Empirical transport transitions for heat-kernel tangent laws

Scott Schubr
Luminary Oddities

November 2026

Abstract

Heat-kernel Sinkhorn metrics compare densities through Gaussian tangent laws whose covariance is set by a heat time. We determine how many samples these tangent laws need and which heat time is optimal. On the circle, for smooth positive densities with positive compact corrections, the quantity $h\log(N)^2$ controls the leading squared empirical Wasserstein constants, crossover profiles have a unique positive optimal cross bandwidth, and supercritical excess has sharp stretched logarithmic scales with explicit constants. On the torus $\mathbb{T}^d$, $d \geq 2$, positive compact heat tangent covariances give exact finite $h(\log N)^{2/d}$ sampling constants for every Sobolev order above $d/2-1$, with population bias and a unique optimal bandwidth throughout the subcritical range. For a positive fourth-order correction and smoothing order m, the critical Sobolev order is $2m$, with universal leading bias $h^{2m}\log(1/h)$ and an exact optimal correction at heat time $t^m \log R$; above it the leading coefficient is the weighted Gaussian tangent action.

Contents

1 Introduction **2**

2 The empirical transport transition for heat-kernel tangent laws **2**
2.1 The transition and optimal heat time 2
2.2 A triangular Gaussian comparison 4
2.3 Comparison with a spectral filter 5
2.4 The logarithmic supercritical decay 7
2.5 Optimizing the cross risk 9
2.6 Linear empirical tangent fluctuations 10

3 Sobolev transitions and optimal heat times for empirical tangent laws **11**
3.1 Tangent laws in a spectral Sobolev metric 11
3.2 Scalar Gaussian sampling with a varying spectrum 13
3.3 Spectral profiles and supercritical sampling 14
3.4 Weighted Gaussian comparison and population bias 17
3.5 The optimal positive heat-time constant 18
3.6 The raw logarithmic borderline 19
3.7 The raw tensor above the logarithmic borderline 21

4 High Sobolev bias and optimal sampling for positive heat corrections **23**
4.1 The high Sobolev regimes 24
4.2 The covariance tangent of an odd hyperpower 25
4.3 The compact fourth-order correction 26

4.4 Population comparison 28
4.5 Cross spectral weights and finite-sample sensitivity 29
4.6 Optimizing the high-regularity risks 31

1 Introduction

The Luminary Research Archive article *Leading bias and positive correction of heat-kernel Sinkhorn metrics* [1] identified the population bias of these metrics, their operator convergence at rate $O(h)$ was proved in [3], and root-n Gaussian inference for the empirical metric was established in [2]. Here we add the sampling side: the tangent-law constants below govern how the bandwidth in that inference may shrink with the sample size. Section 2 treats the circle, Section 3 the torus in every dimension, and Section 4 the high Sobolev regime of positive corrections.

2 The empirical transport transition for heat-kernel tangent laws

On the circle, the positive compact heat-kernel Sinkhorn tensors define Gaussian laws on the full tangent Hilbert space. We determine the leading empirical Wasserstein constants when the heat time decreases with the sample size. The critical regime is heat time of order the inverse square of the log sample size, and the constants are independent of the smooth positive underlying density. For approximating the unregularized tangent law by a regularized empirical cloud, a unique positive heat-time constant minimizes the leading risk. The same transition holds for independent clouds of linear empirical tangent fluctuations when each fluctuation is formed from sufficiently many observations.

2.1 The transition and optimal heat time

Let f be a fixed smooth strictly positive probability density on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. All covariance operators act on the real space $\mathcal{H} = q^{\perp} \subset L^2(\mathbb{T}, dx)$, where $q = \sqrt{f}$. Put

$$S = -\partial_x^2/2 + \frac{\partial_x^2 q}{2q}, \qquad G = (2S)^{-1}|_{\mathcal{H}}.$$

We use the original heat-kernel tensor T_h, the positive compact fourth-order correction $\widehat{C}_h$, or any fixed positive odd hyperpower $C_{h,\ell}$ from [4]. Write A_h for the chosen tensor and F for its corresponding filter

$$v(x) = \frac{x}{\sinh x}, \qquad F_0(x) = v(x), \qquad F_+(x) = \frac{4 - \operatorname{sech} x}{3} v(x).$$

$$F_\ell(x) = 1 - (1 - v(x))^\ell \quad (\ell \geq 3 \text{ odd}).$$

All filters are extended continuously at zero. Set $A_0 = G$ and let $\gamma_h = \mathcal{N}(0, A_h)$ for $h \geq 0$. Empirical laws always give equal mass to each sampled point.

For $a > 0$, let $r(a) \in (0, 1)$ solve $r/(1 - r^2) = 2a$ and define

$$B(a) = -\tfrac{1}{2} \log(1 - r(a)^2) + 2a(1 - r(a)),$$
$$I(a) = B(a) - aB'(a), \qquad B'(a) = 2(1 - r(a)).$$

The scalar functions are extended continuously to zero. For a strictly positive continuous profile g on $(0,\infty)$ with $g(u) \le C/u^2$, define η_g by $\int_0^\infty I(\eta_g g(u))\,du = 1$, and put

$$\begin{aligned} D(g) &= \int_0^\infty g(u) B'(\eta_g g(u))\,du \\ &= \sup_{\eta>0} \frac{\int_0^\infty B(\eta g(u))\,du - 1}{\eta}. \end{aligned} \tag{1}$$

The information integral is continuous, strictly increasing from zero to infinity. The last identity follows by differentiating the quotient, whose derivative has the sign of $1 - \int I(\eta g)$. Finiteness and differentiation use $B(a) \lesssim \min(a, 1+\log(1+a))$, $0 \le I(a) \le B(a)$, and $aB'(a) \le 1/2$.

Theorem 2.1. *Let $R = \log N$ for one empirical law and $R = \log\min(N,M)$ for two independently sampled empirical laws. As $R \to \infty$ along a sequence with $hR^2 \to t \in [0,\infty)$ set*

$$g_0(u) = \frac{1}{\pi^2 u^2}, \qquad g_t(u) = g_0(u) F(\pi^2 t u^2/2), \qquad m_t = \sqrt{g_0 g_t}.$$

For the W_s distance L_s from γ_h to its empirical law, or between two independent empirical laws of γ_h, for every fixed $1 \le s < \infty$,

$$RL_s^2 \;\longrightarrow\; D(g_t), \qquad R(\mathbb{E}L_s^s)^{2/s} \;\longrightarrow\; D(g_t).$$

For the cross distance L_* in W_2 from γ_0 to an empirical law of γ_h, or between independent empirical laws of γ_0 and γ_h,

$$RL_*^2 \;\longrightarrow\; H_F(\sqrt{t}), \qquad R\mathbb{E}L_*^2 \;\longrightarrow\; H_F(\sqrt{t}), \qquad H_F(\sqrt{t}) = \int_0^\infty (\sqrt{g_0} - \sqrt{g_t})^2\,du + D(m_t). \tag{2}$$

Random limits are almost sure for prescribed triangular rows indexed by their smaller sample size, with arbitrary dependence across rows. If $h \to 0$ and $hR^2 \to \infty$, the same-law limits are zero, and

$$L_*^2/\sqrt{h}, \quad \mathbb{E}L_*^2/\sqrt{h} \;\longrightarrow\; A_F = \int_0^\infty (\sqrt{g_0} - \sqrt{g_1})^2\,du$$

with the same almost sure convention.

Theorem 2.2. *For each filter above, $H_F(z)$ is strictly convex for $z > 0$ and has a unique minimum at $z_* > 0$. Its minimum satisfies*

$$\frac{2}{\pi^2} < H_F(z_*) < H_F(0) = \frac{\mathrm{B}(3/4, 1/2)^2}{2\pi^2}.$$

Thus $h \sim z_*^2/R^2$ minimizes the leading squared cross risk among heat times tending to zero. For any two filters ordered strictly pointwise on $(0,\infty)$, the larger filter gives a strictly higher same-law coefficient at each $t > 0$. It gives a strictly higher cross coefficient at small positive t and a strictly lower cross coefficient at large t.

The operator constructions, positivity and Sobolev bias estimates are from [4]. The fixed-dimensional computational inference and calibration in [5, Theorem 2.1 and Corollary 5.3] concern the exact empirical nonlinear estimator evaluated on prescribed scores. Here the comparison is in the full tangent Hilbert law, including high frequencies, and Section 2.6 treats linear empirical fluctuations. Earlier Gaussian functional quantization [9] and empirical Wasserstein convergence [7, 8] give the natural unrestricted benchmark and logarithmic rate context. The fixed-covariance empirical constant in [6] is the $t = 0$ case.

2.2 A triangular Gaussian comparison

Proposition 2.3. *Suppose a centered diagonal Gaussian law in each row has eigenvalues $a_{R,j} \geq 0$ such that $a_{R,j} \leq Cj^{-2}$ and $R^2 a_{R,j}$ converges to $g(j/R)$ uniformly on every band $\varepsilon R \leq j \leq TR$. Then its same-law squared empirical distance has the limits $D(g)/R$ of Theorem 2.1. The hypotheses allow any labeling of the diagonal coordinates.*

Proof. For two standard normals with prescribed marginals, $B(a)$ minimizes mutual information plus $a\mathbb{E}(X-Y)^2$. At a fixed correlation the joint Gaussian minimizes mutual information by maximal entropy, and differentiation gives the equation for $r(a)$. Its joint density relative to the independent pair is

$$e^{f_a(x)+f_a(y)-a(x-y)^2}, \qquad f_a(x) = -\tfrac{1}{4}\log(1-r(a)^2) + a(1-r(a))x^2. \tag{3}$$

Here $2\mathbb{E}f_a = B(a)$. Integration in either marginal gives one. Under the correlated pair, independent standard normals U, V represent the normalized sum and difference. Centered log likelihood is $(r/2)(U^2-V^2)$, and $a(X-Y)^2 = 2a(1-r)V^2$. The identity $a(1-r) = r/[2(1+r)]$ gives uniform bounds for all penalties.

For an upper bound fix $0 < \eta < \eta_g$ and $T > 1$. Use (3) independently on $j \leq TR$ with penalties $a = \eta R^2 a_{R,j}$, and pair the other coordinates independently. The head information for $j \leq \varepsilon R$, divided by R, is $O(\varepsilon(1 + |\log\varepsilon|))$ by the bound on B and summation of $\log(1 + C(R/j)^2)$. The central Riemann sum therefore gives mean log likelihood at most $(1-2\delta)R$ for a fixed $\delta > 0$ and large R. Bounded centered Gaussian-square coefficients on $O_T(R)$ slots give probability e^{-cR} of exceeding $(1-\delta)R$ under the planted coupling. The mean square displacement scaled by R equals the corresponding central distortion integral to within head and tail errors $O(\varepsilon + T^{-1})$. In the head each scaled square before division by R has bounded scale, and in the uncoupled tail their scales are bounded by $2CR^2/j^2$. The sum has variance $O(R)$ and moments of all fixed orders $O(R^k)$ at order k before division by R. This gives convergence of the scaled distortion moments, with the stated tail error.

To pass from the planted coupling to equal masses, cut its likelihood to $k_0 = k\mathbf{1}_{\log k \leq (1-\delta)R}$. With two clouds use the partial matrix $k_0(X_i, Y_j)/(NM)$. Its conditional row sums, scaled by N, have variance at most $e^{-\delta R}$ and conditional bias whose absolute mean is at most e^{-cR}. The same is true for columns. Trim overloaded rows and columns and couple the remaining deficits. The expected fill mass is exponentially small, and its expected order-s cost is exponentially small by Cauchy–Schwarz and the bounded order-$2s$ marginal moments. The retained mass costs at most the original planted mean. With one true marginal use the integral matrix $k_0(x, Y_j)d\mu(x)/N$ and the same variance and deficit bound. Sending T to infinity and η up to η_g proves the upper limit.

For the converse set $\eta = \eta_g$ and use only $\varepsilon R \leq j \leq TR$ in (3). Call a point typical when its sum of marginal potentials on these slots is within δR of the mean. The atypical probability is exponentially small by Gaussian-square concentration. The sum of expected potentials of both points is $R\int B(\eta_g g) + o(R)$ up to arbitrarily small cutoff error. If typical points have squared distance at most $(D(g) - d)/R$ with fixed $d > 0$, their joint likelihood on the band is at least $e^{(1+c_d)R}$ for sufficiently small δ, ε and large T. This uses $\int B(\eta_g g) = 1 + \eta_g D(g)$. Because the likelihood integrates to one conditionally, this bounds the conditional chance of such a pair from the independent reference. In the one-cloud case a union over its atoms gives expected small-cost mass $O(e^{-cR})$. In the two-cloud case the expected number of typical small-cost pairs is at most $NMe^{-(1+c_d)R}$. Each pair carries at most $1/\max(N, M)$ mass. Both expected atypical marginal masses are exponentially small. This proves the moment lower limit by sending d down to zero.

For the almost sure statement project onto the ball of radius R. The uniform trace bound gives uniformly subgaussian Gaussian norms. The expected order-s clipping cost is $O(e^{-cR^2})$ and

Markov's inequality makes it summably negligible, including for an arbitrarily large second sample. After clipping, replacing one point changes optimal order-s cost by at most $(2R)^s/N$ or $(2R)^s/M$. Bounded differences at scale $R^{-s/2}$ give exponent at least ce^R/R^{3s}. Borel–Cantelli for the prescribed rows proves the assertion. □

For later use, the cross Gaussian identity of [6, Section 6] is purely algebraic. If C, D are injective trace-class covariances and w_j are the singular values of $C^{1/2}D^{1/2}$, express the Gaussian points using standard normals in the left and right singular bases before applying $C^{1/2}$ and $D^{1/2}$. The standard coordinates in any singular basis are recovered measurably by taking each covariance eigencoordinate divided by its positive standard deviation and using the square-summable orthogonal change of basis. Their cross inner product is exactly $\sum_j w_j U_j V_j$. The measurable coordinate maps to $(\sqrt{w_j}U_j)_j$ and $(\sqrt{w_j}V_j)_j$ preserve every coupling's cross term. Consequently the cross squared empirical cost equals the population squared Gaussian cost plus the same-law squared empirical cost for covariance $(w_j)_j$, up to centered sample mean differences of squared norms. With one true law there is one such correction, and with two clouds there are two. Under uniform trace bounds the corrections are $o(R^{-1})$ in mean absolute value and almost surely, by uniformly subexponential Gaussian square norms. The identity also follows directly by expansion in common coordinates when C, D commute.

2.3 Comparison with a spectral filter

Let $Q_h = M_{b_h} e^{h\partial_x^2/2} M_{b_h}$ and $E_h = e^{-hS}$, both leaving $\mathcal{H}$ invariant. On $\mathcal{H}$ we have $T_h = hQ_h(I - Q_h^2)^{-1}$. Define $A_h^\circ = GF(hS)$. We prove

$$W_2^2(\mathcal{N}(0, A_h), \mathcal{N}(0, A_h^\circ)) = o(\sqrt{h}). \tag{4}$$

The baseline estimates of [4, Lemma 2.1] give $\|Q_h - E_h\| = O(h^2)$ and $\|(I - Q_h)^{-1}\| = O(h^{-1})$ on $\mathcal{H}$. Also $\|(A_h - G)u\| \le Ch^2\|u\|_{H^2}$ by the same lemma, Theorem 1.2 or Theorem 9.1 as appropriate. The same last bound for A_h° follows by spectral calculus and $1 - F(x) = O(x^2)$ near zero. The two-derivative smoothing and selfadjoint duality for T_h in [4, Lemmas 6.1–6.2] give a bounded map $H^{-1} \to H^1$ by interpolation, and consequently the form bound $0 \le T_h \le C(S + \Pi)^{-1}$, where $\Pi = q \otimes q$. Theorem 1.2 and Theorem 9.1 of [4] transfer this bound to both positive corrections. The reference covariance obeys it as well.

Write $\lambda_j(S)$ for the positive eigenvalues on $\mathcal{H}$, with multiplicity. Bounded-potential perturbation and the real periodic Fourier basis give $\lambda_j(S) \sim \pi^2 j^2/2$. Fix $0 < \varepsilon < 1 < T$ and use the spectral bands

$$L : \lambda_j(S) \le \varepsilon^2/h, \qquad M : \varepsilon^2/h < \lambda_j(S) \le T^2/h, \qquad P = L + M,$$

also writing L, M, P for their orthogonal projections. The high complement costs $O(\sqrt{h}/T)$ in squared Wasserstein distance by the trace tail from the form bound, by coupling a Gaussian vector with its projection. On the low columns the difference satisfies

$$\|(A_h - A_h^\circ)L\| \le Ch^2(1 + \varepsilon^2/h).$$

Thus both low rows and low columns in the P compression contribute at most $O(\varepsilon^3\sqrt{h}) + o(\sqrt{h})$ in trace norm, by their rank $O(1 + \varepsilon/\sqrt{h})$.

For the middle block first compare T_h with $T_h^\circ = hE_h(I - E_h^2)^{-1}$. The resolvent identity gives $\|M(I - Q_h)^{-1}\| \le C_\varepsilon$ by writing it as

$$M(I - E_h)^{-1}[I + (Q_h - E_h)(I - Q_h)^{-1}].$$

Taking adjoints gives the other side. The identity $2T_h/h = (I - Q_h)^{-1} - (I + Q_h)^{-1}$ now gives

$$\|M(T_h - T_h^\circ)M\| = O_\varepsilon(h^3), \qquad \|M(T_h - T_h^\circ)\| = O_\varepsilon(h^2). \tag{5}$$

The plus inverse is uniformly bounded. For the fourth-order compact correction the extra bounded rational term has the form $h\,b(Q_h)$, where b has no poles on $[0, 1]$. Its difference from $h\,b(E_h)$ has norm $O(h^3)$ by telescoping products and inverses. The additional sandwich $-2h^2 T_h \mathcal{R} T_h$ is $O(h^2)$ in operator norm by smoothing, with $\mathcal{R}$ the differential operator in [4, Theorem 1.2]. The reference scalar combination divided by G is F_+, since the extra rational term divided by T_h° on an eigenspace is $(1 - e^{-x})^2/[3(1 + e^{-2x})]$ at $x = h\lambda_j(S)$.

The odd hyperpower is $T_h P_\ell(2ST_h)$ with $P_\ell(z) = \sum_{k=0}^{\ell-1}(1-z)^k$, and its reference is $T_h^\circ P_\ell(2ST_h^\circ)$. Both ST_h and ST_h° are uniformly bounded. In the difference start with $(T_h - T_h^\circ)P_\ell(2ST_h)$, which after multiplication on the left by M is $O_\varepsilon(h^2)$ by (5). In each telescoped term with T_h° on the left, place the powers of $2ST_h^\circ$ to the left of $2S(T_h - T_h^\circ)$. They commute with M. Then $\|MT_h^\circ\| = O_\varepsilon(h)$ and $\|MS(T_h - T_h^\circ)\| = O_{\varepsilon,T}(h)$ by (5). Thus the middle block is $O_{\varepsilon,T}(h^2)$ in all three cases. Its rank $O(T/\sqrt{h})$ makes its trace norm $o(\sqrt{h})$. For the compressed Gaussians the squared Wasserstein distance is bounded by the trace norm of the covariance difference, via the common normal coupling and the Powers–Størmer inequality [10, Lemma 4.1]. Combining the three blocks and the high projection errors, then sending ε to zero and T to infinity, proves (4).

For each fixed s, the coupling used in this proof also gives $o(h^{1/4})$ in the sth mean Hilbert norm. Every displacement used there is a centered Gaussian vector, whose sth mean is bounded by a constant depending only on s times its root mean square. Independent copies of any such coupling give the same bound on the sth mean Wasserstein distance between the empirical laws, by coupling corresponding sample points.

Finite critical parameter. For hR^2 bounded, (4) gives an $o(R^{-1/2})$ comparison in any fixed order of mean distance, since $h^{1/4} = O(R^{-1/2})$. We may thus compute the constants with A_h°. Its sorted spectral coordinates in the basis of S have profile g_t because $\lambda_j(S) \sim \pi^2 j^2/2$ and $0 < F \le 1$. Proposition 2.3 gives the same-law limit. In the cross case the commuting covariance product gives profile m_t for the reduction above, and the squared population cost is

$$\sum_j (2\lambda_j(S))^{-1}\bigl(1 - \sqrt{F(h\lambda_j(S))}\bigr)^2.$$

After multiplication by R its limit is the bias integral in (2). For the Riemann sum the tail is bounded by C/T, and at $j \le \varepsilon R$ we use $0 \le 1 - \sqrt{F(x)} \le Cx^2$ near zero, giving a bound of order ε^7 for bounded hR^2. The same estimates apply to the limit integral. Almost sure limits for the actual Gaussian covariances now follow from the clipped bounded differences argument and uniform trace bounds.

Supercritical parameter. For $hR^2 \to \infty$ with $h \to 0$, the commuting population cost divided by $\sqrt{h}$ tends to A_F by the same Riemann sum with mesh $\sqrt{h}$. Equation (4) transfers this population limit. To control sampling separately, the positive eigenvalues of Q_h on $\mathcal{H}$ obey

$$\lambda_j(Q_h|_{\mathcal{H}}) \le \|b_h\|_\infty^2 e^{-2\pi^2 h\lceil j/2\rceil^2}.$$

Indeed the full eigenvalues are the squares of the singular values of $e^{h\partial_x^2/4}M_{b_h}$, and the first eigenvalue on the full space is 1 with eigenvector q. The gap and $\|b_h\|_\infty \le 1 + Ch$ are from [4, Lemma 2.1]. Thus for $j \ge J$ with J fixed, $\lambda_j(T_h) \le h/[2\sinh(chj^2)]$ for some $c > 0$. For $j < J$ they are

uniformly bounded. The relative form bounds (9) and (48) of [4] give the same envelope up to constants for A_h. For $w_j = s_j(G^{1/2}A_h^{1/2})$, the singular-value product inequality gives

$$w_{2j-1} \le \sqrt{\lambda_j(G)\lambda_j(A_h)}.$$

It follows that each of these decreasing covariance sequences is bounded above by a diagonal envelope of the form $Cj^{-2}\sqrt{v(c'hj^2)}$ beyond a fixed index, and by Cj^{-2} at the earlier indices. The envelope for $h \ge K/R^2$ is bounded above by replacing h by K/R^2. By coordinate contraction its empirical costs are bounded above by those of the corresponding envelope law. For each fixed K, Proposition 2.3 applies to the latter, with profile $g_K^\dagger = Cu^{-2}\sqrt{v(c'Ku^2)}$.

As $K \to \infty$, $g_K^\dagger$ decreases pointwise to zero and $\eta_{g_K^\dagger} \to \infty$, because the associated information integral at each fixed η tends to zero by dominated convergence. For $K \ge 1$, the distortion integrand is bounded by $\min(2C/u^2, 1/(2\eta_{g_1^\dagger}))$ and tends to zero pointwise. Hence $D(g_K^\dagger) \to 0$. The same-law sampling distances and the cross sampling excess have size $o(R^{-1/2})$ and $o(R^{-1})$, respectively, in the corresponding means. The exact cross identity gives the latter assertion, together with the negligible centered norm corrections. Uniform Gaussian tails and the clipped concentration argument again give the stated almost sure limits, including for the reduced cross costs. This proves Theorem 2.1.

2.4 The logarithmic supercritical decay

The supercritical Gaussian sampling terms admit a finer estimate. Define the cross excess $E_* = L_*^2 - W_2^2(\gamma_0, \gamma_h)$. Put $U = R\sqrt{h}$ and let $U \to \infty$ while $h \to 0$.

Theorem 2.4. *In the supercritical regime of Theorem 2.1, for each fixed s,*

$$\frac{-\log(L_s^2/\sqrt{h})}{U^{2/3}}, \quad \frac{-\log((\mathbb{E}L_s^s)^{2/s}/\sqrt{h})}{U^{2/3}} \longrightarrow (3\pi/\sqrt{2})^{2/3}.$$

For one or two independent Gaussian clouds in the cross problem, E_* is eventually positive almost surely, and

$$\frac{-\log(E_*/\sqrt{h})}{U^{2/3}}, \quad \frac{-\log(\mathbb{E}E_*/\sqrt{h})}{U^{2/3}} \longrightarrow (3\pi/2)^{2/3}.$$

Random limits use the prescribed-row convention. These finer Gaussian statements use the exponentially small sampling terms.

Proof. We first describe the eigenvalue comparison for both problems. On the full space, the singular values of $e^{h\partial_x^2/4}M_{b_h}$ are comparable to those of $e^{h\partial_x^2/4}$, by multiplication with $M_{b_h}^{-1}$ for the lower bound. Removing the ground-state eigenvalue of Q_h and using the heat spectrum yields two-sided bounds

$$\log\lambda_j(Q_h|_{\mathcal{H}}) = -\tfrac{\pi^2}{2}hj^2 + O(hj+h),$$

since $\|b_h^{\pm 1}\|_\infty = 1 + O(h)$ and the real heat eigenvalues occur in pairs. On indices $j \ge K/\sqrt{h}$, T_h and hQ_h have comparable eigenvalues. The relative form estimates (9) and (48) of [4] give $A_h \asymp T_h$ with fixed positive constants. In particular, for $\kappa = \pi^2/2$ and $a_{h,j} = \lambda_j(A_h)$,

$$\frac{\log(a_{h,j}/h)}{L} \longrightarrow -\kappa v^2 \quad \text{uniformly for } v = j\sqrt{h/L} \in [\varepsilon, T], \qquad L \longrightarrow \infty. \tag{6}$$

Also $a_{h,j} \le Cj^{-2}$ on all slots and $a_{h,j} \le Che^{-chj^2}$ beyond $K/\sqrt{h}$.

To obtain the same assertions for the cross sequence $a_{h,j} = s_j(G^{1/2}A_h^{1/2})$, use the positive extension $\widetilde{G} = G + \Pi$ on the full space. The norm of $\widetilde{G}^{1/2}M_{b_h}u$ is comparable, uniformly in h, to $\|u\|_{H^{-1}}$. In fact $\widetilde{G} \asymp (I - \partial_x^2)^{-1}$ follows from the equivalent positive H^1 forms of the inverse operators, and $b_h^{\pm 1}$ are uniformly bounded H^{-1} multipliers by the uniform smooth bounds of [4, Lemma 2.1]. The full-space operator $\widetilde{G}^{1/2}Q_h\widetilde{G}^{1/2}$ therefore has square root singular scales given by those of

$$(I - \partial_x^2)^{-1/2}e^{h\partial_x^2/4},$$

up to fixed multiplicative constants. To compare with the cross sequence, first remove the rank-one ground state. Then $T_h \geq hQ_h$ on $\mathcal{H}$, and $T_h \leq ChQ_h$ plus a positive correction of rank at most $K/\sqrt{h}$, because Q_h has eigenvalues at most a fixed number below 1 beyond that index. These inequalities survive conjugation by $G^{1/2}$ and comparison to A_h by relative form bounds. By the min–max principle, the cross sequence is bounded below by a constant times $\sqrt{h}$ times the displayed singular scales with index $j + 1$, and above by a constant times $\sqrt{h}$ times the same scales with index $j - \lceil K/\sqrt{h} \rceil$ for $j > \lceil K/\sqrt{h} \rceil$. The displayed singular scales at index k are comparable to $k^{-1}\exp(-\pi^2 hk^2/4 + O(hk + h))$ for $k \geq 1$. The index shifts are negligible in (6), which holds with $\kappa = \pi^2/4$. The global bound Cj^{-2} follows from the product inequality and the separate Cj^{-2} eigenvalue bounds on G, A_h. The shifted upper estimate also gives $a_{h,j} \leq Che^{-chj^2}$ for $j \geq K'/\sqrt{h}$.

We give the empirical comparison for either sequence. Use the penalty $\eta = e^L/h$, let $J_L = \sqrt{L/h}$, and choose an arbitrarily large fixed T. The Laplace information of the preserving coupling on $j \leq TJ_L$ satisfies

$$\sum_{j \leq TJ_L} I(\eta a_{h,j}) \sim \frac{L^{3/2}}{3\sqrt{\kappa h}}. \tag{7}$$

Indeed $I(a) = \frac{1}{2}\log a + O(1)$ for large a and $I(a) \to 0$ at zero, so the central limit Riemann integrand after dividing by L is $(1 - \kappa v^2)_+/2$. On $j \leq \varepsilon J_L$, the Cj^{-2} bound gives $\sum I(\eta a_{h,j})/(LJ_L) \leq C\varepsilon + o(1)$ by summing $\log(1 + Ce^L/(hj^2))$. Taking $T > \kappa^{-1/2}$ gives the full integral $1/(3\sqrt{\kappa})$. The asymptotic applies for any $L \to \infty, h \to 0$.

Set $L_{\mathrm{crit}} = (3\sqrt{\kappa}\,U)^{2/3}$ and take $L = (1 - \delta)L_{\mathrm{crit}}$ for a fixed $0 < \delta < 1$. Then $J_L = o(R)$, $L = o(R)$, and (7) is strictly below R by a fixed fraction. The likelihood clipping and matrix argument of Proposition 2.3 now give error $e^{-c_\delta R}$, since the centered likelihood on $O(J_L)$ slots has bounded Gaussian-square scales and the cutoff differs from the mean by a fixed fraction of R. In the coupled slots, the squared distortion has at most a constant times he^{-L} per slot in every corresponding fixed moment. Beyond TJ_L the independent squared distortion is bounded in fixed moments by a constant times the sum of the covariance eigenvalues there. Choose T so large that this is $O(\sqrt{h}\,e^{-2L})$ by the heat envelope. The resulting squared moment cost is at most $C\sqrt{h}(1 + \sqrt{L})e^{-L}$ up to the exponentially small clipping error. Because $U \to \infty$ and $h \to 0$, $\sqrt{h} \geq 1/R$ eventually and $L = o(R)$, so the clipping error is negligible.

For the lower bound let $L = (1 + \delta)L_{\mathrm{crit}}$ and use (3) on the same finite band, now with mean information at least $(1 + c_\delta)R$. The expected sum of the two marginal potentials is at least that information. Again the centered potentials have bounded Gaussian-square scales on $O(J_L)$ slots, so deviations by a small fixed fraction of R have probability e^{-cR}. On pairs of typical points at squared distance at most $c'_\delta Rhe^{-L}$, the log likelihood is at least $(1 + c'_\delta)R$, by making c'_δ small. The conditional likelihood identity and the cheap-pair argument of Proposition 2.3 then give negligible expected mass at that cost. This proves matching upper and lower logarithmic asymptotics after δ tends to zero. Clipped bounded differences continue to give summable absolute deviations at the lower cost scale, since it is $e^{-o(R)}$ and the smaller sample size is e^R. The same is true for clipping removal by the Gaussian tails. For the cross problem the exact Gaussian reduction

subtracts the population cost and leaves these same-law results for the cross sequence, with centered norm corrections exponentially small in R both in mean absolute value and almost surely. Those corrections are negligible even on the finer scale, which also gives eventual positivity. Substitution of the two values of κ proves the assertions. □

2.5 Optimizing the cross risk

Fix a filter and put $m = \sqrt{g_0 g_1}$, $\Phi(\xi) = \int B(\xi m)$, and $\mathcal{I}(\xi) = \Phi(\xi) - \xi\Phi'(\xi)$. Integrals without limits in this section are over $(0,\infty)$ in u. For $z = \sqrt{t} > 0$, the change of variables $v = zu$ gives $m_t(u) = z^2 m(zu)$ and $\eta_{m_t} = \xi/z^2$ with $\mathcal{I}(\xi) = z$. In particular

$$H_F(z) = z\,[A_F + \Phi'(\xi)], \qquad H_F'(z) = A_F - J(\xi), \qquad J(\xi) = \frac{\Phi(\xi) - 2\xi\Phi'(\xi)}{\xi}.$$

The derivative follows from $\mathcal{I}'(\xi) = -\xi\Phi''(\xi) > 0$. Integration by parts gives

$$J(\xi) = \int [-um'(u) - 2m(u)]B'(\xi m(u))\,du.$$

Boundary terms vanish since $uB(\xi m(u)) \to 0$ at either end. The weight is $-ug_0(u)(\sqrt{F(\pi^2u^2/2)})'$. It is positive and integrable because the filters are strictly decreasing, are $1 + O(x^2)$ at zero, and decay exponentially at infinity. For F_0, F_ℓ monotonicity follows from that of $x/\sinh x$. For F_+ take the ratio $(4\cosh x - 1)/(3\sinh(2x)/(2x))$. The coefficient ratios for the power series in x^2 are 1 at degree zero and $4(2j+1)/(3\cdot 4^j)$ at degree $j \geq 1$. They are nonincreasing with a strict decrease from $j = 1$ to $j = 2$. Differentiating the quotient and pairing terms of distinct degrees gives strict decrease for $x > 0$.

Since B' decreases strictly from 2 to zero, J decreases strictly to zero. Subtracting the homogeneous profile g_0 in the weight and integrating by parts gives $J(0+) = 2\int(g_0 - m)$. Thus

$$H_F'(0+) = -\int (g_0 - g_1) < 0, \qquad H_F'(+\infty) = A_F > 0. \tag{8}$$

Continuity at zero also follows from (1): the information integrals converge for each η by domination, so $\eta_{m_t} \to \eta_{g_0}$ and the distortion integrals converge. Strict monotonicity of H_F' proves strict convexity and the unique positive minimum. It gives a direct scalar characterization by $J(\xi_*) = A_F$, $z_* = \mathcal{I}(\xi_*)$, and $H_F(z_*) = z_*^2/\xi_*$.

For the lower bound on this minimum, consider a profile g_t at a fixed finite $t > 0$. At each u, its penalized cross objective is

$$\inf_{0\leq r<1} \left[-\tfrac{1}{2}\log(1-r^2) + \eta(g_0 + g_t - 2\sqrt{g_0 g_t}\,r)\right].$$

Freeing g_t pointwise gives the unique minimizing variance $g_0 r^2$, then $r^2 = (1 - 1/(2\eta g_0))_+$. At $\eta = \pi^2/2$ the cutoff in u is 1, the integral of the information term is $\int_0^1 -\log u\,du = 1$, and the integrated distortion is $2/\pi^2$. The constraint $g_t(u) > 0$ gives a strict increase in the penalized objective on the positive-measure region $u > 1$. Taking the dual supremum (1) with the bias term therefore gives $H_F(z) > 2/\pi^2$ for each finite positive z, in particular at its attained minimum. This is the unrestricted functional quantization constant for g_0. The value $H_F(0)$ follows from the fixed-covariance Gaussian constant in [6] with eigenvalues $\lambda_j \sim 1/(\pi^2 j^2)$. The negative slope in (8) gives the upper strict inequality.

Finally if $F_a < F_b$ on $(0,\infty)$, strict increase of B and the variational representation (1) order their same-law coefficients. Their cross coefficients have the same value at zero, and (8) gives the

stated strict order near zero. The large-z order is reversed: $A_{F_a} > A_{F_b}$ and $D(m_t)$ stays bounded by $D(g_0)$ by (1). The filters above satisfy

$$F_0 < F_+ < F_3 < F_5 < \cdots \quad \text{on } (0,\infty).$$

For the middle inequality use $3x\cosh x < \sinh(2x) + \sinh x$ from the power series, so $v(x) < (2+\operatorname{sech} x)/3$. It implies $F_+ < v(2-v) < F_3$. This proves Theorem 2.2.

2.6 Linear empirical tangent fluctuations

For $X \sim f\,dx$, let $Y_h(X)$ be the vector determined by

$$\langle Y_h(X), v\rangle = \frac{(A_h^{1/2}v)(X)}{q(X)} \qquad (v \in \mathcal{H}),$$

including $A_0 = G$. This vector is the heat-tensor image of a centered empirical point mass in the ground-state representation. It is well-defined and uniformly bounded, since the form bound gives $A_h^{1/2} : L^2 \to H^1$ uniformly and point evaluation on the circle is bounded on H^1. It is centered since $\langle q, A_h^{1/2}v\rangle = 0$, and its covariance is exactly A_h. Let

$$Z_{h,n} = n^{-1/2}\sum_{i=1}^{n} Y_h(X_i)$$

for independent X_i. Each empirical draw is constructed using its own independent group of n observations.

Theorem 2.5. *If $n/[R^8(\log R)^8] \to \infty$, the cross and same-law W_2 conclusions of Theorem 2.1 hold for the corresponding ensembles of $Z_{h,n}$ and $Z_{0,n}$. A true marginal may also be kept Gaussian. Each independently sampled type may use a separate group size satisfying the same condition.*

Proof. Project both $Z_{h,n}$ and its Gaussian reference to the first $d = \lceil R\log R\rceil$ eigenmodes of S. Expected squared distance to each projection is $O(1/d)$ by the covariance form bound. Smooth each projected law by convolution with $\mathcal{N}(0,\epsilon^2 I_d)$ for $\epsilon = 1/(R\log R)$. Both squared smoothing costs are $O(d\epsilon^2) = O(1/(R\log R))$.

After smoothing, the total variation discrepancy of the projected laws is at most $Cn^{-1/2}\epsilon^{-3}$. To see this, convolve a bounded measurable test with the Gaussian density, and replace the projected summands one at a time by Gaussians from the same covariance. Constant, linear and quadratic Taylor terms cancel. The third derivative along u has magnitude at most $C\epsilon^{-3}\|u\|^3\|\text{test}\|_\infty$, by rotating u to the first coordinate and integrating the third derivative of the one-dimensional Gaussian density. The original and Gaussian summands have uniformly bounded third norm moments. Summing the Taylor remainders proves the claim, with a constant independent of d, h.

Both projected, smoothed laws have uniformly subgaussian norm tails. For the normalized empirical sum this follows from bounded differences applied to its norm, with sensitivity $O(n^{-1/2})$ per observation and mean bounded by $(\operatorname{tr} A_h)^{1/2}$. The projected Gaussian and the smoothing noise have uniformly bounded covariance traces. If two such laws have total variation δ, their maximal coupling has squared cost at most $C\delta\log(2/\delta)$ by integrating the two norm tails on its mismatch event. Here $n^{-1/2}\epsilon^{-3} = o(1/(R\log R))$. We conclude, uniformly for all small h for each fixed tensor choice, that

$$W_2(Z_{h,n}, \mathcal{N}(0, A_h)) = o(R^{-1/2}),$$

where the notation on the left means the distance from the law of $Z_{h,n}$. Coupling corresponding independent cloud points transfers the same bound in root mean squared Wasserstein distance, and the triangle inequality transfers the mean asymptotics. For the almost sure statements use the uniform subgaussian tails of the tangent observations and the clipped bounded differences proof of Proposition 2.3. This controls deviations on the R^{-1} squared scale even when the leading squared cross cost is of larger order $\sqrt{h}$. □

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, independent internal checks and drafting. The proof uses no data.

3 Sobolev transitions and optimal heat times for empirical tangent laws

On a flat torus of dimension at least two, heat-kernel Sinkhorn tensors define Gaussian tangent laws in a scale of negative Sobolev metrics. We compute the finite and supercritical sampling asymptotics and the population regularization bias throughout the range where high frequencies determine that bias. The cross risk of approximating the unregularized tangent law has a unique positive optimal heat-time constant in this range. At the raw tensor's critical Sobolev index, the population bias gains a logarithm. There we determine the second-order optimal heat time and the exact improvement it gives over unregularized sampling. Above that index the bias is density-dependent, and a quantitative sensitivity argument gives the optimal gain relative to the finite-sample baseline. All results hold for fixed smooth positive densities, including nonconstant densities.

3.1 Tangent laws in a spectral Sobolev metric

Let $d \geq 2$ and let f be a fixed smooth strictly positive probability density on $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. Put $q = \sqrt{f}$, $\mathcal{H} = q^\perp \subset L^2(\mathbb{T}^d, dx)$, $S = -\Delta/2 + (\Delta q)/(2q)$, and $G = (2S)^{-1}$ on $\mathcal{H}$. We use each of the following positive compact tensors A_h from [4], with its scalar filter F and smooth-input order m:

A_h	$F(x), \quad v(x) = x/\sinh x$	m
T_h	$v(x)$	2
$\widehat{C}_h$	$(4 - \operatorname{sech} x)v(x)/3$	4
$C_{h,\ell},\ \ell \geq 3$ odd	$1 - (1 - v(x))^\ell$	2ℓ.

In each case $F(0) = 1$ and $1 - F(x) \sim a_F x^m$ for $x \downarrow 0$, with $a_F > 0$ (for the raw filter $a_F = 1/6$). We set $A_0 = G$.

Fix $r > d/2 - 1$ and set $W = (1 + 2S)^{-r/2}$. We represent the Gaussian law in the Sobolev metric by its image in $\mathcal{H}$, with covariance $C_h = WA_hW$. Thus the distance is ordinary Hilbert Wasserstein distance for the image laws $\gamma_h = \mathcal{N}(0, C_h)$. Write

$$\alpha = r + 1 - d/2 > 0, \qquad b = 1 + 2\alpha/d, \qquad \tau_d = 2\pi^2\omega_d^{-2/d}, \qquad c_0 = (2\tau_d)^{-(r+1)},$$

where ω_d is the Euclidean unit ball volume. For $t \geq 0$ and $u > 0$ let

$$g_0(u) = c_0 u^{-b}, \qquad g_t(u) = g_0(u)F(\tau_d t u^{2/d}), \qquad m_t = \sqrt{g_0 g_t}.$$

For a continuous positive $g \le Cu^{-b}$, define a scalar preserving Gaussian constant by

$$B(a) = -\tfrac{1}{2}\log(1-\rho^2) + 2a(1-\rho), \qquad \rho = \rho(a), \quad \rho/(1-\rho^2) = 2a,$$
$$I(a) = B(a) - aB'(a), \qquad \int_0^\infty I(\eta_g g(u))\,du = 1,$$
$$D(g) = \int_0^\infty g(u)B'(\eta_g g(u))\,du = \sup_{\eta>0} \frac{\int_0^\infty B(\eta g(u))\,du - 1}{\eta}.$$

Here $B'(a) = 2(1-\rho(a))$. The value η_g exists uniquely by monotone convergence at infinity and domination at finite η, using $B(a) \lesssim \min(a, 1+\log(1+a))$, $I(a) \le B(a)$ and $aB'(a) \le 1/2$. Differentiating in η proves the variational identity.

Let $R = \log N$ for a population law compared with an equal-mass empirical cloud of N points. In the two-cloud case all points in both clouds are independent and $R = \log\min(N, M)$. Write L_s for a same-law W_s distance between γ_h and its cloud, or between two clouds from γ_h. Write $L_\times$ for the W_2 distance from γ_0 to a cloud from γ_h, or between clouds from γ_0 and γ_h, and put

$$P(h) = W_2^2(\gamma_0, \gamma_h), \qquad E_\times = L_\times^2 - P(h).$$

Almost sure statements mean one prescribed triangular sampling schedule with rows indexed by their smaller sample size and with any dependence across rows.

Theorem 3.1. *Fix $r > d/2 - 1$ and let $R \to \infty$.*

1. *If $hR^{2/d} \to t \in [0, \infty)$, then for every fixed $1 \le s < \infty$,*

$$R^{b-1}L_s^2, \quad R^{b-1}(\mathbb{E}L_s^s)^{2/s} \longrightarrow D(g_t).$$

   Also $R^{b-1}E_\times$ and $R^{b-1}\mathbb{E}E_\times$ tend to $D(m_t)$. Random limits are almost sure.

2. *If $h \to 0$ and $U = Rh^{d/2} \to \infty$, the exact decay of each squared sampling term D_R is*

$$\frac{-\log(D_R/h^\alpha)}{U^{2/(d+2)}} \longrightarrow [(d+2)\kappa^{d/2}]^{2/(d+2)},$$

   where $\kappa = \tau_d$ for $D_R = L_s^2$ or $(\mathbb{E}L_s^s)^{2/s}$, and $\kappa = \tau_d/2$ for $D_R = E_\times$ or $\mathbb{E}E_\times$. The random cross excess is eventually positive almost surely.

Theorem 3.2. *For the given filter, assume $0 < \alpha < 2m$. Then*

$$P(h) \sim h^\alpha A_F, \qquad A_F = \int_0^\infty (\sqrt{g_0} - \sqrt{g_1})^2\,du.$$

Consequently the normalized total cross risk at $hR^{2/d} \to t < \infty$ is

$$R^{b-1}\mathbb{E}L_\times^2 \longrightarrow K_F(t) := t^\alpha A_F + D(m_t).$$

The analogous limit holds almost surely for the squared distance. As a function of t^α, $K_F(t)$ is strictly convex on $(0, \infty)$ and has a unique minimum at $t_* > 0$. Its attained value lies strictly between

$$\frac{b}{b-1}c_0(2/b)^{1-b} \qquad \text{and} \qquad D(g_0).$$

The bandwidth $h \sim t_* R^{-2/d}$ minimizes the leading risk among heat times tending to zero.

The following borderline result concerns the raw tensor. Let $r = d/2 + 3$, so $\alpha = 4$, use $F = v$, and denote $\eta_0 = \eta_{g_0}$. Define two constants

$$C_* = \frac{c_0 \tau_d^4}{144}, \qquad K_* = \frac{c_0 \tau_d^2}{12} \int_0^\infty u^{4/d-b} B'(\eta_0 g_0(u))\, du. \tag{9}$$

Both are positive and finite.

Theorem 3.3. *At this raw borderline,*

$$P(h) \sim (d/2) C_* h^4 \log(1/h).$$

If $hR^{2/d}(\log R)^{1/4} \to \lambda \in [0, \infty)$, then $R^{b-1} L_\times^2$ and $R^{b-1}\mathbb{E}L_\times^2$ tend to $D(g_0) + C_*\lambda^4$. In particular the bandwidth that improves on unregularized sampling must be smaller than that leading crossover scale.

For $a_R = (hR^{2/d})^2 \log R$ bounded, the finer expected comparison is

$$R^{b-1}\mathbb{E}L_\times^2 = D(g_0) + \frac{C_* a_R^2 - K_* a_R}{\log R} + o(1/\log R). \tag{10}$$

The unique asymptotically optimal value of a_R for the $1/\log R$ improvement is $K_*/(2C_*)$. It improves the normalized risk by $K_*^2/(4C_* \log R) + o(1/\log R)$. These expectations hold for one or two clouds as above.

The underlying compact tensors and their relative form and smooth-input bounds are from [4]. The computational theorem in [5] treats the nonlinear empirical metric estimator on fixed scores. The present sampling problem concerns whole Gaussian laws generated by the corresponding tangent operators. The circle theory Section 2 and fixed-spectrum Gaussian constants [6] motivate the quantitative comparison here. Related Sobolev covariance approximation for Whittle–Matérn fields is treated in [11]; a Hilbert Sinkhorn-manifold approach based on Matérn regularity is developed in [13], whose Section 3.1 distinguishes the periodic heat kernel from its Matérn hypotheses.

3.2 Scalar Gaussian sampling with a varying spectrum

We record the sampling argument and a quantitative estimate used at the raw borderline. Suppose the eigenvalues $a_{R,j}$ of a centered Gaussian law satisfy $a_{R,j} \le C j^{-b}$ and

$$R^b a_{R,j} - g(j/R) \longrightarrow 0 \tag{11}$$

uniformly on each band $\epsilon R \le j \le TR$. Then the squared same-law moment and almost sure sampling constants on the R^{1-b} scale are $D(g)$ for every fixed finite W_s order. This assertion does not require g to be a power law. We give the proof because the required error controls depend on the growing bandwidth.

For a standard Gaussian pair with correlation $\rho(a)$, the density relative to independence is

$$\exp\{f_a(x) + f_a(y) - a(x-y)^2\}, \qquad f_a(x) = -\tfrac{1}{4}\log(1-\rho^2) + a(1-\rho)x^2.$$

It integrates to one in either variable, and $2\mathbb{E}f_a = B(a)$. Under this pair centered log likelihood equals $(\rho/2)(U^2 - V^2)$ for independent standard normals U, V, and $a(X-Y)^2 = 2a(1-\rho)V^2$. The coefficients are uniformly bounded at all $a \ge 0$. This also holds for each centered potential under its Gaussian marginal.

For the upper bound choose $\eta < \eta_g$ and couple the first TR coordinates independently with penalties $a = \eta R^b a_{R,j}$. Couple the remaining coordinates independently without tilting. For small enough ϵ and large R, the mean information divided by R is strictly below one by a fixed margin, since $\int I(\eta g) < 1$. This follows from the central Riemann sum and the head bound $O(\epsilon(1 + |\log \epsilon|))$ from $\sum_{j \le \epsilon R} \log(1 + C(R/j)^b)$. Gaussian-square concentration shows that clipping the total likelihood above $e^{(1-\delta)R}$ loses e^{-cR} coupled probability for some $\delta, c > 0$. The scaled mean distortion is given by the central sum for $\int gB'(\eta g)$, with head error $O(\epsilon)$ and independent tail $O(T^{1-b})$. The scaled squared distortion has limiting constant in each fixed moment, with the same cutoff errors, because its total before division by R has bounded quadratic scales, variance $O(R)$ and bounded normalized higher moments.

For two clouds, give entry (i, j) mass $k_0(X_i, Y_j)/(NM)$ using the clipped likelihood k_0. Conditional row means are at most their target masses and their averaged absolute bias is controlled by the clipped mass. The scaled conditional row variance is at most $e^{-\delta R}$ by $k_0^2 \le e^{(1-\delta)R} k_0$ and $M \ge e^R$. The same holds for columns. Trim overloaded margins and couple their deficits, paying exponentially small expected cost by Cauchy–Schwarz and the uniform marginal norm moments from the trace bound. With a population on one side use the integral matrix with population measure in place of the corresponding atoms. Sending T to infinity and η to η_g proves the upper limit.

For the lower bound use η_g on a central band. A point is typical if its potential sum is within δR of its Gaussian mean; the exceptional probability is e^{-cR}. The expectation of the two sums is as close as desired to $R \int B(\eta_g g) = R[1+\eta_g D(g)]$. For typical pairs below squared cost $(D(g)-\delta')R^{1-b}$, the likelihood on the band is therefore at least $e^{(1+c')R}$ with δ' fixed and suitable cutoffs. Conditional integration of the density gives independent probability at most its inverse for such pairs. In the one-cloud case take a union over its atoms. In the two-cloud case each pair can carry at most $1/\max(N, M)$ mass and its expected number is at most $NMe^{-(1+c')R}$. The expected mass at these cheap distances thus vanishes. To pass to almost sure results, project onto the norm ball of radius R and use bounded differences: one clipped point changes the order-s optimal cost by at most $(2R)^s/N$ or $(2R)^s/M$. The uniformly Gaussian norm tail makes removal of clipping negligible by Markov and Borel–Cantelli, even with arbitrarily large second sample size. This proves the sampling assertion.

For two injective Gaussian covariances C, D in the same Hilbert space, the cross empirical squared cost minus its population value equals the same-law empirical squared cost for a diagonal Gaussian law with eigenvalues $w_j = s_j(C^{1/2}D^{1/2})$, up to centered empirical square-norm corrections. To see this, write the input Gaussian as $C^{1/2}$ applied to white coordinates in the left singular basis, and the output as $D^{1/2}$ applied to white coordinates in the right singular basis. The coordinates are recoverable by whitening in covariance eigenbases and applying orthogonal changes of basis. The cross inner product becomes $\sum_j w_j X_j Y_j$. The common diagonal representation preserves all coupling cross terms, and centering the marginal norm differences gives the identity on expanding squared distances. This gives zero mean corrections, exponentially small in R almost surely and in mean absolute value by uniform Gaussian norm moments. Details and the population identity also appear in [6] and Section 2.

3.3 Spectral profiles and supercritical sampling

The operator S is a bounded potential perturbation of the periodic Laplacian, has ground eigenvalue zero, and has Weyl law

$$\lambda_j(S|_{\mathcal{H}}) = \tau_d j^{2/d} + O(j^{1/d} + 1). \tag{12}$$

Indeed, counting lattice points by their unit cubes gives a ball volume remainder of the order of its boundary area, and the bounded potential and the ground-state index shift preserve the stated estimate. In particular the number of positive eigenvalues up to M is $O(1+M^{d/2})$.

Write $A_h^\circ = GF(hS)$, $C_h^\circ = WA_h^\circ W$. Let $Q_h = M_{b_h} e^{h\Delta/2} M_{b_h}$ and $E_h = e^{-hS}$. The published baseline gives $Q_h q = q$, $\|Q_h - E_h\| = O(h^2)$, and $\|(I-Q_h)^{-1}\| = O(h^{-1})$ on $\mathcal{H}$. All operator norms without a displayed domain are on L^2. On $\mathcal{H}$,

$$T_h = hQ_h(I-Q_h^2)^{-1}, \qquad T_h^\circ = hE_h(I-E_h^2)^{-1}.$$

From the two-derivative smoothing, selfadjoint duality and interpolation in [4, Lemmas 6.1–6.2], T_h maps H^{-1} boundedly to H^1. Its form is bounded above by a constant times G on $\mathcal{H}$. The relative estimates in [4, (9) and (48)] give $A_h \asymp T_h$ in form for each selected correction. By elliptic equivalence, these statements use any fixed spectral or usual Sobolev norm. We will use

$$0 \le A_h, A_h^\circ \le CG, \qquad \|(A_h - A_h^\circ)u\| \le Ch^m \|u\|_{H^{2m-2}}. \tag{13}$$

The second estimate follows by comparing each tensor to G, using the strong Sobolev estimates in [4, Lemma 2.1 and Theorems 1.2, 9.1] and the scalar filter at zero.

For fixed $0 < \epsilon < T$, let M also denote the spectral projection onto $\epsilon^2/h < \lambda_j(S) \le T^2/h$. We need the one-sided bound

$$\|M(A_h - A_h^\circ)\| = O_{\epsilon,T}(h^2). \tag{14}$$

For the raw tensor, use

$$M(I-Q_h)^{-1} = M(I-E_h)^{-1}\big[I + (Q_h - E_h)(I-Q_h)^{-1}\big]$$

and the gap bounds. This left product is uniformly bounded for fixed ϵ. The resolvent identity in $2T_h/h = (I-Q_h)^{-1} - (I+Q_h)^{-1}$ then gives $\|M(T_h - T_h^\circ)\| = O_\epsilon(h^2)$. For the compact correction, the extra rational term in [4, (8)] has the form $h\,\varphi(Q_h)$ with no poles on $[0,1]$, and changing Q_h to E_h costs $O(h^3)$. The remaining differential sandwich is $O(h^2)$ by uniform two-derivative smoothing. For a fixed odd hyperpower, write $A_h = T_h P_\ell(2ST_h)$ with $P_\ell(z) = \sum_{k=0}^{\ell-1}(1-z)^k$. Subtract the same expression at T_h°. The initial difference on the left is handled by the raw bound. In the polynomial difference telescope place powers of $2ST_h^\circ$ to the left of $2S(T_h - T_h^\circ)$, after T_h°. They commute with M, $\|MT_h^\circ\| = O_\epsilon(h)$, and $\|MS(T_h - T_h^\circ)\| = O_{\epsilon,T}(h)$. The other factors are bounded by smoothing. This proves (14).

To determine the finite-t profile, consider both $C_h = WA_hW$ and the positive operator $C_0^{1/2} C_h C_0^{1/2}$ whose eigenvalue square roots are the cross singular values. Each has the form $W_* A_h W_*$ with $W_* = W$ for the first case and $W_* = WC_0^{1/2}$ for the second, commuting with S. Their form bounds are the same as those for their respective unregularized references up to a fixed constant. Write $b_* = b$ or $2b$, respectively. Choose low, middle and high spectral slots at cutoffs $\epsilon^2 R^{2/d}$ and $T^2 R^{2/d}$. Removing the low slots changes eigenvalue indices by at most $O(\epsilon^d R + 1)$, by interlacing for the compression. On the middle compression, the covariance difference from its commuting reference is $o(R^{-b_*})$. For $t > 0$ this follows from (14) with substituted positive cutoffs and the two Sobolev weights. For $t = 0$, (13) gives a bound by $C_{\epsilon,T} R^{-b_*} (hR^{2/d})^m$, including exact equality at $h = 0$. The high-high block norm, multiplied by R^{b_*}, is $O(T^{-db_*})$ by the form bound. Positivity bounds the mixed middle-high block by

$$\|MCH\|^2 \le \|MCM\|\|HCH\| \qquad (C \ge 0).$$

After normalization it vanishes as $T \to \infty$ for fixed ϵ. The corresponding middle block is bounded after scaling by a constant depending on ϵ. These estimates and eigenvalue perturbation in operator

norm give central spectral convergence, first sending T to infinity and then ϵ to zero to remove the rank shifts. The reference profiles follow immediately from (12), and are continuous in their index parameter. Taking the square root in the cross case yields (11) with g_t or m_t. The global Cj^{-b} envelopes follow by min–max from the form bounds, and after taking the square root also in the cross case. The finite-t assertions of Theorem 3.1 now follow from Section 2.

For the supercritical assertions, compare T_h with hQ_h on $\mathcal{H}$. The first dominates the second, and for the upper bound one may add to a constant multiple of hQ_h an operator of rank $O(h^{-d/2})$, because Q_h has eigenvalues uniformly bounded away from one beyond that many slots. Indeed the full eigenvalues are the squares of the singular values of $e^{h\Delta/4}M_{b_h}$; the uniform bounds for $b_h^{\pm 1}$ compare these to the free heat singular values. Ground-state removal costs one index. The relative form estimates transfer the same comparisons to each A_h. Conjugate by W_* as above, extending it on q by 1. The squares of the singular values of $W_*M_{b_h}e^{h\Delta/4}$ give the resulting Q_h spectrum on the full space. The multiplier $b_h^{\pm 1}$ is uniformly bounded in each fixed Sobolev order by the smooth bounds of [4, Lemma 2.1]. The norms $\|W_*M_{b_h}u\|$ are uniformly equivalent to $\|u\|_{H^{-r}}$ or $\|u\|_{H^{-(2r+1)}}$ respectively, by ellipticity and the multiplier bounds. Therefore the corresponding singular values are comparable by min–max to those of $(I-\Delta)^{-\beta/2}e^{h\Delta/4}$ with $\beta = r$ or $2r+1$. Restricting to $\mathcal{H}$ costs at most one index. The weighted covariance eigenvalues or cross singular values $a_{h,j}$ consequently have the following profile for $L \to \infty$, $h \to 0$:

$$\frac{\log(a_{h,j}/h^{r+1})}{L} \longrightarrow -\kappa v^{2/d} \quad \text{uniformly for} \quad v = j(h/L)^{d/2} \in [\epsilon, T], \tag{15}$$

where $\kappa = \tau_d$ in the same-law case and $\tau_d/2$ for cross. To check the constants, the free singular scales at index k are comparable to $k^{-\beta/d}\exp[-(h/2)(\tau_d k^{2/d} + O(k^{1/d}+1))]$. They are squared and multiplied by h for the same-law weights, and multiplied by $\sqrt{h}$ for the cross weights. The $O(h^{-d/2})$ shifts are negligible relative to $(L/h)^{d/2}$, and the prefactors after division by h^{r+1} contribute only $O(\log L)$ to logarithms. These bounds also give $a_{h,j} \le Ch^{r+1}e^{-chj^{2/d}}$ beyond a fixed multiple of $h^{-d/2}$.

Put $J_L = (L/h)^{d/2}$ and use penalties $e^L a_{h,j}/h^{r+1}$ on $j \le TJ_L$ for large fixed T. Since $I(a) = (\log a)/2 + O(1)$ at infinity, (15) gives normalized information integral

$$\sum_{j \le TJ_L} I(e^L a_{h,j}/h^{r+1}) \sim \frac{L^{1+d/2}h^{-d/2}}{(d+2)\kappa^{d/2}}.$$

The integral is $\int_0^\infty (1-\kappa v^{2/d})_+/2\,dv$; the head is negligible as its relative cutoff tends to zero by $a_{h,j} \le Cj^{-b}$ and summing $\log(1 + Ce^L/(h^{r+1}j^b))$. Set $L_c = [(d+2)\kappa^{d/2}U]^{2/(d+2)}$. At $L = (1-\delta)L_c$ the information is below R by a fixed relative slack. Section 2's likelihood matrix now pays exponentially small clipping error, since the tilted terms have bounded centered quadratic scales and $J_L = o(R)$. The coupled squared distortion is bounded in the corresponding fixed moments by $Ch^{r+1}e^{-L}J_L = Ch^\alpha e^{-L}L^{d/2}$; the untilted tail costs $O(h^\alpha e^{-2L})$ for sufficiently large T by the heat envelope. At $L = (1+\delta)L_c$ the mean information exceeds R by a fixed fraction, as does the typical potential sum with high probability. The conditional likelihood argument at squared cost $c_\delta R h^{r+1}e^{-L}$ gives negligible cheap mass. Up to powers of L and U these bounds give the asserted exponential scale. Since $L_c = o(R)$ and $h^\alpha \ge R^{-2\alpha/d}$ eventually, the bounded-differences clipping argument applies at this finer scale, and centered norm corrections in the cross identity remain negligible. This proves Theorem 3.1.

3.4 Weighted Gaussian comparison and population bias

We prove for the full range $0 < \alpha < 2m$ that

$$W_2^2(\mathcal{N}(0, C_h), \mathcal{N}(0, C_h^\circ)) = o(h^\alpha). \tag{16}$$

Use the same spectral cutoffs ϵ^2/h and T^2/h as in (14), with projections L, M, H, and put $P = L + M$. By the form bound, the Gaussian projection error in squared distance is at most $Ch^\alpha T^{-2\alpha}$ for each covariance. This follows by summing the tail of the diagonal reference eigenvalues $c_j(C_0) \asymp j^{-b}$ beyond order $(T^2/h)^{d/2}$.

On P the scalar spectral reference satisfies $PA_h^\circ P \geq c_T PGP$ for $c_T > 0$, since the maximum heat argument there is bounded by T^2. On L, (13) implies

$$\|G^{-1/2} L(A_h - A_h^\circ) L G^{-1/2}\| \leq Ch^m (1 + \epsilon^2/h)^m \leq C' \epsilon^{2m} + o(1).$$

All blocks in the P compression touching M have relative norm $O_{\epsilon,T}(h)$ by (14), since $\|G^{-1/2}P\| = O_T(h^{-1/2})$. Choosing ϵ sufficiently small for each fixed T gives $PA_hP \geq (c_T/2)PGP$. In particular both weighted compressed covariances dominate a fixed positive multiple of PC_0P.

We use the following finite-dimensional comparison. If two positive covariances Σ_0, Σ_1 both dominate $c\Sigma_*$ and $\Delta = \Sigma_1 - \Sigma_0$, then, in the eigenbasis of Σ_* with eigenvalues v_i,

$$W_2^2(\mathcal{N}(0, \Sigma_0), \mathcal{N}(0, \Sigma_1)) \leq \frac{1}{2c} \sum_{i,j} \frac{|\Delta_{ij}|^2}{v_i + v_j}. \tag{17}$$

For its proof, interpolate $\Sigma_s = \Sigma_0 + s\Delta$ and solve the selfadjoint Lyapunov equation $X_s\Sigma_s + \Sigma_s X_s = \Delta$. The deterministic linear flow with velocity $X_s x$ transports the centered Gaussian along this segment, because its covariance solves that differential equation. The mean-square action is $\int_0^1 \operatorname{tr}(X_s \Sigma_s X_s)\, ds = (1/2) \int_0^1 \langle \Delta, \mathcal{L}_{\Sigma_s}^{-1} \Delta \rangle\, ds$, where $\mathcal{L}_\Sigma X = X\Sigma + \Sigma X$ on the selfadjoint matrices. Squared distance is at most this action by Cauchy–Schwarz along the flow. Since $\mathcal{L}_{\Sigma_s} \geq c\mathcal{L}_{\Sigma_*}$ as positive operators on the matrix Hilbert space, inversion and diagonalization give (17). This comparison retains the size of the local covariance in the denominator. The Lyapunov action is the usual Gaussian Wasserstein metric tensor [12].

Partition the positive spectrum on P into dyadic bands of size $M_i = 2^i \geq 1$, with the bounded low spectrum included in the first band and splits at the two projection thresholds. Their ranks are $O(M_i^{d/2})$. For bands $M_i \leq M_j$, the unweighted covariance difference has block norm $O(h^m M_i^{m-1})$ by selfadjointness and (13). The two Sobolev weights contribute $O(M_i^{-r/2} M_j^{-r/2})$. The denominator in (17) for these eigenvectors is at least cM_i^{-r-1}. Thus the summed quotient over the block is at most

$$Ch^{2m} M_i^{d/2+2m-1} M_j^{-r}.$$

Summing over $j \geq i$ costs $O(M_i^{-r})$, since $r > 0$. For blocks touching L the sum over i therefore gives at most $Ch^\alpha \epsilon^{2(2m-\alpha)} + o(h^\alpha)$ when $\alpha < 2m$. In the middle-middle block, (14) and the two weights give norm $O_{\epsilon,T}(h^{r+2})$, rank $O_T(h^{-d/2})$, and inverse denominator $O_T(h^{-r-1})$. Its contribution is $O_{\epsilon,T}(h^{\alpha+2})$. Formula (17), together with the two projection errors, proves (16) by making T large and then ϵ small.

For the commuting reference the population squared distance is exactly

$$P^\circ(h) = \sum_j c_j(C_0) \left(1 - \sqrt{F(h\lambda_j(S))}\right)^2. \tag{18}$$

At mesh $h^{d/2}$ in the index, (12) gives a central Riemann sum $h^{-\alpha}P^{\circ}(h) \to A_F$. The high part is bounded by the summable u^{-b} tail, and the head is bounded by $Cu^{4m/d-b}$ after scaling, using $1-\sqrt{F(x)} = O(x^m)$ at zero. The head is integrable exactly for $\alpha < 2m$. Thus (16) transfers the asymptotic to $P(h)$. The exact cross sampling excess already proved gives the finite-t formula in Theorem 3.2. If $h \to 0$ and $hR^{2/d} \to \infty$, the positive population bias grows after normalization, while the sampling excess is negligible compared with the bias.

3.5 The optimal positive heat-time constant

Set $m_1 = \sqrt{g_0 g_1}$ and $\Phi(\xi) = \int_0^\infty B(\xi m_1(u))\,du$. Write $\mathcal{I}(\xi) = \Phi(\xi) - \xi\Phi'(\xi)$ and use $z = t^{d/2}$. Then $m_t(u) = z^b m_1(zu)$, so the variational optimizer has $\eta_{m_t} = \xi/z^b$ with $\mathcal{I}(\xi) = z$. Consequently

$$K_F(t) = z^{b-1}[A_F + \Phi'(\xi)], \qquad \frac{dK_F}{d(z^{b-1})} = A_F - \frac{J_b(\xi)}{b-1}, \quad J_b(\xi) = \frac{\Phi(\xi) - b\xi\Phi'(\xi)}{\xi}.$$

These derivatives follow from $\mathcal{I}' = -\xi\Phi'' > 0$. Integration by parts gives

$$J_b(\xi) = \int_0^\infty [-um_1'(u) - bm_1(u)]B'(\xi m_1(u))\,du.$$

The boundary terms are zero since $uB(\xi m_1(u))$ tends to zero at both ends, and differentiation on fixed positive ξ intervals is justified by the scalar bounds and exponential decay of the filter at infinity. The factor in brackets equals $-ug_0(u)(\sqrt{F(\tau_d u^{2/d})})'$ and is positive for $u > 0$. The raw and hyperpower filters decrease strictly because $x/\sinh x$ does. For the compact correction express F_+ as the ratio $(4\cosh x - 1)/(3\sinh(2x)/(2x))$. The coefficient ratios in powers of x^2 are 1 at degree zero and $4(2j+1)/(3\cdot 4^j)$ at degree $j \geq 1$. Their monotonicity, with a strict decrease at $j = 2$, proves strict decrease of the function by differentiation of the ratio and pairing terms of distinct degrees.

For fixed $\xi > 0$ the integral for J_b is finite, since near zero $aB'(a) \leq 1/2$ reduces the integrand bound to $O(u^{2m/d})$, and the tail is integrable. By monotonicity of B', J_b decreases strictly to zero. At $\xi \downarrow 0$ its limit is $2(b-1)\int(g_0 - m_1)$ when $\alpha < m$, by removing the homogeneous g_0 term and integrating by parts. At $\alpha \geq m$ the integral of the positive bracket diverges near zero, so monotone convergence gives $J_b(0+) = \infty$. In the finite case the initial derivative in $t^\alpha = z^{b-1}$ is $-\int(g_0 - g_1) < 0$; in the infinite case it is negative infinite. The terminal derivative is $A_F > 0$. Continuity at zero follows from dominated convergence for the information and distortion integrals, and strict increase of the derivative gives the claimed unique positive optimum.

For its strict lower bound, the unconstrained benchmark is the sharp Gaussian quantization constant [9, Theorem 2.2]. It can be recovered in the present notation using the marginal-preserving local Gaussian objective at a fixed positive penalty η, now permitting the reproduction variance profile to be any nonnegative function. With input variance g_0, minimize first over the reproduction variance for fixed correlation ρ. The unique optimum is $g_0\rho^2$, and then $\rho^2 = (1 - 1/(2\eta g_0))_+$. Choose η so that the cutoff in u is $q = 2/b$. The integrated information is $\frac{b}{2}\int_0^q \log(q/u)\,du = 1$, and the distortion is

$$c_0 q^{1-b} + \int_q^\infty c_0 u^{-b}\,du = \frac{b}{b-1}c_0(2/b)^{1-b}.$$

For each finite positive t the prescribed profile g_t stays strictly positive on the tail beyond q, giving a strict increase in the integrated penalized objective. The dual supremum for $D(m_t)$, plus the bias term, therefore gives a strict gap at the attained optimum. The negative initial derivative gives the strict upper bound $D(g_0)$. This proves Theorem 3.2.

3.6 The raw logarithmic borderline

In this section $A_h = T_h$ and $\alpha = 4$. In (18), $1 - \sqrt{v(x)} \sim x^2/12$. Thus the low-frequency summand is asymptotic to

$$h^4 \, \frac{c_0 \tau_d^4}{144 j}.$$

To make this uniform, take first a fixed large lower slot and a fixed small upper heat argument δ. The relative error past that lower slot and below δ can be made arbitrarily small by (12) and the scalar Taylor expansion. The remaining high-frequency sum for each fixed $\delta > 0$ is $O_\delta(h^4)$ by the j^{-b} tail. The finitely many lower slots contribute $O(h^4)$. Since the harmonic sum up to heat argument δ is $(d/2)\log(1/h) + O_\delta(1)$, it follows that

$$P^\circ(h) \sim (d/2) C_* h^4 \log(1/h). \tag{19}$$

We next compare the actual covariance to this reference with a stronger cancellation. The published raw expansion in [4, Theorem 1.1, equations (3) and (39)] gives

$$T_h = G + h^2(2G\mathcal{R}G - S/12) + O(h^4 : H^6 \to L^2),$$

where $\mathcal{R}$ is the specified symmetric second-order operator. The reference $T_h^\circ = Gv(hS)$ has the same expansion with the $2G\mathcal{R}G$ term absent. Put $K = G\mathcal{R}$ on $\mathcal{H}$, with the zero extension of G. The operator K and its adjoint are bounded on every fixed Sobolev order, by ellipticity. Define

$$\bar{T}_h = (I + h^2 K) T_h^\circ (I + h^2 K)^*, \qquad \bar{C}_h = W \bar{T}_h W.$$

Since $KG + GK^* = 2G\mathcal{R}G$, the new difference $T_h - \bar{T}_h$ is $O(h^4 : H^6 \to L^2)$. The middle one-sided bound remains $O_{\epsilon,T}(h^2)$, as the congruence changes T_h° by $O(h^2)$ in operator norm. The form comparison above PGP for $\bar{T}_h$ follows from that of T_h°, since the congruence difference has relative norm $O(h)$ on P; the upper form bound by CG holds because K^* is bounded on H^{-1}. The weighted Gaussian comparison of Section 4 now applies to $T_h - \bar{T}_h$ with $m = 4$ and $\alpha = 4$, giving

$$W_2(\mathcal{N}(0, C_h), \mathcal{N}(0, \bar{C}_h)) = o(h^2). \tag{20}$$

Couple the unweighted reference and its congruence with the same Gaussian white coordinates before applying W. The resulting displacement has squared mean $h^4 \operatorname{tr}(WKT_h^\circ K^* W) = O(h^4)$ by the weighted form bound. Thus the distance between actual and commuting covariance laws is $O(h^2)$. The population formula in Theorem 3.3 follows from (19). The leading crossover then follows from Theorem 3.1 with $t = 0$ on the indicated schedule, since

$$R^{b-1} h^4 \log(1/h) = t^4 \big[(2/d)\log R + \log(1/t)\big], \qquad t = hR^{2/d},$$

and $t^4 \log(1/t) \to 0$ as $t \downarrow 0$.

For the second-order comparison first work with the commuting reference. Its reduced cross covariance has eigenvalues

$$w_{R,j} = \frac{(1 + 2\lambda_j(S))^{-r}}{2\lambda_j(S)} \sqrt{v(h\lambda_j(S))}.$$

We need a quantitative version of Section 2. Uniformly for $0 \le t \le 1$, its same-law mean square empirical error equals

$$R^{1-b}\big[D(m_t) + o(1/\log R)\big]. \tag{21}$$

Here is a proof with explicit slack. Put $\delta_R = (\log R)^{-2}$ and make the lower and upper scaled index cutoffs $(\log R)^{-A}$ and $(\log R)^A$ for a sufficiently large fixed A. The bounds $O(\epsilon(1 + |\log \epsilon|))$ and $O(T^{1-b})$ in Section 2 are then $o(\delta_R)$. On the band, the relative eigenvalue errors in replacing $R^b w_{R,j}$ by $m_t(j/R)$ are $O(R^{-1/d}(\log R)^{C_A})$ from (12), uniformly in $t \le 1$. For this observation use $|(\log v)'| \le 1$ on positive arguments. The same polynomial logarithmic bound divided by R suffices for the Riemann sum error, since the u derivative of $\log m_t(u)$ is bounded by a fixed power of $\log R$ on the band. The functions $B(a)$, $I(a)$, and $aB'(a)$ have bounded derivatives with respect to $\log a$: for I this is $aI'(a) = 2a^2\rho'(a)$, bounded by differentiation of $\rho/(1-\rho^2) = 2a$, and the other two follow directly. Thus all normalized central sums have error $o(\delta_R)$.

The critical penalties η_{m_t} range over a fixed compact positive interval for $t \in [0,1]$, by continuity and the information criterion. The derivative of the information integral in η is bounded below there, uniformly in t, by restricting to $1 \le u \le 2$. For the upper bound in Section 2 decrease η_{m_t} by a fixed multiple of δ_R, and for the converse allow a fixed multiple of δ_R loss in the distortion threshold. The mean information or typical log-likelihood gaps are then of size at least $cR\delta_R$. The bounded centered Gaussian-square terms in at most $R(\log R)^A$ slots give exceptional probabilities $\exp[-cR/(\log R)^{A+4}]$ by exponential concentration. They still pay the rectangular clipping and cheap-pair arguments. The change in the scaled distortion integral from the penalty shift is $O(\delta_R)$, since its derivative in η is uniformly integrable by the same scalar formulas. This proves (21), for one or two clouds.

For small t, the variational expression for D gives

$$D(m_t) = D(g_0) - K_* t^2 + o(t^2). \tag{22}$$

Indeed, $t^{-2}(m_t - g_0) \to -c_0 \tau_d^2 u^{4/d-b}/12$ pointwise. In a bounded positive range of penalties the integrand difference divided by t^2, including the factor B', is dominated at infinity by $Cu^{4/d-b}$ and at zero by $Cu^{4/d}$, using $aB'(a) \le 1/2$ and $m_t \asymp g_0$ for small u. Both are integrable at this index. Evaluating the two variational suprema at their respective optimizers bounds their difference between the two corresponding fixed-penalty differences. The optimizer tends to η_0 and dominated convergence proves (22).

It remains to preserve the second-order scale under the nonconstant density. Let $L_h = W(I + h^2 K)W^{-1}$, so $\bar{C}_h = L_h C_h^\circ L_h^*$. The relative similarity

$$C_0^{1/2} L_h C_0^{-1/2} = I + O(h^2)$$

holds in operator norm by Sobolev boundedness of K and spectral elliptic equivalence. Therefore the cross singular values for $\bar{C}_h$ and C_h° differ at most by factors $1 + O(h^2)$ in both directions, by writing the squared cross operators as a congruence with this bounded invertible operator and applying the singular-value inequalities. The reduced empirical squared costs compare by the same multiplicative factors, using coordinatewise contractions. Centered norm corrections have zero expectation in the exact cross identity. The population squared costs for $\bar{C}_h$ and C_h° differ by at most $O(h^4\sqrt{\log(1/h)})$, by their $O(h^2)$ distance and (19). Finally (20), applied also to clouds by independently coupling corresponding draws, changes the normalized root mean square cross distance by $o(t^2)$.

If $a_R = t^2 \log R$ is bounded, the normalized population cost of the commuting reference is

$$C_* a_R^2 / \log R + o(1/\log R)$$

by (19); the term $t^4 \log(1/t)$ is negligible uniformly for bounded a_R. The transfer errors just proved are $o(1/\log R)$ in normalized squared distance. Combining with (21) and (22) gives (10). Completing the square in a_R proves the stated optimum among bounded a_R.

For completeness, no other heat-time scale tending to zero gives a better second-order improvement. If t stays bounded below along a subsequence, the normalized population cost tends to infinity by (19) and the weighted comparison; the expected cross excess is nonnegative. If $t \to 0$ but $a_R \to \infty$, the normalized population term has order $t^4 \log(1/h)$ and dominates $t^2 + 1/\log R$, since $\log(1/h) \geq (2/d) \log R$. The relative population comparison error tends to zero, the sampling reduction is only $O(t^2)$ by (21)–(22), and the actual-to-congruent normalized squared error is bounded by

$$o(t^2)\,[1 + t^2\sqrt{\log(1/h)}],$$

by the root-distance comparison and the population bound. This too is negligible compared with the population term. Every remaining schedule has bounded subsequences of a_R, where (10) applies. The unique minimizing value is therefore necessary and sufficient to attain the stated improvement on the full sequence. This completes the proof.

3.7 The raw tensor above the logarithmic borderline

When $\alpha > 4$, fixed smooth modes determine the leading population bias. Let c_i be the eigenvalues of C_0 in the eigenbasis (e_i) of S, write $\mathcal{L}_f = 2G\mathcal{R}G - S/12$, and put

$$B_r = \frac{1}{2} \sum_{i,j} \frac{|\langle e_i, W\mathcal{L}_f W e_j\rangle|^2}{c_i + c_j}, \qquad K_{\mathrm{samp}} = \frac{c_0 \tau_d^2}{12} \int_0^\infty u^{4/d-b} B'(\eta_{g_0} g_0(u))\, du.$$

Let $\beta_R = R^{b-1}\mathbb{E}L_{\times,0}^2$, with $L_{\times,0}$ denoting the same cross problem at $h = 0$ and with the same sample sizes. By Theorem 3.1, $\beta_R \to D(g_0)$.

Theorem 3.4. *Suppose $\alpha > 4$ and $A_h = T_h$. Then $0 < B_r < \infty$ and $P(h) \sim B_r h^4$. If $h^2 R^{\alpha/d} \to \lambda < \infty$, the normalized cross squared cost converges, almost surely and in expectation, to $D(g_0) + B_r\lambda^2$.*

Define $q_R = R^{(2\alpha-8)/d}$ and $y_R = q_R(hR^{2/d})^2$. For bounded y_R the expected cross risk satisfies

$$R^{b-1}\mathbb{E}L_\times^2 = \beta_R + \frac{B_r y_R^2 - K_{\mathrm{samp}} y_R}{q_R} + o(q_R^{-1}).$$

Among heat-time schedules tending to zero this is optimized by $y_R \to K_{\mathrm{samp}}/(2B_r)$, with gain $K_{\mathrm{samp}}^2/(4B_r q_R) + o(q_R^{-1})$ over the finite-sample baseline β_R.

Population bias. Fix a cutoff $M > 1$ and put $B = W\mathcal{L}_f W$ and $B_M = P_M B P_M$, where P_M projects onto $\lambda(S) \leq M$. Compare C_h to the covariance $C_0 + h^2 B_M$, positive for small h at each fixed M. The dyadic estimate in Section 4 with $m = 2$ applies to the difference on blocks having at least one index outside P_M: the truncated perturbation vanishes there. Summing the quotients in (17) over these blocks gives a bound by

$$Ch^4 \sum_{\substack{M_i \leq M_j \\ M_j \gtrsim M}} M_i^{d/2+3} M_j^{-r} \leq C' h^4 M^{4-\alpha}. \tag{23}$$

For M_i below M this follows by first summing M_j^{-r} on $M_j \gtrsim M$; above M sum first on $M_j \geq M_i$. The group split at the cutoff only changes constants. The same bound without h^4 and without the high heat cutoff proves that the series defining B_r is finite. On blocks lying entirely in P_M, the covariance difference is $O_M(h^4)$ by the finite-mode raw expansion, so the quotient sum is $O_M(h^8)$.

To apply (17), compress to $\lambda \le T^2/h$ with a fixed $T > 1$. The actual covariance is above a fixed multiple of the compressed C_0 there by the proof of Section 4, regardless of α. The finite perturbation has the same property for small h at fixed M. Projection tails cost $O(h^\alpha) = o(h^4)$ as squared distances. Thus

$$\limsup_{h \downarrow 0} h^{-2} W_2(\mathcal{N}(0, C_h), \mathcal{N}(0, C_0 + h^2 B_M)) \le C M^{(4-\alpha)/2}.$$

The finite perturbed covariance agrees with C_0 in all other coordinates, so its population distance is the finite-dimensional one on P_M. Its small-h coefficient is precisely the truncated B_r. To see this, the positive optimal linear map U_h satisfies $U_h C_0 U_h = C_0 + h^2 B_M$ on the finite block. Smooth differentiation of the positive square root, or the invertible Lyapunov equation at the identity, gives $U_h = I + h^2 X + o(h^2)$, where $XC_0 + C_0 X = B_M$. The cost divided by h^4 tends to $\operatorname{tr}(XC_0X) = (1/2)\sum_{i,j \in P_M} |B_{ij}|^2/(c_i + c_j)$. Sending M to infinity proves $P(h) \sim B_r h^4$. This coefficient is positive, since $2G\mathcal{R}G$ is bounded and cannot equal the unbounded $S/12$ on smooth tangents, so B is nonzero. At $h^2 R^{\alpha/d}$ bounded one has $hR^{2/d} \to 0$. Theorem 3.1 gives the asserted leading cross limit.

Sampling sensitivity at small heat time. The exact finite-sample baseline is relevant because its convergence error need not be of order q_R^{-1}. We use a relative estimate valid for every sequence $t = hR^{2/d} \to 0$. For the commuting reference,

$$R^{b-1} \mathbb{E} E_\times^\circ = \beta_R - K_{\mathrm{samp}} t^2 + o(t^2). \tag{24}$$

For its proof, fix a central band $\epsilon R \le j \le TR$ and write $x = t^2$. Approximate the reduced cross weights on this band by the affine weights

$$c_j \left(1 - x R^{-4/d} \lambda_j(S)^2/12\right)$$

and keep the others equal to c_j. These weights are positive for x in a fixed neighborhood of zero, including negative x. Use the same white-coordinate samples for every set of weights. For each sample realization the optimal squared transport cost for the affine weights is the infimum of affine functions of x, since all admissible couplings can be expressed in white coordinates with the same marginals. This uses the measurable inverse of each positive diagonal scaling. Hence its expectation, normalized by R^{b-1}, is concave in x. Its pointwise limit on an interval about zero is the corresponding $D(g)$ with affine profile $g_0(u)[1 - x\tau_d^2 u^{4/d} \mathbf{1}_{[\epsilon,T]}(u)/12]$, by the profile proof in Section 2. Finite jump points in this continuous-away-from-the-jumps profile do not change the Riemann sum argument. The derivative of the limiting profile objective at zero is minus the integral defining K_{samp} restricted to $[\epsilon, T]$, by the variational argument in (22). For every sequence $x_R \downarrow 0$, the secant slopes from zero of the finite concave functions are bracketed by their slopes at fixed neighboring positive and negative arguments. First take $R \to \infty$, then send those neighboring arguments to zero. This gives convergence of the small secant slope to the stated derivative, with no rate of convergence needed for the baseline.

We must remove the affine approximation. On the central band its defect relative to c_j is $O_{\epsilon,T}(x^2)$ by scalar expansion. In the low band the true relative change is at most $Cx\epsilon^{4/d}$ by (12) and the scalar bound $1 - \sqrt{v(z)} \le Cz^2$. Such relative errors change expected squared cost by at most their size times $O(R^{1-b})$, in each direction, by weighted coordinate contraction and an upper comparison with the unregularized spectrum. For the high band, the expected difference in optimized costs is bounded by at most four times the sum of the absolute changes of weights. Indeed for every coupling the coordinate squared differences are bounded in expectation by twice

the sum of the two marginal second moments, including after averaging the empirical clouds. Thus the normalized high-band error is bounded by $CxT^{1-b+4/d}$, again by the scalar bound, and this tends to zero after division by x and letting $T \to \infty$, as $\alpha > 4$. The integral defining K_{samp} converges at both ends by the same bounds as in (22). Now remove both cutoffs. This proves (24); it also covers zero heat time by exact equality there.

Transferring and optimizing the finer comparison. Retain the congruence $\bar{C}_h$ from Section 6. Its reduced cross cost differs from the commuting cost by a factor $1 + O(h^2)$, by the same relative comparison of singular values. In units of R^{1-b} this is $O(h^2) = o(t^2)$ as $R \to \infty$. For the actual-to-congruent distance, the refined weighted comparison with $m = 4$ gives

$$e_h := W_2(\mathcal{N}(0, C_h), \mathcal{N}(0, \bar{C}_h)) \le Ch^k(1 + \mathbf{1}_{\alpha=8}\sqrt{\log(1/h)}), \qquad k = \min(\alpha/2, 4).$$

Indeed take fixed suitable projection cutoffs as in Section 6, use the dyadic sum for the $O(h^4 : H^6 \to L^2)$ difference, and note that its low exponent is $8 - \alpha$. The sum is of order h^α when $\alpha < 8$, $h^8 \log(1/h)$ at equality, and at most h^8 above it. The middle and the tail are smaller or of the indicated order. In all cases $e_h = o(h^2)$. The population distances for $\bar{C}_h$ and C_h are both $O(h^2)$, by the just-proved population result and this comparison. For the total cross expected cost, independently coupling corresponding draws changes its root mean squared distance by at most e_h. Combining the total and population comparisons, the error after normalization in the *sampling excess* is therefore at most

$$Ce_hR^{\alpha/d}(1 + h^2R^{\alpha/d} + e_hR^{\alpha/d}). \tag{25}$$

Here the normalized commuting or congruent sampling excess stays bounded, by comparison with C_0.

If $y_R = q_Rt^2$ is bounded, then $h = O(R^{-(\alpha-2)/d})$. Direct substitution gives $e_hR^{\alpha/d} = o(q_R^{-1})$: the numerator after multiplication by q_R has exponent at most $(3\alpha - 8) - k(\alpha - 2)$ divided by d, which is $-(\alpha - 4)^2/(2d)$ for $4 < \alpha < 8$ and $-\alpha/d$ at and above 8. The possible logarithm at $\alpha = 8$ is harmless. Thus the transfer error (25) is $o(q_R^{-1})$. The population term after normalization is $B_rq_Rt^4(1 + o(1))$. Combine this with (24) to obtain the expansion of Theorem 3.4.

If t is bounded below along a subsequence, the normalized population cost diverges as $q_R \to \infty$. If $t \to 0$ but $y_R \to \infty$, that cost has order q_Rt^4 and dominates the possible $O(t^2)$ sampling gain. It also dominates (25): the ratio of the first term to the population cost is $e_h/(h^4R^{\alpha/d})$, which tends to zero by the bound for e_h and $h \ge R^{-(\alpha-2)/d}$ eventually. The other two terms are negligible since $e_h/h^2 \to 0$. Thus no such subsequence attains a better improvement. Completing the square on bounded y_R proves the unique optimum.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, independent internal checks and drafting. The proof uses no data.

4 High Sobolev bias and optimal sampling for positive heat corrections

For the positive compact fourth-order heat-kernel Sinkhorn correction and each fixed odd hyperpower of length at least three on a flat torus, we determine the population Gaussian bias at and above its high Sobolev threshold, and the exact first improvement from choosing a positive

heat time for empirical transport. At the threshold the bias has a universal logarithmic factor. Above it a weighted covariance tangent gives a density-specific formula for the coefficient. A direct comparison of cross spectral weights gives sampling sensitivity down to arbitrarily small heat times and avoids requiring a convergence rate for the finite-sample unregularized baseline.

4.1 The high Sobolev regimes

Let f be a fixed smooth strictly positive probability density on $\mathbb{T}^d$ with $d \geq 1$. In the real Hilbert space $\mathcal{H} = q^\perp \subset L^2(dx)$ put

$$q = \sqrt{f}, \qquad S = -\Delta/2 + \frac{\Delta q}{2q}, \qquad G = (2S)^{-1}.$$

Let T_h be the heat-kernel Sinkhorn tangent tensor of [4]. For fixed odd $\ell \geq 3$ consider its positive hyperpower

$$A_h = G[1 - (I - 2ST_h)^\ell], \qquad A_h^\circ = GF(hS), \qquad F(x) = 1 - \big(1 - v(x)\big)^\ell, \quad v(x) = x/\sinh x.$$

The first identity equals the finite positive correction in [4, Theorem 9.1]; for this choice $m = 2\ell$ and $a_F = 6^{-\ell}$. Our second choice is the positive compact correction $A_h = \widehat{C}_h$ of [4, Theorem 1.2], with $m = 4$, $F(x) = (4 - \operatorname{sech} x)v(x)/3$, and $a_F = 7/90$. In either case put $A_0 = G$ and write $1 - F(x) \sim a_F x^m$ at zero.

For $r > d/2 - 1$ consider the negative Sobolev image with $W = (1 + 2S)^{-r/2}$ and Gaussian covariances $C_h = WA_hW$, $C_h^\circ = WA_h^\circ W$. Set

$$\alpha = r + 1 - d/2, \quad b = 1 + 2\alpha/d, \quad \tau = 2\pi^2\omega_d^{-2/d}, \quad c_0 = (2\tau)^{-(r+1)}, \quad g_0(u) = c_0 u^{-b},$$

where ω_d is the volume of the Euclidean unit ball. Let c_i be the eigenvalues of C_0 in the eigenbasis (e_i) of S. These satisfy $c_i \sim c_0 i^{-b}$. The squared population bias is

$$P(h) = W_2^2(\mathcal{N}(0, C_0), \mathcal{N}(0, C_h)).$$

Denote by $L_\times$ the W_2 distance from $\mathcal{N}(0, C_0)$ to an equal-mass empirical cloud of $\mathcal{N}(0, C_h)$, or between two independent clouds from the two laws. Put $R = \log N$ or $\log \min(N, M)$, respectively, $t = hR^{2/d}$, and $\beta_R = R^{b-1}\mathbb{E}L_{\times,0}^2$, where the latter is the same experiment with $h = 0$ and the same sample sizes.

We use the scalar preserving Gaussian function B and information I from Section 3, given by

$$\frac{\rho(a)}{1 - \rho(a)^2} = 2a, \qquad B(a) = -\frac{1}{2}\log(1 - \rho(a)^2) + 2a(1 - \rho(a)), \quad I = B - aB'.$$

In particular $B'(a) = 2(1 - \rho(a))$. The parameter $\eta_0 > 0$ solves $\int_0^\infty I(\eta_0 g_0(u))du = 1$. Define

$$C_* = \frac{a_F^2}{4}c_0\tau^{2m}, \qquad K_* = \frac{a_F}{2}c_0\tau^m \int_0^\infty u^{2m/d-b}B'(\eta_0 g_0(u))\, du.$$

The second integral is finite and positive when $\alpha \geq 2m$.

Theorem 4.1. *Suppose $\alpha = 2m$. Then*

$$P(h) \sim (d/2)C_* h^{2m}\log(1/h).$$

For $a_R = t^m \log R$ bounded and $R \to \infty$, the normalized risk is

$$R^{b-1}\mathbb{E}L_\times^2 = \beta_R + \frac{C_* a_R^2 - K_* a_R}{\log R} + o(1/\log R).$$

Among heat-time schedules tending to zero the unique asymptotically optimal value is $a_R \to K_*/(2C_*)$, giving gain $K_*^2/(4C_* \log R) + o(1/\log R)$ over the unregularized finite-sample baseline.

To state the smoother result, use the symmetric differential operator $\mathcal{R}$ of [4, Theorem 1.1], and write Π_q for orthogonal projection onto $\mathcal{H}$. In the hyperpower case define on smooth tangents

$$D_f = S^2/6 - 2\Pi_q \mathcal{R} G, \qquad L_f = -G D_f^\ell.$$

In the compact case use $L_f = L_c$ as explicitly computed in (32). For either choice let

$$B_f = \frac{1}{2}\sum_{i,j} \frac{|\langle e_i, W L_f W e_j\rangle|^2}{c_i + c_j}.$$

Theorem 4.2. *Suppose $\alpha > 2m$. Then $0 < B_f < \infty$ and $P(h) \sim B_f h^{2m}$. Put $q_R = R^{(2\alpha-4m)/d}$ and $y_R = q_R t^m$. For bounded y_R,*

$$R^{b-1}\mathbb{E}L_\times^2 = \beta_R + \frac{B_f y_R^2 - K_* y_R}{q_R} + o(q_R^{-1}).$$

Among heat-time schedules tending to zero the unique asymptotically optimal value is $y_R \to K_*/(2B_f)$, giving gain $K_*^2/(4B_f q_R) + o(q_R^{-1})$ over the unregularized finite-sample baseline.

The leading low-regularity profile and raw tensor's corresponding high-regularity formulas are given in Section 3. Here the leading covariance tangents and sharper bounds below their scalar principal parts are essential. The proof also provides a direct sampling comparison that applies at every high Sobolev index in these theorems. Both results concern the stated positive corrections and fixed smooth densities. The published fixed-score estimator of [5] is a separate statistical object from the full Gaussian tangent cloud here.

4.2 The covariance tangent of an odd hyperpower

On all fixed Sobolev scales the raw expansion from [4, Theorem 1.1, equations (3) and (39)] reads

$$T_h = G + h^2(2G\mathcal{R}G - S/12) + O(h^4 : H^{s+6} \to H^s). \tag{26}$$

For clarity, the displayed result in the donor source has output L^2 including any fixed number of smooth density parameter derivatives. For nonnegative integer s, adjoin spatial translations of the fixed smooth density as parameters. Spatial derivatives of the operator are then controlled by commutators with translations, and expanding ∂^α past the remainder yields (26), just as in equation (51) of [4]. All the displayed terms are symmetric after extension by zero on the moving ground state, so adjoints and interpolation yield the corresponding estimate at negative and fractional orders. The same argument gives the strong bound $T_h - G = O(h^2 : H^{s+2} \to H^s)$.

Consequently on $\mathcal{H}$,

$$M_h := I - 2ST_h = h^2 D_f + O(h^4 : H^{s+8} \to H^s), \qquad M_h = O(h^2 : H^{s+4} \to H^s).$$

Here $2SG = \Pi_q$, explaining the projection in D_f. Telescope $M_h^\ell - (h^2 D_f)^\ell$, using $H^{s+8} \to H^s$ for the changed factor and the order-four bound for every other factor. This gives an error $O(h^{2\ell+2} : H^{s+4\ell+4} \to H^s)$. Use $s = -2$ before applying G, which gains two derivatives. The covariance then satisfies the strong expansion

$$A_h = G + h^m L_f + O(h^{m+2} : H^{2m+2} \to L^2). \tag{27}$$

The reference has the same expansion with $L^\circ = -a_F G S^m$. This follows by bounding the scalar Taylor remainder for F on all positive arguments by a constant times x^{m+2}. The smooth leading operators are symmetric on $\mathcal{H}$, since they are limits of symmetric covariance differences, so

$$A_h - A_h^\circ = h^m L_\Delta + E_h, \qquad L_\Delta = L_f - L^\circ, \qquad E_h = O(h^{m+2} : H^{2m+2} \to L^2) \tag{28}$$

has symmetric terms. The operator L_Δ maps H^{s+2m-6} boundedly to H^s. Indeed every product in $D_f^\ell - (S^2/6)^\ell$ has at least one factor $\Pi_q \mathcal{R} G$ of order zero in place of a factor of order four, leaving order at most $4\ell - 4$ before the left G. This is the lower-order cancellation used below. By symmetry and interpolation, (28) implies

$$|\langle u, (A_h - A_h^\circ)u\rangle| \le C(h^m \|u\|_{H^{m-3}}^2 + h^{m+2}\|u\|_{H^{m+1}}^2). \tag{29}$$

The compact proof below gives the same estimate, even with H^3 in its second term.

4.3 The compact fourth-order correction

For the compact correction put $m = 4$ and use the scalar filter

$$F_+(x) = \frac{4 - \operatorname{sech} x}{3} \frac{x}{\sinh x} = 1 - \frac{7}{90}x^4 + O(x^6).$$

The donor's exact formula is $\widehat{C}_h = T_h + h\varphi(Q_h) - 2h^2 T_h \mathcal{R} T_h$, where $\varphi(z) = z(1-z)/[3(1+z)(1+z^2)]$. We first prove the following strengthened comparison:

$$|\langle u, (\widehat{C}_h - G F_+(hS))u\rangle| \le C(h^4\|u\|_{H^1}^2 + h^6\|u\|_{H^3}^2). \tag{30}$$

Write $S_h = S - h^2\mathcal{R}$, $E_h^* = e^{-hS_h}$, and denote its raw tensor by T_h^*. The exact Lemma 4.1 of [4] gives $Q_h - E_h^* = O(h^5 : H^6 \to L^2)$. The two operators are symmetric, so adjoint interpolation gives the bilinear form bound $O(h^5 : H^3 \to H^{-3})$. We use improved minus-resolvent mapping on the active complement:

$$\|(I - Q_h)^{-1}u\|_{H^3} + \|(I - E_h^*)^{-1}u\|_{H^3} \le C(h^{-1}\|u\|_{H^1} + \|u\|_{H^3}). \tag{31}$$

For E_h^* this follows by scalar spectral calculus and uniform elliptic Sobolev equivalence, since the scalar multiplier on the positive spectrum is at most $C[1+(h\lambda)^{-1}]$. For the other term use the actual rank-one factorization of [4, equation (33) and Lemma 6.1]. Namely $I - Q_h + h\Pi = hM_{b_h}\mathcal{A}_h M_{b_h}$, with Π the projection onto q, $\mathcal{A}_h = (I - H_h)/h + V_h + L_h$, V_h a uniformly smooth bounded multiplier and L_h a uniformly smooth rank-one term. Its inverse is uniformly bounded on both H^1 and H^3 by that lemma. If $K_h = (I - H_h)/h$, then

$$\mathcal{A}_h^{-1}u = (I + K_h)^{-1}[u + (I - V_h - L_h)\mathcal{A}_h^{-1}u].$$

The free multiplier $(I + K_h)^{-1}$ maps $H^1 \cap H^3$ into H^3 with bound $C(\|u\|_{H^1} + h\|u\|_{H^3})$, since its symbol is at most $C[(1+\lambda)^{-1} + h]$ at free Laplacian eigenvalue λ. The uniformly smooth b_h^{-1} multipliers in the factorization preserve the estimate. This proves (31).

The identity for the difference of the two minus resolvents, paired in the quadratic form of $h[(I-Q_h)^{-1}-(I+Q_h)^{-1}]/2$, now costs at most

$$Ch^6\big(h^{-1}\|u\|_{H^1}+\|u\|_{H^3}\big)^2.$$

The plus resolvents and the plus rational factors in $h\varphi$ are bounded on H^3 by perturbation of their free heat counterparts. Their differences cost $O(h^6\|u\|_{H^3}^2)$ in the quadratic form, by telescoping products and inverses between Q_h and E_h^*. Thus the replacement of Q_h by E_h^* in the raw-plus-rational part contributes at most the right side of (30). For the sandwich, $(T_h-G)\mathcal{R}T_h+G\mathcal{R}(T_h-G)$ is $O(h^2 : H^2\to L^2)$: use the strong baseline, smoothing of T_h, and boundedness of $G\mathcal{R}$ on L^2. This symmetric difference interpolates to a quadratic form bounded by $Ch^2\|u\|_{H^1}^2$. Including its prefactor h^2 costs $Ch^4\|u\|_{H^1}^2$.

The resulting scalar reference and leading sandwich are $G_hF_+(hS_h)-2h^2G\mathcal{R}G$, where $G_h=(2S_h)^{-1}$. Its difference from $GF_+(hS)$ is $O(h^4 : H^2\to L^2)$. To prove this, first use the elliptic inverse expansion $G_h-G-2h^2G\mathcal{R}G=O(h^4)$ in operator norm. Write $GF_+(hS)-G=(h^2/2)Se(hS)$ with $e(z)=(F_+(z)-1)/z^2$, and similarly for S_h. On rays of fixed angles strictly between zero and $\pm\pi/2$, e is holomorphic, $O(|z|^2)$ near zero and $O(|z|^{-2})$ at infinity. The resolvent identity on these rays bounds the difference of the resolvents of hS_h, hS by $Ch^2/|z|$ on H^2, since $\mathcal{R}$ is relative order two and the spectral Sobolev norms of S_h are uniformly equivalent to those of S. Integrating in the holomorphic functional calculus gives $e(hS_h)-e(hS)=O(h^2 : H^2\to H^2)$. Multiplying by the factor S and using $S_h-S=O(h^2 : H^2\to L^2)$ yields the claimed $O(h^4 : H^2\to L^2)$ bound. By symmetry and interpolation its form is bounded by $Ch^4\|u\|_{H^1}^2$. This completes (30).

We also identify the leading compact covariance tangent on smooth inputs. Define $\mathcal{R}_4$ as the coefficient of h^5 in the formal logarithm of Q_h. These formal coefficients stand for actual finite-order expansions on every fixed smooth input for positive h. In fact, the scale coefficients of b_h are determined recursively by $b_hH_h(b_hq)=q$, with invertible coefficient 2 in the new unknown at each order after division by q. A truncated smooth polynomial in h formed this way has the stated order of residual in each fixed Sobolev norm by positive-time heat Taylor expansion. The uniform inverse of the scale linearization near $b_h=1$ in [4, Lemma 2.1] bounds the difference from the actual scale by the same order. This gives as many fixed-input Taylor coefficients as needed.

Formal uniqueness in the scale equation gives $b_{-h}=b_h^{-1}$ as formal series, since the formal inverse of Q_h fixes q. Therefore $\log Q_h$ is odd, and its first terms are $-hS+h^3\mathcal{R}+h^5\mathcal{R}_4$. The donor's comparison to E_h^* identifies the cubic coefficient. The fifth coefficient is symmetric, annihilates q, and has differential order at most four. To see the order, expand the formal symmetric product $M_{b_h}e^{h\Delta/2}M_{b_h}$ by the Baker–Campbell–Hausdorff formula, treating $\log b_h$ as a series of multiplication operators. Any contributing Lie word at degree five, apart from a single multiplier, contains at least one multiplication operator and at most four factors of the Laplacian. Each commutator of scalar differential operators saves one order, so its total order is at most four.

These fixed-input expansions can be inserted into the Laurent expression for T_h. A direct justification for the formal minus inverse is to construct the finite parametrix for $h(I-Q_h)^{-1}$ coefficient by coefficient on smooth tangents, using that $(I-Q_h)/h$ has leading term S with smooth elliptic inverse there. Substituting the truncated inverse into $(I-Q_h)/h$ gives a remainder of the desired order on each fixed smooth input by the just-proved Taylor expansions. The actual inverse has bounded L^2 norm, by the gap, so applying it pays the same error. The plus and rational inverses have nonsingular leading coefficient and their remainders are justified in the same way. The sandwich uses the strong raw expansion in each fixed output Sobolev order from Section 2.

Thus, for every fixed smooth tangent, the raw-plus-rational operator has leading symbolic expression $G_{\mathrm{mod}}F_+(hS_{\mathrm{mod}})$ through order four, where $S_{\mathrm{mod}}=S-h^2\mathcal{R}-h^4\mathcal{R}_4$ and $G_{\mathrm{mod}}=(2S_{\mathrm{mod}})^{-1}$

are only formal at this stage. Expanding the scalar filter and the sandwich gives the actual leading tangent

$$L_{\mathrm{c}} = -\frac{7}{180}S^3 + 2G\mathcal{R}_4G - 4G\mathcal{R}G\mathcal{R}G + \frac{1}{6}(S\mathcal{R}G + G\mathcal{R}S), \qquad h^{-4}(\widehat{C}_h - G)u \longrightarrow L_{\mathrm{c}}u. \tag{32}$$

Indeed G_{mod} has coefficients $2G\mathcal{R}G$ at order two and $2G\mathcal{R}_4G + 4G\mathcal{R}G\mathcal{R}G$ at order four. The sandwich cancels the order-two term and at order four subtracts $2(T_2\mathcal{R}G + G\mathcal{R}T_2)$ with $T_2 = 2G\mathcal{R}G - S/12$. This gives (32). The two terms of order two in the formula are the only possible order-two discrepancy with the scalar reference $-7S^3/180$; all others have order at most zero. No positive-time semigroup for an operator of order four is required in this argument.

4.4 Population comparison

We record the covariance bounds needed to apply a weighted Gaussian comparison. From [4, Lemmas 2.1, 6.1, 6.2 and Theorem 9.1], $0 \le A_h \le CG$ in form, and $A_h^\circ \le G$ by the scalar filter. For the compact correction use the relative form bound of the donor's Theorem 1.2. In fact the two-derivative smoothing and selfadjointness give a map $H^{-1} \to H^1$ for T_h, and positive relative form bounds give the result for its hyperpower. On a middle spectral band $\epsilon^2/h < \lambda(S) \le T^2/h$ with projection M, we also have

$$\|M(A_h - A_h^\circ)\| = O_{\epsilon,T}(h^2). \tag{33}$$

To see this, use the Markov operator Q_h of the donor and $E_h^\circ = e^{-hS}$. They differ by $O(h^2)$ in operator norm, and the minus resolvents have full norm $O(h^{-1})$ on $\mathcal{H}$. The reference minus resolvent has norm $O_\epsilon(1)$ on the middle band, and so does $M(I - Q_h)^{-1}$ by the resolvent identity. In the difference of the two raw tensors $h[(I - Q_h)^{-1} - (I + Q_h)^{-1}]/2$ and its analogue for E_h°, the same identity gives the one-sided $O_\epsilon(h^2)$ bound. For the compact correction the extra rational term costs $O(h^3)$ by comparing Q_h and e^{-hS}, and the sandwich costs $O(h^2)$ by two-derivative smoothing. For the other case express the hyperpower as $T_hP_\ell(2ST_h)$ with $P_\ell(z) = \sum_{k=0}^{\ell-1}(1-z)^k$, and telescope its difference. The reference factors commute with M, the reference tensor on this band is $O_\epsilon(h)$, and MS has norm $O_T(h^{-1})$. The middle bound follows, using the uniform smoothing bounds for products with ST_h.

Compress to $P = \mathbf{1}_{\lambda(S)\le T^2/h}$ at a fixed $T > 1$. On P the reference tensor is bounded below by c_TG, since F is positive on compact arguments. On its low subband $\lambda \le \epsilon^2/h$, the strong estimate $A_h - A_h^\circ = O(h^m : H^{2m-2} \to L^2)$ from [4, Theorems 1.2 and 9.1] gives relative operator norm at most $C\epsilon^{2m} + o(1)$ after conjugating by $G^{-1/2}$. On the blocks touching the middle band (33) gives relative norm $O_{\epsilon,T}(h)$. For sufficiently small fixed ϵ this yields $PA_hP \ge (c_T/2)PGP$. These properties hold for every fixed r under consideration. The squared error of approximating either Gaussian by its projection is $O_T(h^\alpha)$ from the Weyl tail $\sum_{j\gtrsim h^{-d/2}} j^{-b}$.

If positive finite covariances Σ_0, Σ_1 both dominate $c\Sigma_*$, and $\Delta = \Sigma_1 - \Sigma_0$, then

$$W_2^2(\mathcal{N}(0,\Sigma_0), \mathcal{N}(0,\Sigma_1)) \le \frac{1}{2c}\sum_{i,j}\frac{|\Delta_{ij}|^2}{v_i + v_j} \tag{34}$$

in the eigenbasis of Σ_* with eigenvalues v_i. Indeed along the linear covariance segment Σ_s solve $X_s\Sigma_s + \Sigma_sX_s = \Delta$ with X_s selfadjoint. The resulting linear transport flow has energy $\int \mathrm{tr}(X_s\Sigma_sX_s)\,ds$ equal to one half the quadratic form of the inverse Lyapunov operator on Δ. As $\Sigma_s \ge c\Sigma_*$, comparison of the corresponding Lyapunov forms proves (34); see also [12] and Section 3.

Use this with the weighted compressed $\Sigma_* = PC_0P$. The Weyl law $\lambda_j(S) \sim \tau j^{2/d}$ gives rank $O(L^{d/2})$ for a dyadic spectral interval of size $L \ge 1$. As in the dyadic comparison of Section 3, if a

symmetric difference obeys $\|Du\| \le B\|u\|_{H^k}$, its block from intervals of sizes $L_i \le L_j$ has operator norm at most $CBL_i^{k/2}$. The two Sobolev weights give $L_i^{-r/2}L_j^{-r/2}$, the rank of the block is at most $CL_i^{d/2}$, and the inverse denominator in (34) is at most CL_i^{r+1}. Therefore the contribution of this block is bounded by

$$CB^2 L_i^{d/2+k+1} L_j^{-r}. \tag{35}$$

Sum over $j \ge i$ using $r > 0$, which holds at all the high indices here even if $d = 1$.

At the logarithmic threshold $\alpha = 2m$, use the common form bound (29) and let $\Gamma_i = h^m L_i^{m-3} + h^{m+2} L_i^{m+1}$ on a dyadic size L_i. Quadratic form comparison gives a block bound $C\sqrt{\Gamma_i \Gamma_j}$: conjugate the symmetric difference by the inverse square root of the positive diagonal form majorant. With $L_i \le L_j$, its weighted contribution to (34) is therefore at most

$$CL_i^{d/2+1} L_j^{-r} \Gamma_i \Gamma_j.$$

Sum over $j \ge i$, using $r > m+1$ at $\alpha = 2m$. The two pure terms have exponents $2m - 4 - \alpha = -4$ and $2m + 4 - \alpha = 4$ on L_i, respectively. They give $O_T(h^{2m})$, as $L_i \le T^2/h$. The mixed terms have exponent $2m - \alpha = 0$ and contribute $O_T(h^{2m+2}\log(1/h))$. Thus

$$W_2(\mathcal{N}(0, C_h), \mathcal{N}(0, C_h^\circ)) = O(h^m).$$

For the commuting population squared cost the summand is $c_j(1 - \sqrt{F(h\lambda_j)})^2$. At this index it is asymptotic to $h^{2m} C_*/j$ for large j and small $h\lambda_j$. The sum up to any fixed small heat argument is $(d/2)\log(1/h) + O(1)$ in the harmonic weight, and the tail beyond that argument is $O(h^{2m})$ up to a constant depending on the fixed argument. This proves the logarithmic leading bias of Theorem 4.1, since the $O(h^m)$ comparison is smaller than its square root. It also shows the population squared costs differ by $O(h^{2m}\sqrt{\log(1/h)})$.

For $\alpha > 2m$, fix $L > 1$ and compare C_h to $C_0 + h^m P_L W L_f W P_L$, with $P_L = \mathbf{1}_{\lambda(S) \le L}$. This is positive for small h at fixed L. On the heat compression it is above a fixed multiple of PC_0P, as is C_h. When both indices lie in P_L, their covariance difference is $o_L(h^m)$ by (27) or (32). When at least one index is outside P_L, use the global strong bound with $k = 2m - 2$, $B = h^m$ in (35). Summing only these blocks costs at most $Ch^{2m}L^{2m-\alpha}$ in (34). For blocks with the smaller size below L, sum the larger indices first starting at L; for smaller sizes above L, sum first over all larger sizes. The same bound without h^{2m} proves finiteness of the high-index tail of the defining series for B_f. The projection error $O(h^\alpha)$ is negligible here. Thus the limsup of the scaled comparison distance $h^{-m}W_2$ to the finite perturbation is at most $CL^{(2m-\alpha)/2}$.

For the finite perturbation the infinitesimal Gaussian cost at covariance C_0 is precisely the truncated series for B_f. The optimal positive linear map satisfies $UC_0U = C_0 + h^m P_L W L_f W P_L$ on P_L, so its derivative in h^m at zero solves the Lyapunov equation. Its action is one half the quadratic form of the inverse Lyapunov operator on the perturbation, as in (34), now with equality to leading order. Sending L to infinity yields $P(h) \sim h^{2m} B_f$. Positivity follows since $L^\circ e_j$ has size comparable to λ_j^{m-1} whereas $L_\Delta e_j = O(\lambda_j^{m-3})$, so L_f does not vanish. This proves the smoother population assertion.

4.5 Cross spectral weights and finite-sample sensitivity

Let

$$w_{h,j} = s_j(C_0^{1/2} C_h^{1/2}), \qquad w_{h,j}^\circ = c_j\sqrt{F(h\lambda_j(S))}$$

be the decreasing cross weights, using the strict decrease of both filters from Section 3, where s_j denotes singular values. The Gaussian cross identity [6] and Section 3 writes the expected squared

cross transport minus $P(h)$ as the expected squared same-law transport for the diagonal Gaussian with weights $w_{h,j}$. Indeed whiten each law in its respective left or right singular basis, in which the cross inner product is $\sum_j w_{h,j} X_j Y_j$. Passing to coordinates $\sqrt{w_{h,j}} X_j, \sqrt{w_{h,j}} Y_j$ preserves all coupling cross terms. The remaining differences are centered marginal squared norms and vanish in expectation. This argument works with either one true law or two independent clouds.

We prove, uniformly for small h and all j, that

$$|w_{h,j} - w^\circ_{h,j}| \le C c_j \min\left(1,\ h^m \lambda_j^{m-2} + (h\lambda_j)^{m+2}\right), \qquad \lambda_j = \lambda_j(S). \tag{36}$$

The common estimate (29) bounds the difference in quadratic form by

$$C h^m \|u\|^2_{H^{m-3}} + C h^{m+2} \|u\|^2_{H^{m+1}}.$$

Consequently, after conjugating by $W_* = C_0^{1/2} W$, the positive cross squared operator lies between its diagonal reference plus or minus a diagonal error bounded in the jth slot by

$$C c_j^2 \left(h^m \lambda_j^{m-2} + (h\lambda_j)^{m+2}\right).$$

Here $W_*^2 G = C_0^2$, so each term in the unweighted bound is divided by the baseline power of G in the corresponding relative error. The two error weight sequences, up to fixed upper factors, may be taken proportional to $\lambda_j^{-2(r+1)+m-2}$ and $\lambda_j^{-2(r+1)+m+2}$, respectively. Both decrease in j at the indicated high Sobolev indices. Thus the upper min–max bound on the subspace excluding the first $j-1$ modes bounds the eigenvalue in position j of the positive cross operator by $(w^\circ_{h,j})^2$ plus the displayed error. Use the first j modes for the lower bound. If $h\lambda_j$ is sufficiently small, the relative errors up to this mode are bounded by a constant times $h^m \lambda_j^{m-2} + (h\lambda_j)^{m+2}$, since F is uniformly bounded below there and the error factors relative to c_j^2 increase with λ_j. The lower eigenvalue is therefore at least $(w^\circ_{h,j})^2$ times one minus this error factor. Taking square roots gives the desired estimate for small $h\lambda_j$. For the other modes, the global form bound $A_h \le CG$ gives $w_{h,j} \le C c_j$ by min–max, and the minimum in (36) is bounded below by a fixed constant. This completes the proof.

Let $t = hR^{2/d} \to 0$. For a fixed $T > 1$ and $j \le TR$, the right side of (36) is

$$c_j \, O_T(t^m R^{-4/d} + t^{m+2}) = c_j \, o_T(t^m),$$

using the Weyl upper bound. In this range the commuting weight is comparable to c_j, so this relative error costs at most $o_T(t^m)$ times a bounded normalized expected transport cost: both diagonal laws are bounded coordinatewise by a constant times the weights c_j, and the unregularized expected cost normalized by R^{b-1} is bounded [6]. For $j > TR$, use $\min(1, x^{m+2}) \le x^m$ for $x \ge 0$, and use the Weyl upper bound again. The sum of the absolute weight errors is at most

$$C \sum_{j > TR} j^{-b} (h j^{2/d})^m \le C R^{1-b} t^m T^{1-b+2m/d}.$$

For any white-coordinate coupling the squared difference in a coordinate is bounded by twice the sum of its marginal squares, which have combined expectation two, even for the empirical marginals after averaging the clouds. Thus the absolute sum gives an upper bound up to a factor four on the change in expected optimized tail cost. This also proves the claim for the optimal coupling when the coordinates are all considered together, by evaluating either coupling for the other weights. Since $\alpha > m$, sending T to infinity after R proves that the normalized expected reduced cost differs

between the actual and commuting laws by $o(t^m)$. No rate of convergence of the baseline was used here.

It remains to compute the commuting first variation. We give the detail because arbitrarily small t can matter in the smoother regime. With $x = t^m$, on a fixed central band $\epsilon R \le j \le TR$ replace $w^\circ_{h,j}$ by the affine weights

$$c_j \left(1 - x\,\frac{a_F}{2} R^{-2m/d} \lambda_j^m\right),$$

and leave other weights equal to c_j. They are positive on a two-sided fixed neighborhood of zero. For one realization of the white-coordinate samples, the squared optimized transport is the infimum of affine dependence on x, since the feasible couplings in white coordinates do not change. Hence its normalized expectation is a concave function of x. For each fixed x in that neighborhood, the Gaussian triangular sampling theorem Section 3 gives its limit as the scalar profile objective

$$D(g_x) = \sup_{\eta>0} \frac{\int B(\eta g_x(u))du - 1}{\eta}, \qquad g_x(u) = g_0(u)\left(1 - x\,\frac{a_F}{2}\tau^m u^{2m/d}\mathbf{1}_{[\epsilon,T]}(u)\right).$$

The proof of that theorem uses Riemann sums of weights indexed by coordinate, which need not be decreasing, so it also applies to profiles continuous except at these two fixed endpoints. Its other hypotheses are the pure power upper bound Cj^{-b} and uniform convergence of the scaled weights on fixed central bands away from the endpoints, supplied by Weyl. The derivative of $D(g_x)$ at zero equals $-K_*$ with its integral restricted to $[\epsilon, T]$. For this, the optimizer η is unique by the information criterion, tends to η_0, and testing the two suprema with their respective optimizers bounds the difference between the two fixed-penalty derivatives. The derivative of the integrand divided by η is $B'(\eta g_0)$ times the weight derivative.

For every positive sequence $x_R \to 0$, concavity bounds the finite expected secant slopes between their slopes at a fixed positive and negative argument. Take $R \to \infty$ at those fixed arguments and then send them to zero. This recovers the derivative for every shrinking sequence, relative to the exact value at $x = 0$ for that R. The affine approximation error on the central band is $O_{\epsilon,T}(x^{1+2/m})$ in relative weights by scalar Taylor expansion. The head has relative change $O(x\epsilon^{2m/d})$ and costs that factor times the bounded normalized expected cost. The tail costs $O(xT^{1-b+2m/d})$ in normalized absolute weights, by the same coordinate moment argument. These latter errors vanish after division by x and removal of the cutoffs. The integral for K_* converges since $B'(a) \le \min(2, 1/(2a))$, so the integrand after multiplication by the power of u is integrable near zero and infinity. We conclude, for every $t \to 0$, that the actual expected cross excess satisfies

$$R^{b-1}(\mathbb{E}L_\times^2 - P(h)) = \beta_R - K_* t^m + o(t^m). \tag{37}$$

At $h = 0$ this equality is exact with zero perturbation. The proof applies to all the sample schedules in the statement. It compares the sampling excess separately from the population distance.

4.6 Optimizing the high-regularity risks

At $\alpha = 2m$ the normalized population bias is asymptotic to

$$(d/2)C_* t^{2m}\left[(2/d)\log R + \log(1/t)\right].$$

For $a_R = t^m \log R$ bounded this equals $C_* a_R^2/\log R + o(1/\log R)$, because $t^{2m}\log(1/t)$ is uniformly negligible on the scale $1/\log R$. Combining with (37) proves the expansion of Theorem 4.1. At

$\alpha > 2m$ the normalized population term is asymptotic to $B_f q_R t^{2m}$. For bounded $y_R = q_R t^m$ this is $B_f y_R^2/q_R + o(1/q_R)$. Again (37) gives the stated expansion.

In each case the quadratic in the rescaled parameter has a unique positive minimum, by $C_*, B_f, K_* > 0$. To see that the optimizer is global among schedules tending to zero, first exclude every subsequence with t bounded below: the normalized population term diverges and the expected excess is nonnegative by the Gaussian cross identity. If $t \to 0$ but the indicated parameter a_R or y_R tends to infinity, the positive population term dominates t^m by the respective asymptotic. The potential sampling gain is $O(t^m)$ by (37). Such subsequences cannot improve the proposed optimum. On every bounded subsequence the expansions apply, and the unique minimizer of the quadratic gives the indicated necessary and sufficient limit for optimality to this order. This proves both theorems.

Statements and disclosures

OpenAI Codex assisted with mathematical exploration, independent internal checks and drafting. The proof uses no data.

Statements and Declarations

Funding. This research was self-funded.

Competing interests. The author owns Luminary Oddities and has no other relevant competing interests.

AI assistance. OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of each part. Anthropic Claude edited, consolidated and cross-referenced the parts and checked computations. Scott Schubr directed the research and is responsible for the content.

References

[1] S. Schubr, Leading bias and positive correction of heat-kernel Sinkhorn metrics, Luminary Research Archive, edition 5 (2026). `https://research.luminaryoddities.org/leading-bias-positive-correction-heat-kernel-sinkhorn-metrics`.

[2] S. Schubr, Statistical inference for empirical heat-kernel Sinkhorn metrics, Luminary Research Archive, edition 5 (2026). `https://research.luminaryoddities.org/statistical-inference-empirical-heat-kernel-sinkhorn-metrics`.

[3] S. Schubr, Sharp convergence of heat-kernel Sinkhorn metrics, Luminary Research Archive, edition 2 (2026). `https://research.luminaryoddities.org/sharp-convergence-of-heat-kernel-sinkhorn-metrics`.

[4] S. Schubr, *Leading bias and positive correction of heat-kernel Sinkhorn metrics*, Luminary Research Archive, edition 5, 2026.

[5] S. Schubr, *Fourier computation of empirical diffusion metrics*, Luminary Research Archive, edition 3, 2026.

[6] S. Schubr, *Wasserstein transport for Kloosterman paths: optimal rates, prime powers and transport constants*, the part titled *Sharp empirical Wasserstein constants for Gaussian measures on Hilbert space*, Luminary Research Archive, 2026.

[7] E. Boissard and T. Le Gouic, *On the mean speed of convergence of empirical and occupation measures in Wasserstein distance*, Ann. Inst. H. Poincaré Probab. Statist. 50 (2014), 539–563.

[8] J. Lei, *Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces*, Bernoulli 26 (2020), 767–798.

[9] H. Luschgy and G. Pagès, *Sharp asymptotics of the functional quantization problem for Gaussian processes*, Ann. Probab. 32 (2004), 1574–1599.

[10] R. T. Powers and E. Størmer, *Free states of the canonical anticommutation relations*, Comm. Math. Phys. 16 (1970), 1–33.

[11] S. Cox and K. Kirchner, *Regularity and convergence analysis in Sobolev and Hölder spaces for generalized Whittle–Matérn fields*, Numer. Math. 146 (2020), 819–873.

[12] L. Malagò, L. Montrucchio and G. Pistone, *Wasserstein Riemannian geometry of Gaussian densities*, Inf. Geom. 1 (2018), 137–179.

[13] P. Navas, *The information geometry of the Sinkhorn divergence on a Hilbert manifold*, arXiv:2610.39850, 2026.