First-Hit Stopping and Bias-Corrected Score Inference in Dimension Three

Scott Schubr
Luminary Oddities

October 20, 2026

Abstract

We study the actual first residual crossing of the symmetric half-Sinkhorn iteration for observations from a fixed positive C^8 density on the three-dimensional torus. At two heat scales, computed from the same sample, a linear combination removes the first-order population score bias. For $\varepsilon = n^{-\kappa}$ with $2/17 < \kappa < 4/9$ and every fixed $0 < \gamma < 3/4$, stopping each run at its first projective mass residual below $n^{-1/2}h^{-\gamma}$ preserves the corrected estimator's fixed-query Gaussian limit. The stopping index is asymptotic to $(3/4 - \gamma)\log(1/h)/\log 2$, and the attained residual lies between c times the threshold and the threshold with probability tending to one for every fixed $0 < c < 1/2$. The covariance retains the dependence between the two scales. An explicit cubic population coefficient gives a fixed analytic density for which the limit is shifted at $\kappa = 2/17$ and the normalized bias diverges below it, also under actual stopping. The formal Gaussian range extends to $\kappa < 4/9$; consistency and shrinking uncertainty require $\kappa < 2/5$. A sharp heat-spectral trace determines logarithmic corrections to the stopping count and phase-dependent residual limits, proving that the constants $1/2$ and 1 in the residual bracket are sharp.

Keywords: entropic self-transport; symmetric Sinkhorn; score estimation; first-hit stopping; bias correction.

1 Introduction

Numerical kernel normalization is terminated using an observable residual, while score inference concerns a spatial derivative of the normalized off-sample field. The relevant mathematical question is whether a residual-based stopping decision preserves the limiting distribution of the fully converged estimator. It is particularly delicate when the sample, kernel scale, iterates, residual and queried field are all dependent.

We establish such a result in dimension three for a two-scale corrected self-transport score. The stopping rule checks every iterate from the constant initialization. The threshold exceeds $n^{-1/2}$ by a diverging factor, yet the stopped field has the same limiting law as the exact empirical solution. A complete second-order trajectory expansion removes the tolerance-dependent upper restriction and preserves equivalence after any deterministic diverging iteration index within the original budget. We determine the stopping count up to integer rounding, identify its residual phase law, and exhibit a density for which the lower bandwidth endpoint cannot be included.

This paper is a follow-up to the one- and two-dimensional residual-stopping theorem of [1]. That theorem concerns an uncorrected score and a different sufficient bandwidth regime. Substitution of dimension three into its two bandwidth conditions leaves no nonempty power window. Here the cubic remainder after population bias cancellation, the joint two-scale covariance and refined empirical-trajectory estimates supply a nonempty dimension-three result. The preceding residual-certificate work [2] and the lower-dimensional first-hit principle are antecedents; the iteration and the general idea of statistical stopping are not new.

The population expansion identifies the bias remaining after two-scale correction. The trajectory estimates determine when numerical score error is negligible on the same sample. The heat-spectral clock then locates the actual first crossing, including logarithmic corrections and its integer phase. For $\varepsilon = n^{-\kappa}$, inference uses $2/17 < \kappa < 4/9$; consistency and shrinking uncertainty additionally require $\kappa < 2/5$. The sharp-clock theorem states its own bandwidth hypotheses.

1.1 Related work and relation to the preceding paper

The actual first-hit question was studied in dimensions one and two in [1]. That paper proves preservation of pointwise score inference, the leading stopping count, the attained-residual bracket and a coefficient-one tolerance obstruction for fixed positive smooth periodic densities. Its sampling and bias conditions leave no bandwidth window in dimension three. The present paper extends that inferential trajectory analysis to dimension three using a two-band corrected score. The earlier residual certificates for arbitrary candidate scales [2] are a separate prerequisite perspective; a residual certificate alone does not determine the law at its first attainment along a sampled nonlinear iteration.

Pointwise self-transport score central limit theorems with decreasing regularization are already available. In particular, Mordant's Theorem 12 [3] gives an exact, uncorrected score limit under its Euclidean-domain hypotheses when $n^{-2/(d+2)} \ll \varepsilon \ll n^{-2/(d+6)}$. In dimension three this corresponds to $2/9 < \kappa < 2/5$ for $\varepsilon = n^{-\kappa}$. The present result is consequently not a first dimension-three score limit theorem or a first exact-score result beyond $\kappa = 1/3$. Its target is the corrected periodic estimator at two dependent, observable first-hit stopping times. Richardson cancellation of entropic bias is an established idea: Chizat et al. [10] apply it to Sinkhorn-divergence cost estimation. Here a differentiated cubic population expansion is combined with the joint fluctuation law of two empirical self-scales.

Higher-order self-transport expansions also have close precedents. Yao, Kim and Schiebinger [14], Corollary 1.9, give arbitrary-order deterministic expansions of entropic Brenier maps in L^2 for positive C^∞ densities on the flat torus. Goldfeld, Kato, Rioux and Sadhu [13] develop first- and second-order Hadamard differentiability of entropic potentials and statistical limit theory at fixed regularization. These results establish the expansion and functional-differentiation perspectives. Our use of a pointwise differentiated population expansion under C^8 regularity, and of local quadratic and cubic empirical terms with explicit repeated- observation contributions, addresses a different joint limiting regime.

Kokot [4], Chapter 3, develops Gaussian-RKHS scaling differentiability, the plus-resolvent at the self diagonal and fixed-regularization Hilbert limit theory. Chapter 4 gives a general first-order population score expansion and matching after two symmetric updates; its Gaussian example has a stronger explicit remainder. These are important precedents for the analytic and statistical linearizations used here. The distinction is the joint bandwidth–sample–iteration control required by the corrected empirical first-hit law.

Differentiating a convergent numerical iteration requires additional control. Pauwels and Vaiter [15] prove locally uniform linear convergence of derivatives of finite-dimensional Sinkhorn–Knopp transport plans. Li, Lu, Tao and Ye [16] analyze first- and second-order derivatives of finite-data regularized transport losses, including conditioning under early termination. Scieur et al. [17] study convergence rates of unrolled differentiation for quadratic optimization. More directly on the statistical side, Liu, Mehta, Pal and Harchaoui [18], Appendix D, recursively decompose same-sample balancing errors into operator compositions and higher-order terms. Thus neither differentiated Sinkhorn iterations nor recursive statistical expansions are introduced here. The trajectory argument needed below controls the actual initialization, random normalization weights, derivative features and nonlinear empirical products uniformly through the original iteration budget as the heat bandwidth decreases.

Spatial derivatives of Sinkhorn iterates also have direct convergence theorems. Greco, Noble, Conforti and Oliviero-Durmus [22] prove uniform exponential convergence of potentials and their spatial gradients for positive smooth densities on the torus. Greco and Tamanini [23] obtain gradient and Hessian stability and convergence rates under semiconcavity and transport-inequality assumptions. Belomestny [24] gives uniform bounds for exact empirical potentials under entropy and residual-stability hypotheses. Gao, Qu, Ye and Udell [25] establish nonasymptotic local convergence rates for finite-dimensional alternating

scaling under scalability and connectivity assumptions. These results provide analytic, statistical and algorithmic precedents. The additional requirement here is control of a same-sample spatial derivative as bandwidth decreases, uniformly enough to preserve inference at the realized residual crossing.

Wormell and Reich [5] analyze the same periodic Gaussian bistochastic normalization, including smooth weights, empirical scale stability, resolvents and improved operator bias. They also give an accelerated symmetric algorithm with its own stopping guarantee. Cheng and Landa [6] show that approximate normalization at statistical precision suffices for graph-Laplacian consistency. Maeda, Yao and Nitanda [7] combine sampling and iteration errors for two-sample Schrödinger bridges and discuss balancing those errors to select an iteration count. These results establish substantial precedent for statistical termination. Their conclusions differ from preserving a local score fluctuation law at the realized residual hit.

Residual-based statistical stopping and transfer between residual and estimation errors are established in spectral regularization. Blanchard, Hoffmann and Reiss [19] prove oracle adaptation for residual stopping in statistical inverse problems; Celisse and Wahl [20] analyze discrepancy rules for kernelized spectral filters using concentration and changes of norm. Hucker and Reiss [21] explicitly treat first residual crossings, continuous interpolation and the error caused by integer rounding for conjugate gradients. Their conclusions concern different inverse and regression models. The sharper stopping result here uses the specific heat-filter trace to obtain its logarithmic corrections and residual phase law. Laplace asymptotics, inversion of a slowly varying prefactor and ceiling overshoot are standard steps; the model-specific assertion is their validity for this unchanged sampled self-transport trajectory, including sharp attained-residual constants along integer sample sizes.

The symmetric square-root update is also established, appearing in Feydy et al. [9] and in the normalization algorithm of Kessler, Magnet and Feydy [8]. Matrix and operator concentration are standard tools [11, 12]. The contribution here is the combined dimension-three theorem, including the coupled corrected-score limit, genuine first-hit count and residual phase law, the full tolerance–bandwidth rectangle, and the explicit correction-specific necessary lower endpoint. In particular, the useful consistent range $2/17 < \kappa < 2/5$ holds for every fixed $0 < \gamma < 3/4$; numerical equivalence also holds after any deterministic diverging iteration index within the original budget. The upper $4/9$ endpoint remains a sufficient boundary of the present proof, with neither equality nor necessity established.

Related heat-kernel metric results concern different observables. The cubic population coefficient used here also appears in the analysis of metric bias and positive corrections [26]. That work studies population inverse tensors. Empirical metric inference [27] concerns integrated inverse weighted-Laplacian energies for prescribed smooth tangent scores, with its own numerical certificates. The present target is the spatial density-gradient score at two residual-selected stopping times.

2 Model, stopping rule and main results

Write $\mathbb{T}^3 = \mathbb{R}^3/\mathbb{Z}^3$ with unit Lebesgue measure. Let $f \in C^8(\mathbb{T}^3)$ be a fixed probability density with $\inf f > 0$, and let $X_1, \dots, X_n$ be independent with law $\mu = f\,\mathrm{d}x$. All constants may depend on this fixed density and its norms. Set $\mu_n = n^{-1}\sum_i \delta_{X_i}$. For $h > 0$, let k_h be the integral-one periodic heat kernel

$$k_h(z) = \sum_{k\in\mathbb{Z}^3} e^{-2\pi^2 h|k|^2} e^{2\pi i k\cdot z}, \qquad H_h = \exp(h\Delta/2),$$

and define

$$K_{h,n}b(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - X_i)b(X_i).$$

The population scale a_h and empirical scale $\widehat{a}_h$ solve $a_h K_{h,\mu} a_h = 1$ and $\widehat{a}_h(X_i)(K_{h,n}\widehat{a}_h)(X_i) = 1$ respectively. Their off-sample extensions are defined by these convolution equations. The corresponding scores are

$$s_h^{\mathrm{pop}}(x) = -2\,\nabla \log a_h(x), \qquad \widehat{s}_h(x) = 2\,\nabla \log(K_{h,n}\widehat{a}_h)(x).$$

Throughout, x is a fixed deterministic spatial query.

Definition 2.1 (Genuine first-hit rule)**.** At each $h \in \{\varepsilon, 2\varepsilon\}$ start from $b_{0,h}(X_i) = 1$ and update

$$b_{j+1,h}(X_i) = \sqrt{\frac{b_{j,h}(X_i)}{(K_{h,n} b_{j,h})(X_i)}}, \qquad s_{j,h}(x) = 2\,\nabla_x \log(K_{h,n} b_{j,h})(x).$$

Set

$$F_{j,h}(X_i) = \sqrt{b_{j,h}(X_i)(K_{h,n} b_{j,h})(X_i)}, \quad m_{j,h} = \int F_{j,h}\,\mathrm{d}\mu_n, \quad r_{j,h} = \left\| F_{j,h}/m_{j,h} - 1 \right\|_{L^2(\mu_n)}.$$

For fixed $\gamma > 0$, let $t_{n,h} = n^{-1/2} h^{-\gamma}$ and $L_n = \max(3, \lceil 4 \log n \rceil)$, with natural logarithm. Take the first index $j \in \{0, \ldots, L_n\}$ satisfying $r_{j,h} \le t_{n,h}$; if there is no such index return L_n. Denote this index by τ_h. Every residual is checked, including $j = 0$. The two runs use the same observations and separate stopping times.

Define the positive constant

$$J_R = \int_{\mathbb{R}^3} (2\pi\xi_1)^2 [2h_1(\xi) - h_2(\xi)]^2\,\mathrm{d}\xi, \qquad h_c(\xi) = \frac{e^{-2\pi^2 c|\xi|^2}}{1 + e^{-2\pi^2 c|\xi|^2}}. \tag{1}$$

Theorem 2.2 (Corrected score at the actual first hit)**.** *Suppose*

$$\varepsilon = n^{-\kappa}, \qquad \frac{2}{17} < \kappa < \frac{4}{9}, \qquad 0 < \gamma < \frac{3}{4}$$

with fixed exponents. At each $h = \varepsilon, 2\varepsilon$, a genuine hit occurs before L_n with probability tending to one, and

$$\frac{\tau_h}{\log(1/h)} \longrightarrow \frac{3/4 - \gamma}{\log 2} \quad \textit{in probability}, \tag{2}$$

$$\mathbb{P}\{c\, t_{n,h} \le r_{\tau_h,h} \le t_{n,h}\} \longrightarrow 1 \qquad (0 < c < 1/2), \tag{3}$$

$$\sqrt{n}\, h^{5/4} \left\| s_{\tau_h,h}(x) - \widehat{s}_h(x) \right\| \longrightarrow 0 \quad \textit{in probability}. \tag{4}$$

In fact, for every deterministic $J_n \to \infty$ with $J_n \le L_n$, the expression in (4)*, with τ_h replaced by j, tends to zero uniformly over $J_n \le j \le L_n$. For the corrected score,*

$$\sqrt{n}\, \varepsilon^{5/4} \{2 s_{\tau_\varepsilon,\varepsilon}(x) - s_{\tau_{2\varepsilon},2\varepsilon}(x) - \nabla \log f(x)\} \;\Rightarrow\; N_3\left(0, \frac{4J_R}{f(x)} I_3\right). \tag{5}$$

The same limit holds with both stopped scores replaced by exact scores.

Corollary 2.3 (Consistency and interpretation)**.** *In the full rectangle*

$$\frac{2}{17} < \kappa < \frac{2}{5}, \qquad 0 < \gamma < \frac{3}{4},$$

the corrected stopped score is consistent and its Gaussian uncertainty scale shrinks to zero. At $\kappa = 2/5$ the normalization in (5) *is one, so the raw error has a nondegenerate Gaussian limit. For $2/5 < \kappa < 4/9$ the normalized limit remains valid, but the raw error leaves every fixed bounded ball in probability. Thus the formal range beyond $2/5$ is not an improvement in estimation precision.*

Proof. The normalization is $n^{1/2-5\kappa/4}$. Its divergence proves the first assertion by tightness in (5). At equality use (5) directly. Above equality write $Y_n = A_n E_n \Rightarrow Z$ with $A_n \to 0$ and nondegenerate Gaussian Z. For every fixed M and $\eta > 0$, $\mathbb{P}(\|E_n\| \le M) \le \mathbb{P}(\|Y_n\| \le \eta)$ eventually. Portmanteau followed by $\eta \downarrow 0$ proves the last assertion. □

Theorem 2.4 (Sharp first-hit location and residual phase)**.** *The rule in Definition 2.1 has the following additional geometric properties for* $2/19 < \kappa < 2/3$ *and* $0 < \gamma < 3/4$*. Put* $u = 3/4 - \gamma$*,* $L_h = \log(1/h)$ *and* $K = 1/(4\,2^{1/4}\pi)$*, and let* $t_h > 2$ *be the unique large solution of*

$$Kh^{-u}2^{-t_h}t_h^{-1}[\log(2t_h)]^{1/4} = 1.$$

Then

$$t_h = \frac{uL_h - \log L_h + \frac{1}{4}\log\log L_h + \log(K\log 2/u)}{\log 2} + o(1), \tag{6}$$

$$\mathbb{P}\{\lceil t_h - \eta\rceil \le \tau_h \le \lceil t_h + \eta\rceil\} \longrightarrow 1 \quad \textit{for every fixed } \eta > 0, \tag{7}$$

$$\frac{r_{\tau_h,h}}{t_{n,h}} - 2^{t_h - \tau_h} \longrightarrow 0 \quad \textit{in probability.} \tag{8}$$

For each deterministic subsequence with $\{t_h\} \to \theta \in (0,1)$*,* $\mathbb{P}\{\tau_h = \lceil t_h\rceil\} \to 1$ *and* $r_{\tau_h,h}/t_{n,h} \to 2^{\theta-1}$ *in probability. Every interior phase is attained along an integer-sample-size subsequence. Consequently no universal fixed lower constant greater than* $1/2$*, or upper constant less than* 1*, can replace the residual bracket constants. The ratio has no full-sequence probability limit.*

The geometric domain in this theorem does not extend the score CLT beyond Theorem 2.2. The proof is in Section 5.

The uniformity just stated is in iteration at a fixed spatial query. The two stopping times and scores remain statistically dependent; this dependence is essential to (1).

Theorem 2.5 (Necessary lower bandwidth condition)**.** *On* $\mathbb{T}^3$*, fix*

$$f(x) = \frac{32}{33}\left[1 + \frac{1}{4}\cos(2\pi x_1)\right]^2, \qquad x_* = (1/4, 0, 0), \qquad b_* = \big((2\pi)^7/256, 0, 0\big).$$

For any fixed $0 < \gamma < 3/4$*, use the unchanged rule in Definition 2.1. At* $\kappa = 2/17$*, the statistic on the left of* (5)*, evaluated at* x_**, converges to* $N_3(b_*, [4J_R/f(x_*)]I_3)$*. For every fixed* $0 < \kappa < 2/17$*, its first coordinate diverges to* $+\infty$ *in probability. Thus the strict lower condition is necessary for the universal zero-centered conclusion for this specified two-scale estimator and rule.*

Proposition 2.6 (Coefficient-one tolerance obstruction)**.** *For* $f = 1$ *and the same bandwidth range as Theorem 2.2,* $\gamma \ge 3/4$ *makes both runs stop at index zero with probability tending to one. Their corrected score covariance is different from* (5)*.*

This obstruction is for the stated threshold coefficient and rule.

3 Population expansion and second-order empirical inference

We give the analytic and statistical estimates used below, with their dependence on the shrinking heat time explicit. Constants may depend on the fixed density and on finitely many of its norms. All spatial integrals in this section are over the unit torus unless another domain is displayed. Write $\mu = f\,dx$, $r = \sqrt{f}$, and

$$H_h = \exp(h\Delta/2), \qquad \lambda_{h,k} = \exp(-2\pi^2 h|k|^2), \quad k \in \mathbb{Z}^3.$$

Thus H_h has integral-one kernel k_h. The population symmetric scale is denoted by a_h, so that

$$a_h H_h(a_h f) = 1, \qquad q_h = a_h r, \qquad q_h H_h(q_h r) = r. \tag{9}$$

Its reciprocal is $v_h = 1/a_h = H_h(a_h f)$, and its score is $s_h = 2\nabla \log v_h = -2\nabla \log a_h$.

3.1 Heat estimates and a uniform inverse

Lemma 3.1 (Heat and feature estimates)**.** *For $0 < h \le 1$, the periodic Gaussian satisfies*

$$k_h(0) \le Ch^{-3/2}, \qquad |\nabla k_h(z)| \le Ch^{-1/2} k_{2h}(z), \tag{10}$$

$$\|k_h(x-\cdot)\|_{L^2(dx)} \le Ch^{-3/4}, \qquad \|\nabla_x k_h(x-\cdot)\|_{L^2(dx)} \le Ch^{-5/4}. \tag{11}$$

Let $\mathcal{H}_h$ be the real reproducing-kernel Hilbert space of k_h, and write $k_{h,x} = k_h(x-\cdot)$. Then

$$\|k_{h,x}\|_{\mathcal{H}_h} \le Ch^{-3/4}, \qquad \|\partial_{x_j} k_{h,x}\|_{\mathcal{H}_h} \le Ch^{-5/4}. \tag{12}$$

For each fixed nonnegative integer s, $(I + H_h)^{-1}$ is bounded on $C^s(\mathbb{T}^3)$ uniformly in h.

Proof. Periodizing the Gaussian on $\mathbb{R}^3$ gives the first estimate. On each Euclidean lift we have

$$|z| \exp(-|z|^2/(2h)) \le C\sqrt{h} \exp(-|z|^2/(4h)).$$

Summing the lifts proves (10). The squared row norms in (11) are respectively $\sum_k \lambda_{h,k}^2$ and $\sum_k (2\pi k_j)^2 \lambda_{h,k}^2$. Comparing these convergent Gaussian sums with integrals gives the asserted powers. The reproducing property gives $\|k_{h,x}\|^2 = k_h(0)$ and $\|\partial_{x_j} k_{h,x}\|^2 = \sum_k (2\pi k_j)^2 \lambda_{h,k}$, which proves (12).

For the inverse assertion, the function

$$\chi(\xi) = \frac{e^{-|\xi|^2/2}}{1 + e^{-|\xi|^2/2}}$$

is Schwartz. Its inverse Fourier transform L is integrable. The multiplier of $(I + H_h)^{-1}$ on the torus is $1 - \chi(2\pi\sqrt{h}\, k)$. Hence this inverse is the identity minus convolution with the periodization of $h^{-3/2} L(\cdot/\sqrt{h})$. The latter kernel has uniformly bounded L^1 norm. Convolution commutes with derivatives, so the same bound holds on every fixed C^s space. The identities initially hold for trigonometric polynomials and extend by their density in C^s. No operator-norm convergence of H_h to I on C^s is being assumed. □

Lemma 3.2 (Population scale and cubic logarithmic coefficient)**.** *For positive $r \in C^8(\mathbb{T}^3)$ and all sufficiently small $h > 0$, equation* (9) *has a unique positive continuous solution, and*

$$\|q_h - 1\|_{C^s} = O(h), \qquad 0 \le s \le 6. \tag{13}$$

Already for positive $r \in C^7$, define

$$u = -\frac{\Delta r}{4r}, \qquad c = \frac{\Delta^3 r}{192 r} - \frac{\Delta[(\Delta r)^2/r]}{128 r} + \frac{(\Delta r)^3}{384 r^3}. \tag{14}$$

Then

$$\log q_h = hu + h^3 c + o_{C^1}(h^3). \tag{15}$$

In particular,

$$2s_h - s_{2h} = \nabla \log f + 12h^3 \nabla c + o_{C^0}(h^3). \tag{16}$$

Proof. We first establish a uniform inverse for the nonlinear equation. On $X = C^s$, put $F_h(q) = qH_h(qr) - r$. Its derivative at 1 is

$$L_h z = (H_h r) z + H_h(rz) = M_{H_h r}(I + H_h + E_h) z,$$

where M_b means multiplication by b and

$$E_h = M_{r/(H_h r) - 1} H_h + M_{1/(H_h r)} [H_h, M_r].$$

For $r \in C^{s+1}$ the convolution formula and the product rule give

$$\|[H_h, M_r] z\|_{C^s} \le C_s \sqrt{h}\, \|r\|_{C^{s+1}} \|z\|_{C^s}. \tag{17}$$

Indeed, before differentiating, the integrand is $k_h(y)[r(x-y)-r(x)]z(x-y)$. Each derivative produces the same difference with a derivative of r, times a derivative of z; the Gaussian first absolute moment is $O(\sqrt{h})$. Also

$$H_h r - r = \frac{1}{2}\int_0^h H_t \Delta r \, dt = O_{C^s}(h)$$

when $r \in C^{s+2}$. Positivity of r keeps $H_h r$ bounded away from zero. The preceding inverse lemma and a Neumann perturbation therefore give $\|L_h^{-1}\|_{X\to X} \le C$ for small h.

The residual $F_h(1)$ is $O_X(h)$, and the exact nonlinear remainder at any base point p is

$$F_h(p+z) - F_h(p) - DF_h(p)z = zH_h(rz). \tag{18}$$

Since C^s is a Banach algebra, a contraction on a ball of radius Ch around 1 gives a positive solution with (13), whenever $r \in C^{s+2}$. Constructing first in C^6 proves all the stated bounds for C^8 densities.

For uniqueness, write $w_i = q_i r$ for two positive continuous solutions. They satisfy $w_i H_h w_i = r^2$. With $d = w_1 - w_2$,

$$H_h d = -\frac{r^2 d}{w_1 w_2}.$$

Taking the $L^2(dx)$ inner product with d gives a nonnegative left side, since the Fourier eigenvalues of H_h are positive, and a nonpositive right side. Strict positivity of $r^2/(w_1 w_2)$ forces $d = 0$. Thus all the locally constructed solutions agree.

To calculate the cubic term, set $A = \Delta/2$ and use

$$p_h = 1 + hu + \tfrac{1}{2}h^2u^2 + h^3(c + u^3/6).$$

The identity $ru = -Ar/2$ and semigroup Taylor expansion yield the following expansions in C^1:

$$\begin{aligned}
H_h r &= r + hAr + \tfrac{1}{2}h^2A^2r + \tfrac{1}{6}h^3A^3r + o(h^3),\\
H_h(ru) &= ru + hA(ru) + \tfrac{1}{2}h^2A^2(ru) + o(h^2),\\
H_h(ru^2/2) &= ru^2/2 + hA(ru^2)/2 + o(h),\\
H_h[r(c+u^3/6)] &= r(c+u^3/6) + o(1).
\end{aligned}$$

These are Peano expansions, not formal differentiation in negative heat time. For example, the first follows by applying strong continuity on C^1 to $A^3 r$ in the integral Taylor remainder. The needed regularity is, respectively, $r \in C^7$, $ru \in C^5$, $ru^2 \in C^3$ and $rc \in C^1$, all implied by positive $r \in C^7$.

The orders h and h^2 of $F_h(p_h)$ cancel. Its cubic coefficient is

$$2rc + ru^3/3 + A^3r/6 + A^2(ru)/2 + A(ru^2)/2 = 2rc + ru^3/3 - A^3r/12 + A(ru^2)/2.$$

It vanishes for $c = A^3r/(24r) - A(ru^2)/(4r) - u^3/6$, precisely (14). Thus $F_h(p_h) = o_{C^1}(h^3)$. The inverse at p_h is uniformly bounded, since $p_h - 1 = O_{C^1}(h)$. Using (18), a contraction on a ball of radius $O(\|F_h(p_h)\|_{C^1})$ gives $q_h - p_h = o_{C^1}(h^3)$. If the residual is zero, simply take $q_h = p_h$. The Banach-algebra logarithm gives $\log p_h = hu + h^3c + O_{C^1}(h^4)$, proving (15). Finally $s_h = \nabla \log f - 2\nabla \log q_h$; its linear term cancels in $2s_h - s_{2h}$, while the cubic coefficient is $(-4+16)\nabla c = 12\nabla c$. □

3.2 Concentration and the empirical inverse

For the remainder of this section let h be either ε or 2ε, where $\varepsilon = n^{-\kappa}$ and $0 < \kappa < 2/3$ is fixed. Set

$$\zeta_h = n^{-1/2}h^{-3/4}, \qquad \rho_h = \sqrt{\frac{\log n}{nh^{3/2}}} + \frac{\log n}{nh^{3/2}}, \qquad \delta_h = \rho_h \log n. \tag{19}$$

Then $\zeta_h \to 0$, $\rho_h(\log n)^2 \to 0$ and $\delta_h \log n \to 0$, with strict polynomial margins. These are unnormalized geometric estimates. A normalized first-order remainder need not vanish on this domain; we retain its second-order terms below.

Lemma 3.3 (Scalar and weighted operator concentration)**.** *Suppose t_h is deterministic and uniformly bounded. Then*

$$\sup_x |(\mu_n - \mu)[t_h(y)k_h(x-y)]| = O_{\mathbb{P}}(\rho_h), \tag{20}$$

$$\sup_x |\nabla_x(\mu_n - \mu)[t_h(y)k_h(x-y)]| = O_{\mathbb{P}}(h^{-1/2}\rho_h). \tag{21}$$

On $\mathcal{H}_h$, define

$$D_{t,h,n} = \frac{1}{n}\sum_{i=1}^{n} t_h(X_i)k_{h,X_i} \otimes k_{h,X_i}, \qquad D_{t,h} = \int t_h(y)k_{h,y} \otimes k_{h,y}\, d\mu(y).$$

For real t_h, including signed weights,

$$\|D_{t,h,n} - D_{t,h}\|_{\mathrm{op}} = O_{\mathbb{P}}(\rho_h). \tag{22}$$

The conclusions hold jointly for finitely many such weights and for both bandwidths.

Proof. At fixed x, the centered scalar kernel has envelope $Ch^{-3/2}$ and variance at most $Ch^{-3/2}$, since f and t_h are bounded. The corresponding derivative has envelope Ch^{-2} and variance at most $Ch^{-5/2}$. Scalar Bernstein therefore gives the displayed orders with failure probability n^{-A} for any fixed A, after enlarging constants. To obtain a spatial supremum, take a grid of mesh n^{-M} on the unit torus. It has polynomial cardinality; a union bound preserves the estimates. Derivatives of order one and two of k_h are bounded by Ch^{-2} and $Ch^{-5/2}$, respectively. They give deterministic interpolation errors $Cn^{-M}h^{-2}$ and $Cn^{-M}h^{-5/2}$ for the two fields. Choosing a fixed sufficiently large M makes these negligible. This argument also covers deterministic weights depending on h; no derivative of $t_h(y)$ is taken.

Here is a finite-dimensional reduction proving the operator assertion. In the Fourier realization of $\mathcal{H}_h$, a kernel feature has coordinates $\sqrt{\lambda_{h,k}}e^{-2\pi ik\cdot y}$; one may equivalently use a real sine–cosine basis. Project onto $|k| \le R_n = (C_M \log n/h)^{1/2}$. The projection dimension is $O(h^{-3/2}(\log n)^{3/2})$, so its logarithm is $O(\log n)$. Uniformly in y, the omitted squared feature norm is

$$\sum_{|k|>R_n} \lambda_{h,k} \le n^{-M}$$

after increasing C_M, by the Gaussian tail sum and the fixed power assumption on h. If Π is this projection, the difference between $t_h(y)k_{h,y} \otimes k_{h,y}$ and its compression has norm at most $Ch^{-3/4}n^{-M/2}$, uniformly in y. Thus the empirical and expected discarded tails are negligible.

In the finite-dimensional range let $Z_y = t_h(y)\Pi k_{h,y} \otimes \Pi k_{h,y}$. Its norm is at most $Ch^{-3/2}$ and

$$\mathbb{E}Z_X^2 \le Ch^{-3/2}\,\mathbb{E}[\Pi k_{h,X} \otimes \Pi k_{h,X}].$$

The expected unweighted feature operator has norm at most $\|f\|_\infty$: in Fourier coordinates it is $H_h^{1/2}M_f H_h^{1/2}$, up to the evident isometry. Centering subtracts the square of the expectation from the variance operator. Consequently the centered variance norm is at most $Ch^{-3/2}$, including for signed t_h. Finite matrix Bernstein, in the form

$$\mathbb{P}\left(\left\|n^{-1}\sum_i (Z_{X_i} - \mathbb{E}Z_X)\right\| > u\right) \le 2D\exp\left[-\frac{nu^2/2}{Ch^{-3/2} + Ch^{-3/2}u}\right],$$

where D is the projection dimension, proves (22). This is the usual matrix Bernstein inequality; see [11, Theorem 6.1.1]. Restoring the discarded tails finishes the proof. Finite intersections prove the last assertion. □

Lemma 3.4 (A logarithmic sup-norm inverse)**.** *Let Q be a symmetric positive semidefinite $n \times n$ matrix with nonnegative entries, $Q_{ii} < 1$, and row sums in $[1-\eta, 1+\eta]$. If $\eta \log n \to 0$, then, for all sufficiently large n,*

$$\|(I+Q)^{-1}\|_{\ell^\infty \to \ell^\infty} \le C\log n. \tag{23}$$

Proof. Schur's bound gives $\|Q\|_{2\to 2} \le 1+\eta < 2$. For $B_0 = (I-Q)/2$, positivity of the spectrum of Q gives $\|B_0\|_{2\to 2} \le 1/2$. Its absolute row sum is $(1 - Q_{ii} + \sum_{j\neq i} Q_{ij})/2 \le 1+\eta/2$. Put $m = \lceil 3\log_2 n \rceil$. For the first m powers,

$$\sum_{j<m} \|B_0^j\|_{\infty\to\infty} \le m(1+\eta/2)^m = O(\log n).$$

For the rest, Cauchy–Schwarz gives $\|B_0^j\|_{\infty\to\infty} \le \sqrt{n}\, 2^{-j}$, whose sum over $j \ge m$ is $o(1)$. Finally $(I+Q)^{-1} = \frac{1}{2}\sum_{j\ge 0} B_0^j$. □

3.3 Exact empirical scale and its off-sample derivative

For clarity fix one bandwidth and omit h from some operators. Define

$$\begin{aligned} S : L^2(\mu) &\longrightarrow \mathcal{H}_h, & Sz &= \int a_h(y) k_{h,y} z(y)\, d\mu(y), \\ C &= SS^*, & P &= S^*S, \qquad R = (I+C)^{-1}, \\ C_n &= \frac{1}{n}\sum_i a_h(X_i)^2 k_{h,X_i} \otimes k_{h,X_i}, & g &= (\mu_n - \mu)[a_h(y) k_{h,y}]. \end{aligned}$$

Here $S^*w(x) = a_h(x)w(x)$. The kernel of P relative to μ is $a_h(x)k_h(x-y)a_h(y)$. By (9), it is a positive self-adjoint Markov kernel; hence $\|S\| = \|C\| = 1$. Also $Cv_h = v_h$, and the Hilbert-space covariance identities give

$$\mathbb{E}g = 0, \qquad \mathrm{Cov}(g) = n^{-1}(C - v_h \otimes v_h), \qquad \|g\|_{\mathcal{H}_h} = O_{\mathbb{P}}(\zeta_h). \tag{24}$$

Indeed $\int a_h(y) k_{h,y}\, d\mu(y) = v_h$ and $\operatorname{tr} C \le C h^{-3/2}$.

Let $\widehat{a}_h(X_i)$ be the exact empirical symmetric scale and extend its reciprocal off sample by

$$\widehat{v}_h(x) = \frac{1}{n}\sum_i k_h(x - X_i)\widehat{a}_h(X_i), \qquad \widehat{a}_h(x) = 1/\widehat{v}_h(x).$$

These definitions agree at the sample points. Existence and uniqueness at any finite sample can also be seen by minimizing the strictly convex function

$$\Phi(t) = \frac{1}{2n^2}\sum_{i,j} k_h(X_i - X_j) e^{t_i + t_j} - \frac{1}{n}\sum_i t_i.$$

Its derivative is zero precisely when $e^{t_i}(K_{h,n}e^t)(X_i) = 1$. The diagonal terms make its Hessian positive definite. It is coercive: if $\max_i t_i \to +\infty$ the diagonal exponential dominates the possible negative linear contribution, whereas a bounded maximum and an unbounded negative coordinate make the linear term diverge.

Proposition 3.5 (Empirical scale estimates)**.** *Let f be positive C^8 and $0 < \kappa < 2/3$. At sample sites put $z_i = (\widehat{v}_h(X_i) - v_h(X_i))/v_h(X_i)$, and put $w = \widehat{v}_h - v_h$ off sample. Then*

$$\|z\|_\infty = O_{\mathbb{P}}(\delta_h), \qquad \|z\|_{L^2(\mu_n)} = O_{\mathbb{P}}(\zeta_h), \tag{25}$$

$$w = Rg + O_{\mathcal{H}_h,\mathbb{P}}(\delta_h\zeta_h), \qquad \|w\|_{\mathcal{H}_h} = O_{\mathbb{P}}(\zeta_h), \tag{26}$$

$$\|w\|_\infty = O_{\mathbb{P}}(\delta_h), \qquad \|\nabla w\|_\infty = O_{\mathbb{P}}(h^{-1/2}\delta_h). \tag{27}$$

The scales and their reciprocals are uniformly bounded in probability. Define $\widehat{S} : L^2(\mu_n) \to \mathcal{H}_h$ using the sample weights $\widehat{a}_h$, and $\widehat{C} = \widehat{S}\widehat{S}^*$. For $\theta_x = \nabla_x[a_h(x)k_{h,x}]$ and $\widehat{\theta}_x = \nabla_x[\widehat{a}_h(x)k_{h,x}]$,

$$\|\widehat{C} - C\|_{\mathrm{op}} = O_{\mathbb{P}}(\delta_h), \qquad \sup_x \|\widehat{\theta}_x - \theta_x\|_{\mathcal{H}_h} = O_{\mathbb{P}}(h^{-5/4}\delta_h). \tag{28}$$

Proof. Set $a_i = a_h(X_i)$ and $Q_{ij} = a_i k_h(X_i - X_j)a_j/n$. This matrix is symmetric PSD and nonnegative. Its row sums are $1 + E_i$, where $E_i = a_i g(X_i)$. By (20), $\|E\|_\infty = O_{\mathbb{P}}(\rho_h)$; also $Q_{ii} \le C/(nh^{3/2}) = o(1)$. The preceding inverse lemma applies on an event of probability tending to one.

The sample scaling equation is exactly equivalent to

$$(I + Q)z = E + Q\frac{z^2}{1+z}, \tag{29}$$

where operations inside the fraction are coordinatewise. Applying $(I + Q)^{-1}$ defines a contraction on a sup-norm ball of radius $C\rho_h \log n = C\delta_h$: its forcing has that order, and the nonlinear Lipschitz constant is $O(\delta_h \log n) = o(1)$. The constructed branch has $1 + z_i > 0$ and therefore agrees with the unique empirical scale. This proves the first bound in (25).

By the weighted operator estimate, $\|C_n - C\|_{\mathrm{op}} = O_{\mathbb{P}}(\rho_h)$, so $\|C_n\| = O_{\mathbb{P}}(1)$. The exact identity

$$\|E\|^2_{L^2(\mu_n)} = \langle g, C_n g\rangle_{\mathcal{H}_h}$$

and (24) imply $\|E\|_2 = O_{\mathbb{P}}(\zeta_h)$. Since $\|(I + Q)^{-1}\|_{2\to 2} \le 1$ and $\|Q\|_{2\to 2} = O_{\mathbb{P}}(1)$, taking $L^2(\mu_n)$ norms in (29) and absorbing $C\|z\|_\infty\|z\|_2$ proves its second bound.

Off sample, the defining convolution gives

$$w = g - K_{h,n}\frac{a_h z}{1+z}, \qquad (I + C_n)w = g + K_{h,n}\frac{a_h z^2}{1+z}. \tag{30}$$

All terms lie in $\mathcal{H}_h$. The map $b \mapsto n^{-1}\sum_i a_i k_{h,X_i} b_i$ has norm $\|C_n\|^{1/2}$ from $L^2(\mu_n)$ to $\mathcal{H}_h$. Consequently the last term has norm $O_{\mathbb{P}}(\delta_h\zeta_h)$. Positivity gives $\|(I + C_n)^{-1}\| \le 1$; its resolvent identity with R, and $\rho_h \le \delta_h$, prove (26).

For (27), use the first identity in (30), scalar concentration, and the sample sup-norm bound. The derivative term is bounded using (10) and $\sup_x K_{2h,n}1(x) = O_{\mathbb{P}}(1)$, which follows from (20). Since v_h stays bounded away from zero, $\widehat{v}_h = v_h + w$ does also with probability tending to one. Reciprocation gives

$$\|\widehat{a}_h - a_h\|_\infty = O_{\mathbb{P}}(\delta_h), \qquad \|\nabla\widehat{a}_h - \nabla a_h\|_\infty = O_{\mathbb{P}}(h^{-1/2}\delta_h).$$

For $\widehat{C}$, replace its sample weight $\widehat{a}_h^2$ by a_h^2 using these uniform bounds, and then use (22). For the feature derivative, the coefficient error multiplies $\nabla k_{h,x}$, whereas the derivative coefficient error multiplies $k_{h,x}$. Applying (12) gives $\delta_h h^{-5/4} + h^{-1/2}\delta_h h^{-3/4}$, proving (28). No independence between scaling weights and observations is used in these pathwise estimates. □

3.4 Local kernels and repeated-observation terms

We retain the nonlinear terms needed when a first-order norm estimate does not vanish at the statistical normalization. Write h for one of the two bandwidths, omit its subscript on a, v, C, R, and put $m = a^2 f = q_h^2$, $\nu = \mu_n - \mu$ and $D = C_n - C$.

Lemma 3.6 (Physical resolvent and local rows)**.** *The physical action of C on its function arguments is HM_m. The operator $J = (I + HM_m)^{-1}$ is uniformly bounded on C^0 and C^1, and agrees with R on heat-RKHS function arguments. Set*

$$L(x, y) = \langle k_{h,x}, Rk_{h,y}\rangle_{\mathcal{H}_h}, \qquad \ell(x, y) = a(y)L(x, y) - v(x)/2.$$

The integral operator $L = JH$ is symmetric and positive, and $0 \le L \le H$ on $L^2(dx)$. Its rows satisfy, uniformly in x,

$$\|L(x,\cdot)\|_1 \le C, \qquad \|\partial_x L(x,\cdot)\|_1 \le Ch^{-1/2}, \tag{31}$$

$$\|L(x,\cdot)\|_2 \le Ch^{-3/4}, \qquad \|\partial_x L(x,\cdot)\|_2 \le Ch^{-5/4}, \tag{32}$$

$$\|L\|_\infty \le Ch^{-3/2}, \qquad \|\partial_x L\|_\infty \le Ch^{-2}. \tag{33}$$

These bounds hold for rows measured against either dx or μ. The kernel ℓ is centered in its second variable and has the same L^2 and supremum row bounds, including their derivative bounds. Its integral operator on $L^2(\mu)$ has uniformly bounded norm. Finally, with $w_1 = Rg$, one has

$$\|w_1\|_\infty = O_{\mathbb{P}}(\rho_h), \qquad \|\nabla w_1\|_\infty = O_{\mathbb{P}}(h^{-1/2}\rho_h), \qquad w_1(x) = \nu_y \ell(x, y). \tag{34}$$

Proof. The physical formula follows directly from $Cb(x) = \int a(y)^2 k_h(x-y)b(y)\, d\mu(y)$. Since $m - 1 = O_{C^1}(h)$, factorization gives

$$J = [I + (I+H)^{-1}HM_{m-1}]^{-1}(I+H)^{-1}.$$

Lemma 3.1 and a Neumann series on C^0 and C^1 give the claimed uniform norms. The resolvent intertwining identity gives

$$L = JH = H^{1/2}(I + H^{1/2}M_m H^{1/2})^{-1}H^{1/2}.$$

This proves symmetry, positivity and the stated operator order. RKHS evaluation, $\|R\| \le 1$, and the fact that the unweighted feature covariance on Lebesgue space has norm at most one give (32)–(33).

The C^0 norm of the smooth-kernel operator JH is its largest absolute row integral: approximate the sign of a fixed row by continuous functions. Hence its row L^1 norm is bounded. The identity $L = H - HM_mL$ gives

$$\partial_x L = \partial_x H - (\partial_x H)M_m L.$$

The Gaussian derivative has row L^1 norm $O(h^{-1/2})$. Integrating the absolute value on the right proves (31); no derivative of an arbitrary inverse kernel is used. Equivalence of the two measures follows from the fixed positive bounded density. Since $Cv = v$, $Rv = v/2$, and therefore $\int a(y)L(x,y)d\mu(y) = v(x)/2$. This proves centering and the formula for w_1. The bounded L^2 operator norm of ℓ follows from that of L, bounded multiplication by a, f, and its bounded rank-one centering term. Finally $w_1 = Jg$; the scalar field and derivative estimates of Lemma 3.3, followed by the uniform C^0/C^1 bounds for J, prove (34). □

We record collision estimates also used for the trajectory. All kernels in the next two lemmas are deterministic and may depend on h. Suppose $\beta(u, y)$ and $\chi(u, y)$ are centered in y, have bounded integral-operator norms on $L^2(\mu)$, and obey

$$\begin{aligned} \sup_u \|\beta(u,\cdot)\|_2 + \sup_u \|\chi(u,\cdot)\|_2 &\le Ch^{-3/4}, \\ \sup_u \|\partial_u\beta(u,\cdot)\|_2 + \sup_u \|\partial_u\chi(u,\cdot)\|_2 &\le Ch^{-5/4}, \\ \|\beta\|_\infty + \|\chi\|_\infty &\le Ch^{-3/2}, \\ \|\partial_u\beta\|_\infty + \|\partial_u\chi\|_\infty &\le Ch^{-2}. \end{aligned} \tag{35}$$

Let A have L^1, L^2, L^∞ norms bounded respectively by $Ch^{-1/2}, Ch^{-5/4}, Ch^{-2}$. Constants and deterministic geometric weights in these assumptions propagate multiplicatively through the following statements, by homogeneity. The derivative in each statement is one fixed query coordinate.

Lemma 3.7 (Quadratic empirical products)**.** *For a deterministic coefficient b with uniformly bounded C^1 norm, consider any of*

$$A(y)\beta(y,z), \qquad \int A(u)\beta(u,y)\chi(u,z)d\mu(u), \qquad \partial_x[b(x)\beta(x,y)\chi(x,z)].$$

Center the kernel separately in its two variables, and symmetrize it; call the resulting kernel K. Then

$$\|K\|_{L^2(\mu^2)} \le Ch^{-2}, \quad \|K(X,X)\|_{L^2} \le Ch^{-11/4}, \quad |\mathbb{E}K(X,X)| \le Ch^{-2}. \tag{36}$$

Consequently

$$\nu^{\otimes 2}K = O_{\mathbb{P}}(n^{-1}h^{-2} + n^{-3/2}h^{-11/4}), \qquad \sqrt{n}h^{5/4}\nu^{\otimes 2}K = O_{\mathbb{P}}(\zeta_h + \zeta_h^2). \tag{37}$$

Proof. For the first kernel, use $\|A\|_2$ times the largest L^2 row of β for the full norm, and $\|A\|_2\|\beta\|_\infty$ on the diagonal. Its diagonal mean is bounded by $\|A\|_1 \sup_y |\beta(y,y)|$. For the integral kernel, Minkowski and the two row L^2 bounds give its full norm h^{-2}; on the diagonal use one supremum and one row L^2 norm, giving $h^{-11/4}$. Its diagonal mean uses the product of the two row L^2 norms and $\|A\|_1$. For the differentiated product, one derivative row norm times one ordinary row norm gives h^{-2}; one supremum and the other derivative row norm give the diagonal $h^{-11/4}$ bound. Cauchy–Schwarz gives the same h^{-2} diagonal mean bound. Bounded derivatives of b add smaller terms.

The latter two kernels are already centered in both variables. For the first, only its first-variable projection can be nonzero; that projection has zero integral and L^2 norm at most the original joint norm by Jensen. Substituting its single variable on the diagonal adds no new norm factor. Thus centering preserves all three asserted orders and the diagonal mean; symmetrization does too, up to constants. For a canonical symmetric kernel, expanding the square of the all-distinct sum leaves only matching unordered pairs. Its standard deviation is at most $C\|K\|_2/n$. The diagonal contribution has expectation $\mathbb{E}K(X,X)/n$ and standard deviation at most $\|K(X,X)\|_2/n^{3/2}$. This proves both rates. □

Lemma 3.8 (Cubic empirical numerator and all collisions)**.** *Set $T_0(u,y,z) = A(u)\beta(u,y)\chi(u,z)$ and $T = T_0 - \int T_0(t,y,z)d\mu(t)$. Then T is centered in every variable, $\nu^{\otimes 3}T = \nu^{\otimes 3}T_0$, and*

$$\sqrt{n}h^{5/4}\nu^{\otimes 3}T = O_{\mathbb{P}}(\zeta_h^2 + \zeta_h^3). \tag{38}$$

In particular this explicit cubic statistic is negligible throughout the strict occupancy domain $0 < \kappa < 2/3$.

Proof. The full $L^2(\mu^3)$ norm is at most $Ch^{-11/4}$ by $\|A\|_2$ and the two row L^2 norms. Each collapsed-pair kernel $T(U,U,Z), T(U,Y,U), T(U,Y,Y)$ has L^2 norm at most $Ch^{-7/2}$: use one diagonal supremum and one row norm, or a supremum times a row norm for the square. The removed integral has joint norm Ch^{-2} and diagonal norm $Ch^{-11/4}$ by the preceding proof, which are smaller. Also $\mathbb{E}|T(U,U,U)| \le Ch^{-7/2}$, using $\|A\|_1$ and two diagonal suprema. Here independent letters have law μ.

One must control the first projections of the collapsed kernels more accurately than their full norms. For the $u = y$ collision, its nonzero projection is the adjoint integral operator of χ applied to $A(u)\beta(u,u)$; it has L^2 norm at most $Ch^{-11/4}$. The $u = z$ case exchanges β, χ. For $y = z$, integration in u gives zero by its canonicalization; the other projection is

$$A(u)\mathbb{E}[\beta(u,Y)\chi(u,Y)] - \int A(t)\mathbb{E}[\beta(t,Y)\chi(t,Y)]d\mu(t),$$

again of norm $Ch^{-11/4}$. Each collapsed kernel has global mean zero, but need not be canonical in both remaining variables.

Expand $n^{-3}\sum_{i,j,k} T(X_i,X_j,X_k)$ into equality partitions. The all-distinct part has standard deviation at most $Cn^{-3/2}h^{-11/4}$, because canonicity leaves only matching triples. For each of the three exactly-two-equal partitions, its coefficient is $O(1/n)$ times an order-two statistic. Hoeffding's two first projections and its canonical remainder give standard deviation at most

$$C(n^{-3/2}h^{-11/4} + n^{-2}h^{-7/2}).$$

The fully repeated part has coefficient n^{-2} times an empirical mean; its absolute expectation is at most $Cn^{-2}h^{-7/2}$, so Markov's inequality suffices. After multiplication by $\sqrt{n}h^{5/4}$ these orders are respectively ζ_h^2, $\zeta_h^2 + \zeta_h^3$ and ζ_h^3. No repeated-observation mean has been discarded. □

3.5 Second-order scale and score expansion

Proposition 3.9 (Retained empirical scaling numerator)**.** *For $0 < \kappa < 2/3$, put $w_1 = Rg$ and*

$$w_2 = R\{K_{h,n}[a^3 w_1^2] - Dw_1\}.$$

Then $\|w_2\|_{\mathcal{H}_h} + \|w_2\|_{L^2(\mu_n)} = O_{\mathbb{P}}(\delta_h\zeta_h)$ and

$$w = w_1 + w_2 + O_{\mathcal{H}_h,\mathbb{P}}(\delta_h^2\zeta_h). \tag{39}$$

Proof. The exact reciprocal equation is

$$(I + C_n)w = g + K_{h,n}\frac{a^3w^2}{1+aw}.$$

The first-order bound in Proposition 3.5 and bounded sample evaluation give $\|w - w_1\|_{L^2(\mu_n)} = O_{\mathbb{P}}(\delta_h\zeta_h)$. Both w and w_1 have sup norm $O_{\mathbb{P}}(\delta_h)$, so

$$\|w^2 - w_1^2\|_{L^2(\mu_n)} + \|aw^3/(1+aw)\|_{L^2(\mu_n)} = O_{\mathbb{P}}(\delta_h^2\zeta_h).$$

The bounded sample-to-feature map transfers these estimates to $\mathcal{H}_h$. Moving Dw to the right in the exact equation and replacing it by Dw_1 costs $O_{\mathbb{P}}(\rho_h\delta_h\zeta_h)$. Applying R proves the expansion. The displayed norm of w_2 follows from $\|K_{h,n}[a^3w_1^2]\|_{\mathcal{H}_h} \le C_{\mathbb{P}}\|w_1\|_\infty\|w_1\|_{L^2(\mu_n)}$ and $\|Dw_1\| \le O_{\mathbb{P}}(\rho_h)\|w_1\|$. This proof retains $K_{h,n}$; it uses neither multiplication closure of a Gaussian RKHS nor concentration for a random deterministic weight. □

Proposition 3.10 (Normalized exact-score linearization)**.** *For $0 < \kappa < 4/9$ and each fixed query x,*

$$\widehat{s}_h - s_h = 2\nabla(aw_1) + 2\nabla(aw_2) - \nabla(a^2w_1^2) + O_{\mathbb{P}}(h^{-5/4}\delta_h^2\zeta_h), \tag{40}$$

where all fields are evaluated at x. Moreover

$$\sqrt{n}h^{5/4}\{\widehat{s}_h(x) - s_h(x) - 2\langle\theta_x, Rg\rangle_{\mathcal{H}_h}\} \longrightarrow 0 \quad \textit{in probability.}$$

Proof. The expansion of $2\nabla\log(1+aw)$ has linear term $2\nabla(aw)$ and quadratic term $-\nabla(a^2w^2)$. The cubic logarithm remainder is bounded by $C\delta_h^2$ times a derivative evaluation of w, thus by $O_{\mathbb{P}}(h^{-5/4}\delta_h^2\zeta_h)$. Replacing w in the linear term uses Proposition 3.9. For the quadratic replacement, set $e = w - w_1$. The product rule bounds its derivative by a constant times

$$\delta_h|\nabla e(x)| + |e(x)|\,|\nabla w_1(x)| + \delta_h|e(x)|.$$

RKHS evaluation bounds e and its derivative by $h^{-3/4}\delta_h\zeta_h$ and $h^{-5/4}\delta_h\zeta_h$, respectively. The improved derivative bound (34) controls the middle product. Each term is at most the remainder in (40). Its normalized size satisfies

$$\sqrt{n}\,\delta_h^2\zeta_h = O(n^{-1}h^{-9/4}(\log n)^3) \longrightarrow 0 \qquad (\kappa < 4/9). \tag{41}$$

It remains to analyze the explicit second-order terms, not to bound them all by the preceding first-order norm estimate. For one coordinate let $\theta_x = \partial_x[a(x)k_{h,x}]$ and set

$$A_x(u) = 2a(u)^3\langle\theta_x, Rk_{h,u}\rangle, \qquad B_x(u) = 2a(u)^2\langle\theta_x, Rk_{h,u}\rangle.$$

Lemma 3.6 gives both left rows the L^1, L^2, L^∞ orders used in the collision lemmas. Replacing $K_{h,n}$ in the explicit w_2 numerator by its population part gives the quadratic raw kernel

$$K_0(y,z) = \int A_x(u)\ell(u,y)\ell(u,z)d\mu(u) - B_x(y)\ell(y,z) - \partial_x[a(x)^2\ell(x,y)\ell(x,z)].$$

Centering and symmetrizing leave its $\nu^{\otimes 2}$ statistic unchanged. Lemma 3.7 bounds its normalized size by $O_{\mathbb{P}}(\zeta_h + \zeta_h^2)$. The remaining empirical-minus-population numerator is exactly

$$\nu_u\nu_y\nu_z A_x(u)\ell(u,y)\ell(u,z),$$

whose normalized size is $O_{\mathbb{P}}(\zeta_h^2 + \zeta_h^3)$ by Lemma 3.8. In particular the empirical nonlinear numerator has not been inserted into a deterministic-weight covariance bound. Both explicit statistics vanish throughout $\kappa < 2/3$. Together with (41) this proves the assertion. The finite intersection of these events at $h = \varepsilon, 2\varepsilon$ needs no independence. □

3.6 Influence kernel and the same-sample Gaussian law

Theorem 3.11 (Joint corrected exact-score limit)**.** *Let f be a fixed positive C^8 density on $\mathbb{T}^3$, fix x, and take $\varepsilon = n^{-\kappa}$ with $0 < \kappa < 4/9$. At both bandwidths use the same sample and the exact empirical scales above. Put*

$$h_c(\xi) = \frac{e^{-2\pi^2 c|\xi|^2}}{1 + e^{-2\pi^2 c|\xi|^2}}, \qquad J_R = \int_{\mathbb{R}^3} (2\pi\xi_1)^2 [2h_1(\xi) - h_2(\xi)]^2 \, d\xi.$$

Then

$$\sqrt{n}\,\varepsilon^{5/4}\{2\widehat{s}_\varepsilon(x) - \widehat{s}_{2\varepsilon}(x) - [2s_\varepsilon(x) - s_{2\varepsilon}(x)]\} \implies N_3\left(0, \frac{4J_R}{f(x)} I_3\right). \tag{42}$$

If also $\kappa > 2/17$, the population bracket may be replaced by $\nabla \log f(x)$. At $\kappa = 2/17$, its replacement instead gives mean $12\nabla c(x)$, with c from (14)*.*

Proof. We first identify an influence within each bandwidth's own Hilbert space; we never equate the RKHS norms at different bandwidths. The second-order score expansion in Proposition 3.10 gives

$$\widehat{s}_h(x) - s_h(x) = 2\langle \theta_x, Rg\rangle_{\mathcal{H}_h} + o_{\mathbb{P}}((\sqrt{n}\,h^{5/4})^{-1}). \tag{43}$$

The explicit quadratic and cubic empirical terms in that proposition are negligible after normalization; its remaining cubic norm bound vanishes for $\kappa < 4/9$. The differential of $\nabla \log v$ is $\nabla(a_h w)$, giving exactly the derivative feature θ_x.

We put the influences into the common space $L^2(dx)$. Let

$$T_h = M_{q_h} H_h M_{q_h}, \qquad F(T) = T(I+T)^{-1}, \qquad G_h = F(H_h).$$

Multiplication $U : L^2(\mu) \to L^2(dx)$ by r is unitary and $UPU^{-1} = T_h$. The feature factorization and resolvent identity imply $S^*RS = P(I+P)^{-1}$. Therefore the centered iid influence in (43) has kernel

$$\psi_h(x,y) = 2\nabla_x \left\{ \frac{F(T_h)(x,y)}{r(x)r(y)} \right\}. \tag{44}$$

Indeed the kernel of S^*RS relative to μ is $a_h(x)a_h(y)\langle k_{h,x}, Rk_{h,y}\rangle$; differentiating in x gives the feature expression in (43). Also $T_h r = r$, so $F(T_h)r = r/2$, and integration of (44) against $d\mu(y) = r(y)^2 dy$ gives zero. Equivalently $Rv_h = v_h/2$ makes the centering feature contribution vanish. Thus

$$\widehat{s}_h(x) - s_h(x) = \frac{1}{n}\sum_{i=1}^n \psi_h(x, X_i) + o_{\mathbb{P}}((\sqrt{n}\,h^{5/4})^{-1}). \tag{45}$$

We next justify the leading heat multiplier in the covariance, including the derivative row estimate. Write $E_h = T_h - H_h$ and $R_T = (I+T_h)^{-1}$, $R_H = (I+H_h)^{-1}$. By (13),

$$\|E_h\|_{2\to 2} = O(h), \quad \|E_h(x,\cdot)\|_2 = O(hh^{-3/4}), \quad \|\nabla_x E_h(x,\cdot)\|_2 = O(hh^{-5/4}).$$

The row estimates follow directly from the kernel $q_h(x)k_h(x-y)q_h(y) - k_h(x-y)$ and its derivative; derivatives of $q_h - 1$ are $O(h)$. Both resolvents have L^2 norm at most one by positivity. The exact identity

$$F(T_h) - F(H_h) = R_H E_h R_T = E_h R_T - G_h E_h R_T$$

now gives

$$\|\nabla_x [F(T_h) - F(H_h)](x,\cdot)\|_2 = O(hh^{-5/4}), \tag{46}$$

and the corresponding undifferentiated bound is $O(hh^{-3/4})$. For the second term of the identity, differentiate only the convolution row of G_h and multiply its L^2 norm by $\|E_h R_T\|$. The multipliers of G_h are $\lambda_{h,k}/(1+\lambda_{h,k})$; their row and derivative-row norms are bounded by those of H_h. This proves (46) without attempting to differentiate the kernel of an arbitrary bounded operator.

In an $L^2(\mu)$ inner product of influences, the $r(y)^{-1}$ factor in (44) cancels with the density $r(y)^2$. Consequently the leading Lebesgue row is

$$2\{r(x)^{-1}\nabla_x G_h(x,\cdot) + \nabla(r^{-1})(x)G_h(x,\cdot)\}.$$

The undifferentiated term is $O(h^{-3/4})$ and is negligible relative to the derivative row $O(h^{-5/4})$. The error in replacing $F(T_h)$ by G_h has an additional factor h, by (46). Form the SAME-SAMPLE influence $W_\varepsilon = 2\psi_\varepsilon - \psi_{2\varepsilon}$ before taking any covariance. Parseval's identity gives, for coordinates j, ℓ,

$$\begin{aligned}\varepsilon^{5/2}\mathbb{E}[W_{\varepsilon,j}(x,X)W_{\varepsilon,\ell}(x,X)] &= \frac{4}{f(x)}\varepsilon^{5/2}\sum_{k\in\mathbb{Z}^3}(2\pi k_j)(2\pi k_\ell)\left[\frac{2\lambda_{\varepsilon,k}}{1+\lambda_{\varepsilon,k}} - \frac{\lambda_{2\varepsilon,k}}{1+\lambda_{2\varepsilon,k}}\right]^2 + o(1) \qquad (47)\\ &\longrightarrow \frac{4}{f(x)}\int_{\mathbb{R}^3}(2\pi\xi_j)(2\pi\xi_\ell)[2h_1(\xi) - h_2(\xi)]^2\,d\xi.\end{aligned}$$

The passage to the integral is the Gaussian-dominated lattice Riemann sum with $\xi = \sqrt{\varepsilon}\,k$. Radial symmetry makes the off-diagonal integrals zero and all diagonal integrals equal to J_R. This number is finite and positive. In particular the cross covariance between the two bandwidths has been retained, rather than replaced by independent sampling variances.

Finally the feature representation and (12) give $\sup_y |\psi_h(x,y)| \le Ch^{-2}$: the derivative feature norm is $O(h^{-5/4})$, the input feature norm is $O(h^{-3/4})$, and $\|R\| \le 1$. Each normalized iid summand therefore has envelope

$$\sup_y \frac{\varepsilon^{5/4}|W_\varepsilon(x,y)|}{\sqrt{n}} \le \frac{C}{\sqrt{n}\,\varepsilon^{3/4}} \longrightarrow 0.$$

The triangular-array Lindeberg theorem, applied by Cramér–Wold, and (47) prove (42). By (16), the normalized centering difference is

$$\sqrt{n}\,\varepsilon^{17/4}\{12\nabla c(x) + o(1)\}.$$

It tends to zero for $\kappa > 2/17$ and to $12\nabla c(x)$ at equality, which proves the remaining assertions. □

The exponent $4/9$ in this theorem is a sufficient restriction from the remaining cubic norm bound (41). Its equality case is not excluded by a statistical counterexample here. The lower boundary concerns the specified two-bandwidth correction and true-score centering; it is distinct from the population-centered limit, which holds on the larger interval $0 < \kappa < 4/9$. Later sections determine when the actual first residual crossing preserves this exact-estimator law.

4 The actual first residual hit

All logarithms are natural. We use the empirical inner product $\langle u, v\rangle_n = n^{-1}\sum_i u(X_i)v(X_i)$ and its norm $\|u\|_{2,n}$. Constants below can depend on the fixed density, its lower bound, and the fixed exponents, but not on n or the iteration index. Statements involving several bandwidths refer to their common sample.

For a bandwidth h, write $K_{h,n}u(x) = n^{-1}\sum_i k_h(x - X_i)u(X_i)$. Starting with $b_0 = 1$, perform the updates

$$b_{j+1}(X_i) = \left\{\frac{b_j(X_i)}{K_{h,n}b_j(X_i)}\right\}^{1/2}, \qquad s_{j,h}(x) = 2\,\nabla \log K_{h,n}b_j(x). \qquad (48)$$

Define $F_j = (b_j K_{h,n} b_j)^{1/2}$ at the sample sites and

$$r_j = \left\|\frac{F_j}{\langle 1, F_j\rangle_n} - 1\right\|_{2,n}, \quad t_{n,h} = n^{-1/2}h^{-\gamma}, \quad L_n = \max\{3, \lceil 4\log n\rceil\}. \qquad (49)$$

The stopping time τ_h is the first $j \in \{0,\dots,L_n\}$ with $r_j \le t_{n,h}$; if there is no such index, set $\tau_h = L_n$. Every index, including zero, is checked. Denote the exact empirical scale by $\widehat{a}_h$, and its score by $\widehat{s}_h$.

Theorem 4.1. *Let f be a fixed positive C^8 probability density on $\mathbb{T}^3$, bounded away from zero, and let x be fixed and deterministic. Suppose*

$$\varepsilon = n^{-\kappa}, \qquad \frac{2}{17} < \kappa < \frac{4}{9}, \qquad 0 < \gamma < \frac{3}{4}.$$

At each $h \in \{\varepsilon, 2\varepsilon\}$, the rule (49) *genuinely crosses the threshold before L_n with probability tending to one. Moreover,*

$$\frac{\tau_h}{\log(1/h)} \longrightarrow \frac{3/4 - \gamma}{\log 2} \quad \textit{in probability}, \tag{50}$$

$$\mathbb{P}\{c t_{n,h} \le r_{\tau_h} \le t_{n,h}\} \longrightarrow 1 \qquad (0 < c < 1/2), \tag{51}$$

$$\sqrt{n} h^{5/4} \left| s_{\tau_h,h}(x) - \widehat{s}_h(x) \right| \longrightarrow 0 \quad \textit{in probability}. \tag{52}$$

In fact, the normalized difference in (52) *tends to zero uniformly over $J_n \le j \le L_n$ for every deterministic $J_n \to \infty$, $J_n \le L_n$. Consequently*

$$\sqrt{n}\varepsilon^{5/4}\{2 s_{\tau_\varepsilon,\varepsilon}(x) - s_{\tau_{2\varepsilon},2\varepsilon}(x) - \nabla \log f(x)\} \;\Rightarrow\; N_3\left(0, \frac{4 J_R}{f(x)} I_3\right), \tag{53}$$

where

$$J_R = \int_{\mathbb{R}^3} (2\pi\xi_1)^2 [2h_1(\xi) - h_2(\xi)]^2 \, \mathrm{d}\xi, \qquad h_c(\xi) = \frac{e^{-2\pi^2 c|\xi|^2}}{1 + e^{-2\pi^2 c|\xi|^2}}.$$

The two stopping times are allowed to be dependent.

4.1 Feature geometry and a local projective comparison

For the proof fix $h = \varepsilon$ and suppress its subscript. Set

$$\zeta = n^{-1/2}\varepsilon^{-3/4}, \quad \rho = \sqrt{\frac{\log n}{n\varepsilon^{3/2}}} + \frac{\log n}{n\varepsilon^{3/2}}, \quad d_n = \rho \log n, \quad A_n = \sqrt{n}\varepsilon^{5/4}. \tag{54}$$

The letter d_n denotes a deterministic error envelope, whereas d_j below denotes a population iterate. For every fixed $0 < \kappa < 2/3$,

$$d_n \log n = o(1), \qquad \zeta = o(1). \tag{55}$$

The scale and concentration estimates used here hold for every fixed $0 < \kappa < 2/3$. We retain geometric factors on all errors, and never assume $\sqrt{n} d_n \zeta = o(1)$ or $\sqrt{n}\zeta^2 = o(1)$. The deterministic background coefficient $\sqrt{n}\varepsilon^{19/4}$ tends to zero when $\kappa > 2/19$.

Let $\mathcal{H}$ be the RKHS of k_ε and $k_y = k_\varepsilon(\cdot - y)$. Let a be the population scale, $v = 1/a$, $F = H_\varepsilon f$, and define

$$Su = \int a(y) k_y u(y) \, \mathrm{d}\mu(y), \quad C = SS^*, \quad P = S^*S, \quad B = (I - C)/2, \quad R = (I + C)^{-1}.$$

Then $S^* z(y) = a(y) z(y)$ and $P1 = 1$. The operator P is self-adjoint, positive semidefinite, and Markov on $L^2(\mu)$. From this point set $B_0 = (I - P)/2$ on $L^2(\mu)$. Thus

$$\|S\| = \|C\| = 1, \quad \|B\| \le 1/2, \quad \|R\| \le 1, \quad Cv = v, \quad \|v\|_{\mathcal{H}} = 1, \quad S^* v = 1. \tag{56}$$

Here $v = \int a k_y \, \mathrm{d}\mu(y)$, and its squared RKHS norm is $\int a v \, \mathrm{d}\mu = 1$. The empirical analogues $S_n, \widehat{C}, P_n$ use $\widehat{a}$ and μ_n, so $P_n = S_n^* S_n$ is again PSD Markov, $\|S_n\| = 1$, $\left\|\widehat{C}\right\| = 1$, and $B_n = (I - P_n)/2$ has norm at most $1/2$. All empirical operator norms in this section are relative to $L^2(\mu_n)$ when the domain is a sample vector.

For off-sample evaluation use

$$P_n u(x) = \widehat{a}(x) K_n(\widehat{a}u)(x), \qquad P_n 1(x) = 1, \quad \theta_x = \nabla[a(x)k_x], \quad \widehat{\theta}_x = \nabla[\widehat{a}(x)k_x].$$

Proposition 3.5 and Lemma 3.1 give, on common events whose probabilities tend to one,

$$\left\|\widehat{C} - C\right\|_{\mathrm{op}} = O_{\mathbb{P}}(d_n), \qquad \left\|\widehat{\theta}_x - \theta_x\right\|_{\mathcal{H}} = O_{\mathbb{P}}(d_n \varepsilon^{-5/4}), \tag{57}$$

$$|P_n u(x)| \le O_{\mathbb{P}}(\varepsilon^{-3/4}) \, \|u\|_{2,n}, \qquad |\nabla P_n u(x)| \le O_{\mathbb{P}}(\varepsilon^{-5/4}) \, \|u\|_{2,n}. \tag{58}$$

The last estimates hold for every sample vector on the same event; hence they also hold for vectors depending on the observations. The derivative row also has a total-variation bound

$$\sup_x \frac{1}{n} \sum_i |\nabla_x \{\widehat{a}(x) k_\varepsilon(x - X_i) \widehat{a}(X_i)\}| = O_{\mathbb{P}}(\varepsilon^{-1/2}). \tag{59}$$

The scale derivative has size $O_{\mathbb{P}}(1 + \varepsilon^{-1/2} d_n)$, and the heat derivative is dominated by $C\varepsilon^{-1/2} k_{2\varepsilon}$. Bounded positive weights and the KDE bound at 2ε prove the assertion. The row sums to one, so its derivative annihilates constants.

With relative logs $\ell_j = \log(b_j/\widehat{a})$, the iteration is exactly

$$\ell_{j+1} = \Psi_n(\ell_j), \qquad \Psi_n(z) = \tfrac{1}{2}\{z - \log(P_n e^z)\}. \tag{60}$$

Its derivative is $(I - Q_z)/2$, where Q_z is P_n tilted by e^z and renormalized to have row sum one. It follows that Ψ_n is globally one-Lipschitz in sup norm and $\|\Psi_n(z)\|_\infty \le \|z\|_\infty$. For $\|z\|_\infty \le s$ small, positivity and self-adjoint Markov contraction give

$$\|Q_z - P_n\|_{\mathrm{op}} \le Cs, \qquad \|\Psi_n(z) - B_n z\|_{2,n} \le C \, \|z\|_\infty \, \|z\|_{2,n}. \tag{61}$$

Indeed $|(Q_z - P_n)u| \le CsP_n|u|$ pointwise. The same bounds hold for the population map Ψ defined by P.

The residual is invariant under adding constants to z, and Ψ_n kills such constants. If z is centered and has small sup norm, its linear residual is $(I + P_n)z/2$, which is also centered and has norm between $\|z\|_{2,n}/2$ and $\|z\|_{2,n}$. Taylor expansion in the Banach algebra of bounded sample vectors gives a remainder at most $C \, \|z\|_\infty \, \|z\|_{2,n}$. Consequently there are absolute local constants, independent of n and ε, such that

$$c_0 \, \|z - \langle 1, z\rangle_n\|_{2,n} \le r(z) \le C_0 \, \|z - \langle 1, z\rangle_n\|_{2,n}, \qquad r(\Psi_n(z)) \le C_1 r(z) \tag{62}$$

when the oscillation of z is sufficiently small. This comparison will exclude actual early hits, rather than altering the stopping rule.

4.2 Four population updates and their sampled counterparts

Write $d_j = \log(b_{j,\mathrm{pop}}/a)$. The population starts at the same $b_{0,\mathrm{pop}} = 1$ and obeys $d_{j+1} = \Psi(d_j)$. The population estimate in Lemma 3.2, used in C^6, gives $d_1 = \log v - \frac{1}{2}\log F = O_{C^6}(\varepsilon)$. Since $P1 = 1$,

$$(P - I)u = a\{H_\varepsilon(afu) - uH_\varepsilon(af)\}, \qquad \|(P - I)u\|_{C^s} \le C\varepsilon \, \|u\|_{C^{s+2}}.$$

Apply this to $e^d - 1$ in the identity $e^{-d}Pe^d = 1 + e^{-d}(P - I)(e^d - 1)$. Taking a logarithm proves $\|\Psi(d)\|_{C^s} \le C\varepsilon \, \|d\|_{C^{s+2}}$ for small smooth d. It follows that the actual population updates satisfy

$$d_2 = O_{C^4}(\varepsilon^2), \qquad d_3 = O_{C^2}(\varepsilon^3), \qquad d_4 = O_{C^0}(\varepsilon^4). \tag{63}$$

This is an analytic decomposition of the original trajectory, not a mandatory four-step delay.

Put $g = (\mu_n - \mu)(ak)$ and $g_0 = (\mu_n - \mu)k$. Both have RKHS norm $O_{\mathbb{P}}(\zeta)$, and

$$\mathrm{Cov}(g) = n^{-1}(C - v \otimes v), \qquad \langle v, g\rangle = 0. \tag{64}$$

Expansion of the first update using Proposition 3.5 gives, with $u_1 = \ell_1 - d_1$,

$$u_1 = S_n^* BRg + h_c + O_{L^2(\mu_n),\mathbb{P}}(d_n\zeta), \quad \|u_1\|_{2,n} = O_{\mathbb{P}}(\zeta), \quad \|u_1\|_\infty = O_{\mathbb{P}}(d_n), \tag{65}$$

where

$$h_c = \tfrac{1}{2}(ag - g_0/F). \tag{66}$$

For completeness, before using $R = BR + I/2$, the expansion is $u_1 = aRg - g_0/(2F) + O_{L_n^2,\mathbb{P}}(d_n\zeta)$. The logarithm remainders use a sup factor d_n and sample L^2 factor ζ; replacing a by $\widehat{a}$ costs the same product.

We need a quantitative comparison for the next three actual updates. If $d = d_j$, $1 \le j \le 3$, normalized row differences and (63) give

$$\left\|\Psi_n(d_j) - \Psi(d_j)\right\|_{2,n} \le O_{\mathbb{P}}\{\varepsilon^{j+1/2}(d_n + \zeta) + \varepsilon^{j+1} d_n\}. \tag{67}$$

Here are details of the sampled-site estimate. First replace the empirical scale in the normalized row by the population scale. Its relative error is $O_{\mathbb{P}}(d_n)$, and the row mean of the absolute difference of d_j is $O_{\mathbb{P}}(\varepsilon^{j+1/2})$, since $\operatorname{dist}(x,y)k_\varepsilon(x-y) \le C\sqrt{\varepsilon}k_{2\varepsilon}(x-y)$. This costs $O_{\mathbb{P}}(\varepsilon^{j+1/2}d_n)$. For the fixed-scale row numerator the summand is a bounded multiple of $k_\varepsilon(x-y)[d_j(y) - d_j(x)]$. At $x = X_i$ its diagonal is zero. Conditioning on $X_i = x$ gives bias $O(\varepsilon^{j+1}/n)$ and variance at most

$$\frac{C}{n}\varepsilon^{2j}\int k_\varepsilon(x-y)^2 \operatorname{dist}(x,y)^2 \, \mathrm{d}\mu(y) \le \frac{C}{n}\varepsilon^{2j+1-3/2}.$$

Averaging the conditional second moments over i therefore gives $O_{\mathbb{P}}(\varepsilon^{j+1/2}\zeta)$ in $L^2(\mu_n)$ without independence of the sampled queries. Denominator errors cost $O_{\mathbb{P}}(\varepsilon^{j+1}\rho)$, since the population row difference is $O(\varepsilon^{j+1})$. Exponentials of d_j have the same derivative bounds, proving (67) for the logarithmic map as stated.

Taylor expansion of Ψ_n at d_j now gives, for $u_j = \ell_j - d_j$,

$$u_{j+1} = B_n u_j + E_j, \qquad \left\|E_j\right\|_{2,n} \le O_{\mathbb{P}}(\varepsilon^j \left\|u_j\right\|_{2,n} + \left\|u_j\right\|_\infty \left\|u_j\right\|_{2,n} + \varepsilon^{j+1/2} d_n).$$

Sup nonexpansion and the corresponding coarser uniform row estimate give $\left\|u_j\right\|_\infty = O_{\mathbb{P}}(d_n)$ for these finitely many indices. As $\varepsilon^{3/2}d_n = o(\zeta)$, finite induction yields

$$\begin{aligned} \ell_4 = d_4 + B_n^3 u_1 + R_4, \qquad \|R_4\|_{2,n} = O_{\mathbb{P}}(\varepsilon\zeta + d_n\zeta + \varepsilon^{3/2}d_n), \\ \|\ell_4 - d_4\|_{2,n} = O_{\mathbb{P}}(\zeta), \qquad \|\ell_4 - d_4\|_\infty = O_{\mathbb{P}}(d_n). \end{aligned} \tag{68}$$

4.3 A uniform second-order trajectory expansion

In this subsection put $h = \varepsilon$, $\delta = d_n$, and $F = H_h f$. To avoid confusing the density convolution F with a scale error, write $\eta = \widehat{\nu} - \nu$, $\eta_1 = Rg$, $D = C_n - C$, where $C_n = \mu_n[a^2 k \otimes k]$. A symbol $p(j)$ denotes a fixed polynomial with nonnegative coefficients; its degree may be enlarged a finite number of times. All stochastic bounds below hold simultaneously through L_n, with a common bounded-in-probability envelope and, when indicated, a fixed additional power of $\log n$. This distinction matters: the leading linear variance tail will carry no such logarithmic loss.

Lemma 4.2 (Population filters)**.** *Let $m = a^2 f = 1 + O_{C^1}(h)$ and $T = I - H_h M_m$. Uniformly for $0 \le j \le C_0 \log n$ and $q = 0, 1$,*

$$\left\|T^j\right\|_{C^q \to C^q} \le C(j+1)^2.$$

The smoothed kernel $T^j H_h$ has absolute row mass at most $C(j+1)^2$ and derivative absolute row mass at most $Ch^{-1/2}(j+1)^2$.

Proof. For $m = 1$, $(I - H_h)^j$ is the identity minus a rescaled, periodized convolution whose Euclidean multiplier is $\chi_j(\xi) = 1-(1-e^{-|\xi|^2/2})^j$. Differentiating at most twice gives $\sum_{|\alpha|\le 2} \|\partial^\alpha \chi_j\|_2 \le C(j+1)^2$: the derivatives are polynomial factors in ξ times $je^{-t}(1-e^{-t})^{j-1}$ or $j(j-1)e^{-2t}(1-e^{-t})^{j-2}$, $t = |\xi|^2/2$; the two smallest indices are handled separately. Weighted Cauchy–Schwarz and Plancherel, using $(1 + |z|^2)^{-1} \in L^2(\mathbb{R}^3)$, bound the inverse Fourier transform's L^1 norm by the same polynomial. Periodization and rescaling do not increase that norm. Convolution commutes with one spatial derivative. For general m, Duhamel with $T_0 = I - H_h$ gives

$$T^j = T_0^j - \sum_{r<j} T_0^{j-1-r} H_h M_{m-1} T^r.$$

If $N_J = \max_{r\le J} \|T^r\|_{C^q\to C^q}/(r+1)^2$, then $N_J \le C + Ch(J+1)^3 N_J$. Absorb the last term because $h(\log n)^3 \to 0$. Applying the resulting C^0/C^1 bounds to H_h proves the row assertions, including the derivative cost $h^{-1/2}$. □

Lemma 4.3 (Linear influence classes)**.** *Set $\nu = \mu_n - \mu$ and*

$$U_1 = a\eta_1 - g_0/(2F), \quad U_r = B_0^{r-1}U_1, \quad Z = SU_1 = CBRg + W, \qquad W = Sh_c.$$

The centered deterministic influence kernels $\beta_r(u, y)$ of U_r have, with $e_r = p(r)2^{-(r-1)}$, row L^2 norm at most $Ce_r h^{-3/4}$, derivative-row L^2 norm at most $Ce_r h^{-5/4}$, supremum at most $Ce_r h^{-3/2}$, derivative supremum at most $Ce_r h^{-2}$, and integral $L^2(\mu)$ operator norm at most Ce_r. Moreover

$$\|U_r\|_\infty + \sqrt{h}\,\|\nabla U_r\|_\infty = O_{\mathbb{P}}(e_r\rho), \quad \|U_r\|_{2,n} = O_{\mathbb{P}}(e_r\zeta), \quad \|(P_n - P)U_r\|_{2,n} = O_{\mathbb{P}}(e_r\delta\zeta). \tag{69}$$

The same conservative kernel and field bounds hold for $B^r Rg$ and $B^r W$ in the corresponding RKHS representation. In addition,

$$\mathrm{Cov}(W) \le Chn^{-1}I, \quad \|h_c\|_{2,n} = O_{\mathbb{P}}(\sqrt{h}\zeta), \quad \|S_n h_c - W\|_{\mathcal{H}} = O_{\mathbb{P}}(\delta\zeta), \quad \mathrm{Cov}(Z) \le Cn^{-1}I_{\mathcal{H}}. \tag{70}$$

Proof. The kernels of η_1 and g_0/F have the asserted heat row bounds and bounded integral L^2 norms by Lemma 3.6 and boundedness of F^{-1}. Their centered kernels may contain a bounded rank-one term, which does not change any order. Apply Lemma 4.2, its row bounds and Minkowski to the population filter $B_0^{r-1} = M_a B^{r-1} M_{a^{-1}}$. Its integral $L^2(\mu)$ norm is $2^{-(r-1)}$, since P is PSD Markov. This proves all kernel bounds and the sup assertions from scalar concentration.

For the sample comparison, the explicit representation $U_r = t_r g_0 + aJ_r$ satisfies

$$t_r = -2^{-(r-1)}/(2F), \quad J_1 = Rg, \quad J_{r+1} = BJ_r - \tfrac{1}{2}D_{at_r,\mu}g_0,$$

where $D_{t,\mu} = \mu[tk \otimes k]$ is a deterministic weighted covariance. Thus $\|J_r\|_{\mathcal{H}} \le O_{\mathbb{P}}(p(r)2^{-r}\zeta)$. The fixed-scale row difference on U_r is a deterministic weighted covariance difference acting on J_r or g_0, hence has norm bounded by ρ times these RKHS norms. Changing either scale to $\widehat{a}$ costs δ times the sample L^2 norm. Bounded empirical feature maps prove (69). Smooth multiplication has been used in sample or physical L^2, without asserting Gaussian-RKHS closure.

For $h_c = \nu J$, the kernel $J(u, y) = k_h(u - y)[a(u)a(y) - 1/F(u)]/2$ has both absolute Schur bounds $O(\sqrt{h})$: its coefficient is $O(\mathrm{dist}(u, y) + h)$. This proves the covariance of W. The zero-diagonal kernel $k_h(u-y)[1-v(u)a(y)]$ for g_0-vg has conditional sampled variance $O(h\zeta^2)$ and conditional bias $O(h/n)$. Together with $a - v/F = O(h)$ this proves the sampled norm of h_c. Replacing $\widehat{a}$ in its lift costs $O_{\mathbb{P}}(\delta\zeta)$; the rest is

$$(D_{a^2/2,n} - D_{a^2/2,\mu})g - (D_{a/(2F),n} - D_{a/(2F),\mu})g_0,$$

also of this order. Its population lift has conservative heat-row bounds and bounded integral L^2 norm; Lemma 4.2 gives the filtered W class. Finally the influence operator of U_1 is bounded on $L^2(\mu)$. Applying S to its covariance gives $\mathrm{Cov}(Z) \le CSS^*/n \le CI/n$. □

Lemma 4.4 (Quadratic and cubic primitives)**.** *Let β, χ be centered influence kernels from the preceding class, with geometric factors e_β, e_χ. Suppose a deterministic left score row A satisfies*

$$\|A\|_1 \le e_A h^{-1/2}, \qquad \|A\|_2 \le e_A h^{-5/4}, \qquad \|A\|_\infty \le e_A h^{-2}.$$

In each bound below e is the product of the geometric factors of the fields and left row that occur in that expression. Quadratic kernels of the forms

$$A(u)\beta(u,y), \quad \int A(u)\beta(u,y)\chi(u,z)d\mu(u), \quad \partial_x[b(x)\beta(x,y)\chi(x,z)]$$

have joint L^2 norm $O(eh^{-2})$, diagonal L^2 norm $O(eh^{-11/4})$ and absolute diagonal mean $O(eh^{-2})$, with the factors on the local fields included in e in the last expression. Their normalized centered-measure statistics have order $O_{\mathbb{P}}(e(\zeta+\zeta^2))$. For the empirical numerator $A(u)\beta(u,y)\chi(u,z)$, center u. Its joint norm, pair-collision norms and repeated absolute mean have orders $eh^{-11/4}, eh^{-7/2}, eh^{-7/2}$; every nonzero first Hoeffding projection of a pair collision has norm $O(eh^{-11/4})$. The normalized cubic statistic has order $O_{\mathbb{P}}(e(\zeta^2+\zeta^3))$. These estimates allow different influence kernels, bounded smooth weights, and polynomially many logarithmic indices, with an additional fixed logarithmic factor for simultaneous bounds.

Proof. These are the estimates of Lemmas 3.7 and 3.8; the following products specify their application. For a numerator use $\|A\|_1$ times the two row L^2 norms; on its diagonal use one kernel supremum and one row norm. For a covariance insertion use $\|A\|_2$ times a row norm, and on the diagonal use its supremum. The local derivative uses one derivative row norm and one ordinary row norm, or a supremum times the complementary derivative row. The means are paid by the same L^1 or Cauchy–Schwarz products.

For a cubic collision $u = y$ or $u = z$, the first projection is a bounded integral-L^2 adjoint applied to A times the other kernel's diagonal. For $y = z$, it is $A(u)\mathbb{E}[\beta(u,Y)\chi(u,Y)]$ minus its mean. Both have norm $O(eh^{-11/4})$; the other projection vanishes by canonicity. The full pair norms are $O(eh^{-7/2})$, and the fully repeated absolute mean follows from $\|A\|_1$ times two suprema. The all-distinct, exactly-two-equal and all-equal partitions therefore give, after multiplication by $\sqrt{n}h^{5/4}$, respectively $e\zeta^2$, $e(\zeta^2+\zeta^3)$ and $e\zeta^3$. There is no assumption that a collapsed kernel remains canonical or has vanishing diagonal mean. Union bounds over polynomially many logarithmic indices cost only fixed logarithmic powers. □

Lemma 4.5 (Random features and powers on linear inputs)**.** *For $U_0 = Rg$ or W, and $0 \le j \le L_n$,*

$$\left\langle \widehat{\theta}_x, \widehat{B}^j U_0 \right\rangle - \left\langle \theta_x, B^j U_0 \right\rangle$$

is a sum of quadratic and cubic primitives of Lemma 4.4, with total geometric factor $p(j)2^{-j}$, and a remainder bounded by $O_{\mathbb{P}}(h^{-5/4}p(j)2^{-j}\delta^2\zeta)$. The same assertion holds with a fixed left C factor or any fixed finite combination of adjacent B powers on the input.

Proof. The exact reciprocal formula $\widehat{a} = a/(1+a\eta)$ gives

$$\begin{aligned}
&\widehat{C} - C = D + V_\mu + V_\nu + E,\\
&V_\mu = -2\mu[a^3\eta_1 k \otimes k], \qquad V_\nu = -2\nu[a^3\eta_1 k \otimes k], \qquad E = \mu_n[ck \otimes k],\\
&|c| \le C(|\eta - \eta_1| + \eta^2).
\end{aligned}$$

For a filtered input $U_r = B^r U_0$, the sample-to-feature bound gives

$$\|EU_r\|_{\mathcal{H}} \le C\{\|\eta - \eta_1\|_{2,n}\|U_r\|_\infty + \|\eta\|_\infty^2\|U_r\|_{2,n}\} = O_{\mathbb{P}}(p(r)2^{-r}\delta^2\zeta).$$

This is an applied remainder. In the first insertion paired with $B^l\theta_x$, D is the covariance primitive, V_μ the population product primitive, and V_ν the full empirical cubic primitive. Their left row is a bounded weight times $[B^l\theta_x](u)$. Lemma 4.2 and RKHS evaluation give its L^1, L^2, ∞ norms

$p(l)2^{-l}h^{-1/2}, C2^{-l}h^{-5/4}, C2^{-l}h^{-2}$. For the L^1 statement the symmetric smoothed kernel $B^l H_h$ converts the query derivative column to its derivative row. Resolvent and C factors are functions of the same population operator; their physical C^0/C^1 bounds preserve these orders.

Likewise

$$\widehat{\theta}_x = \theta_x - \partial_x[a(x)^2\eta_1(x)k_x] + \partial_x[d(x)k_x], \quad d = -a^2(\eta - \eta_1) + a^3\eta^2/(1 + a\eta).$$

The first variation paired with U_r is a local quadratic primitive. For its remainder, RKHS value and derivative evaluation on $\eta - \eta_1$ combine with the sup and derivative-sup bounds on U_r to give $h^{-5/4}p(r)2^{-r}\delta\rho\zeta$. The squared-scale part costs $h^{-1/2}p(r)2^{-r}\delta^2\rho$, which is smaller than the claimed bound, since $h^{3/4}\rho/\zeta \to 0$.

Finally let $\Delta B = \widehat{B} - B$. Twice telescoping gives the exact finite identity

$$\widehat{B}^j = B^j + \sum_{l+r=j-1} B^l \Delta B B^r + \sum_{p+q+r=j-2} \widehat{B}^p \Delta B B^q \Delta B B^r.$$

Both powers have norm at most one half; the last sum on U_0 has norm $O_{\mathbb{P}}((j+1)^2 2^{-j}\delta^2\zeta)$. Pairing the first power difference with the feature error costs the same normalized order. Lemma 4.4 handles every retained first insertion. A left C factor is a finite difference of adjacent powers, since $C = I - 2B$. □

Lemma 4.6 (Second-order seed from the actual first update)**.** *Let*

$$e_0 = g_0/F, \qquad \eta_2 = R\{K_n[a^3\eta_1^2] - D\eta_1\}, \qquad U_2 = a\eta_2 - a^2\eta_1^2/2 + e_0^2/4.$$

Then, for $u = \ell_1 - d_1$,

$$u = U_1 + U_2 + O_{L_n^2,\mathbb{P}}(\delta^2\zeta), \quad \|u\|_\infty = O_{\mathbb{P}}(\delta), \quad \|u\|_{2,n} = O_{\mathbb{P}}(\zeta), \quad \|U_2\|_{2,n} = O_{\mathbb{P}}(\delta\zeta).$$

In addition

$$S_n u = Z + H_2 + O_{\mathcal{H},\mathbb{P}}(\delta^2\zeta), \tag{71}$$

where

$$H_2 = \nu[aU_1] - K_n[a^2\eta_1 U_1] + (CR - I/2)K_n[a^3\eta_1^2] - CRD\eta_1 + \tfrac{1}{4}K_n[ae_0^2],$$
$$\|H_2\|_{\mathcal{H}} = O_{\mathbb{P}}(\delta\zeta).$$

Proof. Since $b_1 = (K_n 1)^{-1/2}$, exactly

$$\ell_1 = d_1 + \log(1 + a\eta) - \tfrac{1}{2}\log(1 + e_0).$$

Lemma 3.9 gives $\eta = \eta_1 + \eta_2 + O_{\mathcal{H},\mathbb{P}}(\delta^2\zeta)$. Expanding the two scalar logarithms gives the stated seed; the quadratic replacement costs $\delta\|\eta - \eta_1\|_{2,n}$ and cubic terms use two sup factors and one L^2 factor. For the lift, $S_n U_1 = K_n[aU_1] - K_n[a^2\eta U_1/(1 + a\eta)]$. Replacing $\eta/(1 + a\eta)$ by η_1 costs at most $C\|\eta - \eta_1\|_{2,n}\|U_1\|_\infty + C\delta^2\|U_1\|_{2,n}$. In $S_n U_2$ replace $\widehat{a}$ by a; its error is $\delta\|U_2\|_{2,n}$. The term containing η_2 is $C_n\eta_2$; replacing C_n by C costs $\rho\|\eta_2\|_{\mathcal{H}}$. All these are $O_{\mathbb{P}}(\delta^2\zeta)$, giving the exact coefficients in (71). Its centered empirical linear lift is a deterministic weighted covariance difference acting on η_1 or g_0. Every other product has bounded lift and sup-times-L^2 size $O_{\mathbb{P}}(\rho\zeta)$. This proves the coarse norm of H_2 while retaining all empirical cubic pieces. No sup bound of order δ^2 on U_2 or $\eta - \eta_1$ is required. □

Lemma 4.7 (Smooth background and nonlinear lift)**.** *There is a centered linear field $L_j(x) = 2\langle\theta_x, B^{j-1}Z\rangle$ with $\mathbb{E}[A^2|L_j(x)|^2] \le C4^{-j}$ such that the full actual trajectory satisfies, simultaneously for $4 \le j \le L_n$,*

$$A|s_j(x) - \widehat{s}(x)| \le A|L_j(x)| + \mathcal{E}_n + K_n(\log n)^C p(j)2^{-j}\{\zeta + \zeta^2 + \zeta^3 + \sqrt{n}\delta^2\zeta\}, \tag{72}$$

where $K_n = O_{\mathbb{P}}(1)$ and

$$\mathcal{E}_n = O_{\mathbb{P}}\big((\log n)^C\{\sqrt{n}h^{19/4} + \sqrt{n}h\delta + \sqrt{n}h^{3/2}\delta\}\big). \tag{73}$$

In particular $\mathcal{E}_n = o_{\mathbb{P}}(1)$ for $\kappa > 2/19$.

Proof. Define, for analysis only, $p_j = \Psi_n^{j-1}(d_1)$ and $z_j = \Psi_n^{j-1}(u)$. Their sup bounds are $O(h)$ and $O_{\mathbb{P}}(\delta)$. The derivative estimate (61) gives their L^2 bounds $O(h2^{-j})$ and $O_{\mathbb{P}}(\zeta 2^{-j})$, because $\exp(C(h+\delta)L_n)$ is bounded. The mixed Taylor difference is at most $C(\|p\|_\infty\|z\|_2+\|z\|_\infty\|p\|_2)$. Its finite Duhamel recurrence proves

$$\ell_j = p_j + z_j + E_{b,j}, \quad \|E_{b,j}\|_{2,n} = O_{\mathbb{P}}(p(j)h\delta 2^{-j}), \quad \|E_{b,j}\|_\infty = O_{\mathbb{P}}(\delta).$$

The last assertion uses sup nonexpansion of the difference $\ell_j - p_j$ and of z_j, rather than converting an L^2 estimate into a sup bound. By (67), $\|p_4 - d_4\|_{2,n} = O_{\mathbb{P}}(h^{3/2}\delta)$. For $j \geq 4$, compare p_j to $\widetilde{p}_j = \Psi_n^{j-4}(d_4)$; their norm difference is $O_{\mathbb{P}}(p(j)h^{3/2}\delta 2^{-j})$, while $\|\widetilde{p}_j\|_\infty = O(h^4)$. The row bounds (58)–(59), with logarithm denominators bounded away from zero, give the safe background coefficient (73). This uses $\sqrt{n}h\delta = O(h^{1/4}(\log n)^{3/2})$ and $\sqrt{n}h^{3/2}\delta = O(h^{3/4}(\log n)^{3/2})$.

For the zero-background trajectory the exact quadratic update is

$$\Psi_n(z) = B_n z + N_2(z) + E_3(z), \quad N_2(z) = -\{P_n z^2 - (P_n z)^2\}/4, \quad \|E_3(z)\|_{2,n} \leq C\|z\|_\infty^2\|z\|_{2,n}.$$

Put $V_j = B_n^{j-1}U_1$. Its sup norm is $O_{\mathbb{P}}(\rho)$ since $\|B_n\|_{\infty\to\infty} \leq 1$, and its L^2 norm is $O_{\mathbb{P}}(2^{-j}\zeta)$. First Duhamel gives $\|z_j - V_j\|_{2,n} = O_{\mathbb{P}}(p(j)\delta\zeta 2^{-j})$. The Lipschitz constant of N_2 between these fields is $O_{\mathbb{P}}(\delta)$. Second Duhamel therefore yields the simultaneous expansion

$$z_j = B_n^{j-1}(U_1 + U_2) + \sum_{r=1}^{j-1} B_n^{j-1-r} N_2(V_r) + O_{L_n^2,\mathbb{P}}(p(j)\delta^2\zeta 2^{-j}). \tag{74}$$

The U_r of Lemma 4.3 satisfy $\|V_r - U_r\|_{2,n} = O_{\mathbb{P}}(p(r)2^{-r}\delta\zeta)$ by telescoping $(P_n - P)U_r$. Their sup bounds suffice to replace V_r in N_2 with error $O_{\mathbb{P}}(p(r)2^{-r}\delta^2\zeta)$. The exact identities $S_nP_n = \widehat{C}S_n$, $S_nB_n = \widehat{B}S_n$ then give

$$\begin{aligned} S_nN_2(U_r) &= Q_r + O_{\mathcal{H},\mathbb{P}}(p(r)2^{-r}\delta^2\zeta), \\ Q_r &= -\tfrac{1}{4}CK_n[aU_r^2] + \tfrac{1}{4}K_n[a(PU_r)^2]. \end{aligned}$$

In the first term replacing $\widehat{C}$ by C and $\widehat{a}$ by a costs $\delta\|U_r^2\|_{2,n}$. In the second, replacing P_nU_r by PU_r costs their sup bounds times their controlled L^2 difference. Each error is bounded by the claimed cubic remainder. In particular no empirical square numerator is replaced by its expectation.

Lift (74) and use (71). The quadratic fields H_2, Q_r have coarse norm $O_{\mathbb{P}}(\delta\zeta)$ with their input decay. Replacing their outer $\widehat{B}$ powers by B costs another δ, since both are half contractions. Thus

$$S_nz_j = \widehat{B}^{j-1}Z + B^{j-1}H_2 + \sum_{r=1}^{j-1} B^{j-1-r}Q_r + O_{\mathcal{H},\mathbb{P}}(p(j)\delta^2\zeta 2^{-j}). \tag{75}$$

For $G(z) = 2\,\nabla\log(P_ne^z)$ the off-sample quadratic expansion is

$$G(z) = 2\left\langle \widehat{\theta}_x, S_nz \right\rangle + \nabla\{P_nz^2 - (P_nz)^2\} + O_{\mathbb{P}}(h^{-5/4}\|z\|_\infty^2\|z\|_{2,n}).$$

Both ordinary and derivative row norms in (58) are used; the denominator cross term costs $h^{-1/2}h^{-3/4} = h^{-5/4}$ too. Replacing its quadratic argument z_j by U_j leaves an error $O_{\mathbb{P}}(h^{-5/4}p(j)2^{-j}\delta^2\zeta)$ and the explicit observable

$$\nabla\{aK_n(aU_j^2) - (PU_j)^2\}. \tag{76}$$

For the derivative of the squared-row replacement, decompose $P_nU_j - PU_j$ into the deterministic weighted covariance differences and two scaling differences used in Lemma 4.3. Its value and derivative cost respectively $h^{-3/4}$ and $h^{-5/4}$ times $p(j)2^{-j}\delta\zeta$. Multiply by the other field's value or derivative sup $p(j)2^{-j}\rho$ or $h^{-1/2}p(j)2^{-j}\rho$. This proves the asserted remainder also for the nonlinear denominator.

Pair (75) with the feature. On its quadratic fields the feature error costs $h^{-5/4}\delta$ times their $\delta\zeta$ norm. On the linear term apply Lemma 4.5 to $Z = CBRg + W$. The remaining explicit statistics are exactly

the primitives of Lemma 4.4: $\nu[aU_1]$ and $CRD\eta_1$ are covariance insertions; each K_n product in H_2, Q_r and (76) splits into its population quadratic and empirical cubic numerator; the square of PU_j is a local derivative primitive. All their left rows come from B^l, B^lC, B^lCR, $B^l(CR - I/2)$ followed by a heat numerator. Their row bounds follow from Lemma 4.2, the physical resolvent bound and RKHS evaluation. This accounts for every random scaling, quadratic and repeated-observation term. The geometric factors sum to $p(j)2^{-j}$: in the nonlinear sum an outer $2^{-(j-1-r)}$ multiplies an input-square factor $4^{-(r-1)}$.

Finally (70) gives $\mathrm{Cov}(Z) \le CI/n$, so $\mathbb{E}[A^2|2\langle\theta_x, B^{j-1}Z\rangle|^2] \le C4^{-j}$. Combining these estimates with the background bound proves (72) on one common trajectory event. □

Proposition 4.8 (Uniform numerical equivalence)**.** *For $2/19 < \kappa < 4/9$ and every deterministic integer $J_n \to \infty$, $J_n \le L_n$,*

$$\sup_{J_n \le j \le L_n} A|s_j(x) - \widehat{s}(x)| \longrightarrow 0 \quad \textit{in probability.}$$

Proof. The linear variance tail sums to $C4^{-J_n}$. Its bound has no external logarithmic factor and therefore permits an arbitrarily slow J_n. The background vanishes for $\kappa > 2/19$. Since $\sup_j p(j)2^{-j} < \infty$, all other terms in (72) vanish uniformly: $\zeta \to 0$ with a polynomial margin and

$$\sqrt{n}\delta^2\zeta = O(n^{-1}h^{-9/4}(\log n)^3) = o(1) \qquad (\kappa < 4/9).$$

Every fixed additional logarithmic factor is absorbed by the strict power margins. No tolerance exponent is needed for this assertion. □

4.4 Residual proxy and simultaneous spectral concentration

Let

$$Q_j = \tfrac{1}{4}\langle g, CB^{2j}g\rangle_{\mathcal{H}}, \qquad q_j = \sqrt{Q_j}. \tag{77}$$

Lemma 4.9 (Simultaneous residual comparison)**.** *Set $D_h = \varepsilon^4 + \zeta$, $\sigma_h = \varepsilon^4 + d_n$, $H_n = \sqrt{n}d_n\zeta$, and*

$$R_n = \sqrt{n}\varepsilon^4 + \varepsilon^{-1/4} + 1 + H_n + \sqrt{n}\sigma_h D_h + \sqrt{n}\zeta^2 + \sqrt{d_n}\varepsilon^{-3/4}.$$

There is a common $K_n = O_{\mathbb{P}}(1)$ such that

$$\sqrt{n}|r_j - q_j| \le K_n(j+1)2^{-j}R_n, \qquad 4 \le j \le L_n. \tag{78}$$

For every $2/19 < \kappa < 2/3$, $\varepsilon^{3/4}R_n \to 0$ with a strict polynomial margin, including any fixed additional logarithmic factor.

Proof. The first-order estimates (68), propagated using the exact half contraction and (61), give $\|\ell_j\|_{2,n} = O_{\mathbb{P}}(D_h2^{-j})$, $\|\ell_j\|_\infty = O_{\mathbb{P}}(\sigma_h)$. Their Duhamel error is bounded by $O_{\mathbb{P}}((j+1)\sigma_h D_h 2^{-j})$; the same estimate includes the nonlinear projective residual error. The leading feature is $S_n^*B^jRg$, and

$$\frac{I+P_n}{2}S_n^*B^jRg = \tfrac{1}{2}S_n^*B^jg + \tfrac{1}{2}S_n^*(\widehat{C} - C)B^jRg.$$

The second term costs $O_{\mathbb{P}}(d_n\zeta 2^{-j})$. The population norm squared of the first term is Q_j. Its empirical-to-population norm replacement costs $O_{\mathbb{P}}(\sqrt{d_n}\zeta 2^{-j})$ by the square-root operator comparison. The seven displayed coefficients in R_n pay, respectively, the initial smooth background, sampled commutator, smaller finite-row terms, initial and empirical-power errors, nonlinear trajectory and residual errors, actual sample centering, and the norm replacement. The commutator estimates are proved in Lemma 4.3. For centering, the exact identity is

$$\langle 1, S_n^*B^jg\rangle_n = \langle \widehat{\nu} - \nu, B^jg\rangle_{\mathcal{H}};$$

it follows from $S_n1 = \widehat{\nu}$ and $\langle \nu, g\rangle = 0$, and has size $O_{\mathbb{P}}(\zeta^2 2^{-j})$. Thus no potentially large root-n coefficient has been discarded.

After multiplication by $\varepsilon^{3/4}$ the seven terms reduce to $\sqrt{n}\varepsilon^{19/4}, \varepsilon^{1/2}, \varepsilon^{3/4}, d_n$,

$$\sqrt{n}\varepsilon^{35/4} + \varepsilon^4 + \sqrt{n}\varepsilon^{19/4} d_n + d_n, \qquad \zeta, \qquad \sqrt{d_n}.$$

All vanish with strict polynomial margins on the stated domain. □

Theorem 4.10 (Geometry independently of score inference)**.** *For every* $2/19 < \kappa < 2/3$ *and* $0 < \gamma < 3/4$*, genuine crossing,* (50)*, and* (51) *hold for the unchanged rule.*

Proof of Theorems 4.10 and 4.1. The nonzero spectrum of C is that of $M_q H_\varepsilon M_q$ on $L^2(\,\mathrm{d}x)$, where $q = a\sqrt{f} = 1 + O(\varepsilon)$. Their operator difference from H_ε is $O(\varepsilon)$. The min–max principle and the heat eigenvalues $e^{-2\pi^2\varepsilon|k|^2}$ imply that every fixed positive-width interior spectral interval in $(0,1)$ contains at least $c\varepsilon^{-3/2}$ eigenvalues of C, for sufficiently small ε. Indeed choose a smaller heat interval inside it; the eigenvalue perturbation is smaller than the gap, and lattice counting gives the stated number. Also $\operatorname{tr} C \le C\varepsilon^{-3/2}$.

Here is the concentration calculation needed for a growing number of iterations. Put $Y = a(X)k_X - v$ and $Z_j = \frac{1}{2}C^{1/2}B^j Y$, $j \ge 1$. By (64), its covariance is $\Sigma_j = \frac{1}{4}C^2 B^{2j}$, and $Q_j = \left\| n^{-1}\sum_i Z_{j,i} \right\|^2$. For a centered real Hilbert vector with covariance Σ, expansion of the four-index sum yields

$$\operatorname{Var}\left\| \frac{1}{n}\sum_i Z_i \right\|^2 = \frac{\mathbb{E}\,\|Z\|^4 - (\operatorname{tr}\Sigma)^2}{n^3} + \frac{2(n-1)\operatorname{tr}(\Sigma^2)}{n^3}. \tag{79}$$

If $T_j = \operatorname{tr}(C^2 B^{2j})$, then $\mathbb{E}Q_j = T_j/(4n)$ and $\left\|Z_j\right\|^2 \le C\varepsilon^{-3/2}4^{-j}$. Thus

$$\frac{\operatorname{Var} Q_j}{(\mathbb{E}Q_j)^2} \le C\frac{\max_\lambda \lambda^2((1-\lambda)/2)^{2j}}{T_j} + \frac{C\varepsilon^{-3/2}4^{-j}}{nT_j}. \tag{80}$$

The same formula holds after any fixed spectral projection. It retains the fourth-moment diagonal contribution rather than discarding it.

Put $u = 3/4 - \gamma > 0$. Choose a small $\chi > 0$ and set

$$j_- = \left\lfloor \frac{(u-\chi)\log(1/\varepsilon)}{\log 2} \right\rfloor, \qquad j_+ = \left\lceil \frac{(u+\chi)\log(1/\varepsilon)}{\log 2} \right\rceil. \tag{81}$$

Take $\chi < u$ and then a fixed $\beta_* < 1/2$ sufficiently close to $1/2$. An interior band with $\beta = (1-\lambda)/2 \ge \beta_*$ supplies $T_j \ge c\varepsilon^{-3/2}\beta_*^{2j}$. Equation (80) is then bounded by

$$C(\varepsilon^{3/2} + 1/n)(1/(4\beta_*^2))^j.$$

By choosing β_* close enough to $1/2$, its sum over $4 \le j \le j_-$ tends to zero. Chebyshev's inequality and a union bound therefore give, simultaneously,

$$\sqrt{n}q_j \ge c\varepsilon^{-3/4}\beta_*^j, \qquad 4 \le j \le j_-. \tag{82}$$

The choice can also ensure that its value at j_- dominates $\varepsilon^{-\gamma}$: if $\sigma = -\log\beta_*/\log 2$, it suffices that $(u-\chi)\sigma < u$. Both requirements are open inequalities at $\beta_* = 1/2$.

The first crossing, its residual, and inference. Divide (78) by (82). The relative error is at most a common bounded random factor times

$$(j+1)R_n\varepsilon^{3/4}(1/(2\beta_*))^j.$$

The product $(j+1)R_n\varepsilon^{3/4}$ tends to zero with a strict polynomial margin; choose β_* still closer to $1/2$ to absorb the arbitrarily mild last factor, uniformly up to j_-. Thus no index from 4 through j_- hits. At

j_+, $\mathbb{E}[nQ_{j_+}] \le C\varepsilon^{-3/2}4^{-j_+}$, so $q_{j_+}/t_{n,\varepsilon} \to 0$ by Markov's inequality. The error in (78), divided by the threshold, tends to zero there as well. Since $j_+ < L_n$ eventually, a genuine hit occurs by j_+.

It remains essential to exclude indices zero through three. At fixed $j = 4$, (82) gives $r_4/t_{n,\varepsilon} \to \infty$. The relative logs at indices one, two and three have sup norm $O_{\mathbb{P}}(\varepsilon + d_n) = o_{\mathbb{P}}(1)$. Applying (62) backwards excludes each of those actual checks. If f is nonconstant,

$$r_0 \longrightarrow \frac{\mathrm{sd}_\mu(\sqrt{f})}{\int \sqrt{f}\,\mathrm{d}\mu} > 0,$$

because $K_n 1 - H_\varepsilon f$ tends to zero uniformly, $H_\varepsilon f \to f$, and the bounded population moments obey the law of large numbers. The threshold tends to zero since $\kappa\gamma < 1/2$. If $f = 1$, $\ell_0 = -\log \widehat{a} = o_{\mathbb{P}}(1)$ in sup norm, and (62) instead gives $r_1 \le Cr_0$; the lower bound already obtained excludes index zero in this case too. This proves the first-hit bracket for the rule that really checks every index. Letting $\chi \downarrow 0$ proves (50).

For the attained residual, Q_j is a positive spectral sum with factors β^{2j}, $0 \le \beta \le 1/2$. Cauchy–Schwarz shows that q_{j+1}/q_j is nondecreasing and at most $1/2$ whenever $q_j > 0$. We claim that this ratio tends to $1/2$, uniformly for $j \ge j_-$. Fix $\beta_L < 1/2$ and choose $\beta_L < \beta_H < \beta_{H2} < 1/2$. At $J = j_-$ the low-band expectation is at most $C\varepsilon^{-3/2}\beta_L^{2J}/n$, whereas a fixed high band $[\beta_H, \beta_{H2}]$ has expectation at least $c\varepsilon^{-3/2}\beta_H^{2J}/n$. The projected vector in (79) has squared envelope $C\varepsilon^{-3/2}\beta_{H2}^{2J}$. Its relative variance is bounded by

$$C(\varepsilon^{3/2} + 1/n)(\beta_{H2}/\beta_H)^{2J}.$$

Choose the band's positive width sufficiently small that this tends to zero. Markov's inequality for the low modes, and concentration for the high modes, show that the low contribution is $o_{\mathbb{P}}(Q_J)$. Hence $q_{J+1}/q_J \ge \beta_L - o_{\mathbb{P}}(1)$. Monotonicity and then $\beta_L \uparrow 1/2$ prove the claim.

For all $j \ge j_-$, (78) divided by the threshold is $o_{\mathbb{P}}(1)$, provided χ was chosen sufficiently small. Its largest coefficient is $O_{\mathbb{P}}(R_n\varepsilon^{3/4-\chi}\log n)$, which vanishes for a sufficiently small fixed $\chi > 0$ by the strict margin of the residual comparison. At the actual crossing, $r_{\tau-1} > t_{n,\varepsilon}$ and $r_\tau \le t_{n,\varepsilon}$. The ratio result thus gives $r_\tau \ge (1/2 - o_{\mathbb{P}}(1))t_{n,\varepsilon} - o_{\mathbb{P}}(t_{n,\varepsilon})$, which proves (51). A rounding phase may prevent a single limit for $r_\tau/t_{n,\varepsilon}$.

This proves Theorem 4.10 and the geometric assertions of Theorem 4.1. Finally, Proposition 4.8 applies on the common trajectory event after every deterministic diverging lower index. Choose one below the logarithmic lower hit bracket to prove (52) at the actual random hit. Constant multiples of ε preserve each strict margin. Intersect the two bandwidth events and apply the joint exact-score law of Theorem 3.11 and Slutsky's theorem to obtain (53). No independence of scores, stopping times or sample-dependent matrices is used.

□

5 Sharp first-hit location and lattice overshoot

We prove Theorem 2.4 on its geometric domain $2/19 < \kappa < 2/3$, $0 < \gamma < 3/4$. This section uses the actual-residual comparison and first-hit localization, not a score central limit theorem. Put $L = \log(1/h)$ and fix $0 < a < b < \infty$. All trace estimates below are uniform for integers $aL \le j \le bL$. They concern the population covariance C, $B = (I - C)/2$ and the scalar residual proxy $q_j^2 = \langle g, CB^{2j}g\rangle_{\mathcal{H}}/4$ from Section 4.

5.1 A density-independent spectral constant

Lemma 5.1 (Moving-window heat trace)**.** *Uniformly on the stated logarithmic window,*

$$\mathrm{tr}(C^2B^{2j}) = (1 + o(1))\,\frac{h^{-3/2}4^{-j}\sqrt{\log(2j)}}{4\sqrt{2}\pi^2 j^2}. \tag{83}$$

The leading constant is independent of the fixed density f.

Proof. The nonzero eigenvalues of C equal those of $M_q H_h M_q$, where $q = a_h\sqrt{f} = 1 + O_{C^2}(h)$. They also equal those of $H_h^{1/2} M_{q^2} H_h^{1/2}$. The min–max principle therefore gives, for the decreasing eigenvalues λ_i of C and μ_i of H_h,

$$(1 - Ch)\mu_i \le \lambda_i \le (1 + Ch)\mu_i.$$

Both spectra lie in $[0, 1]$; the former assertion follows from the Markov conjugacy. We do not apply operator monotonicity to the nonmonotone filter $W_j(z) = z^2(1 - z)^{2j}$. For $\mu_i \le 1/2$, the eigenvalue comparison gives $W_j(\lambda_i)/W_j(\mu_i) = \exp(O(jh) + O(h)) = 1 + o(1)$. There are $O(h^{-3/2})$ remaining heat eigenvalues. Their combined contribution for either operator is at most $Ch^{-3/2}(1/2 + Ch)^{2j}$, negligible relative to $h^{-3/2} j^{-2}\sqrt{\log(2j)}$. Thus it suffices to calculate the heat trace relatively on the moving window.

Set $F_j(s) = e^{-2s}(1 - e^{-s})^{2j}$. Unit-cube lattice counting yields

$$N_h(s) = \#\{k \in \mathbb{Z}^3 : 2\pi^2 h|k|^2 \le s\} = \frac{\sqrt{2}}{3\pi^2} h^{-3/2} s^{3/2} + O(1 + h^{-1}s).$$

The error is also bounded by $Ch^{-1}(1 + s)$ for $h \le 1$. The nonnegative function F_j is unimodal, with maximum at $s = \log(j + 1)$ of height $O(j^{-2})$, and

$$\int_0^\infty F_j(s)\, ds = \frac{1}{(2j + 1)(2j + 2)}.$$

Splitting its weighted variation at the maximum and integrating by parts gives $\int (1 + s)|F_j'(s)|\, ds \le C j^{-2} \log(2j)$. Stieltjes integration against N_h consequently gives

$$\sum_k F_j(2\pi^2 h|k|^2) = \frac{h^{-3/2}}{\sqrt{2}\pi^2} \int_0^\infty s^{1/2} F_j(s)\, ds + O\Big(h^{-1} j^{-2} \log(2j)\Big). \tag{84}$$

There are no endpoint boundary terms, since $j \ge 1$.

The substitutions $z = e^{-s}$ and $y = 2jz$ show that

$$\int_0^\infty s^{1/2} F_j(s)\, ds = \int_0^1 z(1 - z)^{2j} \sqrt{-\log z}\, dz \sim (2j)^{-2}\sqrt{\log(2j)}. \tag{85}$$

Indeed the last expression divided by its asserted equivalent is

$$\int_0^{2j} y(1 - y/(2j))^{2j} \left(\frac{\log(2j/y)}{\log(2j)}\right)^{1/2} dy.$$

For $y \ge 1$ it is dominated by ye^{-y}, and for $0 < y < 1$ by a constant times $ye^{-y}(1 + \sqrt{|\log y|})$. Dominated convergence gives the limit $\int_0^\infty ye^{-y} dy = 1$. This scalar asymptotic is uniform when $j \in [aL, bL]$ since the smallest such j diverges. The error in (84) divided by the main term is $O(\sqrt{h \log(2j)}) = o(1)$. Multiplication by 4^{-j} from B^{2j} proves (83). The density independence comes from $q^2 = 1 + O(h)$, not from a uniform sampling assumption. □

5.2 Relative concentration and the actual residual

Lemma 5.2 (Uniform sharp residual profile)**.** *For $K = 1/(4\,2^{1/4}\pi)$ and a window $[aL, bL]$ contained in $[4, L_n]$, uniformly for its integer indices,*

$$r_{j,h} = (1 + o_{\mathbb{P}}(1))Kn^{-1/2} h^{-3/4} 2^{-j} j^{-1} [\log(2j)]^{1/4}. \tag{86}$$

Proof. For the centered feature $Y = a(X)k_X - v$ put $Z_j = C^{1/2} B^j Y/2$. Since $Cv = v$ and $B^j v = 0$ for $j \ge 1$,

$$\Sigma_j = \mathrm{Cov}(Z_j) = C^2 B^{2j}/4, \qquad \mathbb{E}q_j^2 = \mathrm{tr}(\Sigma_j)/n.$$

The exact iid Hilbert-space fourth-moment identity (79) applies without independent Fourier coordinates or a Gaussian assumption. Feature evaluation gives $\sup_y \|Z_j(y)\|^2 \le C4^{-j}h^{-3/2}$. Maximizing $z^2(1-z)^{2j}$ on $[0,1]$ gives $\|\Sigma_j\|_{\mathrm{op}} \le C4^{-j}j^{-2}$. Use

$$\mathbb{E}\|Z_j\|^4 \le \sup_y \|Z_j(y)\|^2 \operatorname{tr}\Sigma_j, \qquad \operatorname{tr}\Sigma_j^2 \le \|\Sigma_j\|_{\mathrm{op}} \operatorname{tr}\Sigma_j$$

and Lemma 5.1. The resulting relative variance is bounded by

$$C\left\{\frac{j^2}{n\sqrt{\log(2j)}} + \frac{h^{3/2}}{\sqrt{\log(2j)}}\right\}.$$

Its sum over $O(L)$ indices tends to zero. Chebyshev and a union bound therefore give uniform relative concentration of q_j about $\sqrt{\mathbb{E}q_j^2}$. Taking the square root of (83), including the factor $1/(4n)$, gives the constant K and the profile in (86) for q_j.

To replace the proxy by the actual residual use (78), whose error is bounded by $O_{\mathbb{P}}(1)(j+1)2^{-j}R_n/\sqrt{n}$, with $h^{3/4}R_n \to 0$ at a strict polynomial rate up to logarithms on the geometric domain. Dividing by the sharper profile costs at most $CL^2/(\log L)^{1/4}$. The polynomial margin absorbs this factor, proving (86) on a common event for every index of the window. This step is essential for an actual, sample-dependent hit. □

5.3 Inverting the clock and resolving its lattice phase

Proof of Theorem 2.4. Write $u = 3/4 - \gamma > 0$ and

$$\phi_h(t) = Kh^{-u}2^{-t}t^{-1}[\log(2t)]^{1/4}.$$

For $t > 2$ its logarithmic derivative is $-\log 2 - 1/t + 1/(4t\log(2t)) < 0$ and stays away from zero. For small h, $\phi_h(2) > 1$ and $\phi_h(t) \to 0$ as $t \to \infty$, so there is a unique root $t_h > 2$ of $\phi_h(t_h) = 1$. Taking logarithms first gives $t_h \sim uL/\log 2$, and then

$$t_h \log 2 = uL - \log t_h + \tfrac{1}{4}\log\log(2t_h) + \log K$$

gives (6), including its $o(1)$ error.

Choose $a < u/\log 2 < b$ so the window lies within the original budget. The earlier genuine first-hit localization already puts τ_h in this window with high probability, excluding every earlier actual check. Lemma 5.2 gives $r_{j,h}/t_{n,h} = \phi_h(j)(1 + o_{\mathbb{P}}(1))$ uniformly there. For each fixed $\eta > 0$, every integer $j < t_h - \eta$ therefore has $r_{j,h} > t_{n,h}$, while $j = \lceil t_h + \eta \rceil$ has $r_{j,h} < t_{n,h}$, with probability tending to one. This proves the two-ceiling sandwich (7). A deterministic $\eta_n \downarrow 0$ can be chosen sufficiently slowly, giving $\operatorname{dist}(\tau_h - t_h, [0,1]) \to 0$ in probability. The slowly varying factors have ratio $1 + o_{\mathbb{P}}(1)$ at these bounded offsets, which proves (8).

If $\{t_h\} \to \theta \in (0,1)$, choose η smaller than both distances to the neighboring integers. The two ceilings agree eventually, so $\tau_h = \lceil t_h \rceil$ with high probability and the residual ratio tends to $2^{\theta-1}$. At integer phase this argument does not select a side, which explains the qualification in the statement.

For pure powers $h = n^{-\kappa}$, regard n temporarily as real. Implicit differentiation shows that $t(n)$ is eventually increasing, unbounded and satisfies

$$t'(n) = \frac{\kappa u}{n\{\log 2 + 1/t - 1/(4t\log(2t))\}} = O(1/n).$$

For any fixed $\theta \in (0,1)$ take the first integer n_m with $t(n_m) \ge m + \theta$. Its overshoot is at most $t(n_m) - t(n_m - 1) \to 0$. Thus every interior phase occurs along an integer-sample-size subsequence, with no equidistribution assumption. Phases arbitrarily close to zero rule out any universal lower constant greater than $1/2$; phases close to one rule out any upper constant less than 1. Two distinct interior phases give different constant probability limits, ruling out a full-sequence probability limit. Fixed bandwidth multiples, in particular $h = 2\varepsilon$, preserve every strict polynomial margin. The two-band assertions need only finite intersection of their events, not independent stopping times. □

6 Endpoint obstructions and the remaining bandwidth question

6.1 The cubic coefficient and a necessary lower endpoint

Let $r = \sqrt{f}$. The population equation is $q_\varepsilon H_\varepsilon(q_\varepsilon r) = r$. Its expansion, proved in Lemma 3.2, is

$$\log q_\varepsilon = \varepsilon u + \varepsilon^3 c + o_{C^1}(\varepsilon^3), \qquad u = -\frac{\Delta r}{4r}, \quad c = \frac{\Delta^3 r}{192r} - \frac{\Delta[(\Delta r)^2/r]}{128r} + \frac{(\Delta r)^3}{384r^3}. \tag{87}$$

The coefficient needs only $r \in C^7$: the highest heat Taylor term has a C^1 Peano remainder, and the term involving c needs only strong continuity in C^1. In particular it is available for the present C^8 class. Since $s_{\mathrm{pop},\varepsilon} = \nabla \log f - 2\nabla \log q_\varepsilon$,

$$2s_{\mathrm{pop},\varepsilon} - s_{\mathrm{pop},2\varepsilon} - \nabla \log f = \varepsilon^3 b + o(\varepsilon^3), \qquad b = 12\nabla c. \tag{88}$$

The positive factor 12 is $-4 + 16$; it is not the uncorrected cubic coefficient.

Proposition 6.1. *Use the same first-hit rule and correction as Theorem 4.1. There is a single positive analytic density and a deterministic query for which, for every fixed $0 < \gamma < 3/4$, the following hold. At $\kappa = 2/17$ the normalized true-score statistic has a nonzero Gaussian mean. For every fixed $0 < \kappa < 2/17$ its first coordinate diverges to positive infinity in probability. Thus $\kappa > 2/17$ is necessary for universal zero-centered inference of the specified form.*

Proof. Choose

$$f(x) = \frac{32}{33}\left[1 + \frac{1}{4}\cos(2\pi x_1)\right]^2, \qquad x_* = (1/4, 0, 0). \tag{89}$$

Its integral is one, its minimum is $6/11$, and it is analytic and independent of x_2, x_3. Such independence violates no dimension-three hypothesis. Multiplication of r by a constant changes neither u nor c, so put $r_0 = 1 + a\cos y$, $y = 2\pi x_1$, and $a = 1/4$. Write $z = a\cos y$ and $c = (2\pi)^6 C$. Equation (87) becomes

$$C = -\frac{z}{192(1+z)} - \frac{(\mathrm{d}^2/\mathrm{d}y^2)[z^2/(1+z)]}{128(1+z)} - \frac{z^3}{384(1+z)^3}.$$

At $y = \pi/2$, $z = 0$, $z' = -a$, $z'' = 0$, $z''' = a$. For $g(z) = z^2/(1+z)$, $g'(0) = 0$, $g''(0) = 2$, $g'''(0) = -6$. Consequently, if $H = (g(z))''$, then $H = 2a^2$ and $H' = 6a^3$ at that point, so $(H/(1+z))' = 8a^3$. Differentiating the three terms yields

$$C'(\pi/2) = \frac{a}{192} - \frac{a^3}{16} = \frac{a(1-12a^2)}{192} = \frac{1}{3072}, \qquad b(x_*) = \left(\frac{(2\pi)^7}{256}, 0, 0\right).$$

This coefficient is nonzero by an exact calculation.

We must justify numerical equivalence below the sufficient range of Theorem 4.1, since a counterexample for an exact solver alone does not prove one for an actual first hit. Fix $0 < \kappa \le 2/17$ and choose an integer $m \ge 4$ with

$$\kappa(m + 3/4) > 1/2. \tag{90}$$

This index is fixed as $n \to \infty$ and is used only in the proof. The same density (89) serves for every choice of the exponents; it does not depend on n or on m.

Because this density is smooth to every finite order, the population argument leading to (63) can be performed in C^{2m-2}: $d_1 = O_{C^{2m-2}}(\varepsilon)$, and then

$$d_j = O_{C^{2m-2j}}(\varepsilon^j), \quad 1 \le j \le m, \qquad \|d_m\|_\infty = O(\varepsilon^m).$$

The normalized-row proof of (67) applies through every $j < m$, because d_j still has C^2 norm $O(\varepsilon^j)$. The smooth-background comparison used in Section 4 extends through these finitely many steps: each sampled normalized row annihilates the constant $e^{d_j(x)}$, its diagonal self-term is zero, and conditioning

gives the same $O_{\mathbb{P}}(\varepsilon^{j+1/2}d_n)$ row error. Thus the empirical background at index m differs from d_m in L^2_n by $O_{\mathbb{P}}(\varepsilon^{3/2}d_n)$. The mixed background/stochastic trajectory error has size $\mathrm{poly}(j)\varepsilon d_n 2^{-j}$. The full second-order stochastic recurrence and all its kernel estimates are unchanged. Its normalized remainder coefficient $\sqrt{n}d_n^2\zeta = O(n^{-1}\varepsilon^{-9/4}(\log n)^3)$ tends to zero in the present range. The background score is now $O_{\mathbb{P}}(\sqrt{n}\varepsilon^{m+3/4}) = o_{\mathbb{P}}(1)$ by (90); the other background costs are $\sqrt{n}\varepsilon d_n$ and $\sqrt{n}\varepsilon^{3/2}d_n$, which also vanish. The same common-event argument therefore gives numerical equivalence uniformly after every deterministic diverging lower index. This fixed proof index never changes the actual initialization or omits a residual check.

For completeness the actual hit still diverges. The residual comparison (78) has $\sqrt{n}\varepsilon^m$ in place of $\sqrt{n}\varepsilon^4$; relative to its high spectral band the deterministic factor becomes $\sqrt{n}\varepsilon^{m+3/4} = o(1)$. All other relative factors are unchanged and vanish with polynomial margins. The proof of (82) and the upper trace bound therefore bracket the genuine hit at the same diverging logarithmic indices. At fixed m, $r_m/t_{n,\varepsilon} \to \infty$. The relative logs at all fixed indices $1, \dots, m-1$ are small, so (62) applied backwards excludes every early positive index. At index zero the density is nonconstant, r_0 has a positive limit, and $t_{n,\varepsilon} \to 0$ because $\kappa\gamma < 3/34 < 1/2$. All checks of the original rule are retained. The argument applies at both bandwidths on their common sample.

The population-centered joint exact-score limit in Theorem 3.11 applies for these powers. Set $\lambda_n = \sqrt{n}\varepsilon^{17/4}$. Combining that limit, the numerical equivalence just proved and (88), the actual stopped statistic is

$$Z_n + \lambda_n[b(x_*) + o(1)], \qquad Z_n \Rightarrow N_3(0, [4J_R/f(x_*)]I_3).$$

At $\kappa = 2/17$, $\lambda_n = 1$; below it $\lambda_n \to \infty$. The nonzero positive first component of b gives both assertions. □

This is a universal-class obstruction for this particular two-band correction and target centering. It is not a minimax lower bound or an impossibility for other corrections. A special density or query can have $b(x) = 0$ and admit a different bias regime.

6.2 The coefficient-one tolerance ceiling

Proposition 6.2. *For every $2/17 < \kappa < 4/9$ and every fixed $\gamma \geq 3/4$, the uniform density $f = 1$ makes both bandwidth runs stop at index zero with probability tending to one. Their corrected score has a different limiting covariance from* (53)*.*

Proof. Put $g_0 = H_\varepsilon(\mu_n - \mu)$, so $F_0 = (1 + g_0)^{1/2}$. Uniform KDE concentration gives $\|g_0\|_\infty = O_{\mathbb{P}}(\rho)$. The actual empirical residual is consequently

$$r_0 = \tfrac{1}{2}\left\|g_0 - \langle 1, g_0\rangle_n\right\|_{2,n} + O_{\mathbb{P}}(\rho\,\|g_0\|_{2,n}).$$

The empirical/population norm difference is $O_{\mathbb{P}}(\sqrt{\rho}\zeta)$ by the operator concentration bound, and $\langle 1, g_0\rangle_n = \|g_0\|^2_{\mathcal{H}} = O_{\mathbb{P}}(\zeta^2)$. In population L^2, Fourier orthogonality gives

$$\mathbb{E}\,\|g_0\|^2_{L^2} = \frac{1}{n}\sum_{k\neq 0} e^{-4\pi^2\varepsilon|k|^2} \sim \frac{1}{n}\varepsilon^{-3/2}(4\pi)^{-3/2}.$$

The fourth-moment expansion (79) gives relative variance $O(\varepsilon^{3/2} + 1/n)$ for this norm. Hence

$$\frac{r_0}{n^{-1/2}\varepsilon^{-3/4}} \longrightarrow \frac{1}{2}(4\pi)^{-3/4} < 1. \tag{91}$$

The same holds at 2ε. For $\gamma \geq 3/4$ the coefficient-one threshold is at least the denominator scale in (91), proving that both actual first hits equal zero with probability tending to one.

At iteration zero $s_{0,h} = 2\,\nabla \log K_{h,n}1$. For $f = 1$, its linearization is $2\,\nabla H_h(\mu_n - \mu)$. At a fixed query, $g_0(x) = O_{\mathbb{P}}(\zeta)$ and $\nabla g_0(x) = O_{\mathbb{P}}(n^{-1/2}h^{-5/4})$ by their scalar second moments. Since $\|g_0\|_\infty = o_{\mathbb{P}}(1)$,

the exact logarithm denominator gives a normalized nonlinear error $O_{\mathbb{P}}(\zeta) = o_{\mathbb{P}}(1)$ for $\kappa < 2/3$. The usual kernel triangular-array calculation therefore yields the corrected covariance

$$4 \int_{\mathbb{R}^3} (2\pi\xi)(2\pi\xi)^T [2\Lambda(\xi) - \Lambda(\xi)^2]^2 \, \mathrm{d}\xi, \qquad \Lambda(\xi) = e^{-2\pi^2|\xi|^2}.$$

The exact corrected factor is $2\Lambda/(1+\Lambda) - \Lambda^2/(1+\Lambda^2)$. Both factors are positive, and their difference is

$$\frac{2\Lambda^2}{1+\Lambda} - \frac{\Lambda^4}{1+\Lambda^2} > 0 \qquad (0 < \Lambda \le 1).$$

The derivative-weighted covariance difference is strictly positive definite, proving the claimed failure of covariance preservation. □

The endpoint assertion concerns the specified threshold coefficient and rule. It does not rule out different coefficients or other stopping rules. The upper exponent $4/9$ is a strict sufficient bound from the second-order exact-score remainder. All powers below $2/5$ in the theorem's rectangle yield shrinking uncertainty; formal powers above it do not.

7 Scope and open questions

The results concern a fixed positive smooth density, a fixed query and the exact periodic heat kernel. They give asymptotic score inference and iteration counts.

The strict upper exponent $4/9$ comes from the differentiated second-order scaling remainder. The full trajectory estimate pays that same remainder and removes the earlier coupling between tolerance and bandwidth. No upper endpoint optimality is needed for the stated result. In contrast, the lower obstruction comes from a fixed nonzero cubic bias coefficient for the two-scale correction. Further bias correction changes the estimator and requires new statistical and stopping arguments. The sharp residual constants do not decide the inequality at exactly $1/2$ when the continuous stopping phase approaches an integer on a shrinking scale.

Funding

This research received no external funding.

AI assistance

AI tools assisted mathematical development, literature discovery, drafting and checking. The author is responsible for the content.

References

[1] S. Schubr. Residual stopping and pointwise inference for nonuniform entropic self-transport. Luminary Research Archive, edition 1, 18 October 2026. https://luminaryoddities.com/researcharchive/residual-stopping-pointwise-inference-nonuniform-entropic-self-transport/v/1.

[2] S. Schubr. Residual certificates and empirical sharpness for entropic self-transport scores. Luminary Research Archive, edition 1, 18 October 2026. https://luminaryoddities.com/researcharchive/residual-certificates-empirical-sharpness-entropic-self-transport-scores/v/1.

[3] G. Mordant. The statistics of entropic optimal transport with decreasing regularisation. Preprint, arXiv:2412.12007v2, 29 April 2026. https://arxiv.org/abs/2412.12007v2.

[4] A. Kokot. *Variational Problems in Feature Selection and Data Compression*. Ph.D. dissertation, University of Washington, 2026. https://hdl.handle.net/1773/57893.

[5] C. L. Wormell and S. Reich. Spectral convergence of diffusion maps: improved error bounds and an alternative normalization. *SIAM Journal on Numerical Analysis* **59**(3) (2021), 1687–1734. doi:10.1137/20M1344093.

[6] X. Cheng and B. Landa. Bi-stochastically normalized graph Laplacian: convergence to manifold Laplacian and robustness to outlier noise. *Information and Inference: A Journal of the IMA* **13**(4) (2024), iaae026. doi:10.1093/imaiai/iaae026.

[7] I. Maeda, R. Yao and A. Nitanda. Statistical analysis of the Sinkhorn iterations for two-sample Schrödinger bridge estimation. Preprint, arXiv:2510.22560v1, 26 October 2025. https://arxiv.org/abs/2510.22560v1.

[8] N. Kessler, R. Magnet and J. Feydy. Sinkhorn normalization of diffusion kernels. In *Proceedings of the 43rd International Conference on Machine Learning*, Proceedings of Machine Learning Research **306** (2026), 56764–56795. https://proceedings.mlr.press/v306/kessler26a.html.

[9] J. Feydy, T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouvé and G. Peyré. Interpolating between optimal transport and MMD using Sinkhorn divergences. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research **89** (2019), 2681–2690. https://proceedings.mlr.press/v89/feydy19a.html.

[10] L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard and G. Peyré. Faster Wasserstein distance estimation with the Sinkhorn divergence. In *Advances in Neural Information Processing Systems* **33** (2020). https://proceedings.neurips.cc/paper/2020/hash/17f98ddf040204eda0af36a108cbdea4-Abstract.html.

[11] J. A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning* **8**(1–2) (2015), 1–230. doi:10.1561/2200000048.

[12] S. Minsker. On some extensions of Bernstein's inequality for self-adjoint operators. *Statistics & Probability Letters* **127** (2017), 111–119. doi:10.1016/j.spl.2017.03.020.

[13] 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**(1) (2024), 980–1041. doi:10.1214/24-EJS2217.

[14] R. Yao, Y.-H. Kim and G. Schiebinger. Quantitative stability of many-marginal Schrödinger bridge. Preprint, arXiv:2604.15191v1, 16 April 2026. https://arxiv.org/abs/2604.15191v1.

[15] E. Pauwels and S. Vaiter. The derivatives of Sinkhorn–Knopp converge. *SIAM Journal on Optimization* **33**(3) (2023), 1494–1517. doi:10.1137/22M1512703.

[16] X. Li, F. Lu, M. Tao and F. X.-F. Ye. Robust first and second-order differentiation for regularized optimal transport. *SIAM Journal on Scientific Computing* **47**(3) (2025), C630–C654. doi:10.1137/24M1674030.

[17] D. Scieur, Q. Bertrand, G. Gidel and F. Pedregosa. The curse of unrolling: rate of differentiating through optimization. Preprint, arXiv:2209.13271v2, 19 May 2023. https://arxiv.org/abs/2209.13271v2.

[18] L. Liu, R. Mehta, S. Pal and Z. Harchaoui. The benefits of balance: from information projections to variance reduction. *Advances in Neural Information Processing Systems* **37** (2024). Appendix version: arXiv:2408.15065v2, 11 February 2025. https://arxiv.org/abs/2408.15065v2.

[19] G. Blanchard, M. Hoffmann and M. Reiss. Optimal adaptation for early stopping in statistical inverse problems. *SIAM/ASA Journal on Uncertainty Quantification* **6**(3) (2018), 1043–1075. doi:10.1137/17M1154096.

[20] A. Celisse and M. Wahl. Analyzing the discrepancy principle for kernelized spectral filter learning algorithms. *Journal of Machine Learning Research* **22**(76) (2021), 1–59. https://jmlr.org/papers/v22/20-358.html.

[21] L. Hucker and M. Reiss. Early stopping for conjugate gradients in statistical inverse problems. *Numerische Mathematik* **157** (2025), 1739–1791. doi:10.1007/s00211-025-01469-4.

[22] G. Greco, M. Noble, G. Conforti and A. Oliviero-Durmus. Non-asymptotic convergence bounds for Sinkhorn iterates and their gradients: a coupling approach. In *Proceedings of the 36th Conference on Learning Theory*, Proceedings of Machine Learning Research **195** (2023), 716–746. https://proceedings.mlr.press/v195/greco23a.html.

[23] G. Greco and L. Tamanini. Hessian stability and convergence rates for entropic and Sinkhorn potentials via semiconcavity. *Bernoulli* **32**(3) (2026), 2351–2378. doi:10.3150/25-BEJ1956.

[24] D. Belomestny. Uniform statistical convergence of empirical Sinkhorn potentials with exponential and polynomial dependence on the regularization parameter. Preprint, arXiv:2608.29152v1, 29 August 2026. https://arxiv.org/abs/2608.29152v1.

[25] W. Gao, Z. Qu, Y. Ye and M. Udell. Tight nonasymptotic local convergence of Sinkhorn–Knopp. Preprint, arXiv:2608.11760v3, 28 September 2026. https://arxiv.org/abs/2608.11760v3.

[26] S. Schubr. Leading bias and positive correction of heat-kernel Sinkhorn metrics. Luminary Research Archive, edition 2, 2026. https://research.luminaryoddities.com/leading-bias-positive-correction-heat-kernel-sinkhorn-metrics/v/2.

[27] S. Schubr. Statistical inference for empirical heat-kernel Sinkhorn metrics. Luminary Research Archive, edition 3, 2026. https://research.luminaryoddities.com/statistical-inference-empirical-heat-kernel-sinkhorn-metrics/v/3.