's inequality applied to the squared coefficient sums therefore gives

$$\|H_{\beta'\Gamma}h\|_{L^2(N(0,PI-\beta'\Gamma))} \le \|h\|^{1-\beta}_{L^2(N(0,PI))}\,\|H_\Gamma h\|^{\beta}_{L^2(N(0,PI-\Gamma))}. \tag{47}$$

The calculation also holds for Hilbert-valued coefficients, hence for passive-vector differences. Boundedness controls the first norm; (39) controls the second. No Lebesgue L^2 assumption on the nonperiodic likelihood is introduced.

It remains to justify pointwise target evaluation and target gradients. In (40), apply the additional smoothing H_Δ and let $Y = DZ + \xi$ with $\xi \sim N(0,\Delta)$. Then $V_0 = \mathrm{Cov}(Y) = I - \beta'\Gamma \succ 0$. Write $\Lambda = PI - \beta'\Gamma$, enlarging the already fixed P if necessary to have $2V_0^{-1} - \Lambda^{-1} \succ 0$. For every fixed integer $k \ge 0$,

$$\int_{\mathbb{R}^r} \frac{\phi_{V_0}(y)^2}{\phi_\Lambda(y)}(1+|y|^k)\,dy < \infty. \tag{48}$$

The scalar integration kernel is bounded by a constant times ϕ_{V_0}. Its u- and v-gradient kernels carry weights $\Delta^{-1}\xi$ and $Z - D^\top\Delta^{-1}\xi$, respectively. Their conditional moments given $Y = y$ grow at most polynomially in $|y|$, since (Y, Z, ξ) is jointly Gaussian. Cauchy–Schwarz using (48) shows that the scalar target, active gradients, and passive-vector targets are bounded functionals on the left-hand Hilbert space of (47). This argument does not require S and Γ to commute.

The same telescoping and denominator estimates as above prove

$$\left\|\widehat{Z} - Z_R\right\| \le C_\beta(E_A + E_B + \delta + e^{-cn})^\beta. \tag{49}$$

Using feasible comparison residuals establishes the general upper statement of Theorem 2. Constants may diverge as $\beta \uparrow \alpha_*$. Explicit intermediate estimates do not by themselves constitute an implemented numerical error certifier.

C Information lower bound and the exact noise threshold

We prove the remaining assertions of Theorem 1. The finite construction above, with its order tending to infinity, gives a complete-field upper bound $C_\beta\delta^\beta$ for every $\beta < \alpha_*$. It gives $C\delta$ directly when both source covariances are at most S.

For the lower bound, make the likelihoods independent of passive coordinates and choose

$$f_\pm(u) = 1 \pm a\sin(k^\top u), \quad g(v) = 1 + b\cos(l^\top v), \quad l = Dk, \quad a = a_0\min\{1, \delta e^{k^\top\Gamma_A k/2}/|k|\}. \tag{50}$$

Here $k \in \mathbb{R}^r \setminus \{0\}$ is a real frequency, and fixed positive a_0, b are small enough that all likelihoods lie in $[m, M]$. Source smoothing multiplies the perturbation by $e^{-k^\top\Gamma_A k/2}$. Therefore

$$\|q_{A,\pm}\|_\infty \le \frac{a|k|e^{-k^\top\Gamma_A k/2}}{1 - ae^{-k^\top\Gamma_A k/2}} \le \delta \tag{51}$$

after reducing a_0 by a fixed factor. Both worlds accept the same zero source field for condition A and the same exact field for B.

At the target mean zero, parity makes the two joint denominators identical, equal to $1 + be^{-|l|^2/2}$. Their u-score difference is exactly

$$\frac{2ak}{1 + be^{-|l|^2/2}}\left[e^{-|k|^2/2} + be^{-(|k|^2+|l|^2)/2}\cosh(k^\top Dl)\right]. \tag{52}$$

There is no cancellation in the bracket. Since $l = Dk$ and $\cosh z \ge e^z/2$ for $z \ge 0$, completing the square yields

$$\|Z_+ - Z_-\| \ge c\min\left\{|k|e^{-k^\top Sk/2}, \delta e^{k^\top(\Gamma_A - S)k/2}\right\}. \tag{53}$$

Any deterministic reconstruction given the common input must err by at least half this separation in one world. The same complete-field information lower bound also holds for worst-case expected

error of randomized reconstructions, by the triangle inequality for their common output distribution. This observation is distinct from a randomized *finite-query* lower bound.

If $\Gamma_A \not\preceq S$, choose a direction v with $v^\top(\Gamma_A - S)v > 0$ and set $k = tv$. Taking δ_t proportional to $te^{-t^2 v^\top \Gamma_A v/2}$ makes the ratio in (53) to δ_t diverge exponentially. Thus linear stability is impossible. Repeat the argument for condition B. Conversely, both matrix inequalities imply the linear upper bound already proved, establishing the if and only if including equality.

For the critical exponent, minimize $v^\top S v/(v^\top \Gamma_A v)$ over nonzero v, or the analogous expression for B. When this minimum is below one, choose t on its minimizing ray so that the two branches in (53) meet. Then $t^2 \asymp \log(1/\delta)$ and the logarithm of the lower bound divided by $\log \delta$ tends to that Rayleigh quotient; powers of t are subleading on this scale. If both minima are at least one, a fixed-frequency small-amplitude perturbation gives $\Omega(\delta)$. Combining with every upper exponent $\beta < \alpha_*$ proves the stated logarithmic limit. For $\Gamma_A = \Gamma_B = I$, the smallest eigenvalue of S is $1 - \rho_1^2$.

Why the joint-likelihood specification is necessary. An unrelated identifiability obstruction exists without a prescribed clean joint likelihood. For instance, keep $p(A \mid X_0) = p(B \mid X_0) = 1/2$ but let $p(A, B \mid X_0 = x) = 1/4 + \epsilon h(x)$ with $|\epsilon h| < 1/4$. Changing h leaves both single-condition distributions unchanged and changes the joint distribution. This is not the obstruction proved above: our two worlds both obey the product-likelihood specification, so the difficulty persists even after arbitrary unknown joint semantics have been excluded.

D Deterministic adaptive lower bound for one output point

The following proof uses Hermite *functions in Lebesgue* L^2. The weighted interpolation in Appendix B.5 instead uses Hermite polynomials in Gaussian L^2. Their eigenvalues are not interchangeable.

D.1 Mehler representation and a nondegenerate derivative matrix

The density kernel of a scalar standard Gaussian pair with correlation $0 < \rho < 1$ is

$$K_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}\right]. \tag{54}$$

Set

$$b_\rho = (1-\rho^2)^{-1/2}, \quad \zeta_\rho = \frac{\rho}{1+\sqrt{1-\rho^2}}, \quad \phi_n^b(x) = \frac{b^{1/4} H_n(\sqrt{b}\, x) e^{-bx^2/2}}{\pi^{1/4}\sqrt{2^n n!}}. \tag{55}$$

The Mehler generating identity, after matching its quadratic exponent, gives

$$K_\rho(x, y) = \sum_{n \geq 0} \lambda_{0,\rho} \zeta_\rho^n \phi_n^{b_\rho}(x) \phi_n^{b_\rho}(y), \qquad \lambda_{0,\rho} = \frac{\sqrt{1-\zeta_\rho^2}}{2\sqrt{\pi}(1-\rho^2)^{1/4}}. \tag{56}$$

This is a classical kernel identity; the factor here is ζ_ρ^n, not ρ^n.

For each of r independent correlated pairs, use its own scale b_{ρ_j}. Let $d_N = N^r$. In the A trial space take odd Hermite functions ϕ_{2i+1} in the first coordinate and even functions ϕ_{2i} in all others; in the B space take even functions in every coordinate, with $0 \leq i < N$. The matrix of the scalar functional $\partial_{u_1} T(h, g)(0, 0)$ is a tensor product. By

$$(\phi_n^b)' = \sqrt{bn/2}\, \phi_{n-1}^b - \sqrt{b(n+1)/2}\, \phi_{n+1}^b, \tag{57}$$

its first factor has nonzero entries

$$(M_1)_{j,j} = \sqrt{b(2j+1)/2}\, \lambda_0 \zeta^{2j}, \qquad (M_1)_{j-1,j} = -\sqrt{b(2j)/2}\, \lambda_0 \zeta^{2j} \quad (j \geq 1). \tag{58}$$

Every other factor is $\mathrm{diag}(\lambda_0 \zeta^{2j})_{j=0}^{N-1}$. The placement of the superdiagonal in (58) follows by differentiating the odd A basis, not by differentiating an unrelated marginal kernel.

Remove the geometric column weights from M_1. The remaining upper-bidiagonal matrix has diagonal $\sqrt{b(2j+1)/2}$ and superdiagonal $-\sqrt{b(2j+2)/2}$. Its inverse entries are bounded by products of consecutive ratios at most $\sqrt{2}$ times a fixed inverse-diagonal bound. Thus its inverse norm is at most $CN2^{N/2}$. Restoring the smallest column weight and taking the fixed r-fold tensor product yields

$$\sigma_{\min}(M_N) \geq e^{-C_D N}. \tag{59}$$

The constant depends on the fixed nonzero correlations and on r. It is not uniform as a correlation vanishes.

D.2 A ZERO-ANSWER PATH FOR ANY DETERMINISTIC ADAPTIVE ALGORITHM

Fix an algorithm making at most Q calls, and follow the path on which every supplied attribute-score vector is zero. This fixes its adaptive query locations and which source is queried at each step. In our construction the likelihoods do not depend on passive coordinates, so one full-vector call imposes at most r nontrivial gradient constraints.

On the A trial coefficients require $\nabla H_{\Gamma_A} h = 0$ at every A query. Also require $\partial_{u_1} H_I h(0) = 0$, removing the single-likelihood target contribution. On B require $\nabla H_{\Gamma_B} g = 0$ at every B query. If $U_N, V_N \subset \mathbb{R}^{d_N}$ are the resulting coefficient spaces, then

$$\dim U_N + \dim V_N \geq 2d_N - rQ - 1 > d_N \quad \text{whenever } N^r > rQ + 1. \tag{60}$$

Since M_N is invertible, (60) implies $V_N \cap M_N^{-1} U_N \neq \{0\}$. Choose a unit v in this intersection and let $u = M_N v / \|M_N v\|$. Then

$$u \in U_N, \quad \|u\| = 1, \quad u^\top M_N v \geq \sigma_{\min}(M_N). \tag{61}$$

Thus the null space of the observed source information cannot also remove every target cross-interaction.

D.3 UNIFORMLY BOUNDED LIKELIHOODS AND EXACT TARGET SEPARATION

The trial functions must be bounded globally; bare Hermite polynomials would not meet the model assumptions. For Hermite functions, $\|\phi_n\|_2 = 1$ and $\|\phi_n'\|_2^2 = n + 1/2$. The one-dimensional Sobolev estimate $\|v\|_\infty^2 \leq 2\|v\|_2 \|v'\|_2$ gives $\|\phi_n^b\|_\infty \leq C_b(n+1)^{1/4}$. Tensor products and Cauchy–Schwarz for unit coefficient vectors therefore give

$$\|h\|_\infty, \|g\|_\infty \leq B_N := C_D N^{3r/4}. \tag{62}$$

Choose a fixed $\epsilon_0 > 0$ sufficiently small and define

$$f_\pm = 1 \pm \epsilon_0 h / B_N, \qquad g_* = 1 + \epsilon_0 g / B_N. \tag{63}$$

Both worlds lie in the same fixed interval $[m, M]$. Their source-score numerators vanish exactly at every queried point. Their denominators are positive, so the actual scores there are exactly the predetermined zero answers. Induction over the queries establishes the indistinguishability of the entire adaptive transcript.

Simultaneously changing the signs of the first correlated pair preserves its Gaussian density, changes h's sign, and preserves g_*. Hence $T(h, g_*)(0) = 0$, so the two target denominators coincide. The extra marginal constraint eliminates the linear contribution to the selected target numerator. By (61), the remaining score separation is at least

$$cB_N^{-2} \sigma_{\min}(M_N) \geq cN^{-3r/2} e^{-CN}. \tag{64}$$

One of the worlds causes error at least half this amount. Choosing the smallest integer N with $N^r > rQ + 1$ and absorbing the polynomial factor proves the lower bound in (7). Coupled with the finite construction, this gives the matching query order. A fixed-frequency $\Omega(\delta)$ information lower bound combined with this exact-query lower bound gives the left side of (8).

This proof fixes a hard input after following a *deterministic* zero-answer path. Choosing a different input for each random seed would not prove a randomized minimax lower bound. No such extension is claimed. The same construction tolerates any prescribed finite order of likelihood-derivative bounds by an additional polynomial scaling; this observation does not supply uniform analytic-class bounds or eliminate the fixed-parameter dependence.

E LOW-RANK REPLACEMENT AND CROSS-CORRELATION PERTURBATIONS

Here no active-coordinate periodicity or likelihood differentiability is needed. The two marginal Gaussian measures remain fixed and standard. Translate the target mean into the likelihoods; the bounds $[m, M]$ are preserved. Write γ_p, γ_q for standard Gaussian measures and define

$$(\mathsf{T}_R h)(u) = \mathbb{E}_R[h(V) \mid U = u]. \tag{65}$$

We distinguish this conditional-expectation operator from the scalar bilinear evaluation $T(h, g)$ used above.

E.1 EXACT OPERATOR TAIL AND THE FULL-SCORE BOUND

Orthogonal changes of variables reduce R to its singular-value diagonal. In normalized multivariate Hermite *polynomials* under Gaussian measure, T_R maps a paired multi-index to itself with multiplier $\prod_j \rho_j^{\nu_j}$; nonconstant unpaired modes are annihilated. This follows, for example, by comparing coefficients in

$$\mathbb{E}_R e^{a^\top U - |a|^2/2} e^{b^\top V - |b|^2/2} = e^{a^\top R b}. \tag{66}$$

Under the SVD truncation R_k, multipliers agree on every mode not involving a discarded direction, and vanish on the remaining modes. Each difference is at most ρ_{k+1}, and a first-order Hermite mode in that direction attains it. Hence

$$\|\mathsf{T}_R - \mathsf{T}_{R_k}\|_{L^2(\gamma_q) \to L^2(\gamma_p)} = \rho_{k+1}. \tag{67}$$

This is a direct application of classical Gaussian conditional-expectation structure, consistent with maximal-correlation inequalities (Nourdin et al., 2019).

Put $\tau = \rho_{k+1}$, $\lambda = 1 - \rho_1 > 0$, and $\kappa = (M/m)^2$. With w_R, b_R, n_R as in Section 5, boundedness and (67) give

$$|w_R - w_{R_k}| \le M^2 \tau. \tag{68}$$

For a unit vector $e = (a, b)$, the first-moment difference is

$$\begin{aligned} e^\top (b_R - b_{R_k}) = {} & \left\langle (a^\top U) f, (\mathsf{T}_R - \mathsf{T}_{R_k}) g \right\rangle_{L^2(\gamma_p)} \\ & + \left\langle f, (\mathsf{T}_R - \mathsf{T}_{R_k})[(b^\top V) g] \right\rangle_{L^2(\gamma_p)}. \end{aligned} \tag{69}$$

Since $\left\|(a^\top U) f\right\|_2 \le M \|a\|$, taking a supremum over unit e proves

$$\|b_R - b_{R_k}\| \le \sqrt{2} M^2 \tau. \tag{70}$$

For either correlation matrix $H \in \{R, R_k\}$,

$$w_H \ge m^2, \qquad \left\|\Sigma_H^{-1}\right\| \le \lambda^{-1}, \qquad \|n_H\| \le M^2 \lambda^{-1/2}. \tag{71}$$

The last estimate follows by testing each unit direction against the centered Gaussian score $e^\top \Sigma_H^{-1}(U, V)$ and using Cauchy–Schwarz.

Gaussian mean differentiation gives $n_H = \Sigma_H^{-1} b_H$. It follows that

$$n_R - n_{R_k} = \Sigma_R^{-1}(b_R - b_{R_k}) - \Sigma_R^{-1}(\Sigma_R - \Sigma_{R_k}) n_{R_k}. \tag{72}$$

Because $\|\Sigma_R - \Sigma_{R_k}\| = \tau$, (70)–(71) imply

$$\|n_R - n_{R_k}\| \le M^2 \tau \left(\frac{\sqrt{2}}{\lambda} + \frac{1}{\lambda^{3/2}} \right). \tag{73}$$

Taking the difference of $Z_H = n_H / w_H$ proves (14) with the explicit constant

$$K_{\mathrm{tr}} = \frac{\sqrt{2}\kappa}{\lambda} + \frac{\kappa}{\lambda^{3/2}} + \frac{\kappa^2}{\sqrt{\lambda}}. \tag{74}$$

The covariance-inverse term is essential: controlling only w_R or b_R would not yet control the full score. The absence of a dimension factor relies on separability of the tests in (70); it is not a bound on total variation or KL between the two entire Gaussian backgrounds.

E.2 A MATCHING LOWER BOUND FOR EVERY PRESCRIBED-RANK REPLACEMENT

Fix any $\widetilde{R}$ of rank at most k and norm below one. The singular-value min–max principle gives a unit $v \in \ker \widetilde{R}$ for which $\|Rv\| \geq \rho_{k+1}$. When $\rho_{k+1} > 0$, put $u = Rv/\|Rv\|$ and $c = u^\top Rv$. Choose fixed small $a, b > 0$ such that

$$f(U) = 1 + a\sin(u^\top U), \qquad g(V) = 1 + b\cos(v^\top V) \tag{75}$$

lie in $[m, M]$. The two scalar views have unit variances in both backgrounds. Their true correlation is $c \geq \rho_{k+1}$, while their replacement correlation is zero. At target mean zero, both normalizers equal $1 + be^{-1/2}$. Differentiating along the v direction of the second mean gives zero in the replacement and

$$-\frac{abe^{-1}\sinh c}{1 + be^{-1/2}} \tag{76}$$

in the true model. Since $\sinh c \geq c$, the score difference is at least $c_0\rho_{k+1}$ with $c_0 = abe^{-1}/(1 + be^{-1/2})$. The zero-tail case is immediate. Taking the infimum over all replacements, and using the SVD upper bound, proves (15).

Consequently, a rank-k Gaussian replacement can uniformly meet a target bias η only if $\rho_{k+1} \leq \eta/c_0$, and SVD truncation suffices if $\rho_{k+1} \leq \eta/K_{\mathrm{tr}}$. This characterizes the scale of an effective rank inside the stated replacement class. It does not prove the optimality of hard truncation among arbitrary nonlinear recovery procedures.

E.3 PERTURBATIONS WITHOUT A SINGULAR-VALUE GAP

Assume $\|R\|, \left\|\widetilde{R}\right\| \leq \bar{\rho} < 1$ and set $\varepsilon_R = \left\|R - \widetilde{R}\right\|$. On the nth homogeneous Gaussian chaos, (66) identifies the conditional expectation with the symmetric n-fold tensor map. Telescoping tensor factors gives

$$\left\|R^{\otimes n} - \widetilde{R}^{\otimes n}\right\| \leq n\bar{\rho}^{n-1}\varepsilon_R \leq \frac{\varepsilon_R}{1 - \bar{\rho}}. \tag{77}$$

The constant mode has zero difference. Orthogonality of the chaos decomposition therefore gives $\left\|\mathsf{T}_R - \mathsf{T}_{\widetilde{R}}\right\| \leq \varepsilon_R/(1 - \bar{\rho})$. Repeat the scalar and first-moment bounds above with this operator difference, while the covariance difference itself remains ε_R. With $\lambda = 1 - \bar{\rho}$,

$$\sup_{f,g,\mu} \left\|Z_R - Z_{\widetilde{R}}\right\| \leq K_{\mathrm{est}}\varepsilon_R, \qquad K_{\mathrm{est}} = \frac{\sqrt{2}\kappa}{\lambda^2} + \frac{\kappa + \kappa^2}{\lambda^{3/2}}. \tag{78}$$

No singular-vector accuracy or spectral gap is assumed. Fixed marginals are essential: errors in marginal whitening, source-coordinate transformations, or unknown likelihood blocks require separate control.

E.4 COMBINING ESTIMATION, TRUNCATION, AND QUERY ERRORS

Same-noise source marginals are unchanged by changing cross-correlation. Thus oracles for the original likelihoods also serve a rotated, truncated cross-correlation model without additional single-condition training. Apply (78), then (74) to $\widetilde{R}_k$, then (49). The triangle inequality gives (17). In the selective Schur interface use (45) instead, with power one. The actual diffusion-score error carries the factor from (4); its known background score is not changed.

This combination is valid for each chosen retained rank and fixed covariance regime. Choosing a retained rank that grows as the error tolerance decreases does not make $C_{k,\beta}$ a rank-uniform constant. In addition, merely estimating small node residuals is not an executable output certificate until all tail constants, source-error evidence, constraint errors, and floating-point effects have been included.

F FROM SCORE RECOVERY TO A DISCRETE SAMPLER AND A CLEAN TARGET

The following statements use standard Gaussian KL, change-of-measure, and transport tools. They close the error accounting for composition; they are not new general diffusion sampling-complexity results. Related diffusion discretization analyses include Chen et al. (2023).

F.1 AN EXACT IDENTITY FOR A JOINT-SCORE EULER REFERENCE

Let $T = t_0 > \cdots > t_K = a > 0$, $h_k = t_k - t_{k+1} > 0$, and let two Markov chains share their initial law. Their transition kernels at state x are

$$\widehat{K}_k(x,\cdot) = N(x + h_k \widehat{s}_{t_k}(x), h_k I), \qquad K_{*,k}(x,\cdot) = N(x + h_k s^{AB}_{t_k}(x), h_k I). \tag{79}$$

For deterministic measurable score functions, the Gaussian mean-shift formula gives

$$\mathrm{KL}(\widehat{K}_k(x,\cdot) \| K_{*,k}(x,\cdot)) = \frac{h_k}{2} \left\| \widehat{s}_{t_k}(x) - s^{AB}_{t_k}(x) \right\|^2. \tag{80}$$

The chain rule for relative entropy then proves (19), with extended value $+\infty$ when necessary. With different initial laws, add their KL in the same direction. Data processing yields

$$\mathrm{KL}(\widehat{P}_K \| P_{*,K}) \le \mathrm{KL}(\widehat{P} \| P_*). \tag{81}$$

The expectation in (19) is under the approximate chain, not the reference chain. No restriction $h_{\max} \le 1$ is needed, because this compares two discrete Gaussian-transition chains rather than a discretization to a continuous diffusion. An empirical trajectory average estimates the identity's right side but is not a certified upper bound without an uncertainty guarantee.

For paired sample sets, their pairing is a valid transport plan. Thus with $d^2_{\mathrm{pair}} = N^{-1} \sum_i \|Y_i - Y^*_i\|^2$,

$$W_2(\widehat{\nu}, \widehat{\nu}_*) \le d_{\mathrm{pair}}, \qquad \left| W_2(\widehat{\nu}, \widehat{\pi}) - W_2(\widehat{\nu}_*, \widehat{\pi}) \right| \le d_{\mathrm{pair}}. \tag{82}$$

Nearly identical empirical distances to the same reference are therefore not evidence independent of a tiny paired RMS difference. If an RMS is averaged over coordinates instead of vector norms, the corresponding dimension factor must first be restored.

F.2 A SEPARATE FINITE-STEP BOUND AGAINST THE CLEAN DISTRIBUTION

Let

$$\pi_0(dx) = \frac{f(x_U) g(x_V)}{Z_0} N(0, Q_0^{-1})(dx), \qquad \pi_t = \pi_0 * N(0, tI), \qquad \kappa = (M/m)^2. \tag{83}$$

Start the approximate Euler chain from the available background $p_T = N(0, Q_0^{-1} + TI)$ and use

$$Y_{k+1} = Y_k + h_k \widehat{s}_{t_k}(Y_k) + \sqrt{h_k} \xi_k. \tag{84}$$

Assume $h_{\max} \le 1$ and global errors $\left\| \widehat{s}_{t_k} - s^{AB}_{t_k} \right\|_\infty \le e_k \le 1$. For fixed a, T, Q_0, m, M, d, there is a finite constant C_{disc} such that the final law ν_{alg} satisfies

$$\mathrm{KL}(\nu_{\mathrm{alg}} \| \pi_a) \le \mathrm{KL}(p_T \| \pi_T) + \frac{3}{2} \sum_k h_k e_k^2 + C_{\mathrm{disc}} h_{\max}. \tag{85}$$

To justify the regularity needed here, write the noisy target as its Gaussian background density times the posterior expectation of the bounded product likelihood, divided by Z_0. On $[a, T]$ this expectation has a uniform positive lower bound. Its spatial derivatives are Gaussian Hermite-weighted integrals of a bounded function, and its time derivative includes the linear factor $A'_t x$. This gives constants B, L, H independent of the particular likelihoods for which

$$\|s_t(x)\| \le B(1 + \|x\|), \qquad \|s_t(x) - s_t(y)\| \le L \|x - y\|, \tag{86}$$

$$\|s_t(x) - s_s(x)\| \le H|t - s|(1 + \|x\|). \tag{87}$$

No derivative of the clean likelihood is assumed. A measurable approximate correction can be projected into a known ball containing the true correction without increasing its error, ensuring linear growth.

Let $u_k = T - t_k$ be reverse time and continuously interpolate (84) with frozen drift on each interval. Discrete Gronwall gives a uniform second moment M_2 and an increment bound $\mathbb{E} \|Y_{u_k+v} - Y_{u_k}\|^2 \le C_{\mathrm{inc}} v$ for $0 \le v \le h_k$. Split its drift error relative to the true reverse process into spatial movement, time movement, and the score-recovery error at the frozen state. Its squared expectation is at most

$$3L^2 C_{\mathrm{inc}} v + 6H^2(1 + M_2)v^2 + 3e_k^2. \tag{88}$$

Half its integral over the grid is bounded by the last two terms of (85), because $\sum_k h_k^2 \le h_{\max}(T-a)$ and $\sum_k h_k^3 \le h_{\max}^2(T-a)$.

For the change of measure use the approximate path relative to the exact reverse path. Stop when the path norm exceeds a finite radius and, after stopping, switch the auxiliary process to the true drift. On each stopped interval the difference is bounded, so Girsanov's formula applies. The uniform moment bound and continuous paths make the stopping probability vanish as the radius increases. Lower semicontinuity of relative entropy yields the unstopped inequality, and endpoint data processing gives (85). The uniform source-to-score error permits the required expectation under the approximate process; an unverified target-distribution L^2 error would not suffice for this step.

F.3 Initialization, transport, and early stopping

For any two initial measures with finite second moments, couple their initial variables and compare the Gaussian kernels with common covariance TI. Joint convexity of KL, followed by the infimum over couplings, gives

$$\mathrm{KL}(\nu * N(0,TI)\|\pi * N(0,TI)) \le \frac{W_2^2(\nu,\pi)}{2T}. \tag{89}$$

In particular $\mathrm{KL}(p_T\|\pi_T) \le W_2^2(p_0,\pi_0)/(2T)$. Since $\pi_0/p_0 \le \kappa$ and p_0 is centered, the independent coupling gives the explicit coarse bound $W_2^2(p_0,\pi_0) \le (1+\kappa)\,\mathrm{tr}(Q_0^{-1})$.

The ratio tilting $N(0, Q_0^{-1} + aI)$ into π_a has maximum-to-minimum ratio at most κ. Gaussian logarithmic Sobolev, bounded perturbation, and the implication from logarithmic Sobolev to T_2 (Gozlan et al., 2011; Otto and Villani, 2000) therefore imply

$$W_2^2(\nu,\pi_a) \le 2\kappa \left\|Q_0^{-1} + aI\right\|_{\mathrm{op}} \mathrm{KL}(\nu\|\pi_a). \tag{90}$$

This requires exactly the KL direction in (85). Finally, coupling the clean sample to its forward-noised version gives $W_2(\pi_0,\pi_a) \le \sqrt{da}$. Thus

$$W_2(\nu_{\mathrm{alg}},\pi_0) \le \sqrt{da} + \left\{2\kappa \left\|Q_0^{-1} + aI\right\|_{\mathrm{op}} \left[\frac{W_2^2(p_0,\pi_0)}{2T} + \frac{3}{2}\sum_k h_k e_k^2 + C_{\mathrm{disc}} h_{\max}\right]\right\}^{1/2}. \tag{91}$$

This closes error accounting to the original clean joint distribution rather than renaming the early-stopped distribution. The constants may be poor as $a \downarrow 0$, $T \uparrow \infty$, or dimension grows. This is neither an optimal total-computation guarantee nor a guarantee of hard semantic success. Finite-sample empirical distances do not certify (91).

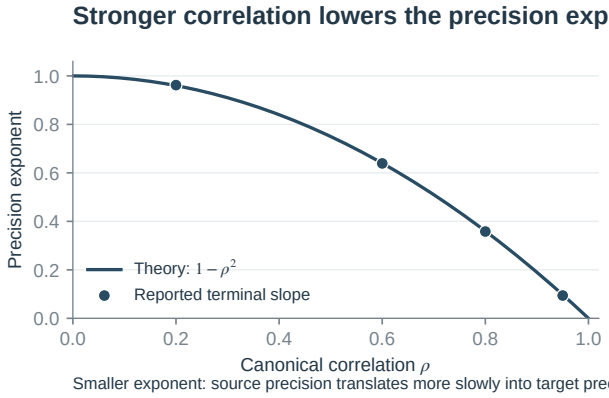
G Historical benchmark reports and baseline definitions

G.1 Evidence provenance and the historical comparison

This appendix preserves the historical numerical reports without treating them as case-level reproducible executions. The available artifacts contain the rounded benchmark aggregates and figure scripts, but not the complete original likelihood coefficients, target means, perturbation parameters, seeds, or trajectory records. The independent executable reference studies in Section 6 and Appendix J have their own fully specified inputs and outputs. They neither reconstruct the historical runs nor re-estimate their confidence intervals. The history is retained for context and to preserve adverse results; the primary reproducible empirical conclusions use the reference studies.

The historical Gaussian panels report 200 information-resolution cases, 7,040 rank-one recovery cases, 64 rank-two recovery cases, 552 spectral/replacement/perturbation cases, 16 targeted spectral witnesses, and 18 finite-query lower-bound witnesses. These are numerical cases, not training runs. The rank-two study reports mean error decreasing from 1.12×10^{-3} at 50 queries to 1.56×10^{-7} at 578 queries. The separate non-Gaussian scope reports appear in Appendix H.5.

The historical comparison in Table 3 uses eight attribute pairs and 16 means per pair. Its reported order-4 mean is 2.16×10^{-7}, with pair-cluster bootstrap 95% interval $[1.95, 2.38] \times 10^{-7}$, median



ρ	Predicted	Observed
0.20	0.96	0.962
0.60	0.64	0.639
0.80	0.36	0.358
0.95	0.0975	0.094

Figure 3: Stronger correlation reduces the same-noise precision exponent: a smaller exponent means slower translation of source accuracy into target accuracy. The table separates predicted exponents from reported terminal local slopes. The solid curve is the theorem, not an experimental fit. Each error budget uses its own two-world instance; the four markers summarize the 200-case resolution study, not a recovery curve for one fixed attribute pair.

Table 3: Historical exact-source panel (aggregate records only): eight attribute pairs, 16 means per pair, $d = 4$, $r = 1$, $\rho = 0.8$. Error is the mean full-vector Euclidean error. Queries are unique vector evaluations; time is the historical reported CPU implementation mean, not neural-network throughput.

Method	Queries	Score error	Time/point (ms)
Score addition	2	5.15×10^{-3}	0.096
JAC (full matrix)	10	6.40×10^{-4}	0.243
Full Gaussian overlap	26	1.89×10^{-4}	1.803
FQR, $n = 1$	10	1.67×10^{-3}	1.262
FQR, $n = 2$	18	3.99×10^{-5}	1.240
FQR, $n = 4$	34	2.16×10^{-7}	1.553
FQR, $n = 32$	258	3.90×10^{-8}	20.658

2.12×10^{-7}, and maximum 3.45×10^{-7}. The reported 875-fold advantage over full overlap belongs to this historical panel, not to the independent reference study. The bootstrap resamples attribute pairs, retaining their positions. No bootstrap replicates or case-level distributions are inferred from these aggregates. Passive-dimension panels hold total amplitude and direction norms fixed; rank two is a separate dimension-six study.

G.2 Strong Gaussian baselines and fair information access

Score addition uses $s_A + s_B - s_0$. JAC and the complete Gaussian-overlap gradient follow the Gaussian backward-posterior construction of Linhart et al. (2026). To fix the mathematical baseline, let (m_0, C_0) be the background denoising-posterior moments and $(m_A, C_A), (m_B, C_B)$ the corresponding single-condition moments. For additive variance-time Gaussian observations, the exact moment identities are

$$m_A = x + ts_A(x,t), \qquad C_A = tI + t^2 D_x s_A(x,t), \tag{92}$$

and similarly for B and the background. The moment identities do not assert that the conditionals themselves are Gaussian. Define

$$P_i = C_i^{-1}, \quad \Lambda = P_A + P_B - P_0, \quad h = P_A m_A + P_B m_B - P_0 m_0, \quad m_J = \Lambda^{-1} h. \tag{93}$$

When the fused precision is positive definite, JAC uses $(m_J - x)/t$. The full Gaussian overlap instead differentiates the entire scalar approximation

$$\mathcal{O}(x) = \int \frac{N(z; m_A, C_A) N(z; m_B, C_B)}{N(z; m_0, C_0)}\, dz \tag{94}$$

and adds $\nabla_x \log \mathcal{O}$ to score addition. Explicitly,

$$\log \mathcal{O} = \tfrac{1}{2}\big(\log\det C_0 - \log\det C_A - \log\det C_B - \log\det \Lambda\big) - \tfrac{1}{2}\big(m_A^\top P_A m_A + m_B^\top P_B m_B - m_0^\top P_0 m_0 - h^\top \Lambda^{-1} h\big). \tag{95}$$

Its derivative includes the spatial changes in conditional covariances. This term, and high-order corrections to Gaussian approximations more generally, are not new contributions of this work (Linhart et al., 2026; Opper et al., 2013).

The compared implementations use the same single-condition score values, cache finite-difference points, and report unique vector evaluations. The main finite-difference step is 10^{-4}, selected on independent validation attributes. Baseline identities were checked on quadratic-log likelihoods against exact Gaussian targets. The global-covariance GAUSS variant was not implemented; the comparisons are not a claim to dominate every compositional sampling method.

A separate high-precision diagnostic differentiates the analytic sources, including all covariance derivatives of (95), at three selected targets. At 45-digit precision the largest vector differences from the ordinary implementations are 8.64×10^{-12} for JAC and 1.84×10^{-11} for full overlap, whereas baseline target errors at those points remain approximately 10^{-4}–8×10^{-4}. This rules out finite-difference inaccuracy as the main explanation at these points. The high-precision derivative calculation is an accuracy check, not an extra same-cost baseline.

G.3 Finite-query candidates and numerical feasibility

The practical method uses a truncated-SVD feasible coefficient candidate, not a certified solution of the minimax SOCP. It records the actual maximum vector residual, normalization discrepancy, and finite auxiliary-grid constraints. A full-vector passive norm is checked as a Euclidean norm; coordinatewise boxes would not establish the same condition. Scalar bounds require both signs. Symmetry-duplicate source points are cached, and a fixed error field returns the same perturbation at repeated locations.

The historical 192-case SOCP screening reference contains 64 rows with the solver state `optimal_inaccurate`. Constraints were checked separately and passed numerically. This is not exact optimality certification, and the reference's target error is not always smaller than the SVD candidate's. A variable-projection prototype in the historical report produced an approximately 7×10^{-5} full-vector error in a pure-passive diagnostic and was superseded; those screening outputs are not part of the historical recovery table. The retained implementation's reported check is approximately 1.07×10^{-6}, with approximately 0.198 error when J is removed in that particular check. These numbers are from a separate diagnostic and are not substituted for the main ablation mean near 0.102.

The implementation reports `residual_checked` and measured error against analytic truth. It does not report `numeric_bound`: the executable constants, region tails, source-error certification, and floating-point enclosures required for a calibrated output bound remain incomplete. A tiny node residual cannot by itself certify a tiny target error because representation and region-tail terms persist.

G.4 Resolution experiments and the scope of construction checks

The source budgets range from 10^{-2} to 10^{-64}. Different budgets generally correspond to different hard likelihood pairs. Stable-region verification includes a fixed frequency with linearly decreasing amplitude: using only a moving high-frequency family could mistakenly suggest a critical exponent greater than one. The terminal slope comparisons are

ρ	**0.20**	**0.60**	**0.80**	**0.95**
Theory $1-\rho^2$	0.960	0.640	0.360	0.0975
Reported local slope	0.962	0.639	0.358	0.094

These checks assess the analytic mechanism, not the recoverable precision of a double-precision score network. The 18 90-digit adaptive-lower-bound witnesses use query counts $2, 4, 6, 8, 10, 12$. Each numerical witness corresponds to a fixed query schedule, with residual below 10^{-55} and positive

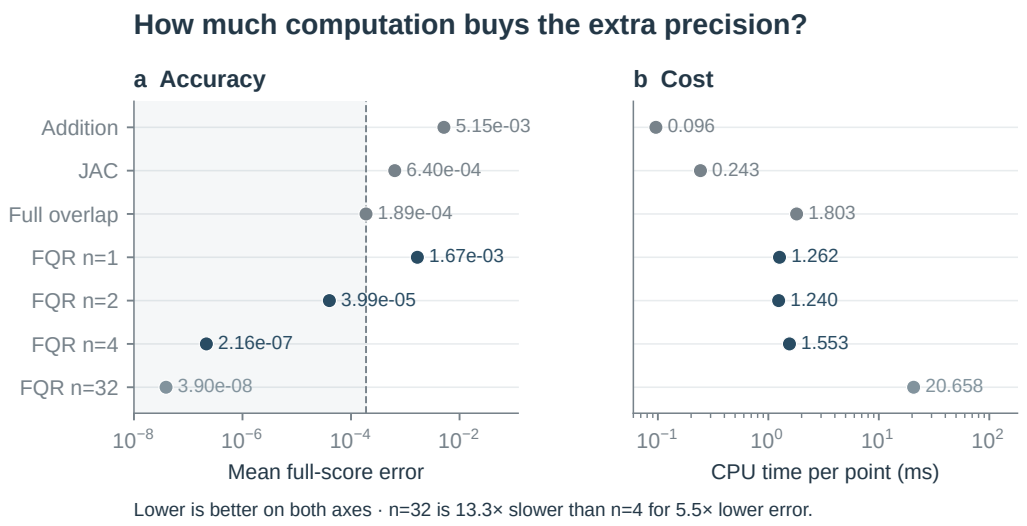


Figure 4: Exact-source accuracy and CPU cost on aligned method rows; lower is better on both axes. The dashed line is the smallest reported Gaussian-baseline error, and the region to its left marks lower errors. Points and labels are aggregates from Table 3, not a fitted Pareto frontier or neural-model throughput. Relative to $n = 4$, the $n = 32$ candidate uses 13.3 times the CPU time for 5.5 times lower mean error. These are reported implementation means, not a universal efficiency law; $n = 1$ does not beat the best baseline.

target separation. Universality over deterministic adaptive algorithms follows from Appendix D, not from testing those schedules.

The 16 targeted spectral witnesses exhibit a score-error-to-ρ_{k+1} ratio in $[0.01597, 0.01602]$. Random attributes may have cancellations and much smaller errors. The explicit spectral upper bounds hold on the tested cases but have loose constants. Tests involving nonaligned rank replacements and cross-correlation perturbations check the distinctions in Theorem 3; they are not a certification over all likelihoods.

G.5 Representation and plateau diagnostics

The historical full-vector ablation reports mean error near 0.102 when J is discarded, and a plateau near 5.6×10^{-6} when the query box is fixed. These are distinct from the passive-output ablation in the executable reference grid, whose model and target means differ.

Three diagnostic targets are $(\text{seed}, \text{point}) = (0, 0), (1, 9), (4, 15)$. The first is ordinary; the other two were selected for the largest $n = 8/n = 4$ rebound and the largest $n = 32$ error in the original panel. They are intentionally chosen diagnostics, not new independent statistical samples. The intervention grid contains 252 candidates (three targets, four orders, three region scales, seven truncation thresholds), and 36 records selected by the pre-existing rule. Six infeasible diagnostic candidates are retained in the record and excluded from selection. Twelve original-configuration results reproduce their earlier values.

The 55- and 80-digit joint references differ by at most 4.03×10^{-57}; the double-precision references differ from high precision by at most 5.80×10^{-17}. The historical panel does not activate denominator clipping. At fixed region scale 0.5 and SVD threshold 10^{-12}, reconstructing the values, matrices, SVD, and target at 55 digits does not remove the observed plateau:

Target	$n = 4$ **error**	$n = 8$ **error**
Ordinary diagnostic	1.9333×10^{-7}	2.9327×10^{-7}
Rebound diagnostic	1.3718×10^{-7}	5.3203×10^{-7}

At the same $n = 4$ and 34 queries, changing `box_scale` from 0.5 to 1 increases the radius by $\sqrt{2}$ and changes the Fourier period. Errors at the three targets change from $1.93 \times 10^{-7}, 1.37 \times 10^{-7}, 1.58 \times 10^{-7}$ to $3.64 \times 10^{-10}, 2.40 \times 10^{-10}, 1.82 \times 10^{-10}$. Smaller truncation thresholds also change the $n = 8$ error, but expose differences between double and high precision. These interventions implicate representation geometry and truncation, not a pure reference or machine-precision floor. They do not isolate each theoretical error term, because changing the region also changes the frequency representation.

On the historical panel, $n = 4$ has median, 95th percentile, and maximum errors 2.12×10^{-7}, 2.87×10^{-7}, and 3.45×10^{-7}. For $n = 2$ and $n = 4$, both Gaussian baselines are beaten at every one of the 128 positions and in every attribute-pair mean. The $n = 1$ win rate is only 3.125% against JAC and zero against full overlap. Thus the study does not establish pointwise improvement at every budget or monotonicity in order.

G.6 All methods see the same fixed source-error field

The supplemental noise panel fixes four original attribute pairs and four positions per pair. The exact source and low/high-frequency perturbations are shared across all methods, with Euclidean budget δ for each source. Repeated coordinates return identical errors. There are 640 method-result rows with no failures. Baseline difference steps are separately checked on distinct validation attributes/positions over $\{10^{-4}, 10^{-3}, 10^{-2}, 0.03\}$; 10^{-4} is selected in every tested setting.

Table 4: Complete reported fixed-field noise panel. All entries are mean full-vector errors over the same 16 targets. This is a different panel from Table 3.

Method	**Exact**	**Low** 10^{-8}	**High** 10^{-8}	**High** 10^{-6}
Addition	5.30×10^{-3}	5.30×10^{-3}	5.30×10^{-3}	5.30×10^{-3}
JAC	5.16×10^{-4}	5.16×10^{-4}	5.16×10^{-4}	5.16×10^{-4}
Full overlap	1.77×10^{-4}	1.77×10^{-4}	1.77×10^{-4}	1.77×10^{-4}
FQR, $n = 2$	4.09×10^{-5}	4.09×10^{-5}	4.08×10^{-5}	5.63×10^{-5}
FQR, $n = 4$	2.05×10^{-7}	2.06×10^{-7}	7.37×10^{-4}	4.28×10^{-3}
FQR, $n = 8$	2.80×10^{-7}	2.82×10^{-7}	2.36×10^{-4}	1.80×10^{-2}

The exact-source advantage of a high-order candidate can disappear under a small high-frequency perturbation. A low-order candidate retains more approximation bias but is less sensitive in this panel. Theorem 1 does not establish a matching lower bound for these particular perturbations; the observed amplification may include avoidable numerical regularization effects. These results motivate noise-aware representation and regularization selection, which is not a completed method contribution of this paper.

G.7 Sampling batches and the right error comparison

The original sampling batch uses a two-dimensional Gaussian background, three seeds, 1,024 samples per seed, a 64-step geometric variance-time grid from $T = 20$ to $a = 0.01$, and background p_T initialization. With shared random numbers, $n = 4$ has mean paired RMS 3.24×10^{-7} to the exact-score Euler sampler; both have mean empirical W_2 about 0.17545, versus 0.17831 for addition. The reported $n = 4$ time is about 56 seconds per group, versus 0.055 for the analytic joint oracle and 0.011 for addition. These implementation paths have different vectorization opportunities; none establishes a neural throughput advantage.

The supplemental batch couples all coarse Brownian increments to one 512-step path and shares initial states. This changes the original random-number mapping, so its empirical distances are a separate batch, not improvements obtained by modifying an algorithm. It includes three seeds and 1,024 samples for each of six methods/configurations: exact-score Euler at 64, 256, and 512 steps, and addition, $n = 2$, and $n = 4$ at 64 steps. All 18 groups complete without fallback.

Independent reference-to-reference distances are 0.14673–0.15560, and cannot be subtracted from Table 6 as an error correction. Coupled exact-score RMS differences are 0.04254 for 64 versus 256

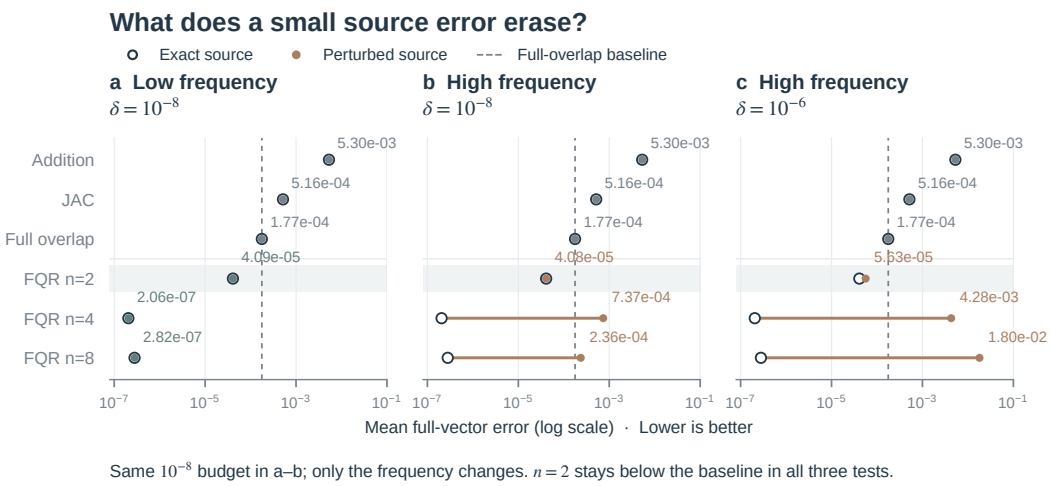


Figure 5: Historical fixed-source-error comparison (aggregate records only) (four attribute pairs, four positions each). Each segment joins a method's exact-source error (open marker) to its perturbed error (filled marker), on a shared logarithmic scale. Panels (a,b) hold the vector error budget at 10^{-8} while changing frequency; (c) raises the high-frequency budget. Left of the dashed full-overlap baseline is better. High-order precision is largely lost in (b,c), while $n = 2$ remains below the baseline. Points are reported means, not reconstructed trials; repeated queries return the same error field.

Table 5: Historical coupled sampling report (aggregate records only): 3,072 trajectories, 64 Euler steps, $T = 20$, $a = 0.01$, and background initialization. Energy estimates the right side of (19) along the approximate chain. Standard errors use independent trajectories, not time points. None is a certified upper bound.

Method	**Mean energy**	**MC standard error**	**Paired RMS to P_***
Score addition	3.2083×10^{-4}	1.79×10^{-6}	2.2259×10^{-2}
FQR, $n = 2$	1.2047×10^{-8}	9.03×10^{-11}	1.0235×10^{-4}
FQR, $n = 4$	3.0022×10^{-13}	1.93×10^{-15}	3.2533×10^{-7}

steps and 0.007712 for 256 versus 512. The 64-step grid violates $h_{\max} \leq 1$ in (85); the finer grids satisfy that single condition, not automatically its source-error or other assumptions.

For each approximate trajectory the supplemental code accumulates

$$\mathcal{E}_i = \tfrac{1}{2} \sum_k h_k \left\| \widehat{s}_{t_k}(X_k^{(i)}) - s_{t_k}^{AB}(X_k^{(i)}) \right\|^2 \tag{96}$$

at the approximate chain's actual original-coordinate states. Its expectation is precisely (19). Table 5 uses 3,072 independent trajectory values, with Monte Carlo standard errors computed across trajectories rather than across time points. There is no need to satisfy $h_{\max} \leq 1$ for this discrete-reference identity.

About 90.8% of the $n = 4$ energy comes from $t > 5$, but its total is already approximately 3×10^{-13}; the fraction alone does not indicate a large practical high-noise failure. The reported $n = 2$ and $n = 4$ times, including trajectory diagnostics, are approximately 52.1 and 56.9 seconds per group. Their matched-RNG RMS values measure substitution for the reference chain. The clean-target empirical W_2 differences (addition minus exact-score 64 steps) are $-0.001597, +0.001393, +0.003393$ across the seeds. They do not establish a significant clean-distribution improvement. All saved pairings satisfy (82); close empirical W_2 values are not independent validation beyond the close paired samples.

Table 6: Exact-score grid refinement in the supplemental batch. The empirical distances include finite-reference sampling error; they are not certified population distances.

Steps	$h_{\max}$	**Empirical** W_2 **to** π_0	**Empirical** W_2 **to** π_a
64	2.23965	0.162066	0.170633
256	0.58509	0.150768	0.158376
512	0.29472	0.149423	0.156798

G.8 Reported execution environment and reproducibility boundaries

The report records NumPy 2.4.6, SciPy 1.18.1, and mpmath 1.3.0, with `OPENBLAS_NUM_THREADS=1` and `OMP_NUM_THREADS=1`. The aggregate report does not specify the CPU model; it is therefore not inferred from the timing values. High-precision checks use the decimal precisions stated above, whereas the practical historical recovery comparisons use double precision. Solver tolerances, case-specific coefficients, and all original configurations require the execution archive for exact reconstruction.

The historical reports retain failures and limitations: nonmonotone order curves, source-noise sensitivity, solver `optimal_inaccurate` states, infeasible diagnostic candidates, and methods/budgets where JAC wins. No raw distributions or bootstrap replicates have been synthesized for presentation. The historical quantitative figures read only the accompanying reported aggregate tables. Figure 6 is replotted from the reported aggregate table in `data/`, not reconstructed from missing raw observations. The work does not evaluate a trained conditional score, image quality, unknown likelihood views, arbitrary non-Gaussian backgrounds, or randomized-query lower bounds.

H Background misspecification and rare conditions

This section separates two assumptions that play different roles. Gaussianity specifies a solvable dependence channel; a positive likelihood floor controls conditioning. The following results strengthen the scope analysis, but do not replace the hypotheses of Theorems 1–2. In particular, the hard-condition result below concerns low-rank *replacement bias*, not a finite-query recovery theorem for rare events.

H.1 A dimension-independent transfer for bounded non-Gaussian tilts

Let $q_0 = N(0, Q_0^{-1})$ be the known reference background, and let the actual background be

$$p_0(dx) = \frac{e^{V(x)}}{\mathbb{E}_{q_0} e^V}\, q_0(dx), \qquad \operatorname{osc}(V) := \operatorname{ess\,sup} V - \operatorname{ess\,inf} V \le \varepsilon. \tag{97}$$

The potential may be nonsmooth and need not separate over the two blocks. For a measurable likelihood F with $0 < \ell \le F \le u < \infty$, write $s^p_{F,t}$ and $s^q_{F,t}$ for the scores of the respective normalized F-tilted backgrounds after adding $N(0, tI)$ noise. Set $C_t = (Q_0 + t^{-1}I)^{-1}$ as before.

Proposition 4 (Bounded-potential score transfer)**.** *For every $t > 0$,*

$$\sup_x \left\| s^p_{F,t}(x) - s^q_{F,t}(x) \right\| \le \sqrt{u/\ell}\,(e^{\varepsilon} - 1)\frac{\sqrt{\|C_t\|_{\mathrm{op}}}}{t}. \tag{98}$$

There is no trace or ambient-dimension factor, and no derivative bound on V or F is required.

Proof. Fix x. The reference backward distribution before conditioning is $G_x = N(A_t x, C_t)$. Its F-tilt, denoted $Q_{x,F}$, has density at most u/ℓ with respect to G_x. The actual F-conditioned backward law $P_{x,F}$ has density ratio

$$r(y) = \frac{e^{V(y)}}{\mathbb{E}_{Q_{x,F}} e^V} \in [e^{-\varepsilon}, e^{\varepsilon}]$$

with respect to $Q_{x,F}$. Since $\mathbb{E}_Q(r-1)=0$, for any unit vector e,

$$|e^\top(\mathbb{E}_P Y - \mathbb{E}_Q Y)| = |\mathbb{E}_Q[(r-1)e^\top(Y-A_t x)]| \\ \le (e^\varepsilon-1)\sqrt{\mathbb{E}_Q[(e^\top(Y-A_t x))^2]} \le (e^\varepsilon-1)\sqrt{(u/\ell)e^\top C_t e}.$$

Taking the supremum over unit e and applying the Gaussian observation identity $s_{F,t}(x) = (\mathbb{E}[Y \mid x, F] - x)/t$ proves the claim. Importantly, the supremum over directions is taken *after* the scalar second-moment bound; bounding the full squared displacement first would introduce an unnecessary $\operatorname{tr} C_t$. □

Transfer to the actual score interface. This result compares genuine diffusion scores under the two backgrounds, not separately provided Gaussian-mixture component scores. Let $L_{\mathrm{in,t}}$ be the largest norm of the reference source projection in Appendix A, and let $L_{\mathrm{out,t}} = \|W_t A_t\|_{\mathrm{op}}$. The algorithm receives the actual conditional scores $s^p_{A,t}, s^p_{B,t}$ with raw vector error at most δ_s, subtracts the *known reference* background score $s^q_{0,t}$, and uses the reference maps. For $m \le f, g \le M$, define

$$b_t = (e^\varepsilon - 1)\frac{\sqrt{\|C_t\|_{\mathrm{op}}}}{t}, \qquad \delta_{\mathrm{eff}} = L_{\mathrm{in,t}}\big(\delta_s + \sqrt{M/m}\, b_t\big).$$

Whenever δ_{eff} lies in the fixed-parameter regime of Theorem 2, an exact-reference-rank-r construction obeys

$$\left\|\widehat{s}_t - s^p_{AB,t}\right\| \le (M/m)b_t + L_{\mathrm{out,t}} C_\beta\big(\delta_{\mathrm{eff}}^\beta + e^{-c_\beta Q^{1/r}}\big), \quad 0<\beta<1-\rho_1^2. \tag{99}$$

For a residual-checked candidate, use the residual version of Theorem 2; any reference rank-truncation term is added separately. The formula retains the reference background in the output. It does not silently substitute the unknown actual Gaussian parameters. For fixed $\varepsilon > 0$, this upper bound retains a model-mismatch term. It does not prove a positive asymptotic error floor or the optimal recovery rate for the misspecified problem. The result is a perturbation guarantee, not a continuation of the sharp Gaussian laws to arbitrary backgrounds.

H.2 Why covariance alone cannot replace the Gaussian hypothesis

A fixed-time kernel example identifies information absent from covariance. Let

$$K = \tfrac{1}{2}N(0,\Sigma_\rho) + \tfrac{1}{2}N(0,\Sigma_{-\rho}), \qquad \Sigma_\rho = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad 0<\rho<1,$$

and compare it with $K_0 = N(0, I_2)$. Both kernels have standard Gaussian marginals and covariance I_2. Consequently their complete single-factor *marginal convolution* score fields agree. For the fixed positive factors

$$f(u) = 1 + a\sin u, \qquad g(v) = 1 + b\cos v, \qquad 0<a,b<1,$$

consider $Z_K(\mu) = \nabla_\mu \log \mathbb{E}_K[f(\mu_U + U)g(\mu_V + V)]$. Direct Gaussian Fourier integration yields

$$[Z_K(0) - Z_{K_0}(0)]_U = \frac{abe^{-1}(\cosh\rho - 1)}{1 + be^{-1/2}} > 0. \tag{100}$$

Indeed, both denominators are $1 + be^{-1/2}$, while $\mathbb{E}_K[\cos U \cos V] = e^{-1}\cosh\rho$ and $\mathbb{E}_{K_0}[\cos U \cos V] = e^{-1}$. This is a counterexample to replacing the known Gaussian kernel by its covariance in a general convolution problem. It is *not* an assertion that these translation families are the backward laws of one Brownian diffusion at every location, or that complete genuine diffusion score fields under the two backgrounds are identical. If the full non-Gaussian background is known, that is additional information; our Gaussian covariance-only formulas do not use it.

H.3 Low-rank bias with hard conditions and local evidence

The positive floor in the main theorems is not a smoothness assumption. It prevents conditioning on a vanishing normalizer at every query and target location. For the replacement theorem alone, one can replace it by a *local joint-evidence condition*. Fix μ, take $0 \le f, g \le 1$, and let

$$z_R = w_R(\mu) = \mathbb{E}_R[f(\mu_U + U)g(\mu_V + V)].$$

Indicators are permitted, in which case z_R is a joint event probability under the background posterior. For likelihoods bounded by another known constant, first divide by their upper bounds; this does not change scores.

Proposition 5 (Evidence-conditioned spectral truncation)**.** *Suppose $\rho_1 \le \bar{\rho} < 1$, $z_R \ge \zeta > 0$, and $\tau = \rho_{k+1} \le \zeta/2$. Put $\lambda = 1 - \bar{\rho}$. Then $w_{R_k}(\mu) \ge \zeta/2$ and*

$$\|Z_R(\mu) - Z_{R_k}(\mu)\| \le \frac{\tau}{\zeta}\left[\frac{\sqrt{2}}{\lambda} + \left(1 + \frac{3}{2\lambda}\right)\sqrt{\frac{2\log(2/\zeta)}{\lambda}}\right]. \tag{101}$$

The constants do not depend on block dimensions. The same statement is uniform on any region where $z_R \ge \zeta$. No matching dependence on ζ is claimed.

Proof. Write $S = R_k$, $w_T = \mathbb{E}_T fg$, $b_T = \mathbb{E}_T[(U,V)fg]$, and $n_T = \Sigma_T^{-1} b_T$ for $T = R, S$, absorbing the fixed translation into the factors. The operator proof of Theorem 3 does not require a positive lower bound before taking ratios. It gives

$$|w_R - w_S| \le \tau, \qquad \|b_R - b_S\| \le \sqrt{2}\tau.$$

Thus $w_S \ge \zeta/2$ and $w_S/w_R \le 3/2$.

We also need a sharper score bound than dividing a uniform numerator by a rare probability. If $G = N(0,\Sigma)$, $0 \le L \le 1$, and $w = \mathbb{E}_G L > 0$, the tilted law $P = LG/w$ satisfies $\mathrm{KL}(P\|G) \le \log(1/w)$. The entropy variational inequality applied to $a^\top Y$, and the Gaussian moment-generating function, give

$$\tfrac{1}{2}(\mathbb{E}_P Y)^\top \Sigma^{-1}(\mathbb{E}_P Y) \le \mathrm{KL}(P\|G).$$

This inequality also follows by substituting $a = \Sigma^{-1}\mathbb{E}_P Y$ into $a^\top \mathbb{E}_P Y \le \mathrm{KL}(P\|G) + a^\top \Sigma a/2$. Consequently

$$\|\nabla \log w\| \le \sqrt{2\log(1/w)/\lambda_{\min}(\Sigma)}. \tag{102}$$

All moments exist because $dP/dG \le 1/w$. This argument differentiates the Gaussian kernel, not an indicator.

Using the inverse-covariance identity (16) and then the ratio identity gives

$$\begin{aligned}\|Z_R - Z_S\| &\le \frac{\sqrt{2}\tau}{\lambda w_R} + \tau\left(\frac{w_S}{\lambda w_R} + \frac{1}{w_R}\right)\|Z_S\| \\ &\le \tau\left[\frac{\sqrt{2}}{\lambda\zeta} + \left(\frac{3}{2\lambda} + \frac{1}{\zeta}\right)\sqrt{\frac{2\log(2/\zeta)}{\lambda}}\right].\end{aligned}$$

Since $\zeta \le 1$, this implies (101). □

What this extension does not grant. Target evidence does not bound the source normalizers over an expanding query box. Proposition 5 therefore does not remove $m > 0$ from Theorems 1 and 2, nor from their constants. A score field also fixes a likelihood only up to scale; obtaining a usable evidence bound can require model information not supplied by scores alone. We do not introduce likelihood flooring as a free score-only operation.

H.4 A RARE-EVENT OBSTRUCTION EVEN WHEN THE DISCARDED CORRELATION TENDS TO ZERO

The dependence on conditioning cannot be eliminated merely by improving the proof. For $L \to \infty$, set

$$f_L(u) = \mathbf{1}_{\{u \ge L\}}, \qquad g_L(v) = \mathbf{1}_{\{v \ge L\}}, \qquad R_L = 1/L, \qquad \widetilde{R}_L = 0.$$

These are legitimate hard factors with positive, but vanishing, Gaussian joint probability. At the target mean zero,

$$\left\|R_L - \widetilde{R}_L\right\|_{\mathrm{op}} \longrightarrow 0, \qquad \|Z_{R_L}(0) - Z_0(0)\| \longrightarrow \sqrt{2}. \tag{103}$$

Thus no uniform modulus tending to zero with covariance error can hold over all bounded nonnegative factors without a conditioning restriction.

Proof. By symmetry let $m_{L,\rho} = \mathbb{E}[U \mid U \geq L, V \geq L]$. Gaussian mean differentiation gives

$$[Z_\rho(0)]_U = [Z_\rho(0)]_V = m_{L,\rho}/(1+\rho).$$

For $0 \leq \rho \leq 1/2$, change variables to $U = L + x/L$, $V = L + y/L$. The conditional excess density on $x, y \geq 0$ is proportional to

$$\exp\left[-\frac{x+y}{1+\rho} - \frac{x^2+y^2-2\rho xy}{2L^2(1-\rho^2)}\right].$$

For $L \geq 2$, its normalization is bounded below by integrating over $[0,1]^2$, uniformly in L, ρ. Its first x moment is bounded above by integrating $xe^{-2(x+y)/3}$ over the quadrant. Hence $m_{L,\rho} = L + O(L^{-1})$ uniformly. The one-dimensional tail mean likewise equals $L + O(L^{-1})$. At $\rho = 1/L$, each score-component difference is

$$\frac{L}{1+1/L} - L + O(L^{-1}) \longrightarrow -1.$$

This proves (103). □

This is a worst-case statement with the factors changing with L, not instability of every rare event. Conversely, the main information and deterministic-query lower bounds already hold for factors arbitrarily close to one: for each fixed $\epsilon > 0$, their constructions may be scaled into $[1-\epsilon, 1+\epsilon]$. Thus the resolution obstruction exists without rare conditioning, while rare conditioning creates a separate obstruction.

Separate algebra and quadrature checks. The executable script `revision_checks/check_scope_results.py` evaluates the scope identities and bounds separately from the recovery benchmarks. It checks 11 evidence-qualified hard-rectangle truncations, six rare-event tail instances, 54 bounded-potential score comparisons, and three equal-covariance kernel witnesses. The potential comparisons use independent 48- and 80-node-per-coordinate Gaussian quadrature; their maximum score-difference discrepancy is below 2×10^{-12}. The rare-event calculations use 65-digit arithmetic and scaled excess integrals rather than subtraction of tiny tail probabilities. These are finite numerical checks of identities and bounds, not statistical validation, a uniform numerical certificate, or an evaluation of FQR on learned/non-Gaussian sources.

H.5 Reported recovery and scope diagnostics

This subsection preserves historical deterministic CPU recovery and scope summaries. Their FQR execution archive is not included; the provenance boundary is the same as in Appendix G. The executable calculations in Appendix I.5 evaluate the corresponding analytic scope results, not these FQR outputs.

Bounded-potential design and score access. The reference precision and actual perturbation are

$$Q_0 = \begin{pmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}, \qquad V_\varepsilon(u,v) = \frac{\varepsilon}{2}\sin(uv),$$

with $f(u) = 1 + 0.30\,\mathrm{erf}(0.9u + 0.2)$ and $g(v) = 1 + 0.35\,\mathrm{erf}(1.1v - 0.3)$. The grid is $\varepsilon \in \{0, 0.01, 0.03, 0.10\}$, $t \in \{0.5, 1, 2\}$, $x \in \{(0,0), (0.6, -0.4)\}$, and $n \in \{2, 4\}$: 48 recovery settings with 12 reused Gaussian controls. The two orders use 18 and 34 unique source-vector queries. The reported box scale, feasible-residual selection, SVD cutoff candidates, and denominator floor remain fixed. Theoretical comparisons use $m = 0.6, M = 1.4$; the implementation retains its more conservative $0.55, 1.45$ constants.

For each original-space source query, quadrature under $N(A_t x, C_t)$ computes the posterior mean weighted by $e^{V_\varepsilon}F$, with $F = f$ or g, and returns the genuine absolute score $(\mathbb{E}[X_0 \mid x, F] - x)/t$. The joint factor fg is used only by the evaluator. With $B_t = W_t A_t$, an active-block query fixes the other canonical mean at the target, maps μ' to $x' = B_t^{-1}\mu'$, subtracts $s^q_{0,t}(x')$, applies B_t^{-T}, and selects the requested block. Output is $s^q_{0,t}(x) + B_t^\top \widehat{Z}$. Signed posterior correlations are $-1/7, -1/5, -1/4$; reflecting the second block gives positive canonical correlations and transforms its factor consistently.

Table 7: Reported non-Gaussian recovery: mean original-coordinate Euclidean error over six fixed (t, x) settings per cell. The target shift is a model difference, not an error component that may be added as a scalar to the other columns. These 48 historical FQR evaluations are aggregate reports; their execution archive is not included.

ε	**FQR,** $n = 2$	**FQR,** $n = 4$	$\|s^p_{AB} - s^q_{AB}\|$
0	1.34×10^{-6}	3.25×10^{-12}	0
0.01	1.96×10^{-4}	1.95×10^{-4}	4.46×10^{-4}
0.03	5.86×10^{-4}	5.85×10^{-4}	1.34×10^{-3}
0.10	1.96×10^{-3}	1.96×10^{-3}	4.45×10^{-3}

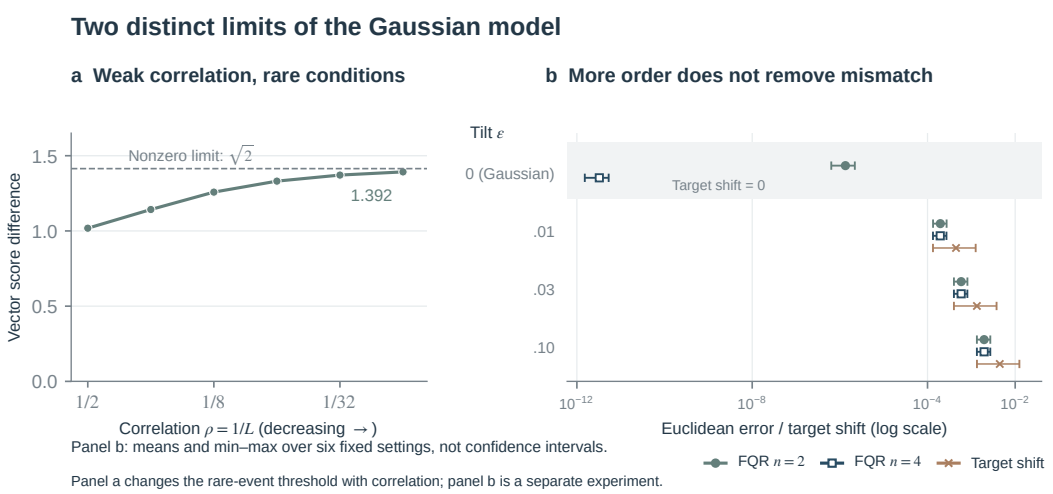


Figure 6: Two distinct scope limits, using the reported aggregates. (a) In the one-sided rare-condition witness, correlation decreases as $\rho = 1/L$, yet the score gap tends to $\sqrt{2}$ rather than zero; the condition also becomes rarer. (b) The Gaussian control ($\varepsilon = 0$) separates the two FQR orders, whereas their mean errors nearly coincide at nonzero tilt. Markers show means and intervals show min–max over six fixed settings, not confidence intervals. Target shift is a separate model difference, not an additive error component; its zero control is annotated off the log scale. These finite observations do not establish an asymptotic error floor.

Reference checks and observed recovery error. The reported run cross-checks all sources and targets with 48 and 80 Gaussian nodes per coordinate, increasing the order if necessary until the discrepancy is below one percent of the smallest displayed nonzero error, capped at 10^{-10}. It checks recovered vectors separately. All 48 settings pass at 48/80 nodes. Reported maximum raw-source, raw-target, and raw-recovery discrepancies are 2.80×10^{-13}, 8.30×10^{-16}, and 1.32×10^{-15}. Zero-tilt sources agree with independent analytic erf sources within 1.27×10^{-15}; joint targets agree with a Gaussian-CDF expression within 3.93×10^{-16}. At $\varepsilon = 0.1$, five-point differentiation at two step sizes agrees with the posterior-mean score within 4.39×10^{-12}.

All cases reportedly return a feasible-residual candidate without failure or denominator clipping. At zero perturbation, $n = 4$ errors range from 1.51×10^{-12} to 5.31×10^{-12}. At nonzero perturbations they range from 1.33×10^{-4} to 2.70×10^{-3}, and differ little between the budgets. This establishes neither a positive asymptotic error floor nor an optimal mismatch exponent. Observed-to-Proposition 4 ratios are at most 0.1323 for raw source differences at queried points, 0.0524 after source projection, and 0.0464 for the joint target. These maxima are not global source-error estimates; the proposition alone is not a calibrated FQR certificate.

Rare events and the original evidence-qualified bound. For $L = 2, 4, 8, 16, 32, 64$ and $\rho = 1/L$, the reported run saves both complete scores, their difference, and log joint probabilities. A 65-digit scaled-excess calculation is cross-checked against an 85-digit conditional one-dimensional integral;

largest discrepancies are 1.32×10^{-64} for a score component and 6.29×10^{-63} for a log probability. The score gap increases from 1.01815 to 1.39246 while ρ decreases to 0.015625.

The separate evidence grid uses $\rho \in \{0.01, 0.03, 0.10\}$ and $L \in \{-1, 0, 0.5, 1\}$, with $\bar{\rho} = 0.2$, $\lambda = 0.8$, and $\zeta = w_\rho(0)$. Nine of 12 settings satisfy the condition $\rho \le \zeta/2$; their error/bound ratios lie in $[0.00392, 0.02188]$. The other three, $(\rho, L) = (0.03, 1), (0.10, 0.5), (0.10, 1)$, remain inapplicable to Proposition 5, not violations or successful checks of that proposition. Appendix I derives a different bound that does apply to them. No hard-condition oracle was passed to FQR.

Execution records. The reported runner is `python -m experiments.scope -part all`, followed by `python -m experiments.scope_summary`; its records contain queried vectors, quadrature comparisons, rejected candidates, residuals, and separate timings. The reported environment was an Intel Xeon Gold 6430 CPU, Python 3.12.3, NumPy 2.4.6, SciPy 1.18.1, mpmath 1.3.0, and one BLAS/OpenMP thread. The reported 7.09-second subprocess time excludes development and reporting. The original scope execution archive is distinct from the independent verification script; it is not reconstructed from rounded means.

I A SHARP LOGARITHMIC CONDITIONING PENALTY AND A PERTURBATION DIAGNOSTIC

Proposition 5 controls unnormalized differences and then divides by the evidence, introducing an inverse-probability factor. Retaining product structure throughout covariance interpolation removes that loss. The resulting bound applies to arbitrary cross-correlation perturbations with fixed Gaussian marginals, not only SVD truncations. It does not extend the finite-query theorem to hard conditions.

I.1 STATEMENT: PROBABILITY ENTERS THROUGH ITS LOGARITHM

Let $0 \le f, g \le 1$ be measurable, with neither factor identically zero almost everywhere. Fix a target mean μ. For $\|R\|_{\mathrm{op}}, \left\|\widetilde{R}\right\|_{\mathrm{op}} \le \bar{\rho} < 1$, write

$$\lambda = 1 - \bar{\rho}, \qquad \tau = \left\|R - \widetilde{R}\right\|_{\mathrm{op}}, \qquad w_R = \mathbb{E}_{\Sigma_R}[f(\mu_U + U)g(\mu_V + V)], \qquad a_0 = \log(1/w_R).$$

Thus $0 < w_R \le 1$; the scalar a_0 is the negative log evidence. Define $\Psi(x) = (e^x - 1)/x$ for $x > 0$ and $\Psi(0) = 1$.

Theorem 6 (Logarithmic conditioning bound)**.** *Under the preceding assumptions,*

$$w_R^{e^{\tau/\lambda}} \le w_{\widetilde{R}} \le w_R^{e^{-\tau/\lambda}}, \tag{104}$$

and

$$\left\|Z_R(\mu) - Z_{\widetilde{R}}(\mu)\right\| \le (4 + \sqrt{2})\frac{\tau}{\lambda^{3/2}}\Psi\left(\frac{3\tau}{2\lambda}\right)\sqrt{a_0}(1 + a_0), \tag{105}$$

$$\left\|Z_R(\mu) - Z_{\widetilde{R}}(\mu)\right\| \le \sqrt{\frac{2a_0}{\lambda}}\left(1 + e^{\tau/(2\lambda)}\right). \tag{106}$$

One may take the smaller of the two bounds. There is no ambient-dimension factor and no hypothesis $\tau \le w_R/2$. If only $w_R \ge \zeta > 0$ is known, replace a_0 by $\log(1/\zeta)$. If both endpoint evidences are available, the smaller endpoint negative log evidence may be used by reversing the path.

For fixed $\bar{\rho}$, this implies

$$\left\|Z_R - Z_{\widetilde{R}}\right\| \le C_{\bar{\rho}}\tau\sqrt{\log(1/\zeta)}\big(1 + \log(1/\zeta)\big),$$

because $\tau \le 2\bar{\rho}$. In particular, the small-evidence penalty is logarithmic, rather than ζ^{-1}. Section I.3 shows that the logarithmic power $3/2$ cannot be uniformly decreased in a bound linear in τ. This is sharpness of that power, not a claim that every constant or the complete two-parameter minimax modulus is optimal.

I.2 Proof by smoothing, entropy, and covariance interpolation

We supply the elementary estimates so that the probability dependence is explicit. Gaussian covariance differentiation is the classical Price identity (Voigtlaender, 2021); the entropy estimates below are proved directly.

Smoothed likelihood derivatives. For $0 \le h \le 1$ and $v > 0$, put $H = H_{vI}h(x) > 0$ and $a = \log(1/H)$. Let P tilt a standard Gaussian G by $h(x + \sqrt{v}\,G)/H$. Then

$$\mathrm{KL}(P\|\gamma) \le a.$$

The entropy variational inequality and the linear Gaussian moment-generating function give $\|\mathbb{E}_P G\| \le \sqrt{2a}$. For any unit e, using $\mathbb{E}e^{(e^\top G)^2/4} = \sqrt{2}$ gives

$$\mathbb{E}_P(e^\top G)^2 \le 4a + 2\log 2.$$

Differentiating the Gaussian kernel, rather than h, consequently yields

$$\frac{\|\nabla H\|}{H} \le \sqrt{\frac{2a}{v}}, \qquad \frac{\left\|\nabla^2 H\right\|_{\mathrm{op}}}{H} \le \frac{1+4a}{v}. \tag{107}$$

For the Hessian, $v\nabla^2 H/H = \mathbb{E}_P GG^\top - I$; its negative eigenvalues have magnitude at most one, and its positive eigenvalues are at most $4a + 2\log 2 - 1 \le 4a + 1$. Thus no trace estimate is used.

A tilted log-weight estimate. On an arbitrary probability space let $0 < W \le 1$, $w = \mathbb{E}W$, $a = \log(1/w)$, and let P have density W/w. Write $L = -\log W$. Nonnegativity of relative entropy and a direct tail estimate give

$$\mathbb{E}_P L \le a, \qquad P(L \ge a + t) \le e^{-t} \quad (t \ge 0), \qquad \mathbb{E}_P L^2 \le a^2 + 2a + 2. \tag{108}$$

Indeed $\mathrm{KL}(P\|P_0) = a - \mathbb{E}_P L$, and on $\{L \ge a + t\}$ the weight W/w is at most e^{-t}. Integrating the tail proves the second moment bound. Cauchy–Schwarz also gives

$$\mathbb{E}_P L^{3/2} \le \sqrt{a(a^2 + 2a + 2)}. \tag{109}$$

The estimates apply by limits when W can be zero.

Keep an independent smoothing reserve. Take any $0 < v < \lambda$ and define

$$F = H_{vI_p} f, \qquad G = H_{vI_q} g, \qquad R_s = R + s(\widetilde{R} - R), \quad 0 \le s \le 1.$$

The residual covariance $\Sigma_{R_s} - vI$ is positive definite. The original w_s is exactly the expectation of $F(\mu_U + U)G(\mu_V + V)$ under this residual Gaussian. The smoothed factors are fixed along s, positive, and at most one. In the normalized residual expectation denote

$$A = -\log F, \qquad B = -\log G, \qquad L = A + B, \qquad a_s = -\log w_s.$$

Put $\Delta = \widetilde{R} - R$. Covariance differentiation gives

$$\dot{w}_s = \mathbb{E}_s[\nabla F^\top \Delta \nabla G]. \tag{110}$$

By (107), $2\sqrt{AB} \le A + B$, and (108),

$$|\dot{a}_s| = |\dot{w}_s|/w_s \le (\tau/v)\mathbb{E}_{P_s} L \le (\tau/v)a_s. \tag{111}$$

Thus $e^{-s\tau/v}a_0 \le a_s \le e^{s\tau/v}a_0$. Sending $v \uparrow \lambda$ proves (104).

The complete score derivative. Let $n_s = \nabla_\mu w_s$. Differentiating (110) in both mean blocks gives the stacked vector

$$\dot{n}_s = \mathbb{E}_s \begin{pmatrix} (\nabla^2 F)\Delta\nabla G \\ (\nabla^2 G)\Delta^\top \nabla F \end{pmatrix}.$$

No derivative of an indicator is taken: all derivatives act on the smoothed factors. Pointwise after dividing by FG, the integrand norm is at most

$$\frac{\sqrt{2}\tau}{v^{3/2}}\sqrt{(1+4A)^2 B + (1+4B)^2 A} \le \frac{\sqrt{2}\tau}{v^{3/2}}(1+2L)\sqrt{L}.$$

The last step uses $AB \le L^2/4$. Moreover, $Z_s = \mathbb{E}_{P_s}(\nabla \log F, \nabla \log G)$, so Jensen's inequality and $\mathbb{E}_{P_s} L \le a_s$ give

$$\|Z_s\| \le \sqrt{2a_s/v}.$$

Applying (108)–(109) and differentiating $Z_s = n_s/w_s$ now yields

$$\begin{aligned}\left\|\dot{Z}_s\right\| &\le \frac{\sqrt{2}\tau}{v^{3/2}}\left[\sqrt{a_s} + 2\sqrt{a_s(a_s^2+2a_s+2)} + a_s^{3/2}\right] \\ &\le (4+\sqrt{2})\frac{\tau}{v^{3/2}}\sqrt{a_s}(1+a_s).\end{aligned}$$

The last inequality uses $\sqrt{a_s^2+2a_s+2} \le \sqrt{2}(1+a_s)$. The score normalizer derivative is included in the $a_s^{3/2}$ term. Combining with (111), integrating $e^{3s\tau/(2v)}$, and taking $v \uparrow \lambda$ proves (105). This limit is of bounds on the same original quantities, so no differentiation at a singular residual covariance is required.

Finally, the original Gaussian entropy bound (102) gives $\|Z_R\| \le \sqrt{2a_0/\lambda}$ and $\left\|Z_{\widetilde{R}}\right\| \le \sqrt{2a_1/\lambda}$. Use (111) and the triangle inequality to obtain (106). If $a_0 = 0$, the evidence inequality gives $a_s = 0$ and both scores vanish; if $\tau = 0$, both models coincide. Thus the degenerate endpoints are also included. This proves the theorem.

Scope and use. SVD truncation is the special case $\tau = \rho_{k+1}$. Proposition 5 remains applicable on its stated domain; where both bounds apply, their minimum may be used. For positive factors in $[m, M]$, divide by M and use $a_0 \le 2\log(M/m)$ to obtain a logarithmic-contrast alternative to the constants in Theorem 3. Scaling does not change a score, but a useful evidence bound still requires a specified likelihood normalization. No score-only estimate of evidence, no hard-event query theorem, and no non-Gaussian replacement theorem are asserted.

I.3 The power $3/2$ is necessary: a rare mode-switching family

The one-sided example in Appendix H.4 proves that conditioning cannot be omitted, but does not saturate the logarithmic upper bound. A two-branch first condition does.

Fix $\theta > 0$ and for sufficiently large L set

$$f_L(u) = \mathbf{1}_{\{|u| \ge L\}}, \qquad g_L(v) = \mathbf{1}_{\{v \ge L\}}, \qquad R_L = \rho_L = \theta/L^2, \qquad \widetilde{R}_L = 0.$$

Let $a_L = -\log w_{\rho_L}(0)$. Then

$$\frac{a_L}{L^2} \longrightarrow 1, \qquad \frac{\|Z_{\rho_L}(0) - Z_0(0)\|}{L} \longrightarrow \tanh\theta, \qquad \frac{\|Z_{\rho_L}(0) - Z_0(0)\|}{\rho_L a_L^{3/2}} \longrightarrow \frac{\tanh\theta}{\theta} > 0. \tag{112}$$

Consequently no exponent smaller than $3/2$ can uniformly replace it in a bound of the form $C\tau(1 + \log(1/w_R))^b$ at fixed $\bar{\rho} < 1$.

Proof. Write $P_+(\rho) = P_\rho(U \ge L, V \ge L)$ and $P_-(\rho) = P_\rho(U \le -L, V \ge L) = P_+(-\rho)$. The change of variables $U = L + x/L, V = L + y/L$ gives

$$P_+(\rho) = \frac{e^{-L^2/(1+\rho)}}{2\pi L^2\sqrt{1-\rho^2}} \int_0^\infty \int_0^\infty e^{-(x+y)/(1+\rho)} e^{-(x^2+y^2-2\rho xy)/(2L^2(1-\rho^2))}\, dx\, dy.$$

At $\rho = \theta/L^2$, dominated convergence, also with a first-moment weight, implies

$$P_\pm(\rho_L) = \frac{e^{-L^2 \pm \theta}}{2\pi L^2}(1 + o(1)), \qquad w_{\rho_L} = \frac{e^{-L^2}\cosh\theta}{\pi L^2}(1 + o(1)).$$

Conditional on the positive branch, $(\mathbb{E}U, \mathbb{E}V) = (L, L) + O(L^{-1})$; on the negative branch it is $(-L, L) + O(L^{-1})$. The normalized branch-weight difference converges to $\tanh\theta$. Hence the total conditional mean divided by L tends to $(\tanh\theta, 1)$. Since $\Sigma_{\rho_L}^{-1} \to I$, the correlated score divided by L has the same limit. The independent score divided by L tends to $(0, 1)$, using the scalar Gaussian tail mean. These observations prove (112). □

The mechanism is a change in weights between two rare alternatives, not just movement of one tail. The parameter $\theta = \rho L^2$ controls that competition. The family changes with L; this is not instability of every fixed event. The witness and theorem remain Gaussian statements, and do not replace the main positive-factor query theorem.

I.4 A FIRST-RESPONSE IDENTITY FOR THE NON-GAUSSIAN DIAGNOSTIC

The measured equality of two query budgets at nonzero tilt does not prove a positive asymptotic recovery floor. To analyze the model change separately, parameterize $p_\varepsilon \propto q_0 e^{\varepsilon v}$ with $\operatorname{osc} v \le 1$. For a fixed $F \in [\ell, u]$, let $Q_{x,F}$ be the reference Gaussian backward law tilted by F, as in Proposition 4, and define

$$B_{F,t}(x) = \frac{1}{t} \operatorname{Cov}_{Q_{x,F}}(X_0, v(X_0)).$$

This covariance is a vector and is not supplied by the original score-only oracle.

Proposition 7 (Uniform first response, without differentiating the potential)**.** *For every real ε and fixed $t > 0$,*

$$s^{\varepsilon}_{F,t}(x) = s^{0}_{F,t}(x) + \varepsilon B_{F,t}(x) + \mathcal{E}_{F,t}(x), \qquad \sup_x \|\mathcal{E}_{F,t}(x)\| \le \frac{\varepsilon^2 e^{|\varepsilon|/2}}{2t} \sqrt{(u/\ell)\,\|C_t\|_{\mathrm{op}}}. \quad (113)$$

Proof. At fixed x, let $m_\theta = \mathbb{E}_\theta X_0$ for the posterior tilted by $F e^{\theta v}$. Exponential-tilt differentiation gives

$$m'_\theta = \operatorname{Cov}_\theta(X_0, v), \qquad m''_\theta = \mathbb{E}_\theta[(X_0 - m_\theta)(v - \mathbb{E}_\theta v)^2].$$

Because $\operatorname{osc} v \le 1$, $|v - \mathbb{E}_\theta v| \le 1$. For every unit e, the absolute second derivative in that direction is at most $\sqrt{\operatorname{Var}_\theta(e^\top X_0)}$. The posterior density relative to $N(A_t x, C_t)$ is at most $e^{|\theta|} u/\ell$, so this variance is at most $e^{|\theta|}(u/\ell) e^\top C_t e$. Taylor's integral remainder, the supremum over e, and $s^\theta = (m_\theta - x)/t$ prove the result. □

This is a sensitivity identity with an explicit uniform remainder, not a new oracle-only non-Gaussian recovery algorithm. For $F = fg$, the required covariance is under the joint reference posterior. Using the evaluator's access to it inside FQR would change the information model. In particular, subtracting this response from the true/reference target difference tests its leading term; it does not establish that FQR's own misspecification error has been removed.

I.5 EXECUTABLE CONDITIONING AND RESPONSE CHECKS

The CPU script `revision_checks/check_conditioning.py` evaluates Theorem 6 and Proposition 7 without calling FQR. It does not rerun or overwrite the original recovery experiments. Its saved outputs contain all evaluated values and precision checks.

Evidence grid. On exactly the 12 (ρ, L) half-space settings reported in Appendix H.5, all 12 satisfy Theorem 6. All measured score gaps are below the logarithmic bound; error/bound ratios are 0.02339–0.03343. The three points excluded by Proposition 5 remain excluded from that proposition. On its nine applicable points, the logarithmic bound is smaller by factors 1.38–6.42. These are finite numerical checks, not calibrated interval certificates or estimates of worst-case sharp constants.

Sharp logarithmic power. The two-branch family uses $L \in \{2, 4, 8, 16, 32, 64\}$ and $\theta \in \{0.1, 0.5, 1\}$, for 18 checks. Scaled Gaussian excess integrals use 60-digit arithmetic; all six $\theta = 1$ cases are repeated at 85 digits and agree in the 35 reported significant digits. A representative slice is

L	$\rho = 1/L^2$	**Score gap**	**Gap**$/[\rho a_L^{3/2}]$
4	0.0625	3.17803	0.58761
8	0.015625	6.16499	0.69002
16	0.00390625	12.22271	0.73672
32	0.0009765625	24.38976	0.75372
64	0.000244140625	48.75142	0.75923

The limiting ratio is $\tanh(1) = 0.761594\ldots$. This is a different family from the one-sided $\rho = 1/L$ witness with a $\sqrt{2}$ limit; those original observations are retained unchanged.

First response of the genuine target. Use exactly the fixed background, factors, times, and means of Appendix H.5, with $v(u, v) = \frac{1}{2}\sin(uv)$. For $F = f, g, fg$, six positions/time settings, and $\varepsilon \in \{\pm 0.01, \pm 0.03, \pm 0.1\}$, 108 comparisons satisfy (113). Independent 48- and 80-node-per-coordinate quadrature differs by at most 2.55×10^{-15} for the tested response and score-shift vectors. The positive-tilt joint-target means are

ε	$\left\|s^{\varepsilon}_{AB} - s^{0}_{AB}\right\|$	$\left\|s^{\varepsilon}_{AB} - s^{0}_{AB} - \varepsilon B_{AB}\right\|$
0.01	4.46499×10^{-4}	4.20378×10^{-7}
0.03	1.33859×10^{-3}	3.78354×10^{-6}
0.10	4.45148×10^{-3}	4.20439×10^{-5}

The remainder divided by ε^2 is approximately 0.004204 throughout these three rows. This confirms the first-order target response in this diagnostic, but the covariance uses additional evaluator information. These rows must not be presented as improved FQR recovery errors, a fair additional baseline, or neural-source results.

J Executable studies and independent verification

This appendix specifies the primary reference studies and their complete execution records. The information and recovery studies are independent executions of the stated constructions, not reconstructions of the historical aggregates in Appendix G. All inputs, query responses, measured outputs, and numerical checks are available in the source package.

J.1 A fully specified same-noise recovery study

The reference model is the nonperiodic erf example, with $U = (u, a)$, $V = (v, b)$,

$$f(u, a) = 1 + 0.35\ \mathrm{erf}(1.1u + 0.7a + 0.25),$$
$$g(v, b) = 1 + 0.40\ \mathrm{erf}(0.9v - 0.6b - 0.50).$$

Its marginal covariances are identities, the only cross-correlation is $\mathrm{Corr}(u, v) = 0.8$, and declared bounds are $m = 0.55, M = 1.45$. We evaluate all 16 means

$$\mu = (s, 0.2t, t, -0.2s), \qquad (s, t) \in \{-0.75, -0.25, 0.25, 0.75\}^2.$$

Thus both active and passive means vary, while the likelihood pair and its rank remain fixed. The errors in Table 2 are $\left\|\widehat{Z} - Z_R\right\|$ in these canonical coordinates; original diffusion-score errors would also include the map in (4).

For a factor $1 + A\,\mathrm{erf}(d^\top z + c)$, the source callable returns the full two-vector

$$q(z) = \frac{2A\exp(-v^2)}{\sqrt{\pi}\,\sigma[1 + A\,\mathrm{erf}(v)]}\,d, \qquad \sigma = \sqrt{1 + 2\,\|d\|^2}, \quad v = (d^\top z + c)/\sigma.$$

The joint evaluator is separate: it integrates the two linear attribute views using a bivariate Gaussian-CDF formula, and independently checks every target by 80-node-per-coordinate Gaussian quadrature. The maximum full-vector discrepancy over the 16 targets is 3.22×10^{-16}. Neither the clean factors nor the joint evaluator are arguments to the fitter or to the Gaussian baselines.

Fixed source errors. For each source, at its absolute active coordinate z_1, add

$$e_{\phi,\omega}(z_1) = \frac{\delta}{\sqrt{2}}\begin{pmatrix}\sin(1.7\omega z_1 + \phi)\\ \sin(2.3\omega z_1 - \phi)\end{pmatrix}, \qquad \phi_A = 0.3, \quad \phi_B = -0.7.$$

The full-vector norm is at most δ at every real location. The four settings are exact sources, $(\delta, \omega) = (10^{-8}, 1), (10^{-8}, 10), (10^{-6}, 10)$. All methods see the same deterministic field, including when finite differences revisit a location. This is a controlled stress test, not a learned-score error model; the error field need not itself be a gradient.

Candidate construction and selection. The supplied file `reference/fqr_rank_one.py` implements one active direction and any finite number of passive output coordinates. It accepts only a callable returning full single-condition score vectors. For orders $n \in \{1, 2, 4, 8, 16\}$, it uses radius $\sqrt{n/2}$, period $8\sqrt{n/2}$, degree $2n$, and $4n + 1$ distinct cosine nodes per source. The seven relative SVD cutoffs are fixed at

$$\{10^{-12}, 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 10^{-3}, 10^{-2}\}.$$

On the auxiliary grid of $16n + 1$ points, the scalar bound is $3M/m + 1.1$ and the passive Euclidean bound is $3(M/m)\sqrt{2/\pi} + 0.1$. Origin normalization is checked to 10^{-10}. Among numerically feasible candidates, the smallest maximum vector residual selects the output, breaking exact residual ties toward the larger cutoff. The denominator floor is $(m/M)^2/2$. No setting is selected by joint-target error. This is a residual-checked truncated-SVD candidate, not an exact minimax SOCP solve or a calibrated target-error certificate.

The three baselines and five FQR orders, at four source settings and 16 means, give $8 \cdot 4 \cdot 16 = 512$ evaluations. All complete, all selected FQR candidates pass the finite constraints, and none clips the denominator. The order-4 passive ablation sets both recovered passive entries to zero while leaving the active reconstruction fixed; its mean error is 0.184579. This tests output retention, not a separately refitted method.

J.2 Matched score-value access and baseline checks

The Gaussian baselines are those defined in Appendix G.2, expressed in the canonical mean coordinates. Let $\bar{q}_i$ and $\bar{H}_i$ embed a source score and its Jacobian into the full four-dimensional space. For background covariance Σ_R, the conditional moment identities are

$$m_i = \mu + \Sigma_R \bar{q}_i, \qquad C_i = \Sigma_R + \Sigma_R \bar{H}_i \Sigma_R.$$

JAC fuses these Gaussian moments and returns $\Sigma_R^{-1}(m_J - \mu)$. Full overlap differentiates the complete scalar Gaussian-overlap integral, including covariance variation, and adds that gradient to the two embedded source scores.

Both methods use central finite differences with fixed step 10^{-4}; integer stencil keys cache repeated source locations. A noisy Jacobian is symmetrized before forming a covariance. Source covariances and fused precision are checked for positivity, with failures explicitly recorded rather than replaced by a different method. Addition, JAC, and overlap use 2, 10, and 26 unique source-vector calls respectively. Every baseline sees only callable score values; analytic derivatives are used exclusively for separate validation.

An independent Gaussian-likelihood algebra test checks both fusion identities against the exact Gaussian joint target. At grid indices $0, 5, 15$ in lexicographic (s, t) order, 60-digit analytic differentiation gives maximum discrepancies 4.20×10^{-11} for JAC and 9.53×10^{-11} for full overlap. A separate, off-grid mean $(0.11, -0.23, 0.37, 0.19)$ checks steps $3 \times 10^{-5}, 10^{-4}, 3 \times 10^{-4}, 10^{-3}$ against the same high-precision formulas. The fixed 10^{-4} step has errors below 3.50×10^{-11} for both methods there. These diagnostics separate baseline approximation bias from finite-difference error without selecting FQR settings on target truth.

Records and descriptive statistics. The command `python experiments/reference_study.py` writes its complete design before evaluating results. The files are

```
experiments/results/reference_study_config.json
experiments/results/reference_study_cases.jsonl
experiments/results/reference_study_summary.json.
```

Each case records its mean, source setting, method, target, recovered vector, full and passive errors, query count, and wall time. FQR records also include queried vectors and positions, every tested cutoff, selected coefficients, residuals, finite constraints, and denominator status. Baseline records contain all stencil positions and queried vectors. The summary records means, medians, minima, maxima, failures, clipping, independent reference checks, and the passive ablation. No raw observations are synthesized from aggregate means, and this deterministic single-pair grid is not treated as independent attribute draws for confidence intervals.

J.3 Executable information-resolution construction

The command `python experiments/information_study.py` evaluates Appendix C's two worlds at 100-digit precision. Set $a = b = 0.1$,

$$f_{\pm}(u) = 1 \pm a\sin(ku), \qquad g(v) = 1 + b\cos(\rho k v),$$

with $\rho \in \{0.2, 0.6, 0.8, 0.95\}$. For each $j = 0, \dots, 49$, choose budget $\delta_j = 10^{-2-62j/49}$ and solve on $k > 1$ for the explicit uniform source-score upper bound

$$\frac{ake^{-k^2/2}}{1 - ae^{-k^2/2}} = \delta_j.$$

Bisection brackets the root in $[1, 32]$ and takes the upper endpoint, preserving the source-error inequality. Both worlds can return the zero A field and the same exact B field. Formula (52) and its v derivative give the full-vector target separation; half is the two-world minimax lower bound. The terminal slope uses the final two adjacent log-budget/log-bound values. The four slopes reported in Section 6 are generated from these 200 recorded cases, saved in `experiments/results/information_study.json`. The likelihoods vary with the budget; all remain in $[0.9, 1.1]$.

J.4 Same-noise and selective-interface controls at one target

The separate command `python reference/run_example.py` uses the same erf pair at mean zero with source smoothing $\mathrm{diag}(\gamma, 1)$, for $\gamma = 1$ or $\gamma = 0.36 = 1 - 0.8^2$. The selective columns require the extended covariance interface of Section 2; they are not obtained merely by changing an ordinary scalar diffusion time. It evaluates $n = 4, 8, 16$, with either exact sources or the low-frequency field above at $\delta = 10^{-6}$. The independent target check differs by 8.89×10^{-17}.

Table 8: Executable one-target controls. Entries are measured full-vector errors, not fitted minimax rates. The same-noise and selective columns use different information interfaces.

		Same-noise, $\gamma = 1$		Selective, $\gamma = 0.36$	
n	**Calls**	$\delta = 0$	$\delta = 10^{-6}$	$\delta = 0$	$\delta = 10^{-6}$
4	34	8.76×10^{-7}	1.65×10^{-6}	2.36×10^{-6}	2.30×10^{-6}
8	66	4.47×10^{-7}	9.81×10^{-7}	2.04×10^{-7}	2.99×10^{-7}
16	130	2.35×10^{-7}	9.70×10^{-7}	3.61×10^{-9}	2.91×10^{-7}

All twelve configurations return numerically feasible candidates and retain passive components. Complete records are saved in `verification/results/reference_example.json`. Finite same-noise precision remains sensitive to truncation and regularization; this control is not evidence of a sharp numerical error floor.

J.5 Independent checks of the proof mechanisms

The command `python verification/run_all.py` executes four independently integrated Hermite derivative matrices, a 70-digit zero-transcript witness, 32 spectral truncations, and four nonaligned rank-replacement witnesses. The largest high-order Hermite matrix discrepancy is 9.51×10^{-16}; the lower-order quadrature refinement discrepancy reaches 2.61×10^{-9} at $\rho = 0.95$, and is retained in the records. The zero-transcript residual is 4.60×10^{-72} with positive target separation $3.59499885549 \times 10^{-6}$. These are finite construction checks, not empirical proofs of universal quantifiers.

The suite also executes the 12 evidence-grid cases, 18 two-branch rare-event cases, and 108 signed perturbation-response comparisons in Appendix I.5, together with the 74 scope checks in Appendix H. An additional 40 finite-Fourier cases use arbitrary, generally nonaligned cross-correlation matrices in block dimensions $(1, 1), (2, 3), (4, 4), (16, 20), (64, 96)$. They test both endpoint evidence inequalities, the complete-vector logarithmic bound, and Price differentiation against independent central differences. The largest full-vector derivative discrepancy is below 4.70×10^{-12}. Six Gaussian-transition quadratures check (19), including a step larger than one. All assertions pass; per-case results and the execution summary are saved under `verification/results/`.

Environment and reproduction. The recorded reference execution uses Python 3.13.5, NumPy 2.3.5, SciPy 1.17.0, mpmath 1.3.0, and one BLAS/OpenMP thread. The recovery configuration records the CPU model and source-file hashes. The dependencies are listed in `verification/requirements-tested.txt`. Commands require no network once dependencies are installed; compilation uses the included bibliography, figures, and style file. Table 2 is generated from the saved summary by `python experiments/render_tables.py`. High-precision arithmetic is not interval arithmetic, and passing finite checks is not formal proof verification. Numerical results can vary slightly across floating-point libraries; failure states, residuals, and reference discrepancies remain part of the output rather than being suppressed.