Statistical inference for empirical heat-kernel Sinkhorn metrics

Scott Schubr
Luminary Oddities
scottschubr@luminaryoddities.com

October 31, 2026
Revision draft

Abstract

For a smooth positive density on a flat torus, the inverse weighted Laplacian defines energies of prescribed smooth tangent directions. We estimate a fixed finite matrix of these energies from independent observations, using symmetric heat-kernel Sinkhorn normalization without evaluating the density. The nonconstant spectral gap shrinks with the heat bandwidth, so statistical linearization must control the inverse uniformly in that bandwidth. We prove a scalar Sobolev Taylor estimate that supplies this control. Fixed multiband extrapolation removes the population bias and yields a joint, population-centered root-n Gaussian limit in every fixed dimension under explicit smoothness and bandwidth conditions. The influence function includes normalization, inversion and the changing integration measure. It gives a consistent covariance estimator, an exact nonparametric bootstrap and the pointwise information bound along fixed smooth density submodels. Separate observed error certificates transfer metric and covariance inference to finite numerical outputs. Their finite-arithmetic realization uses effective input enclosures and eventually informative spectral-gap enclosures. A circle example shows why freezing the normalized operator gives an incorrect asymptotic variance; exact rational fixtures check the numerical covariance error budgets. An explicit covariance rate also permits joint confidence cylinders with unknown limiting rank and certified numerical error. A circle example exhibits a noncompact second variation in a direction with zero first-order influence.

Keywords: Sinkhorn metric; diffusion maps; inverse weighted Laplacian; asymptotic linearity; semiparametric efficiency; observed stopping.

1 Introduction

The local metric of a Sinkhorn divergence contains an inverse whose nonconstant spectral gap shrinks with regularization. Consequently a fixed-bandwidth delta method, or consistency of a normalized graph operator, does not by itself give a central limit theorem for the limiting inverse metric. A further distinction is important statistically: replacing the outer population integral by a sample average while freezing the normalized operator omits part of the influence of the sample.

We study a specific periodic heat-kernel model. For an unknown smooth positive density f and fixed smooth functions u_a, the target is the finite matrix of inverse weighted-Laplacian energies of the centered scores $u_a - \int u_a f$. The sample supplies both the normalization and the integration measure. We prove a bandwidth-uniform scalar expansion, remove population bias by fixed-band extrapolation, and identify the complete joint limiting distribution. A deterministic certificate makes the numerical metric target explicit, including the change of mass and centering caused by an imperfect scaling iterate.

A motivating example. On the unit circle with uniform density, take the tangent direction $u(x) = \sqrt{2}\cos(2\pi kx)$, $k \neq 0$, and write $\omega = 4\pi^2 k^2$. The limiting energy is $G^0 = 1/\omega$. Freezing the normalized operator and varying only the outer integral gives influence $\cos(4\pi kx)/\omega$. Varying the entire metric

gives instead $3\cos(4\pi kx)/\omega$, and therefore variance $9/(2\omega^2)$. Section 9 derives this difference. It illustrates why a complete same-sample derivative, rather than a sample average with a fixed inverse, is needed for inference.

Relation to earlier work. Sinkhorn divergence geometry and its Hessian metric are established objects [1]. Symmetric normalization of diffusion kernels, spectral convergence and observed stopping were studied by Wormell–Reich [2]; approximate bistochastic normalization and graph-Laplacian convergence are treated by Cheng–Landa [3]. Kessler–Magnet–Feydy [4], in particular their mass-variation formulas, provide direct antecedents for differentiating the normalization. These are predecessors of the present construction, rather than consequences of it.

Fixed-regularization differentiability and distributional limits for entropic transport are developed by Goldfeld–Kato–Rioux–Sadhu [5]. Root-n graph eigenvalue fluctuations, variance estimation and information-geometric interpretations are studied by Li–García Trillos–Li–Suchan [6]. Efficient inference for smooth transport functionals already appears in Manole et al. [7] and, using elliptic functional derivatives, in Kono [8]. Our observable is instead a fixed-score inverse metric with empirical Sinkhorn normalization and a simultaneous vanishing heat bandwidth. Horizontal metric limits on a fixed finite support [9] concern a different perturbation and sampling regime.

Saidi–Gunsilius–Mena [10] develop differentiability and population-centered Gaussian limits for geometric operations in Sinkhorn geometry at fixed positive regularization. Their Section 1.3 fixes the regularization throughout; Sections 2.2 and 3.5 establish metric-operator differentiability and limit theorems for geodesics, logarithmic maps, parallel transport and the induced distance. Our target is instead the limiting inverse weighted-Laplacian energy as the heat bandwidth vanishes. The uniform shrinking-gap remainder and multiband bias control below provide the additional estimates needed for this joint limit.

The population inverse-metric limit and smooth-input bias were studied in [12, 13]. In particular the reversibility mechanism behind the even tensor expansion belongs to that population analysis. Here we give its finite-jet form at every required order and use it to simplify the empirical bias correction. The new statistical work is the shrinking-bandwidth scalar remainder, complete same-sample influence and covariance, and their combination with certified numerical equivalence. Sobolev perturbation, extrapolation, iid central limits, LAN information bounds and residual-Gram identities are classical tools.

Comparison of statistical targets. The distinction is the observable and its simultaneous limiting regime. The following comparison records the relevant scope of the cited results.

Work	Observable and regime	Relation to the present theorem
Wormell–Reich [2]	Normalized operators and eigendata; shrinking heat scale	Operator and spectral consistency, rather than the population-centered inverse-energy limit considered here.
Goldfeld et al. [5]	Entropic transport maps, potentials and divergences; fixed positive regularization	Differentiability, bootstrap and efficiency precede this work; a bandwidth-uniform inverse estimate is still required here.
Li et al. [6]	Simple graph eigenvalues; shrinking graph radius	Joint Gaussian limits and variance estimation centered at finite-sample expectations; their Remark 8 leaves population centering unresolved in the proved bandwidth range. The operator and observable also differ.
Kono [8]	Scalar quadratic-transport functionals	Elliptic influence and bootstrap precedents; the empirical population-centered CLT has the dimension and remainder restrictions in Theorem 4.3.
Present work	Fixed-score inverse heat-Sinkhorn energies	Population centering in every fixed dimension under smooth positivity, high-order bias removal and the stated conservative bandwidth budget.

From the limit law to computed inference The principal statistical result is the population-centered joint limit in Theorem 2.1. Covariance consistency, fixed-contrast studentization and the multinomial bootstrap give its inferential consequences. There are then two numerical tasks: a metric-error certificate transfers the same root-n limit to the returned metric, while the later weighted-norm certificate controls inexact influence solves and covariance under the original law. The latter pays residual tilting and changing centering as well as inverse error. These tasks have separate inputs and stopping conditions; each is stated with the scale and gap data it uses.

Navas [11] develops the smooth-stratum information geometry and a spectral description of its cotangent form. The small-regularization density limit and sample-estimation issues discussed there remain distinct from the shrinking-bandwidth scalar inference proved here.

2 Model and main results

Let $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ have volume one. Throughout d is fixed, $\mu = f\,\mathrm{d}x$ is a probability measure with $f \in C^\infty(\mathbb{T}^d)$ strictly positive, and $u_1, \dots, u_p$ are fixed real smooth functions. Neither p nor these functions grow with the sample size. Write $\nu(v) = \int v\,\mathrm{d}\nu$ and use the periodic heat kernel

$$H_h = e^{h\Delta/2}, \qquad H_h v(x) = \int k_h(x-y)v(y)\,\mathrm{d}y, \qquad h > 0.$$

For a probability measure ν let $a = a_{h,\nu} > 0$ solve

$$aH_h(a\nu) = 1, \qquad P_{h,\nu}v = aH_h(av\nu). \tag{1}$$

For smooth μ the small-h solution is constructed below. For a finite positive measure on distinct points it is the usual unique symmetric positive scaling; its extension to all x is specified by (1). The operator preserves constants and ν, is selfadjoint and nonnegative in $L^2(\nu)$, and has a simple constant eigenvalue. Define $\Pi_\nu v = \nu(v)1$ and

$$T_{h,\nu} = \frac{h}{2}\left\{(I - P_{h,\nu} + \Pi_\nu)^{-1} - (I + P_{h,\nu})^{-1} - \frac{\Pi_\nu}{2}\right\}. \tag{2}$$

Thus $T1 = 0$, while on the mean-zero space $T = hP(I - P^2)^{-1}$. In particular the last projection in (2) cannot be omitted when the measure varies. The empirical pullback metric is

$$G_{ab}(h,\nu) = \nu[u_a T_{h,\nu} u_b]. \tag{3}$$

For the entropic interpretation use $\varepsilon = 2h$ and the periodic cost $c_h(x,y) = -2h\log[k_h(x-y)/k_h(0)]$. If $\mathrm{OT}_{\varepsilon,c}(\nu,\rho)$ denotes the infimum of $\int c\,\mathrm{d}\pi + \varepsilon\,\mathrm{KL}(\pi \mid \nu\otimes\rho)$ over couplings, its debiased divergence subtracts the two half self-costs. The established diagonal quadratic-coefficient formula is $(\varepsilon/2)P(I - P^2)^{-1}$ on mean-zero scores [1, 12]; hence the convention (2). The analysis below can equivalently take (3) as its definition. Equivalently one may center each input at ν. Its prescribed statistical family is the exponential tilt

$$\nu_\theta = \frac{\exp(\sum_a \theta_a u_a)\nu}{\nu[\exp(\sum_a \theta_a u_a)]}.$$

The prescribed smooth functions therefore determine the local family without evaluations of an unknown density ratio.

Put $u_a^0 = u_a - \mu(u_a)$ and

$$D_f = -f^{-1}\,\mathrm{div}(f\nabla), \qquad D_f\phi_a = u_a^0, \quad \mu\phi_a = 0, \qquad G_{ab}^0 = \mu(u_a^0\phi_b) = \mu(\nabla\phi_a\cdot\nabla\phi_b). \tag{4}$$

Smooth elliptic theory gives the indicated solution. Define

$$\psi_{ab} = u_a^0\phi_b + u_b^0\phi_a - \nabla\phi_a\cdot\nabla\phi_b - G_{ab}^0, \qquad \Sigma_{ab,cd} = \mu(\psi_{ab}\psi_{cd}). \tag{5}$$

The vectorization below uses entries $a \le b$.

Theorem 2.1 (Population-centered empirical inference)**.** *Let $X_1, \dots, X_n$ be iid with law μ and $\mu_n = n^{-1} \sum_i \delta_{X_i}$. Choose an integer $s > d/2$ and set*

$$q = 1 + s/2, \qquad M = 2 + 5s/2. \tag{6}$$

Take an integer K with $2K > M$, a constant $1/(4K) < \beta < 1/(2M)$, and $h_n = n^{-\beta}$. Let $c_0, \dots, c_{K-1}$ be the unique real coefficients satisfying

$$\sum_{\ell=0}^{K-1} c_\ell = 1, \qquad \sum_{\ell=0}^{K-1} c_\ell 4^{-\ell j} = 0 \quad (1 \le j < K). \tag{7}$$

Then

$$\sqrt{n}\ \mathrm{vec}\left\{\widehat{G}_n - G^0\right\} \implies N(0, \Sigma), \qquad \widehat{G}_n = \sum_{\ell=0}^{K-1} c_\ell G(h_n/2^\ell, \mu_n). \tag{8}$$

More precisely, the left side is $n^{-1/2} \sum_i \mathrm{vec}\,\psi(X_i) + o_{\mathbb{P}}(1)$. The density is not an input to $\widehat{G}_n$. The covariance may be singular. There is a consistent sample-computable covariance estimator using exact empirical influence solves, given in Section 6.

An admissible bandwidth choice. The inequalities in Theorem 2.1 prescribe the cost of population centering. For example, in dimension one take $s = 1$, $M = 9/2$, $K = 3$ and $\beta = 1/10$; in dimension three take $s = 2$, $M = 7$, $K = 4$ and $\beta = 1/15$. In each case $1/(4K) < \beta < 1/(2M)$ holds strictly. The dimension and the number of directions remain fixed, and smoothness permits the required finite population expansion. The estimates below retain these sufficient bandwidth budgets explicitly.

Corollary 2.2 (Certified numerical equivalence)**.** *Set $C_K \ge \sum_\ell |c_\ell|$. If the computed metric at each band has a certified operator-norm error at most $n^{-3/4}/C_K$, replacing the exact band metrics in* (8) *preserves its conclusion. Section A constructs a finite first-hit algorithm with this property on every finite distinct sample set. A certified finite-arithmetic realization requires effective convergent enclosures of sites, kernel and score values, and eventually informative spectral-gap enclosures. It asserts neither fast runtime nor fixed-precision success. The metric certificate alone does not certify numerical covariance or studentization; the separate certificate in Section B provides the additional error budget.*

Theorem 2.3 (Information bound)**.** *In the model of all smooth strictly positive densities, the vector* $\mathrm{vec}\,\psi$ *is the canonical gradient of* $\mathrm{vec}\,G^0$*. The estimators in Theorem 2.1 and Corollary 2.2 are regular along each fixed finite smooth exponential density submodel under local $n^{-1/2}$ alternatives. They attain the pointwise semiparametric covariance bound Σ for this notion of regularity. Singular directions are allowed. The inference concerns the fixed density and prescribed tangent directions stated above.*

The metric is a periodic heat-kernel metric, whose cost changes with h; fixed Euclidean squared-cost problems and boundary domains are not included. Signed extrapolation can produce an indefinite matrix at finite n. If G^0 is positive definite, consistency implies positive definiteness with probability tending to one.

3 A fixed-space measure perturbation

We give the analytic estimates needed for the simultaneous limit. Constants may depend on the fixed density, scores, dimension and finitely many derivatives, but not on small h or n. Use the Sobolev norm with Fourier weight $(1 + 4\pi^2|j|^2)^s$. Since $s > d/2$, H^s is an algebra, embeds in C^0, and its multipliers act on H^{-s}. In particular

$$\|H_h\|_{H^{-s} \to H^s} \le C h^{-s}. \tag{9}$$

All measure perturbations have total mass zero. Signed measures and distributions are used locally to define derivatives; probability measures are used for the statistical values.

Lemma 3.1 (Population scale and gap)**.** *At every fixed integer Sobolev order $m > d/2$,*

$$a_{h,\mu} = f^{-1/2} + O_{H^m}(h).$$

The plus inverse $(I + P_{h,\mu})^{-1}$ is uniformly bounded on H^m. The nonconstant $L^2(\mu)$ gap of $P_{h,\mu}$ is at least ch; on H^m,

$$\|(I - P_{h,\mu} + \Pi_\mu)^{-1}\| \le Ch^{-1-m/2}. \tag{10}$$

Proof. Write $a = f^{-1/2}r$. At $r = 1$ the normalization residual is $f^{-1/2}H_h\sqrt{f} - 1 = O_{H^m}(h)$. The derivative in r is the sum of multiplication by $1 + O_{H^m}(h)$ and $P_0 = f^{-1/2}H_h(\sqrt{f}\,\cdot)$. Conjugating by multiplication by $\sqrt{f}$ turns $I + P_0$ into $I + H_h$, whose inverse has norm at most one on H^m. Algebra estimates bound the second derivative uniformly on a fixed neighborhood. A contraction/Newton argument therefore gives $r = 1 + O_{H^m}(h)$ and a uniformly bounded local normalization inverse. This construction works at any prescribed finite high order; uniqueness identifies the solutions. Positivity follows from the C^0 bound.

For the exact scale, conjugation by $\sqrt{f}$ gives $Q = bH_hb$, $b = a\sqrt{f} = 1 + O_{H^m}(h)$. Thus $I + Q$ is an $O(h)$ bounded perturbation of $I + H_h$ on H^m, proving the plus assertion. In $L^2(\mu)$, kernel symmetry gives

$$\mu[v(I - P)v] = \frac{1}{2}\iint k_h(x - y)a(x)a(y)f(x)f(y)(v(x) - v(y))^2 \,\mathrm{d}x\,\mathrm{d}y.$$

The coefficients are uniformly bounded above and below. Fourier Poincaré for $I - H_h$ gives a lower bound ch times the Lebesgue variance. The latter controls the μ variance, since f is bounded above and below and variance is the minimum over constant centers. This proves the gap. Positivity of P follows from that of H_h. Its constant eigenspace is simple by strict positivity of the kernel. Finally, writing $R_- = (I - P + \Pi_\mu)^{-1}$,

$$R_- = I + PR_- - \Pi_\mu, \qquad \|R_-\|_{L^2(\mu)\to L^2(\mu)} \le C/h, \qquad \|P\|_{L^2(\mu)\to H^m} \le Ch^{-m/2}$$

gives (10). □

Lemma 3.2 (Local normalization and inverses)**.** *Let $\epsilon = \|\zeta\|_{H^{-s}}$ and $\nu = \mu + \zeta$. There is a constant $c > 0$ such that on*

$$\epsilon \le ch^{s+q} \tag{11}$$

the normalization and (2) *are smooth maps on fixed Sobolev spaces. Uniformly there,*

$$\|a_{h,\nu}/a_{h,\mu} - 1\|_{H^s} \le C\epsilon h^{-s}, \tag{12}$$

$$\|DP_{h,\nu}\| \le Ch^{-s}, \qquad \|D^2P_{h,\nu}\| \le Ch^{-2s}, \tag{13}$$

$$\|R_{-,\nu}\| \le Ch^{-q}, \qquad \|R_{+,\nu}\| \le C. \tag{14}$$

Derivative norms take H^{-s} directions and output bounded H^s operators; $R_{-,\nu} = (I - P_{h,\nu} + \Pi_\nu)^{-1}$ and $R_{+,\nu} = (I + P_{h,\nu})^{-1}$.

Proof. Put $a = a_{h,\mu}$ and $a_{h,\nu} = ar$. The equation is

$$r\{Pr + aH_h(ar\zeta)\} = 1.$$

Its derivative at $(r, \zeta) = (1, 0)$ is $I + P$, with the uniform inverse from Lemma 3.1. The residual is $O(\epsilon h^{-s})$. On a fixed H^s neighborhood the second r derivative is $O(1 + \epsilon h^{-s})$, and the perturbation of its first derivative is $O(\|r - 1\|_{H^s} + \epsilon h^{-s})$. A quantitative implicit argument gives (12) and

$$\|Dr\| \le Ch^{-s}, \qquad \|D^2r\| \le Ch^{-2s}$$

already on $\epsilon \le ch^s$. Product differentiation then gives (13) and $\|P_{h,\nu} - P_{h,\mu}\| \le C\epsilon h^{-s}$. The rank-one projection perturbation has norm at most $C\epsilon$. By (10) the Neumann criterion for the minus inverse is precisely (11) with sufficiently small c; the plus inverse has a weaker criterion. This proves the claims.

All identities $P1 = 1$, $\nu P = \nu$ and kernel symmetry persist on the signed local extension. On probability measures the local scale is positive by the C^0 estimate. For a finite positive measure it agrees with the unique finite scale: that scale is the critical point of the strictly convex logarithmic scaling energy

$$\frac{1}{2}\sum_{i,j} w_i w_j K_{ij} e^{x_i+x_j} - \sum_i w_i x_i.$$

Strict positivity of the entries, including the diagonal, makes this energy strictly convex and coercive. Its critical equation is the scaling equation. The extension away from the atoms is $a(x) = 1/H_h(a\nu)(x)$ and hence also agrees. Restriction to the atoms intertwines P, Π_ν, both shifted operators and their inverses. Thus the fixed-space functional is the exact finite sample metric, not a quadrature proxy. □

4 Finite smooth expansions and even bias

The next argument also supplies the strong smooth-input estimates used in the scalar Hessian and influence calculations. We include the remainder justification because a formal series alone would not suffice for population centering.

Lemma 4.1 (Finite smooth expansions)**.** *For each prescribed integer order N and each fixed Sobolev order, the population scale, $h(I - P_h)^{-1}u^0$ and $(I + P_h)^{-1}u^0$ have finite asymptotic expansions through order N, with controlled next-order remainders, for every fixed smooth u. Their leading terms are respectively $f^{-1/2}$, $2D_f^{-1}u^0$ and $u^0/2$.*

Proof. On smooth functions the heat Taylor formula is

$$H_h v = \sum_{j=0}^{N} \frac{h^j}{2^j j!}\Delta^j v + O_{H^m}(h^{N+1});$$

its remainder is bounded by a constant times a sufficiently high finite Sobolev norm of v, by the integral semigroup remainder. Insert $a^{[N]} = \sum_{j=0}^{N} h^j a_j$ in $aH_h(af) = 1$. At order zero $a_0 = f^{-1/2}$; at each order j the new coefficient is multiplied by $2\sqrt{f}$. This determines smooth a_j recursively. The residual is $O_{H^m}(h^{N+1})$. The uniform local normalization inverse upgrades this to the same order of error for the exact scale. Consequently on fixed smooth inputs

$$P_h = I - \frac{h}{2}D_f + \sum_{j=2}^{N} h^j L_j + O(h^{N+1}). \tag{15}$$

Each L_j is a finite-order smooth differential operator. Since the exact P_h fixes constants and preserves μ, uniqueness of finite asymptotic coefficients gives $L_j 1 = 0$ and $\mu L_j = 0$.

Let $y_h = h(I - P_h)^{-1}u^0$ with $\mu y_h = 0$. Construct $y^{[N]} = \sum_{j=0}^{N} h^j y_j$ with $y_0 = 2D_f^{-1}u^0$ and all coefficients mean zero. Matching successive orders of $(I - P_h)y = hu^0$ requires

$$\frac{D_f}{2}y_j = \sum_{\ell=2}^{j+1} L_\ell y_{j+1-\ell}.$$

Every right side has mean zero; smooth elliptic solvability applies. The residual of a sufficiently long truncation is $O_{H^m}(h^{N+2})$. Applying the coarse inverse of Lemma 3.1 bounds its error by $Ch^{N+2-(1+m/2)}$. For any desired expansion order, start with a longer truncation than that order plus this fixed loss. The low coefficients are unchanged, giving an expansion to every fixed order. The plus equation is treated by its uniformly bounded inverse and a residual expansion beginning with $u^0/2$. These arguments use finite truncations only. □

Proposition 4.2 (Even population jets)**.** *For each fixed integer $K \geq 1$,*

$$G_{ab}(h,\mu) = G^0_{ab} + \sum_{j=1}^{K-1} b_{ab,2j} h^{2j} + O(h^{2K}). \tag{16}$$

This is the all-finite-orders form of the population reversal mechanism used in [13]; it is used here for empirical bias removal. It does not define backward heat evolution.

Proof. By Lemma 4.1, genuine positive-h expansions exist to every fixed order. We identify their unique coefficients using formal series with smooth coefficients. In this paragraph negative h denotes substitution in finite jets, never a bounded heat operator at negative time. The formal differential-operator series $H(h)$ satisfies $H(h)H(-h) = I$ coefficientwise by the binomial identity. The normalization coefficients are unique. From $H(h)(a(h)f) = 1/a(h)$ one obtains formally $H(-h)(1/a(h)) = a(h)f$. Hence $\widetilde{a} = 1/(fa(h))$ solves the normalization at $-h$ and has the same leading coefficient; uniqueness gives

$$a(-h) = \frac{1}{fa(h)}.$$

If $A(h)$ is multiplication by $a(h)$ and M_f by f, then

$$P(h) = A(h)H(h)A(h)M_f, \qquad P(h)^{-1} = M_f^{-1}A(h)^{-1}H(-h)A(h)^{-1} = P(-h).$$

The order of multiplication is essential here. The fixed μ-mean-zero smooth space is invariant coefficientwise. On it $I - P(h)^2$ has leading term hD_f, so elliptic invertibility gives a unique formal Laurent inverse, also selected by the finite-residual construction. Since all these factors are functions of the same $P(h)$,

$$(-h)P(h)^{-1}(I - P(h)^{-2})^{-1} = hP(h)(I - P(h)^2)^{-1}.$$

The factor h removes the pole. Thus the tensor jet is even; extend it by zero on constants. Every odd coefficient of each fixed smooth bilinear form vanishes. The controlled expansions from Lemma 4.1 now prove (16) for actual positive h. The result is a finite-order asymptotic statement. □

5 Scalar Hessian and exact influence

For a fixed smooth real u set $\Phi_h(\nu) = \nu[uT_{h,\nu}u]$.

Proposition 5.1 (Uniform scalar Taylor bound)**.** *On the neighborhood* (11)*,*

$$\|D^2\Phi_h(\nu)\| \leq Ch^{-M}, \qquad M = 2 + 5s/2. \tag{17}$$

In particular

$$|\Phi_h(\mu+\zeta) - \Phi_h(\mu) - D\Phi_h(\mu)[\zeta]| \leq Ch^{-M}\|\zeta\|^2_{H^{-s}}. \tag{18}$$

Proof. The improved scalar estimate uses the fixed smooth input before applying a coarse inverse bound. Let B_ν be either $I - P_\nu + \Pi_\nu$ or $I + P_\nu$ and $R_\nu = B_\nu^{-1}$. Lemma 4.1 gives $\|R_\mu u\|_{H^s} \leq C/h$, including its bounded constant part. The identity

$$R_\nu u = (I + R_\mu(B_\nu - B_\mu))^{-1}R_\mu u$$

and the same Neumann neighborhood give $\|z\|_{H^s} \leq C/h$ for $z = R_\nu u$, uniformly in ν. Only an inverse acting on a general perturbed input will need the larger loss h^{-q}.

We account explicitly for the moving pairing. Along $\nu_t = \nu + t\zeta$ put $A = \dot{B}_\nu$ and $F(\nu) = \nu[uR_\nu u]$. Then

$$\dot{z} = -R_\nu A z, \qquad \dot{F} = \zeta(uz) - \nu(zAz).$$

Although B_ν is symmetric in the ν pairing, its derivative A need not be. Differentiating the symmetry identity for fixed v, w yields

$$\nu(vAw) - \nu(wAv) = \zeta(wB_\nu v - vB_\nu w).$$

With $v = z$, $w = \dot{z}$, $B_\nu z = u$ and $B_\nu \dot{z} = -Az$, this becomes

$$\nu(zA\dot{z}) - \nu(\dot{z}Az) = \zeta(u\dot{z} + zAz).$$

Differentiating $\dot{F}$ and cancelling the indicated terms gives the exact identity

$$\ddot{F} = 2\nu[(Az)R_\nu(Az)] - \nu[z\ddot{B}_\nu z] - 2\zeta(zAz). \tag{19}$$

It is valid on the signed extension: kernel symmetry and the rank-one projection give the algebraic symmetry there as well. All pairings are continuous by the H^s algebra and duality. For a unit H^{-s} direction, Lemma 3.2 bounds the three terms by

$$Ch^{-2-2s-q}, \qquad Ch^{-2-2s}, \qquad Ch^{-2-s},$$

respectively. Here ν has bounded H^{-s} norm on the ball. Finally the exact uncentered representation is

$$\Phi_h(\nu) = \frac{h}{2}\left\{F_-(\nu) - F_+(\nu) - \frac{\nu(u)^2}{2}\right\}.$$

The last term has a bounded second derivative times h. Multiplying the preceding bounds by h gives $M = 1 + q + 2s = 2 + 5s/2$. Polarization bounds mixed directions; integrating the second derivative along the segment proves (18). A full operator Hessian estimate would have spent three coarse inverse factors and would give a weaker scalar budget; (19) avoids that loss. □

Proposition 5.2 (Influence at smooth and atomic laws)**.** *At a probability measure ν in the local neighborhood, center $g = u - \nu u$ and set*

$$z_- = (I - P_\nu)^{-1} g \quad (\nu z_- = 0), \qquad z_+ = (I + P_\nu)^{-1} g, \qquad C_\nu = (I - P_\nu)(I + P_\nu)^{-1}.$$

The derivative has three sources: the changing outer measure, the normalization scale, and the inverse operator. The first gives the $g(z_- - z_+)$ term below; the latter two combine into the Cayley-operator term. Retaining both is essential even on the uniform circle. The centered derivative representative is

$$\Psi_{h,\nu} = \frac{h}{2}\{g(z_- - z_+) + C_\nu(z_- P_\nu z_- + z_+ P_\nu z_+)\}, \tag{20}$$

$$\psi_{h,\nu} = \Psi_{h,\nu} - \Phi_h(\nu), \qquad D\Phi_h(\nu)[\zeta] = \zeta(\psi_{h,\nu}), \quad \nu\psi_{h,\nu} = 0. \tag{21}$$

These formulas hold for every mass-zero H^{-s} direction, including directions off the support of an atomic ν. At $\nu = \mu$,

$$\psi_{h,\mu} \longrightarrow 2u^0\phi - |\nabla\phi|^2 - G^0(u) \tag{22}$$

in every prescribed finite Sobolev order as $h \downarrow 0$.

Proof. Define $S_\nu\zeta = a_\nu H_h(a_\nu\zeta)$. Differentiating normalization gives $\alpha = \dot{a}/a = -(I + P_\nu)^{-1} S_\nu\zeta$ and

$$\dot{P}_\nu z = \alpha P_\nu z + P_\nu(\alpha z) + S_\nu(z\zeta).$$

Kernel symmetry directly gives

$$\nu[zS_\nu(z\zeta)] = \zeta[zP_\nu z],$$
$$\nu[B(I + P_\nu)^{-1} S_\nu\zeta] = \zeta[P_\nu(I + P_\nu)^{-1} B].$$

Therefore

$$\nu[z\dot{P}_\nu z] = \zeta[C_\nu(zP_\nu z)]. \tag{23}$$

These are dual-kernel identities, not a density argument using perturbations $\nu\nu$. Such perturbations are not dense at a fixed finite atomic law. All their kernels and products are smooth and their pairings extend continuously to H^{-s}.

Differentiate the quadratic resolvent forms. Derivatives of the input center multiply a mean-zero resolvent output and vanish; the derivative of the rank-one projection in the minus inverse has zero quadratic pairing for the same reason. The minus form has derivative $\zeta(gz_-) + \nu(z_-\dot{P}z_-)$; the plus form has derivative $\zeta(gz_+) - \nu(z_+\dot{P}z_+)$. Equation (23) proves (20). Since $\nu C_\nu = 0$, $\nu\Psi_{h,\nu} = \Phi_h(\nu)$, which proves the centering and derivative assertions.

At the population law Lemma 4.1 gives $y_h = hz_- \to 2\phi$ and $z_+ \to u^0/2$ in arbitrarily high prescribed Sobolev orders. Equation (15) and the uniform plus inverse give, on smooth inputs,

$$C_h/h \longrightarrow D_f/4, \qquad \|C_h/h\|_{H^{m+2}\to H^m} \le C.$$

The bound follows from $(I - P_h)/h$ and the plus inverse on H^{m+2}; the limit follows first on fixed smooth functions and then on convergent high-Sobolev inputs. Rewriting the singular contribution to (20) as $\frac{1}{2}(C_h/h)(y_h P_h y_h)$ shows

$$\Psi_{h,\mu} \longrightarrow u^0\phi + \tfrac{1}{2}D_f(\phi^2) = 2u^0\phi - |\nabla\phi|^2.$$

The identity $D_f(\phi^2) = 2\phi D_f\phi - 2|\nabla\phi|^2$ fixes both the sign and coefficient. Its mean is $G^0(u)$ by integration by parts. This proves (22). □

6 Statistical limit and covariance

We first prove the finite-band centered expansion and then address the limit target and covariance.

Proof of Theorem 2.1. For $\zeta_n = \mu_n - \mu$, Fourier duality and independence give

$$\mathbb{E}\|\zeta_n\|^2_{H^{-s}} \le \frac{1}{n}\sum_{j\in\mathbb{Z}^d}(1 + 4\pi^2|j|^2)^{-s} \le C/n. \tag{24}$$

Thus $\epsilon_n = O_{\mathbb{P}}(n^{-1/2})$. The condition $\sqrt{n}h^M \to \infty$ implies that (11) contains ζ_n with probability tending to one. Indeed $M - (s+q) = 1 + s > 0$, so $n^{-1/2}h^{-(s+q)} = (n^{-1/2}h^{-M})h^{1+s} \to 0$. Propositions 5.1 and 5.2 give

$$\sqrt{n}\{\Phi_h(\mu_n) - \Phi_h(\mu)\} = n^{-1/2}\sum_{i=1}^n \psi_{h,\mu}(X_i) + O_{\mathbb{P}}(n^{-1/2}h^{-M}).$$

The remainder tends to zero. The centered triangular sum differs from the sum of the limiting influence by a random variable with variance $\|\psi_{h,\mu} - \psi_{0,\mu}\|^2_{L^2(\mu)} \to 0$. The iid central limit theorem applies, including variance zero. Polarization, for instance $G_{ab} = \{\Phi(u_a + u_b) - \Phi(u_a - u_b)\}/4$, gives the joint expansion with influence (5).

The same event and argument control every fixed number of scores and bands, since all bandwidths are fixed multiples of h_n. Equation (16) and (7) give $\sum_\ell c_\ell G(h_n/2^\ell, \mu) = G^0 + O(h_n^{2K})$. Its scaled bias vanishes since $\beta > 1/(4K)$; the empirical remainders vanish since $\beta < 1/(2M)$. All bands use the same sample and have the same limiting influence. Their weights sum to one, so their linear terms combine into that influence, not into a sum of independent-band variances. This proves (8) and its asymptotic linear representation. □

For the covariance estimator compute (20) at μ_n using a single bandwidth h_n and exact finite solves, and polarize to obtain $\widehat{\psi}_{ab,h_n}$. Set

$$\widehat{\Sigma}_{ab,cd} = \mu_n[\widehat{\psi}_{ab,h_n}\widehat{\psi}_{cd,h_n}]. \tag{25}$$

Proposition 6.1 (Covariance consistency)**.** *Under Theorem 2.1, $\widehat{\Sigma} \to \Sigma$ entrywise in probability. For any fixed contrast with positive limiting variance, the corresponding studentized statistic tends to $N(0,1)$. Inverse-covariance inference requires nonsingularity of the covariance on the selected contrast space.*

Proof. For a scalar score let χ_ν be the unique zero-Lebesgue-mean H^s representative of $D\Phi_h(\nu)$ modulo constants. Duality of mass-zero H^{-s} with H^s modulo constants and (17) give

$$\|\chi_\nu - \chi_\mu\|_{H^s} \le Ch^{-M}\|\nu - \mu\|_{H^{-s}}.$$

The direct atomic calculation in Proposition 5.2 identifies the derivative at an atomic law as well. Thus $\psi_{h,\nu} = \chi_\nu - \nu\chi_\nu$. The population χ_μ is bounded uniformly in H^s by the high-order smooth-input expansions. Since ν is bounded in H^{-s} on the local ball,

$$\|\psi_{h,\nu} - \psi_{h,\mu}\|_{H^s} \le Ch^{-M}\|\nu - \mu\|_{H^{-s}}. \tag{26}$$

This uses the mean-value integral on the whole segment, not differentiation of a pointwise asymptotic formula. For $\nu = \mu_n$ the right side is $o_{\mathbb{P}}(1)$. Split the variance error into

$$\mu_n(\psi^2_{h,\mu_n} - \psi^2_{h,\mu}) + (\mu_n - \mu)(\psi^2_{h,\mu}) + \mu(\psi^2_{h,\mu} - \psi^2_{0,\mu}).$$

The first vanishes by (26) and the algebra bound, the second is $O_{\mathbb{P}}(\epsilon_n)$, and the third vanishes by smooth influence convergence. Polarization and the same product estimates prove every covariance entry. The finite-band limiting influence is unchanged, so the one-band covariance serves (8). Slutsky's theorem gives the studentized conclusion when the limiting denominator is positive. □

Corollary 6.2 (Exact nonparametric bootstrap)**.** *Under Theorem 2.1, let $X_1^*, \dots, X_n^*$ be conditionally iid from μ_n, and compute the same exact multiband functional $\widehat{G}_n^*$ at $\mu_n^* = n^{-1}\sum_i \delta_{X_i^*}$, with the same n, bandwidths and coefficients. Its conditional law satisfies*

$$d_{\mathrm{BL}}\left(\mathcal{L}\left[\sqrt{n}\,\mathrm{vec}(\widehat{G}_n^* - \widehat{G}_n) \mid X_1, \dots, X_n\right], N(0,\Sigma)\right) \longrightarrow 0$$

in probability, where d_{BL} is the bounded-Lipschitz metric on laws in the fixed entry space. The covariance may be singular. This assertion concerns exact empirical metrics, not an uncertified numerical bootstrap implementation.

Proof. Repeated bootstrap sites are merged and given their positive multinomial weights; zero weights are omitted. This leaves the measure and normalized metric unchanged. A one-site law has the zero mean-zero metric. Thus duplicates cause no definitional problem.

Let $E = H^{-s}$ and $C_s = \|\delta_x\|_E^2 < \infty$, independent of x. Conditional Hilbert-space variance gives

$$\mathbb{E}^*\|\mu_n^* - \mu_n\|_E^2 = \frac{1}{n}\{\mu_n\|\delta_x\|_E^2 - \|\mu_n\|_E^2\} \le C_s/n.$$

The cross term with $\mu_n - \mu$ has conditional mean zero. Together with (24) this yields $\mathbb{E}\|\mu_n^* - \mu\|_E^2 \le 2C_s/n$. Both measures consequently lie in the Taylor neighborhood centered at the deterministic true law μ, at all fixed bands, with joint probability tending to one. No uniform Taylor estimate centered at a random atomic law is needed.

Let ψ_h denote the finite-band population influence vector, and put $b_n = \sum_\ell c_\ell \psi_{h_n/2^\ell}$. It has mean zero, converges uniformly to $\mathrm{vec}\,\psi$, and is uniformly bounded. Apply (18) separately to μ_n and μ_n^* at the same μ and subtract:

$$\sqrt{n}\,\mathrm{vec}(\widehat{G}_n^* - \widehat{G}_n) = \sqrt{n}(\mu_n^* - \mu_n)b_n + r_n^*, \qquad r_n^* = o_{\mathbb{P}}(1)$$

under the joint original/bootstrap law. Indeed, on the good neighborhood the expected remainder is at most $Cn^{-1/2}h_n^{-M} \to 0$, while its complement has probability tending to zero. The population tensor

and its bias cancel exactly. For each $\epsilon > 0$, $\mathbb{EP}^*(\|r_n^*\| > \epsilon) = \mathbb{P}(\|r_n^*\| > \epsilon) \to 0$. Markov's inequality therefore gives conditional negligibility in original-sample probability.

Conditionally the leading term is a sum of n iid centered vectors $b_n(X_i^*) - \mu_n b_n$, divided by $\sqrt{n}$. They are uniformly bounded, and their conditional covariance

$$\mu_n(b_n b_n^T) - (\mu_n b_n)(\mu_n b_n)^T \longrightarrow \Sigma$$

in probability by uniform convergence of b_n and the ordinary laws of large numbers for the fixed limiting influence. Conditional Lindeberg holds. To obtain convergence of laws, take any subsequence and a further subsequence on which these covariances converge almost surely. On almost every such sample path the usual triangular-array CLT applies to the conditional arrays, including singular covariance. The subsequence principle proves bounded-Lipschitz convergence in probability on the original sequence. Finally, uniformly over bounded 1-Lipschitz tests, the effect of r_n^* is at most $\epsilon + 2\mathbb{P}^*(\|r_n^*\| > \epsilon)$. Letting $\epsilon \downarrow 0$ proves the corollary.

The bootstrap mechanism is the classical one for asymptotically linear functionals; fixed-regularization Sinkhorn and transport-functional precedents include [5, 8]. Here the shrinking-bandwidth remainder justifies its use for this inverse-metric statistic. □

7 Joint inference with unknown covariance rank

The joint limit permits singular covariance. Spectral regularization of Wald statistics is classical; see Dufour and Valéry [14]. Here the influence estimates give an explicit shrinking-bandwidth error budget, which can also include the paper's numerical covariance certificate.

Vectorize the fixed finite set of metric entries as $\theta \in \mathbb{R}^m$, with target θ_0 and multiband estimator $\widehat{\theta}_n$. Let $\widehat{\Sigma}_n$ be (25). Put

$$a_n = h_n + n^{-1/2} h_n^{-M}.$$

Proposition 7.1 (Covariance rate and spectral selection)**.** *Under Theorem 2.1,*

$$\|\widehat{\Sigma}_n - \Sigma\|_{\mathrm{op}} = O_{\mathbb{P}}(a_n).$$

Suppose a symmetric numerical covariance $\widetilde{\Sigma}_n$ has a nonnegative allowance $\eta_n \to 0$ in probability such that $\|\widetilde{\Sigma}_n - \widehat{\Sigma}_n\|_{\mathrm{op}} \le \eta_n$ on events with probability tending to one. Exact covariance computation corresponds to $\eta_n = 0$. Set $t_n = \sqrt{a_n + \eta_n}$, retain the eigenvalues of $\widetilde{\Sigma}_n$ strictly greater than t_n, and invert only these eigenvalues. Denote the resulting matrix by $\widetilde{\Sigma}^+_{n,t_n}$ and their number by $\widehat{r}_n$. For the fixed rank $r = \mathrm{rank}(\Sigma)$,

$$\mathbb{P}(\widehat{r}_n = r) \longrightarrow 1, \qquad \widetilde{\Sigma}^+_{n,t_n} \longrightarrow \Sigma^+ \quad \textit{in probability.}$$

Proof. The finite smooth expansions, taken one order beyond the limiting calculation in Proposition 5.2, give

$$hz_- = 2\phi + O_{H^{s+2}}(h), \qquad z_+ = u^0/2 + O_{H^{s+2}}(h), \qquad C_h/h = D_f/4 + O(h)$$

on sufficiently smooth inputs. Substitution into (20) yields $\|\psi_{h,\mu} - \psi_{0,\mu}\|_{H^s} = O(h)$. The derivative-stability argument in Proposition 6.1 gives $\|\widehat{\psi}_h - \psi_{h,\mu}\|_\infty = O_{\mathbb{P}}(n^{-1/2}h^{-M})$, including the empirical centering. Split the covariance difference into the empirical influence error, this population influence difference, and the empirical mean error of the fixed smooth products $\psi_0\psi_0^\top$. All influences are bounded in probability, so these three contributions have orders

$$O_{\mathbb{P}}(n^{-1/2}h^{-M}), \qquad O(h), \qquad O_{\mathbb{P}}(n^{-1/2}),$$

respectively. The vector has a fixed finite number of entries; equivalence of its matrix norms gives the stated operator-norm rate.

The total covariance error divided by t_n tends to zero in probability: $a_n/t_n \le \sqrt{a_n}$ and $\eta_n/t_n \le \sqrt{\eta_n}$. Weyl's inequality separates all zero eigenvalues from all fixed positive eigenvalues. Continuous spectral functional calculus on that separated spectrum then proves rank selection and convergence to the Moore–Penrose inverse, including repeated positive eigenvalues. □

Corollary 7.2 (Joint confidence cylinders)**.** *In addition, suppose a computed metric vector satisfies* $\sqrt{n}\|\widetilde{\theta}_n - \widehat{\theta}_n\| \to 0$ *in probability, as provided by the separate metric-error certificate. Then*

$$n(\widetilde{\theta}_n - \theta_0)^\top \widetilde{\Sigma}^+_{n,t_n}(\widetilde{\theta}_n - \theta_0) \;\Rightarrow\; \chi^2_r.$$

Fix $0 < \alpha < 1$*. If* $\widehat{r}_n = 0$*, define* $C_n = \mathbb{R}^m$*. Otherwise use the positive-rank chi-square quantile and set*

$$C_n = \{\theta :\; n(\widetilde{\theta}_n - \theta)^\top \widetilde{\Sigma}^+_{n,t_n}(\widetilde{\theta}_n - \theta) \le \chi^2_{\widehat{r}_n, 1-\alpha}\}$$

For $r \ge 1$ *this region has limiting coverage* $1 - \alpha$ *and is a cylinder in the discarded directions. At* $r = 0$ *the full-space convention gives coverage tending to one.*

Proof. Theorem 2.1, the metric-error condition and Proposition 7.1 give the quadratic-form limit by Slutsky. A Gaussian vector with covariance Σ has $Z^\top \Sigma^+ Z \sim \chi^2_r$. Consistent rank selection gives the stated quantile and coverage when $r \ge 1$, where the limiting distribution is continuous. At rank zero the selected inverse is zero with probability tending to one and the stated full-space convention is conservative. □

The covariance allowance concerns the original empirical law, as in Appendix B; it does not replace the root-n metric-error requirement. The rank and coverage conclusions are pointwise at the fixed density, not uniform over sequences whose positive covariance eigenvalues vanish.

7.1 A singular circle metric and its second variation

On the unit circle take $f = 1$, $u_1 = \sqrt{2}\cos(2\pi kx)$, $u_2 = \sqrt{2}\sin(2\pi kx)$, $k \ge 1$, and $\omega = (2\pi k)^2$. In the order (G_{11}, G_{12}, G_{22}) the limiting influence and covariance are

$$\psi = \frac{3}{\omega}(\cos(4\pi kx), \sin(4\pi kx), -\cos(4\pi kx)), \qquad \Sigma = \frac{9}{2\omega^2}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

Thus $r = 2$, and the trace direction $(1, 0, 1)$ has zero first-order fluctuation. The following computation identifies a distinct high-frequency feature of that trace.

Proposition 7.3 (Noncompact trace Hessian)**.** *Let* $F(f) = G^0_f(u_1, u_1) + G^0_f(u_2, u_2)$*. Its second variation along smooth density perturbations at* $f = 1$ *extends to a bounded self-adjoint operator on mean-zero real* L^2 *of the circle. For* $v_\ell = \sqrt{2}\cos(2\pi\ell x)$ *with* $\ell > k$*, write*

$$F(1 + tv_\ell) = \frac{2}{\omega} + t^2 Q_\ell + o(t^2).$$

Then

$$Q_\ell = \frac{\ell^2}{\omega}\left[\frac{1}{(\ell - k)^2} + \frac{1}{(\ell + k)^2}\right] \longrightarrow \frac{2}{\omega}.$$

In particular, this Hessian is not compact and is not Hilbert–Schmidt.

Proof. For a smooth mean-zero v, let $f_t = 1 + tv$, $L_t = -\partial_x(f_t \partial_x) = L_0 + tL_1$, and $R = L_0^{-1}$ on mean-zero functions. For either score u, put $m_u = \int vu$, $\phi_0 = u/\omega$, and

$$b_t = f_t(u - \mu_{f_t} u) = u + t(vu - m_u) - t^2 m_u v.$$

The energy is $\langle b_t, L_t^{-1} b_t\rangle$. Expanding its inverse gives the quadratic coefficient

$$Q(v) = \sum_{u = u_1, u_2}\left[\langle a_u, Ra_u\rangle - \frac{2m_u^2}{\omega}\right], \qquad a_u = vu - m_u - L_1\phi_0 = v'u'/\omega - m_u.$$

For $\ell > k$, both m_u vanish. Product-to-sum formulas put a_{u_1} in the cosine modes $\ell - k, \ell + k$ with coefficients ℓ/k and $-\ell/k$, and a_{u_2} in the corresponding sine modes with coefficients $-\ell/k, -\ell/k$. Their squared norms and the eigenvalues of R give the displayed Q_ℓ.

For general mean-zero $v \in L^2$, write $a_u = \partial_x(vu')/\omega + vu - m_u$. This has H^{-1} norm bounded by $C\|v\|_2$ and mean zero. Hence Q is a continuous quadratic form on mean-zero L^2 and defines a bounded self-adjoint Hessian. The orthonormal sequence v_ℓ converges weakly to zero, but its Hessian quadratic values are $2Q_\ell \to 4/\omega > 0$. Compactness would force those values to zero. This proves the final assertions. □

The vanishing trace influence therefore coexists with a non-Hilbert–Schmidt continuum second variation. A finite-variance Gaussian quadratic-form limit cannot be obtained by simply substituting this continuum Hessian into a second-order delta argument. A shrinking-bandwidth second-order limit would require its own finite-h centering, normalization and remainder estimates. The projected joint inference above uses the established first-order limit on its nondegenerate range.

8 Canonical gradient and local regularity

Proof of Theorem 2.3. For a smooth centered density score v, let $f_t = fe^{tv}/\mu(e^{tv})$. Work modulo constants in the weak Poisson equation

$$E_f(\phi_a, z) = L_a(f; z), \qquad E_f(x, z) = \mu(\nabla x \cdot \nabla z), \quad L_a(f; z) = \mu[(u_a - \mu u_a)z].$$

Uniform ellipticity for small t and the inverse on the quotient give differentiable smooth solutions; this can also be obtained by an implicit inverse between fixed Sobolev quotient spaces. The derivatives are

$$\dot{E}_f(x, z) = \mu[v\nabla x \cdot \nabla z], \qquad \dot{L}_a(f; z) = \mu(vu_a^0 z) - \mu(vu_a)\mu z.$$

Using the centered baseline solutions as tests, differentiation of $G^0_{ab} = E_f(\phi_a, \phi_b)$ and the two weak equations gives

$$\begin{aligned} DG^0_{ab}(f)[v] &= \dot{L}_a(f; \phi_b) + \dot{L}_b(f; \phi_a) - \dot{E}_f(\phi_a, \phi_b) \\ &= \mu[v(u_a^0\phi_b + u_b^0\phi_a - \nabla\phi_a \cdot \nabla\phi_b)] = \mu(v\psi_{ab}). \end{aligned} \tag{27}$$

This includes derivatives of the score means; they vanish here because the test solutions have baseline mean zero. Smooth centered functions are dense in $L^2_0(\mu)$, since f is bounded above and below. The smooth centered vector ψ is therefore the canonical gradient.

We state regularity precisely rather than presume a uniform density-class remainder. Fix a finite vector v of smooth centered functions and a fixed local vector t. Under the local exponential law $f_{n,t} = f\exp(t^T v/\sqrt{n})/\mu\exp(t^T v/\sqrt{n})$ the likelihood satisfies

$$\log L_n = t^T n^{-1/2} \sum_i v(X_i) - \tfrac{1}{2} t^T \mu(vv^T)t + o_{\mathbb{P}}(1).$$

Boundedness of v also gives $\mathbb{E}_f L_n^2 = O(1)$, by a second-order expansion of its exponential moments. Hence these alternatives are contiguous to the baseline product laws. The baseline asymptotically linear representation in Theorem 2.1 retains its negligible remainder under each such alternative. The joint iid CLT for (ψ, v) and Le Cam's third lemma shift the mean of its linear term to $\mu(\psi v^T)t$, leaving covariance Σ unchanged. Alternatively this follows by expanding

$$\left\{\frac{\mu\exp[(t^T v + iz^T\psi)/\sqrt{n}]}{\mu\exp[t^T v/\sqrt{n}]}\right\}^n.$$

Equation (27) gives exactly the same shift for $\sqrt{n}\{G^0(f_{n,t}) - G^0(f)\}$. After target centering the limit is $N(0, \Sigma)$, independently of t. This proves regularity along every fixed finite smooth exponential submodel. The certified numerical error also transfers by contiguity.

The information conclusion uses classical LAN convolution theory, not the covariance calculation alone; see [15, Theorems 25.20–25.21], and the transport applications [7, 8]. For a scalar contrast $b^T \operatorname{vec} G^0$, write $\chi = b^T \operatorname{vec} \psi$. The supremum of the parametric bounds $(\mu\chi v)^2/\mu(v^2)$ over smooth centered v is $\mu\chi^2$; if nonzero it is attained by the smooth score $v = \chi$. The convolution bound gives this Gaussian component for regular limiting laws, and the present estimator has exactly that normal limit. For the vector, choose a basis of the finite span of its smooth influence components and apply the multivariate LAN bound, or test every contrast. Thus the covariance bound is Σ, including singular directions, and it is attained. This is the claimed pointwise smooth-submodel efficiency; it entails no uniform coverage or minimax smoothing-rate conclusion. □

9 A circle example and the complete variance

Let $f = 1$ on $\mathbb{T}$, $u(x) = \sqrt{2}\cos(2\pi kx)$ with $k \neq 0$, and $\omega = 4\pi^2 k^2$. Then $\phi = u/\omega$, $G^0 = 1/\omega$ and

$$\psi_0(x) = \frac{3}{\omega}\cos(4\pi kx), \qquad \Sigma = \frac{9}{2\omega^2}.$$

The exact finite-band population multiplier is

$$\lambda = e^{-h\omega/2}, \qquad t_h = \frac{h\lambda}{1-\lambda^2} = \frac{h}{2\sinh(h\omega/2)}.$$

Here the exact centered influence is $A_h \cos(4\pi kx)$ with

$$A_h = t_h \left\{ 1 + \frac{(1+\lambda^2)^2}{1+\lambda^4} \right\}. \tag{28}$$

For completeness, the scaling derivative in a mass direction ζ solves $(I + H_h)\alpha = -H_h\zeta$. Since $u^2 = 1 + \cos(4\pi kx)$ and the heat multiplier on the doubled frequency is λ^4, kernel duality gives

$$\langle u, \dot{P}u \rangle_{L^2} = \lambda \frac{1-\lambda^4}{1+\lambda^4}\zeta[\cos(4\pi kx)].$$

The filter $q_h(z) = hz/(1-z^2)$ has derivative $q_h'(\lambda) = h(1+\lambda^2)/(1-\lambda^2)^2$. Adding the outer-measure derivative $t_h\zeta(u^2)$ gives (28), hence $A_h \to 3/\omega$.

Taking $s = 1$ gives $M = 9/2$. Three bands suffice:

$$(c_0, c_1, c_2) = (1/45, -4/9, 64/45), \qquad \sum |c_\ell| = 17/9.$$

They cancel h^2 and h^4. With $\beta = 2/21$, both $\sqrt{n}h^6$ and $n^{-1/2}h^{-9/2}$ equal $n^{-1/14}$. This is a conservative sufficient balancing example.

Freezing the population operator and merely averaging $u\phi$ would instead produce variance $1/(2\omega^2)$. That is an oracle quadrature proxy for a different experiment; it omits normalized-operator and inverse sensitivity. It does not contradict the full-model information bound. If uniformity is known as a model restriction, the target $1/\omega$ is already known and no sample estimator is needed. Efficiency here is relative to unknown nearby smooth positive densities.

10 Parameter regime and numerical inputs

The bandwidth budget gives a nonempty regime in every fixed dimension for an infinitely smooth strictly positive density and a fixed finite collection of prescribed tangent directions. Smoothness pays the finite-order population bias expansion, while positivity supplies the elliptic inverse and the local perturbation estimates. More economical common-space inverse bounds could reduce the sufficient number of bands and enlarge the admissible bandwidth interval.

Computed inference uses two separate observed budgets. The metric certificate controls error in the extrapolated energy matrix; the covariance certificate additionally controls influence solves, residual tilting and recentering under the original law. Their finite-arithmetic realization assumes convergent input enclosures and gap enclosures that eventually detect the positive finite gap. The analytic statistical conclusions are fixed-density asymptotic limits, with fixed contrasts used for studentization.

A Observed metric-error stopping

We now give an actual numerical target and a finite first-hit construction. This section is deterministic and uses no asymptotic empirical gap lower bound. All matrix products and inequalities in the certificates must be evaluated exactly or enclosed rigorously; floating eigenvalues alone are not certificates.

Fix distinct sites, positive probability weights w_i, a symmetric positive definite matrix K with strictly positive entries, a bandwidth $h > 0$, and finitely many prescribed score columns. The periodic heat Gram has these properties. Indeed its Fourier quadratic form has strictly positive multipliers; vanishing of that form forces the finite signed atomic measure to have all Fourier coefficients zero, hence to vanish. Duplicate sites can instead be merged.

A.1 Backward interpretation of a scaling iterate

For any positive vector b put

$$r_i = b_i \sum_j K_{ij} b_j w_j, \quad R = \sum_i w_i r_i, \quad \nu_i = \frac{w_i r_i}{R}, \quad a_i^{\mathrm{tilt}} = \frac{\sqrt{R}\, b_i}{r_i}, \quad P_{ij} = \frac{K_{ij} b_i b_j w_j}{r_i}. \tag{29}$$

Direct substitution proves $a_i^{\mathrm{tilt}} \sum_j K_{ij} a_j^{\mathrm{tilt}} \nu_j = 1$. Thus P is exactly normalized at the residual-tilted measure ν, with detailed balance $\nu_i P_{ij} = w_i w_j K_{ij} b_i b_j / R$. Its Euclidean symmetric representative is

$$\widetilde{Q}_{ij} = \frac{K_{ij} b_i b_j \sqrt{w_i w_j}}{\sqrt{r_i r_j}}, \qquad \widetilde{\Pi} = (\sqrt{\nu})(\sqrt{\nu})^T. \tag{30}$$

Perron simplicity and positivity imply that its nonconstant eigenvalues lie in $(0, 1)$. The correct score columns at this intermediate law are centered at ν, not at w. For example, if $w = (1/2, 1/2)$ but $\nu = (1/2 + e, 1/2 - e)$, the old-centered vector $(1, -1)$ has new mean $2e$ and cannot be the right side of a mean-zero inverse equation. This issue cannot be removed by merely making a normalization residual small.

A.2 A posteriori scale and tensor bounds

Let a be the exact scale at the original weights w, and let $m = \min_i w_i$. Define observed quantities

$$\eta_r = \|r - 1\|_{L^2(w)}, \quad r_{\min} = \min r_i, \quad r_{\max} = \max r_i, \quad R_0 = 2\eta_r / r_{\min}, \quad M_0 = R_0/\sqrt{m}. \tag{31}$$

Lemma A.1 (Observed logarithmic scale bound)**.** *If*

$$q_0 = 2(r_{\max}/r_{\min})(e^{2M_0} - 1) \le 1/2, \tag{32}$$

then $\|\log(a/b)\|_\infty \le M_0$.

Proof. In $L^2(w)$ write $B = \mathrm{diag}(b) K_w \mathrm{diag}(b)$, where $(K_w v)_i = \sum_j K_{ij} w_j v_j$, and $F(x) = e^x b K_w (e^x b) - 1$. Positivity gives $B \ge 0$ and $\|B\| \le r_{\max}$, the latter by the nonnegative row-sum bound and symmetric similarity. $J_0 = DF(0) = \mathrm{diag}(r) + B \ge r_{\min} I$. On the closed $L^2(w)$ ball of radius R_0 we have $\|x\|_\infty \le M_0$. If $D = \mathrm{diag}(e^x)$, then

$$DF(x) = \mathrm{diag}(r(x)) + DBD, \qquad \|DF(x) - J_0\| \le 2 r_{\max}(e^{2M_0} - 1).$$

For the diagonal part use $e^{-2M_0} r_i \le r_i(x) \le e^{2M_0} r_i$; for the matrix part expand $DBD - B = (D - I)BD + B(D - I)$. The Newton map $x - J_0^{-1} F(x)$ is q_0-Lipschitz on the ball and its value at zero has norm at most $R_0/2$. It therefore maps the ball to itself and contracts. Its fixed point is the unique exact positive scale. The zero-residual case is immediate. The ℓ^∞ estimate follows from the $L^2(w)$ radius. No smallest kernel eigenvalue is used. □

Let Q be the exact symmetric matrix at w and $\Pi_w = (\sqrt{w})(\sqrt{w})^T$. With a certified upper bound $p_0 \ge \|\widetilde{\Pi} - \Pi_w\|$ define

$$L = M_0 + \tfrac{1}{2} \max_i |\log r_i|, \quad \epsilon_0 = e^{2L} - 1, \quad e_0 = \epsilon_0 + p_0. \tag{33}$$

For example $p_0 = 2\|\sqrt{\nu} - \sqrt{w}\|_2$ is valid. Equation (30) gives $\widetilde{Q} = DQD$ with $D_{ii} = b_i/(a_i\sqrt{r_i})$. Hence Lemma A.1 implies $\|\widetilde{Q} - Q\| \le \epsilon_0$. Let $g \in (0, 1]$ be a certified lower bound on the nonconstant gap of $\widetilde{Q}$ and require $e_0 < g$. Both shifted matrices $I - Q + \Pi_w$ and $I - \widetilde{Q} + \widetilde{\Pi}$ are positive definite; Weyl's inequality bounds the former below by $(g - e_0)I$. The resolvent identity and (2) give

$$\|T(\widetilde{Q}, \widetilde{\Pi}) - T(Q, \Pi_w)\| \le B_0 := \frac{h}{2} \left\{ \frac{e_0}{g(g - e_0)} + \epsilon_0 + \frac{p_0}{2} \right\}. \tag{34}$$

Indeed the plus inverse norms are at most one, while the minus inverse product costs $1/[g(g - e_0)]$.

Let S be the matrix of the prescribed score values and form the common-Euclidean columns

$$\begin{aligned} Y_w &= \operatorname{diag}(\sqrt{w})[S - 1(w^T S)], \quad Y_\nu = \operatorname{diag}(\sqrt{\nu})[S - 1(\nu^T S)], \\ D_Y &= Y_\nu - Y_w, \qquad\qquad y = \|Y_w\|_F, \qquad d_Y = \|D_Y\|_F. \end{aligned}$$

The exact tilted tensor has norm at most $h/(2g)$. Expanding the difference of the two metric matrices as

$$Y_w^T(\widetilde{T} - T)Y_w + D_Y^T \widetilde{T} Y_w + Y_w^T \widetilde{T} D_Y + D_Y^T \widetilde{T} D_Y$$

therefore proves the observed allowance

$$\|G(h, \nu) - G(h, w)\|_{\mathrm{op}} \le E_{\mathrm{scale}} := B_0 y^2 + \frac{h}{2g}(2y d_Y + d_Y^2). \tag{35}$$

Changing only the operator and keeping the old mass factors or the old center would not establish this inequality.

A.3 Inverse residuals and the complete certificate

Work now in $L^2(\nu)$ with score columns $Y = S - 1(\nu^T S)$. On its mean-zero subspace let $A_- = I - P$; let $A_+ = I + P$. Choose minus trial columns X_- centered at ν and arbitrary plus trial columns X_+. Set

$$\begin{aligned} Z_\pm &= Y - A_\pm X_\pm, \\ W_\pm &= Y^* X_\pm + X_\pm^* Y - X_\pm^* A_\pm X_\pm, \qquad G^{\mathrm{trial}} = \frac{h}{2}(W_- - W_+), \end{aligned}$$

where stars include the actual ν weights. Expanding the residual proves the classical identity

$$Y^* A_\pm^{-1} Y - W_\pm = Z_\pm^* A_\pm^{-1} Z_\pm.$$

As $A_- \ge gI$ on the complement and $A_+ \ge I$, it follows in full matrix Loewner order that

$$-\frac{h}{2} Z_+^* Z_+ \;\le\; G(h, \nu) - G^{\mathrm{trial}} \;\le \frac{h}{2g} Z_-^* Z_-.$$

In particular

$$E_{\text{solve}} := \max\left\{\frac{h}{2g}\|Z_-\|^2_{L^2(\nu),F}, \frac{h}{2}\|Z_+\|^2_{L^2(\nu),F}\right\} \geq \|G(h,\nu) - G^{\text{trial}}\|_{\text{op}}. \tag{36}$$

The full residual matrices retain off-diagonal correlations; no entrywise independence is assumed. Combining (35)–(36) gives the complete observed certificate

$$E_{\text{scale}} + E_{\text{solve}} \leq \eta \quad \Longrightarrow \quad \|G^{\text{trial}} - G(h,w)\|_{\text{op}} \leq \eta. \tag{37}$$

Its target is the metric at the original empirical law.

A.4 Actual first hit and finite termination

Proposition A.2 (Finite first-hit construction)**.** *For every finite distinct sample set and every $\eta > 0$, there is an actual scaling and inverse iteration that stops with* (37)*. In ideal real arithmetic it terminates finitely. It has a certified finite-arithmetic realization under effective convergent enclosures of inputs and gap certification that eventually detects the positive gap.*

Proof. Start $b = 1$ and perform damped symmetric scaling

$$b_i^{\text{next}} = \sqrt{\frac{b_i}{(K_w b)_i}}. \tag{38}$$

To see actual convergence, let a be the exact scale and $t = b/a$. Then $t^{\text{next}} = \sqrt{t/(P_* t)}$, where P_* is the exact balanced matrix. For a strictly positive matrix its Hilbert projective contraction gives

$$\operatorname{osc}(\log(P_* t)) \leq \theta \operatorname{osc}(\log t), \qquad \theta = \tanh(\Delta(K)/4) < 1,$$

where $\Delta(K) = \max_{i,j,k,l} \log(K_{ik}K_{jl}/(K_{il}K_{jk}))$. Consequently $\operatorname{osc}(\log t^{\text{next}}) \leq (1+\theta)\operatorname{osc}(\log t)/2$. Moreover each new log coordinate lies between $-\operatorname{osc}(\log t)/2$ and $\operatorname{osc}(\log t)/2$, since $P_* t$ lies between the minimum and maximum of t. Thus $\log(b/a) \to 0$, not merely convergence modulo constants. Classical positive-kernel contraction is also discussed in [2, 4].

At each iterate evaluate (32), a certified gap g for $\widetilde{Q}$, and (35). Stop scaling at the first iterate for which

$$q_0 \leq 1/2, \qquad e_0 < g, \qquad E_{\text{scale}} < \eta/2.$$

This hit is finite: convergence implies $r \to 1$, $\nu \to w$ and $M_0, q_0, e_0, p_0, d_Y, B_0, E_{\text{scale}} \to 0$. The exact original finite matrix has a positive gap, and so do the nearby tilted matrices. An eventually informative gap enclosure can have a lower bound at least half their positive limiting gap. These strict tests are then satisfied.

Freeze that iterate and its tilted law. Start $X_- = X_+ = 0$ and apply

$$X_-^{\text{next}} = PX_- + Y, \qquad X_+^{\text{next}} = (I-P)X_+/2 + Y/2.$$

Minus columns remain centered in exact arithmetic. On that space the first iteration contracts with nonconstant eigenvalues in $(0,1)$; the second has eigenvalues $(1-\lambda)/2 \in [0, 1/2]$. Their actual residuals tend to zero. Stop at the first inverse iteration with $E_{\text{solve}} < \eta/2$. This hit is also finite and proves (37).

For certified finite arithmetic, arbitrary real observations do not alone specify effective numerical input. Assume convergent rigorous enclosures for sites, scores and kernel values. Spectral enclosures must eventually detect a positive gap; an arbitrary valid but excessively small lower bound is not enough. Increase precision or dovetail it with iterate count. For an existence-level realization, represent each exact finite iterate as a finite real-arithmetic expression, enclose it at increasing precision, choose positive rational approximations, and freshly certify the displayed tests at those concrete vectors. A sufficiently late exact iterate satisfies all strict tests with a positive margin. Continuity ensures sufficiently close rational approximations do too. Treat the inverse trials in the same way, recentering minus columns and certifying their actual residuals. This avoids assuming convergence of an arbitrarily rounded recurrence. The dovetailing finds a valid finite hit. □

Proof of Corollary 2.2. Use Proposition A.2 at each band with $\eta = n^{-3/4}/C_K$. The combined numerical error is at most $n^{-3/4}$ pathwise on every finite distinct sample set. Multiplying by $\sqrt{n}$ gives $n^{-1/4} \to 0$; Slutsky's theorem applies. Distinctness holds almost surely under the smooth density. No statistical lower bound on the realized empirical gap was used in this numerical transfer. □

There is no practical runtime conclusion hidden in finite termination. In particular $\Delta(K)$ can grow as $1/h$, making the projective bound extremely poor. Likewise a normalization first hit does not perform an inverse solve. A raw unnormalized inverse is unsafe: for a two-point exact matrix with tangent eigenvalue $\lambda = e^{-ch}$, replacing the exact scale by $b = \lambda^{-1/2}a$ gives residual $r_i = \lambda^{-1} = 1 + O(h)$ but raw tangent eigenvalue one, so its supposed tensor inverse is singular, although $\nabla \log r = 0$ for its constant residual. The residual-tilted row normalization and the separate inverse residual test are substantive parts of the certificate.

B Certified numerical covariance at the original law

Numerical output	Required observed inputs	Error target
Metric matrix	Scaling enclosure, gap bound, inverse residuals	Original empirical-law metric; root-n transfer uses error $o(n^{-1/2})$.
Influence covariance	Scale and two gap enclosures, weighted residuals for the influence solves	Original empirical-law covariance; consistency uses error $o(1)$.

The metric certificate does not by itself control its variance: the influence in (20) contains products of inverse outputs, a further plus inverse, and a moving centering functional. We give a separate deterministic certificate for that problem. All norms below are finite weighted norms. Validated lower gaps and a validated logarithmic scale enclosure are inputs; their computation is not assigned an atom-independent cost.

B.1 Weighted resolvents and the influence

For a positive probability vector μ, write $\|v\|_{p,\mu} = (\sum_i \mu_i |v_i|^p)^{1/p}$ and $\Pi_\mu v = (\mu v)1$. Assume P is a nonnegative Markov matrix, reversible at μ, positive semidefinite in $L^2(\mu)$, and its nonconstant spectrum is contained in $[0, 1-g]$, $0 < g \le 1$. Set

$$R_- = (I - P + \Pi_\mu)^{-1}, \quad R_+ = (I + P)^{-1}, \quad C = (I - P)(I + P)^{-1} = 2R_+ - I,$$

and use the constants

$$c_+ = 1 + \frac{1}{\sqrt{3}} < \frac{8}{5}, \qquad c_-(g) = 1 + \frac{\sqrt{2}}{1 - \sqrt{1-g}} \le 1 + \frac{3}{g}. \tag{39}$$

Lemma B.1 (Weighted inverse bounds)**.** *Under these assumptions,* $\|R_+\|_{4\to4} \le c_+$, $\|R_-\|_{4\to4} \le c_-(g)$, *and* $\|R_+\|_{2\to2}, \|C\|_{2\to2} \le 1$.

Proof. Jensen's inequality and stationarity make P and Π_μ contractions on every $L^p(\mu)$. For $T = (I - 2P)/3$, $\|T^k\|_{2\to2} \le 3^{-k}$ and $\|T^k\|_{\infty\to\infty} \le 1$. Weighted finite-dimensional Riesz–Thorin interpolation gives $\|T^k\|_{4\to4} \le 3^{-k/2}$. Summing $R_+ = (2/3)\sum_{k\ge0} T^k$ gives c_+. Also $\|P^k - \Pi_\mu\|_{2\to2} \le (1-g)^k$ and its infinity norm is at most2. Interpolation and $R_- = \Pi_\mu + \sum_{k\ge0}(P^k - \Pi_\mu)$ give $c_-(g)$. The rational upper bound follows from $1 - \sqrt{1-g} \ge g/2$ and $2\sqrt{2} < 3$. The L^2 bounds follow directly from the spectrum. □

For a prescribed real score u put $\gamma_\mu = u - \mu u$. The exact scalar influence at bandwidth h can be written

$$\begin{aligned} z_- &= R_-\gamma_\mu, \quad z_+ = R_+\gamma_\mu, \\ U &= z_-Pz_- + z_+Pz_+, \\ F_\mu &= \frac{h}{2}\{\gamma_\mu(z_- - z_+) + CU\}, \qquad \psi_\mu = F_\mu - \mu F_\mu. \end{aligned} \tag{40}$$

Products of vectors here and below are componentwise. This is (20), not a frozen-normalization approximation.

Proposition B.2 (Inexact influence at one law)**.** *Let x_-, x_+, v be concrete trial vectors, with $\mu x_- = 0$. Define*

$$\begin{aligned} \rho_- &= \gamma_\mu - (I-P)x_-, \qquad \rho_+ = \gamma_\mu - (I+P)x_+, \\ \widehat{U} &= x_-Px_- + x_+Px_+, \qquad \rho_v = \widehat{U} - (I+P)v, \\ a_- &= c_-(g)\|\rho_-\|_{4,\mu}, \quad a_+ = c_+\|\rho_+\|_{4,\mu}, \\ E_U &= \sum_{\sigma\in\{-,+\}} a_\sigma(2\|x_\sigma\|_{4,\mu} + a_\sigma), \\ E_F &= \frac{h}{2}\{\|\gamma_\mu\|_\infty(a_- + a_+) + E_U + 2\|\rho_v\|_{2,\mu}\}. \end{aligned}$$

Put $\widehat{F} = (h/2)\{\gamma_\mu(x_- - x_+) + 2v - \widehat{U}\}$ and $\widehat{\psi} = \widehat{F} - \mu\widehat{F}$. Then $\|\psi_\mu - \widehat{\psi}\|_{2,\mu} \le E_F$.

Proof. Lemma B.1 bounds $\|z_\sigma - x_\sigma\|_{4,\mu}$ by a_σ. The identity $zPz - xPx = (z-x)Pz + xP(z-x)$, Hölder's inequality and Markov L^4 contraction give the stated E_U. Moreover

$$CU - (2v - \widehat{U}) = C(U - \widehat{U}) + 2(R_+\widehat{U} - v).$$

Both operators on the right have L^2 norm at most1, and $R_+\widehat{U} - v = R_+\rho_v$. This proves the bound on the uncentered influence. Centering at the same law is an orthogonal L^2 projection and has norm1. □

B.2 Transport from the residual-tilted law

Use w, b, r, R, ν, P_ν from (29), writing P_ν for the matrix denoted P there. Let a be the exact scale at w and P_w its Markov matrix. Assume validated bounds

$$\|\log(a/b)\|_\infty \le M, \qquad \mathrm{gap}(P_w) \ge g_w > 0, \quad \mathrm{gap}(P_\nu) \ge g_\nu > 0. \tag{41}$$

For example Lemma A.1 and the common-Euclidean spectral perturbation test of Section A supply such bounds when their displayed tests pass. Strictly positive entries and positive semidefiniteness give the Markov hypotheses of Lemma B.1. They do not by themselves supply a numerical positive lower gap.

Choose any nonnegative D satisfying

$$D \ge \max\{|r_{\min}e^{-2M} - 1|, |r_{\max}e^{2M} - 1|\}. \tag{42}$$

A convenient alternative upper bound is $r_{\max}(e^{2M} - 1) + \max_i |r_i - 1|$. Introduce only observed relative-law quantities

$$\begin{aligned} a_i^0 &= w_i/\nu_i, \quad A = \max_i a_i^0, \quad \tau_\infty = \max_i |a_i^0 - 1|, \quad t = \sum_i |w_i - \nu_i|, \\ \chi_p &= (r_{\max}/r_{\min})^{1/p}, \quad W_2 = (\sum_i w_i^2/\nu_i)^{1/2}, \quad \tau_2 = (\sum_i \nu_i(a_i^0 - 1)^2)^{1/2}. \end{aligned}$$

For $U_0 = \|u\|_\infty$ and $G_0 = 2U_0$ define

$$\begin{aligned}
Z_- &= c_-(g_\nu)G_0, \quad Z_+ = c_+G_0,\\
d_- &= \chi_4 c_-(g_w)\{tU_0 + (D+\tau_\infty)Z_-\},\\
d_+ &= \chi_4 c_+\{tU_0 + DZ_+\},\\
U_\nu^0 &= Z_-^2 + Z_+^2,\\
d_U &= \sum_{\sigma\in\{-,+\}} \{(1+D)d_\sigma(2Z_\sigma + d_\sigma) + DZ_\sigma^2\},\\
d_F &= \frac{h}{2}\{tU_0(Z_- + Z_+ + d_- + d_+) + G_0(d_- + d_+) + \chi_2 d_U + 2\chi_2 DU_\nu^0\},\\
B_F &= \frac{h}{2}\{G_0(Z_- + Z_+) + U_\nu^0\}, \qquad L_u = (1+W_2)d_F + \tau_2 B_F.
\end{aligned} \tag{43}$$

Proposition B.3 (Original-law influence enclosure)**.** *With* (41)–(43)*,* $\|\psi_w - \psi_\nu\|_{2,\nu} \le L_u$*.*

Proof. Write $y_i = \log(a_i/b_i)$. Exact normalization gives

$$(P_w)_{ij}/(P_\nu)_{ij} = r_i e^{y_i + y_j}.$$

Thus $|P_w - P_\nu| \le DP_\nu$ entrywise, and for every $1 \le p \le \infty$, $\|(P_w - P_\nu)f\|_{p,\nu} \le D\|f\|_{p,\nu}$. In particular $\|P_w\|_{p,\nu} \le 1 + D$. Changing between the two measures costs at most χ_p in an operator norm: the product of the two conversion constants is $(\max(w/\nu)/\min(w/\nu))^{1/p} = \chi_p$. Consequently the inverse norms at w, measured on $L^4(\nu)$, are bounded by $\chi_4 c_-(g_w)$ and $\chi_4 c_+$. Also $\|\Pi_w - \Pi_\nu\|_{4,\nu} \le \tau_\infty$.

The two inverse equations and their resolvent identities now give $\|z_{\sigma,w} - z_{\sigma,\nu}\|_{4,\nu} \le d_\sigma$: the changed right side has norm at most tU_0, and the changed operators contribute respectively $(D+\tau_\infty)Z_-$ and DZ_+. Expanding each quadratic product gives $\|U_w - U_\nu\|_{2,\nu} \le d_U$. For example its three terms are bounded by $(1+D)d_\sigma(2Z_\sigma + d_\sigma)$ and DZ_σ^2. The Cayley operators satisfy

$$\|C_w\|_{2,\nu} \le \chi_2, \qquad \|C_w - C_\nu\|_{2,\nu} \le 2\chi_2 D,$$

since $C = 2R_+ - I$ and $R_{+,w} - R_{+,\nu} = R_{+,w}(P_\nu - P_w)R_{+,\nu}$. Expansion of (40) proves $\|F_w - F_\nu\|_{2,\nu} \le d_F$ and $\|F_\nu\|_{2,\nu} \le B_F$. Finally

$$\psi_w - \psi_\nu = (F_w - F_\nu) - w(F_w - F_\nu) - (w - \nu)F_\nu.$$

Weighted Cauchy–Schwarz bounds the last two terms by $W_2 d_F$ and $\tau_2 B_F$. This is a change of centering law; the same-law orthogonal-projection argument cannot be used here. □

B.3 Joint covariance and numerical studentization

Fix p score functions. The diagonal metric entries use the scalar influence directly. For an off-diagonal entry use $\psi_{ab} = (\psi_{u_a+u_b} - \psi_{u_a-u_b})/4$. Apply Propositions B.2 and B.3 to these finitely many scalar scores. For each vectorized matrix entry j, let E_j and L_j be the corresponding error allowances, with the sum of the two scalar allowances divided by4 for off-diagonal entries. Let $m = p(p+1)/2$ and use this fixed ordering, with no additional $\sqrt{2}$ scaling. Define

$$E = (\sum_{j=1}^m E_j^2)^{1/2}, \quad L = (\sum_{j=1}^m L_j^2)^{1/2}, \quad \widehat{S} = (\nu\|\widehat{\psi}\|^2)^{1/2}, \quad B_\nu = \widehat{S} + E, \quad \widehat{\Sigma} = \nu[\widehat{\psi}\widehat{\psi}^T].$$

Theorem B.4 (Simultaneous covariance certificate)**.** *For the exact original-law influence covariance* $\Sigma_w = w[\psi_w\psi_w^T]$*, put*

$$C_\Sigma = 2\widehat{S}E + E^2 + A(2B_\nu L + L^2) + \tau_\infty B_\nu^2. \tag{44}$$

Then

$$\|\widehat{\Sigma} - \Sigma_w\|_{\mathrm{op}} \le C_\Sigma, \qquad \widehat{\Sigma} - C_\Sigma I \le \Sigma_w \le \widehat{\Sigma} + C_\Sigma I.$$

This includes singular covariances. For every vector v *simultaneously,* $v^T\Sigma_w v$ *lies within* $C_\Sigma\|v\|^2$ *of* $v^T\widehat{\Sigma}v$*.*

Proof. For vector-valued f, e, Cauchy–Schwarz gives $\|\nu[(f+e)(f+e)^T - ff^T]\|_{\mathrm{op}} \le 2\|f\|_{2,\nu}\|e\|_{2,\nu} + \|e\|_{2,\nu}^2$. Use this first with the trial and exact tilted-law influence, giving $2\widehat{S}E + E^2$ and $\|\psi_\nu\|_{2,\nu} \le B_\nu$. Next change the influence from ψ_ν to ψ_w under w; $w/\nu \le A$ bounds this change by $A(2B_\nu L + L^2)$. The remaining change in averaging law, with ψ_ν fixed, is at most $\tau_\infty B_\nu^2$. Sum these three bounds. □

Corollary B.5 (Finite covariance hit and asymptotic transfer)**.** *For each fixed finite distinct sample set and $\eta > 0$, the effective input-enclosure and eventually informative gap-enclosure conditions of Proposition A.2 allow a finite computation with $C_\Sigma < \eta$. In Theorem 2.1, compute the covariance at the single bandwidth h_n used in Proposition 6.1 and choose any deterministic $\eta_n \downarrow 0$. Its certified approximation is consistent for Σ. Combining it with Corollary 2.2 gives the same studentized limit for each fixed contrast with positive limiting variance.*

Proof. Damped scaling converges to the exact finite scale by Proposition A.2. The observed logarithmic enclosure can tend to zero; relative laws tend to1, so $D, t, \tau_\infty, \tau_2, L \to 0$. The two finite gaps remain bounded below near the fixed exact matrix. Solve the two linear equations and the further plus equation with residuals tending to zero, recentering the minus trial. In finite dimension the displayed weighted norms then tend to zero and $E \to 0$, while the exact influence norms stay bounded. Thus $C_\Sigma \to 0$. Strict stopping tests with convergent rigorous input enclosures have a finite hit by the same precision-dovetailing argument as in Proposition A.2. Covariance consistency and Slutsky's theorem give the statistical assertion. □

No explicit minimum atom weight occurs in (39)–(44). The scale and gap certification inputs can nevertheless be badly conditioned; for example Lemma A.1 itself uses a minimum weight.

Simultaneous deterministic contrast bounds do not justify Gaussian inference after choosing a contrast from the same data.

The accompanying standard-library program `check_covariance.py` corroborates the algebra on six exact rational problems with rare weights down to 10^{-12}, perturbed scalings and three inexact solves. Three joint covariances are exactly singular. All six certify a positive fixed contrast. The negative control computes the strictly positive excess of an uncentered second moment over the covariance and verifies that this excess is exactly the square of the omitted mean. These are finite checks of the certificate, not premises of the analytic theorems or a heat-kernel runtime experiment.

Acknowledgments and disclosures

ChatGPT Sol 6.1 assisted with mathematical research, proof development, literature discovery, drafting and verification through workflows developed in the author's AI governance research. The author is responsible for all statements, proofs, attribution and presentation.

The research was self-funded and received no external grant or sponsorship. The author is the founder of Luminary Oddities, which operates the Research Archive hosting the preprint. No other competing interests are declared. The finite rational fixtures and standard-library verification code are included in the accompanying source package. No empirical data from human participants were collected.

References

[1] H. Lavenant, J. Luckhardt, G. Mordant, B. Schmitzer and L. Tamanini, *The Riemannian geometry of Sinkhorn divergences*, Ann. Inst. H. Poincaré Anal. Non Linéaire 43 (2026), 1079–1144. doi:10.4171/AIHPC/165; preprint version arXiv:2405.04987v2.

[2] C. L. Wormell and S. Reich, *Spectral convergence of diffusion maps: improved error bounds and an alternative normalisation*, SIAM J. Numer. Anal. 59(3) (2021), 1687–1734. doi:10.1137/20M1344093; preprint version arXiv:2006.02037v3.

[3] X. Cheng and B. Landa, *Bi-stochastically normalized graph Laplacian: convergence to manifold Laplacian and robustness to outlier noise*, Information and Inference 13(4) (2024), iaae026. doi:10.1093/imaiai/iaae026; preprint version arXiv:2206.11386v3.

[4] N. Kessler, R. Magnet and J. Feydy, *Sinkhorn Normalization of Diffusion Kernels*, Proceedings of Machine Learning Research 306 (2026), 56764–56795; arXiv:2507.06161v2.

[5] Z. Goldfeld, K. Kato, G. Rioux and R. Sadhu, *Limit Theorems for Entropic Optimal Transport Maps and the Sinkhorn Divergence*, Electronic Journal of Statistics 18 (2024), 980–1041; doi:10.1214/24-EJS2217; arXiv:2207.08683.

[6] C. Li, N. García Trillos, H. Li and L. Suchan, *Central limit theorems for the eigenvalues of graph Laplacians on data clouds*, arXiv:2507.18803v1, 2025.

[7] T. Manole, S. Balakrishnan, J. Niles-Weed and L. Wasserman, *Plugin Estimation of Smooth Optimal Transport Maps*, Ann. Statist. 52(3) (2024), 966–998. doi:10.1214/24-AOS2379; preprint version arXiv:2107.12364v3.

[8] H. Kono, *Central Limit Theorems for Optimal Transport Functionals*, arXiv:2610.27857v1, 2026.

[9] G. Mordant and L. Xu, *Convergence of the Sinkhorn Riemannian metric for finitely supported measures*, arXiv:2609.14651v1, 2026.

[10] T. L. Saidi, F. Gunsilius and G. Mena, *Second-order Sinkhorn Geometry and $\sqrt{n}$-Inference*, arXiv:2610.36488v1, 2026.

[11] P. Navas,*The information geometry of the Sinkhorn divergence on a Hilbert manifold*,arXiv:2610.39850v1, 2026.

[12] S. Schubr, *Sharp convergence of heat-kernel Sinkhorn metrics*, Luminary Research Archive, edition 1, 2026. `https://research.luminaryoddities.com/sharp-convergence-of-heat-kernel-sinkhorn-metrics/v/1`.

[13] S. Schubr, *Leading bias and positive correction of heat-kernel Sinkhorn metrics*, Luminary Research Archive, edition 1, 2026. `https://research.luminaryoddities.com/leading-bias-positive-correction-heat-kernel-sinkhorn-metrics/v/1`.

[14] J.-M. Dufour and P. Valéry, *Rank-robust Wald-type tests: a regularization approach*, working paper, 13 February 2016. `https://econ.la.psu.edu/wp-content/uploads/sites/5/2023/05/Dufour_Valery_Regularized_Wald_Tests_Short_2016_01_19.pdf`.

[15] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998.